Massachusetts Institute of Technology Michel X. Goemans

18.409: Topics in TCS: Embeddings of Finite Metric Spaces

November 15, 2006 Presentation and notes by David Sontag and Krzysztof Onak

---

Lecture 18

In today's lecture, we present some of the results contained in [1, 2].

1 Good news: A general theorem on embeddings with slack

We will show how to obtain embeddings into ℓ_p with small dimension where all but an ϵ-fraction of the embedded distances have small distortion. This result has various applications in networking. The embedding is achieved by randomly choosing a small set of beacons, constructing a good embedding for these nodes (using, e.g., Borgain's embedding), and then extending this embedding to the remaining nodes. Thus, we need the following definition:

Definition 1. *A family $\mathcal{X}$ of metrics is* subset-closed *if any metric in $\mathcal{X}$ restricted to any subset of nodes is also in $\mathcal{X}$.*

Definition 2. *Given ϵ, an embedding $\varphi : V \to V'$ has* distortion D with ϵ-slack *if a set of all but an ϵ-fraction of edges has distortion at most D under φ.*

Theorem 1. *Consider a fixed space ℓ_p, $p \geq 1$. Let $\mathcal{X}$ be a subset-closed family of finite metric spaces such that for any $n \geq 1$ and any n-point metric space $X \in \mathcal{X}$ there exists an embedding $\varphi_X : X \hookrightarrow \ell_p$ with distortion $\alpha(n)$ and dimension $\beta(n)$.*

Then there exists a universal constant $C > 0$ such that for any metric space $X \in \mathcal{X}$ and any $\epsilon > 0$ we have an embedding into ℓ_p with ϵ-slack, distortion $\alpha(\frac{C}{\epsilon}\log\frac{1}{\epsilon})$ and dimension $\beta(\frac{C}{\epsilon}\log\frac{1}{\epsilon}) + C\log\frac{1}{\epsilon}$.

A (modified) Bourgain's theorem gives $\alpha(n) = O(\log n)$ and $\beta(n) = O(\log^2 n)$. Using an improved result by Bartal gives the following corollary.

Corollary 2. *Any metric space $X \in \mathcal{X}$ can be embedded into ℓ_p with ϵ-slack, distortion $O(\log\frac{1}{\epsilon})$ and dimension $O(\log\frac{1}{\epsilon})$.*

Proof. Consider some metric $X = (V, d) \in \mathcal{X}$, where V is a set of n nodes. Given $\epsilon > 0$ let $\hat{\epsilon} = \epsilon/20$, and $t = 100\log\frac{1}{\hat{\epsilon}}$. The first step is to sample $\frac{t}{\hat{\epsilon}}$ beacons, B, uniformly at random from V.

Let g be a contracting embedding from B into ℓ_p with distortion $\alpha(\frac{t}{\hat{\epsilon}})$ and dimension $\beta(\frac{t}{\hat{\epsilon}})$. We will extend g to all of V by defining, $\forall u \in V$, $f(u) = g(b)$, where $b \in B$ is the beacon closest to node u. Also, let

$$\{\sigma_j(u) | u \in V, 1 \leq j \leq t\} \tag{1}$$

be independently sampled $\{0, 1\}$-valued Bernoulli random variables. For all $u \in V$ and $j = 1, \ldots, t$, define the function

$$h_j(u) = \sigma_j(u)\rho_u(\hat{\epsilon})t^{-1/p}. \tag{2}$$

Definition 3. *Let $\rho_u(\epsilon)$ be the radius of the smallest ball around u that contains at least ϵn nodes.*

The embedding $\varphi(u)$ will be the concatenation of $f(u)$ and $(h_1(u), \ldots, h_t(u))$, giving a dimension of $\beta(\frac{t}{\hat{\epsilon}}) + t$.

Let E be the set of all unordered node pairs. We will remove from consideration three sets D_1, D_2, and D_3 of "difficult" node pairs, whose distortion we will not bound, and will show that they are of size only $O(\epsilon)$.

$$D_1 = \{(u,v) | d(u,v) < \max\{\rho_u(\hat{\epsilon}), \rho_v(\hat{\epsilon})\}\} \tag{3}$$
$$D_2 = \{(u,v) | d(u,B) \geq \rho_u(\hat{\epsilon}) \vee d(v,B) \geq \rho_v(\hat{\epsilon})\}. \tag{4}$$

At an intuitive level, D_1 is the set of node pairs that are too close together, and D_2 is the set of node pairs that are too far from the beacons. The third set, D_3, will be given later. Note that $|D_1| \leq \hat{\epsilon} n^2$, since for any node $u \in V$ there are at most $\hat{\epsilon} n$ nodes in the ball of radius $\rho_u(\hat{\epsilon})$ around u. Also, for any node $v \in V$ we have:

$$\Pr[d(u,B) \geq \rho_u(\hat{\epsilon})] \leq \Pr[\forall v \in B_{\rho_u(\hat{\epsilon})}(u), v \notin B] \leq \left(1 - \left(\frac{t}{\hat{\epsilon}}\right)\frac{1}{n}\right)^{\hat{\epsilon} n} \leq e^{-t} \leq \hat{\epsilon}, \tag{5}$$

so $\mathbb{E}[|D_2|] \leq \hat{\epsilon} n^2$ and by Markov's inequality we have $|D_2| \leq 2\hat{\epsilon} n^2$ with probability at least 1/2.

1.1 Upper Bound on $d(\varphi(u), \varphi(v))$

Let $G' = E \backslash (D_1 \cup D_2)$. We will now upper bound $d(\varphi(u), \varphi(v))$ for all $(u,v) \in G'$. Since $G' \subseteq \overline{D_1}$, $d(u,v) > \rho_u(\hat{\epsilon})$ and $d(u,v) > \rho_v(\hat{\epsilon})$. Since $G' \subseteq \overline{D_2}$, $d(u,B) < \rho_u(\hat{\epsilon})$ and $d(v,B) < \rho_v(\hat{\epsilon})$. Together these imply that $d(u,B) < d(u,v)$ and $d(v,B) < d(u,v)$. We thus have:

$$\begin{aligned}
||\varphi(u) - \varphi(v)||_p^p &= ||f(u) - f(v)||_p^p + \sum_{j=1}^{t} |h_j(u) - h_j(v)|^p \\
&= ||g(b_u) - g(b_v)||_p^p + \sum_{j=1}^{t} |h_j(u) - h_j(v)|^p \\
&\leq (d(b_u, b_v))^p + \sum_{j=1}^{t} |h_j(u) - h_j(v)|^p \\
&\leq (d(b_u, u) + d(u,v) + d(v, b_v))^p + \sum_{j=1}^{t} |h_j(u) - h_j(v)|^p \\
&< (3d(u,v))^p + \sum_{j=1}^{t} |t^{-1/p} \max\{\rho_u(\hat{\epsilon}), \rho_v(\hat{\epsilon})\} - 0|^p \\
&< (3^p + 1)(d(u,v))^p.
\end{aligned} \tag{6}$$

Where we used the fact that g is a contractive embedding, and the notation b_u refers to the node in B which is closest to u.

All that remains is to give a lower bound on $d(\varphi(u), \varphi(v))$. We consider two separate cases:

$$G_1 = \{(u,v) \in G' : \max\{\rho_u(\hat{\epsilon}), \rho_v(\hat{\epsilon})\} \geq d(u,v)/4\} \tag{7}$$
$$G_2 = G' \backslash G_1. \tag{8}$$

1.2 Lower Bound on $d(\varphi(u), \varphi(v))$ for G_1

Since for all edges in G_1 either $\rho_u(\hat{\epsilon})$ or $\rho_v(\hat{\epsilon})$ is $\Omega(d(u,v))$, the h_j coordinates will provide the desired lower bound. The σ_j random variables help ensure that in some non-negligible fraction of the dimensions we get $\rho_u(\hat{\epsilon}) + \rho_v(\hat{\epsilon})$. More concretely, consider an edge $(u,v) \in G_1$, and without loss of generality assume $\rho_u(\hat{\epsilon}) \geq \rho_v(\hat{\epsilon})$, so that $\rho_u(\hat{\epsilon}) \geq d(u,v)/4$.

Let $\mathcal{E}_j(u,v)$ be the event that $\sigma_j(v) = 0$ and $\sigma_j(v) = 1$. This event happens with probability $\frac{1}{4}$. Let $A(u,v) = \sum_{j=1}^{t} \mathbf{1}_{\mathcal{E}_j(u,v)}$. Then $\mathbb{E}[A(u,v)] = t/4$, and by Chernoff's bound we have:

$$\Pr\left[A(u,v) \leq \frac{\mathbb{E}[A(u,v)]}{2}\right] \leq e^{-t/50} \leq \hat{\epsilon}. \tag{9}$$

We now define D_3, the third set to be removed:

$$D_3 \;\; = \;\; \{(u,v) \in G_1 | A(u,v) \leq t/8\} \tag{10}$$

Note that there is nothing inherently difficult about the edges in D_3 – we were just unlucky with the coin tosses. By Markov's inequality, $|D_3| \leq 2\hat{\epsilon}n^2$ with probability $\geq 1/2$.

We now get the desired lower bound for $(u,v) \in G_1 \backslash D_3$:

$$\begin{aligned} ||\varphi(u) - \varphi(v)||_p^p \;\; &\geq \;\; \sum_{j=1}^{t} |h_j(u) - h_j(v)|^p \\ &= \;\; \sum_{j=1}^{t} |\sigma_j(u)\rho_u(\hat{\epsilon})t^{-1/p} - \sigma_j(v)\rho_v(\hat{\epsilon})t^{-1/p}|^p \\ &\geq \;\; \frac{t}{8}\left(\rho_u(\hat{\epsilon})t^{-1/p}\right)^p \geq \frac{1}{8}\left(\frac{1}{4}d(u,v)\right)^p. \end{aligned} \tag{11}$$

1.3 Lower Bound on $d(\varphi(u), \varphi(v))$ for G_2

Recall that b_u refers to the beacon in B which is closest to u. Since $G_2 \subseteq \overline{D_2}$, $d(u,B) < \rho_u(\hat{\epsilon})$ and $d(v,B) < \rho_v(\hat{\epsilon})$. Since $G_2 \subseteq \overline{G_1}$, $d(u,v)/4 > \rho_u(\hat{\epsilon})$ and $d(u,v)/4 > \rho_v(\hat{\epsilon})$. Together these imply that $d(u,B) < d(u,v)/4$ and $d(v,B) < d(u,v)/4$. Applying the triangle inequality,

$$\begin{aligned} d(b_u, b_v) \;\; &\geq \;\; d(u,v) - d(u,b_u) - d(v,b_v) \\ &> \;\; d(u,v) - d(u,v)/4 - d(u,v)/4 = d(u,v)/2. \end{aligned} \tag{12}$$

We thus have the following lower bound for $(u,v) \in G_2$:

$$\begin{aligned} ||\varphi(u) - \varphi(v)||_p^p \;\; &\geq \;\; ||f(u) - f(v)||_p^p \\ &= \;\; ||g(b_u) - g(b_v)||_p^p \\ &\geq \;\; \frac{1}{\alpha(\frac{t}{\hat{\epsilon}})} \cdot d(b_u, b_v) \\ &\geq \;\; \frac{d(u,v)}{2\alpha(\frac{t}{\hat{\epsilon}})}. \end{aligned} \tag{13}$$

To finish the proof, note that D_2 and D_3 are independent (D_3 is a function of the σ_j random variables). Thus, in total from D_1, D_2, and D_3 we have removed at most $5\hat{\epsilon}n^2$ nodes with probability at least $\frac{1}{4}$. Let $G = E\backslash(D_1 \cup D_2 \cup D_3)$ be the set of edges whose distortion we bounded in the previous two sections. We conclude that, with probability $\geq \frac{1}{4}$,

$$|G| \geq \binom{n}{2} - 5\hat{\epsilon}n^2 \geq \binom{n}{2} - \frac{\epsilon n^2}{4} \geq (1-\epsilon)\binom{n}{2}. \tag{14}$$

□

2 Bad news: Lower bounds for embedding with slack

2.1 Why distortion $\Omega(\log \frac{1}{\epsilon})$ is necessary

Theorem 3. *There exists a finite metric (X, d) on arbitrary many nodes that requires distortion $\Omega(\frac{1}{p}\log\frac{1}{\epsilon})$ for embedding with ϵ-slack into ℓ_p, where $p \geq 1$.*

Proof. Suppose that for every finite metric, there exists an embedding into ℓ_p with ϵ-slack and distortion D. We will show that $D = \Omega(\frac{1}{p}\log\frac{1}{\epsilon})$.

Suppose without loss of generality that $0 < \epsilon \leq 1/4$. Let k be $1/(2\sqrt{\epsilon})$. Let $G = (V, E)$ be a constant-degree expander on k nodes, and (V, dist_G) be the corresponding shortest-path metric. For each node v in V, create a new path on $n/k = 2\sqrt{\epsilon}n$ nodes, and attach one of its ends to v. This way we get a graph on n nodes. Let δ such that $\delta \cdot D \leq 1/3$ be the length of each attached path. The new weighted graph $H = (V', E')$ induces the shortest-path metric (V', dist_H).

There exists an embedding ϕ of (V', dist_H) into ℓ_p with ϵ-slack and distortion D. We can assume without loss of generality that

$$\mathrm{dist}_H(v, w) \leq ||\phi(v) - \phi(w)||_p \leq D \cdot \mathrm{dist}_H(v, w) \tag{15}$$

for all but an ϵ-fraction of pairs v and w. Let I be the set of the pairs for which the above inequality does not hold. We have $|I| \leq \epsilon n^2/2$.

We will show that the expander metric (V, dist_G) embeds into ℓ_p with distortion $3D$ in the standard sense (i.e. with 0-slack). Remove from H nodes that belong to at least $\sqrt{\epsilon}n$ pairs in I. There are at most $\sqrt{\epsilon}n$ such nodes, which means that at least $\sqrt{\epsilon}n$ nodes survive in each path that was attached to the expander. For each node v of the expander choose a node $v^\star$ in the path attached to v that survived, and define an embedding ψ of (V, dist_G) into ℓ_p as

$$\psi(v) \stackrel{\mathrm{def}}{=} \phi(v^\star). \tag{16}$$

Let v and w be two different nodes in V. An easy counting argument shows that there exists a node $u \in V'$ that belongs to the path that was attached to v, and such that neither $(v^\star, u)$ nor

$(w^\star, u)$ belongs to I, the set of "bad" pairs. We have

$$\begin{aligned}
||\psi(v) - \psi(w)||_p &= ||\phi(v^\star) - \phi(w^\star)||_p \\
&\le ||\phi(v^\star) - \phi(u)||_p + ||\phi(u) - \phi(w^\star)||_p \\
&\le D \cdot \mathrm{dist}_H(v^\star, u) + D \cdot \mathrm{dist}_H(u, w^\star) \\
&\le D \cdot \mathrm{dist}_H(v, w) + 3D\delta \\
&\le D \cdot \mathrm{dist}_G(v, w) + 1 \\
&\le D \cdot \mathrm{dist}_G(v, w) + \mathrm{dist}_G(v, w) \\
&\le 2D \cdot \mathrm{dist}_G(v, w), \qquad (17)
\end{aligned}$$

and also

$$\begin{aligned}
||\psi(v) - \psi(w)||_p &= ||\phi(v^\star) - \phi(w^\star)||_p \\
&\ge ||\phi(u) - \phi(w^\star)||_p - ||\phi(u) - \phi(v^\star)||_p \\
&\ge \mathrm{dist}_H(u, w^\star) - D \cdot \mathrm{dist}_H(u, v^\star) \\
&\ge \mathrm{dist}_H(v, w) - D\delta \\
&\ge \mathrm{dist}_G(v, w) - 1/3 \\
&\ge \mathrm{dist}_G(v, w) - 1/3 \cdot \mathrm{dist}_G(v, w) \\
&\ge 2/3 \cdot \mathrm{dist}_G(v, w), \qquad (18)
\end{aligned}$$

which implies that ψ is an embedding of the expander metric into ℓ_p with distortion $3D$. It is known that to embed a bounded-degree expander metric on k nodes into ℓ_p we need distortion $\Omega(\frac{1}{p}\log k)$ ([3], see Lecture 3 for a proof of distortion $\Omega(\log k)$ for ℓ_2). This implies that $D = \Omega(\frac{1}{p}\log\frac{1}{\epsilon})$. □

2.2 Contracting embeddings with slack

Suppose we wanted to construct a *contracting embedding ϕ with ϵ-slack* of a finite metric into ℓ_p, that, is a contracting embedding such that ϕ contracts by at most D on all but an ϵ-fraction of the pairs. What D can we hope for? It turns out that for bounded-degree expanders this still implies distortion $\Omega(\log n)$, that is, we do not gain anything over the standard notion of an embedding.

Theorem 4. *A contracting embedding of a bounded-degree expander on n nodes into ℓ_p requires distortion $\Omega(\frac{1}{p}\log n)$ even with 1/2-slack.*

We will only prove it for $p = 2$, using our knowledge from Lecture 3.

Theorem 5. *Let ϵ be a fixed constant in $(0, 1)$. A contracting embedding of a constant-degree expander on n nodes into ℓ_2 with ϵ-slack requires distortion $\Omega(\log n)$.*

Proof. Let $G = (V, E)$ be an r-regular expander, where r is a constant, and let ϕ be a contracting embedding with ϵ-slack and distortion D.

We know that $\mu_2(G) = \Theta(1)$, and in Lecture 3 in the proof of Theorem 4 we showed that for any embedding ϕ of G into ℓ_2, we have

$$\sum_{(v,w)\in E} ||\phi(v) - \phi(w)||^2 \ge \frac{\mu_2(G)}{n} \sum_{(v,w)\in\binom{V}{2}} ||\phi(v) - \phi(w)||^2. \qquad (19)$$

Note that because the embedding is contracting, we have

$$\sum_{(v,w)\in E} ||\phi(v) - \phi(w)||^2 \leq \frac{r}{2}n, \tag{20}$$

and it follows from the last two equations that

$$\sum_{(v,w)\in\binom{V}{2}} ||\phi(v) - \phi(w)||^2 = O(n^2). \tag{21}$$

On the other hand, for all but an ϵ-fraction of pairs v and w, it holds

$$||\phi(v) - \phi(w)||^2 \geq \frac{\mathrm{dist}_G^2(v,w)}{D^2}. \tag{22}$$

For all but an $o(1)$-fraction of pairs v and w, we have $\mathrm{dist}_G(v,w) = \Theta(\log n)$, and therefore, even if the embedding completely contracts the distance on some ϵ-fraction of pairs, we still have

$$\sum_{(v,w)\in\binom{V}{2}} ||\phi(v) - \phi(w)||^2 = \Omega\left(\frac{n^2 \log^2 n}{D^2}\right), \tag{23}$$

which means that $D = \Omega(\log n)$. □

References

[1] Ittai Abraham, Yair Bartal, T.-H. Hubert Chan, Kedar Dhamdhere, Anupam Gupta, Jon M. Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *FOCS*, pages 83–100. IEEE Computer Society, 2005.

[2] T.-H. Hubert Chan, Kedar Dhamdhere, Anupam Gupta, Jon M. Kleinberg, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. Full version of the FOCS paper. Available at http://www.cs.brown.edu/people/slivkins/focs05_full.pdf.

[3] Jiří Matoušek. On embedding expanders into ℓ_p spaces. *Israel J. Math.*, 102:189–197, 1997.