EE270
Large scale matrix computation, optimization and learning

Instructor : Mert Pilanci

Stanford University

Thursday, Jan 28 2020

Randomized Linear Algebra
Lecture 6: Johnson Lindenstrauss Lemma and Applications

JL - Lemma

Dimension Reduction

- map a high dimensional vector to low dimensions such that certain properites are preserved

Diag.

- examples so far:
- Approximate Matrix Multiplication $AS^T SB \approx AB$ where S is random

$SB^{(1)}\ SB^{(2)}$

- Freivalds Algorithm $ABr - Mr$ where r is random
- Trace estimation $r^T Mr \approx \mathbf{tr}(M)$ where r is random

Dimension Reduction

- map a high dimensional vector to low dimensions such that certain properites are preserved
- examples so far:
- Approximate Matrix Multiplication $AS^TSB \approx AB$ where S is random
- Freivalds Algorithm $ABr - Mr$ where r is random
- Trace estimation $r^TMr \approx \mathbf{tr}(M)$ where r is random

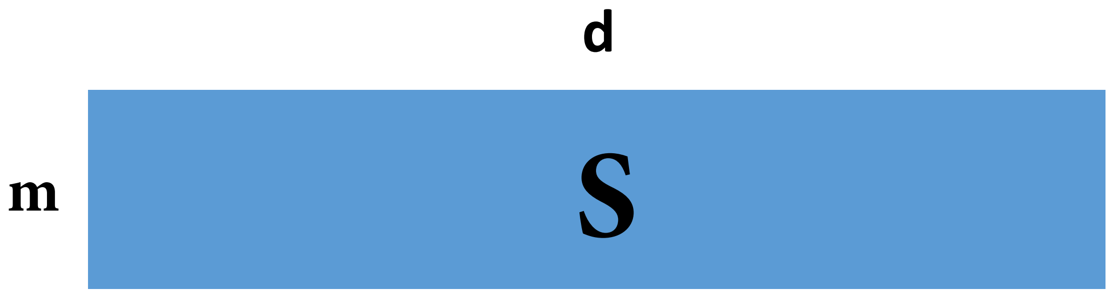
x^T

n samples x_1

- Generic dimension reduction problem
- Given vectors $x_1, \ldots, x_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $y_1, \ldots, y_n \in \mathbb{R}^m$ where $m \ll d$
- another instance is Principal Component Analysis

Randomized Dimension Reduction

- Given vectors $x_1, \ldots, x_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $y_1, \ldots, y_n \in \mathbb{R}^m$ where $m < d$

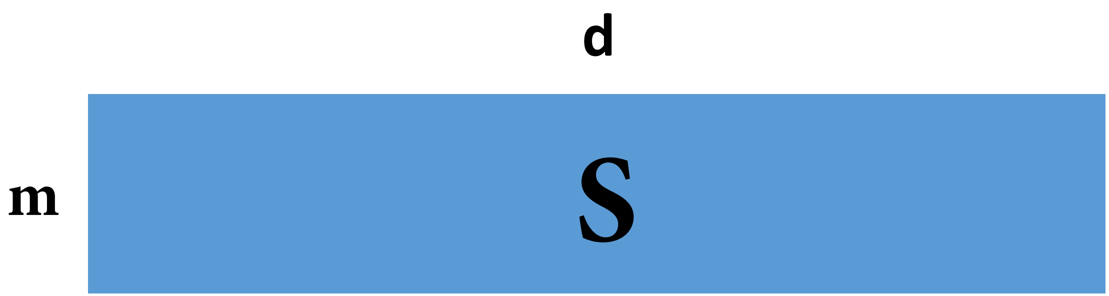


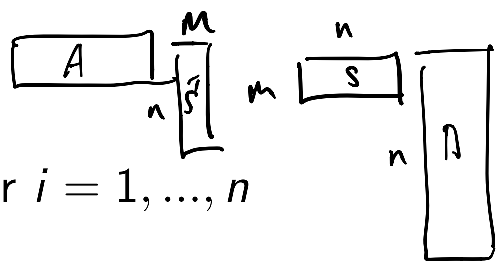
- Linear transformation $y_i = Sx_i$ for $i = 1, \ldots, n$
- S is chosen randomly

$O(m \cdot d)$

Randomized Dimension Reduction

- Given vectors $x_1, \ldots, x_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $y_1, \ldots, y_n \in \mathbb{R}^m$ where $m < d$

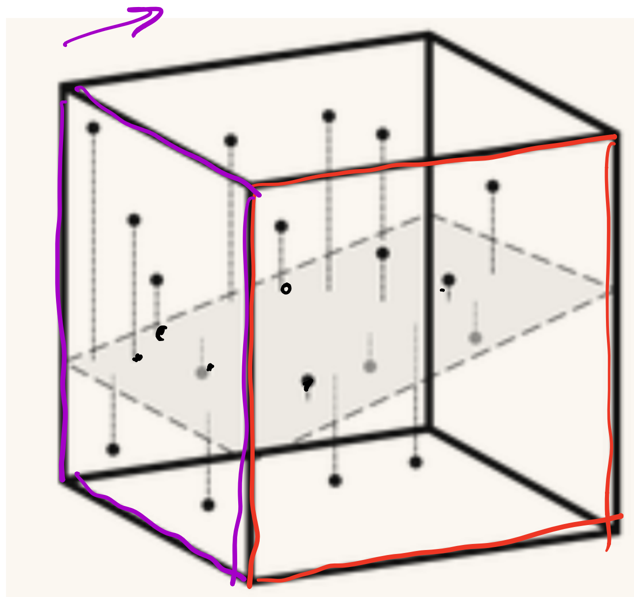




- Linear transformation $y_i = Sx_i$ for $i = 1, \ldots, n$
- S is chosen randomly
- Approximate Matrix Multiplication: $AS^TSB \approx AB$ where S is random matrix

Geometry of Random Projections

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ \text{[illegible]} \end{bmatrix}$$



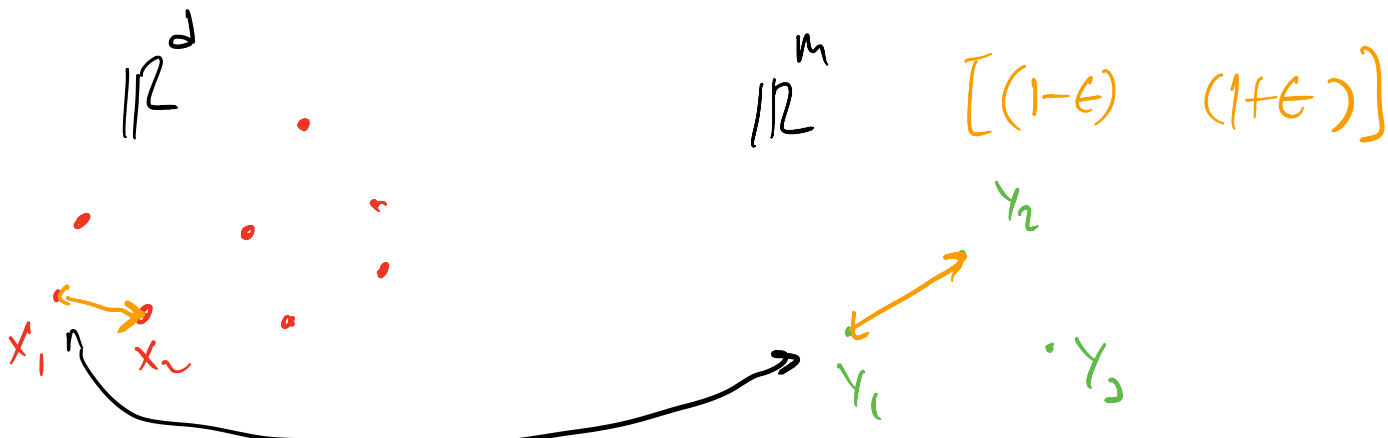
Johnson Lindenstrauss Lemma

$y_i = S(x_i) = Sx_i$

- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^d \to \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that

$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon$$

$(1-\epsilon) \cdot \|x_i - x_j\|_2^2 \leq \|y_i - y_j\|_2^2 \leq (1+\epsilon) \cdot \|x_i - x_j\|_2^2$



Johnson Lindenstrauss Lemma

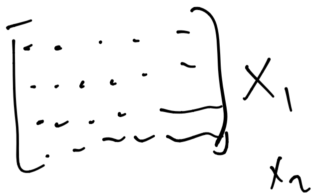
- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^d \to \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that

$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon$$

- Note that the target dimension m is **independent of the original dimension** d, and depends **only on the number of points** n and the accuracy parameter.

Johnson Lindenstrauss Lemma $1-\epsilon \leq \left\| \frac{S(x_i-x_j)}{\|x_i-x_j\|_2} \right\|_2^2 \leq 1+\epsilon$

- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^d \to \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that



$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon$$

- Note that the target dimension m is **independent of the original dimension** d, and depends **only on the number of points** n and the accuracy parameter.
- more surprises: picking an $m \times d$ random matrix $S = \frac{1}{\sqrt{m}} G$ with $G_{ij} \sim N(0,1)$ standard normal works with high probability!

Sx

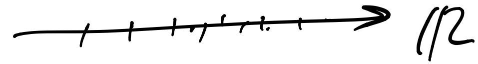
Johnson Lindenstrauss (JL) Lemma

- Define $u_{ij} \triangleq \frac{x_i - x_j}{\|x_i - x_j\|_2}$.
- note that $\|u_{ij}\|_2 = 1$
- JL Lemma:
  $\mathbb{P}\big[\|Su_{ij}\|_2^2 \in (1 \pm \epsilon) \text{ for all } i, j \in \{1, ..., n\}\big] \geq 1 - \delta$

  where $\delta \in (0, 1)$ for large enough m

Warm-up: Geometry of Concentration of Measure on the Sphere

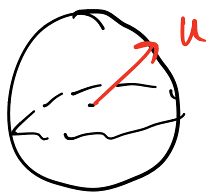
$S =$



- Suppose $m = 1$, i.e., we project to dimension one
- S is a uniformly random row vector on the sphere, i.e., $S = \frac{g^T}{\|g\|_2}$ where $g \sim N(0, I)$
- For any fixed unit norm vector u, how large is the product Su?
- Su is distributed identically to $Se_1 = S_1 = \frac{g_1}{\|g\|_2}$ where e_1 is the first ordinary basis vector

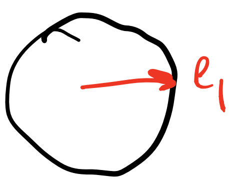
$g \overset{(d)}{=} Rg$

$$Su = \frac{g^T u}{\|g\|_2} \overset{(d)}{=} \frac{g^T Ru}{\|\cancel{R}g\|_2}$$



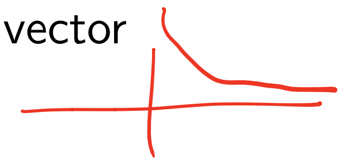
identical dist

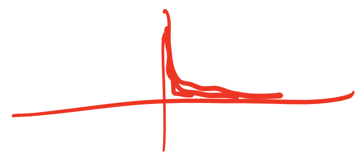
$$Se_1 = S_1 = \frac{g_1}{\|g\|_2} = \frac{g_1}{\sqrt{\sum g_i^2}}$$



Warm-up: Geometry of Concentration of Measure on the Sphere

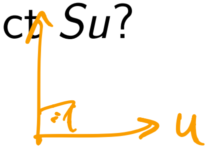
- Suppose $m = 1$, i.e., we project to dimension one
- S is a uniformly random row vector on the sphere, i.e., $S = \frac{g^T}{\|g\|_2}$ where $g \sim N(0, I)$
- For any fixed unit norm vector u, how large is the product Su?
- Su is distributed identically to $Se_1 = S_1 = \frac{g_1}{\|g\|_2}$ where e_1 is the first ordinary basis vector
- **Lemma:** $\mathbb{P}\left[|S_1| \geq t\right] \leq 2e^{-\frac{dt^2}{2}}$.



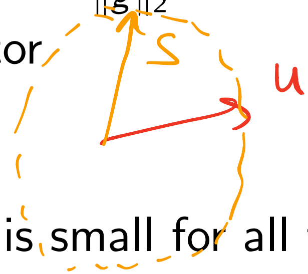


Warm-up: Geometry of Concentration of Measure on the Sphere

- Suppose $m = 1$, i.e., we project to dimension one
- S is a uniformly random row vector on the sphere, i.e., $S = \frac{g^T}{\|g\|_2}$ where $g \sim N(0, I)$
- For any fixed unit norm vector u, how large is the product Su?



- Su is distributed identically to $Se_1 = S_1 = \frac{g_1}{\|g\|_2}$ where e_1 is the first ordinary basis vector
- **Lemma:** $\mathbb{P}\left[|S_1| \geq t\right] \leq 2e^{-\frac{dt^2}{2}}$.

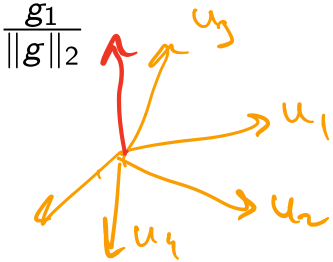


- The inner product $Su = \left(\frac{g}{\|g\|_2}\right)^T \frac{u}{\|u\|_2}$ is small for all fixed directions $\frac{u}{\|u\|_2}$, implying near-orthogonality

Warm-up: Geometry of Concentration of Measure on the Sphere

- Suppose $m = 1$, i.e., we project to dimension one
- S is a uniformly random row vector on the sphere, i.e., $S = \frac{g^T}{\|g\|_2}$ where $g \sim N(0, I)$
- For any fixed unit norm vector u, how large is the product Su?
- Su is distributed identically to $Se_1 = S_1 = \frac{g_1}{\|g\|_2}$ where e_1 is the first ordinary basis vector
- **Lemma:** $\mathbb{P}\left[|S_1| \geq t\right] \leq 2e^{-\frac{dt^2}{2}}$.
- The inner product $Su = \left(\frac{g}{\|g\|_2}\right)^T \frac{u}{\|u\|_2}$ is small for all fixed directions $\frac{u}{\|u\|_2}$, implying near-orthogonality
- For n fixed vectors $u_1, \dots, u_n$, we can apply union bound

$n \leq e^{\frac{d\varepsilon^2}{2}}$



$\log(2n) < \frac{d\varepsilon^2}{2}$

$\max_{i=1,\dots,n} x_i \geq t$

$\Updownarrow$

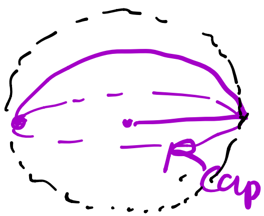
$\exists i$ s.t. $x_i \geq t$

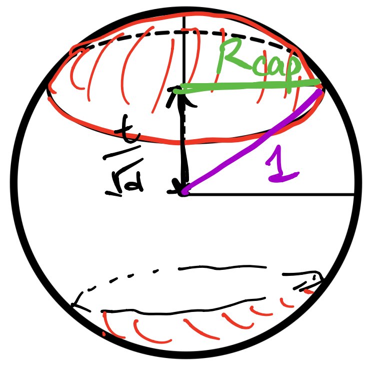
$$\mathbb{P}\left[\max_{i=1,\dots,n} |Su_i| \geq t\right] \leq \sum_{i=1}^{n} 2e^{-\frac{dt^2}{2}} = 2ne^{-\frac{dt^2}{2}} = e^{\overbrace{-\frac{d\varepsilon^2}{2} + \log(2n)}}$$

$$\leq \mathbb{P}\left[\bigcup_{i=1,\dots,n} \{|Su_i| \geq t\}\right] \leq \sum \mathbb{P}\left[|Su_i| \geq t\right]$$

Concentration of Measure for Uniform Distribution on the Sphere

- **Lemma (rephrased):** $\mathbb{P}\left[\left|\frac{g_1}{\|g\|_2}\right| \geq \frac{t}{\sqrt{d}}\right] \leq 2e^{-\frac{t^2}{2}}$.
- Note that $\frac{g}{\|g\|_2}$ is distributed uniformly on the unit sphere

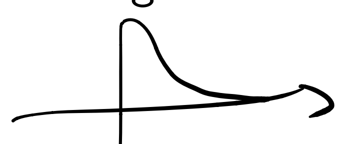




Surface area of sphere in $\mathbb{R}^d$ radius R $= R^{d-1}$

- Pythagorean theorem: $\frac{t^2}{d} + R_{\text{cap}}^2 = 1$ implies $R_{\text{cap}} = \sqrt{1 - \frac{t^2}{d}}$
- $\mathbb{P}\left[\left|\frac{g_1}{\|g\|_2}\right| \geq \frac{t}{\sqrt{d}}\right] \leq \frac{\text{area of the spherical cap}}{\text{area of the sphere}} \leq \frac{\left(\sqrt{1-\frac{t^2}{d}}\right)^{d-1}}{\cancel{1^{d-1}}\ 1}$
- using the fact $(1 - \frac{x}{n})^n \leq e^{-x}$ we get

$$\mathbb{P}\left[\left|\frac{g_1}{\|g\|_2}\right| \geq \frac{t}{\sqrt{d}}\right] \leq 2e^{-\frac{t^2}{2}} .$$



Proof of JL Lemma

- We need to show $\|Su_{ij}\|_2^2$ is concentrated around 1
- **Lemma** Let $S = \frac{1}{\sqrt{m}} G \in \mathbb{R}^{m \times n}$ where $G_{ij} \sim N(0, 1)$ and u be any fixed vector. Then

$$\mathbb{E}\|Su\|_2^2 = \|u\|_2^2$$

- implies that the distance between two points is preserved in expectation
- **Proof:**

$$\mathbb{E} \left\| \frac{1}{\sqrt{m}} G u \right\|_2^2 = \mathbb{E} \frac{1}{m} u^T G^T G u$$

$$= \mathbb{E} \frac{1}{m} u^T \cdot \sum_{k=1}^{m} G_{(k)} G_{(k)}^T \cdot u$$

$$= \frac{1}{m} u^T \cdot \underbrace{\sum_{k=1}^{m} \underbrace{\mathbb{E}\, G_{(k)} G_{(k)}^T}_{I}}_{m \cdot I} \cdot u$$

$$= \frac{1}{\cancel{m}} u^T \cdot \cancel{m} I u = u^T u$$

$$\mathbb{E} \begin{bmatrix} G_{(k),1} \cdot G_{(k),1} & \cdots \\ \cancel{G_{(k),1} \cdot G_{(k),2}} & \cdot \end{bmatrix}$$

Proof of JL Lemma

- Set $S = \frac{1}{\sqrt{m}} G$ where $G \in \mathbb{R}^{m \times d}$ and $G_{ij} \sim N(0,1)$
- Consider the probability that $\|Su\|_2^2$ deviates from 1, i.e., projected vectors are stretched more than their expectation

we first show that $\mathbb{P}\left[\|Su\|_2^2 \geq (1+\epsilon)\|u\|_2^2\right] \leq e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

$$\mathbb{P}\left(\frac{1}{m}\|Gu\|_2^2 > (1+\epsilon)\right) = \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m}(G_{(i)}^T u)^2 > 1+\epsilon\right)$$

$$= \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m}\Big(\underbrace{\sum_{j=1}^{d} G_{ij}u_j}_{z_i}\Big)^2 > 1+\epsilon\right)$$

$$z_i \triangleq \sum_{j=1}^{d} G_{ij}u_j$$

$N(0,1)$ Gaussian distributed

$$= \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} z_i^2 > 1+\epsilon\right)$$

$$\mathbb{E}\, z_i = \mathbb{E}\sum_{j=1}^{d} G_{ij}u_j$$

$$= \sum (\mathbb{E}\, G_{ij})u_j$$

$$= 0$$

$$\mathrm{Var}(z_i) = \sum_j \mathrm{Var}(\underbrace{G_{ij}u_j}_{N(0,u_j^2)}) = \sum_{j=1}^{d} u_j^2 = 1$$

Proof of JL Lemma

- Set $S = \frac{1}{\sqrt{m}} G$ where $G \in \mathbb{R}^{m \times d}$ and $G_{ij} \sim N(0,1)$
- Consider the probability that $\|Su\|_2^2$ deviates from 1, i.e., projected vectors are stretched more than their expectation

we first show that $\mathbb{P}\left[\|Su\|_2^2 \geq (1+\epsilon)\|u\|_2^2\right] \leq e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

$$\mathbb{P}\left[\frac{\lambda}{m}\sum_{i=1}^{m} z_i^2 > (1+\epsilon)\cdot\lambda\right] = \mathbb{P}\left[e^{\frac{\lambda}{m}\sum_{i=1}^{m} z_i^2} > e^{(1+\epsilon)\cdot\lambda}\right]$$

Apply Markov's Inequality $\leq \dfrac{\mathbb{E}\, e^{\frac{\lambda}{m}\sum_{i=1}^{m} z_i^2}}{e^{(1+\epsilon)\cdot\lambda}}$

$$= \mathbb{E}\prod_{i=1}^{m} e^{\lambda \cdot z_i^2} \cdot e^{-(1+\epsilon)\lambda \cdot m} = \prod_{i=1}^{m} \mathbb{E}\left[e^{\lambda z_i^2}\right] \cdot e^{-(1+\epsilon)\cdot\lambda m}$$

$\lambda \leftarrow \lambda \cdot m$

Moment Generating Function of $z_i^2 \cdot \lambda$

Proof of JL Lemma

$$\int e^{\lambda \cdot z^2} \cdot p(z)\, dz = \sqrt{\frac{1}{1-2\lambda}}$$

- Set $S = \frac{1}{\sqrt{m}} G$ where $G \in \mathbb{R}^{m \times d}$ and $G_{ij} \sim N(0,1)$
- Consider the probability that $\|Su\|_2^2$ deviates from 1, i.e., projected vectors are stretched more than their expectation

  we first show that $\mathbb{P}\left[\|Su\|_2^2 \geq (1+\epsilon)\|u\|_2^2\right] \leq e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

- $\mathbb{P}\left[\|Su_{ij}\|_2^2 \geq (1+\epsilon)\|u_{ij}\|_2^2\right] \leq \sum_{i,j} e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}} = n^2 e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

$$\leq \left(\frac{1}{1-2\lambda}\right)^{m/2} e^{-(1+\epsilon)\cdot\lambda m} \leq e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$$

pick

$$\lambda^* = \frac{\epsilon}{2(1+\epsilon)}$$

Proof of JL Lemma

- Set $S = \frac{1}{\sqrt{m}} G$ where $G \in \mathbb{R}^{m \times d}$ and $G_{ij} \sim N(0,1)$
- Consider the probability that $\|Su\|_2^2$ deviates from 1, i.e., projected vectors are stretched more than their expectation

  we first show that $\mathbb{P}\left[\|Su\|_2^2 \geq (1+\epsilon)\|u\|_2^2\ \right] \leq e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

  union bound

- $\mathbb{P}\left[\|Su_{ij}\|_2^2 \geq (1+\epsilon)\|u_{ij}\|_2^2\ \right] \leq \sum_{i,j} e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}} = n^2 e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

  Set error probability $= \frac{1}{2} = \cancel{n^2} e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}} + \log(n^2)$
- $m = \frac{9 \log n}{\epsilon^2 - \epsilon^3}$

Proof of JL Lemma

- Set $S = \frac{1}{\sqrt{m}} G$ where $G \in \mathbb{R}^{m \times d}$ and $G_{ij} \sim N(0,1)$
- Consider the probability that $\|Su\|_2^2$ deviates from 1, i.e., projected vectors are stretched more than their expectation

  we first show that $\mathbb{P}\left[\|Su\|_2^2 \geq (1+\epsilon)\|u\|_2^2\right] \leq e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

- $\mathbb{P}\left[\|Su_{ij}\|_2^2 \geq (1+\epsilon)\|u_{ij}\|_2^2\right] \leq \sum_{i,j} e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}} = n^2 e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$

  Set error probability $= \frac{1}{2} = n^2 e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$
- $m = \frac{9 \log n}{\epsilon^2 - \epsilon^3}$

  for smaller error probability $0.01 = n^2 e^{-(\epsilon^2-\epsilon^3)\frac{m}{4}}$
- $m = \frac{\text{constant} \times \log n}{\epsilon^2 - \epsilon^3}$

True 'projections': random subspaces also work

- Pick $S_{(i)}$ uniformly random on the unit sphere
- Pick $S_{(i+1)}$ uniformly random on the unit sphere and $\perp S_{(i)}, ...S_{(1)}$
- S is a projection matrix, which projects onto a uniformly random subspace

$$\mathbb{P}\left\{ \left| \|Su\|_2 - \sqrt{\frac{m}{d}} \right| > t \right\} \leq 2e^{\frac{-t^2 d}{2}}$$

- Applying union bound for all points $i, j = 1, ..., d$ gives a similar result
- Random i.i.d. S matrices are easier to generate and approximately orthogonal: $\mathbb{E}S^T S = I$

Computationally cheaper random matrices

- Gaussian $S_{ij} = \frac{1}{\sqrt{m}} N(0,1)$
- Rademacher

$$S_{ij} = \begin{cases} +\frac{1}{m} & \text{with probability } \frac{1}{\sqrt{m}} \\ -\frac{1}{\sqrt{m}} & \text{with probability } \frac{1}{2} \end{cases} \tag{1}$$

- Bernoulli-Rademacher

$$S_{ij} = \begin{cases} +\frac{\sqrt{3}}{\sqrt{m}} & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{2}{3} \\ -\frac{\sqrt{3}}{\sqrt{m}} & \text{with probability } \frac{1}{2} \end{cases} \tag{2}$$

- other sparse matrices (e.g. one non-zero per column)
- Fourier transform based matrices

Optimality of the JL Embedding

- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^n \to \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that

$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon \qquad (\star)$$

- Can we embed to a **smaller dimension**?
- maybe using a **nonlinear** embedding?

Optimality of the JL Embedding

- Let $\epsilon \in (0, \frac{1}{2})$. Given any set of points $\{x_1, ..., x_n\}$ in $\mathbb{R}^d$, there exists a map $S : \mathbb{R}^n \to \mathbb{R}^m$ with $m = \frac{9\log(n)}{\epsilon^2 - \epsilon^3}$ such that

$$1 - \epsilon \leq \frac{\|Sx_i - Sx_j\|_2^2}{\|x_i - x_j\|_2^2} \leq 1 + \epsilon \qquad (\star)$$

- Can we embed to a **smaller dimension**?
- maybe using a **nonlinear** embedding?
- **No**

  Johnson-Lindenstrauss Embedding is optimal
- There exists a set of n points $\{x_1, ..., x_n\}$ such that any linear/nonlinear embedding satisfying $(\star)$ must have $m \geq O(\frac{\log n}{\epsilon^2})$.

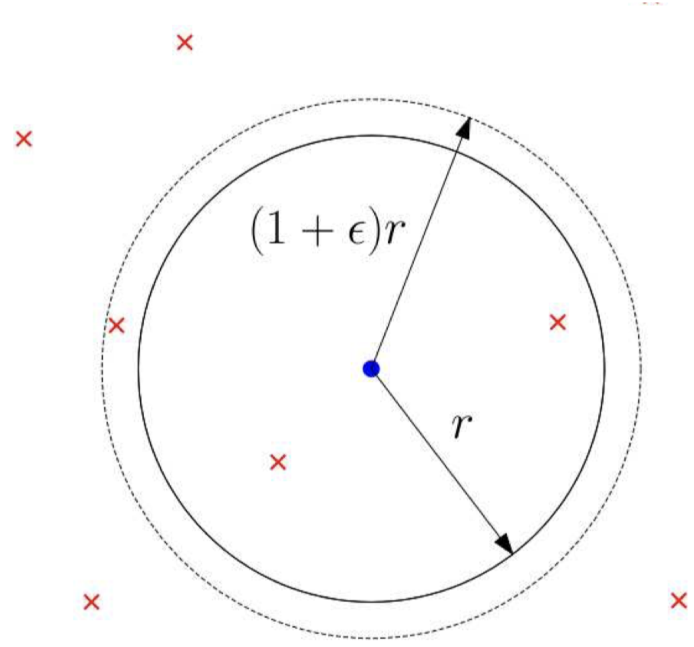
Optimality of the Johnson-Lindenstrauss Lemma, Larsen and Nelson, 2016

Applications of JL Embeddings

- General idea: run algorithms on $Sx_1, \ldots, Sx_n \in \mathbb{R}^m$ instead of $x_1, \ldots, x_n$
- Examples:
  - approximate nearest neighbor search
  - estimating norms and frequency moments
  - regression
  - classification
  - randomized matrix operations (matrix multiplication, decomposition etc)
  - optimization
  - ...

Approximate Nearest Neighbors

- Given a point set $P = \{x_1, ..., x_n\} \in \mathbb{R}^d$
- and a query point $q \in \mathbb{R}^d$
- Find an ϵ-approximate nearest neighbor to q from P



Estimating p-norms

- Streaming data

  $x_{t+1} = x_t + \delta_t$
- Estimate $\|x\|_2$
- second moment

Estimating p-norms

- Streaming data

  $x_{t+1} = x_t + \delta_t$
- Estimate $\|x\|_2$
- second moment
- **linear sketch**

  Generate S randomly such that $\mathbb{E}S^TS = I$

  Let $y_t = Sx_t$

  $y_t = Sx_t + S\delta_t$
- $\|Sy\|_2^2 \approx \|Sx\|_2^2$

Estimating p-norms

- Streaming data

  $x_{t+1} = x_t + \delta_t$
- Estimate $\|x\|_2$
- second moment
- **linear sketch**

  Generate S randomly such that $\mathbb{E}S^TS = I$

  Let $y_t = Sx_t$

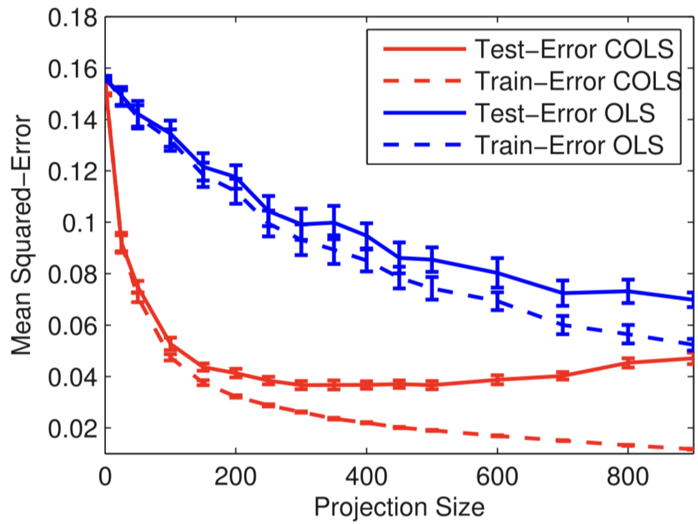
  $y_t = Sx_t + S\delta_t$
- $\|Sy\|_2^2 \approx \|Sx\|_2^2$
- Can also be extended to $\|x\|_p$

Music similarity prediction

- Predict the similarity score $\in [0, 1]$ between 30 second tracks
- Frequency based features from each 200ms segment results in 10^6 features
- OLS: randomly pick m features
- COLS: apply random projection to dimension m

Fard et al. Compressed Least-Squares Regression on Sparse Spaces, 2012

Music similarity prediction



- Predict the similarity score $\in [0, 1]$ between 30 second tracks
- Frequency based features from each 200ms segment results in 10^6 features
- OLS: randomly pick m features
- COLS: apply random projection to dimension m

Fard et al. Compressed Least-Squares Regression on Sparse Spaces, 2012

Random Projection for Approximate Matrix Multiplication

- matrices A and B

  output AS^TSB where S is a random projection matrix

Random Projection for Approximate Matrix Multiplication

- matrices A and B

  output AS^TSB where S is a random projection matrix
- need to characterize $\|Sx\|_2^2 - \|x\|_2^2$ for vectors x
- **Definition:** (ϵ, δ, p) JL moment property

$$\mathbb{E}\left|\|Sx\|_2^2 - 1\right|^p \leq \epsilon^p \delta$$

  for any unit norm x where $p \geq 2$

Random Projection for Approximate Matrix Multiplication

- matrices A and B

  output AS^TSB where S is a random projection matrix
- need to characterize $\|Sx\|_2^2 - \|x\|_2^2$ for vectors x
- **Definition:** (ϵ, δ, p) JL moment property

$$\mathbb{E}\left|\|Sx\|_2^2 - 1\right|^p \leq \epsilon^p \delta$$

  for any unit norm x where $p \geq 2$
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$random i.i.d. sub-Gaussian, e.g., ± 1, or $N(0,1)$ with $m = \frac{c_1}{\epsilon^2} \log \frac{1}{\delta}$ satisfies $(\epsilon, \delta, \log \frac{1}{\delta})$ JL moment property

Random Projection for Approximate Matrix Multiplication

- matrices A and B
  output AS^TSB where S is a random projection matrix
- need to characterize $\|Sx\|_2^2 - \|x\|_2^2$ for vectors x
- **Definition:** (ϵ, δ, p) JL moment property

$$\mathbb{E}\left|\|Sx\|_2^2 - 1\right|^p \leq \epsilon^p \delta$$

  for any unit norm x where $p \geq 2$
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$random i.i.d. sub-Gaussian, e.g., ± 1, or $N(0,1)$ with $m = \frac{c_1}{\epsilon^2} \log \frac{1}{\delta}$ satisfies $(\epsilon, \delta, \log \frac{1}{\delta})$ JL moment property
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$CountSketch matrix (one nonzero per column, which is ± 1 at a uniformly random location) with $m = \frac{c_2}{\epsilon^2 \delta}$ satisfies $(\epsilon, \delta, 2)$ JL moment property

Random Projection for Approximate Matrix Multiplication

- matrices A and B
  output AS^TSB where S is a random projection matrix
- need to characterize $\|Sx\|_2^2 - \|x\|_2^2$ for vectors x
- **Definition:** (ϵ, δ, p) JL moment property

$$\mathbb{E}\left|\|Sx\|_2^2 - 1\right|^p \leq \epsilon^p \delta$$

  for any unit norm x where $p \geq 2$
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$random i.i.d. sub-Gaussian, e.g., ± 1, or $N(0,1)$ with $m = \frac{c_1}{\epsilon^2} \log \frac{1}{\delta}$ satisfies $(\epsilon, \delta, \log \frac{1}{\delta})$ JL moment property
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$CountSketch matrix (one nonzero per column, which is ± 1 at a uniformly random location) with $m = \frac{c_2}{\epsilon^2 \delta}$ satisfies $(\epsilon, \delta, 2)$ JL moment property
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$Fast JL Transform with $m = \frac{c_3}{\epsilon} \log \frac{1}{\delta}$ satisfies $(\epsilon, \delta, \log \frac{n}{\delta})$ JL moment property

Approximating inner products

- **Lemma**

$$\mathbb{E}\left|\|Sx\|_2^2 - 1\right|^p \leq \epsilon^p \delta$$

for any unit norm x implies that

$$\mathbb{E}\left|x^T S^T S y - x^T y\right|^p \leq 3\epsilon^p \delta$$

since

$$x^T y = \frac{1}{2}\left(\|x\|_2^2 + \|y\|_2^2 - \|x - y\|_2^2\right)$$

$$x^T S^T S y = \frac{1}{2}\left(\|Sx\|_2^2 + \|Sy\|_2^2 - \|S(x - y)\|_2^2\right)$$

Random Projection for Approximate Matrix Multiplication

- Let $C = AS^TSB$

$$\mathbb{P}\left[\|AB - C\|_F > 3\epsilon\|A\|_F\|B\|_F\right] = \left[\|AB - C\|_F^p > (3\epsilon)^p\|A\|_F^p\|B\|_F^p\right]$$
$$\leq \frac{\mathbb{E}\|AB - C\|_F^p}{(3\epsilon\|A\|_F\|B\|_F)^p}$$

- Let $a_i = A_{(i)}$ and $b_i = B_{(i)}$

$$\|AB - C\|_F^2 = \sum_{ij}\left|(Sa_i)^T(Sb_j) - a_i^Tb_j\right|^2$$

Random Projection for Approximate Matrix Multiplication

- Let $C = AS^T SB$

$$\mathbb{P}\left[\|AB - C\|_F > 3\epsilon\|A\|_F\|B\|_F\right] = \left[\|AB - C\|_F^p > (3\epsilon)^p\|A\|_F^p\|B\|_F^p\right]$$
$$\leq \frac{\mathbb{E}\|AB - C\|_F^p}{(3\epsilon\|A\|_F\|B\|_F)^p}$$

- Let $a_i = A_{(i)}$ and $b_i = B_{(i)}$

$$\|AB - C\|_F^2 = \sum_{ij} \left|(Sa_i)^T(Sb_j) - a_i^T b_j\right|^2$$

- we can normalize $\frac{a_i}{\|a_i\|_2}, \frac{b_i}{\|b_i\|_2}$ and apply JL moment property to get

$$\mathbb{P}\left[\|AB - C\|_F > 3\epsilon\|A\|_F\|B\|_F\right] \leq \delta$$

Final error bound for random projection

- Let the approximate product of AB be $C = AS^TSB$

$$\mathbb{P}\left[\|AB - C\|_F > 3\epsilon\|A\|_F\|B\|_F\right] \leq \delta$$

- Follows from JL Moment property
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$random i.i.d. sub-Gaussian, e.g., ± 1, or $N(0, 1)$ with $m = \frac{c_1}{\epsilon^2} \log \frac{1}{\delta}$
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$CountSketch matrix (one nonzero per column, which is ± 1 at a uniformly random location) with $m = \frac{c_2}{\epsilon^2 \delta}$
- $S \in \mathbb{R}^{m \times n} \sim \frac{1}{\sqrt{m}} \times$Fast JL Transform with $m = \frac{c_3}{\epsilon} \log \frac{1}{\delta}$

Final error bound for random projection

- Let the approximate product of AB be $C = AS^TSB$

$$\mathbb{P}\left[\|AB - C\|_F > 3\epsilon\|A\|_F\|B\|_F\right] \leq \delta$$

- Follows from JL Moment property
- $S \in \mathbb{R}^{m\times n} \sim \frac{1}{\sqrt{m}} \times$random i.i.d. sub-Gaussian, e.g., ± 1, or $N(0,1)$ with $m = \frac{c_1}{\epsilon^2}\log\frac{1}{\delta}$
- $S \in \mathbb{R}^{m\times n} \sim \frac{1}{\sqrt{m}} \times$CountSketch matrix (one nonzero per column, which is ± 1 at a uniformly random location) with $m = \frac{c_2}{\epsilon^2\delta}$
- $S \in \mathbb{R}^{m\times n} \sim \frac{1}{\sqrt{m}} \times$Fast JL Transform with $m = \frac{c_3}{\epsilon}\log\frac{1}{\delta}$
- Sparse JL and Fast JL are more efficient
- advantages: doesn't require any knowledge about matrices A and B (**oblivious**)
- optimal sampling probabilities depend on the column/row norms of A and B

Questions?