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.

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.

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. 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

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.

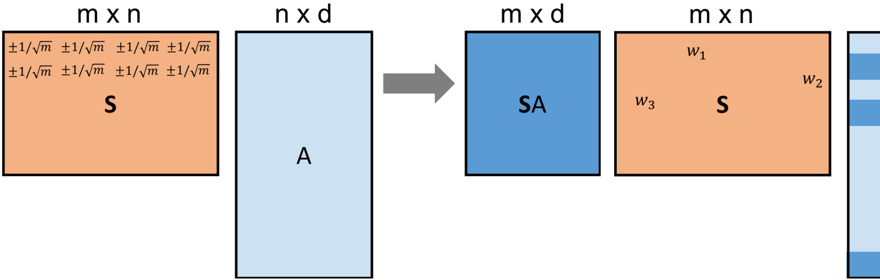
- ☐ a. Markov bound
- ☐ b. Bernstein bound
- ☐ c. Chebyshev inequality
- ☐ d. Hanson-Wright Inequality

Check

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 $x \in \mathbb{R}^d$,

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

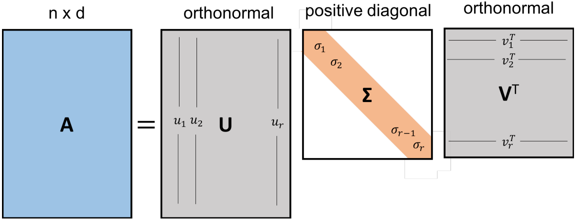


So Far: If $\mathbf{S}$ is a random sign matrix, and $m = O\left(\frac{d+\log(1/\delta)}{\epsilon^2}\right)$, then for any A, $\mathbf{S}$ is an ϵ-subspace embedding with probability $\geq 1-\delta$.

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

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 $x \in \mathbb{R}^d$,

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

Loewner Ordering

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

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

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

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

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

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

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?

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.

$$\begin{aligned}
\mathbb{E}[U^T\mathbf{S}^T\mathbf{S}U] = & = \sum_{j=1}^m \mathbb{E}[U^T\mathbf{S}_{:,j}^T\mathbf{S}_{:,j}U] \\
& = \sum_{j=1}^m \sum_{i=1}^n p_i \cdot (\frac{1}{\sqrt{mp_i}}U_{i,:}^T)(\frac{1}{\sqrt{mp_i}}U_{i,:}) \\
& = \sum_{j=1}^m \frac{1}{m}U^TU = I.
\end{aligned}$$

Matrix Concentration

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:

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

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

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.
- $M = \mathbb{E}[\mathsf{X}] =$
- $R =$
- $\Pr[U^T \mathsf{S}^T \mathsf{S} U \succeq (1+\epsilon) I] \leq d \cdot \left[\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}}\right]^{m/d} \lesssim d \cdot e^{-\epsilon^2 \cdot m/d}$
- If we set $m = O\left(\frac{d \log(d/\delta)}{\epsilon^2}\right)$ we have $\Pr[U^T \mathsf{S}^T \mathsf{S} U \succeq (1+\epsilon) I] \leq \delta$.

Subpace Embedding via Sampling

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 $\mathsf{S} \in \mathbb{R}^{m \times n}$ have $\mathsf{S}_{:,j}$ independently set to $\frac{1}{\sqrt{m p_i}} \cdot e_i^T$ with probability p_i.

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

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