Lecture 5

Last time we discussed the sparsest cut problem and its connection to metric embeddings. If we can efficiently embed any metric (i.e. l_∞-embeddable) into l_1 with distortion α then we can derive an α-approximation algorithm for the sparsest cut problem. Similarly, for embedding a negative-type metric into l_1. In 1985, Bourgain [1] proved that any n-point metric space can be embedded into l_2 (hence, l_1 or any l_p for $p \geq 1$) with distortion $O(\log n)$. Metrics from expander graphs show this is tight. Recently, Arora, Lee and Naor [2] have shown that a distortion of $O(\sqrt{\log n} \cdot \log\log n)$ can be achieved to embed a negative type metric into Euclidean space. Today, we describe Bourgain's result and prove it.

Theorem 1. *Every metric space (X, d) with $|X| = n$ can be embedded into l_2 with distortion $O(\log n)$.*

Frechet Embedding. For a given metric space (X, d), distance between a point $x \in X$ and a set $A \subset X$ is defined as:

$$d(x, A) = \min_{y \in A} \; d(x, y).$$

Also, we can check that a function $f_A : X \to \mathbb{R}^1$ which maps x to $d(x, A)$ has the following non-expanding property. For $x \in X$, let $s \in X$ be such that $d(x, s) = d(x, A)$ and, similarly for $y \in X$, let $t \in X$ be such that $d(y, t) = d(y, A)$. Thus,

$$\begin{aligned} f_A(x) - f_A(y) &= d(x, A) - d(y, A) \\ &= d(x, s) - d(y, t) \\ &\leq d(x, t) - d(y, t) \leq d(x, y). \end{aligned}$$

The same is true if we exchange the roles of x and y and thus we have $|f_A(x) - f_A(y)| \leq d(x, y)$.

A *Frechet Embedding* of (X, d) is a probability distribution μ over all non-empty subsets of X. If A is a random subset distributed according to μ, then we associate, to every $x \in X$, the real-valued random variable $F_\mu(x) = d(x, A)$. Then, the following lemma holds.

Lemma 2. *Let (X, d) be a metric space with $|X| = n$. If, for a Frechet embedding μ, we can show that*

$$\forall x, y \in X, d(x, y) \leq \gamma E_\mu |F_\mu(x) - F_\mu(y)| = \gamma E_\mu[|d(x, A) - d(y, A)|]$$

then the mapping $G : X \to \mathbb{R}^{2^n}$ with[1] $G(x) = (\sqrt{\mu(A)}\; d(x, A))_{A \subset X}$ embeds (X, d) into l_2 with distortion γ.

[1]The coordinate value of $G(x)$ corresponding to A in $\mathbb{R}^{2^n}$ is $\sqrt{\mu(A)}\; d(x, A)$

Proof.

$$\forall x, y \in X,\ ||G(x) - G(y)||_2 = \left(\sum_{A \subseteq X} \mu(A)(d(x,A) - d(y,A))^2 \right)^{\frac{1}{2}}$$
$$\leq \left(\sum_{A \subseteq X} \mu(A) d(x,y)^2 \right)^{\frac{1}{2}}$$
$$= d(x,y).$$

Also, if we set $u, v \in \mathbb{R}^{2^n}$ as $u = (\sqrt{\mu(A)}\ |d(x,A) - d(y,A)|)_{A \subseteq X}, v = (\sqrt{\mu(A)})_{A \subseteq X}$, we can get the other direction of the inequality as follows:

$$\forall x, y \in X,\ ||G(x) - G(y)||_2 = ||u||_2 = ||u||_2 ||v||_2$$
$$\geq \langle u, v \rangle \text{ (by the Cauchy Schwartz inequality)}$$
$$= \sum_{A \subseteq X} \mu(A) |d(x,A) - d(y,A)|$$
$$= \mathbb{E}_\mu[|d(x,A) - d(y,A)|]$$
$$\geq \frac{1}{\gamma} d(x,y).$$

Therefore, for all $x, y \in X$, we have

$$||G(x) - G(y)||_2 \leq d(x,y) \leq \gamma ||G(x) - G(y)||_2.$$

□

For proving the main theorem, this lemma says that it suffices to find a Frechet embedding μ which satisfies the condition of the lemma with $\gamma = O(\log n)$.

Let $K = \{1, 2, \ldots, 2^p\}$ where $p = \lfloor \log_2(n) \rfloor$, and construct a Frechet embedding μ as follows:

$$\mu(A) = \begin{cases} \frac{1}{p+1} \frac{1}{\binom{n}{|A|}} & \text{if } |A| \in K \\ 0 & \text{if } |A| \notin K. \end{cases}$$

We need to verify the condition of Lemma 2 for every x and y. So, fix $x, y \in X$, and define some notions.

Definition 1.

- $B(x, \rho) = \{y \in X, d(x,y) \leq \rho\}$
- $\bar{B}(x, \rho) = \{y \in X, d(x,y) < \rho\}$
- $\rho_t = \min\{\rho\ :\ |B(x,\rho)| \geq 2^t\ \&\ |B(y,\rho)| \geq 2^t\}$.

By definition, $\rho_0 = 0$. Let l be a least index with $\rho_l \geq \frac{d(x,y)}{4}$. Then, $B(x, \rho_{l-1}) \cap B(y, \rho_{l-1}) = \phi$ because $\rho_{l-1} < \frac{d(x,y)}{4}$. Therefore, $2^l \leq n$ because $n \geq |B(x, \rho_{l-1}) \cup B(y, \rho_{l-1})| = |B(x, \rho_{l-1})| + |B(y, \rho_{l-1})| \geq 2^{l-1} + 2^{l-1} = 2^l$. Thus $l \leq p$. Now, fix $k = 2^j$ with $p - l \leq j \leq p - 1$, and let

$t = p - j$. (Hence, $1 \le t \le l$.) We can assume $|\bar{B}(x, \rho_t)| < 2^t$ without loss of generality, because either $|\bar{B}(x, \rho_t)| < 2^t$ or $|\bar{B}(y, \rho_t)| < 2^t$. Set

$$R_k = \{A \subseteq X : |A| = k,\ \bar{B}(x, \rho_t) \cap A = \phi,\ B(y, \rho_{t-1}) \cap A \neq \phi\}.$$

Then, the following lemma holds.

Lemma 3. *There exists a constant β (independent from k) such that*

$$|R_k| \ge \frac{\binom{n}{k}}{\beta}.$$

Proof. Let P and Q be $\bar{B}(x, \rho_t)$ and $B(y, \rho_{t-1})$ respectively for convenience. P and Q are disjoint, and $|P| < 2^t$, $|Q| \ge 2^{t-1}$. If we generate A uniformly among sets of size $k = 2^j$, then

$$\begin{aligned}
&Pr[A \cap P = \phi \ \&\ A \cap Q \neq \phi] = R_k / \binom{n}{k} \\
&= \left[\binom{n - |P|}{k} - \binom{n - |P| - |Q|}{k}\right] / \binom{n}{k} \\
&= \frac{(n - |P|)!(n-k)!}{n!(n - |P| - k)!} - \frac{(n - |P| - |Q|)!(n-k)!}{n!(n - |P| - |Q| - k)!} \\
&= (1 - \frac{|P|}{n})(1 - \frac{|P|}{n-1}) \ldots (1 - \frac{|P|}{n-k+1}) - (1 - \frac{|P| + |Q|}{n}) \ldots (1 - \frac{|P| + |Q|}{n-k+1}) \\
&\approx (e^{-\frac{|P|}{n}})^k - (e^{-\frac{|P|+|Q|}{n}})^k = e^{-\frac{|P|k}{n}}(1 - e^{-\frac{|Q|k}{n}}) \\
&\ge e^{-\frac{2^t 2^{p-t}}{n}}(1 - e^{-\frac{2^{t-1} 2^{p-t}}{n}}) = e^{-\frac{2^p}{n}}(1 - e^{-\frac{2^{p-1}}{n}}) \\
&\ge e^{-1}(1 - e^{-4}) = 1/\beta,
\end{aligned}$$

where the approximation can be made formal. □

Now, we are ready to prove the main theorem.

Proof. μ satisfies the condition of Lemma 2 for the following reason. If $A \subset R_k$, $|d(x, A) - d(y, A)| \ge \rho_t - \rho_{t-1}$ because $d(x, A) \ge \rho_t$, $d(y, A) \le \rho_{t-1}$. Thus we have:

$$\begin{aligned}
&E_\mu[|d(x, A) - d(y, A)|] \\
&= \sum_{A \subseteq X} \mu(A) |d(x, A) - d(y, A)| \ge \sum_{\substack{j=p-l \\ (k=2^j, t=p-j)}}^{p-1} \sum_{A \in R_k} |d(x, A) - d(y, A)| \mu(A) \\
&\ge \sum_{\substack{j=p-l \\ (k=2^j, t=p-j)}}^{p-1} \sum_{A \in R_k} (\rho_t - \rho_{t-1}) \frac{1}{p+1} \frac{1}{\binom{n}{k}} = \sum_{\substack{j=p-l \\ (k=2^j, t=p-j)}}^{p-1} |R_k| (\rho_t - \rho_{t-1}) \frac{1}{p+1} \frac{1}{\binom{n}{k}} \\
&\ge \sum_{\substack{j=p-l \\ (k=2^j, t=p-j)}}^{p-1} (\rho_t - \rho_{t-1}) \frac{1}{p+1} \frac{1}{\beta} \text{ (by Lemma 3)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{p+1}\frac{1}{\beta}\rho_l \geq \frac{1}{p+1}\frac{1}{\beta}\frac{d(x,y)}{4} \\
&\geq \frac{1}{O(\log n)} d(x,y).
\end{aligned}
$$

Therefore, this completes the proof of the main theorem. □

Remark 1. *(X, d) can be embedded into any l_p ($p \geq 1$) with distortion $O(\log n)$ because l_2 can be isometrically embedded into l_p. Also, lemma 2 can be slightly modified to give directly the l_p-embedding with distortion γ.*

Remark 2. *We have embedded (X, d) into $\mathbb{R}^{2^n}$. The Frechet embedding μ we constructed considered all subset A of size $k = 2^j$, and gave the same probability($\frac{1}{p+1}\frac{1}{\binom{n}{k}}$) to their sets. But, instead of looking all subsets A of size $k = 2^j$, if we choose $O(\log n)$ sets among them, the second lemma ($Pr[A \cap P = \phi \;\&\; A \cap Q \neq \phi] \geq constant$) holds with high probability for each $x, y \in X$ (using a Chernoff bound [3]). Therefore, we can do the embedding efficiently (in polynomial time) into $\mathbb{R}^{O(p \log n)} = \mathbb{R}^{O(\log^2 n)}$ with the same distortion $O(\log n)$. This was observed by London, Linial and Rabinovich (1985).*

Remark 3. *The embedding we constructed gives a $O(\log n)$-approximation algorithm for the general sparsest cut problem. However, if we focus on the source and sink vertices ($T = \{(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)\}$) and do the embedding G as we did but restricted to T (thus considering only sets $A \subseteq T$), we can get*

$$\forall x, y \in T,\; ||f(x) - f(y)||_2 \geq \frac{d(x,y)}{O(\log k)}.$$

Furthermore, the non-expanding property of our embedding says

$$\forall x, y \in X,\; ||f(x) - f(y)||_2 \leq d(x,y).$$

These two inequality are enough to analyze that this gives $O(\log k)$-approximation algorithm because the denominator of the formulas[2] of $\alpha(G)$ and $\beta(G)$ only depend on vertices in T.

References

[1] J. Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. Israel Journal of Mathematics, 52:46–52, 1985.

[2] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. In 37th STOC, 2005.

[3] H. Chernoff. A measure of asymptotic efficiency for test of hypothesis based on the sum of observations. Annals of Mathematical Statistics, vol. 23, pp. 493–507, 1952.

[2] $\alpha(G) = \min_{(V,d),\; l_\infty-embeddable} \frac{\sum_{e \in E} c(e) \cdot d(e)}{\sum_{i=1}^{k} D_i \cdot d(s_i, t_i)}$