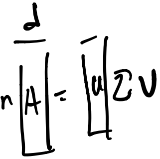
EE270
Large scale matrix computation, optimization and learning

Instructor : Mert Pilanci

Stanford University

Randomized Linear Algebra and Optimization
Lecture 11: Spectral Approximation, Subspace Embedding and Fast JL Transforms

Approximating Matrices



Approximate matrix product $A^T A \approx A^T S^T S A$

sampling based vs projection based methods

Let $A = U \Sigma V^T$ be the Singular Value Decomposition of A

- **Sampling based**
  - Uniform
  - Row norm scores $p_i = \frac{\|a_i\|_2^2}{\sum_j \|a_j\|_2^2}$
  - Leverage scores $p_i = \frac{\|u_i\|_2^2}{\sum_j \|u_j\|_2^2}$ $\equiv$ row norm scores of U

Approximating Matrices

Approximate matrix product $A^T A \approx A^T S^T S A$

sampling based vs projection based methods

Let $A = U \Sigma V^T$ be the Singular Value Decomposition of A

- **Sampling based**
  - Uniform
  - Row norm scores $p_i = \frac{\|a_i\|_2^2}{\sum_j \|a_j\|_2^2}$
  - Leverage scores $p_i = \frac{\|u_i\|_2^2}{\sum_j \|u_j\|_2^2}$
- **Projection based**
  - Gaussian $N(0,1)$ random projection
  - Rademacher ± 1 random projection
  - Haar (uniform orthogonal) random projection
  - Sparse Johnson Lindenstrauss (CountSketch) Embeddings 
  - Fast Johnson Lindenstrauss (Randomized Hadamard) Transform we'll see more today

Leverage Scores

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A implies Least Squares cost approximation
- Importance sampling: proportional to the rows norms of U
- Leverage scores: $\ell_i := \|u_i\|_2^2$ for $i = 1, \ldots n$
- $\sum_i \ell_i = \sum_i \|u_i\|_2^2 = \|U\|_F^2 = trU^TU = trI_d = d$ when A is full column rank
- Sampling probabilities: $p_i = \frac{1}{d}\|u_i\|_2^2$ $\sum_i p_i = 1$

$$\begin{array}{cc} & d \\ \begin{array}{c} d \\ n-d \end{array} & \begin{bmatrix} I \\ 0 \end{bmatrix} \end{array} \qquad \begin{bmatrix} I \\ I \\ I \\ I \end{bmatrix}$$

- Can be non-uniform or uniform $A = [I; 0]$
- Approximate Matrix Multiplication for U^TU i.e, $\|U^TS^TSU - U^TU\|_F = \|U^TS^TSU - I\|_F \leq \epsilon$

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for $U^T U$ i.e,

$$\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon \tag{1}$$

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for $U^T U$ i.e,

$$\|U^T S^T SU - U^T U\|_F = \|U^T S^T SU - I\|_F \le \epsilon \qquad (1)$$

$$\|M\|_2 = \sigma_{max} = \max_i \sigma_i \le \|M\|_F = \sqrt{\sum \sigma_i^2}$$

$\uparrow$ spectral norm

- (1) implies $\sigma_{max}\left(U^T S^T SU - I\right) \le \epsilon$

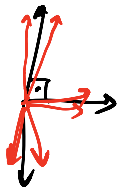
  Singular values of a symmetric matrix are the absolute values of the eigenvalues

$$\|M\|_2 = \max_{\|x\|_2=1} \|Mx\|_2 = \sigma_{max}$$

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for U^TU i.e,

$$\|U^TS^TSU - U^TU\|_F = \|U^TS^TSU - I\|_F \leq \epsilon \tag{1}$$



- (1) implies $\sigma_{max}\left(U^TS^TSU - I\right) \leq \epsilon$

  Singular values of a symmetric matrix are the absolute values of the eigenvalues
- $\max_{i=1,..d}\left|\lambda_i\left(U^TS^TSU - I\right)\right| \leq \epsilon$ $\Leftrightarrow$ $|\lambda_i(U^TS^TSU) - 1| \leq \epsilon$
- (1) implies $1 - \epsilon \leq \lambda_i\left(U^TS^TSU\right) \leq 1 + \epsilon$ for all i

$\lambda_i(M - I) = \lambda_i(M) - 1$

$\|$

$1 - \epsilon \leq \sigma_i(SU) \leq 1 + \epsilon$

left sketching
preconditioning
randomized QR (SVD)

$\sigma_i(U) = 1 \quad \forall i$

$U^TS^TSU \neq I$

Interpretation of Leverage Scores: Spectral Approximation

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of A
- S be the leverage score sampling matrix
- Approximate Matrix Multiplication for $U^T U$ i.e,

$$\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \leq \epsilon \qquad (1)$$

- (1) implies $\sigma_{max}\left(U^T S^T S U - I\right) \leq \epsilon$

  Singular values of a symmetric matrix are the absolute values of the eigenvalues
- $\max_{i=1,..d} \left|\lambda_i\left(U^T S^T S U - I\right)\right| \leq \epsilon$ $\Rightarrow$ $U^T S^T S U$ is invertible when $\epsilon < 1$
- (1) implies $1 - \epsilon \leq \lambda_i\left(U^T S^T S U\right) \leq 1 + \epsilon$ for all i
- $(A^T S^T S A)^{-1}$ exists whenever $(A^T A)^{-1}$ exists

  $A = U\Sigma V^T$

  $(A^T S^T S A)^{-1} = V \Sigma^{-1} (U^T S^T S U)^{-1} \Sigma^{-1} V^T$

  Σ is invertible
- sketched least squares solution $\arg\min_x \|SAx - Sb\|_2 = (A^T S^T S A)^{-1} S^T S b$ is well defined

Preserving Spectral Properties

$A = U \Sigma V^T$

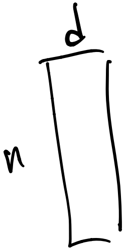
$$\max_x |x^T(U^T S^T S U - I)x| \le \epsilon \cdot \|x\|_2^2 \iff \big| \|SUx\|_2^2 - \|x\|_2^2 \big| \le \epsilon \cdot \|x\|_2^2$$

$$x = \Sigma V^T \cdot x' \Rightarrow \|x\|_2^2 = \|Ax'\|_2^2 = x'^T V \Sigma \underbrace{U^T U}_{I} \Sigma V^T x'$$

$$= \|\Sigma V^T x'\|_2^2$$

$$\big| \|SAx'\|_2^2 - \|Ax'\|_2^2 \big| \le \epsilon \|Ax'\|_2^2$$

$$\|U^T S^T S U - I\|_2 \le \|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \le \epsilon \qquad (2)$$



- also implies that

$$(1 - \epsilon)\|Ax\|_2^2 \le \|SAx\|_2^2 \le (1 + \epsilon)\|Ax\|_2^2$$

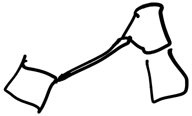
for all $x \in \mathbb{R}^d$ (not just for finitely many points)

Johnson-Lindenstrauss embedding property for the whole subspace range(A)

- we utilized this in the basic inequality method

$$\max \left| \frac{\|SAx\|_2^2}{\|Ax\|_2^2} - 1 \right| \le \epsilon$$

Interpretation of Leverage Scores: Subspace Embedding



row norm sampling is optimal min $\mathbb{E} \|A^T S^T S A - A^T A\|_F^2$

$\mathbb{E}\, A^T S^T S A = A^T A$ ✓

$$\|U^T S^T S U - U^T U\|_F = \|U^T S^T S U - I\|_F \le \epsilon$$

implies

$$(1-\epsilon)\|Ax\|_2^2 \le \|SAx\|_2^2 \le (1+\epsilon)\|Ax\|_2^2$$

for all $x \in \mathbb{R}^d$

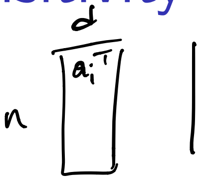
- Weyl's Inquality $\left|\lambda_i(M) - \lambda_i(M')\right| \le \sigma_{\max}\left(M - M'\right)$ for all i
- $\left|\lambda_i(A^T S^T S A) - \lambda_i(A^T A)\right| \le \epsilon$, i.e., all eigenvalues are approximately preserved in leverage score sampling method

not true for uniform / row-norm sampling

Interpretation of $\boxed{\text{Leverage Scores:}}$ Sensitivity of the loss function

$$a_k^T = e_k^T A$$

$\hat{}$ k'th ord. basis beetor



- Consider $\|Ax - b\|_2^2 = \sum_i (a_i^T x - b_i)^2$ suppose that $\boxed{b = Ax^*}$ for simplicity (planted model assumption)
- Consider the worst-case ratio

$$\max_x \frac{(a_k^T x - b_k)^2}{\sum_i (a_i^T x - b_i)^2} = \max_x \frac{(a_k^T x - b_k)^2}{\|Ax - b\|_2} = \max_x \frac{(a_k^T \cdot (x - x^*))^2}{\|A(x - x^*)\|_2^2}$$

$x' \leftarrow x + x^*$

$$\max_{x'} \frac{(a_k^T \cdot x')^2}{\|Ax'\|_2^2} = \frac{(e_k^T A \cdot x')^2}{\|Ax'\|_2^2} = \frac{(e_k^T \cdot U x'')^2}{\|x''\|_2^2} = \frac{(u_k^T \cdot x'')^2}{\|x''\|_2} \overset{\text{C.S.}}{\leq} \frac{\|u_k\|_2^2 \|x''\|_2^2}{\|x''\|_2^2}$$

$$A = U \Sigma V^T$$
$$x'' = \Sigma V^T x$$

$$= \|u_k\|_2^2$$

Fast Johnson Lindenstrauss Transform

$n \times n$ $\square$ $\|n$ $O(n^2)$

- Let H denote the $n \times n$ Hadamard Transform matrix constructed as follows Fourier Transform

log n steps $\Rightarrow O(n \log n)$

H_k is $2^{k-1} \times 2^{k-1}$

$$H_2 := \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

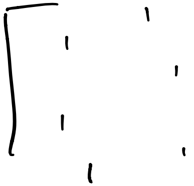
$$H_{n+1} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}$$

$$H_{n+1} \cdot x = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} H_n(x_1+x_2) \\ H_n(x_1-x_2) \end{bmatrix}$$

- let D be an $n \times n$ diagonal matrix of random ± 1 uniform signs
- Uniform $m \times n$ sub-sampling matrix P scaled with $\frac{\sqrt{n}}{\sqrt{m}}$
- Let $S = \frac{1}{\sqrt{n}} PHD. =$
- Note that $\mathbb{E}S^T S = I$ since $DH^T HD = nI$ and $\mathbb{E}P^T P = I$

$$\frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

HD random



Sx = uniformly subsample { Hadamard transform of D·x }

$SA = P \cdot \overbrace{H D A}$

Fast Johnson Lindenstrauss Transform Analysis

- Leverage scores of a matrix $A = U\Sigma V^T$ are given by $\ell_i = \|U^T e_i\|_2^2 = e_i^T U U^T e_i$ $= e_i^T P_A \cdot e_i$ where $P_A = UU^T$ projection to the range of A.
- Another expression: $\ell_i = e_i^T \underbrace{A(A^T A)^{-1} A^T}_{} e_i$

$$\underbrace{U \cancel{\Sigma} \cancel{V^T V} \cancel{\Sigma^{-1}} \cancel{U^T U} \cancel{\Sigma^{-1}} \cancel{V} \cdot \cancel{V} \cancel{\Sigma} U^T}_{UU^T}$$

Fast Johnson Lindenstrauss Transform Analysis

$$\mathcal{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

$$D \cdot h_2^{(d)} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

I

D

$\boxed{H^T H = I \cdot n}$

$\boxed{\frac{1}{n} H^T H = I}$

- Leverage scores of a matrix $A = U\Sigma V^T$ are given by $\ell_i = \|U^T e_i\|_2^2 = e_i^T UU^T e_i$ $e_i^T DA(A^T \cancel{D}\cancel{D}A)^{-1} A^T \underline{De_i} = \underbrace{D_{ii}^2}_{1} \cdot l_i$
- Another expression: $\ell_i = e_i^T A(A^T A)^{-1} A^T e_i$
- Compare with leverage scores of $\underbrace{\cancel{\frac{1}{\sqrt{n}}} HDA}_{\text{leverage score uniformizer}}$ denoted by $\tilde{\ell}_i$

$$\tilde{\ell}_i := e_i^T HDA(A^T D\underbrace{H^T H}_{I}DA)^{-1} A^T DH^T e_i \tag{3}$$

$$= \frac{1}{n} e_i^T HDA(A^T A)^{-1} A^T DH^T e_i \tag{4}$$

$$= \frac{1}{n} e_i^T HDUU^T DH^T e_i \tag{5}$$

$$= \frac{1}{n} h_i^T DUU^T Dh_i \tag{6}$$

h_i is a row of $\mathcal{H}$

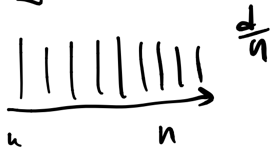
$= \frac{1}{n} \|U^T D h_i\|_2^2$

$D \cdot h_i$ is random ± 1

n l. scores

$\approx \frac{d}{n}$

$\sum l_i = d$



$\overset{2^k}{H} \begin{bmatrix} A \\ 0 \end{bmatrix}$

$\left\| \begin{bmatrix} A \\ 0 \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2$

$$\begin{bmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & +1 \end{bmatrix}$$

- where we have used $H^T H = nI$
- $\tilde{\ell}_i$ is distributed as $\frac{1}{n} \underbrace{r^T UU^T r}_{\text{trace estimation!}}$ where r is i.i.d. ± 1
- $\mathbb{E} \frac{1}{n} r^T UU^T r = \frac{d}{n}$

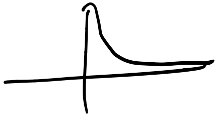
$\text{tr}\, UU^T = \text{tr}\, U^T U = \text{tr}\, I_d = d$

Fast Johnson Lindenstrauss Transform Analysis

$$\|U^T D h_i\|_2^2 = \sum (u_j^T D h_i)^2$$

- Chernoff's method (as in Chernoff Bound) implies that

$$\longrightarrow \quad \mathbb{P}\left[\left|\frac{1}{n}\underbrace{h_i^T D u_j}_{\text{sum of iid variables}}\right| \geq t\right] \leq \underline{2e^{-t^2 n/2}}$$



for every fixed i and j.

- Applying union bound

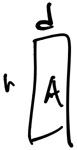
$$\tilde{\ell}_i = \frac{1}{n} h_i^T D U U^T D h_i \leq \text{const} \, \frac{d \log(nd)}{n}$$

with high probability

note that $\ell_i = \frac{d}{n}$ for all i when leverage scores are exactly uniform

Randomized Hadamard Transform *HD* preconditions leverage scores

$S = PHD$ Fast JL Transform



$\min_x \|S(Ax-b)\|_2^2$

ϵ- approx : runtime: $SA : n d \log n$

$(SA)^\dagger Sb : m \cdot d^2$

$A^\dagger b : nd^2$

Gaussian Sketch

$SA : m \cdot n \cdot d$

Apply *HD* to data *A*

- *PHDA* is a uniformly subsampled version *HDA*

Leverage scores of $\frac{1}{\sqrt{n}} HDA$ are near uniform

leverage score sampling = uniform sampling $\frac{1}{\sqrt{n}} HDA$ works!

in other words *SA* where $S = \frac{1}{\sqrt{n}} PHD$ is a subspace embedding

$$(1-\epsilon) \cdot \|Ax\|_2^2 \le \|SAx\|_2^2 \le \|Ax\|_2^2 \cdot (1+\epsilon)$$

$$\Updownarrow$$

$$(1-\epsilon) \cdot \|x\|_2^2 \le \|SUx\|_2^2 \le \|x\|_2^2 \cdot (1+\epsilon)$$

Questions?