COMPSCI 690RA: Randomized Algorithms and Probabilistic Data Analysis

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2022.

Lecture 7

Logistics

- I'll return midterms at the end of class.
- Overall the class did very well – mean was a 29.75 out of 36 ($\approx 83\%$).
- 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 3 by end of this week.
- 1 page progress report on Final Project due 4/8.
- Weekly quiz due next Tuesday at 8pm.

Summary

Randomized Linear Algebra Before Break:

$AB = C$?

$(AB - C)y$

- Freivald's algorithm for matrix product testing.
- Hutchinson's method for trace estimation. Analysis via linearity of variance for pairwise-independent random variables.
- Approximate matrix multiplication via norm-based sampling. Analysis via outer-product view of matrix multiplication.
- Application to fast randomized low-rank approximation.
- Related ideas for sampling for initializing k-means clustering – the k-means++ algorithm.

Summary

Randomized Linear Algebra Before Break:

- Freivald's algorithm for matrix product testing.
- Hutchinson's method for trace estimation. Analysis via linearity of variance for pairwise-independent random variables.
- Approximate matrix multiplication via norm-based sampling. Analysis via outer-product view of matrix multiplication.
- Application to fast randomized low-rank approximation.
- Related ideas for sampling for initializing k-means clustering – the k-means++ algorithm.

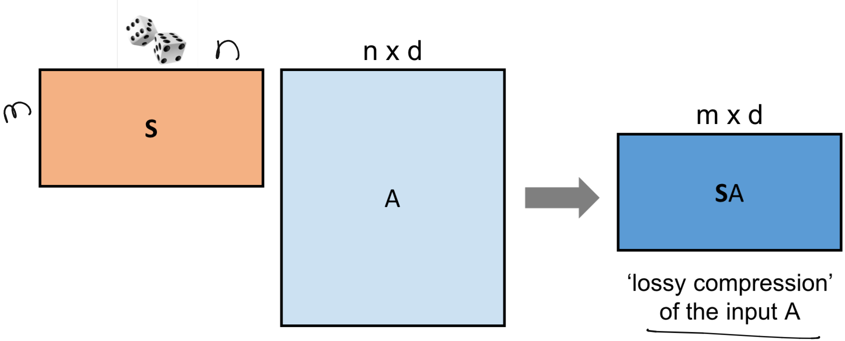
Today: Random sketching and the Johnson-Lindenstrauss lemma.

- Subspace embedding and ϵ-net arguments.
- Application to fast over-constrained linear regression.

Linear Sketching

Linear Sketching

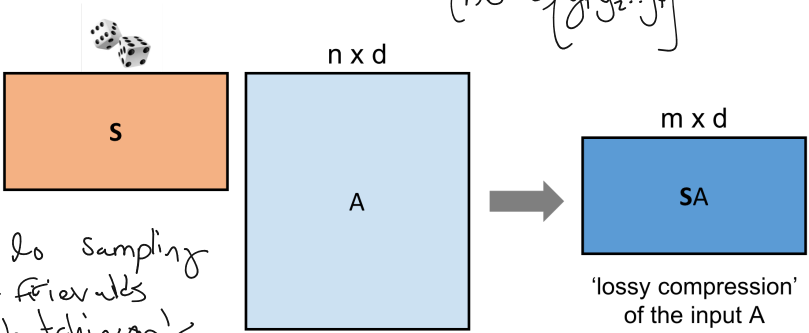
Given a large matrix $A \in \mathbb{R}^{n \times d}$, we pick a random linear transformation $\mathbf{S} \in \mathbb{R}^{m \times n}$ and compute $\mathbf{S}A$ (alternatively, pick $\mathbf{S} \in \mathbb{R}^{d \times m}$ and compute $A\mathbf{S}$). Using $\mathbf{S}A$ we can approximate many computations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



Linear Sketching

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.

$(AB - C)[y_1 \, y_2 \ldots y_t]$

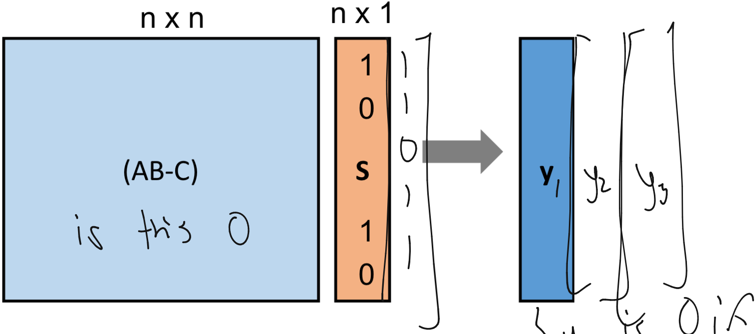


- ℓ_0 sampling
- Frievalds
- hutchinson's

What algorithms have we seen in class that are based on linear sketching?

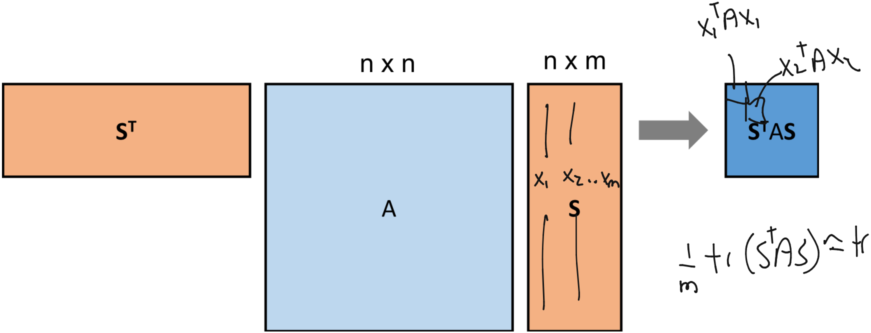
Linear Sketching Examples

Freivald's Algorithm:



y is 0 if AB-C is 0
else y prob not 0.

Hutchinson's Trace Estimator:



$$\frac{1}{m}\,\mathrm{tr}(S^T A S) \approx \mathrm{tr}(A)$$

$$\mathrm{tr}(A) = \frac{1}{m}\sum_{i=1}^{m} x_i^T A x_i$$

$x_1, x_2 \ldots x_m$ are i.i.d ± 1

Linear Sketching Examples

Graph Connectivity via ℓ_0 sampling:

l_0 sampling matrix **S**

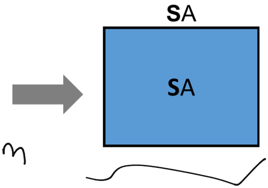
1	**-1**	**0**	**0**	**1**	**-1**	**0**	**1**
-1	**0**	**1**	**1**	**0**	**0**	**-1**	**0**
1	**1**	**-1**	**0**	**-1**	**-1**	**0**	**1**
0	**-1**	**-1**	**-1**	**1**	**1**	**1**	**0**

n

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

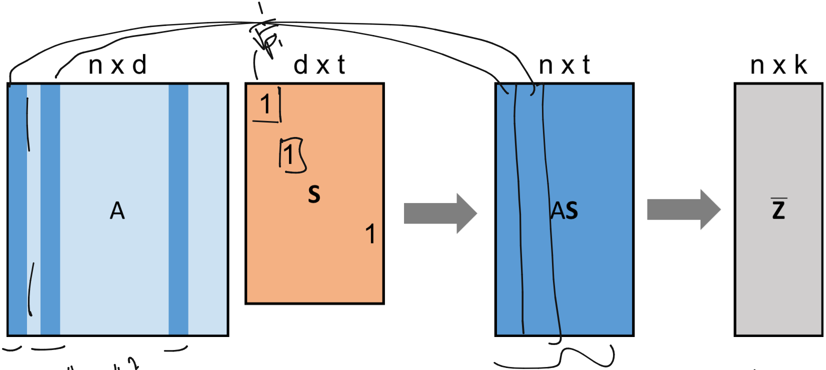
m

vertex-edge incidence matrix A



does this represent a connected graph

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



$p_i = \frac{\|a_i\|_2^2}{\|A\|_F^2}$

preserve spectral info including singular vectors

Subspace Embedding

Subspace Embedding

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

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

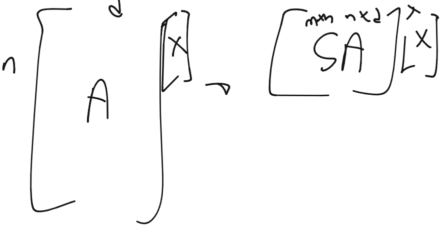
Definition (Subspace Embedding)

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

sketch matrix

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

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



Subspace Embedding

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

Definition (Subspace Embedding)

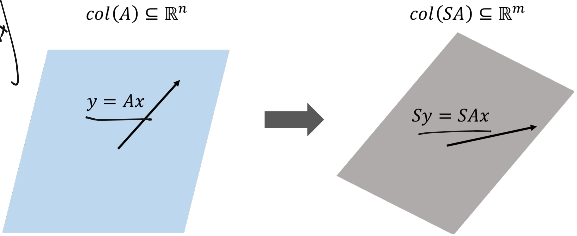
$S \in \mathbb{R}^{m \times d}$ is an ϵ-subspace embedding for $A \in \mathbb{R}^{n \times d}$ if, for all $x \in \mathbb{R}^d$,

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

Ax

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

$$\begin{bmatrix} A \end{bmatrix}\begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} Ax \end{bmatrix}$$



Subspace Embedding

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

$A[x]$

Definition (Subspace Embedding)

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

$m = n$

$S = I_{n \times n}$

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

how big is m? $m << n$

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

- Fast linear regression (next) and preconditioning. SA
- Approximation of A's singular values.
- Approximate matrix multiplication and near optimal low-rank approximation. → streaming mode

SA

- Compressed sensing/sparse recovery (related to ℓ_0 sampling).

Subspace Embedding Application

Theorem (Sketched Linear Regression)

Consider $A \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^n$. We seek to find an approximate solution to the linear regression problem:

$$\operatorname*{arg\,min}_{x \in \mathbb{R}^d} \underbrace{\|\overset{n\times d}{A}x - \overset{n\times 1}{b}\|_2}. \quad , \; O(nd^2)$$

Let $S \in \mathbb{R}^{m \times d}$ be an ϵ-subspace embedding for $[A; b] \in \mathbb{R}^{n \times d+1}$. Let $\tilde{x} = \arg\min_{x \in \mathbb{R}^d} \|\overset{m\times d}{SA}x - \overset{m\times 1}{Sb}\|_2$. Then we have:

$$\begin{bmatrix} A & b \end{bmatrix}$$

$$\underline{\|A\tilde{x} - b\|_2} \leq \overset{\approx 1+2\epsilon}{\frac{1+\epsilon}{1-\epsilon}} \cdot \min_{x \in \mathbb{R}^d} \|Ax - b\|_2.$$

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x \end{bmatrix} \approx \begin{bmatrix} b \end{bmatrix}$$

Subspace Embedding Application

Theorem (Sketched Linear Regression)

Consider $A \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^n$. We seek to find an approximate solution to the linear regression problem:

$$\arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_2.$$

$$\left[\, A \;\vdots\; b \,\right]$$

Let $S \in \mathbb{R}^{m \times d}$ be an ϵ-subspace embedding for $[A; b] \in \mathbb{R}^{n \times d+1}$. Let $\tilde{x} = \arg\min_{x \in \mathbb{R}^d} \|SAx - Sb\|_2$. Then we have:

$$\|A\tilde{x} - b\|_2 \leq \frac{1+\epsilon}{1-\epsilon} \cdot \min_{x \in \mathbb{R}^d} \|Ax - b\|_2.$$

- Time to compute $x^* = \arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$ is $O(nd^2)$.
- Time to compute $\tilde{x}$ is just $O(md^2)$. For large n (i.e., a highly over-constrained problem) can set $m \ll n$.

Sketched Regression Proof

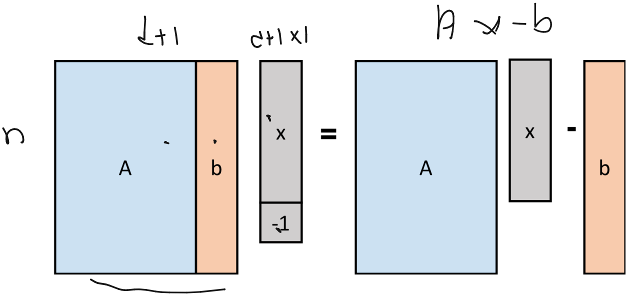
Claim: Since S is a ϵ- subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

$$(1-\epsilon)\|\underbrace{Ax-b}\|_2 \leq \|\underbrace{SAx-Sb}\|_2 \leq (1+\epsilon)\|Ax-b\|_2.$$

Sketched Regression Proof

Claim: Since S is a subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

$$(1-\epsilon)\underbrace{\overset{[A;b]\begin{bmatrix}x\\-1\end{bmatrix}}{\|Ax-b\|_2}} \leq \underbrace{\overset{S[A;b]\begin{bmatrix}x\\-1\end{bmatrix}}{\|SAx-Sb\|_2}} \leq (1+\epsilon)\overset{[A;b]\begin{bmatrix}x\\-1\end{bmatrix}}{\|Ax-b\|_2}.$$



Sketched Regression Proof

Claim: Since S is a subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

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

Sketched Regression Proof

Claim: Since S is a subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

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

Let $x^* = \arg\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$ and $\tilde{x} = \arg\min_{x \in \mathbb{R}^d} \|SAx - Sb\|_2$.

Sketched Regression Proof

Claim: Since S is a subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

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

Let $x^* = \arg\min_{x\in\mathbb{R}^d} \|Ax - b\|_2$ and $\tilde{x} = \arg\min_{x\in\mathbb{R}^d} \|SAx - Sb\|_2$. We have:

$$\|A\tilde{x} - b\|_2 \leq \frac{1}{1-\epsilon}\|SA\hat{x} - Sb\|_2$$

$$(1-\varepsilon)\|A\tilde{x}-b\| \leq \|SA\hat{x} - Sb\|_2$$

Sketched Regression Proof

Claim: Since S is a subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

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

Let $x^* = \arg\min_{x\in\mathbb{R}^d} \|Ax-b\|_2$ and $\tilde{x} = \arg\min_{x\in\mathbb{R}^d} \|SAx-Sb\|_2$. We have:

$$\|A\tilde{x}-b\|_2 \leq \frac{1}{1-\epsilon}\underbrace{\|SA\tilde{x} - Sb\|_2}_{\hat{x} = \arg\min \|SAx - Sb\|} \overset{?}{\leq} \frac{1}{1-\epsilon}\cdot\|SAx^*-Sb\|_2$$

Claim: Since S is a subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

positive

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

Let $x^* = \arg\min_{x\in\mathbb{R}^d} \|Ax-b\|_2$ and $\tilde{x} = \arg\min_{x\in\mathbb{R}^d} \|SAx-Sb\|_2$.
We have:

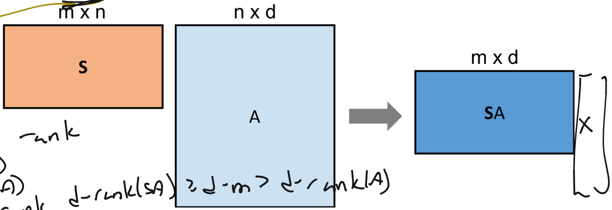
$$\boxed{\|A\tilde{x}-b\|_2} \leq \frac{1}{1-\epsilon}\|SAx-Sb\|_2 \leq \frac{1}{1-\epsilon}\cdot\|SAx^*-Sb\|_2$$
$$\leq \frac{1+\epsilon}{1-\epsilon}\cdot\|Ax^*-b\|_2.$$

$$\simeq (1+2\epsilon)\|Ax^*-b\|_2$$

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.

$m \geq \text{rank}(A) = d$



$SA\,x = 0$

null(A) has rank $d - \text{rank}(A)$

if $m < \text{rank}(A)$ null(SA) has rank $d - \text{rank}(SA) \geq d - m > d - \text{rank}(A)$

for any x, $\|x\| > 0$

$Ax \neq 0$

$\|Ax\| \neq 0 \quad \|SAx\| = 0$

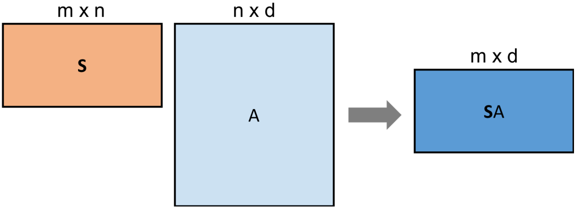
if $m < d$

$\text{rank}(SA) \leq \min(m, d) = m$

so $\exists x$, $x \neq 0$ s.t. $\|SAx\| = 0$

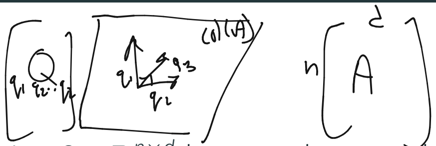
Subspace Embedding Intuition

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 $y \in \mathbb{R}^d$.

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

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., $\underline{m = d}$) and further, for any $x \in \mathbb{R}^d$

$$\|\underline{SAx}\|_2^2 = \|\overset{I}{\underbrace{Q^T}_{S}\underbrace{Qy}_{Ax}}\|_2^2 \qquad .$$

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

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

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

$$\|SAx\|_2^2 = \|Q^TQy\|_2^2 = \underline{\|y\|_2^2} = \underline{\|Ax\|_2^2}.$$

$$\|Qy\|_2^2 = y^TQ^TQy$$
$$= y^Ty = \|y\|_2^2$$

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?

– $O(nd^2)$

SVD

Gaussian

gram-schidmt, QR

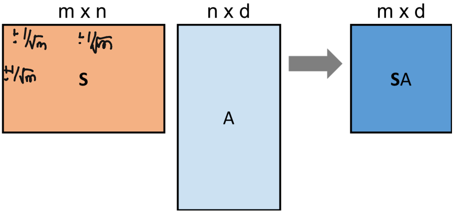
$A = [Q][R]$

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*.*

picked independt of A.

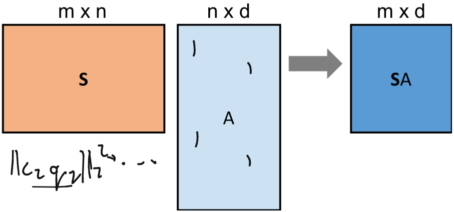


Randomized Subspace Embedding

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.

$n \geq d$ $\quad m \geq nd^2$



$\|Ax\|_2^2 = \|c_1 q_1\|_2^2 + \|c_2 q_2\|_2^2 \ldots$

How check if it is a subspace embedding:

let $q_1 \ldots q_d \in \mathbb{R}^n$ be a basis for col(A)

check if $\|Sq\| \approx_\epsilon \|q\|$ $\forall q$ in basis

check in $O(nd^2)$ time

- $\mathbf{S}$ 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) $O(\text{nnz}(A))$

Oblivious Subspace Embedding Proof

Proof Outline

/Gaussian

1. **Distributional Johnson-Lindenstrauss**: For $\mathbf{S} \in \mathbb{R}^{m \times n}$ 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, $(1-\epsilon)\|y\|_2 \leq \|\mathbf{S}y\|_2 \leq (1+\epsilon)\|y\|_2$. $m = O\left(\frac{\log(1/\delta)}{\varepsilon^2}\right)$

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 $y \in \mathcal{N}$.

$$\begin{array}{cccc} n\times 1 & n\times 1 & & n\times 1 \\ \begin{bmatrix} \\ y_1 \\ \\ \end{bmatrix} & \begin{bmatrix} \\ y_2 \\ \\ \end{bmatrix} & \cdots & \begin{bmatrix} \\ y_{|N|} \\ \\ \end{bmatrix} \end{array}$$

$m = O\left(\frac{\log |N|}{\epsilon^2}\right)$

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!

$y \in col(A)$

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! col(A)
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.

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.

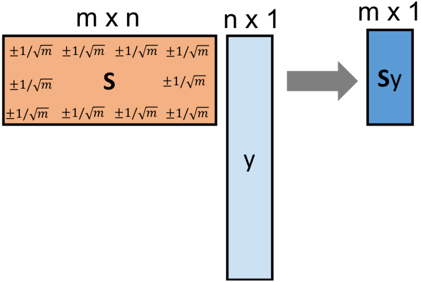
Remark: ϵ-nets are a key proof technique in theoretical computer science, learning theory (generalization bounds), random matrix theory, and beyond. They are a key take-away from this lecture.

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.



Step 1: Distributional JL Lemma

Theorem (Distributional JL)

Let $\mathsf{S} \in \mathbb{R}^{m \times d}$ 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 $\geq 1 - \delta$, $(1-\epsilon)\|y\|_2 \leq \|\mathsf{S}y\|_2 \leq (1+\epsilon)\|y\|_2$.

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.

Expectation:

$$\mathbb{E}[\|\mathsf{S}y\|_2^2] = \sum_{i=1}^{m} \mathbb{E}[\langle \mathsf{S}_{i,:}, y \rangle^2]$$

$$\begin{bmatrix} - s_i - \end{bmatrix} \begin{bmatrix} | \\ y \\ | \end{bmatrix} \qquad \begin{bmatrix} [Sy](1) \\ [Sy](2) \\ \vdots \\ [Sy](m) \end{bmatrix}$$

Step 1: Distributional JL Lemma

Theorem (Distributional JL)

Let $S \in \mathbb{R}^{m \times d}$ 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 $\geq 1 - \delta$, $(1-\epsilon)\|y\|_2 \leq \|Sy\|_2 \leq (1+\epsilon)\|y\|_2$.

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.

Expectation:

$$\mathbb{E}[\|Sy\|_2^2] = \sum_{i=1}^{m} \mathbb{E}[\langle S_{i,:}, y\rangle^2] = \sum_{i=1}^{m} \mathbb{E}\left[\left(\sum_{j=1}^{n} \overset{\pm\frac{1}{\sqrt{m}}}{S_{ij}} \cdot y_j\right)^2\right]$$

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.

Expectation:

$$\mathbb{E}[\|\mathbf{S}y\|_2^2] = \sum_{i=1}^m \mathbb{E}[\langle \mathbf{S}_{i,:}, y\rangle^2] = \sum_{i=1}^m \mathbb{E}\left[\overbrace{\left(\underbrace{\sum_{j=1}^n \mathbf{S}_{ij} \cdot y_j}_{\text{independent}}\right)^2}^{= \mathrm{Var}\left(\sum_{j=1}^n S_{ij} y_j\right)}\right]$$

$$= \sum_{i=1}^m \sum_{j=1}^n \mathrm{Var}(\mathbf{S}_{ij} \cdot y_j)$$

$\pm \frac{1}{\sqrt{m}}$

$y_j^2 \mathrm{Var}(S_{ij}) = \frac{y_j^2}{m}$

$$= \sum_{i=1}^m \sum_{j=1}^n \frac{y_j^2}{m} = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n y_j^2 = \sum_{j=1}^n y_j^2 = \|y\|_2^2$$

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.

Expectation:

$$\mathbb{E}[\|\mathbf{S}y\|_2^2] = \sum_{i=1}^{m} \mathbb{E}[\langle \mathbf{S}_{i,:}, y\rangle^2] = \sum_{i=1}^{m} \mathbb{E}\left[\left(\sum_{j=1}^{n} \mathbf{S}_{ij} \cdot y_j\right)^2\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Var}(\mathbf{S}_{ij} \cdot y_j)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{1}{m} \cdot y_j^2$$

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.

Expectation:

$$\mathbb{E}[\|\mathbf{S}y\|_2^2] = \sum_{i=1}^{m} \mathbb{E}[\langle \mathbf{S}_{i,:}, y \rangle^2] = \sum_{i=1}^{m} \mathbb{E}\left[\left(\sum_{j=1}^{n} \mathbf{S}_{ij} \cdot y_j\right)^2\right]$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \mathrm{Var}(\mathbf{S}_{ij} \cdot y_j)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{1}{m} \cdot y_j^2 = \|y\|_2^2.$$

Restriction to Unit Ball

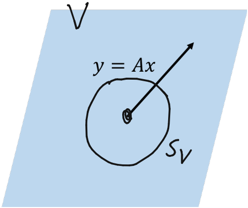
Want to show that with high probability, $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2$ for all $y \in \{Ax : x \in \mathbb{R}^d\}$. I.e., for all $y \in \mathcal{V}$, where $\mathcal{V}$ is A's column span.

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

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

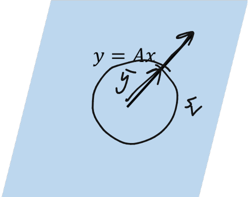


Restriction to Unit Ball

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 $\bar{y} = y/\|y\|_2 \in S_\mathcal{V}$.

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

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

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

$\|S\bar{y}\|_2 \cdot \|y\|_2$

$= \|S\bar{y} \cdot \|y\|_2\|_2$

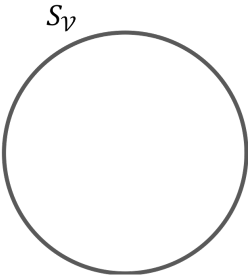
$= \|Sy\|_2$

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 $y \in S_{\mathcal{V}}$,

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



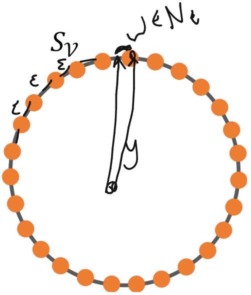
Discretization of Unit Ball

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 $y \in S_\mathcal{V}$,

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

$N_\varepsilon \leq$



$\#\text{points} \cdot \varepsilon = 2\pi$

$\#\text{points} \simeq \frac{2\pi}{\varepsilon}$

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 $y \in S_\mathcal{V}$,

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

$\delta / |\mathcal{N}_\epsilon|$

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 all $w \in \mathcal{N}_\epsilon$,

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

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 all $w \in \mathcal{N}_\epsilon$,

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

Requires $\mathbf{S} \in \mathbb{R}^{m \times n}$ where

$$m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right)$$

$\log(1/\delta') = \log\left(\frac{1}{\delta} \cdot \left(\frac{4}{\epsilon}\right)^d\right)$

$\log(1/\delta) + d\log(4/\epsilon)$

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 all $w \in \mathcal{N}_\epsilon$,

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

Requires $\mathbf{S} \in \mathbb{R}^{m \times n}$ where

can set rid of.

$$m = O\left(\frac{\log(1/\delta')}{\epsilon^2}\right) = O\left(\frac{d\log(4/\epsilon) + \log(1/\delta)}{\epsilon^2}\right) = \tilde{O}\left(\frac{d}{\epsilon^2}\right).$$

$$O\left(\frac{d + \log(1/\delta)}{\epsilon^2}\right)$$

Proof Via ϵ-net

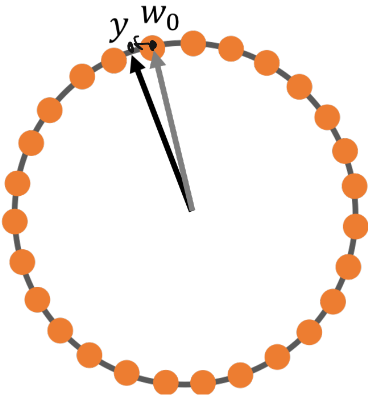
So Far: If we set $m = \tilde{O}(d/\epsilon^2)$ and pick random $\mathbf{S} \in \mathbb{R}^{m \times n}$, then with probability $\geq 1 - \delta$, $\|\mathbf{S}w\|_2 \approx_\epsilon \|w\|_2$ for all $w \in \mathcal{N}_\epsilon$.

Proof Via ϵ-net

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 $w \in \mathcal{N}_\epsilon$. $\Rightarrow$ $\|Sy\| \approx \|y\|$ $\forall y \in S_V$

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

$y = w_0 + \underbrace{(y - w_0)}_{<\epsilon}$ for $w_0 \in \mathcal{N}_\epsilon$



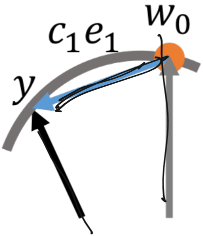
Proof Via ϵ-net

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:

$$y = w_0 + (y - w_0) \qquad \text{for } w_0 \in \mathcal{N}_\epsilon$$

$$= w_0 + \underset{\leq \varepsilon}{c_1} \cdot e_1 \qquad \text{for } c_1 = \|y - w_0\|_2 \text{ and } e_1 = \frac{y - w_0}{\|y - w_0\|_2} \in S_\mathcal{V}$$



Proof Via ϵ-net

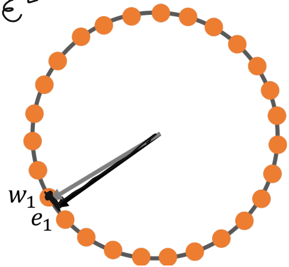
So Far: If we set $m = \tilde{O}(d/\epsilon^2)$ and pick random $\mathbf{S} \in \mathbb{R}^{m \times n}$, then with probability $\geq 1 - \delta$, $\|\mathbf{S}w\|_2 \approx_\epsilon \|w\|_2$ for all $w \in \mathcal{N}_\epsilon$.

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

$$y = w_0 + (y - w_0) \qquad \text{for } w_0 \in \mathcal{N}_\epsilon$$

$$= w_0 + \underbrace{c_1}_{\leq \epsilon} \cdot e_1 \qquad \text{for } c_1 = \|y - w_0\|_2 \text{ and } e_1 = \frac{y - w_0}{\|y - w_0\|_2} \in S_\mathcal{V}$$

$$= w_0 + c_1 \cdot w_1 + \underbrace{c_1 \cdot (e_1 - w_1)}_{\leq \epsilon^2} \qquad \text{for } w_1 \in \mathcal{N}_\epsilon$$



Proof Via ϵ-net

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:

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

Proof Via ϵ-net

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:

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

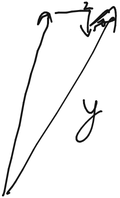
Proof Via ϵ-net

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:

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

For all i, have $c_i \leq \epsilon^i$.



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 $c_i \leq \epsilon^i$.

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:

$$\|\mathbf{S}y\|_2 = \|\mathbf{S}w_0 + c_1\mathbf{S}w_1 + c_2\mathbf{S}w_2 + \ldots\|_2$$

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 $c_i \le \epsilon^i$. By triangle inequality:

$$\begin{aligned}\|\mathbf{S}y\|_2 &= \|\mathbf{S}w_0 + c_1\mathbf{S}w_1 + c_2\mathbf{S}w_2 + \ldots\|_2 \\ &\le \|\mathbf{S}w_0\|_2 + c_1\|\mathbf{S}w_1\|_2 + c_2\|\mathbf{S}w_2\|_2 + \ldots \\ &\le 1+\varepsilon + \varepsilon(1+\varepsilon) + \varepsilon^2(1+\varepsilon) \ldots\end{aligned}$$

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:

$$\begin{aligned}
\|\mathbf{S}y\|_2 &= \|\mathbf{S}w_0 + c_1\mathbf{S}w_1 + c_2\mathbf{S}w_2 + \ldots\|_2 \\
&\leq \|\mathbf{S}w_0\|_2 + c_1\|\mathbf{S}w_1\|_2 + c_2\|\mathbf{S}w_2\|_2 + \ldots \\
&\leq (1+\epsilon) + \epsilon(1+\epsilon) + \epsilon^2(1+\epsilon) + \ldots
\end{aligned}$$

(since via the union bound, $\|\mathbf{S}w\|_2 \approx \|w\|_2$ for all $w \in \mathcal{N}_\epsilon$)

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:

$$\begin{aligned}
\|\mathbf{S}y\|_2 &= \|\mathbf{S}w_0 + c_1\mathbf{S}w_1 + c_2\mathbf{S}w_2 + \ldots\|_2 \\
&\leq \|\mathbf{S}w_0\|_2 + c_1\|\mathbf{S}w_1\|_2 + c_2\|\mathbf{S}w_2\|_2 + \ldots \\
&\leq (1+\epsilon) + \epsilon(1+\epsilon) + \epsilon^2(1+\epsilon) + \ldots
\end{aligned}$$

$$(1+\epsilon)\big(\overbrace{1+\epsilon+\epsilon^2+\ldots}^{\frac{1}{1-\epsilon}}\big)$$

(since via the union bound, $\|\mathbf{S}w\|_2 \approx \|w\|_2$ for all $w \in \mathcal{N}_\epsilon$)

$$\leq \frac{1+\epsilon}{1-\epsilon} \approx 1 + 2\epsilon$$

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:

$\|S(w_0 + c_1 w_1 + \cdots)\|$

$(1-\varepsilon) \leq \|Sy\| \leq (1+\varepsilon)$

$\|a+b\| \leq \|a\| + \|b\|$

$$\begin{aligned}
\|\mathbf{S}y\|_2 &= \|\mathbf{S}w_0 + c_1 \mathbf{S}w_1 + c_2 \mathbf{S}w_2 + \ldots\|_2 \\
&\leq \|\mathbf{S}w_0\|_2 + c_1\|\mathbf{S}w_1\|_2 + c_2\|\mathbf{S}w_2\|_2 + \ldots \\
&\leq (1+\epsilon) + \epsilon(1+\epsilon) + \epsilon^2(1+\epsilon) + \ldots
\end{aligned}$$

(since via the union bound, $\|\mathbf{S}w\|_2 \approx \|w\|_2$ for all $w \in \mathcal{N}_\epsilon$)

$$\leq \frac{1+\epsilon}{1-\epsilon} \approx 1 + 2\epsilon$$

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

$1-2\varepsilon$ $\quad$ $1+2\varepsilon$

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

$\sqrt{n}$

$\|Sy\|_2 = \|Sw_0 + S(y-w_0)\| \leq \|Sw_0\| + \|S(y-w_0)\|_2 \leq (1+\varepsilon) \cdot \|S\|_2 \cdot \|y-w_0\|$

$\leq (1+\varepsilon) + \|S(y-w_0)\| \; \} \leq \|S\|_2 \cdot \|y-w_0\|$

$\leq (1+\varepsilon) + 2\varepsilon\sqrt{n}$

Full Argument

- There exists an ϵ-net $\mathcal{N}_\epsilon$ over the unit ball in A's column span, $S_{\mathcal{V}}$ with $|\mathcal{N}_\epsilon| \leq \left(\frac{4}{\epsilon}\right)^d$.

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 $w \in \mathcal{N}_\epsilon$, $\|\mathbf{S}w\|_2 \approx_\epsilon \|w\|_2$.

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

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 $x \in \mathbb{R}^d$, $\|\mathbf{S}y\|_2 \approx_\epsilon \|y\|_2$.

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 $y \in S_\mathcal{V}$,

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

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 $\mathcal{N}_\epsilon := \mathcal{N}_\epsilon \cup \{v\}$.

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?

Net Construction

How large is the net constructed by our theoretical algorithm?

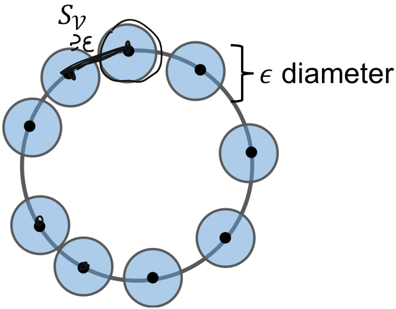
Consider $w, w' \in \mathcal{N}_\epsilon$. We must have $\|w - w'\|_2 > \epsilon$, or we would have not added both to the net.

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.

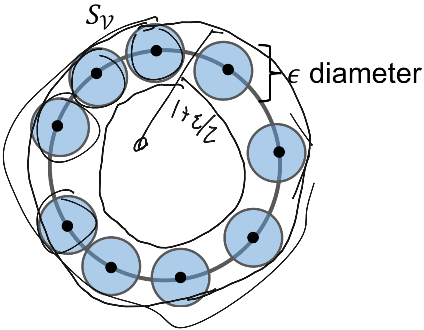


Net Construction

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

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.

Volume Argument

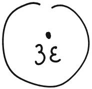
We have $|\mathcal{N}_\epsilon|$ disjoint balls with radius $\epsilon/2$, lying within a ball of radius $(1+\epsilon/2)$.

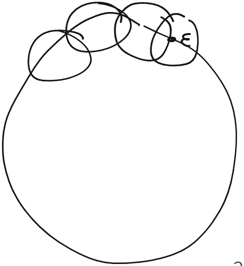
In d dimensions, the radius r ball has volume $c_d \cdot r^d$, where c_d is a constant that depends on d but not r.

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

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

$$\left(\frac{4}{\epsilon}\right)^{d-1}$$





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.

Distributional JL Lemma Proof

Proofs of Distributional JL Lemma

There are many proofs of the distributional JL Lemma:

- Let $\mathbf{S} \in \mathbb{R}^{m \times n}$ have i.i.d. Gaussian entries. Observe that each entry of $\mathbf{S}y$ is distributed as $\mathcal{N}(0, \|y\|_2^2)$, and give a proof via concentration of independent Chi-Squared random variables (see 514 slides). $\Sigma \left[\underbrace{[Sy](i)^2}_{} \right]$

squared Gaussians, χ-squared random variable

Proofs of Distributional JL Lemma

There are many proofs of the distributional JL Lemma:

- Let $\mathbf{S} \in \mathbb{R}^{m \times n}$ have i.i.d. Gaussian entries. Observe that each entry of $\mathbf{S}y$ is distributed as $\mathcal{N}(0, \|y\|_2^2)$, and give a proof via concentration of independent Chi-Squared random variables (see 514 slides).
- Write $\|\mathbf{S}y\|_2^2 = \sum_{i=1}^{m} \overbrace{\sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{S}_{i,j}\mathbf{S}_{i,k} y_j y_k}^{\langle S_{i,:}, y\rangle^2}$ and prove concentration of this sum, even though the terms are not all independent of each other (only pairwise independent within one row).

Proofs of Distributional JL Lemma

There are many proofs of the distributional JL Lemma:

- Let $\mathbf{S} \in \mathbb{R}^{m \times n}$ have i.i.d. Gaussian entries. Observe that each entry of $\mathbf{S}y$ is distributed as $\mathcal{N}(0, \|y\|_2^2)$, and give a proof via concentration of independent Chi-Squared random variables (see 514 slides).
- Write $\|\mathbf{S}y\|_2^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{S}_{i,j}\mathbf{S}_{i,k}y_j y_k$ and prove concentration of this sum, even though the terms are not all independent of each other (only pairwise independent within one row).
- Apply the Hanson-Wright inequality – an exponential concentration inequality for random quadratic forms. $x^T A x$
- This inequality comes up in a lot of places, including in the tight analysis of Hutchinson's trace estimator.

Hanson Wright Inequality

Theorem (Hanson-Wright Inequality)

Let $\mathbf{x} \in \mathbb{R}^n$ *be a vector of i.i.d. random* ± 1 *values. For any matrix* $A \in \mathbb{R}^{n \times n}$,

$\mathbb{E}\, x^T A x = tr(A)$

constant

$$\Pr[|\mathbf{x}^T A \mathbf{x} - \text{tr}(A)| \geq t] \leq 2 \exp\left(-c \cdot \min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_2}\right\}\right).$$

$x^T A x = \sum x_i x_j A_{ij}$

Hanson Wright Inequality

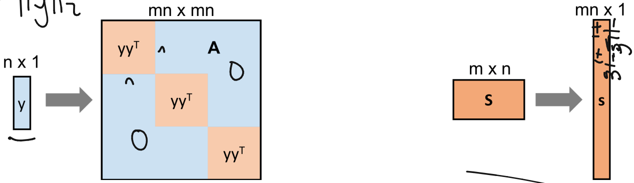
Theorem (Hanson-Wright Inequality)

Let $\mathbf{x} \in \mathbb{R}^n$ be a vector of i.i.d. random ± 1 values. For any matrix $A \in \mathbb{R}^{n \times n}$,

$$\Pr[|\mathbf{x}^T A\mathbf{x} - \operatorname{tr}(A)| \geq t] \leq 2\exp\left(-c \cdot \min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_2}\right\}\right).$$

$\|Sy\|_2^2 \approx \|y\|_2^2$

$\mathbb{E}\|Sy\|_2^2 = \|y\|_2^2$



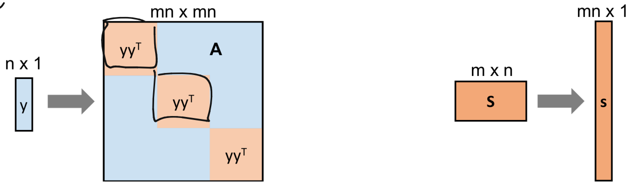
Hanson Wright Inequality

Theorem (Hanson-Wright Inequality)

Let $\mathbf{x} \in \mathbb{R}^n$ be a vector of i.i.d. random ± 1 values. For any matrix $A \in \mathbb{R}^{n \times n}$,

$$\Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \operatorname{tr}(A)\right| \geq t\right] \leq 2\exp\left(-c \cdot \min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_2}\right\}\right).$$

$\|Sy\|_2^2$



$\operatorname{tr}(yy^T)$

$$\begin{bmatrix} y_1^2 & & y_i y_j \\ & y_2^2 & \\ & & \ddots \\ & & & y_n^2 \end{bmatrix}$$

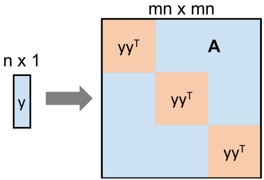
Observe that $\mathbf{s}^T A \mathbf{s} = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{S}_{i,j}\mathbf{S}_{i,k} y_j y_k = \|\mathbf{S}y\|_2^2$ and that

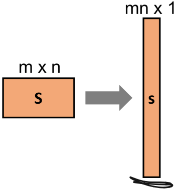
$$\operatorname{tr}(A) = m \cdot \operatorname{tr}(yy^T) =$$

Theorem (Hanson-Wright Inequality)

Let $\mathbf{x} \in \mathbb{R}^n$ *be a vector of i.i.d. random* ± 1 *values. For any matrix* $A \in \mathbb{R}^{n \times n}$,

$$\Pr[|\mathbf{x}^T A \mathbf{x} - \operatorname{tr}(A)| \geq t] \leq 2 \exp\left(-c \cdot \min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_2}\right\}\right).$$





$$\operatorname{tr}(yy^T) = \operatorname{tr}(\underbrace{y^T y}_{\|y\|_2^2})$$

Observe that $\mathbf{s}^T A \mathbf{s} = \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^n \mathbf{S}_{i,j} \mathbf{S}_{i,k} y_j y_k = \|\mathbf{S}y\|_2^2$ and that

$$\operatorname{tr}(A) = m \cdot \operatorname{tr}(yy^T) = m \cdot \|y\|_2^2.$$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\Pr[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon] = \Pr[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon]$$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned} \Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\ &= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \end{aligned}$$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$
\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq \epsilon m\right]
\end{aligned}
$$

$\mathrm{tr}(A) = m \cdot \|y\|_2^2$

$= m$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}$$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}$$

$= \mathrm{tr}(yy^T)$

$\|A\|_F^2 = m \cdot \|yy^T\|_F^2 = \|y\|_2^4 \cdot m$

$\qquad\quad \|$

$\quad \|y\|_F^2 \cdot \|y\|_F^2$

$$m \text{ copies} \left\{ \begin{bmatrix} yy^T & & \\ & yy^T & \\ & & \ddots \end{bmatrix} \right.$$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}$$

$\|A\|_F^2 = m \cdot \|yy^T\|_F^2$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$
\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}
$$

$\|A\|_F^2 = m \cdot \|yy^T\|_F^2 = m \cdot \|y\|_2^4 = m$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \operatorname{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}$$

$\|A\|_F^2 = m \cdot \|yy^T\|_F^2 = m \cdot \|y\|_2^2 = m$

$\|A\|_2 =$

$$\begin{bmatrix} yy^T & \\ & yy^T \end{bmatrix}$$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}$$

$\|A\|_F^2 = m \cdot \|yy^T\|_F^2 = m \cdot \|y\|_2^2 = m$

$\|A\|_2 = \|yy^T\|_2$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \operatorname{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\underbrace{\|A\|_F^2}_{m}}, \frac{\epsilon m}{\underbrace{\|A\|_2}_{1}}\right\}\right).
\end{aligned}$$

$\|A\|_F^2 = m \cdot \|yy^T\|_F^2 = m \cdot \|y\|_2^2 = m$

$\|A\|_2 = \|yy^T\|_2 = \|y\|_2^2 = 1$

$\|M\|_2 = \max_{x : \|x\|_2 = 1} \|Mx\|_2 = \sigma_1(M)$

$\|yy^T\| = \max_{x : \|x\| = 1} \|yy^T x\|_2 = \|y\|_2 \cdot \langle y, x \rangle$

$\|yy^T \frac{y}{\|y\|_2}\|_2 = \|y\|_2 \cdot \langle y, \frac{y}{\|y\|_2} \rangle = \|y\|_2^2$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}$$

$$\|A\|_F^2 = m \cdot \|yy^T\|_F^2 = m \cdot \|y\|_2^2 = m$$

$$\frac{\epsilon^2 m^2}{m} = \epsilon^2 m$$

$$\|A\|_2 = \|yy^T\|_2 = \|y\|_2 = 1$$

$$\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] \leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{m}, \frac{\epsilon m}{1}\right\}\right) = 2 \exp(-c\epsilon^2 m)$$

Distributional JL via Wright Inequality

Let $\mathbf{x} = \sqrt{m} \cdot \mathbf{s}$, so $\mathbf{x}$ has i.i.d. ± 1 entries. Assume w.l.o.g. that $\|y\|_2 = 1$.

$$\begin{aligned}
\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] &= \Pr\left[\left|\mathbf{s}^T A \mathbf{s} - 1\right| \geq \epsilon\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - m\right| \geq \epsilon m\right] \\
&= \Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \operatorname{tr}(A)\right| \geq \epsilon m\right] \\
&\leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{\|A\|_F^2}, \frac{\epsilon m}{\|A\|_2}\right\}\right).
\end{aligned}$$

$\|A\|_F^2 = m \cdot \|yy^T\|_F^2 = m \cdot \|y\|_2^2 = m$

$= 2\exp(\log(\delta))$

$\|A\|_2 = \|yy^T\|_2 = \|y\|_2 = 1$

$2\exp(-\log(1/\delta))$

$$\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] \leq 2 \exp\left(-c \cdot \min\left\{\frac{(\epsilon m)^2}{m}, \frac{\epsilon m}{1}\right\}\right) = 2\exp(-c\epsilon^2 m)$$

If we set $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, $\Pr\left[\left|\|\mathbf{S}y\|_2^2 - 1\right| \geq \epsilon\right] \leq \delta$, giving the distributional JL lemma.

Questions?