Day 4 Talk 2

Jelani Nelson

"Optimality of the Johnson-Lindenstrauss Lemma"

Joint w Kasper Green Larsen (Aarhus)

= Johnson-Lindenstrauss

JL-lemma '84

For every $X \subset \ell_2$ $|X| = n$ $\exists f : X \mapsto \ell_2^m$ with $m = O(\epsilon^{-2} \log n)$ with distortion $1+\epsilon$, i.e.

$\forall x, y \in X$ $(1-\epsilon)\|x-y\|_2^2 \le \|f(x) - f(y)\|_2^2 \le (1+\epsilon)\|x-y\|_2^2$

- Many applications: e.g. high dimensional nearest neighbor with preprocessing.

Distributional JL: $\exists$ distribution $D_{\epsilon,\delta}$ on $\mathbb{R}^{m \times d}$ for $m = O(\epsilon^{-2} \log \frac{1}{\delta})$ s.t.

$\forall u \in S^{n-1}$ $\mathbb{P}_{\Pi \sim D_{\epsilon,\delta}} \left(\left| \|\Pi u\|_2^2 - 1 \right| > \epsilon \right) < \delta$

Proof of JL: Union bound. □

These are
- upper bounds
- are they tight

Optimality (for refs see slides)

if ϵ not too small

Thm $m = \Theta(\epsilon^{-2} \log \frac{1}{\delta})$ is optimal for distributional

Thm Alon '03

$m = \Omega(\epsilon^{-2}/\log(\frac{1}{\epsilon}) \log n)$ is required

Example: simplex

This talk

Thm (Larsen, Nelson '17)

For JL: $m = \Omega(\epsilon^{-2} \log n)$ is required

embedding

prior work: if f linear

Thm $d, n \geq 2$, $\frac{1}{(\min\{n,d\})^{0.4999}} < \epsilon < 1$, there exists $X \subset \ell_2^d$ s.t. any $f: X \to \ell_2^m$ with dist $1+\epsilon$ must have $m = \Omega(\epsilon^{-2} \log n)$

Precice

Prior Lower bounds:

- JL'84: volume bound $\leadsto m = \Omega(\log n)$
- Alon '03: $m = \Omega\left(\epsilon^{-2} \frac{\log n}{\log(\frac{1}{\epsilon})}\right)$
- Larsen, Nelson

net argument + probabilistic method:

$m = \Omega(\epsilon^{-2} \log n)$ for linear f

Barrier in prior proofs

previously didn't have many "bad" point sets X

$\rightsquigarrow$ could always find good non-linear f.

ε-incoherent point sets could be dealt with using Reed-Solomon with JL not tight.

Code: $C \subseteq \{1 \ldots q\}^t$ ← block length

(Black-board)

- C: code
- $\{1 \ldots q\}$: alphabet size

- relative dist of $\mathcal{C}$

$$\min_{c - c' \notin \mathcal{C}} \frac{\Delta(c, c')}{t}$$

Example simplex $0, e_1, \ldots e_n$ construct low dimensional incoherent subsets

$\rightsquigarrow$ can do that (Recall work of Alon)

$0 \quad e_1 \ldots e_n$

$\downarrow$

$0 \quad x_1 \ldots x_n$

$C_i = (\alpha_1 \ldots \alpha_t)$, with $\alpha_j \in [q]$

$x_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix}$ } t

<u>Reed-Solomon</u>

- $q = t =$ prime power
- each codeword deg-d poly over $\mathbb{F}_q[x]$
- $C_p = (p(\beta))_{\beta \in \mathbb{F}_q}$

- $d = \varepsilon q \Rightarrow$ relative dist $1-\epsilon$
- $q^d = q^{\epsilon q} \geq n$

$$\Rightarrow q \simeq \frac{1}{\epsilon}\left(\frac{\lg n}{\lg\frac{1}{\epsilon} + \lg\lg n}\right)$$

Better than $\epsilon < 2^{-c\sqrt{\lg n}}$.

if Reed-Solomon works ⇝ why not all ? X
– turns out to be false
⇝ thm

Encoding argument

- Define large collection $\mathcal{X}$ of n-sized sets $X \subset \mathbb{R}^n$

if all can be embedded into $< 10^{-4} \epsilon^{-2} \log n$

⇝ encode $\mathcal{X}$ by few bits ⇝ contradiction

<u>Obs:</u> preserving lengths ⇝ preserving dot-products.

- $X \in \mathcal{X}$ of the form

$$X = (\underbrace{0, e_1 \ldots e_n}_{\text{simplex}}, y_{S_1} \ldots y_{S_{n+d-1}})$$

$S \subset [n]$, $|S| = k$ $\qquad y_S = \frac{1}{\sqrt{k}} \sum_{i \in S} e_i$

Idea: encode by discretizing $f(e_i)$, $f(y_{S_i})$ by an ϵ-net

⇝ not enough, extra: $\lg\frac{1}{\epsilon}$ – factor

⇝ better encoding

Related to static approximate dot product

Open problems on slides.

Optimality of the Johnson-Lindenstrauss lemma

Jelani Nelson
Harvard

November 16, 2017

joint work with Kasper Green Larsen (Aarhus)

Johnson-Lindenstrauss (JL) lemma

JL lemma [Johnson, Lindenstrauss '84]

For every $X \subset \ell_2$ of size n, there is an embedding $f : X \to \ell_2^m$ for $m = O(\varepsilon^{-2} \log n)$ with distortion $1 + \varepsilon$. That is, for each $x, y \in X$,

$$(1 - \varepsilon)\|x - y\|_2^2 \leq \|f(x) - f(y)\|_2^2 \leq (1 + \varepsilon)\|x - y\|_2^2$$

Johnson-Lindenstrauss (JL) lemma

JL lemma [Johnson, Lindenstrauss '84]

For every $X \subset \ell_2$ of size n, there is an embedding $f : X \to \ell_2^m$ for $m = O(\varepsilon^{-2} \log n)$ with distortion $1 + \varepsilon$. That is, for each $x, y \in X$,

$$(1-\varepsilon)\|x-y\|_2^2 \le \|f(x) - f(y)\|_2^2 \le (1+\varepsilon)\|x-y\|_2^2$$

Uses in computer science:

- Speed up geometric algorithms by first reducing dimension of input [Indyk, Motwani '98], [Indyk '01]
- Faster/streaming numerical linear algebra algorithms [Sarlós '06], [LWMRT '07], [Clarkson, Woodruff '09]
- Essentially equivalent to RIP matrices from compressed sensing [Baraniuk et al. '08], [Krahmer, Ward '11] (used for recovery of sparse signals)
- Volume-preserving embeddings (applications to projective clustering) [Magen '02]
-

How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma (DJL lemma [Johnson, Lindenstrauss '84])

For any $0 < \varepsilon, \delta < 1/2$ *and* $d \geq 1$ *there exists a distribution* $\mathcal{D}_{\varepsilon,\delta}$ *on* $\mathbb{R}^{m \times d}$ *for* $m = O(\varepsilon^{-2} \log(1/\delta))$ *such that for any* $u \in S^{d-1}$

$$\mathop{\mathbb{P}}_{\Pi \sim \mathcal{D}_{\varepsilon,\delta}} \left(\left| \|\Pi u\|_2^2 - 1 \right| > \varepsilon \right) < \delta.$$

How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma (DJL lemma [Johnson, Lindenstrauss '84])

For any $0 < \varepsilon, \delta < 1/2$ *and* $d \geq 1$ *there exists a distribution* $\mathcal{D}_{\varepsilon,\delta}$ *on* $\mathbb{R}^{m\times d}$ *for* $m = O(\varepsilon^{-2}\log(1/\delta))$ *such that for any* $u \in S^{d-1}$

$$\mathbb{P}_{\Pi\sim\mathcal{D}_{\varepsilon,\delta}}\left(\left|\|\Pi u\|_2^2 - 1\right| > \varepsilon\right) < \delta.$$

Proof of JL: Set $\delta = 1/n^2$ in DJL and u as the normalized difference vector of some pair of points. Union bound over the $\binom{n}{2}$ pairs. Thus, in fact, the map $f : X \to \ell_2^m$ can be linear.

How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma (DJL lemma [Johnson, Lindenstrauss '84])

For any $0 < \varepsilon, \delta < 1/2$ *and* $d \geq 1$ *there exists a distribution* $\mathcal{D}_{\varepsilon,\delta}$ *on* $\mathbb{R}^{m \times d}$ *for* $m = O(\varepsilon^{-2} \log(1/\delta))$ *such that for any* $u \in S^{d-1}$

$$\mathop{\mathbb{P}}_{\Pi \sim \mathcal{D}_{\varepsilon,\delta}} \left(\left| \|\Pi u\|_2^2 - 1 \right| > \varepsilon \right) < \delta.$$

Proof of JL: Set $\delta = 1/n^2$ in DJL and u as the normalized difference vector of some pair of points. Union bound over the $\binom{n}{2}$ pairs. Thus, in fact, the map $f : X \to \ell_2^m$ can be linear.

Theorem (Jayram-Woodruff, 2011; Kane-Meka-Nelson, 2011)

For DJL, $m = \Theta(\varepsilon^{-2} \log(1/\delta))$ *is optimal.*

Theorem (Alon, 2003)

For JL, $m = \Omega((\varepsilon^{-2}/\log(1/\varepsilon)) \log n)$ *is required.*

Theorem (Larsen, Nelson 2016)

For JL, $m = \Omega(\varepsilon^{-2} \log n)$ *is required if* f *must be a linear map.*

How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma (DJL lemma [Johnson, Lindenstrauss '84])

For any $0 < \varepsilon, \delta < 1/2$ *and* $d \geq 1$ *there exists a distribution* $\mathcal{D}_{\varepsilon,\delta}$ *on* $\mathbb{R}^{m \times d}$ *for* $m = O(\varepsilon^{-2} \log(1/\delta))$ *such that for any* $u \in S^{d-1}$

$$\mathbb{P}_{\Pi \sim \mathcal{D}_{\varepsilon,\delta}} \left(\left| \|\Pi u\|_2^2 - 1 \right| > \varepsilon \right) < \delta.$$

Proof of JL: Set $\delta = 1/n^2$ in DJL and u as the normalized difference vector of some pair of points. Union bound over the $\binom{n}{2}$ pairs. Thus, in fact, the map $f : X \to \ell_2^m$ can be linear.

Theorem (Jayram-Woodruff, 2011; Kane-Meka-Nelson, 2011)

For DJL, $m = \Theta(\varepsilon^{-2} \log(1/\delta))$ *is optimal.*

Theorem (Alon, 2003)

For JL, $m = \Omega((\varepsilon^{-2}/\log(1/\varepsilon)) \log n)$ *is required.*

Theorem (Larsen, Nelson 2016)

For JL, $m = \Omega(\varepsilon^{-2} \log n)$ *is required if* f *must be a linear map.*

How to prove the JL lemma

Distributional JL (DJL) lemma

Lemma (DJL lemma [Johnson, Lindenstrauss '84])

For any $0 < \varepsilon, \delta < 1/2$ and $d \geq 1$ there exists a distribution $\mathcal{D}_{\varepsilon,\delta}$ on $\mathbb{R}^{m \times d}$ for $m = O(\varepsilon^{-2} \log(1/\delta))$ such that for any $u \in S^{d-1}$

$$\underset{\Pi \sim \mathcal{D}_{\varepsilon,\delta}}{\mathbb{P}} \left(\left| \|\Pi u\|_2^2 - 1 \right| > \varepsilon \right) < \delta.$$

Proof of JL: Set $\delta = 1/n^2$ in DJL and u as the normalized difference vector of some pair of points. Union bound over the $\binom{n}{2}$ pairs. Thus, in fact, the map $f : X \to \ell_2^m$ can be linear.

Theorem (Jayram-Woodruff, 2011; Kane-Meka-Nelson, 2011)

For DJL, $m = \Theta(\varepsilon^{-2} \log(1/\delta))$ is optimal.

Theorem (Alon, 2003)

For JL, $m = \Omega((\varepsilon^{-2}/\log(1/\varepsilon)) \log n)$ is required.

Theorem (Larsen, Nelson 2017)

For JL, $m = \Omega(\varepsilon^{-2} \log n)$ is required ~~if f must be a linear map~~.

JL lower bound

Theorem ([Larsen, Nelson '17])

For any integers $d, n \geq 2$ *and any* $\frac{1}{(\min\{n,d\})^{0.4999}} < \varepsilon < 1$, *there exists a set* $X \subset \ell_2^d$ *such that any embedding* $f : X \to \ell_2^m$ *with distortion at most* $1 + \varepsilon$ *must have*

$$m = \Omega(\varepsilon^{-2} \log n)$$

JL lower bound

Theorem ([Larsen, Nelson '17])

For any integers $d, n \geq 2$ and any $\frac{1}{(\min\{n,d\})^{0.4999}} < \varepsilon < 1$, there exists a set $X \subset \ell_2^d$ such that any embedding $f : X \to \ell_2^m$ with distortion at most $1 + \varepsilon$ must have

$$m = \Omega(\varepsilon^{-2} \log n)$$

- **Can always achieve $m = d$:** f is the identity map.
- **Can always achieve $m = n - 1$:** translate so one vector is 0. Then all vectors live in $(n-1)$-dimensional subspace.

JL lower bound

Theorem ([Larsen, Nelson '17])

For any integers $d, n \geq 2$ and any $\frac{1}{(\min\{n,d\})^{0.4999}} < \varepsilon < 1$, there exists a set $X \subset \ell_2^d$ such that any embedding $f : X \to \ell_2^m$ with distortion at most $1 + \varepsilon$ must have

$$m = \Omega(\varepsilon^{-2} \log n)$$

- **Can always achieve $m = d$:** f is the identity map.
- **Can always achieve $m = n - 1$:** translate so one vector is 0. Then all vectors live in $(n-1)$-dimensional subspace.
- So can only hope JL optimal for $\varepsilon^{-2} \log n \leq \min\{n, d\}$, can view theorem assumption as $\varepsilon^{-2} \log n \ll \min\{n, d\}^{0.999}$.

Lower bound techniques over time

Lower bounds over time

- **Volume argument.** $m = \Omega(\log n)$ [Johnson, Lindenstrauss '84]

Lower bounds over time

- **Volume argument.** $m = \Omega(\log n)$ [Johnson, Lindenstrauss '84]
- **Incoherence + tensor trick.** $m = \Omega(\frac{1}{\varepsilon^2} \frac{\log n}{\log(1/\varepsilon)})$ [Alon '03]

Lower bounds over time

- **Volume argument.** $m = \Omega(\log n)$ [Johnson, Lindenstrauss '84]
- **Incoherence + tensor trick.** $m = \Omega(\frac{1}{\varepsilon^2} \frac{\log n}{\log(1/\varepsilon)})$ [Alon '03]
- **Net argument + probabilistic method.** $m = \Omega(\frac{1}{\varepsilon^2} \log n)$ (only against linear maps $f(x) = \Pi x$) [Larsen, Nelson '16]

Lower bounds over time

- **Volume argument.** $m = \Omega(\log n)$ [Johnson, Lindenstrauss '84]
- **Incoherence + tensor trick.** $m = \Omega(\frac{1}{\varepsilon^2}\frac{\log n}{\log(1/\varepsilon)})$ [Alon '03]
- **Net argument + probabilistic method.** $m = \Omega(\frac{1}{\varepsilon^2}\log n)$ (only against linear maps $f(x) = \Pi x$) [Larsen, Nelson '16]
- **Encoding argument.** $m = \Omega(\frac{1}{\varepsilon^2}\log n)$ [Larsen, Nelson '17]

Barrier in previous proofs.

(The limits of incoherence.)

Barrier in previous proofs.

(The limits of incoherence.)

We say $x_1, \ldots, x_n$ are ε-*incoherent* if

- $\forall i \; \|x_i\|_2 = 1$
- $\forall i \neq j \; |\langle x_i, x_j \rangle| \leq \varepsilon$

Barrier in previous proofs.

(The limits of incoherence.)

We say $x_1, \ldots, x_n$ are ε-*incoherent* if

- $\forall i \ \|x_i\|_2 = 1$
- $\forall i \neq j \ |\langle x_i, x_j \rangle| \leq \varepsilon$

Previous hard point sets were incoherent, so could just map to another set of incoherent vectors in smaller dimension.

Barrier in previous proofs.

(The limits of incoherence.)

We say $x_1, \ldots, x_n$ are ε-*incoherent* if

- $\forall i \ \|x_i\|_2 = 1$
- $\forall i \neq j \ |\langle x_i, x_j \rangle| \leq \varepsilon$

Previous hard point sets were incoherent, so could just map to another set of incoherent vectors in smaller dimension.

$m = O(\varepsilon^{-2}(\frac{\log n}{\log(1/\varepsilon)+\log\log n})^2)$ achievable via Reed-Solomon codes.

Barrier in previous proofs.

(The limits of incoherence.)

We say $x_1, \ldots, x_n$ are ε-*incoherent* if

- $\forall i \ \|x_i\|_2 = 1$
- $\forall i \neq j \ |\langle x_i, x_j \rangle| \leq \varepsilon$

Previous hard point sets were incoherent, so could just map to another set of incoherent vectors in smaller dimension.

$m = O(\varepsilon^{-2}(\frac{\log n}{\log(1/\varepsilon)+\log\log n})^2)$ achievable via Reed-Solomon codes.

Our new lower bound must use more than just incoherence.

Encoding argument.

[Larsen, Nelson '17]

JL is optimal even against non-linear maps

We define a large collection $\mathcal{X}$ of n-sized sets $X \subset \mathbb{R}^d$ s.t. if all $X \in \mathcal{X}$ can be embedded into dimension $\leq 10^{-10} \cdot \varepsilon^{-2} \log_2 n$, then there is an encoding of elements of $\mathcal{X}$ into $< \log_2 |\mathcal{X}|$ bits (i.e. a surjection from $\mathcal{X}$ to $\{0,1\}^t$ for $t < \log_2 |\mathcal{X}|$). **Contradiction.**

Encoding argument.

[Larsen, Nelson '17]

Encoding argument.

[Larsen, Nelson '17]

For now: assume $d = n/\lg(1/\varepsilon)$

Observation

- Preserving distances implies preserving dot products. Say $\|x\|_2 = \|y\|_2 = 1$.

$$\|x - y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle \ (*)$$

$$\|f(x) - f(y)\|_2^2 = \|f(x)\|_2^2 + \|f(y)\|_2^2 - 2\langle f(x), f(y)\rangle$$

Observation

- Preserving distances implies preserving dot products. Say $\|x\|_2 = \|y\|_2 = 1$.

$$\|x - y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle \ (*)$$
$$\|f(x) - f(y)\|_2^2 = \|f(x)\|_2^2 + \|f(y)\|_2^2 - 2\langle f(x), f(y)\rangle$$

$$\implies (1 \pm \varepsilon)\|x - y\|_2^2 = (1 \pm \varepsilon)\overbrace{\|x\|_2^2}^{1} + (1 \pm \varepsilon)\overbrace{\|y\|_2^2}^{1} - 2\langle f(x), f(y)\rangle \ (**)$$

- Now subtract (*) from (**): $\langle f(x), f(y)\rangle = \langle x, y\rangle \pm O(\varepsilon)$

JL lower bound outline

- Pick $k = \frac{1}{100\varepsilon^2}$.
- For $S \subset [d]$ of size k, define vector $y_S = \frac{1}{\sqrt{k}} \sum_{j \in S} e_j$. Note

$$\langle y_S, e_i \rangle = \begin{cases} 10\varepsilon, & i \in S \\ 0, & \text{otherwise} \end{cases}$$

- **Idea:** low-distortion embedding preserves dot products up to $\pm\varepsilon$, which is enough to distinguish the two cases

JL lower bound outline

- Pick $k = \frac{1}{100\varepsilon^2}$.
- For $S \subset [d]$ of size k, define vector $y_S = \frac{1}{\sqrt{k}} \sum_{j \in S} e_j$. Note

$$\langle y_S, e_i \rangle = \begin{cases} 10\varepsilon, & i \in S \\ 0, & \text{otherwise} \end{cases}$$

- **Idea:** low-distortion embedding preserves dot products up to $\pm\varepsilon$, which is enough to distinguish the two cases
- $\mathcal{X}$ is set of all ordered tuples of points, possibly with repetition $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$ with the $S_i \in \binom{[d]}{k}$.

JL lower bound outline

- Pick $k = \frac{1}{100\varepsilon^2}$.
- For $S \subset [d]$ of size k, define vector $y_S = \frac{1}{\sqrt{k}} \sum_{j \in S} e_j$. Note

$$\langle y_S, e_i \rangle = \begin{cases} 10\varepsilon, & i \in S \\ 0, & \text{otherwise} \end{cases}$$

- **Idea:** low-distortion embedding preserves dot products up to $\pm\varepsilon$, which is enough to distinguish the two cases
- $\mathcal{X}$ is set of all ordered tuples of points, possibly with repetition $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$ with the $S_i \in \binom{[d]}{k}$.
- $|\mathcal{X}| = \binom{d}{k}^{n-d-1}$, thus any encoding of $X \in \mathcal{X}$ requires $\geq (n-d-1)\lg\binom{d}{k} = (1 - o_\varepsilon(1))nk\lg(d/k)$ bits.

JL lower bound outline

- Pick $k = \frac{1}{100\varepsilon^2}$.
- For $S \subset [d]$ of size k, define vector $y_S = \frac{1}{\sqrt{k}} \sum_{j\in S} e_j$. Note

$$\langle y_S, e_i \rangle = \begin{cases} 10\varepsilon, & i \in S \\ 0, & \text{otherwise} \end{cases}$$

- **Idea:** low-distortion embedding preserves dot products up to $\pm\varepsilon$, which is enough to distinguish the two cases
- $\mathcal{X}$ is set of all ordered tuples of points, possibly with repetition $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$ with the $S_i \in \binom{[d]}{k}$.
- $|\mathcal{X}| = \binom{d}{k}^{n-d-1}$, thus any encoding of $X \in \mathcal{X}$ requires $\geq (n-d-1) \lg \binom{d}{k} = (1 - o_\varepsilon(1)) nk \lg(d/k)$ bits.
- Will show any $(1+\varepsilon)$-distortion embedding into ℓ_2^m implies encoding into $O(nm)$ bits, hence $nm = \Omega(nk \log(d/k))$ $\Rightarrow m = \Omega(k \log(d/k)) = \Omega(\varepsilon^{-2} \log n)$ for ε not too small.

Problem: Encoding of $X \in \mathcal{X}$ can't just be a description of $f(0), f(e_1), \ldots, f(e_d), f(y_{S_1}), \ldots, f(y_{S_{n-d-1}})$.

Why not?

Problem: Encoding of $X \in \mathcal{X}$ can't just be a description of $f(0), f(e_1), \ldots, f(e_d), f(y_{S_1}), \ldots, f(y_{S_{n-d-1}})$.

Why not? Want to violate pigeonhole principle, so range of the encoding must be of size $< \lg |\mathcal{X}|$. But $f(x)$ has real entries, so the range is infinite!

Problem: Encoding of $X \in \mathcal{X}$ can't just be a description of $f(0), f(e_1), \ldots, f(e_d), f(y_{S_1}), \ldots, f(y_{S_{n-d-1}})$.

Why not? Want to violate pigeonhole principle, so range of the encoding must be of size $< \lg |\mathcal{X}|$. But $f(x)$ has real entries, so the range is infinite!

Fix attempt 1: Round entries of $f(x)$ to integer multiples of γ. Can show $\gamma = O(\frac{\varepsilon}{\sqrt{m}})$ suffices $\Longrightarrow O(nm \log(m/\varepsilon))$ bit encoding

Problem: Encoding of $X \in \mathcal{X}$ can't just be a description of $f(0), f(e_1), \ldots, f(e_d), f(y_{S_1}), \ldots, f(y_{S_{n-d-1}})$.

Why not? Want to violate pigeonhole principle, so range of the encoding must be of size $< \lg |\mathcal{X}|$. But $f(x)$ has real entries, so the range is infinite!

Fix attempt 1: Round entries of $f(x)$ to integer multiples of γ. Can show $\gamma = O(\frac{\varepsilon}{\sqrt{m}})$ suffices $\implies O(nm \log(m/\varepsilon))$ bit encoding $\implies m = \Omega(\varepsilon^{-2} \frac{\log n}{\log(m/\varepsilon)})$ final lower bound

Problem: Encoding of $X \in \mathcal{X}$ can't just be a description of $f(0), f(e_1), \ldots, f(e_d), f(y_{S_1}), \ldots, f(y_{S_{n-d-1}})$.

Why not? Want to violate pigeonhole principle, so range of the encoding must be of size $< \lg|\mathcal{X}|$. But $f(x)$ has real entries, so the range is infinite!

Fix attempt 1: Round entries of $f(x)$ to integer multiples of γ. Can show $\gamma = O(\frac{\varepsilon}{\sqrt{m}})$ suffices $\implies O(nm\log(m/\varepsilon))$ bit encoding $\implies m = \Omega(\varepsilon^{-2}\frac{\log n}{\log(m/\varepsilon)})$ final lower bound

Slightly better fix: Round each $f(x)$ to a point $\widetilde{f(x)}$ in an ε-net in ℓ_2 instead of a γ-net in ℓ_∞ as above.

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \leq 1 + \varepsilon$ for $x \in X$
  i.e. $\forall x \in X,\ f(x) \in (1+\varepsilon)B_{\ell_2^m}$

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \leq 1 + \varepsilon$ for $x \in X$
  i.e. $\forall x \in X$, $f(x) \in (1+\varepsilon)B_{\ell_2^m}$
- Pick $c\varepsilon$-net T of $(1+\varepsilon)B_{\ell_2^m}$ in ℓ_2; has size $N = O(1/\varepsilon)^m$.

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \le 1 + \varepsilon$ for $x \in X$
  i.e. $\forall x \in X$, $f(x) \in (1+\varepsilon)B_{\ell_2^m}$
- Pick $c\varepsilon$-net T of $(1+\varepsilon)B_{\ell_2^m}$ in ℓ_2; has size $N = O(1/\varepsilon)^m$.
- Encode $f(x)$ as $\widetilde{f(x)} \in T$: $|X| \cdot \lg N = nm \lg(1/\varepsilon)$ bits

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \le 1 + \varepsilon$ for $x \in X$
  i.e. $\forall x \in X$, $f(x) \in (1+\varepsilon)B_{\ell_2^m}$
- Pick $c\varepsilon$-net T of $(1+\varepsilon)B_{\ell_2^m}$ in ℓ_2; has size $N = O(1/\varepsilon)^m$.
- Encode $f(x)$ as $\widetilde{f(x)} \in T$: $|X| \cdot \lg N = nm \lg(1/\varepsilon)$ bits
- **Remember:** $\langle e_i, y_S \rangle \in \{0, 10\varepsilon\}$ (depends on whether $i \in S$)

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \le 1 + \varepsilon$ for $x \in X$

  i.e. $\forall x \in X,\ f(x) \in (1+\varepsilon)B_{\ell_2^m}$
- Pick $c\varepsilon$-net T of $(1+\varepsilon)B_{\ell_2^m}$ in ℓ_2; has size $N = O(1/\varepsilon)^m$.
- Encode $f(x)$ as $\widetilde{f(x)} \in T$: $|X| \cdot \lg N = nm \lg(1/\varepsilon)$ bits
- **Remember:** $\langle e_i, y_S \rangle \in \{0, 10\varepsilon\}$ (depends on whether $i \in S$)
- Low-distortion embedding preserves dot products, so $\langle f(e_i), f(y_S) \rangle \in \{\pm\varepsilon, 10\varepsilon \pm \varepsilon\}$

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \le 1 + \varepsilon$ for $x \in X$ i.e. $\forall x \in X,\ f(x) \in (1+\varepsilon) B_{\ell_2^m}$
- Pick $c\varepsilon$-net T of $(1+\varepsilon)B_{\ell_2^m}$ in ℓ_2; has size $N = O(1/\varepsilon)^m$.
- Encode $f(x)$ as $\widetilde{f(x)} \in T$: $|X| \cdot \lg N = nm \lg(1/\varepsilon)$ bits
- **Remember:** $\langle e_i, y_S \rangle \in \{0, 10\varepsilon\}$ (depends on whether $i \in S$)
- Low-distortion embedding preserves dot products, so $\langle f(e_i), f(y_S) \rangle \in \{\pm\varepsilon, 10\varepsilon \pm \varepsilon\}$
- Mapping to $c\varepsilon$-net points again preserves dot products, so $\left\langle \widetilde{f(e_i)}, \widetilde{f(y_S)} \right\rangle \in \{\pm 2\varepsilon, 10\varepsilon \pm 2\varepsilon\}$

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \le 1 + \varepsilon$ for $x \in X$ i.e. $\forall x \in X$, $f(x) \in (1+\varepsilon)B_{\ell_2^m}$
- Pick $c\varepsilon$-net T of $(1+\varepsilon)B_{\ell_2^m}$ in ℓ_2; has size $N = O(1/\varepsilon)^m$.
- Encode $f(x)$ as $\widetilde{f(x)} \in T$: $|X| \cdot \lg N = nm \lg(1/\varepsilon)$ bits
- **Remember:** $\langle e_i, y_S \rangle \in \{0, 10\varepsilon\}$ (depends on whether $i \in S$)
- Low-distortion embedding preserves dot products, so $\langle f(e_i), f(y_S) \rangle \in \{\pm\varepsilon, 10\varepsilon \pm \varepsilon\}$
- Mapping to $c\varepsilon$-net points again preserves dot products, so $\left\langle \widetilde{f(e_i)}, \widetilde{f(y_S)} \right\rangle \in \{\pm 2\varepsilon, 10\varepsilon \pm 2\varepsilon\}$
- Thus from encodings can recover $\langle e_i, y_S \rangle$ to know which $i \in S$ (dot product either $< 2\varepsilon$ in magnitude, or $> 8\varepsilon$)

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Since distances to 0 preserved, $\|f(x)\|_2^2 \leq 1 + \varepsilon$ for $x \in X$ i.e. $\forall x \in X,\ f(x) \in (1+\varepsilon)B_{\ell_2^m}$
- Pick $c\varepsilon$-net T of $(1+\varepsilon)B_{\ell_2^m}$ in ℓ_2; has size $N = O(1/\varepsilon)^m$.
- Encode $f(x)$ as $\widetilde{f(x)} \in T$: $|X| \cdot \lg N = nm \lg(1/\varepsilon)$ bits
- **Remember:** $\langle e_i, y_S \rangle \in \{0, 10\varepsilon\}$ (depends on whether $i \in S$)
- Low-distortion embedding preserves dot products, so $\langle f(e_i), f(y_S) \rangle \in \{\pm\varepsilon, 10\varepsilon \pm \varepsilon\}$
- Mapping to $c\varepsilon$-net points again preserves dot products, so $\left\langle \widetilde{f(e_i)}, \widetilde{f(y_S)} \right\rangle \in \{\pm 2\varepsilon, 10\varepsilon \pm 2\varepsilon\}$
- Thus from encodings can recover $\langle e_i, y_S \rangle$ to know which $i \in S$ (dot product either $< 2\varepsilon$ in magnitude, or $> 8\varepsilon$)
- Can decode X, implies $nm \lg(1/\varepsilon) = \Omega(n\varepsilon^{-2} \log(\varepsilon^2 d))$

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Can decode X, implies $nm \lg(1/\varepsilon) = \Omega(n\varepsilon^{-2} \log(\varepsilon^2 d))$

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, ys_1, \ldots, ys_{n-d-1})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Can decode X, implies $nm \lg(1/\varepsilon) = \Omega(n\varepsilon^{-2} \log(\varepsilon^2 d))$
- Thus $m = \Omega(\varepsilon^{-2} \log(\varepsilon^2 d)/\log(1/\varepsilon))$.

  for ε not too small, this is $m = \Omega(\varepsilon^{-2} \frac{\log n}{\log(1/\varepsilon)})$

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Can decode X, implies $nm \lg(1/\varepsilon) = \Omega(n\varepsilon^{-2} \log(\varepsilon^2 d))$
- Thus $m = \Omega(\varepsilon^{-2} \log(\varepsilon^2 d) / \log(1/\varepsilon))$.

  for ε not too small, this is $m = \Omega(\varepsilon^{-2} \frac{\log n}{\log(1/\varepsilon)})$
- Same lower bound as [Alon '03], but very different argument.

  ... but not what I promised you!

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Can decode X, implies $nm\lg(1/\varepsilon) = \Omega(n\varepsilon^{-2}\log(\varepsilon^2 d))$
- Thus $m = \Omega(\varepsilon^{-2}\log(\varepsilon^2 d)/\log(1/\varepsilon))$.

  for ε not too small, this is $m = \Omega(\varepsilon^{-2}\frac{\log n}{\log(1/\varepsilon)})$
- Same lower bound as [Alon '03], but very different argument.

  ... but not what I promised you!

A warmup lower bound

Recall $X = (0, e_1, \ldots, e_d, y_{S_1}, \ldots, y_{S_{n-d-1}})$. For $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^m$, wlog $f(0) = 0$ (by translating).

- Can decode X, implies $nm\lg(1/\varepsilon) = \Omega(n\varepsilon^{-2}\log(\varepsilon^2 d))$
- Thus $m = \Omega(\varepsilon^{-2}\log(\varepsilon^2 d)/\log(1/\varepsilon))$.

  for ε not too small, this is $m = \Omega(\varepsilon^{-2}\frac{\log n}{\log(1/\varepsilon)})$
- Same lower bound as [Alon '03], but very different argument.

  ... but not what I promised you!
- Will now show a better encoding.

  remember, we are for now assuming $d = n/\lg(1/\varepsilon)$

An encoding of X into $O(nm)$ bits

Sufficed for decoding X: knowing $\left\langle \widetilde{f(e_i)}, \widetilde{f(y_{S_j})} \right\rangle$ for each i, j

An encoding of X into $O(nm)$ bits

Sufficed for decoding X: knowing $\left\langle \widetilde{f(e_i)}, \widetilde{f(y_{S_j})} \right\rangle$ for each i, j

$$A \left\{ \begin{bmatrix} \widetilde{f(e_1)}^T \\ \widetilde{f(e_2)}^T \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \widetilde{f(e_d)}^T \end{bmatrix} \right. \cdot \begin{bmatrix} \\ \widetilde{f(y_{S_j})} \\ \\ \end{bmatrix} = \left. \begin{bmatrix} \left\langle \widetilde{f(e_1)}, \widetilde{f(y_{S_j})} \right\rangle \\ \left\langle \widetilde{f(e_2)}, \widetilde{f(y_{S_j})} \right\rangle \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \left\langle \widetilde{f(e_d)}, \widetilde{f(y_{S_j})} \right\rangle \end{bmatrix} \right\} v_j$$

- Knowing $v_1, \dots, v_{n-d-1}$ would allow us to decode.

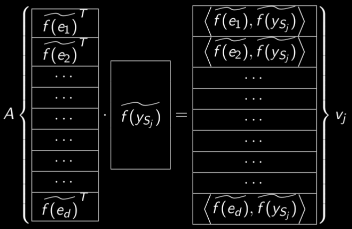
An encoding of X into $O(nm)$ bits

Sufficed for decoding X: knowing $\left\langle \widetilde{f(e_i)}, \widetilde{f(y_{S_j})} \right\rangle$ for each i, j

$$A \left\{ \begin{bmatrix} \widetilde{f(e_1)}^T \\ \widetilde{f(e_2)}^T \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \widetilde{f(e_d)}^T \end{bmatrix} \right. \cdot \begin{bmatrix} \widetilde{f(y_{S_j})} \end{bmatrix} = \left. \begin{bmatrix} \left\langle \widetilde{f(e_1)}, \widetilde{f(y_{S_j})} \right\rangle \\ \left\langle \widetilde{f(e_2)}, \widetilde{f(y_{S_j})} \right\rangle \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \left\langle \widetilde{f(e_d)}, \widetilde{f(y_{S_j})} \right\rangle \end{bmatrix} \right\} v_j$$

- Knowing $v_1, \ldots, v_{n-d-1}$ would allow us to decode.
- In fact, suffices to know $\tilde{v}_j$ such that $\|v_j - \tilde{v}_j\|_\infty < \varepsilon$.
  (then each entry of $\tilde{v}_j$ is $< 3\varepsilon$ or $> 7\varepsilon$ in magnitude)

An encoding of X into $O(nm)$ bits



- $A \in \mathbb{R}^{d \times m}$
- $\widetilde{f(ys_j)} \in \mathbb{R}^m$
- $v_j \in \mathbb{R}^d$

- Let E denote the column space of A $dim(E) \leq m$.

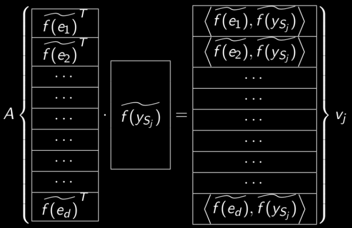
An encoding of X into $O(nm)$ bits

$$A\left\{\begin{bmatrix}\widetilde{f(e_1)}^T\\ \widetilde{f(e_2)}^T\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \widetilde{f(e_d)}^T\end{bmatrix}\right. \cdot \begin{bmatrix}\widetilde{f(ys_j)}\end{bmatrix} = \left.\begin{bmatrix}\left\langle \widetilde{f(e_1)}, \widetilde{f(ys_j)}\right\rangle\\ \left\langle \widetilde{f(e_2)}, \widetilde{f(ys_j)}\right\rangle\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \left\langle \widetilde{f(e_d)}, \widetilde{f(ys_j)}\right\rangle\end{bmatrix}\right\} v_j$$

- Let E denote the column space of A $dim(E) \leq m$.
- Define $K = E \cap (12\varepsilon B_{\ell_\infty^d})$, $\forall j \; v_j \in K$

- $A \in \mathbb{R}^{d \times m}$
- $\widetilde{f(ys_j)} \in \mathbb{R}^m$
- $v_j \in \mathbb{R}^d$

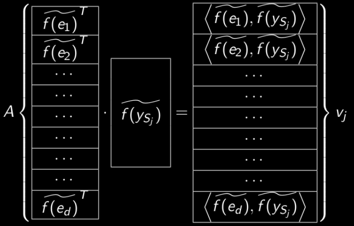
An encoding of X into $O(nm)$ bits



- Let E denote the column space of A $dim(E) \leq m$.
- Define $K = E \cap (12\varepsilon B_{\ell_\infty^d})$, $\forall j\ v_j \in K$
- K has $\frac{1}{12}$-net in K-norm of size $\leq 2^{O(m)}$

- $A \in \mathbb{R}^{d \times m}$
- $\widetilde{f(ys_j)} \in \mathbb{R}^m$
- $v_j \in \mathbb{R}^d$

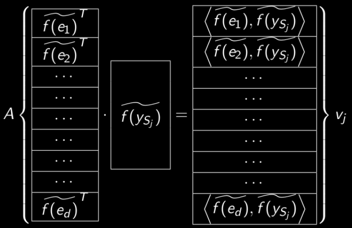
An encoding of X into $O(nm)$ bits



- $A \in \mathbb{R}^{d \times m}$
- $\widetilde{f(y_{S_j})} \in \mathbb{R}^m$
- $v_j \in \mathbb{R}^d$

- Let E denote the column space of A $dim(E) \leq m$.
- Define $K = E \cap (12\varepsilon B_{\ell_\infty^d})$, $\forall j\ v_j \in K$
- K has $\frac{1}{12}$-net in K-norm of size $\leq 2^{O(m)}$
- Define $\tilde{v}_j$ as closest point in this net to v_j which contains v_j. $O(m)$ bits.
  $\implies \|v_j - \tilde{v}_j\|_\infty < \varepsilon$. Done?

An encoding of X into $O(nm)$ bits



- $A \in \mathbb{R}^{d \times m}$
- $\widetilde{f(y_{S_j})} \in \mathbb{R}^m$
- $v_j \in \mathbb{R}^d$

- Let E denote the column space of A
  $dim(E) \leq m$.
- Define $K = E \cap (12\varepsilon B_{\ell_\infty^d})$, $\forall j \; v_j \in K$
- K has $\frac{1}{12}$-net in K-norm of size $\leq 2^{O(m)}$
- Define $\tilde{v}_j$ as closest point in this net to v_j
  which contains v_j. $O(m)$ bits.
  $\implies \|v_j - \tilde{v}_j\|_\infty < \varepsilon$. Done?
- Encoding needs to specify E (i.e. A).
  Encode $\widetilde{f(e_i)}$ using $O(m \log(1/\varepsilon))$ bits
  For the e_i: $O(dm \lg(1/\varepsilon)) = O(nm)$ bits

An encoding of X into $O(nm)$ bits

$$A\left\{\begin{bmatrix}\widetilde{f(e_1)}^T\\ \widetilde{f(e_2)}^T\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \widetilde{f(e_d)}^T\end{bmatrix}\right. \cdot \begin{bmatrix}\widetilde{f(y_{S_j})}\end{bmatrix} = \left.\begin{bmatrix}\langle \widetilde{f(e_1)}, \widetilde{f(y_{S_j})}\rangle\\ \langle \widetilde{f(e_2)}, \widetilde{f(y_{S_j})}\rangle\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \cdots\\ \langle \widetilde{f(e_d)}, \widetilde{f(y_{S_j})}\rangle\end{bmatrix}\right\} v_j$$

- $A \in \mathbb{R}^{d\times m}$
- $\widetilde{f(y_{S_j})} \in \mathbb{R}^m$
- $v_j \in \mathbb{R}^d$

- Let E denote the column space of A
  $dim(E) \leq m$.
- Define $K = E \cap (12\varepsilon B_{\ell_\infty^d})$, $\forall j\ v_j \in K$
- K has $\frac{1}{12}$-net in K-norm of size $\leq 2^{O(m)}$
- Define $\tilde{v}_j$ as closest point in this net to v_j
  which contains v_j. $O(m)$ bits.
  $\implies \|v_j - \tilde{v}_j\|_\infty < \varepsilon$. Done?
- Encoding needs to specify E (i.e. A).
  Encode $\widetilde{f(e_i)}$ using $O(m\log(1/\varepsilon))$ bits
  For the e_i: $O(dm\lg(1/\varepsilon)) = O(nm)$ bits
- **Total:** $O(nm)$ bit encoding

QED

QED

What about when $d \neq n/\lg(1/\varepsilon)$?

QED

What about when $d \neq n/\lg(1/\varepsilon)$?

(We were originally doing something a little more complicated, but Oded Regev pointed out the following simple argument.)

Extending to arbitrary d, n

- Suppose $X \subset \ell_2^{d'}$, $|X| = n$, is a hard set for some ε where $d' = \Theta(n/\log(1/\varepsilon))$ (X has $\Omega(\varepsilon^{-2}\log n)$ lower bound).

Extending to arbitrary d, n

- Suppose $X \subset \ell_2^{d'}$, $|X| = n$, is a hard set for some ε where $d' = \Theta(n/\log(1/\varepsilon))$ (X has $\Omega(\varepsilon^{-2}\log n)$ lower bound).
- $d > d'$: we have a hard set for any $d > d'$ by zero-padding vectors in X. Thus, name of the game: for fixed n, ε, show lower bound for as small a d as possible.

Extending to arbitrary d, n

- Suppose $X \subset \ell_2^{d'}$, $|X| = n$, is a hard set for some ε where $d' = \Theta(n/\log(1/\varepsilon))$ (X has $\Omega(\varepsilon^{-2}\log n)$ lower bound).
- $d > d'$: we have a hard set for any $d > d'$ by zero-padding vectors in X. Thus, name of the game: for fixed n, ε, show lower bound for as small a d as possible.
- $d < d'$: For $d \geq C\varepsilon^{-2}\log n$: we know $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^d$ exists (JL *upper* bound).

Extending to arbitrary d, n

- Suppose $X \subset \ell_2^{d'}$, $|X| = n$, is a hard set for some ε where $d' = \Theta(n/\log(1/\varepsilon))$ (X has $\Omega(\varepsilon^{-2}\log n)$ lower bound).
- $d > d'$: we have a hard set for any $d > d'$ by zero-padding vectors in X. Thus, name of the game: for fixed n, ε, show lower bound for as small a d as possible.
- $d < d'$: For $d \geq C\varepsilon^{-2}\log n$: we know $(1+\varepsilon)$-distortion embedding $f : X \to \ell_2^d$ exists (JL *upper* bound).

  $f(X)$ must be hard, else if good low-dimensional/low-distortion embedding g exists, then $g \circ f$ is a good low-distortional embedding for X (which we know doesn't exist).

What next?

Static approximate dot product

Two days after [Larsen, Nelson '17]

- Noga Alon: "Hi Jelani, Kasper, I wonder ... if you can get a tight estimate for the number of possibilities for the $\binom{n}{2}$ distances among n vectors of length at most 1 ..."

Static approximate dot product

Two days after [Larsen, Nelson '17]

- Noga Alon: "Hi Jelani, Kasper, I wonder ... if you can get a tight estimate for the number of possibilities for the $\binom{n}{2}$ distances among n vectors of length at most 1 ..."
- 4 later: problem solved! (for knowing up to additive ε)

  [Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\texttt{query}(i, j) = \langle x_i, x_j \rangle \pm \varepsilon$ for any $x_i, x_j \in X$.

Static approximate dot product

Two days after [Larsen, Nelson '17]

- Noga Alon: "Hi Jelani, Kasper, I wonder ... if you can get a tight estimate for the number of possibilities for the $\binom{n}{2}$ distances among n vectors of length at most 1 ..."
- 4 later: problem solved! (for knowing up to additive ε)

  [Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\texttt{query}(i, j) = \langle x_i, x_j \rangle \pm \varepsilon$ for any $x_i, x_j \in X$.
- Here $f(n, d, \varepsilon)$ is a bound they prove optimal for this problem

$$f(n,d,\varepsilon) = \begin{cases} \frac{n \log n}{\varepsilon^2}, & \frac{\log n}{\varepsilon^2} \le d \le n \\ nd \log(2 + \frac{\log n}{\varepsilon^2 d}), & \log n \le d \le \frac{\log n}{\varepsilon^2} \\ nd \log(1/\varepsilon), & 1 \le d \le \log n \end{cases}$$

Static approximate dot product

Two days after [Larsen, Nelson '17]

- Noga Alon: "Hi Jelani, Kasper, I wonder ... if you can get a tight estimate for the number of possibilities for the $\binom{n}{2}$ distances among n vectors of length at most 1 ..."
- 4 later: problem solved! (for knowing up to additive ε)

  [Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\texttt{query}(i, j) = \langle x_i, x_j \rangle \pm \varepsilon$ for any $x_i, x_j \in X$.
- Here $f(n, d, \varepsilon)$ is a bound they prove optimal for this problem

$$f(n, d, \varepsilon) = \begin{cases} \frac{n \log n}{\varepsilon^2}, & \frac{\log n}{\varepsilon^2} \le d \le n \\ nd \log(2 + \frac{\log n}{\varepsilon^2 d}), & \log n \le d \le \frac{\log n}{\varepsilon^2} \\ nd \log(1/\varepsilon), & 1 \le d \le \log n \end{cases}$$

- First case for d, upper bound for this data structural problem achieved earlier by [Kushilevitz, Ostrovsky, Rabani '98]

Static approximate dot product

[Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\texttt{query}(i, j) = \langle x_i, x_j \rangle + O(\varepsilon)$ for any $x_i, x_j \in X$.

Static approximate dot product

[Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\mathtt{query}(i, j) = \langle x_i, x_j \rangle + O(\varepsilon)$ for any $x_i, x_j \in X$.

- Proof also via encoding argument.

  Essentially the problem is equivalent to the following: let $\mathcal{G}$ be the set of all $n \times n$ Gram matrices of rank d and diagonal entries ≤ 1. What is the logarithm of the size of the smallest ε-net of $\mathcal{G}$ under entrywise ℓ_∞-norm?

  Encode X as name of closest net point to its Gram matrix.

Static approximate dot product

[Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\mathtt{query}(i, j) = \langle x_i, x_j \rangle + O(\varepsilon)$ for any $x_i, x_j \in X$.

- Proof also via encoding argument.

  Essentially the problem is equivalent to the following: let $\mathcal{G}$ be the set of all $n \times n$ Gram matrices of rank d and diagonal entries ≤ 1. What is the logarithm of the size of the smallest ε-net of $\mathcal{G}$ under entrywise ℓ_∞-norm?

  Encode X as name of closest net point to its Gram matrix.

- **Also implies optimal JL lower bound!**

  $f(n, n, 2\varepsilon) \leq f(n, m, \varepsilon)$ if low-distortion embedding into ℓ_2^m existed (first embed points then build data structure)

Static approximate dot product

[Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\mathtt{query}(i, j) = \langle x_i, x_j \rangle + O(\varepsilon)$ for any $x_i, x_j \in X$.

- Proof also via encoding argument.

  Essentially the problem is equivalent to the following: let $\mathcal{G}$ be the set of all $n \times n$ Gram matrices of rank d and diagonal entries ≤ 1. What is the logarithm of the size of the smallest ε-net of $\mathcal{G}$ under entrywise ℓ_∞-norm?

  Encode X as name of closest net point to its Gram matrix.

- **Also implies optimal JL lower bound!**

  $f(n, n, 2\varepsilon) \leq f(n, m, \varepsilon)$ if low-distortion embedding into ℓ_2^m existed (first embed points then build data structure)

- But [AK'17] gave upper bound on $f(n, m, \varepsilon)$, so m can't be too small lest their lower bound on $f(n, n, 2\varepsilon)$ be violated.

Static approximate dot product

[Alon, Klartag '17]: Given $X \subset S^{d-1}$, $|X| = n$, can create a data structure consuming $f(n, d, \varepsilon)$ bits such that can answer $\texttt{query}(i, j) = \langle x_i, x_j \rangle + O(\varepsilon)$ for any $x_i, x_j \in X$.

- **OPEN:**
  - **dynamic** approx. dot product with fast update/query?
  - approximate distance query with **relative** $1 + \varepsilon$ error? (see [Indyk, Wagner '17]; potential gap of $\lg(1/\varepsilon)$ remains)

And yet there's more

- **Conjecture:** ([Larsen, Nelson '17]) If $s(n, d, \varepsilon)$ is the optimal m for distortion $1 + \varepsilon$ for n-point subsets of ℓ_2^d, then $s(n, d, \varepsilon) = \Theta(\min\{n, d, \varepsilon^{-2} \log(2 + \varepsilon^2 n)\})$ for all ε, n, d.
  (i.e. JL is suboptimal for ε approaching $1/\sqrt{n}$)

- **Conjecture:** ([Larsen, Nelson '17]) If $s(n, d, \varepsilon)$ is the optimal m for distortion $1 + \varepsilon$ for n-point subsets of ℓ_2^d, then $s(n, d, \varepsilon) = \Theta(\min\{n, d, \varepsilon^{-2}\log(2 + \varepsilon^2 n)\})$ for all ε, n, d.

  (i.e. JL is suboptimal for ε approaching $1/\sqrt{n}$)
- [Alon, Klartag '17]: some progress toward conjecture. Proved lower bound. As for upper bound ...

- **Conjecture:** ([Larsen, Nelson '17]) If $s(n, d, \varepsilon)$ is the optimal m for distortion $1 + \varepsilon$ for n-point subsets of ℓ_2^d, then $s(n, d, \varepsilon) = \Theta(\min\{n, d, \varepsilon^{-2} \log(2 + \varepsilon^2 n)\})$ for all ε, n, d.

  (i.e. JL is suboptimal for ε approaching $1/\sqrt{n}$)
- [Alon, Klartag '17]: some progress toward conjecture. Proved lower bound. As for upper bound ...

  for *bipartite* version of problem with $x_1, \ldots, x_n, y_1, \ldots, y_n$ of unit norm, can show there exist $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}^m$ for $m = O(\varepsilon^{-2} \log(2 + \varepsilon^2 n))$ with

$$\forall i, j \; |\langle x_i, y_j \rangle - \langle a_i, b_j \rangle| \le \varepsilon$$

- **Conjecture:** ([Larsen, Nelson '17]) If $s(n, d, \varepsilon)$ is the optimal m for distortion $1 + \varepsilon$ for n-point subsets of ℓ_2^d, then $s(n, d, \varepsilon) = \Theta(\min\{n, d, \varepsilon^{-2} \log(2 + \varepsilon^2 n)\})$ for all ε, n, d.

  (i.e. JL is suboptimal for ε approaching $1/\sqrt{n}$)
- [Alon, Klartag '17]: some progress toward conjecture. Proved lower bound. As for upper bound ...

  for *bipartite* version of problem with $x_1, \ldots, x_n, y_1, \ldots, y_n$ of unit norm, can show there exist $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}^m$ for $m = O(\varepsilon^{-2} \log(2 + \varepsilon^2 n))$ with

$$\forall i, j \mid \langle x_i, y_j \rangle - \langle a_i, b_j \rangle \mid \leq \varepsilon$$

- [Alon, Klartag '17] positive result on bipartite problem makes use of low M^*-estimate [Pajor, Tomczak-Jaegermann '86] and Khatri-Sidak lemma [Khatri '67], [Sidak '67].

More open problems

Open problems

- Improved upper bound for constructing incoherent vectors?

  Maybe [Alon '03] sharp and **Gilbert-Varshamov bound always suboptimal!?**
- Instance-wise optimality for ℓ_2 dimensionality reduction?

  What's the right m in terms of X itself? Bicriteria results?
- JL map that can be applied to x in time $\tilde{O}(m + \|x\|_0)$?

  $\|\cdot\|_0$ denotes support size
- Explicit DJL distribution with seed length $O(\log \frac{d}{\delta})$?
- **Rasmus Pagh:** Las Vegas algorithm for computing a JL map for set of n points faster than repeated random projections then checking?