COMPSCI 614: Randomized Algorithms with Applications to Data Science

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2024.

Lecture 15

Logistics

- I'll release the weekly quiz later this afternoon. Due Monday as usual.
- I'll also release Pset 4 shortly.
- 2 page project progress report due 4/16.

Summary

Subspace Embedding:

- Given $A \in \mathbb{R}^{n \times d}$, want $S \in \mathbb{R}^{m \times n}$ such that $\|SAx\|_2 \approx \|Ax\|_2$ for all x. I.e., $\|Sy\|_2 \approx \|y\|_2$ for all $y \in col(A)$. Want $m \ll n$.
- For a single y, we can apply the Johnson-Lindenstrauss Lemma. Here, we want to preserve the norms of infinite y.
- Proof via Johnson-Lindenstrauss Lemma and ϵ-net argument.

Summary

Subspace Embedding:

- Given $A \in \mathbb{R}^{n \times d}$, want $S \in \mathbb{R}^{m \times n}$ such that $\|SAx\|_2 \approx \|Ax\|_2$ for all x. I.e., $\|Sy\|_2 \approx \|y\|_2$ for all $y \in col(A)$. Want $m \ll n$.
- For a single y, we can apply the Johnson-Lindenstrauss Lemma. Here, we want to preserve the norms of infinite y.
- Proof via Johnson-Lindenstrauss Lemma and ϵ-net argument.

Today:

- Finish the subspace embedding proof.
- Prove the Johnson-Lindenstrauss lemma itself via the Hanson-Wright inequality.
- Possibly give a simple application of subspace embedding to fast linear regression.

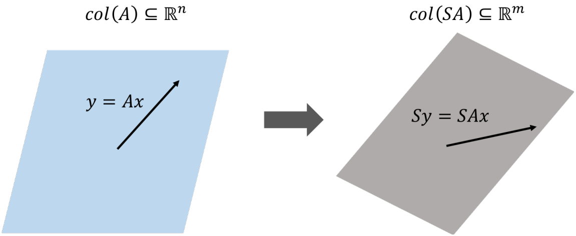
Subspace Embedding

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

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

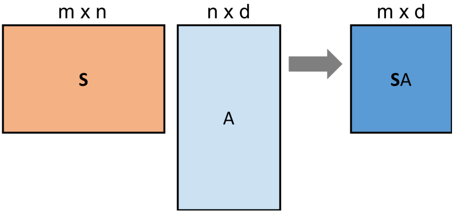
I.e., S preserves the norm of any vector Ax in the column span of A.



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)

Proof Outline

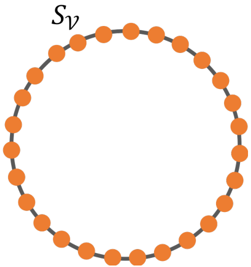
1. **Distributional Johnson-Lindenstrauss**: For $\mathbf{S} \in \mathbb{R}^{m \times d}$ with i.i.d. $\pm 1/\sqrt{m}$ entries, for any fixed $y \in \mathbb{R}^n$, with probability $1 - \delta$ for very small δ, $(1 - \epsilon)\|y\|_2 \leq \|\mathbf{S}y\|_2 \leq (1 + \epsilon)\|y\|_2$.
2. Via a union 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.

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

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



Proof last class via volume argument.

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

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

Proof Via ϵ-net

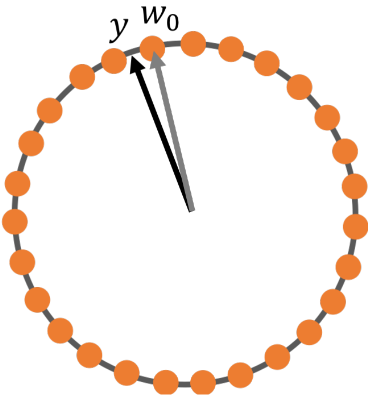
So Far: If we set $m = \tilde{O}(d/\epsilon^2)$ and pick random $\mathbf{S} \in \mathbb{R}^{m \times n}$, then with probability $\geq 1 - \delta$, $\|\mathbf{S}w\|_2 \approx_\epsilon \|w\|_2$ for all $w \in \mathcal{N}_\epsilon$.

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 $w \in \mathcal{N}_\epsilon$.

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

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



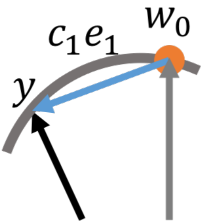
Proof Via ϵ-net

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:

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

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

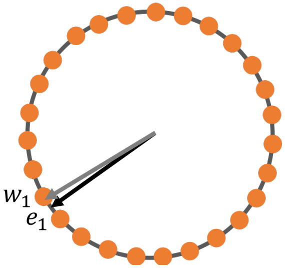


Proof Via ϵ-net

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:

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



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:

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

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:

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

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:

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

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

$$\begin{aligned}\|\mathbf{S}y\|_2 &= \|\mathbf{S}w_0 + c_1\mathbf{S}w_1 + c_2\mathbf{S}w_2 + \ldots\|_2 \\ &\leq \|\mathbf{S}w_0\|_2 + c_1\|\mathbf{S}w_1\|_2 + c_2\|\mathbf{S}w_2\|_2 + \ldots\end{aligned}$$

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:

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

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:

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

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

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.

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).

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).

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.
- 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)

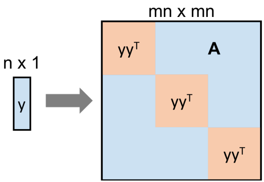
Let $\mathbf{x} \in \mathbb{R}^n$ *be a vector of i.i.d. random* ± 1 *values. For any matrix* $A \in \mathbb{R}^{n \times n}$*,*

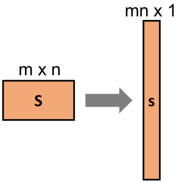
$$\Pr\left[\left|\mathbf{x}^T A \mathbf{x} - \mathrm{tr}(A)\right| \geq t\right] \leq 2 \exp\left(-c \cdot \min\left\{\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|_2}\right\}\right).$$

Theorem (Hanson-Wright Inequality)

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

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



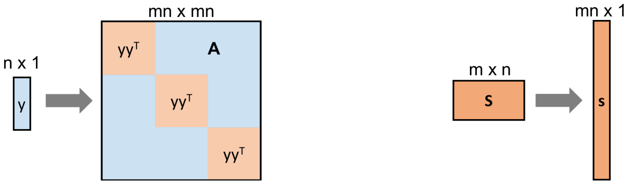


Hanson Wright Inequality

Theorem (Hanson-Wright Inequality)

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

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



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

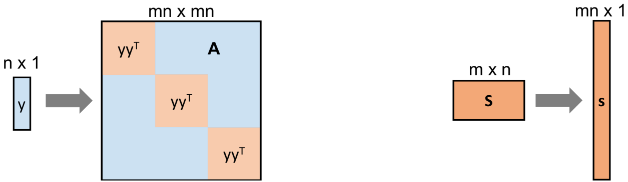
$$\mathrm{tr}(A) = m \cdot \mathrm{tr}(yy^T)$$

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* $A \in \mathbb{R}^{n\times n}$,

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



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 $\|y\|_2 = 1$.

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

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 $\|y\|_2 = 1$.

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

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 $\|y\|_2 = 1$.

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

$\|A\|_F^2 =$

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 $\|y\|_2 = 1$.

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

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

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 $\|y\|_2 = 1$.

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

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

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 $\|y\|_2 = 1$.

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

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

$\|A\|_2 =$

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 $\|y\|_2 = 1$.

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

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

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

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 $\|y\|_2 = 1$.

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

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

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

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 $\|y\|_2 = 1$.

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

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

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

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

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 $\|y\|_2 = 1$.

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

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

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

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

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.

Application to Linear Regression

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:*

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

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:*

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

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:

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

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

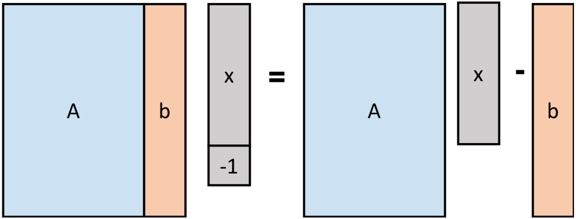
Claim: Since S is a subspace embedding for $[A; b]$, for all $x \in \mathbb{R}^d$,

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

Sketched Regression Proof

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

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



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:

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

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:

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

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:

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