COMPSCI 614: Randomized Algorithms with Applications to Data Science

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2024.

Lecture 16

Logistics

- Problem Set 4 was released on Friday – it is due 4/22.
- Project progress report due on 4/16.
- No class on next Tuesday
- Zoom next Thursday

Summary

Last Week: Subspace embedding via random sketching.

- Finish proof of subspace embedding from the distributional Johnson-Lindenstrauss lemma and an ϵ-net argument.
- Proof of distributional JL via the Hanson-Wright inequality.
- Application to fast over-constrained linear regression.

SAx

Summary

Last Week: Subspace embedding via random sketching.

- Finish proof of subspace embedding from the distributional Johnson-Lindenstrauss lemma and an ϵ-net argument.
- Proof of distributional JL via the Hanson-Wright inequality.
- Application to fast over-constrained linear regression.

$$\begin{bmatrix} \pm1 & S & \pm1 \end{bmatrix} \begin{bmatrix} A \end{bmatrix}$$

Today:

- Subspace embedding via sampling.
- The matrix leverage scores.
- Analysis via matrix concentration bounds.
- ~~Spectral graph sparsifiers~~.

Quiz Review

Question 3

Not complete

Points out of 1.00

Flag question

Edit question

Assume that $S \in \mathbb{R}^{m \times n}$ is an ϵ-subspace embedding for $A \in \mathbb{R}^{n \times d}$. Do the following guarantees hold, for some small constant c (e.g., c = 1, or c = 2, etc.)?

(1) $(1 - c\epsilon)\|A\|_F \leq \|SA\|_F \leq (1 + c\epsilon)\|A\|_F$

and

(2) $(1 - c\epsilon)\|A\|_2 \leq \|SA\|_2 \leq (1 + c\epsilon)\|A\|_2$.

Recall that the spectral norm of a matrix is defined $\|M\|_2 = \max_{x:\|x\|_2=1} \|Mx\|_2$.

Hint: Try to prove these bounds using the guarantee that $\|SAx\| \approx_\epsilon \|Ax\|_2$ for all $x \in \mathbb{R}^d$.

$\|A\|_F^2 = \sum \sigma_i(A)^2$

$\approx \sum \sigma_i(SA)^2$

$= \|SA\|_F^2$

$\|SA\|_2 \leq (1+c\epsilon)\|A\| \ ?$

$x^* = \operatorname{argmax}_{x:\|x\|=1} \|Ax\|$

$[S][A] = [SA]$

$\|SA\|_2 \geq \|SAx^*\| \geq (1-\epsilon)\|Ax^*\|$

$\geq (1-\epsilon)\|A\|_2$

- a. Yes, both always hold.
- b. (1) always holds but (2) may not.
- c. (2) always holds but (1)) may not.
- d. Neither is guaranteed to always hold.

Check

$\|SA\|_F^2 = \sum_{i=1}^{d} \|SA_{:,i}\|_2^2 \overset{?}{\approx} \sum_{i=1}^{d} \|A_{:,i}\|_2^2 \leq \sum_{i=1}^{d} (1+\epsilon)\|SA_{:,i}\|_2^2$

$\|A\|_F^2 \leq (1+\epsilon)\|SA\|_F^2 \leq (1+\epsilon)\sum\|SA_{:,i}\|_2^2$

Quiz Review

Question 5

Not complete

Points out of 1.00

Flag question

Edit question

Which of the following concentration bounds can be apply to show that, for a random $x \in \mathbb{R}^n$ with i.i.d. ± 1 entries, and some fixed $A \in \mathbb{R}^{n \times n}$, that $x^T A x$ is concentrated around its mean? Select all that apply.

- [ ] a. ~~Markov bound~~
- [ ] b. Bernstein bound — boundedness, not independent
- [ ] c. Chebyshev inequality
- [ ] d. Hanson-Wright Inequality

Check

$\|x\|_\infty \leq 1$

$\|A\|_\infty$

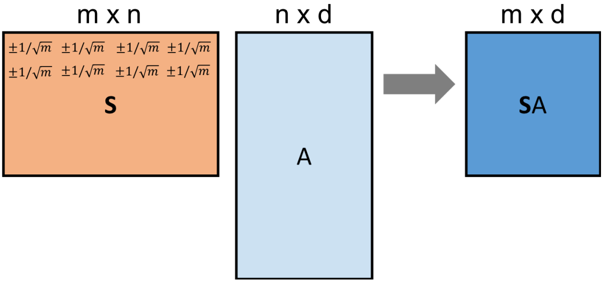
$$\sum_i \sum_j x(i)\, x(j)\, A_{ij}$$

Subspace Embedding

$\mathbf{S} \in \mathbb{R}^{m\times n}$ is an ϵ-subspace embedding for $A \in \mathbb{R}^{n\times d}$, if for all $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|Ax\| \le \|\mathbf{S}Ax\|_2 \le (1+\epsilon)\|Ax\|_2.$$

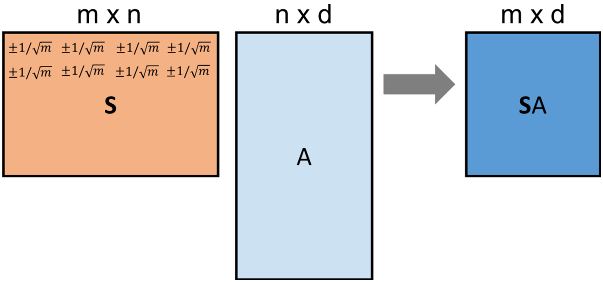
$m = \tilde{O}\left(\frac{d}{\varepsilon^2}\right)$



Subspace Embedding

$\mathbf{S} \in \mathbb{R}^{m \times n}$ is an ϵ-subspace embedding for $A \in \mathbb{R}^{n \times d}$, if for all $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|Ax\| \leq \|\mathbf{S}Ax\|_2 \leq (1+\epsilon)\|Ax\|_2.$$

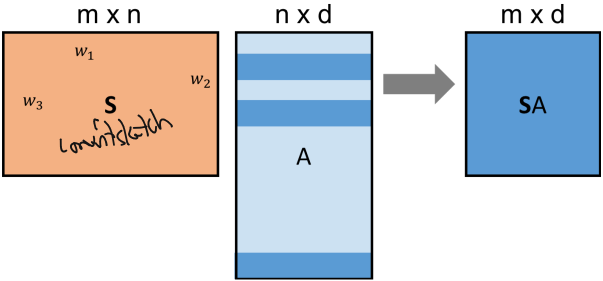


So Far: If $\mathbf{S}$ is a random sign matrix, and $m = O\left(\frac{d+\log(1/\delta)}{\epsilon^2}\right)$, then for any A, $\mathbf{S}$ is an ϵ-subspace embedding with probability $\geq 1-\delta$.

Subspace Embedding

$\mathbf{S} \in \mathbb{R}^{m \times n}$ is an ϵ-subspace embedding for $A \in \mathbb{R}^{n \times d}$, if for all $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|Ax\| \leq \|\mathbf{S}Ax\|_2 \leq (1+\epsilon)\|Ax\|_2.$$



So Far: If $\mathbf{S}$ is a random sign matrix, and $m = O\left(\frac{d+\log(1/\delta)}{\epsilon^2}\right)$, then for any A, $\mathbf{S}$ is an ϵ-subspace embedding with probability $\geq 1-\delta$.

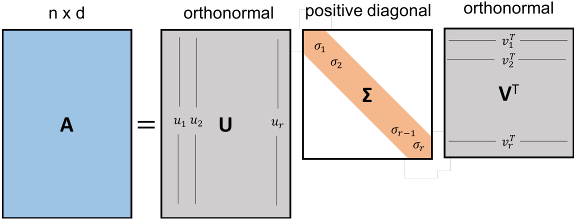
In many applications it is preferable for $\mathbf{S}$ to be a row sampling matrix. The sample can preserve sparsity, structure, etc.

more interpretable, active regression

Problem Reformulation

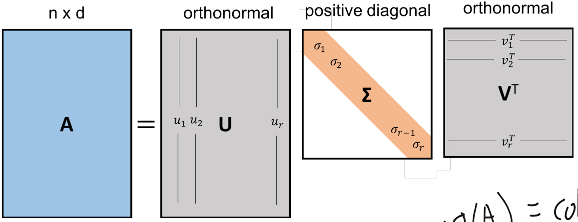
For $A \in \mathbb{R}^{n \times d}$, let $A = U\Sigma V^T$ be its SVD. $U \in \mathbb{R}^{n \times \text{rank}(A)}$, $V \in \mathbb{R}^{d \times \text{rank}(A)}$ are orthonormal, and $\Sigma \in \mathbb{R}^{\text{rank}(A) \times \text{rank}(A)}$ is positive diagonal.)

qr, orthogonal basis for A's column



Problem Reformulation

For $A \in \mathbb{R}^{n \times d}$, let $A = U\Sigma V^T$ be its SVD. $U \in \mathbb{R}^{n \times \text{rank}(A)}$, $V \in \mathbb{R}^{d \times \text{rank}(A)}$ are orthonormal, and $\Sigma \in \mathbb{R}^{\text{rank}(A) \times \text{rank}(A)}$ is positive diagonal.)



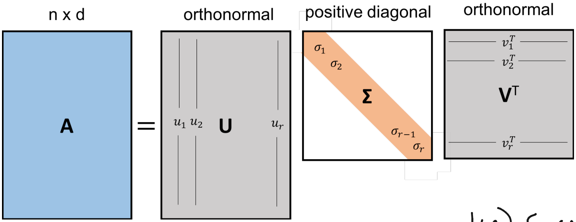
$\text{col}(A) = \text{col}(U)$

$d \times 1$ $\quad$ $d \times d$ $\quad$ $d \times 1$

- For any $x \in \mathbb{R}^d$, let $z = \Sigma V^T x$. Observe that: $\|Ax\|_2 = \|Uz\|_2$ and $\|SAx\|_2 = \|SUz\|_2$.

Problem Reformulation

For $A \in \mathbb{R}^{n \times d}$, let $A = U\Sigma V^T$ be its SVD. $U \in \mathbb{R}^{n \times \text{rank}(A)}$, $V \in \mathbb{R}^{d \times \text{rank}(A)}$ are orthonormal, and $\Sigma \in \mathbb{R}^{\text{rank}(A) \times \text{rank}(A)}$ is positive diagonal.)

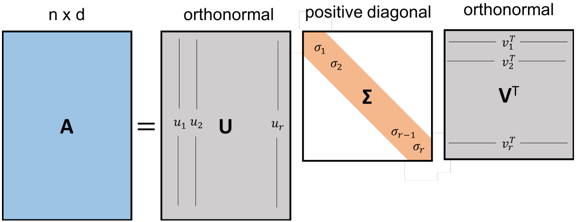


$\text{col}(A) \subseteq \text{col}(U)$
$\text{col}(U) \subseteq \text{col}(A)$ $\Rightarrow \text{col}(A) = \text{col}(U)$

- For any $x \in \mathbb{R}^d$, let $z = \Sigma V^T x$. Observe that: $\|Ax\|_2 = \|Uz\|_2$ and $\|\mathbf{S}A\|_2 = \|\mathbf{S}Uz\|_2$. $\quad Ux \to Az \quad z = V\Sigma^{-1}x$
- Thus, to prove that S is an ϵ-subspace embedding for A, it suffices to show that it is an ϵ-subspace embedding for U.

Problem Reformulation

For $A \in \mathbb{R}^{n \times d}$, let $A = U\Sigma V^T$ be its SVD. $U \in \mathbb{R}^{n \times \text{rank}(A)}$, $V \in \mathbb{R}^{d \times \text{rank}(A)}$ are orthonormal, and $\Sigma \in \mathbb{R}^{\text{rank}(A) \times \text{rank}(A)}$ is positive diagonal.)



- For any $x \in \mathbb{R}^d$, let $z = \Sigma V^T x$. Observe that: $\|Ax\|_2 = \|Uz\|_2$ and $\|\mathbf{S}A\|_2 = \|\mathbf{S}Uz\|_2$.
- Thus, to prove that S is an ϵ-subspace embedding for A, it suffices to show that it is an ϵ-subspace embedding for U.
- I.e., it suffices to show that for any $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|Ux\|_2^2 \le \|\mathbf{S}Ux\|_2^2 \le (1+\epsilon)\|Ux\|_2^2.$$

$\|Ux\|_2^2 = \|x\|_2^2$

because U is orthonormal

Loewner Ordering

Suffices to show that for any $x \in \mathbb{R}^d$,

$\|Ux\|_2^2$

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

Loewner Ordering

$\|y\|_2^2 = y^T y$

Suffices to show that for any $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|x\|_2^2 \leq \|\mathsf{S}Ux\|_2^2 \leq (1+\epsilon)\|x\|_2^2 \implies (1-\epsilon)x^T I x \leq x^T U^T \mathsf{S}^T \mathsf{S} U x \leq (1+\epsilon)x^T I x.$$

Loewner Ordering

Suffices to show that for any $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|x\|_2^2 \leq \|\mathbf{S}Ux\|_2^2 \leq (1+\epsilon)\|x\|_2^2 \implies (1-\epsilon)x^T I x \leq x^T U^T \mathbf{S}^T \mathbf{S} U x \leq (1+\epsilon)x^T I x.$$

This condition is typically denoted by $(1-\epsilon)I \preceq U^T \mathbf{S}^T \mathbf{S} U \preceq (1+\epsilon)I$.

$$M \preceq N \text{ iff } \forall x \in \mathbb{R}^d \; x^T M x \leq x^T N x \quad \text{(Loewner Order)}$$

$AS^TSB \simeq AB$ even though $S^TS \neq I$

Suffices to show that for any $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|x\|_2^2 \leq \|\mathbf{S}Ux\|_2^2 \leq (1+\epsilon)\|x\|_2^2 \implies (1-\epsilon)x^T I x \leq x^T U^T \mathbf{S}^T \mathbf{S} U x \leq (1+\epsilon)x^T I x.$$

This condition is typically denoted by $(1-\epsilon)I \preceq \underline{U^T\mathbf{S}^T\mathbf{S}U} \preceq (1+\epsilon)I$.

$$M \preceq N \text{ iff } \forall x \in \mathbb{R}^d \ x^T M x \leq x^T N x \quad \text{(Loewner Order)}$$

When $(1-\epsilon)N \preceq M \preceq (1+\epsilon)N$, I will write $M \approx_\epsilon N$ as shorthand.

$U^TS^TSU \approx_\epsilon I$

$\|U^TS^TSU - I\|_F \leq \frac{\epsilon}{d^2}$ (AMM)

$U^TU \stackrel{?}{=} I$

$\cancel{S^TS \approx I}$

$$\begin{array}{cc} & m \\ n & \begin{bmatrix} S^T \end{bmatrix} \end{array} \begin{array}{cc} & n \\ m & \begin{bmatrix} S \end{bmatrix} \end{array} = \begin{array}{cc} & n \\ n & \begin{bmatrix} \ddots \end{bmatrix} \end{array}$$

Loewner Ordering

Suffices to show that for any $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|x\|_2^2 \le \|\mathbf{S}Ux\|_2^2 \le (1+\epsilon)\|x\|_2^2 \implies (1-\epsilon)x^T I x \le x^T U^T \mathbf{S}^T \mathbf{S} U x \le (1+\epsilon) x^T I x.$$

This condition is typically denoted by $(1-\epsilon)I \preceq U^T\mathbf{S}^T\mathbf{S}U \preceq (1+\epsilon)I$.

$$M \preceq N \text{ iff } \forall x \in \mathbb{R}^d \; x^T M x \le x^T N x \quad \text{(Loewner Order)}$$

When $(1-\epsilon)N \preceq M \preceq (1+\epsilon)N$, I will write $M \approx_\epsilon N$ as shorthand.

$(1-\epsilon)I \preceq U^T\mathbf{S}^T\mathbf{S}U \preceq (1+\epsilon)I$ is equivilant to all eigenvalues of $U^T\mathbf{S}^T\mathbf{S}U$ lying in $[1-\epsilon, 1+\epsilon]$.

$\lambda_{min}(U^T S^T S U) = \min x^T U^T S^T S U x$

$\ge 1-\epsilon$

$x^T U^T S^T S U x \simeq 1$

Sampling from U

So Far: We have an orthonormal matrix $U \in \mathbb{R}^{n\times d}$ and we want to sample rows so that $U^T \mathbf{S}^T \mathbf{S} U \approx_\epsilon I$.

Sampling from U

So Far: We have an orthonormal matrix $U \in \mathbb{R}^{n \times d}$ and we want to sample rows so that $U^T \mathbf{S}^T \mathbf{S} U \approx_\epsilon I$. What are some possible sampling strategies?

n rows to sample

sample each w.p. p_i repeat trials

$p_i = \|U_{i,:}\|_2^2$

$p_i = \frac{1}{n}$

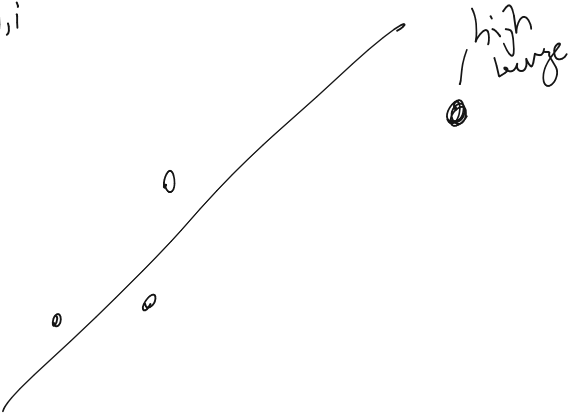
$$\begin{array}{cc} & d \\ d & \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \\ \hline & U & & \\ & 0 & & \end{bmatrix} \end{array} \qquad \begin{bmatrix} A \\ \hline 0 \end{bmatrix}$$

$U^T S^T S U = 0$ usually

- $\tau_i = \|U_{i,:}\|_2^2$ is known as the i^{th} leverage score of U.

$(UU^T)_{i,i} = [P_{col(A)}]_{i,i}$

$[U][U^T]$



Leverage Score Sampling

- $\tau_i = \|U_{i,:}\|_2^2$ is known as the i^{th} leverage score of U.
- Let $p_i = \frac{\tau_i}{\sum_{i=1}^n \tau_i}$. $\rightarrow$ $\sum \|U_{i,:}\|_2^2 = \|U\|_F^2 = \text{rank}(A) = d$
- Let $\mathbf{S}_{:,j} = e_i^T \cdot \frac{1}{\sqrt{mp_i}}$ with probability p_i. $\rightarrow$ sample row i w.p. p_i, reweight by $\frac{1}{\sqrt{mp_i}}$

$$\begin{bmatrix} S \quad \frac{1}{\sqrt{mp_i}} \end{bmatrix} \begin{bmatrix} U \end{bmatrix}$$

Leverage Score Sampling

- $\tau_i = \|U_{i,:}\|_2^2$ is known as the i^{th} leverage score of U.
- Let $p_i = \frac{\tau_i}{\sum_{i=1}^n \tau_i}$.
- Let $\mathbf{S}_{:,j} = e_i^T \cdot \frac{1}{\sqrt{mp_i}}$ with probability p_i.
- repeat m times independently

$$\mathbb{E}[U^T\mathbf{S}^T\mathbf{S}U] =$$

Leverage Score Sampling

- $\tau_i = \|U_{i,:}\|_2^2$ is known as the i^{th} leverage score of U.
- Let $p_i = \frac{\tau_i}{\sum_{i=1}^n \tau_i}$.
- Let $\mathbf{S}_{j,:} = e_i^T \cdot \frac{1}{\sqrt{mp_i}}$ with probability p_i.

jth row of S

1×d

jth row

(indices we swap)

$$\mathbb{E}[U^T\mathbf{S}^T\mathbf{S}U] = \; = \sum_{j=1}^{m} \mathbb{E}[U^T \underbrace{\mathbf{S}_{:,j}^T}_{d\times 1} \underbrace{\mathbf{S}_{:,j}}_{1\times d} U]$$

Leverage Score Sampling

- $\tau_i = \|U_{i,:}\|_2^2$ is known as the i^{th} leverage score of U.
- Let $p_i = \frac{\tau_i}{\sum_{i=1}^{n} \tau_i}$.
- Let $\mathbf{S}_{:,j} = e_i^T \cdot \frac{1}{\sqrt{mp_i}}$ with probability p_i.

$$\mathbb{E}[U^T \mathbf{S}^T \mathbf{S} U] = \; = \sum_{j=1}^{m} \mathbb{E}[U^T \mathbf{S}_{:,j}^T \mathbf{S}_{:,j} U]$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} \cancel{p_i} \cdot \left(\frac{1}{\sqrt{m\cancel{p_i}}} U_{i,:}^T\right)\left(\frac{1}{\sqrt{m\cancel{p_i}}} U_{i,:}\right)$$

Leverage Score Sampling

- $\tau_i = \|U_{i,:}\|_2^2$ is known as the i^{th} leverage score of U.
- Let $p_i = \frac{\tau_i}{\sum_{i=1}^n \tau_i}$.
- Let $\mathbf{S}_{:,j} = e_i^T \cdot \frac{1}{\sqrt{mp_i}}$ with probability p_i.

$$\mathbb{E}[U^T\mathbf{S}^T\mathbf{S}U] = \; = \sum_{j=1}^m \mathbb{E}[U^T\mathbf{S}_{:,j}^T\mathbf{S}_{:,j}U]$$

$$\sum_{i=1}^n U_{i,:}^T U_{i,:} = U^T U$$

$$= \sum_{j=1}^m \sum_{i=1}^n p_i \cdot (\frac{1}{\sqrt{mp_i}} U_{i,:}^T)(\frac{1}{\sqrt{mp_i}} U_{i,:})$$

$$= \sum_{j=1}^m \frac{1}{m} U^T U = I.$$

Matrix Concentration

AMM: looked at entries of $U^T S^T S U$

We want to show that $U^T \mathbf{S}^T \mathbf{S} U$ is close to $\mathbb{E}[U^T \mathbf{S}^T \mathbf{S} U] = I$. Will apply a matrix concentration bound.

Theorem (Matrix Chernoff Bound)

Consider independent symmetric random matrices $\mathbf{X}_1, \ldots, \mathbf{X}_m \in \mathbb{R}^{d \times d}$, with $\mathbf{X}_i \succeq 0$, $\lambda_{\max}(\mathbf{X}_i) \leq R$, and $\mathbf{X} = \sum_{i=1}^m \mathbf{X}_i$. Let $M = \mathbb{E}[\mathbf{X}]$. Then:

$\frac{1}{mp_i} U_{i,:}^T U_{i,:}$

$d \times d$, symmetric, PSD, rank-1

$$\Pr\left[\lambda_{\min}(\mathbf{X}) \leq (1-\epsilon)\lambda_{\min}(M)\right] \leq d \cdot \left[\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right]^{\lambda_{\min}(M)/R}$$

$$\Pr\left[\lambda_{\max}(\mathbf{X}) \geq (1+\epsilon)\lambda_{\max}(M)\right] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\lambda_{\min}(M)/R}$$

(typo?)

$\lambda(X) \in [1-\varepsilon, 1+\varepsilon]$

$$(1-\varepsilon)\underset{I}{M} \preceq X \preceq (1+\varepsilon)\underset{I}{M}$$

Matrix Concentration Applied to Leverage Score Sampling

Theorem (Matrix Chernoff Bound)

Consider independent symmetric random matrices $\mathsf{X}_1, \ldots, \mathsf{X}_m \in \mathbb{R}^{d \times d}$, *with* $\mathsf{X}_i \succeq 0$, $\lambda_{\max}(\mathsf{X}_i) \leq R$, *and* $\mathsf{X} = \sum_{i=1}^m \mathsf{X}_i$. *Let* $M = \mathbb{E}[\mathsf{X}]$. *Then:*

$$\Pr\left[\lambda_{\max}(\mathsf{X}) \geq (1+\epsilon)\lambda_{\max}(M)\right] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\lambda_{\min}(M)/R}$$

Matrix Concentration Applied to Leverage Score Sampling

Theorem (Matrix Chernoff Bound)

Consider independent symmetric random matrices $\mathsf{X}_1, \ldots, \mathsf{X}_m \in \mathbb{R}^{d \times d}$*, with* $\mathsf{X}_i \succeq 0$*,* $\lambda_{\max}(\mathsf{X}_i) \leq R$*, and* $\mathsf{X} = \sum_{i=1}^m \mathsf{X}_i$*. Let* $M = \mathbb{E}[\mathsf{X}]$*. Then:*

$$\Pr\left[\lambda_{\max}(\mathsf{X}) \geq (1+\epsilon)\lambda_{\max}(M)\right] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\lambda_{\min}(M)/R}$$

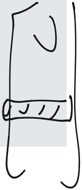
- In our setting, $\mathsf{X}_i = U^T \mathsf{S}_{:,j}^T \mathsf{S}_{:,j} U$. $\mathsf{X}_i = \underline{\frac{1}{mp_i} U_{i,:}^T U_{i,:}}$ with probability p_i.

i^{th} sample

Theorem (Matrix Chernoff Bound)

Consider independent symmetric random matrices $X_1, \ldots, X_m \in \mathbb{R}^{d \times d}$, with $X_i \succeq 0$, $\lambda_{\max}(X_i) \leq R$, and $X = \sum_{i=1}^m X_i$. Let $M = \mathbb{E}[X]$. Then:

$$\Pr\left[\lambda_{\max}(X) \geq (1+\epsilon)\lambda_{\max}(M)\right] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\lambda_{\min}(M)/R}$$



- In our setting, $X_i = U^T S_{:,j}^T S_{:,j} U$. $X_i = \underline{\frac{1}{mp_i} U_{i,:}^T} U_{i,:}$ with probability p_i.
- $M = \mathbb{E}[X] = \mathbb{E} \sum X_i = I$
- $R = \frac{1}{mp_i} \lambda_{\max}\left([U_{i,:}^T][U_{i,:}] \right)$

$\mathrm{tr}\left([U_{i,:}^T][U_{i,:}] \right) = [U_{i,:}][U_{i,:}^T] = \|U_{i,:}\|_2^2$

$= \frac{1}{m} \cdot \frac{\mathrm{rank}(A)}{\|U_{i,:}\|_2^2} \cdot \|U_{i,:}\|_2^2 = \boxed{\frac{\mathrm{rank}(A)}{m}}$

Matrix Concentration Applied to Leverage Score Sampling

Theorem (Matrix Chernoff Bound)

Consider independent symmetric random matrices $\mathsf{X}_1, \ldots, \mathsf{X}_m \in \mathbb{R}^{d \times d}$, *with* $\mathsf{X}_i \succeq 0$, $\lambda_{\max}(\mathsf{X}_i) \leq R$, *and* $\mathsf{X} = \sum_{i=1}^m \mathsf{X}_i$. *Let* $M = \mathbb{E}[\mathsf{X}]$. *Then:*

$$\Pr\left[\lambda_{\max}(\mathsf{X}) \geq (1+\epsilon)\lambda_{\max}(M)\right] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\lambda_{\min}(M)/R}$$

- In our setting, $\mathsf{X}_i = U^T \mathsf{S}_{:,j}^T \mathsf{S}_{:,j} U$. $\mathsf{X}_i = \frac{1}{mp_i} U_{i,:}^T U_{i,:}$ with probability p_i.
- $M = \mathbb{E}[\mathsf{X}] = I$
- $R = \frac{\mathrm{rank}(A)}{m} \leq \frac{d}{m}$
- $\Pr[U^T \mathsf{S}^T \mathsf{S} U \succeq (1+\epsilon)I] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{m/d}$

Matrix Concentration Applied to Leverage Score Sampling

Theorem (Matrix Chernoff Bound)

Consider independent symmetric random matrices $\mathsf{X}_1, \ldots, \mathsf{X}_m \in \mathbb{R}^{d\times d}$, with $\mathsf{X}_i \succeq 0$, $\lambda_{\max}(\mathsf{X}_i) \leq R$, and $\mathsf{X} = \sum_{i=1}^m \mathsf{X}_i$. Let $M = \mathbb{E}[\mathsf{X}]$. Then:

$$\Pr\left[\lambda_{\max}(\mathsf{X}) \geq (1+\epsilon)\lambda_{\max}(M)\right] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\lambda_{\min}(M)/R}$$

- In our setting, $\mathsf{X}_i = U^T \mathsf{S}_{:,j}^T \mathsf{S}_{:,j} U$. $\mathsf{X}_i = \frac{1}{mp_i} U_{i,:}^T U_{i,:}$ with probability p_i.
- $M = \mathbb{E}[\mathsf{X}] =$
- $R =$
- $\Pr[U^T \mathsf{S}^T \mathsf{S} U \succeq (1+\epsilon) I] \leq d \cdot \Bigg[\underbrace{\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}}_{e^{-\varepsilon^2}}\Bigg]^{m/d} \lesssim d \cdot e^{-\epsilon^2 \cdot m/d}$

Matrix Concentration Applied to Leverage Score Sampling

Theorem (Matrix Chernoff Bound)

Consider independent symmetric random matrices $X_1, \ldots, X_m \in \mathbb{R}^{d\times d}$, with $X_i \succeq 0$, $\lambda_{\max}(X_i) \leq R$, and $X = \sum_{i=1}^m X_i$. Let $M = \mathbb{E}[X]$. Then:

$$\Pr\left[\lambda_{\max}(X) \geq (1+\epsilon)\lambda_{\max}(M)\right] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\lambda_{\min}(M)/R}$$

- In our setting, $X_i = U^T S_{:,j}^T S_{:,j} U$. $X_i = \frac{1}{mp_i} U_{i,:}^T U_{i,:}$ with probability p_i.
- $M = \mathbb{E}[X] =$
- $R =$
- $\Pr[U^T S^T S U \succeq (1+\epsilon)I] \leq d \cdot \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{m/d} \lesssim d \cdot e^{-\epsilon^2 \cdot m/d} \simeq d \cdot e^{-\log(d/\delta)}$
  $\simeq d \cdot \frac{\delta}{d} = \delta$
- If we set $m = O\left(\frac{d\log(d/\delta)}{\epsilon^2}\right)$ we have $\Pr[U^T S^T S U \succeq (1+\epsilon)I] \leq \delta$.

intuiton??

Theorem (Subspace Embedding via Leverage Score Sampling)

For any $A \in \mathbb{R}^{n\times d}$ with left singular vector matrix U, let $\tau_i = \|U_{i,:}\|_2^2$ and $p_i = \frac{\tau_i}{\sum \tau_i}$. Let $\mathbf{S} \in \mathbb{R}^{m\times n}$ have $\mathbf{S}_{:,j}$ independently set to $\frac{1}{\sqrt{mp_i}} \cdot e_i^T$ with probability p_i.

Then, if $m = O\left(\frac{d \log(d/\delta)}{\epsilon^2}\right)$, with probability $\geq 1-\delta$, $\mathbf{S}$ is an ϵ-subspace embedding for A.

Subpace Embedding via Sampling

coupon collector

$A = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = U$

~~O~~(d log d) samples needed

Theorem (Subspace Embedding via Leverage Score Sampling)

For any $A \in \mathbb{R}^{n\times d}$ with left singular vector matrix U, let $\tau_i = \|U_{i,:}\|_2^2$ and $p_i = \frac{\tau_i}{\sum \tau_i}$. Let $\mathbf{S} \in \mathbb{R}^{m\times n}$ have $\mathbf{S}_{:,j}$ independently set to $\frac{1}{\sqrt{mp_i}} \cdot e_i^T$ with probability p_i.

Then, if $m = O\left(\frac{d\log(d/\delta)}{\epsilon^2}\right)$, with probability $\geq 1-\delta$, $\mathbf{S}$ is an ϵ-subspace embedding for A.

Matches oblivious random projection up to the $\log d$ factor.

- computing U is expensive & we can approximate τ_i very efficiently

- log d is required

Leverage Score Intuition

Check-In

Check-in Question: Would row-norm sampling from A directly rather than its left singular vectors U have worked to give a subspace embedding?

$$A = \begin{bmatrix} 1 &&&& \\ & 1 &&& \\ && 1 && \\ &&& 1 & \\ &&&& \alpha \end{bmatrix} \qquad \alpha \to 0$$

$$U = \begin{bmatrix} 1 &&& \\ & 1 && \\ && 1 & \\ &&& 1 \end{bmatrix}$$