16 Embeddings of the Euclidean metric

In today's lecture, we will consider how well we can embed n points in the Euclidean metric (l_2) into other l_p metrics. More formally, we ask the following question.

Question 16.1 (Isometric Embedding $l_2 \to l_p$) *Given a n-point metric $(X, ||\cdot||_2) \subseteq \mathbb{R}^n$, does there exist an* isometric *embedding of X into l_p ($p > 1$), i.e., is there a $k \in \mathbb{Z}$ such that*

$$(X, ||\cdot||_2) \xhookrightarrow{1} l_p^k \qquad ?$$

We will answer this question in the affirmative in today's lecture.

The above question is trivial true when the target space is also l_2 (i.e., $p = 2$). In this case, we will instead ask the following question. Can we reduce the dimension of the target l_2 space significantly if we allow for a distortion D. More formally,

Question 16.2 (Dimensionality Reduction for l_2) *Given a n-point metric $(X, ||\cdot||_2) \subset \mathbb{R}^n$ and a distortion factor $D \in (1, \infty)$, what is the smallest $k = k(n, D)$ such that there exists an embedding*

$$(X, ||\cdot||_2) \xhookrightarrow{D} l_2^k \qquad ?$$

We will show in today's lecture that such a dimensionality reduction is in fact possible (as shown by a seminal result of Johnson and Lindenstrauss [JL84]).

Theorem 16.3 (Johnson Lindenstrauss [JL84]) *For any n-point metric $(X, ||\cdot||_2) \subset \mathbb{R}^n$ and $\epsilon > 0$, there exists an embedding*

$$(X, ||\cdot||_2) \xhookrightarrow{1+\epsilon} l_2^{O(\log n/\epsilon^2)}.$$

In other words, the dimension can be dramatically reduced from n to $O\left(\frac{\log n}{\epsilon^2}\right)$ if we are willing to allow a distortion of at most $1 + \epsilon$.

Question 16.2 can also be asked for other l_p spaces when $p \neq 2$. However, such a significant dimension reduction does not seem to be possible when $p \neq 2$.

It turns out that when $p = \infty$, almost no dimension reduction is possible. For the case $p = 1$, Brinkman and Charikar [BC05] showed the following lower bound. (A simpler exposition of the same bound was given by Lee and Naor [LN04].)

Theorem 16.4 *There exists a n-point metric $(X, ||\cdot||_1)$ in l_1 such that any embedding $X \xhookrightarrow{D} l_1^k$ with distortion D satisfies $k = n^{\Omega(1/D^2)}$.*

16.1 Random Projection Method

We will use what is known as the random projection method for proving both the dimensionality reduction of l_2 as well as isometric embeddings of l_2 into l_p. Informally, a random projection from $\mathbb{R}^n$ to $\mathbb{R}^k$ can be described as follows: Let $r_1, \ldots, r_k$ be "random" vectors in $\mathbb{R}^n$ obtained independently from some random process. Consider the map φ described as follows:

$$v \in \mathbb{R}_n \xrightarrow{\varphi} (\langle v, r_1 \rangle, \ldots \langle v, r_k \rangle) \in \mathbb{R}^k$$

The map φ describes a random projection of the space $\mathbb{R}^n$ into $\mathbb{R}^k$. If A is the $k \times n$ matrix whose rows are the k vectors $r_1, \ldots, r_k$, then the map φ can be written as $\varphi(v) = Av$. The nice property about such a map is that it is linear. In other words, for all $v_1, v_2 \in \mathbb{R}^n$, we have $\varphi(v_1 - v_2) = \varphi(v_1) - \varphi(v_2)$.

The following are some of the typical random processes considered to generate the random vectors $r_1, \ldots, r_k$.

- Choose $r_i = (r_i^1, \ldots, r_i^n), i = 1 \ldots, k$ with each $r_i^j \sim N(0, 1)$ i.e., each co-ordinate is chosen independently according to the normal distribution with mean 0 and variance 1. Let us call the random projection corresponding to this random process φ_N.
- Choose $r_i = (b_i^1, \ldots, b_i^n), i = 1, \ldots, k$ with each b_i^j is chosen independently to be $+1$ or -1 with probability $1/2$ each. Let us call this random projection φ_B.
- Choose $r_1, \ldots, r_k$ to be a set of k orthogonal vectors from the unit sphere $\mathcal{S}^{n-1}$ of radius 1[1]. Call this random projection φ_S.

In fact, each of these random projections gives a different proof of the The Johnson-Lindenstrauss Theorem. The original proof of Johnson-Lindenstrauss [JL84] used the random projection φ_S, while Indyk and Motwani [IM98] gave a proof using the projection φ_N. Achlioptas [Ach03] showed that a even simpler random process such as φ_B suffices to prove the theorem. In today's lecture, we will follow the exposition of Dasgupta and Gupta [DG03], which is along the lines of that of Indyk and Motwani [IM98].

16.2 Proof of Johnson-Lindenstrauss Theorem

Lemma 16.5 *Let φ_N be the random projection as defined in the earlier section where $k = O\left(\frac{\log n}{\epsilon^2}\right)$. For any point $v \in \mathbb{R}^n$, we have the following*

$$\mathbf{Pr}\left[(1-\epsilon) \leq \frac{||\varphi_N(v)||_2}{\sqrt{k}||v||_2} \leq 1+\epsilon\right] \geq 1 - \frac{1}{n^2}.$$

The Johnson-Lindenstrauss theorem is then obtained by applying the above lemma to all points $u - v$ where $u, v \in X$ and a simple union bound.

[1] Random vectors according to this vector are generated as follows: Choose k vectors $r_1', \ldots, r_k'$ by choosing each co-ordinate according to the normal distribution $N(0, 1)$ (as in the random projection φ_N). Perform the Gram-Schmidt orthogonalization process to these vectors obtain k orthogonal vectors. These vectors are then orthogonalized to obtain the required vectors $r_1, \ldots, r_k$

Proof.

Recall that the random projection φ_N maps the point $v \in \mathbb{R}^n$ to the the point $(\langle v, r_1\rangle, \ldots \langle v, r_k\rangle) \in \mathbb{R}^k$ where each co-ordinate r_i^j is chosen according to the normal distribution $N(0,1)$.

Wlog, we will assume that the length of the vector $v = (v_1, \ldots, v_n)$ is 1. Consider the random variables $X_i = \langle v, r_i\rangle$. Due to the properties of the normal distribution, each of these random variables X_i is also distributed according to the normal distribution $N(0,1)$. Let Y denote the random variable $||\varphi_N(v)||_2^2$. Note that $Y = \sum X_i^2$. We now make the following simple observations.

$$
\begin{aligned}
X_i &= \langle v, r_i\rangle = \sum_j v_j \cdot r_i^j \\
\mathbf{E}[X_i] &= \sum_j v_j \mathbf{E}[r_j^i] = 0 \\
\mathbf{Var}[X_i] &= \sum_j v_j^2 \mathbf{Var}\left[r_i^j\right] = \sum_j v_j^2 = 1 \\
\mathbf{E}[Y] &= \sum \mathbf{E}[X_i^2] = \sum \left(\mathbf{Var}[X_i] + (\mathbf{E}[X_i])^2\right) \\
&= \sum \mathbf{Var}[X_i] = k
\end{aligned}
$$

Thus, the random variable $||\varphi_N(v)||_2^2 = Y$ has expectation k. We will now show that in fact, Y is concentrated around its mean.

$$
\begin{aligned}
\mathbf{Pr}[Y \geq (1+\epsilon)k] &= \mathbf{Pr}\left[e^{sY} \geq e^{s(1+\epsilon)k}\right] \qquad \text{[for all } s > 0] \\
&\leq \mathbf{E}\left[e^{sY}\right] / e^{s(1+\epsilon)k} \qquad \text{[By Markov's inequality]} \\
&= e^{-s(1+\epsilon)k} \cdot \mathbf{E}\left[\prod e^{sX_i^2}\right] \qquad \text{[since } Y = \sum X_i^2] \\
&= e^{-s(1+\epsilon)k} \cdot \left(\mathbf{E}\left[e^{sX_1^2}\right]\right)^2
\end{aligned}
$$

We know that X_1 is distributed according to the normal distribution $N(0,1)$. In other words, X_1 has the density function $e^{-t^2/2}$.

$$
\begin{aligned}
\mathbf{E}\left[e^{sX_1^2}\right] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{st^2} \cdot e^{-t^2/2} dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{t^2(1-2s)/2} dt \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-2s}} \int_{\mathbb{R}} e^{-z^2/2} dz \qquad \text{[Substituting } z = \sqrt{1-2s} \text{ assuming } s < \tfrac{1}{2}] \\
&= \frac{1}{\sqrt{1-2s}}
\end{aligned}
$$

We thus have

$$
\mathbf{Pr}[Y \geq (1+\epsilon)k] \leq \frac{e^{-s(1+\epsilon)k}}{(1-2s)^{k/2}} = \left(\frac{e^{-s(1+\epsilon)}}{\sqrt{1-2s}}\right)^k.
$$

The choice of s that minimizes the above expression is $s = \frac{\epsilon}{2(1+\epsilon)}$. Substituting this s into the above expression, we obtain

$$\begin{aligned}
\mathbf{Pr}[Y \geq (1+\epsilon)k] &\leq \left(\frac{e^{\epsilon}}{1+\epsilon}\right)^{-\frac{k}{2}} \\
&= \exp\left[-\frac{k}{2}\left(\epsilon - \log(1+\epsilon)\right)\right] \\
&\leq \exp\left[-\frac{k}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)\right] \qquad \text{[Using the Taylor expansion of } \log(1+\epsilon)] \\
&\leq \frac{1}{2n^2} \qquad \text{if } k \geq \frac{2\log 2n^2}{\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}}
\end{aligned}$$

A similar argument shows that $\mathbf{Pr}[Y \leq (1-\epsilon)k] \leq 1/2n^2$. This completes the proof of the Lemma. ■

16.3 Isometric Embeddings of l_2

We now turn to our first question of whether there exists an isometric embedding from l_2 to l_1. We will first demonstrate that there exists an isometric embedding of every n-point metric in l_2 into $l_1^{\mathcal{S}^{n-1}}$. $l_1^{\mathcal{S}^{n-1}}$ is an infinite dimensional l_1 metric which consists of a co-ordinate for each vector r in the unit sphere $\mathcal{S}^{n-1}$. Thus, a point in $l_1^{\mathcal{S}^{n-1}}$ is given by a function of the form $f : \mathcal{S}^{n-1} \to \mathbb{R}$. The l_1 norm of any such point f is defined as follows:

$$||f||_1 = \int_{r \in \mathcal{S}^{n-1}} |f(r)| dr.$$

Note that the space $l_1^{\mathcal{S}^{n-1}}$ is the set of Lebesgue integrable functions over the unit sphere $\mathcal{S}^{n-1}$ and must be actually denoted by $\mathcal{L}^1(\mathcal{S}^{n-1})$. Consider the embedding $\varphi : \mathbb{R}^n \to l_1^{\mathcal{S}n-1}$ defined as follows:

$$\begin{aligned}
v \in l_2^n &\xrightarrow{\varphi} f_v \in l_1^{\mathcal{S}^{n-1}} \\
f_v(r) &= \langle v, r \rangle
\end{aligned}$$

For any $v \in l_1^n$ we observe that

$$\begin{aligned}
||f_v||_1 &= \int_{r \in \mathcal{S}^{n-1}} |\langle v, r \rangle| dr \\
&= \int_{\theta \in \mathcal{S}^{n-1}} ||v||_2 \cdot |\cos\theta| d\theta \\
&= ||v||_2 \int_{\theta \in \mathcal{S}^{n-1}} |\cos\theta| d\theta \\
&= \alpha_1 ||v||_2
\end{aligned}$$

for some universal constant α_1. For any two point $u, v \in \mathbb{R}^n$, we have $||\varphi(u) - \varphi(v)||_1 = ||f_u - f_v||_1 = ||f_{u-v}||_1 = \alpha_1 ||u - v||_2$. Thus, φ is an isometric embedding from l_2 to $l_1^{\mathcal{S}^{n-1}}$.

Almost the same proof also shows that φ is also an isometric embedding from l_2 to $l_p^{\mathcal{S}^{n-1}}$.

$$\begin{aligned} ||f_v||_p &= \left(\int_{r \in \mathcal{S}^{n-1}} |\langle v, r \rangle|^p dr \right)^{\frac{1}{p}} \\ &= \left(\int_{\theta \in \mathcal{S}^{n-1}} ||v||_2^p \cdot |\cos\theta|^p d\theta \right)^{\frac{1}{p}} \\ &= ||v||_2 \left(\int_{\theta \in \mathcal{S}^{n-1}} |\cos\theta|^p d\theta \right)^{\frac{1}{p}} \\ &= \alpha_p ||v||_2 \end{aligned}$$

for some universal constant α_p.

Now using a large deviation bound similar to that of the Johnson-Lindenstrauss Lemma 16.5, we can show that if we pick $k = O\left(\log n / \epsilon^2\right)$ random vectors $r_1, \ldots, r_k$ on the unit sphere $\mathcal{S}^{n-1}$ and set $\varphi(v) = (\langle v, r_1 \rangle, \ldots, \langle v, r_k \rangle)$, then with high probability

$$(1 - \epsilon) \leq \frac{||\varphi(u)||_p}{\alpha_p ||u||_2} \leq (1 + \epsilon).$$

The above fact that there exists a $(1+\epsilon)$-near-isometric embedding of a n-point metric in l_2 into a *finite dimensional* l_p metric for every $\epsilon > 0$ can be used to show that there actually exists an *isometric embedding* of l_2 into a *finite dimensional* l_p. We will however not present this proof in lecture. We will instead present an alternate proof for the case when $p = 1$.

Given the n-point metric $X = \{x_1, \ldots, x_n\}$, partition the unit sphere $\mathcal{S}^{n-1}$ into $n!$ regions as follows: for every permutation $\pi \in \mathfrak{S}_n$ of the elements $\{1, \ldots, n\}$, define

$$S_\pi = \{r \in \mathcal{S}^{n-1} | \langle x_{\pi(1)}, r \rangle < \langle x_{\pi(2)}, r \rangle < \cdots < \langle x_{\pi(n)}, r \rangle \}.$$

For any pair of points $x_i, x_j \in X$, we observe that

$$\begin{aligned} ||f_{x_i} - f_{x_j}||_1 &= ||f_{x_i - x_j}||_1 \\ &= \int_{r \in \mathcal{S}^{n-1}} |\langle x_i - x_j, r \rangle| dr \\ &= \int_{r \in \mathcal{S}^{n-1}} |\langle x_i, r \rangle - \langle x_j, r \rangle| dr \\ &= \sum_\pi \int_{r \in S_\pi} |\langle x_i, r \rangle - \langle x_j, r \rangle| dr \\ &= \sum_\pi \left| \int_{r \in S_\pi} (\langle x_i, r \rangle - \langle x_j, r \rangle)\, dr \right| \\ &\qquad \text{(since the sign of } (\langle x_i, r \rangle - \langle x_j, r \rangle) \text{ is constant within each } S_\pi) \\ &= \sum_\pi \left| \int_{r \in S_\pi} \langle x_i, r \rangle dr - \int_{r \in S_\pi} \langle x_j, r \rangle dr \right| \end{aligned}$$

The above observation tells us that the following map from l_2^n to $l_1^{n!}$ is an isometric embedding.

$$\begin{aligned} x &\longmapsto (s_x^\pi : \pi \in \mathfrak{S}_n) \\ s_x^\pi &= \int_{r \in S_\pi} \langle x, r \rangle dr \end{aligned}$$

References

[Ach03] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.

[BC05] Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in l_1. *Journal of the ACM*, 52(5):766–788, 2005.

[DG03] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.

[IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th ACM Symposium on Theory of Computing (STOC)*, pages 604–613, 1998.

[JL84] William B. Johnson and Joram Lindenstrauss. Extensions of the Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[LN04] James Lee and Assaf Naor. Embedding the diamond graph in L_p and dimension reduction in l_1. *Geometric and Functional Analysis*, 14(4):745–747, 2004.