4 Matoušek's Result: Embeddings into ℓ_∞

In the last class we have seen that any finite metric on n points embeds isometrically into ℓ_∞ with $n-1$ dimensions. In this class we show a theorem due to Matoušek that gives an embedding into a low-dimensional ℓ_∞ space with some distortion. We first recall the definition of a Fréchet-type embedding.

Fréchet-type (Subset) Embeddings: Given a metric d on points V, a *Fréchet-type* embedding is a map

$$f(x) = \bigoplus_{S \subseteq V} \beta_S f_S(X)$$

where $f_S(x) = d(x, S) = min_{v \in S} d(x, v)$. Such mappings are also sometimes called *subset embeddings.*

Lemma 4.1 *When $\beta_S = 1$, $\forall S \subseteq V$, $\|f(x) - f(y)\|_\infty \leq d(x, y)$ which means that the subset-embedding is a contraction when distances are measured in the ℓ_∞-norm.*

Proof. For all $S \subseteq V$, let S_x (resp S_y) denote the point in S closest to x (resp y). Then

$$d(x, S) - d(y, S) \leq d(x, S_y) - d(y, S_y) \leq d(x, y),$$

the last step following by the triangle inequality. Similarly,

$$d(y, S) - d(x, S) \leq d(y, S_x) - d(x, S_x) \leq d(x, y)$$

Thus $\|f(x) - f(y)\|_\infty = |d(x, S) - d(y, S)| \leq d(x, y)$. ∎

We now prove an important result by Matoušek which uses subset embeddings in a nice way.

Theorem 4.2 (Matoušek (1996)) *Given an arbitrary metric space (X, d) on n points and an integral value D. There is an embedding $\phi : (X, d) \to \ell_\infty^{O(Dn^{2/D} \log n)}$ with distortion D.*

Corollary 4.3 *For the special case of $D = \Theta(\log n)$, this implies that $(X, d) \to \ell_\infty^{O(\log^2 n)}$.*

The embedding used to prove Theorem 4.2 will be a subset embedding. Before describing the construction in detail and giving a formal proof, we first give a high level idea of the approach.

High Level Idea: We will choose about $O(Dn^{2/D} \log(n))$ subsets according to a procedure specified later. Each subset will correspond to a dimension in the range metric space. For a point $x \in X$, the jth coordinate of $\phi(x)$ will be $d(x, S)$. Clearly, ϕ is a contraction. The crucial step is to show that ϕ does not contract too much. The idea to show this is as following.

Consider two points x and y at a distance $d(x, y)$. Suppose we have two disjoint balls, the first $B_r(x)$, which is centered at x and has radius r and second $B_{r+\Delta}(y)$ which is centered

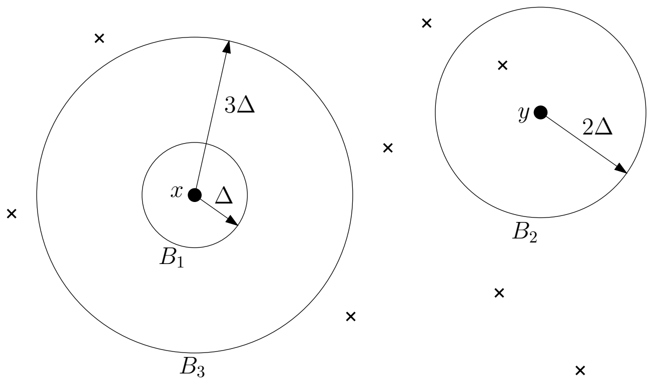

Figure 4.1: Matoušek's analysis. Crosses indicate nodes from a random set S_{ij}.

at y and has radius $r + \Delta$. Suppose we had a subset $S \subset V$ such that $B_r(x) \cap S \neq \emptyset$ and $B_{r+\Delta}(j) \cap S = \emptyset$. Then, this implies that the points $\phi(x)$ and $\phi(y)$ differ by at least Δ in the coordinate corresponding to the subset S.

Our goal will be to show that one can find a family of $O(Dn^{2/D} \log(n))$ subsets $S \subseteq V$ such that for every pair of vertices x, y there is at least one subset (depending on x, y) in the family that satisfies the property mentioned above with Δ about $d(x, y)/D$.

The idea of the algorithm is the following: Let $\Delta = d(x, y)/D$. For $i \in \{0, \ldots, \lceil D/2 \rceil\}$ we define a ball B_i with radius $i\Delta$. If i is odd then B_i is centered around x and if i is even B_i is centered around y. See Figure 4.

We first claim that there exists a $t \in \{0, \ldots, \lceil D/2 \rceil - 1\}$ such that $|B_t| \geq n^{2t/D}$ and $|B_{t+1}| \leq n^{2t+1/D}$ (see proof of Lemma 4.4).

If we choose sets S by randomly sampling nodes at rates $p^j = n^{-2j/D}$ for $j = 1, \ldots, \lceil D/2 \rceil$, then for $j = t + 1$, B_{t+1} will be empty with constant probability. Moreover, the probability that there is a point in B_t will be about $n^{-2/D}$. So, if we repeat this experiment say about $m = O(n^{2/D} \log n)$ times, this will guarantee that for each vertex pair (x, y) we get a separation of Δ in some coordinate.

We now describe and analyze the algorithm formally.

Algorithm:

$p := n^{-2/D}$
$m := O(n^{2/D} \log n)$
for $j = 1$ **to** $\lceil D/2 \rceil$ **do** /* levels of density */
  for $i = 1$ **to** m **do** /* repeat for high probability */
    choose set S_{ij} by sampling each node
    independently with probability $p^j = n^{-2j/D}$
  end
end
$\phi(x) := \bigoplus_{j=1}^{\lceil D/2 \rceil} \bigoplus_{i=1}^{m} d(x, S_{ij})$

Lemma 4.4 *Fix $x, y \in X$. There exists an index $j \in \{1, \ldots, \lceil D \rceil /2\rceil\}$ such that if S_{ij} is chosen randomly as above,*

$$P(|d(x, S_{ij}) - d(y, S_{ij})| \geq \frac{d(x,y)}{D}) \geq \frac{p}{12} = \frac{1}{12n^{2/D}} \quad .$$

We first see how the lemma implies Theorem 4.2.

Proof. The probability that none of the m trials for the appropriate index j generate $d(u,v)/D \leq (1-p/12)^m \leq e^{-pm/12} \leq 1/n^2$ as $m = O(n^{2/D} \ln n)$ and $p = n^{-2/D}$.

Thus by the union bound, the probability that any pair (x, y) is not well separated is at most $\binom{n}{2} \cdot n^{-2} \leq 1/2$. ∎

We now prove Lemma 4.4.

Proof of Lemma 4.4. There exists $t \in \{0, \ldots, \lceil D/2 \rceil - 1\}$ such that $|B_t| \geq n^{2t/D}$ and $|B_{t+1}| \leq n^{2(t+1)/D}$. To see this assume for contradiction that the statement does not hold for any $t \in \{0, \ldots, \lceil D/2 \rceil - 1\}$. Then, in particular it does not hold for $t = 0$. However, $|B_0| = 1 \geq n^{2 \cdot 0/D}$. Therefore, the second requirement (the one on $|B_{t+1}|$) must be false. This means $|B_1| > 1 \geq n^{2 \cdot 1/D}$. Now, checking the statement for $t = 1$ we see again that the first requirement ($|B_1| \geq n^{2 \cdot 1/D}$) is true, therefore the second has to be false. Continuing in this manner gives that the second requirement for $t = \lceil D/2 \rceil - 1$ has to be false. But this means $|B_{\lceil D/2 \rceil}| > n^{2\lceil D/2 \rceil/D} \geq n$ which is a contradiction.

In the following let t be chosen such that $|B_t| \geq n^{2t/D}$ and $|B_{t+1}| \leq n^{2(t+1)/D}$. Now for density $j = t+1$ we consider the event that S_{ij} hits B_t and avoids B_{t+1}. The balls B_t and B_{t+1} are disjoint (check!). Therefore the events that S_{ij} hits B_t and the event that S_{ij} avoids B_{t+1} are independent.

The probability of S_{ij} hitting B_t is $1 - P(S_{ij} \cap B_t = \emptyset)$ which is simply

$$1 - (1-p^j)^{n^{2(j-1)/D}} \geq 1 - e^{-p^j n^{2(j-1)/D}} = 1 - e^{-p}$$

Assuming $p \in [0, 1/2]$, we have that $1 - e^{-p} \geq p/3$.

Remark: Note that we can assume (w.l.o.g.) that $p \in [0, 1/2]$, because if $p > 1/2$, then $D = \Omega(\log n)$. In Theorem 4.2 if $D = \Omega(\log n)$, then it is strictly better to reset $D = \Theta(\log n)$ as we get a better bound both in terms of distortion and in terms of the dimensions required.

Similarly, the probability of avoiding B_{t+1} is

$$(1-p^j)^{n^{2j/D}} = (1-p^j)^{1/p^j} \geq 1/4$$

for $p^j \leq 1/4$. By independence of the events it follows that S_{ij} hits B_t and avoids B_{t+1} with probability at least $p/12$. ∎

Corollary 4.5 *There exists an embedding $(X, d) \to \ell_1^{O(\log^2 n)}$ with distortion $O(\log^2 n)$.*

Proof (sketch). Using the theorem for $D = \Theta(\log n)$ we get an ℓ_∞ embedding into $O(\log^2 n)$ dimensions with distortion D. Let K denote the number of dimensions. Calculating the ℓ_1-distortion of this embedding gives the following.

Expansion: $\|\phi(x) - \phi(y)\|_1 \leq K \cdot d(x, y)$ since every coordinate is a subset-embedding and is therefore contracting.

Contraction: For the ℓ_∞-analysis we analyzed the event that at least one of the m trials for a distance scale are successful when trying to create distance between x and y at 'their' distance scale. We obtained that *with high probability* (probability $> 1 - 1/n^2$) this event holds.

By using Chernoff bounds one can show that with high probability at least $\log n$ trials are successful (one might have to choose m slightly larger but still $O(\log n)$). This means that with high probability our embedding fulfills

$$\|\phi(x) - \phi(y)\|_1 \geq \log n \cdot \Delta = \log n \cdot d(x, y)/\Theta(\log n) = \Theta(d(x, y)) \ .$$

Altogether this gives a distortion of $O(\log^2 n)$. ∎

One can improve the above theorem slightly when embedding into ℓ_2 instead.

Corollary 4.6 *There exists an embedding* $(X, d) \to \ell_2^{O(\log^2 n)}$ *with distortion* $O(\log^{1.5} n)$.

Proof. The proof is similar to the analysis above.

Expansion: $\|\phi(x) - \phi(y)\|_2 = \sqrt{\sum_s |\phi_s(x) - \phi_s(y)|^2} \leq \sqrt{K} \cdot d(x, y)$ since again every coordinate is contracting.

Contraction: With high probability our embedding fulfills

$$\|\phi(x) - \phi(y)\|_2 = \sqrt{\sum_s |\phi_s(x) - \phi_s(y)|^2} \geq \sqrt{\log n \cdot \big(\frac{d(x, y)}{\Theta(\log n)}\big)^2} \geq \frac{1}{\Theta(\sqrt{\log n})} d(x, y) \ .$$

This gives a distortion of $O(\log^{1.5} n)$. ∎

We will see in a later class that any ℓ_2-metric embeds isometrically into ℓ_1. By using the observation that if $X \overset{a}{\hookrightarrow} Y$ and $Y \overset{b}{\hookrightarrow} Z$ then $X \overset{a \cdot b}{\hookrightarrow} Z$ we get an embedding into ℓ_1 with distortion $O(\log^{1.5} n)$.

5 Application to Sparsest Cut

On reason for our interest in ℓ_1-embeddings is that they provide an elegant way to approximate the so-called *sparsest cut problem*, which is defined as follows. An instance to the *sparsest cut* *problem* is a graph $G = (V, V \times V)$ with *edge-capacities* given by $\mathrm{cap} : V \times V \to \mathbb{R}^+$ and pairwise *demands* by $\mathrm{dem} : V \times V \to \mathbb{R}^+$. The *sparsity* of a cut $S \subset V$ is given by

$$\alpha_S = \frac{\mathrm{cap}(S, V \setminus S)}{\mathrm{dem}(S, V \setminus S)} \ ,$$

where we extend the functions cap and dem by defining for two disjoint subsets $A, B \subset V$, $\mathrm{cap}(A, B) := \sum_{(a,b)\in A\times B} \mathrm{cap}(a, b)$ and $\mathrm{dem}(A, B) := \sum_{(a,b)\in A\times B} \mathrm{dem}(a, b)$.

The *sparsity of* G is defined as

$$\alpha = \min_{S\subseteq V} \alpha_S = \min_{\text{elementary cut metrics } d} \frac{\sum_{x,y} \mathrm{cap}(x, y) \cdot d(x, y)}{\sum_{x,y} \mathrm{dem}(x, y) \cdot d(x, y)} \tag{5.1}$$

Computing α using the right-hand-side identity above amounts to solving an integer program (IP). Let $S^* \subseteq V$ achieve the optimal flux, which we denote by $IP^* = \alpha_{S^*} = \alpha$. We can also relax the IP to obtain the following cut (primal) LP, by requiring that d merely be a metric, and not necessarily an elementary cut metric:

$$\begin{aligned}
\text{minimize} \quad & \sum_{x,y} \mathrm{cap}(x, y) \cdot d(x, y) \\
\text{subject to} \quad & \sum_{x,y} \mathrm{dem}(x, y) \cdot d(x, y) = 1 \\
& d(x, y) \le \sum_{(u,v)\in P} d(u, v) \qquad \forall \text{ path } P \text{ between } x \text{ and } y \\
& d(x, y) \ge 0
\end{aligned} \tag{LP 1}$$

Let d^* be the metric that achieves the optimal value of this LP, and let LP^* denote the optimal value. The dual (or flow) LP is given by:

$$\begin{aligned}
\text{maximize} \quad & \lambda \\
\text{subject to} \quad & \sum_{\substack{\text{path } P \text{ betw.} \\ x \text{ and } y}} f_{xy}(P) \ge \lambda \cdot \mathrm{dem}(x, y) \qquad \forall x, y \\
& \sum_{x,y} \sum_{P:(u,v)\in P} f_{xy}(P) \le \mathrm{cap}(u, v) \qquad \forall \text{ edges } (u, v) \\
& \lambda \ge 0 \\
& f_{xy}(P) \ge 0 \qquad \forall \text{ path } P \text{ between } x \text{ and } y, \forall x, y
\end{aligned} \tag{LP 2}$$

Let the optimal value of the dual program be λ^*. Recall the following basic result from linear programming duality:

Theorem 5.7 (Strong duality) *The optimal values of the primal and dual linear programs are equal. I.e., $LP^* = \lambda^*$.*

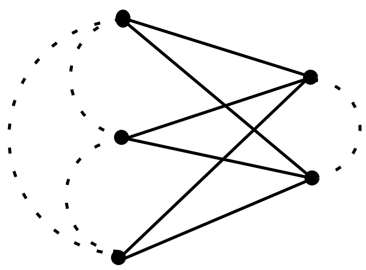

Figure 5.2: An instance in which $IP^* > LP^*$.

On the other hand, $LP^* \leq IP^*$, since LP is a relaxation of the integer program IP. The inequality may be strict for some instances, such as the bipartite graph in Figure 5.2; here all the edge capacities (indicated by solid lines) and the symmetric demands (indicated by the dotted lines) are set to one. By separating any one of the 3 vertices from the larger set in the bipartition from the rest of the graph, we achieve $LP^* = \alpha = 1$. However, a simple volume argument gives us $\lambda^* \leq \frac{3}{4}$: each demand has to travel two edges to reach its destination, and hence every unit of demand sent must occupy two units of volume. Since each edge has unit capacity, and there are six edges, there are only six units of volume available. Hence $8\,\lambda \leq 6$, and thus $\lambda \leq \frac{3}{4}$. Hence, on this instance, $IP^* \geq \frac{4}{3} LP^*$.

Definition 5.8 *Given an integer program, and a linear programming relaxation for it, the* integrality gap *on an instance is the ratio IP^*/LP^* for that instance. The supremum of the ratio IP^*/LP^* over* all *instances is called the* integrality gap of the relaxation.

5.1 The Integrality Gap and Distortion into ℓ_1

We can use a low-distortion embedding of a finite metric into ℓ_1.

Theorem 5.9 *Given an instance of the sparsest cut problem, let d^* be the metric given by LP 1. If $(V, d^*) \overset{D}{\hookrightarrow} \ell_1$, then*

$$IP^* \leq D \cdot LP^* \ ,$$

or in other words there exist a cut with sparsity at most $D \cdot LP^*$.

Proof.

We know that

$$LP^* = \sum_{x,y} \mathrm{cap}(x, y) \cdot d^*(x, y) = \frac{\sum_{x,y} \mathrm{cap}(x, y) \cdot d^*(x, y)}{\sum_{x,y} \mathrm{dem}(x, y) \cdot d^*(x, y)} \ .$$

By our assumption, there is a metric μ in ℓ_1 such that

$$\mu \leq d^* \leq D \cdot \mu \ .$$

Hence,

$$LP^* \geq \frac{\sum_{x,y} \operatorname{cap}(x,y) \cdot \mu(x,y)}{D \cdot \sum_{x,y} \operatorname{dem}(x,y) \cdot \mu(x,y)} \ .$$

But since μ is an ℓ_1 metric, it is also a cut metric; i.e., $\mu(x,y) = \sum_{S \subseteq V : (x,y) \in S \times V \setminus S} y_S$. So we have:

$$LP^* \geq \frac{\sum_S y_S \operatorname{cap}(S, V \setminus S)}{D \cdot \sum_S y_S \operatorname{dem}(S, V \setminus S)} \geq \min_{S : y_S \neq 0} \frac{\operatorname{cap}(S, V \setminus S)}{D \cdot \operatorname{dem}(S, V \setminus S)} \geq \frac{IP^*}{D} \tag{5.4}$$

Therefore, $LP^* \leq IP^* \leq D \cdot LP^*$. ■

This means if we can construct an embedding with distortion D into ℓ_1 using only a polynomial number of non-zero y_S's we directly get an approximation algorithm for the sparsest cut problem with approximation guarantee D.

References

[AR98] Yonatan Aumann and Yuval Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM Journal on Computing*, 27(1):291–301, 1998.

[BK97] Avrim Blum and David Karger. An $O(n^{3/4})$-coloring algorithm for 3-colorable graphs. *Information Processing Letters*, 61(1):49–53, 1997.

[BL84] Sandeep N. Bhatt and Frank Thomson Leighton. A framework for solving VLSI graph layout problems. *Journal of Computer and System Sciences*, 28(2):300–343, 1984.

[Bou85] Jean Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1-2):46–52, 1985.

[Che95] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23(4):493–507, 1995.

[GLS88] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization.* Springer, New York, 1988.

[HIL98] Johan Håstad, Lars Ivansson, and Jens Lagergren. Fitting points on the real line and its application to RH mapping. In *Proceedings of the 6th European Symposium on Algorithms (ESA)*, pages 465–476, 1998.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. Also in *Proc. 35th FOCS*, 1994, pp. 577–591.

[Mat96] Jiří Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics*, 93:333–344, 1996.

[Mat97] Jiří Matoušek. On embedding expander graphs into ℓ_p spaces. *Israel Journal of Mathematics*, 102:189–197, 1997.