Optimal JL lower bound

Saeed Odak

April 4, 2022

1 Introduction

Based on the different applications, dimensionality reduction scheme is a way to reduce the dimenstion of high-dimensional modern data while preserving the geometry of it. One of the highly impactful results in dimensionality reduction is called "Johnson-Lindenstrauss lemma".

Theorem 1 (Johnson and Lindenstrauss [2]) *Let $X \subset \mathbb{R}^d$ be any set of size n, and let $\epsilon \in (0, 1/2)$ be arbitrary. Then there exists a map $f : X \to \mathbb{R}^m$ for some $m = \mathcal{O}(\epsilon^{-2} \log n)$ such that*

$$\forall x, y \in X, (1-\epsilon)\|x-y\|_2^2 \leq \|f(x)-f(y)\|_2^2 \leq (1+\epsilon)\|x-y\|_2^2. \tag{1}$$

Besides the applications on streaming algorithms and nearest neighbor search, it has also applications in design of approximation algorithm for computational geometry problems; instead of solving the problem for high-dimensional data, we can have faster algorithm on the low-dimension data set.

Natural question that comes to one's mind is: "Is the $m = O(\epsilon^{-2} \log n)$ optimal?" or "Is there a point set X such that any $(1+\epsilon)$ distortion of X on lower dimensions based on the ℓ_2 metric requires dimension $m = O(\epsilon^{-2} \log n)$". In order to answer this type of questions and for the purpose of this project, we are going to study the result of Larsen and Nelson Larsen and Nelson [4] as shown below:

Theorem 2 (Larsen and Nelson [4]) *For any integers $n, d \geq 2$ and $\epsilon \in (\log^{0.5001} n/\sqrt{min\{n, d\}}, 1)$, there exists a set of points $X \subset \mathbb{R}^d$ of size n, such that any map $f : X \to \mathbb{R}^m$ providing the guarantee (1) must have*

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

Briefly, this theorem states that there is a hard point set of size n such that the reduction preserving the distance within error $(1+\epsilon)$ has image dimension at least $\Omega(\log \epsilon^2 n/\epsilon^2)$.

Rest of this report is organized in the following way: In section 2 we will see some notations that is required in the proof of theorem 2. Section 3 is dedicated to study the proof of theorem 2.

2 Preliminaries

To start the proof of theorem 2, we need the following definitions and notations.

Definition 1 *A convex body is a compact, convex subset of $\mathbb{R}^d$ with non-empty interior. A convex body is symmetric if $x \in K \Leftrightarrow -x \in K$.*

For a given metric space (X, d), we can define a symmetric convex body as follows:

$$K_d = \{x : \|x\|_d \leq 1\}.$$

On the other hand, for a given symmetric convex body, K, we can define a metric,

$$\|x\|_K = \max\{t : tx \in K\}.$$

Definition 2 *Let (X,d) be a metric space and $S \subseteq X$. For $\epsilon \geq 0$, an ϵ-net of S is a subset $S' \subseteq X$ such that for all $x \in S$, there exists $x' \in S'$ such that $d(x,x') \leq \epsilon$. Transversal of S, $\tau(S,d,\epsilon)$ is the minimum size of an ϵ-net of S with respect to metric space (X,d).*

For a set $S \subset \mathbb{R}^d$ and symmetric convex body $K \subset \mathbb{R}^d$, we denote $\tau(A,K)$ as minimum number of translations of K to cover A. it is not hard to see that

$$\tau(A, \|\cdot\|_K, \epsilon) = \tau(A, \epsilon K).$$

Following lemma brings an upper bound on the size of transversal of unit ball in $\mathbb{R}^d$.

Lemma 1 *Let $B(0,1)$ be the unit ball of some norm $\|\cdot\|$ in $\mathbb{R}^d$. Then $\tau(B, \|\cdot\|, \epsilon) \leq (1+2/\epsilon)^d$.*

Following is the lemma that introduce a property based on inner-product that we are willing to use instead of property (1).

Lemma 2 *Let $X \subset S^{d-1}$ be such that $0 \in X$ and $f : X \to \mathbb{R}^m$ satisfies $f(0) = 0$ and property (1), then*

$$\forall x, y \in X \setminus \{0\} \quad : \quad |\langle f(x), f(y)\rangle - \langle x, y\rangle| \leq 3\epsilon. \tag{2}$$

This lemma states that if an embedding preserves the distance with error $1+\epsilon$ (Property (1)) then it would preserve the inner-product also with in additive error of 3ϵ (Property (2)).

3 Proof of Optimal JL Lemma

Let $k = 1/(400\epsilon^2)$. For a given subset $S \subseteq [d]$ of size k, define $y_S = (1/\sqrt{k})\sum_{i\in S} e_i$ where e_i is the ith standard vector of $\mathbb{R}^d$. We define a collection $\mathcal{P}$ of subsets of size n of $\mathbb{R}^d$. Let $\mathcal{P}$ consist of all n-point sets of form $X = \{0, e_1, e_2, \ldots, e_d, y_{S_1}, y_{S_2}, \ldots, y_{S_{n-d-1}}\}$, where S_j is k-subset of $[d]$. It is clear that $|\mathcal{P}| = \binom{d}{k}^{(n-d-1)}$. Another observation is that $\langle e_i, y_{S_j}\rangle = 20\epsilon$ if $i \in S_j$ and otherwise $\langle e_i, y_{S_j}\rangle = 0$. Therefore knowing the value of $\langle e_i, y_{S_j}\rangle$ for $1 \leq i \leq d$ and $1 \leq j \leq n-d-1$ is enough to determine X completely.

First, we prove theorem 2 with assumption $d = n/\log(1/\epsilon)$ and at the end we extend the argument to to other values of dimension d. Assume the contrary, we consider that there is an embedding f_X for every $X \in \mathcal{P}$ that is satisfying property (1) thus by lemma 2, satisfying property (2) such that $m \ll \epsilon^{-2}\log n$, where m is the dimension of the image space. We will introduce an injective function $g : \mathcal{P} \to \{0,1\}^{O(nm)}$. Thus $nm = \Omega(\log|\mathcal{P}|) = \Omega(nk\log d/k)$. Assuming $d = n/\log(1/\epsilon)$ and $\epsilon > \log^{0.5001} n/\sqrt{d} \geq \log^{0.5001} n/\sqrt{n}$, we get that

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

which is the contradiction we are looking for.

Construction of g

We are aiming to encode each set X in the collection $\mathcal{P}$ using $O(nm)$ bits. we are aware that for $1 \leq i \leq d$ and $1 \leq j \leq n-d-1$, knowing the values of $\langle e_i, y_{S_j}\rangle$ is enough to determine X. On the other hand by lemma 2, $|\langle f(e_i), f(y_{S_j})\rangle - \langle e_i, y_{S_j}\rangle| \leq 3\epsilon$. So $\langle f(e_i), f(y_{S_j})\rangle \leq 3\epsilon$ if $i \notin S_j$ and $\langle f(e_i), f(y_{S_j})\rangle \geq 17\epsilon$ if $i \in S_j$. Thus knowing the values of $\langle f(e_i), f(y_{S_j})\rangle$ for $1 \leq i \leq d$ and $1 \leq j \leq n-d-1$ is also enough to determine X. Very naive idea to encode set X is to concatenate the entries of $f(X) := \{(0), f(e_1), f(e_2), \ldots, f(e_d), f(y_{S_1}), f(y_{S_2}), \ldots, f(y_{S_{n-d-1}})\}$ which is obviously not an $O(nm)$ bit string.

First attempt: As a first attempt, we take a small value γ and round the every entry of vectors in $f(X)$ to the closes multiple of γ. Since each vector in the image space has norm at most $1 \pm \epsilon$, we know that every entry of $f(X)$ is within interval $[-1-\epsilon, 1+\epsilon]$. So there are at most $\lceil (2+2\epsilon)/\gamma \rceil = O(1/\gamma)$ values that each entry can be rounded to. In consequence, encoding a n-point set X consumes at most $O(nd \log(1/\gamma))$ bits. In order to decode the rounded vectors, we have to study the error in the inner product. For $1 \le i \le d$ and $1 \le j \le n-d-1$,

$$\sum_{t=1}^{m} (f_t(e_i) \pm \gamma)(f_t(y_{S_j}) \pm \gamma) = \langle f(e_i), f(y_{S_j}) \rangle \pm \gamma \|f(e_i)\|_1 \pm \gamma \|f(y_{S_j})\|_1 \pm m\gamma^2$$

By setting $\gamma = \Theta(\epsilon/\sqrt{m})$ and using Cauchy-Schwarz, we can guarantee an $\pm\epsilon$ error in the inner products in the rounded vectors which is still recoverable due to 14ϵ distance. So with the value of $\gamma = \Theta(\epsilon/\sqrt{m})$, there is a requirement of $O(nd \log{(m/\epsilon)})$ bits to store all the rounded vectors and this quantity should be at least $\Omega(n\epsilon^{-2} \log n)$, resulting in $m = \Omega(\epsilon^{-2} \frac{\log n}{\log{(1/\epsilon)} + \log\log n})$.

Second attempt: With a close look at the previous definition of g, we are actually picking a γ-net B' of $B_{\ell_2}(0, 1+\epsilon) \subset \mathbb{R}^m$ under ℓ_∞ norm and we round each element of $f(X)$ to the closest element in B'. As a second attempt we are considering a γ-net B'' of $B_{\ell_2}(0, 1+\epsilon) \subset \mathbb{R}^m$ under ℓ_2 norm and round each element of $f(X)$ to the closest element of B''. For the inner-product error we have:

$$\langle \overline{f}(x), \overline{f}(y) \rangle = \langle f(x) + (\overline{f}(x) - f(x)), f(y) + (\overline{f}(y) - f(y)) \rangle$$
$$= \langle f(x), f(y) \rangle + \langle f(x), \overline{f}(y) - f(y) \rangle + \langle \overline{f}(x) - f(x), f(y) \rangle + \langle \overline{f}(x) - f(x), \overline{f}(y) - f(y) \rangle$$
$$= \langle f(x), f(y) \rangle + O(\gamma).$$

Again by setting $\gamma = c\epsilon$ for some small constant c and Cauchy-Schwarz, we get $\langle \overline{f}(x), \overline{f}(y) \rangle = \langle f(x), f(y) \rangle \pm \epsilon$. Thus $\langle \overline{f}(x), \overline{f}(y) \rangle \le 4\epsilon$ if $i \in S_j$ and $\langle \overline{f}(x), \overline{f}(y) \rangle \ge 16\epsilon$ if $i \notin S_j$. Cardinality of B'', by lemma 1 is at most $O(1/\epsilon)^m$. so we need $O(nm \log 1/\epsilon)$ bits to encode every element of $\mathcal{P}$. Now with the lower bound $\Omega(\log |\mathcal{P}|)$, we get $m = \Omega(\epsilon^{-2} \frac{\log n}{\log{(1/\epsilon)}})$ which is the result of Alon [1].

Final attempt Based on the previous attempt, we define $v_j \in \mathbb{R}^d$, by $v_j(i) = \langle \overline{f}(e_i), \overline{f}(y_{S_j}) \rangle$ for $i \in [d]$. Then it is enough to know v_j's to restore the X. Even an approximation of v_j is also enough to restore X. For each $j \in [n-d-1]$, we round each v_j to $\overline{v}_j$ such that $\|\overline{v}_j - v_j\|_\infty \le \epsilon$. This is similar to the first attempt, we are rounding each v_j to the closest element in an ϵ-net under ℓ_∞ norm. For $i \in [d]$ and $j \in [n-d-1]$, to recognize whether $i \in S_j$ or $i \notin S_j$, $\overline{v}_j(i)$ is at least 15ϵ or at most 5ϵ, respectively.

Let A be a $d \times m$ matrix with ith row equal to $\overline{f}(e_i)$. Then for each $j \in [n-d-1]$, $v_j = Ay_{s_j}$. This means v_j is column space, C, of A. so $dim(C) \le m$. On other hand, $\langle \overline{f}(e_i), \overline{f}(y_{S_j}) \rangle \le \langle e_i, y_{S_j} \rangle \pm 4\epsilon$ implies that $\|v_j\|_\infty \le 24\epsilon$. Let $K = C \cap B_{\ell_\infty^d}(0, 24\epsilon)$. So $v_j \in K$ and K is a symmetric convex body. Let K' be the 1/24th net of K. That means for each $v_j \in K$ there is $\overline{v}_j \in K'$ such that $\|\overline{v}_j - v_j\| \le \epsilon$. By lemma 1, $|K'|$ is at most $O(1 + (2/(1/24)))^m = O(1)^m$. So each $\overline{v}_j$'s can be encoded by $O(m)$ bits, resulting in $O(nm)$ bit encoding of all values of $\overline{v}_j$. To decode this scheme, we need to store the matrix A too, that requires $O(dm \log{(1/\epsilon)}) = O(nm)$ extra bits, since $d = n/\log{(1/\epsilon)}$. This finishes the proof for the case when $d = n/\log{(1/\epsilon)}$.

Handling other values of d:

- $d > n/\log{(1/\epsilon)}$: Take the hard point set in dimension $n/\log{(1/\epsilon)}$ and add zeroes to the vectors to make them dimension d.
- $d < n/\log{(1/\epsilon)}$: Assume that there is no hard set for $d < n/\log{(1/\epsilon)}$. Let $d' = n/\log{(1/\epsilon)}$. We take any n-point set P in $\mathbb{R}^{d'}$ and apply JL Lemma (theorem 1) to P to get new point set P' on

dimension d with distance preserved within error $O(1+\sqrt{\frac{\log n}{d}})$. Now we use the assumption that P' is not a hard set for dimension d. Therefore, there is a embedding that uses $m = o(\epsilon^{-2} \log n)$ dimension. Thus the combined embedding of JL lemma and hypothetical embedding will give an embedding on P with distance preservation $O(1+\sqrt{\frac{\log n}{d}})O(1+\epsilon) = O(1+\epsilon)$, since $\epsilon > \log^{0.5001} n/\sqrt{d}$. Contradiction with lower bound for the embedding of dimension d'.

References

[1] N. Alon. Problems and results in extremal combinatorics–i. *Discret. Math.*, 273(1-3):31–53, 2003. doi: 10.1016/S0012-365X(03)00227-9. URL https://doi.org/10.1016/S0012-365X(03)00227-9.

[2] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[3] K. G. Larsen and J. Nelson. The johnson-lindenstrauss lemma is optimal for linear dimensionality reduction. In I. Chatzigiannakis, M. Mitzenmacher, Y. Rabani, and D. Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 82:1–82:11. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. doi: 10.4230/LIPIcs.ICALP.2016.82. URL https://doi.org/10.4230/LIPIcs.ICALP.2016.82.

[4] K. G. Larsen and J. Nelson. Optimality of the johnson-lindenstrauss lemma. In C. Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 633–638. IEEE Computer Society, 2017. doi: 10.1109/FOCS.2017.64. URL https://doi.org/10.1109/FOCS.2017.64.

[5] J. Nelson. Sketching algorithms. *December 3*, 2020.