JL Lemma, Dimensionality Reduction, and Subspace Embeddings

Lecture 11
September 29, 2020
AMS-$\ell_2$-Estimate:

Let $Y_1, Y_2, \ldots, Y_n$ be $\{-1, +1\}$ random variables that are 4-wise independent.

$z \leftarrow 0$

While (stream is not empty) do

$a_j = (i_j, \Delta_j)$ is current update

$z \leftarrow z + \Delta_j Y_{i_j}$

endWhile

Output $z^2$

Claim: Output estimates $||x||^2_2$ where $x$ is the vector at end of stream of updates.
Z = \sum_{i=1}^{n} x_i Y_i \text{ and output is } Z^2

Z^2 = \sum_{i} x_i^2 Y_i^2 + 2 \sum_{i \neq j} x_i x_j Y_i Y_j

\text{and hence}

\mathbb{E}[Z^2] = \sum_{i} x_i^2 = ||x||_2^2.

\text{One can show that } \text{Var}(Z^2) \leq 2(\mathbb{E}[Z^2])^2.
Linear Sketching View

Recall that we take average of independent estimators and take median to reduce error. Can we view all this as a sketch?

AMS-$\ell_2$-Sketch:

\[ k = c \log(1/\delta)/\epsilon^2 \]

Let \( M \) be a $\ell \times n$ matrix with entries in \( \{-1, 1\} \) s.t.

(i) rows are independent and

(ii) in each row entries are 4-wise independent

\( z \) is a $\ell \times 1$ vector initialized to 0

While (stream is not empty) do

\[ a_j = (i_j, \Delta_j) \text{ is current update} \]

\[ z \leftarrow z + \Delta_j M e_{i_j} \]

endWhile

Output vector \( z \) as sketch.

\( M \) is compactly represented via \( k \) hash functions, one per row, independently chosen from 4-wise independent hash family.
Geometric Interpretation

Given vector \( x \in \mathbb{R}^n \) let \( M \) the random map \( z = Mx \) has the following features

- \( \mathbb{E}[z_i] = 0 \) and \( \mathbb{E}[z_i^2] = ||x||_2^2 \) for each \( 1 \leq i \leq k \) where \( k \) is number of rows of \( M \)
- Thus each \( z_i^2 \) is an estimate of length of \( x \) in Euclidean norm
- When \( k = \Theta(\frac{1}{\epsilon^2 \log(1/\delta)}) \) one can obtain an \((1 \pm \epsilon)\) estimate of \( ||x||_2 \) by averaging and median ideas

Thus we are able to compress \( x \) into \( k \)-dimensional vector \( z \) such that \( z \) contains information to estimate \( ||x||_2 \) accurately
Geometric Interpretation

Given vector $x \in \mathbb{R}^n$ let $M$ the random map $z = Mx$ has the following features

- $E[z_i] = 0$ and $E[z_i^2] = \|x\|^2_2$ for each $1 \leq i \leq k$ where $k$ is number of rows of $M$
- Thus each $z_i^2$ is an estimate of length of $x$ in Euclidean norm
- When $k = \Theta(\frac{1}{\epsilon^2} \log(1/\delta))$ one can obtain an $(1 \pm \epsilon)$ estimate of $\|x\|_2$ by averaging and median ideas

Thus we are able to compress $x$ into $k$-dimensional vector $z$ such that $z$ contains information to estimate $\|x\|_2$ accurately

Question: Do we need median trick? Will averaging do?
Lemma (Distributional JL Lemma)

Fix vector $x \in \mathbb{R}^d$ and let $\Pi \in \mathbb{R}^{k \times d}$ matrix where each entry $\Pi_{ij}$ is chosen independently according to standard normal distribution $\mathcal{N}(0, 1)$ distribution. If $k = \Omega\left(\frac{1}{\epsilon^2} \log(1/\delta)\right)$, then with probability $(1 - \delta)$

$$(1 - \epsilon) \|x\|_2 \leq \|\frac{1}{\sqrt{k}} \Pi x\|_2 \leq (1 + \epsilon) \|x\|_2.$$

Can choose entries from $\{-1, 1\}$ as well.

Note: unlike $\ell_2$ estimation, entries of $\Pi$ are independent.

Letting $z = \frac{1}{\sqrt{k}} \Pi x$ we have projected $x$ from $d$ dimensions to $k = O\left(\frac{1}{\epsilon^2} \log(1/\delta)\right)$ dimensions while preserving length to within $(1 \pm \epsilon)$-factor.
\[
\begin{align*}
\mathbf{w} & = \begin{bmatrix}
- & - & - & - & - \\
- & - & - & - & - \\
\end{bmatrix} \\
& \in \mathbb{R}^{k \times n} \\
\mathbf{H}_k & \sim \mathcal{N}(0,I) \\
\mathbf{H}_k & \in \mathbb{R}^{k \times n} \\
\mathbf{z} & \in \mathbb{R}^{k} \\
\|\mathbf{z}\|_2 & \leq (1 \pm \epsilon) \|\mathbf{x}\|_2 \\
\text{with probability} & \geq (1-\delta).
\end{align*}
\]
**Theorem (Metric JL Lemma)**

Let $v_1, v_2, \ldots, v_n$ be any $n$ points/vectors in $\mathbb{R}^d$. For any $\epsilon \in (0, 1/2)$, there is linear map $f : \mathbb{R}^d \to \mathbb{R}^k$ where $k \leq 8 \ln n/\epsilon^2$ such that for all $1 \leq i < j \leq n$,

$$
(1 - \epsilon) \|v_i - v_j\|_2 \leq \|f(v_i) - f(v_j)\|_2 \leq \|v_i - v_j\|_2.
$$

Moreover $f$ can be obtained in randomized polynomial-time.

Linear map $f$ is simply given by random matrix $\Pi$: $f(v) = \Pi v$. 
Lemma: If you choose \( T \in \mathbb{R}^{k \times d} \) with 
\[ K = \frac{1}{\varepsilon^2} \ln \frac{1}{\delta} \] 
then for any fixed vector \( \tilde{x} \in \mathbb{R}^d \), 
\[ \frac{1}{\sqrt{2}} \| T \tilde{x} \|_2 \approx (1 \pm \varepsilon) \| \tilde{x} \|_2 \]

\[ \frac{\| \tilde{x} \|_2}{\sqrt{2}} \sum_{i=1}^{2} \| \tilde{v}_i - \tilde{v}_j \|_2 \] 

If we choose \( \delta = \frac{1}{n^2} \), then with probability \( 1 - \frac{1}{n^3} \) 
\[ \frac{1}{\sqrt{2}} \| T (\tilde{v}_i - \tilde{v}_j) \|_2 \approx (1 \pm \varepsilon) \| \tilde{v}_i - \tilde{v}_j \|_2 \]

By union bound, all \( \tilde{v}_i - \tilde{v}_j \) vectors are preserved with probability \( 1 - \left( \binom{n}{2} \right) \cdot \frac{1}{n^3} \) 
\[ \geq 1 - \frac{1}{n} \].
Theorem (Metric JL Lemma)

Let \( v_1, v_2, \ldots, v_n \) be any \( n \) points/vectors in \( \mathbb{R}^d \). For any \( \epsilon \in (0, 1/2) \), there is linear map \( f : \mathbb{R}^d \rightarrow \mathbb{R}^k \) where \( k \leq 8 \ln n / \epsilon^2 \) such that for all \( 1 \leq i < j \leq n \),

\[
(1 - \epsilon) \|v_i - v_j\|_2 \leq \|f(v_i) - f(v_j)\|_2 \leq \|v_i - v_j\|_2.
\]

Moreover \( f \) can be obtained in randomized polynomial-time.

Linear map \( f \) is simply given by random matrix \( \Pi \): \( f(v) = \Pi v \).

Proof.

Apply DJL with \( \delta = 1/n^2 \) and apply union bound to \( \binom{n}{2} \) vectors \( (v_i - v_j), i \neq j \).
DJL and Metric JL

Key advantage: mapping is *oblivious* to data!
Normal Distribution

Density function: \( f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \)

Standard normal: \( \mathcal{N}(0, 1) \) is when \( \mu = 0, \sigma = 1 \)
Normal Distribution

Cumulative density function for standard normal:
\[ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \text{ (no closed form)} \]
Sum of independent Normally distributed variables

**Lemma**

Let $X$ and $Y$ be independent random variables. Suppose $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Let $Z = X + Y$. Then $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. 

Sum of independent Normally distributed variables

Lemma
Let $X$ and $Y$ be independent random variables. Suppose $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Let $Z = X + Y$. Then $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Corollary
Let $X$ and $Y$ be independent random variables. Suppose $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$. Let $Z = aX + bY$. Then $Z \sim \mathcal{N}(0, a^2 + b^2)$.
Sum of independent Normally distributed variables

**Lemma**

Let $X$ and $Y$ be independent random variables. Suppose $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Let $Z = X + Y$. Then $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

**Corollary**

Let $X$ and $Y$ be independent random variables. Suppose $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$. Let $Z = aX + bY$. Then $Z \sim \mathcal{N}(0, a^2 + b^2)$.

Normal distribution is a *stable* distribution: adding two independent random variables within the same class gives a distribution inside the class. Others exist and useful in $F_p$ estimation for $p \in (0, 2)$.
Concentration of sum of squares of normally distributed variables

\( \chi^2(k) \) distribution: distribution of sum of \( k \) independent standard normally distributed variables

\( Y = \sum_{i=1}^{k} Z_i \) where each \( Z_i \sim \mathcal{N}(0, 1) \).
Concentration of sum of squares of normally distributed variables

$\chi^2(k)$ distribution: distribution of sum of $k$ independent standard normally distributed variables

$Y = \sum_{i=1}^{k} Z_i^2$ where each $Z_i \sim \mathcal{N}(0, 1)$. 


$E[Z_i] = 0$
Concentration of sum of squares of normally distributed variables

$\chi^2(k)$ distribution: distribution of sum of $k$ independent standard normally distributed variables

$Y = \sum_{i=1}^{k} Z_i$ where each $Z_i \sim \mathcal{N}(0, 1)$.


Lemma

Let $Z_1, Z_2, \ldots, Z_k$ be independent $\mathcal{N}(0, 1)$ random variables and let $Y = \sum_i Z_i^2$. Then, for $\epsilon \in (0, 1/2)$, there is a constant $c$ such that,

$$\Pr[(1 - \epsilon)^2 k \leq Y \leq (1 + \epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k}.$$
$\chi^2$ distribution

Density function

$f_k(x)$
$\chi^2$ distribution

Cumulative density function

$$F_k(x)$$

$$\chi^2_k$$

- $k=1$
- $k=2$
- $k=3$
- $k=4$
- $k=6$
- $k=9$
Concentration of sum of squares of normally distributed variables

\( \chi^2(k) \) distribution: distribution of sum of \( k \) independent standard normally distributed variables

Lemma

Let \( Z_1, Z_2, \ldots, Z_k \) be independent \( \mathcal{N}(0, 1) \) random variables and let \( Y = \sum_i Z_i^2 \). Then, for \( \epsilon \in (0, 1/2) \), there is a constant \( c \) such that, \( \Pr[(1 - \epsilon)^2 k \leq Y \leq (1 + \epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k} \).

Recall Chernoff-Hoeffding bound for \textit{bounded} independent non-negative random variables. \( Z_i^2 \) is not bounded, however Chernoff-Hoeffding bounds extend to sums of random variables with exponentially decaying tails.
Proof of DJL Lemma

Without loss of generality assume $\|x\|_2 = 1$ (unit vector)

$$Z_i = \sum_{j=1}^{n} \prod_{ij} x_i$$

- $Z_i \sim \mathcal{N}(0, 1)$

$$C(1 - \delta) \|x\|_2 \leq \left\| \frac{1}{\sqrt{k}} \prod \bar{x} \right\|_2 \leq C(1 + \delta) \|x\|_2$$

$$\prod_{ij} \sim \mathcal{N}(0, 1).$$

with $\mu_k = \frac{1}{\varepsilon^2} \ln \frac{1}{\delta}$
Proof of DJL Lemma

Without loss of generality assume \( \|x\|_2 = 1 \) (unit vector)

\[ Z_i = \sum_{j=1}^n \prod_{ij} x_j \]

- \( Z_i \sim \mathcal{N}(0, 1) \)

- Let \( Y = \sum_{i=1}^k Z_i^2 \). \( Y \)'s distribution is \( \chi^2 \) since \( Z_1, \ldots, Z_k \) are iid.
Proof of DJL Lemma

Without loss of generality assume \( \|x\|_2 = 1 \) (unit vector)

\[
Z_i = \sum_{j=1}^{n} \prod_{ij} x_i = \mathcal{N}(0, 1)
\]

\( Z_i \) is uniformly distributed as \( Z_i \sim \mathcal{N}(0, 1) \).

Let \( Y = \sum_{i=1}^{k} Z_i^2 \). \( Y \)'s distribution is \( \chi^2 \) since \( Z_1, \ldots, Z_k \) are iid.

Hence \( \Pr[(1 - \epsilon)^2 k \leq Y \leq (1 + \epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k} \)

\[
\frac{1 - 2e^{c\epsilon^2 k}}{\epsilon^2} \geq 1 - \delta
\]
Proof of DJL Lemma

Without loss of generality assume $\|x\|_2 = 1$ (unit vector)

$Z_i = \sum_{j=1}^{n} \Pi_{ij} x_i$

- $Z_i \sim \mathcal{N}(0, 1)$
- Let $Y = \sum_{i=1}^{k} Z_i^2$. $Y$’s distribution is $\chi^2$ since $Z_1, \ldots, Z_k$ are iid.
- Hence $\Pr[(1 - \epsilon)^2 k \leq Y \leq (1 + \epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k}$
- Since $k = \Omega(\frac{1}{\epsilon^2 \log(1/\delta)})$ we have $\Pr[(1 - \epsilon)^2 k \leq Y \leq (1 + \epsilon)^2 k] \geq 1 - \delta$
Proof of DJL Lemma

Without loss of generality assume $\|x\|_2 = 1$ (unit vector)

$$Z_i = \sum_{j=1}^n \prod_{ij} x_i$$

- $Z_i \sim \mathcal{N}(0, 1)$
- Let $Y = \sum_{i=1}^k Z_i^2$. $Y$’s distribution is $\chi^2$ since $Z_1, \ldots, Z_k$ are iid.
- Hence $\Pr[(1 - \epsilon)^2 k \leq Y \leq (1 + \epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k}$
- Since $k = \Omega(\frac{1}{\epsilon^2} \log(1/\delta))$ we have $\Pr[(1 - \epsilon)^2 k \leq Y \leq (1 + \epsilon)^2 k] \geq 1 - \delta$
- Therefore $\|z\|_2 = \sqrt{Y/k}$ has the property that with probability $(1 - \delta)$, $\|z\|_2 = (1 \pm \epsilon)\|x\|_2$. 
JL lower bounds

**Question:** Are the bounds achieved by the lemmas tight or can we do better? How about non-linear maps?

Essentially optimal modulo constant factors for worst-case point sets.

\[ n \text{ vectors in } \mathbb{R}^d \rightarrow n \text{ vectors in } \mathbb{R}^k \]

\[ k = \frac{\text{dim } n}{\varepsilon^2} \]

s.t. distances are preserved in \( \ell_2 \)-norm
Fast JL and Sparse JL

Projection matrix $\Pi$ is dense and hence $\Pi x$ takes $\Theta(kd)$ time.

**Question:** Can we find $\Pi$ to improve time bound?

Two scenarios: $x$ is dense and $x$ is sparse

$$\bar{v}_1, \bar{v}_2 \ldots \bar{v}_n \rightarrow v_1', v_2', \ldots, v_n'$$

$$\prod \bar{v}_i = v_i'$$

$\Theta(kd) \leq \frac{1}{\varepsilon^2} \ln n \cdot d$
Fast JL and Sparse JL

Projection matrix $\Pi$ is dense and hence $\Pi x$ takes $\Theta(kd)$ time.

**Question:** Can we find $\Pi$ to improve time bound?

Two scenarios: $x$ is dense and $x$ is sparse

**Known results:**

- Choose $\Pi_{ij}$ to be $\{-1, 0, 1\}$ with probability $1/6, 1/3, 1/6$. Also works. Roughly $1/3$ entries are $0$.
- Fast JL: Choose $\Pi$ in a dependent way to ensure $\Pi x$ can be computed in $O(d \log d + k^2)$ time. For dense $x$.
- Sparse JL: Choose $\Pi$ such that each column is $s$-sparse. The best known is $s = O(\frac{1}{\epsilon} \log(1/\delta))$. Helps in sparse $x$. 
Part I

(Oblivious) Subspace Embeddings
Subspace Embedding

**Question:** Suppose we have linear subspace $E$ of $\mathbb{R}^n$ of dimension $d$. Can we find a projection $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ such that for every $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

$x$ is a vector in $\mathbb{R}^n$

Then $\|\Pi x\|_2 = \|x\|_2$

$k < \ll n$.

$$y = c \bar{x} \quad \forall c$$

$\|\Pi y\|_2 = (1 \pm \epsilon)\|y\|_2$
Two vector \( \overrightarrow{x_1}, \overrightarrow{x_2} \).

Fix \( \overrightarrow{x_1}, \overrightarrow{x_2} \).

Want projection matrix \( \Pi \).

\[ \overrightarrow{y} = c_1 \overrightarrow{x_1} + c_2 \overrightarrow{x_2} \]

\[ \overrightarrow{y} \in \text{span}(\overrightarrow{x_1}, \overrightarrow{x_2}) \]

\[ \| \overrightarrow{y} \|_2 \approx (1 \pm \epsilon) \| \overrightarrow{x} \|_2 \]
Subspace Embedding

**Question:** Suppose we have linear subspace $E$ of $\mathbb{R}^n$ of dimension $d$. Can we find a projection $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ such that for every $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why?

$\text{Chandra (UIUC)}$
Subspace Embedding

**Question:** Suppose we have linear subspace $E$ of $\mathbb{R}^n$ of dimension $d$. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for every $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? $\Pi$ maps $E$ to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $0$. 

Chandra (UIUC)  
CS498ABD  
20  
Fall 2020  
20 / 25
Subspace Embedding

**Question:** Suppose we have linear subspace $E$ of $\mathbb{R}^n$ of dimension $d$. Can we find a projection $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ such that for every $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? $\Pi$ maps $E$ to a lower dimension. Implies some non-zero vector $x \in E$ mapped to 0.
- Possible if $k = d$. Why?
Subspace Embedding

**Question:** Suppose we have linear subspace $E$ of $\mathbb{R}^n$ of dimension $d$. Can we find a projection $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ such that for every $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? $\Pi$ maps $E$ to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $0$.
- Possible if $k = d$. Why? Pick $\Pi$ to be an orthonormal basis for $E$. 
**Subspace Embedding**

**Question:** Suppose we have linear subspace $E$ of $\mathbb{R}^n$ of dimension $d$. Can we find a projection $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ such that for every $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? $\Pi$ maps $E$ to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $0$.
- Possible if $k = d$. Why? Pick $\Pi$ to be an orthonormal basis for $E$. **Disadvantage:** This requires knowing $E$ and computing orthonormal basis which is slow.
Subspace Embedding

**Question:** Suppose we have linear subspace $E$ of $\mathbb{R}^n$ of dimension $d$. Can we find a projection $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ such that for every $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? $\Pi$ maps $E$ to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $0$.
- Possible if $k = d$. Why? Pick $\Pi$ to be an orthonormal basis for $E$. **Disadvantage:** This requires knowing $E$ and computing orthonormal basis which is slow.

**What we really want:** Oblivious subspace embedding ala JL based on random projections
Oblivious Supspace Embedding

**Theorem**

Suppose $E$ is a linear subspace of $\mathbb{R}^n$ of dimension $d$. Let $\Pi$ be a DJL matrix $\Pi \in \mathbb{R}^{k \times n}$ with $k = O\left(\frac{d}{\epsilon^2} \log (1/\delta)\right)$ rows. Then with probability $(1 - \delta)$ for every $x \in E$,

$$\frac{1}{\sqrt{k}} \|\Pi x\|_2 = (1 \pm \epsilon) \|x\|_2.$$

In other words JL Lemma extends from one dimension to arbitrary number of dimensions in a graceful way.
Proof Idea

How do we prove that $\Pi$ works for $\forall x \in E$ which is an infinite set?

Several proofs but one useful argument that is often a starting hammer is the “net argument”

- Choose a large but finite set of vectors $T$ carefully (the net)
- Prove that $\Pi$ preserves lengths of vectors in $T$ (via naive union bound)
- Argue that any vector $x \in E$ is sufficiently close to a vector in $T$ and hence $\Pi$ also preserves length of $x$
Sufficient to focus on unit vectors in $E$. Why?
Net argument

Sufficient to focus on unit vectors in $E$. Why?

Also assume wlog and ease of notation that $E$ is the subspace formed by the first $d$ coordinates in standard basis.

$E$ is linear subspace of $d$-dim in $\mathbb{R}^n$. 

Claim: There is a net $T$ of size $\mathcal{O}(d)$ such that preserving lengths of vectors in $T$ succeeds.

Assuming claim: use DJL with $k = \mathcal{O}(d \varepsilon^2 \log(1/\varepsilon))$ and union bound to show that all vectors in $T$ are preserved in length up to $(1 \pm \varepsilon)$ factor.
Net argument

Sufficient to focus on unit vectors in $E$. Why?

Also assume wlog and ease of notation that $E$ is the subspace formed by the first $d$ coordinates in standard basis.

**Claim:** There is a net $T$ of size $e^{O(d)}$ such that preserving lengths of vectors in $T$ suffices.
Net argument

Sufficient to focus on unit vectors in $E$. Why?

Also assume wlog and ease of notation that $E$ is the subspace formed by the first $d$ coordinates in standard basis.

**Claim:** There is a net $T$ of size $e^{O(d)}$ such that preserving lengths of vectors in $T$ suffices.

Assuming claim: use DJL with $k = O\left(\frac{d^4}{\epsilon^2 \log(1/\delta)}\right)$ and union bound to show that all vectors in $T$ are preserved in length up to $(1 \pm \epsilon)$ factor.

$$\text{any fixed vector is pure } \to 1 - \exp(-\frac{d}{\delta})$$
Claim: If all vectors in $T$ are

\[ \frac{\pi}{d} \cdot \left( \frac{2d}{\pi} \right)^d \]

then all unit vectors per to $(1 \pm 2\kappa)$.

\[ K = \frac{1}{2} \ln \left( \left( \frac{2d}{\pi} \right)^d + \frac{1}{8} \right) \]

\[ d \ln d \]

If $x$ on the circle, we know

\[ \left( \frac{1}{\sqrt{\pi}} \right) \| x \|_{1/2} \approx \left( \frac{1}{\sqrt{\pi}} \right) \| x \|_{1/2} \]
Net argument

Sufficient to focus on unit vectors in $E$.

Also assume wlog and ease of notation that $E$ is the subspace formed by the first $d$ coordinates in standard basis.

A weaker net:

- Consider the box $[-1, 1]^d$ and make a grid with side length $\epsilon/d$
- Number of grid vertices is $(2d/\epsilon)^d$
- Sufficient to take $T$ to be the grid vertices
- Gives a weaker bound of $O(\frac{1}{\epsilon^2} d \log(d/\epsilon))$ dimensions
- A more careful net argument gives tight bound
Fix any $x \in E$ such that $\|x\|_2 = 1$ (unit vector).

There is grid point $y$ such that $\|y\|_2 \leq 1$ and $x$ is close to $y$.

Let $z = x - y$. We have $|z_i| \leq \epsilon / d$ for $1 \leq i \leq d$ and $z_i = 0$ for $i > d$.
Fix any $x \in E$ such that $\|x\|_2 = 1$ (unit vector)
There is grid point $y$ such that $\|y\|_2 \leq 1$ and $x$ is close to $y$
Let $z = x - y$. We have $|z_i| \leq \epsilon/d$ for $1 \leq i \leq i \leq d$ and $z_i = 0$ for $i > d$

$$\|\Pi x\| = \|\Pi y + \Pi z\| \leq \|\Pi y\| + \|\Pi z\|$$

$$\leq (1 + \epsilon) + (1 + \epsilon) \sum_{i=1}^{d} |z_i|$$

$$\leq (1 + \epsilon) + \epsilon(1 + \epsilon) \leq 1 + 3\epsilon$$

$\|\Pi x\| \geq 1 - 3\epsilon$
Net argument: analysis

Fix any \( x \in E \) such that \( \|x\|_2 = 1 \) (unit vector)

There is grid point \( y \) such that \( \|y\|_2 \leq 1 \) and \( x \) is close to \( y \)

Let \( z = x - y \). We have \( |z_i| \leq \epsilon/d \) for \( 1 \leq i \leq i \leq d \) and \( z_i = 0 \) for \( i > d \)

\[
\|\Pi x\| = \|\Pi y + \Pi z\| \leq \|\Pi y\| + \|\Pi z\|
\]

\[
\leq (1 + \epsilon) + (1 + \epsilon) \sum_{i=1}^{d} |z_i|
\]

\[
\leq (1 + \epsilon) + \epsilon(1 + \epsilon) \leq 1 + 3\epsilon
\]

Similarly \( \|\Pi x\| \geq 1 - O(\epsilon) \).
Application of Subspace Embeddings

Faster algorithms for approximate

- matrix multiplication
- regression
- SVD

**Basic idea:** Want to perform operations on matrix $A$ with $n$ data columns (say in large dimension $\mathbb{R}^h$) with small effective rank $d$. Want to reduce to a matrix of size roughly $\mathbb{R}^{d \times d}$ by spending time proportional to $\text{nnz}(A)$.

Later in course.