COMPSCI 614: Randomized Algorithms with Applications to Data Science

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2024.

Lecture 7

Logistics

- I'll return midterms at the end of class.
- Overall the class did well – mean was a 25.5 out of 34 ($\approx 75\%$).
- Generally speaking people felt the test was a bit rushed.
- If you are not happy with your performance, message me and we can chat about it. I'm also happy to review solutions in office hours.
- I plan to release Problem Set 4 by end of this week.
- 2 page progress report on Final Project due 4/16.

Summary

Randomized Linear Algebra Before Break:

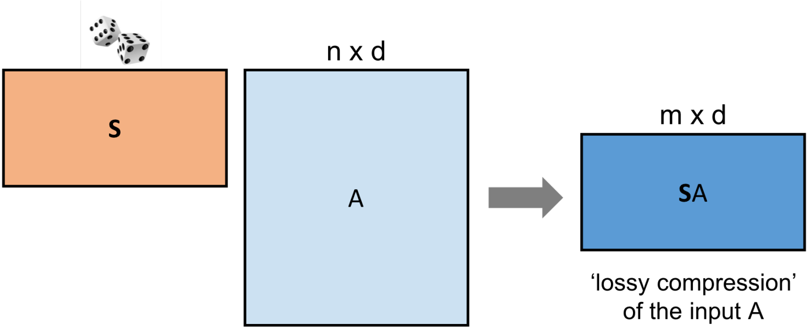
- Approximate matrix multiplication via norm-based sampling. Analysis via outer-product view of matrix multiplication.
- Application to fast randomized low-rank approximation.
- Hutchinson's method for trace estimation. Analysis via linearity of variance for pairwise-independent random variables.
- Random linear sketching for ℓ_0 sampling and ℓ_2 heavy-hitters (Count Sketch).

Today:

- Linear sketching for dimensionality reduction and the Johnson-Lindenstrauss lemma.
- Subspace embedding and ϵ-net arguments.

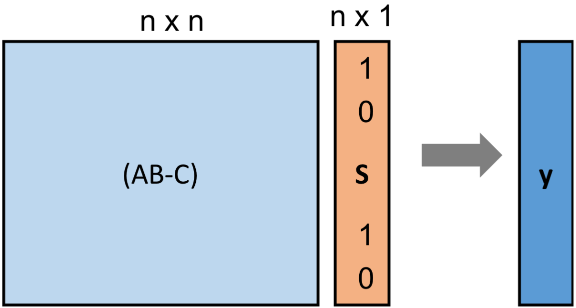
Linear Sketching

Given a large matrix $A \in \mathbb{R}^{n \times d}$, we pick a random linear transformation $\mathbf{S} \in \mathbb{R}^{m \times n}$ and compute $\mathbf{S}A$ (alternatively, pick $\mathbf{S} \in \mathbb{R}^{d \times m}$ and compute $A\mathbf{S}$). Using $\mathbf{S}A$ we can approximate many computations involving A.

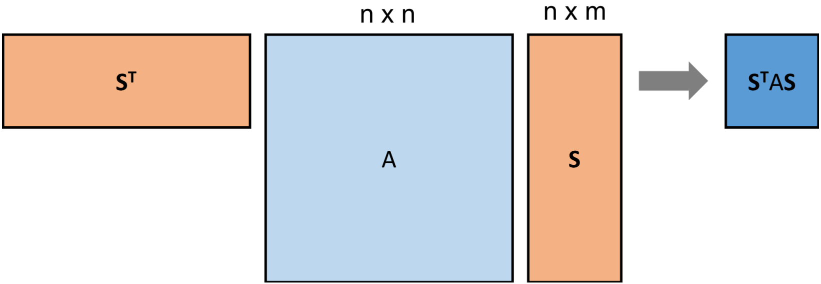


Linear Sketching Examples

Freivald's Algorithm:



Hutchinson's Trace Estimator:



Linear Sketching Examples

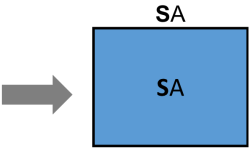
Graph Connectivity via ℓ_0 sampling:

l_0 sampling matrix **S**

1	**-1**	**0**	**0**	**1**	**-1**	**0**	**1**
-1	**0**	**1**	**1**	**0**	**0**	**-1**	**0**
1	**1**	**-1**	**0**	**-1**	**-1**	**0**	**1**
0	**-1**	**-1**	**-1**	**1**	**1**	**1**	**0**

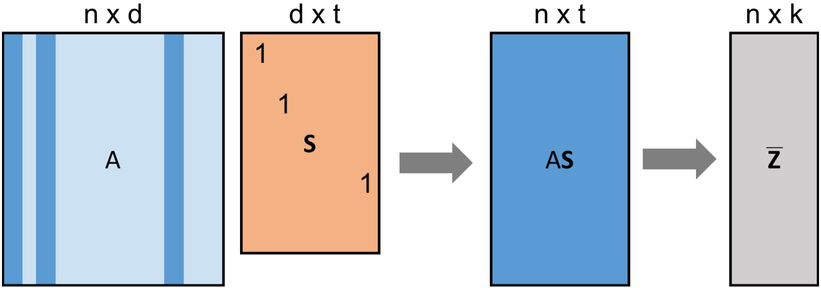
v_1	v_2	v_3	v_4
1	-1	0	0
0	1	0	-1
0	0	1	-1
-1	0	1	0
1	0	-1	0
0	1	-1	0
1	0	0	-1
0	0	1	-1

vertex-edge incidence matrix A



Linear Sketching Examples

Norm-Based Sampling for AMM/Low-Rank Approximation:



Subspace Embedding

Subspace Embedding

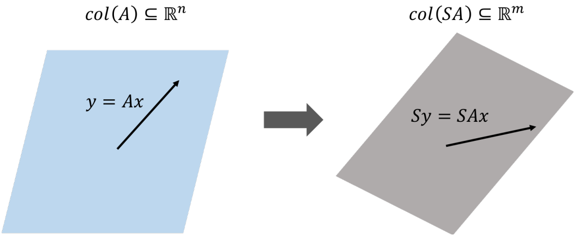
It is helpful to define general guarantees for sketches, that are useful in many problems.

Definition (Subspace Embedding)

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

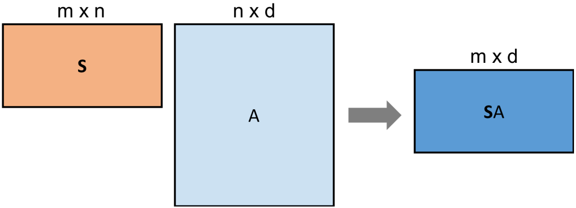
$$(1-\epsilon)\|Ax\|_2 \leq \|SAx\|_2 \leq (1+\epsilon)\|Ax\|_2.$$

I.e., S preserves the norm of any vector Ax in the column span of A.



Subspace Embedding Intuition

Think-Pair-Share 1: Assume that $n > d$ and that $\text{rank}(A) = d$. If $S \in \mathbb{R}^{m \times n}$ an is an ϵ-subspace embedding for A with $\epsilon < 1$, how large must m be? **Hint:** Think about $\text{rank}(SA)$ and/or the nullspace of SA.



Think-Pair-Share 2: Describe how to deterministically compute a subspace embedding S with $m = d$ and $\epsilon = 0$ in $O(nd^2)$ time.

Optimal Subspace Embedding

Let $Q \in \mathbb{R}^{n \times d}$ be an orthonormal basis for the columns of A. Then any vector Ax in A's column span can be written as Qy for some $y \in \mathbb{R}^d$.

Let $S = Q^T$. $S \in \mathbb{R}^{d \times n}$ (i.e., $m = d$) and further, for any $x \in \mathbb{R}^d$

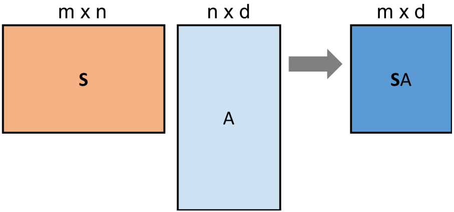
$$\|SAx\|_2^2 = \|Q^T Qy\|_2^2 = \|y\|_2^2 = \|Ax\|_2^2.$$

How would you compute Q?

Randomized Subspace Embedding

Theorem (Oblivious Subspace Embedding)

Let $\mathbf{S} \in \mathbb{R}^{m \times d}$ be a random matrix with i.i.d. $\pm 1/\sqrt{m}$ entries. Then if $m = O\left(\frac{d+\log(1/\delta)}{\epsilon^2}\right)$, for any $A \in \mathbb{R}^{n \times d}$, with probability $\geq 1 - \delta$, $\mathbf{S}$ is an ϵ-subspace embedding of A.



- $\mathbf{S}$ can be computed without any knowledge of A.
- Still achieves near optimal compression.
- Constructions where $\mathbf{S}$ is sparse or structured, allow efficient computation of $\mathbf{S}A$ (fast JL-transform, input-sparsity time algorithms via Count Sketch)

Oblivious Subspace Embedding Proof

Proof Outline

1. **Distributional Johnson-Lindenstrauss**: For $\mathbf{S} \in \mathbb{R}^{m \times d}$ with i.i.d. $\pm 1/\sqrt{m}$ entries, for any fixed $y \in \mathbb{R}^n$, with probability $1 - \delta$ for very small δ, $(1-\epsilon)\|y\|_2 \leq \|\mathbf{S}y\|_2 \leq (1+\epsilon)\|y\|_2$.
2. Via a union bound, have that for any fixed set of vectors $\mathcal{N} \subset \mathbb{R}^n$, with probability $1 - |\mathcal{N}| \cdot \delta$, $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2$ for all $y \in \mathcal{N}$.
3. But we want $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2$ for all $y = Ax$ with $x \in \mathbb{R}^d$. This is a linear subspace, i.e., an infinite set of vectors!
4. 'Discretize' this subspace by rounding to a finite set of vectors $\mathcal{N}$, called an ϵ-net for the subspace. Then apply union bound to this finite set, and show that the discretization does not introduce too much error.

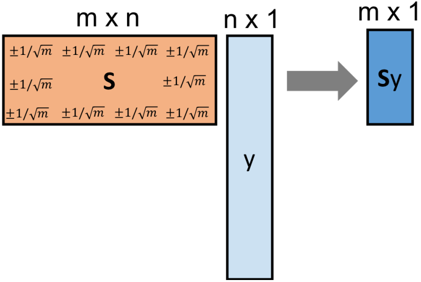
Remark: ϵ-nets are a key proof technique in theoretical computer science, learning theory (generalization bounds), random matrix theory, and beyond. They are a key take-away from this lecture.

Step 1: Distributional JL Lemma

Theorem (Distributional JL)

Let $\mathbf{S} \in \mathbb{R}^{m \times d}$ *be a random matrix with i.i.d.* $\pm 1/\sqrt{m}$ *entries. Then if* $m = O(\log(1/\delta)/\epsilon^2)$, *for any fixed* $y \in \mathbb{R}^n$, *with probability* $\geq 1 - \delta$, $(1-\epsilon)\|y\|_2 \leq \|\mathbf{S}y\|_2 \leq (1+\epsilon)\|y\|_2$.

I.e., via a random matrix, we can compress any vector from n to $\approx \log(1/\delta)/\epsilon^2$ dimensions, and approximately preserve its norm. A bit surprising maybe that m does not depend on n at all.

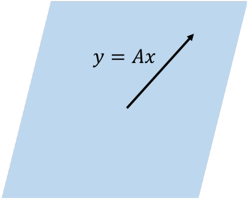


Restriction to Unit Ball

Want to show that with high probability, $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2$ for all $y \in \{Ax : x \in \mathbb{R}^d\}$. I.e., for all $y \in \mathcal{V}$, where $\mathcal{V}$ is A's column span.

Observation: Suffices to prove $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2 = 1$ for all $y \in S_\mathcal{V}$ where

$$S_\mathcal{V} = \{y \,:\, y \in \mathcal{V} \text{ and } \|y\|_2 = 1\}.$$



Proof: For any $y \in \mathcal{V}$, can write $y = \|y\|_2 \cdot \bar{y}$ where $\bar{y} = y/\|y\|_2 \in S_\mathcal{V}$.

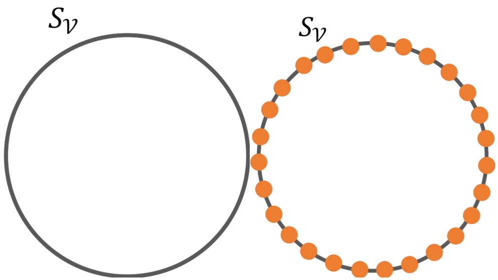
$$(1-\epsilon) \leq \|\mathbf{S}\bar{y}\|_2 \leq (1+\epsilon) \implies$$
$$(1-\epsilon)\cdot\|y\|_2 \leq \|\mathbf{S}\bar{y}\|_2 \cdot \|y\|_2 \leq (1+\epsilon)\cdot\|y\|_2 \implies$$
$$(1-\epsilon)\|y\|_2 \leq \|\mathbf{S}y\|_2 \leq (1+\epsilon)\|y\|_2.$$

Discretization of Unit Ball

Theorem

For any $\epsilon \leq 1$, there exists a set of points $\mathcal{N}_\epsilon \subset S_\mathcal{V}$ with $|\mathcal{N}_\epsilon| = \left(\frac{4}{\epsilon}\right)^d$ such that, for all $y \in S_\mathcal{V}$,

$$\min_{w \in \mathcal{N}_\epsilon} \|y - w\|_2 \leq \epsilon.$$

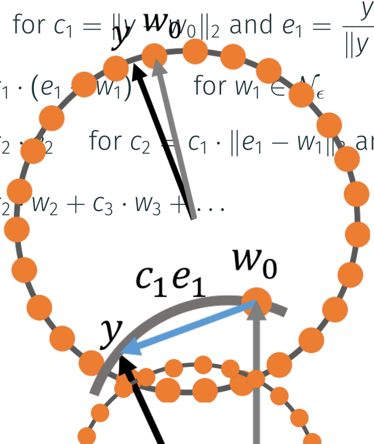


By the distributional JL lemma, if we set $\delta' = \delta \cdot \left(\frac{\epsilon}{4}\right)^d$ then, via a union bound, with probability at least $1 - \delta' \cdot |\mathcal{N}_\epsilon| = 1 - \delta$, for

So Far: If we set $m = \tilde{O}(d/\epsilon^2)$ and pick random $\mathbf{S} \in \mathbb{R}^{m \times n}$, then with probability $\geq 1 - \delta$, $\|\mathbf{S}w\|_2 \approx_\epsilon \|w\|_2$ for all $w \in \mathcal{N}_\epsilon$.

Expansion via net vectors: For any $y \in \mathcal{S}_\mathcal{V}$, we can write:

$$
\begin{aligned}
y &= w_0 + (y - w_0) && \text{for } w_0 \in \mathcal{N}_\epsilon \\
&= w_0 + c_1 \cdot e_1 && \text{for } c_1 = \|y - w_0\|_2 \text{ and } e_1 = \frac{y - w_0}{\|y - w_0\|_2} \in S_\mathcal{V} \\
&= w_0 + c_1 \cdot w_1 + c_1 \cdot (e_1 - w_1) && \text{for } w_1 \in \mathcal{N}_\epsilon \\
&= w_0 + c_1 \cdot w_1 + c_2 \cdot e_2 && \text{for } c_2 = c_1 \cdot \|e_1 - w_1\|_2 \text{ and } e_2 = \frac{e_1 - w_1}{\|e_1 - w_1\|_2} \in S_\mathcal{V} \\
&= w_0 + c_1 \cdot w_1 + c_2 \cdot w_2 + c_3 \cdot w_3 + \ldots
\end{aligned}
$$



Proof Via ϵ-net

Have written $y \in S_{\mathcal{V}}$ as $y = w_0 + c_1 w_1 + c_2 w_2 + \ldots$ where $w_0, w_1, \ldots \in \mathcal{N}_\epsilon$, and $c_i \leq \epsilon^i$. By triangle inequality:

$$\begin{aligned}
\|\mathbf{S}y\|_2 &= \|\mathbf{S}w_0 + c_1 \mathbf{S}w_1 + c_2 \mathbf{S}w_2 + \ldots\|_2 \\
&\leq \|\mathbf{S}w_0\|_2 + c_1\|\mathbf{S}w_1\|_2 + c_2\|\mathbf{S}w_2\|_2 + \ldots \\
&\leq (1+\epsilon) + \epsilon(1+\epsilon) + \epsilon^2(1+\epsilon) + \ldots \\
&\quad \text{(since via the union bound, } \|\mathbf{S}w\|_2 \approx \|w\|_2 \text{ for all } w \in \mathcal{N}_\epsilon) \\
&\leq \frac{1+\epsilon}{1-\epsilon} \approx 1 + 2\epsilon
\end{aligned}$$

Similarly, can prove that $\|\mathbf{S}y\|_2 \geq 1 - 2\epsilon$, giving, for all $y \in S_{\mathcal{V}}$ (and hence all $y \in \mathcal{V}$):

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

Full Argument

- There exists an ϵ-net $\mathcal{N}_\epsilon$ over the unit ball in A's column span, $S_{\mathcal{V}}$ with $|\mathcal{N}_\epsilon| \leq \left(\frac{4}{\epsilon}\right)^d$.
- By distributional JL, for $m = O\left(\frac{d\log(1/\epsilon)+\log(1/\delta)}{\epsilon^2}\right)$, with probability $\geq 1-\delta$, for all $w \in \mathcal{N}_\epsilon$, $\|\mathbf{S}w\|_2 \approx_\epsilon \|w\|_2$.

$\implies$ for all $y \in \mathcal{S}_{\mathcal{V}}$, $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2$.

$\implies$ for all $y \in \mathcal{V}$, i.e., for all $y = Ax$ for $x \in \mathbb{R}^d$, $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2$.

$\implies$ $\mathbf{S} \in \mathbb{R}^{m \times n}$ is an ϵ-subspace embedding for A.

Net Construction

Theorem (ϵ-net over ℓ_2 ball)

For any $\epsilon \leq 1$, there exists a set of points $\mathcal{N}_\epsilon \subset S_\mathcal{V}$ with $|\mathcal{N}_\epsilon| = \left(\frac{4}{\epsilon}\right)^d$ such that, for all $y \in S_\mathcal{V}$,

$$\min_{w \in \mathcal{N}_\epsilon} \|y - w\|_2 \leq \epsilon.$$

Theoretical algorithm for constructing $\mathcal{N}_\epsilon$:

- Initialize $\mathcal{N}_\epsilon = \{\}$.
- While there exists $v \in S_\mathcal{V}$ where $\min_{w \in \mathcal{N}_\epsilon} \|v - w\|_2 > \epsilon$, pick an arbitrary such v and let $\mathcal{N}_\epsilon := \mathcal{N}_\epsilon \cup \{v\}$.

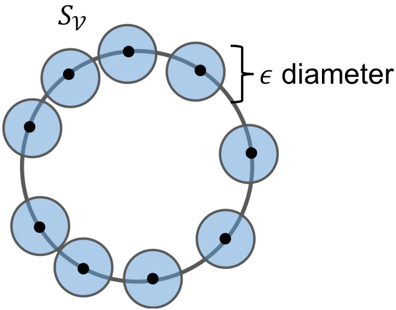
If the algorithm terminates in T steps, we have $|\mathcal{N}_\epsilon| \leq T$ and $\mathcal{N}_\epsilon$ is a valid ϵ-net.

Net Construction

How large is the net constructed by our theoretical algorithm?

Consider $w, w' \in \mathcal{N}_\epsilon$. We must have $\|w - w'\|_2 > \epsilon$, or we would have not added both to the net.

Thus, we can place an $\epsilon/2$ radius ball around each $w \in \mathcal{N}_\epsilon$, and none of these balls will intersect.



Note that all these balls lie within the ball of radius $(1 + \epsilon/2)$.

Volume Argument

We have $|\mathcal{N}_\epsilon|$ disjoint balls with radius $\epsilon/2$, lying within a ball of radius $(1 + \epsilon/2)$.

In d dimensions, the radius r ball has volume $c_d \cdot r^d$, where c_d is a constant that depends on d but not r.

Thus, the total number of balls is upper bounded by:

$$|\mathcal{N}_\epsilon| \leq \frac{(1+\epsilon/2)^d}{(\epsilon/2)^d} \leq \left(\frac{4}{\epsilon}\right)^d.$$

Remark: We never actually construct an ϵ-net. We just use the fact that one exists (the output of this theoretical algorithm) in our subspace embedding proof.