CSE 392: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin
Spring 2024

Lecture 8: Sketching, types of sketching matrices

Outline

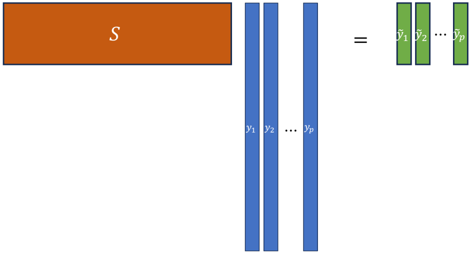
Recall: Embeddings

Embedding

A matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ is an ϵ-embedding of set $\mathcal{P} \subset \mathbb{R}^n$ if, for all $\boldsymbol{y} \in \mathcal{P}$,

$$\|\boldsymbol{S}\boldsymbol{y}\|_2 = (1 \pm \epsilon)\|\boldsymbol{y}\|_2.$$

We will call $\boldsymbol{S}$ a "sketching matrix".



Gaussian sketching matrix

Vector embedding also known as Distributional JL Lemma.

Distributional JL

Let $\boldsymbol{S} \in \mathbb{R}^{m \times d}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$. If $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for *any vector* $\boldsymbol{y} \in \mathbb{R}^d, \epsilon \in (0,1]$, with probability $(1-\delta)$:

$$\|\boldsymbol{S}\boldsymbol{y}\|_2^2 = (1 \pm \epsilon)\|\boldsymbol{y}\|_2^2.$$

Gaussian sketching matrix

Vector embedding also known as Distributional JL Lemma.

Distributional JL

Let $\boldsymbol{S} \in \mathbb{R}^{m \times d}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$. If $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then for *any vector* $\boldsymbol{y} \in \mathbb{R}^d, \epsilon \in (0,1]$, with probability $(1-\delta)$:

$$\|\boldsymbol{S}\boldsymbol{y}\|_2^2 = (1 \pm \epsilon)\|\boldsymbol{y}\|_2^2.$$

Proof: We know that $\mathbb{E}[\|\boldsymbol{S}\boldsymbol{y}\|_2^2] = \|\boldsymbol{y}\|_2^2$. We have

$$\|\boldsymbol{S}\boldsymbol{y}\|_2^2 = \frac{1}{m}\sum_{i=1}^{m}(\langle \boldsymbol{s}_i, \boldsymbol{y}\rangle)^2 = \frac{1}{m}\sum_{i=1}^{m}(\mathcal{N}(0, \|\boldsymbol{y}\|_2^2))^2$$

Chi-squared random variable with m degrees of freedom.

Chi-squared random variable

Let z be a Chi-squared random variable with m degrees of freedom

$$\Pr\{|\mathrm{z} - \mathbb{E}[\mathrm{z}]| \geq \epsilon \mathbb{E}[\mathrm{z}]\} \leq 2\exp(-\epsilon^2 m/8).$$

We have $\mathbb{E}[\mathrm{z}] = \mathbb{E}[\|\boldsymbol{S}\boldsymbol{y}\|_2^2] = \|\boldsymbol{y}\|_2^2$.

So, setting $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, we obtain the result.

Gaussian - JL property

JL Lemma

Let $\boldsymbol{S} \in \mathbb{R}^{m \times d}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$. If $m = O\left(\frac{\log(n)}{\epsilon^2}\right)$, then for any set of n data points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$, with probability at least 9/10:

$$(1-\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq \|\boldsymbol{S}\boldsymbol{x}_i - \boldsymbol{S}\boldsymbol{x}_j\|_2 \leq (1+\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$

Gaussian - JL property

JL Lemma

Let $\boldsymbol{S} \in \mathbb{R}^{m \times d}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$. If $m = O\left(\frac{\log(n)}{\epsilon^2}\right)$, then for any set of n data points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$, with probability at least $9/10$:

$$(1-\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq \|\boldsymbol{S}\boldsymbol{x}_i - \boldsymbol{S}\boldsymbol{x}_j\|_2 \leq (1+\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$

Proof: Fix $i, j \in [d]$, let $\boldsymbol{y} = \boldsymbol{x}_i - \boldsymbol{x}_j$. By the Distributional JL Lemma, with probability $1-\delta$:

$$\|\boldsymbol{S}(\boldsymbol{x}_i - \boldsymbol{x}_j)\|_2^2 = (1 \pm \epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2.$$

Set $\delta = 1/n^2$. Since there are $< n^2$ total i, j pairs, by a union bound we have that with probability $9/10$, the above will hold for *all* i, j, for:

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

Gaussian - Subspace embedding

Subspace embedding

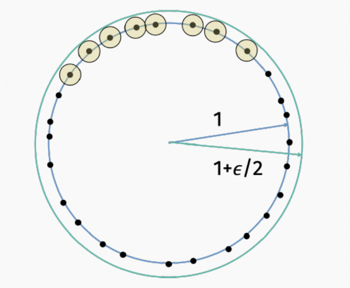
Let $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}} \mathcal{N}(0,1)$. If $m = O\left(\frac{d \log(1/\delta)}{\epsilon^2}\right)$, then for a given $\boldsymbol{A} \in \mathbb{R}^{n \times d}$, with probability at least $1-\delta$:

$$\|\boldsymbol{S}\boldsymbol{A}\boldsymbol{x}\|_2 = (1 \pm \epsilon)\|\boldsymbol{A}\boldsymbol{x}\|_2.$$

Embedding a d-dimensional subspace $\mathcal{U} \equiv span(\boldsymbol{A}) = span(\boldsymbol{U}) \subset \mathbb{R}^n$.

$$\|\boldsymbol{S}\boldsymbol{U}\boldsymbol{x}\|_2 = (1 \pm \epsilon)\|\boldsymbol{x}\|_2 \quad \text{or} \quad \|\boldsymbol{U}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{U} - \boldsymbol{I}\|_2 \leq \epsilon$$

Recall the ϵ-Net argument.



We know $|\mathcal{N}(\epsilon)| \leq (1 + \frac{2}{\epsilon})^d$.
If $\boldsymbol{S}$ is distributional JL with failure probability δ', taking union of the ϵ-net size, we get the result, with

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

AMM to embedding

Given $\boldsymbol{A} \in \mathbb{R}^{n\times d}$ with $n \geq d, \operatorname{rank}(\boldsymbol{A}) = r$; $\boldsymbol{B} \in \mathbb{R}^{d'\times n}$; $\epsilon, \delta > 0$. Let $\boldsymbol{S}$ be chosen (with oblivious distribution) such that, with probability at least $1-\delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq \epsilon \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F.$$

Then, $\boldsymbol{S}$ is an $\epsilon * r$-embedding of $span(\boldsymbol{A})$.

AMM to embedding

Given $\boldsymbol{A} \in \mathbb{R}^{n\times d}$ with $n \geq d, \operatorname{rank}(\boldsymbol{A}) = r$; $\boldsymbol{B} \in \mathbb{R}^{d'\times n}; \epsilon, \delta > 0$. Let $\boldsymbol{S}$ be chosen (with oblivious distribution) such that, with probability at least $1 - \delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top\boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq \epsilon\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F.$$

Then, $\boldsymbol{S}$ is an $\epsilon * r$-embedding of $span(\boldsymbol{A})$.

Set $\boldsymbol{B} = \boldsymbol{A}^\top$, and since $\boldsymbol{S}$ is oblivious, let us assume $\boldsymbol{A}$ is orthonormal. Then,

$$\|\boldsymbol{A}^\top\boldsymbol{S}^\top\boldsymbol{S}\boldsymbol{A} - \boldsymbol{I}\|_2 \leq \|\boldsymbol{A}^\top\boldsymbol{S}^\top\boldsymbol{S}\boldsymbol{A} - \boldsymbol{I}\|_F \leq \epsilon\|\boldsymbol{A}\|_F^2 = \epsilon r.$$

JL moment property

JL moment

A distribution on $\boldsymbol{S} \in \mathbb{R}^{m \times d}$, has the (ϵ, δ, ℓ)-JL moment property if for all $\boldsymbol{y} \in \mathbb{R}^d$ with $\|\boldsymbol{y}\|_2 = 1$,

$$\mathbb{E}_{\boldsymbol{S}}[|\|\boldsymbol{S}\boldsymbol{y}\|_2^2 - 1|^\ell] \leq \epsilon^\ell \cdot \delta.$$

JL moment property

JL moment

A distribution on $\boldsymbol{S} \in \mathbb{R}^{m\times d}$, has the (ϵ, δ, ℓ)-JL moment property if for all $\boldsymbol{y} \in \mathbb{R}^d$ with $\|\boldsymbol{y}\|_2 = 1$,

$$\mathbb{E}_{\boldsymbol{S}}[|\|\boldsymbol{S}\boldsymbol{y}\|_2^2 - 1|^\ell] \leq \epsilon^\ell \cdot \delta.$$

For $\ell = 2$, and if $\mathbb{E}[\|\boldsymbol{S}\boldsymbol{y}\|_2] = 1$ we have

$$\mathrm{Var}(\|\boldsymbol{S}\boldsymbol{y}\|_2^2) \leq \epsilon^2\delta \quad \text{or} \quad sd(\|\boldsymbol{S}\boldsymbol{y}\|_2^2) \leq \epsilon\sqrt{\delta}.$$

JL moment and AMM

Given $\boldsymbol{A} \in \mathbb{R}^{n \times d}$; $\boldsymbol{B} \in \mathbb{R}^{d' \times n}$; $\epsilon, \delta > 0$. Let $\boldsymbol{S}$ satisfy the (ϵ, δ, ℓ)-JL moment property for $\ell \geq 2$, then with probability at least $1 - \delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq 3\epsilon \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F.$$

JL moment and AMM

Given $\boldsymbol{A} \in \mathbb{R}^{n \times d}$; $\boldsymbol{B} \in \mathbb{R}^{d' \times n}$; $\epsilon, \delta > 0$. Let $\boldsymbol{S}$ satisfy the (ϵ, δ, ℓ)-JL moment property for $\ell \geq 2$, then with probability at least $1 - \delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq 3\epsilon \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F.$$

Proof: We follow proof of Theorem 2.8 in Dr. Woodruff's monograph.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\frac{\langle \boldsymbol{S}\boldsymbol{x}, \boldsymbol{S}\boldsymbol{y} \rangle}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2} =$$

JL moment and AMM

Given $\boldsymbol{A} \in \mathbb{R}^{n \times d}$; $\boldsymbol{B} \in \mathbb{R}^{d' \times n}$; $\epsilon, \delta > 0$. Let $\boldsymbol{S}$ satisfy the (ϵ, δ, ℓ)-JL moment property for $\ell \geq 2$, then with probability at least $1 - \delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq 3\epsilon \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F.$$

Proof: We follow proof of Theorem 2.8 in Dr. Woodruff's monograph.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\frac{\langle \boldsymbol{S}\boldsymbol{x}, \boldsymbol{S}\boldsymbol{y} \rangle}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2} =$$

Minkowski's inequality : $\|\|\mathrm{x} + \mathrm{y}\|\|_p \leq \|\|\mathrm{x}\|\|_p + \|\|\mathrm{y}\|\|_p$.

For unit vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$⦀\langle \boldsymbol{Sx}, \boldsymbol{Sy} \rangle - \langle \boldsymbol{x}, \boldsymbol{y} \rangle⦀_\ell =$$

For unit vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\left\|\!\left| \langle \boldsymbol{Sx}, \boldsymbol{Sy} \rangle - \langle \boldsymbol{x}, \boldsymbol{y} \rangle \right|\!\right\|_\ell =$$

Define a random variable

$$X_{ij} = \frac{1}{\|\boldsymbol{B}_{i*}\|_2 \|\boldsymbol{A}_{*j}\|_2} (\langle \boldsymbol{SB}_{i*}, \boldsymbol{SA}_{*j} \rangle - \langle \boldsymbol{B}_{i*}, \boldsymbol{A}_{*j} \rangle)$$

Then,

$$\left\|\!\left| \|\boldsymbol{BS}^\top \boldsymbol{SA} - \boldsymbol{BA}\|_F^2 \right|\!\right\|_{\ell/2} =$$

For unit vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$\left\|\left|\langle \boldsymbol{S}\boldsymbol{x}, \boldsymbol{S}\boldsymbol{y}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle\right|\right\|_\ell =$

Define a random variable

$$X_{ij} = \frac{1}{\|\boldsymbol{B}_{i*}\|_2 \|\boldsymbol{A}_{*j}\|_2}(\langle \boldsymbol{S}\boldsymbol{B}_{i*}, \boldsymbol{S}\boldsymbol{A}_{*j}\rangle - \langle \boldsymbol{B}_{i*}, \boldsymbol{A}_{*j}\rangle)$$

Then,

$\left|\!\left|\!\left| \|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F^2 \right|\!\right|\!\right|_{\ell/2} =$

Using

$$\mathbb{E}\|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F^\ell = \left|\!\left|\!\left| \|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F^2 \right|\!\right|\!\right|_{\ell/2}^{\ell/2},$$

amd Markov's inequality we get the result.

Gaussian sketch and AMM

Given $\boldsymbol{A} \in \mathbb{R}^{n \times d}$; $\boldsymbol{B} \in \mathbb{R}^{d' \times n}$; $\epsilon, \delta > 0$.
Let $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$ and $m = O(\epsilon^{-2}\delta^{-1})$, then with probability at least $1 - \delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq \epsilon \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F.$$

Gaussian sketch and AMM

Given $\boldsymbol{A} \in \mathbb{R}^{n\times d}$; $\boldsymbol{B} \in \mathbb{R}^{d'\times n}$; $\epsilon, \delta > 0$.
Let $\boldsymbol{S} \in \mathbb{R}^{m\times n}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$ and $m = O(\epsilon^{-2}\delta^{-1})$, then with probability at least $1-\delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top\boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq \epsilon\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F.$$

For Gaussian sketch, with $\ell = 2$, JL moment is

$$\mathrm{Var}(\|\boldsymbol{S}\boldsymbol{y}\|_2^2) \leq 2/m.$$

Since $\mathrm{Var}(\frac{1}{m}\chi_m^2) = \frac{1}{m^2}\mathrm{Var}(\chi_m^2) = 2m/m^2 = 2/m$.

We set $2/m \leq \epsilon^2\delta/6$.

Gaussian sketch and AMM

Given $\boldsymbol{A} \in \mathbb{R}^{n \times d}$; $\boldsymbol{B} \in \mathbb{R}^{d' \times n}$; $\epsilon, \delta > 0$.
Let $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$ and $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, then with probability at least $1-\delta$:

$$\|\boldsymbol{B}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{A}\|_F \leq \epsilon \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F.$$

Consider $\ell = \Theta(\log(1/\delta))$. Then, the ℓ-th central moment of χ^2_m is of the form $2^\ell(c_1 m^{\ell/2} + c_2)$. So, if we choose $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, we have:

$$\frac{2^\ell}{m^{\ell/2}} = \epsilon^\ell 2^{\ell/2}(2/\ell)^{\ell/2} \leq \epsilon^\ell \delta.$$

Two approaches

We have seen two approaches to go from vector embeddings to subspace embeddings.
Let $\|\boldsymbol{U}\|_F^2 = d, \text{rank}(\boldsymbol{A})$.

- **Using ϵ-nets:**

$$\Pr[\|\boldsymbol{U}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{U} - \boldsymbol{I}\|_{\mathcal{N}_\epsilon} \geq \epsilon] \leq C^d e^{-m\epsilon^2}$$

$$\implies \Pr[\|\boldsymbol{U}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{U} - \boldsymbol{I}\|_2 \geq 2\epsilon] \leq C^d e^{-m\epsilon^2}$$

- **using JL moment:**

$$\left\|\frac{1}{d}\|\boldsymbol{U}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{U} - \boldsymbol{I}\|_2\right\|_\ell^\ell \leq \epsilon^\ell \delta$$

$$\implies \Pr[\frac{1}{d}\|\boldsymbol{U}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{U} - \boldsymbol{I}\|_2 \geq \epsilon] \leq \delta$$

SHRT: Subsampled Randomized Hadamard Transform

Original JL:

- $\boldsymbol{S}$ is picked to be random matrix (orthogonal columns), i.i.d entries.
- Computing $\boldsymbol{SA}$ takes $O(mnd)$ time.

Faster scheme: pick a random orthogonal matrix, but:

- fewer random bits.
- faster to apply.

Fast JL: Using Subsampled Randomized Hadamard Transform (SRHT)

SRHT

The SRHT is a matrix $\boldsymbol{PHD}$, where

- $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ is diagonal matrix with i.i.d ± 1 on diagonal
- $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is a Hadamard matrix
- $\boldsymbol{P} \in \mathbb{R}^{m\times n}$ uniformly samples the rows of $\boldsymbol{HD}$.

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & -1 & 1 & \cdots & -1 \\ 1 & 1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \cdots & 1 \end{bmatrix}}_{\sqrt{n}\boldsymbol{H}} \underbrace{\begin{bmatrix} \pm 1 & 0 & 0 & \cdots & 0 \\ 0 & \pm 1 & 0 & \cdots & 0 \\ 0 & 0 & \pm 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \pm 1 \end{bmatrix}}_{\boldsymbol{D}}$$

Hadamard matrices

Hadamard matrices have recursive structure.

- Let $\boldsymbol{H}_0 \in \mathbb{R}^{1\times 1}$ be $[1]$.
- Let $\boldsymbol{H}_{i+1} = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{H}_i & \boldsymbol{H}_i \\ \boldsymbol{H}_i & -\boldsymbol{H}_i \end{bmatrix}$ for $i \geq 0$.

So,

$$\boldsymbol{H}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \boldsymbol{H}_2 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

In general, $\boldsymbol{H}_k$ is $2^k \times 2^k$ matrix with ± 1 entries scaled by $1/2^{k/2}$.

Hadamard properties

- Hadamard matrices are orthogonal.

$$\boldsymbol{H}_i^\top \boldsymbol{H}_i = \boldsymbol{H}_i^2 = \boldsymbol{I}.$$

- For any $\boldsymbol{x} \in \mathbb{R}^n, n = 2^k$, we have $\|\boldsymbol{H}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$, also $\|\boldsymbol{H}\boldsymbol{D}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$.
- Matvecs $\boldsymbol{H}\boldsymbol{x}$ can be computed in $O(n \log n)$ time for $\boldsymbol{x} \in \mathbb{R}^n, n = 2^k$.

Suppose $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} \in \mathbb{R}^{2^k}$, where $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^{2^{k-1}}$.

Then, $\boldsymbol{H}_{i+1}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{H}_i\boldsymbol{x}_1 + \boldsymbol{H}_i\boldsymbol{x}_2 \\ \boldsymbol{H}_i\boldsymbol{x}_1 - \boldsymbol{H}_i\boldsymbol{x}_2 \end{bmatrix}$.
So, we can compute $\boldsymbol{H}_{i+1}\boldsymbol{x}$ in linear time from $\boldsymbol{H}_i\boldsymbol{x}_1, \boldsymbol{H}_i\boldsymbol{x}_2$.

Randomized Hadamard analysis

SHRT mixing lemma

Let $\boldsymbol{H}$ be an $(n \times n)$ Hadamard matrix and $\boldsymbol{D}$ a random ± 1 diagonal matrix. Let $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$. With probability $1 - \delta$, for all i simultaneously,

$$z_i^2 \leq \frac{c \log(n/\delta)}{n} \|\boldsymbol{z}\|_2^2.$$

for some fixed constant c.

The vector is very close to uniform with high probability.

$$\|\boldsymbol{z}\|_2^2 = \|\boldsymbol{H}\boldsymbol{D}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2.$$

Randomized Hadamard analysis

z_i is a random variable with mean 0 and variance $\|\boldsymbol{x}\|_2^2/n$, which is a sum of independent random variables.

Can apply Bernstein type concentration inequality to prove the bound:

Rademacher Concentration

Let $r_1, \dots, r_n$ be Rademacher random variables (i.e. uniform ± 1's). Then for any vector $\boldsymbol{a} \in \mathbb{R}^n$,

$$\Pr\left[\sum_{i=1}^{n} r_i a_i \geq t\|\boldsymbol{a}\|_2\right] \leq e^{-t^2/2}$$

$z_i = \boldsymbol{h}_i^\top \boldsymbol{D}\boldsymbol{x}$ and let $\boldsymbol{h}_i^\top \boldsymbol{D} = \frac{1}{\sqrt{n}}[r_1, r_2, \dots, r_n]$, where r_i's are random ± 1's.

$t = \sqrt{\log(n/\delta)}$ and apply union bounds over all n entries.

Fast JL

The Fast JL Lemma

Let $\boldsymbol{S} = \boldsymbol{PHD} \in \mathbb{R}^{m\times n}$ be a subsampled randomized Hadamard transform with $m = O\left(\frac{\log(n/\delta)\log(1/\delta)}{\epsilon^2}\right)$. Then for any fixed $\boldsymbol{x} \in \mathbb{R}^n$. With probability $1-\delta$,

$$\|\boldsymbol{Sx}\|_2^2 = (1 \pm \epsilon)\|\boldsymbol{x}\|_2^2.$$

Proof: Apply Hoeffding's inequality for the sum of m entries.

SRHT embeddings

SRHT - subspace embedding

For $\boldsymbol{S} = \boldsymbol{PHD} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{A} \in \mathbb{R}^{n \times d}$, if $m = O\left(\frac{d \log(n/\delta) \log(1/\delta)}{\epsilon^2}\right)$, then with probability at least $1 - \delta$:

$$\|\boldsymbol{SAx}\|_2 = (1 \pm \epsilon)\|\boldsymbol{Ax}\|_2.$$

We can compute the sketch $\boldsymbol{SA}$ in $O(mn \log(d))$ time.

Faster Embeddings: Countsketch

- Gaussian sketching matrix is goods, but is expensive to apply.
- SRHT is faster, but for dense matrices.
- **Sparse Embeddings:** Adaption of `CountSketch` from streaming algorithms.
- $\boldsymbol{S}$ is of the form:

$$\begin{bmatrix} 0 & -1 & 0 & 0 & \cdots & 0 \\ +1 & 0 & 0 & +1 & \cdots & 0 \\ 0 & 0 & -1 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

- One random ± 1 per column.
- Row $\boldsymbol{A}_{i*}$ of $\boldsymbol{A}$ contributes $\pm \boldsymbol{A}_{i*}$ to one of the rows of $\boldsymbol{SA}$.

Sparse Embeddings

- **Sparse sketching matrix:** For $i \in [n]$, pick uniformly and independently: $h_i \in [m]$, $s_i \in \{-1, +1\}$, and define $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ as:

$$\boldsymbol{S}_{h_i,i} \to s_i \text{ for } i \in [n],$$

  and $\boldsymbol{S}_{j,i} \to 0$ otherwise.
- $\boldsymbol{s}$ is a sign (Radamacher) vector. The vector $\boldsymbol{h}$ hashes to m "hash buckets". That is,

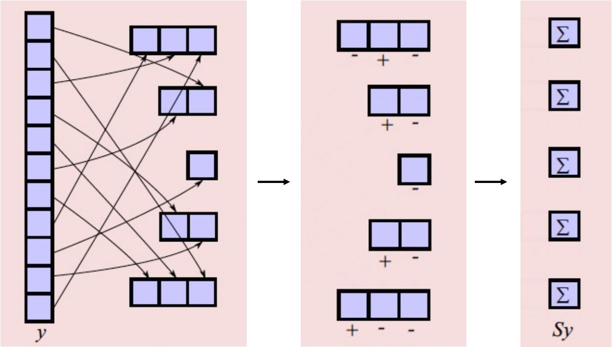
$$\boldsymbol{S}_{j*} = \sum_{i:h_i=j} s_i \boldsymbol{e}_i^\top,$$

  and so

$$[\boldsymbol{S}\boldsymbol{A}]_{j*} = \sum_{i:h_i=j} s_i \boldsymbol{e}_i^\top \boldsymbol{A} = \sum_{i:h_i=j} s_i \boldsymbol{A}_{i*}.$$

- **Fast sketching:** Can compute $\boldsymbol{S}\boldsymbol{A}$ in $O(nnz(\boldsymbol{A}))$ time.

- If $\boldsymbol{s}$ is a sign (Radamacher) vector, then $\mathbb{E}[\boldsymbol{s}^\top \boldsymbol{y}] = \|\boldsymbol{y}\|_2^2$.
- For $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, each row of $\boldsymbol{S}$:
  (a) collects a subset of entries y_i's; (b) applies the signs, and (c) adds
- $\mathbb{E}[\|\boldsymbol{S}\boldsymbol{y}\|_2^2] = \|\boldsymbol{y}\|_2^2$.



Analysis of sparse embeddings

Variance of Countsketch

For $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\boldsymbol{y} \in \mathbb{R}^n$ a unit vector,

$$\mathrm{Var}[\|\boldsymbol{S}\boldsymbol{y}\|_2^2] \leq \frac{3}{m}.$$

Analysis of sparse embeddings

Variance of Countsketch

For $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\boldsymbol{y} \in \mathbb{R}^n$ a unit vector,

$$\mathrm{Var}[\|\boldsymbol{S}\boldsymbol{y}\|_2^2] \leq \frac{3}{m}.$$

Proof: Let $\boldsymbol{z} = \boldsymbol{S}\boldsymbol{y}$. We have $\mathbb{E}[\|\boldsymbol{z}\|_2^2] =$

$\mathrm{Var}[\|\boldsymbol{z}\|_2^2] =$

Variance of Countsketch

For $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\boldsymbol{y} \in \mathbb{R}^n$ a unit vector,

$$\mathrm{Var}[\|\boldsymbol{S}\boldsymbol{y}\|_2^2] \leq \frac{3}{m}.$$

Proof: Let $\boldsymbol{z} = \boldsymbol{S}\boldsymbol{y}$. We have $\mathbb{E}[\|\boldsymbol{z}\|_2^2] =$

$\mathrm{Var}[\|\boldsymbol{z}\|_2^2] =$

$\mathbb{E}[\|\boldsymbol{z}\|_2^4] =$

Analysis of sparse embeddings

Variance of Countsketch

For $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\boldsymbol{y} \in \mathbb{R}^n$ a unit vector,

$$\mathrm{Var}[\|\boldsymbol{S}\boldsymbol{y}\|_2^2] \leq \frac{3}{m}.$$

Proof: Let $\boldsymbol{z} = \boldsymbol{S}\boldsymbol{y}$. We have $\mathbb{E}[\|\boldsymbol{z}\|_2^2] =$

$\mathrm{Var}[\|\boldsymbol{z}\|_2^2] =$

$\mathbb{E}[\|\boldsymbol{z}\|_2^4] =$

$\mathbb{E}_{s,h}[z_j^4] =$

Countsketch Embedding

Countsketch - subspace embedding

For $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ a countsketch matrix and $\boldsymbol{A} \in \mathbb{R}^{n \times d}$, if $m = O\left(\frac{d^2}{\delta \epsilon^2}\right)$, then with probability at least $1 - \delta$:

$$\|\boldsymbol{SAx}\|_2 = (1 \pm \epsilon)\|\boldsymbol{Ax}\|_2.$$

Countsketch Embedding

Countsketch - subspace embedding

For $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ a countsketch matrix and $\boldsymbol{A} \in \mathbb{R}^{n \times d}$, if $m = O\left(\frac{d^2}{\delta \epsilon^2}\right)$, then with probability at least $1 - \delta$:

$$\|\boldsymbol{SAx}\|_2 = (1 \pm \epsilon)\|\boldsymbol{Ax}\|_2.$$

We use the AMM and JL moment result.

We have $\mathrm{Var}[\|\boldsymbol{Sy}\|_2^2] \leq \frac{K}{m}$.

If $\frac{K}{m} \leq \epsilon^2 \delta$, we know $\boldsymbol{S}$ is ϵd-embedding with probability at least $1 - \delta$.

Types of sketching matrices

Sketching matrix	Sketch size m	Cost to sketch $\boldsymbol{SA}$
JL - i.i.d subGaussians	$m = O\left(\frac{d\log(1/\delta)}{\epsilon^2}\right)$	$O(mnd)$
Fast JL -SRHT	$m = O\left(\frac{d\log(d)\log(1/\delta)}{\epsilon^2}\right)$	$O(mn\log(d))$
Countsketch	$m = O\left(\frac{d^2}{\delta\epsilon^2}\right)$	$O(nnz(\boldsymbol{A}))$

We have other sparse embeddings where nnz per column is > 1, e..g, OSNAPs, sparse graphs.

Can improve $m = O\left(\frac{d\log(d)\log(1/\delta)}{\epsilon^2}\right)$ with $s = \Theta(\log(1/\delta))$ nonzero entries per column.

Further Reading:

- Woodruff, David P. "Sketching as a tool for numerical linear algebra." Foundations and Trends® in Theoretical Computer Science 10.1–2 (2014): 1-157.
- Kane, Daniel M., and Jelani Nelson. "Sparser johnson-lindenstrauss transforms." Journal of the ACM (JACM) 61.1 (2014): 1-23.
- Tropp, Joel A. "Improved analysis of the subsampled randomized Hadamard transform." Advances in Adaptive Data Analysis 3.01n02 (2011): 115-126.

Questions?