15-859(J): Metric Embeddings and Methods **Lecturer:** R. Ravi
Topic: Metrics in algorithms, Embeddings into ℓ_∞ **Date:** September 3, 2003
Scribe: Nikhil Bansal

We continue our discussion on metric relaxations from the last lecture. In the last lecture we saw the application of metric relaxations to the min s-t cut problem and the minimum multiway cut problem. We will now consider the application to the problem of finding minimum balanced edge separators in a graph, via the *sparsest cut* problem.

3.1 Metric Relaxations

3.1.1 Sparsest Cut Problem

The sparsity of a capacitated undirected graph is defined as

$$\alpha = \min_{S \subseteq V} \frac{\mathrm{cap}(\partial S)}{\min(|S|, |V \setminus S|)}$$

where ∂S denotes the set of edges with exactly one end point in S. The sparsest cut problem is to find the cut S which achieves the minimum α.

The sparsity is closely related to the flux of a graph which is defined as

$$flux = \min_{S \subseteq V} \frac{\mathrm{cap}(\partial S)}{\min(|S| \cdot |V \setminus S|)}$$

It is easy to see that the value of the sparsest cut is within $n/2$ and n times that of the flux. Thus obtaining a good approximation to sparsest cut is equivalent (up to a constant factor) to obtaining a good approximation to the finding the cut with minimum flux.

Finding the sparsest cut has use in finding balanced (edge) separators. In particular,

Claim 3.1.1 *A ρ approximation for sparsity implies an $O(\rho)$ approximation for balanced separators.*

More precisely, using a greedy type algorithm one can show that if we have a ρ approximation for sparsity then we can obtain a 3/4-balanced cut whose value is within $O(\rho)$ times that of the best 2/3-balanced separator[1]. Observe that the guarantee on the approximation is slightly different from the usual notion of approximation, but this is sufficient for most applications; see Bhatt and Leighton (1984) for more details.

[1] A $b-$balanced separator is a minimum capacity set of edges whose deletion splits the graph into components none of which has more than $b|V|$ nodes.

3.1.2 Generalized Flux Problem

Given a graph $G = (V, E)$ with capacities on the edges cap : $E \to R_+$, and a demand function on pairs of vertices dem : $V \times V \to R_+$, the generalized flux is defined as

$$\alpha = \min_{S \subseteq V} \frac{\mathrm{cap}(\partial S)}{\mathrm{dem}(\partial S)}$$

It is easy to see that if the pairwise demand between each vertex pair is 1, then the quantity above is equal to the flux as defined in the previous section. Thus, this definition generalizes the flux problem.

We will see how to use metric relaxations to solve the generalized flux problem. It is easy to see that generalized flux can be rewritten as

$$\min_{\text{Elementary cut metrics } d} \frac{\sum_e \mathrm{cap}_e \cdot d_e}{\sum_{i,j \in V \times V} \mathrm{dem}_{ij} \cdot d_{ij}}$$

Since we cannot solve this problem in polynomial time, we relax the condition that we require d to be an elementary cut metric to the condition that it is just a metric.

This condition can simply be enforced by the linear constraints $d_{ij} \le d_{ik} + d_{kj}, \forall i, j, k \in V$ and $d_{ij} \ge 0, \forall i, j \in V$.

However, the objective that we want to minimize is a rational function. To convert it into a linear program observe that both the numerator and the denominator are linear in d_{ij} and the solution remains unchanged on scaling the $d_{ij}'s$. Therefore we add the following constraint

$$\sum_{i,j \in V \times V} \mathrm{dem}_{ij} \cdot d_{ij} = 1$$

We now have a linear program which we can solve in polynomial time. Let z^* denote the value of the optimum solution of this LP. The question now is how to round this linear program to obtain a cut without losing too much. We will see how this can be done in an elegant way using metric embeddings. In fact, this was an early introduction of metric embedding methods in computer science; the proof is independently due to Linial, London and Rabinovich (1995), and Aumann and Rabani (1995).

Recall Theorem (3.2) from Lecture 1, which tells us that any metric on n points can be embedded into an ℓ_1 metric on $O(\log^2 n)$ coordinates with distortion $O(\log(n))$. Formally, Given a set of n points V and distance d, there exists a map ϕ such that $\forall i, j \in V$,

$$\frac{d_{ij}}{4H(n)} \le |\phi(i) - \phi(j)|_1 \le d_{ij} \tag{3.1.1}$$

where $H(n)$ is the nth Harmonic number. We will see that this theorem gives what we want as a simple corollary.

Corollary 3.1.2 *$\exists$ a cut $S \subseteq V$, $\alpha_S = 4H(n)z^*$. Moreover, this cut is computable in polynomial time.*

We will use the following two facts:

1. For any $a_1, \ldots, a_n$ and $b_1, \ldots, b_n > 0$,

$$\min_i \frac{a_i}{b_i} \leq \frac{\sum_i a_i}{\sum_i b_i} \leq \max_i \frac{a_i}{b_i} \tag{3.1.2}$$

2. Any ℓ_1 metric on n points and k coordinates can be embedded isometrically into a cut metric on the same points. Moreover the cut metric can be expressed as a linear combination of at most nk elementary cuts.

   Formally, there exists y_S for $S \subseteq V$ such that for every i, j,

$$|\phi(i) - \phi(j)|_1 = \sum_{S \subseteq V \text{ and } i,j \in \partial S} y_S$$

   This can be seen by considering all the cuts which are obtained by projecting the points on each of the coordinates and considering the n possible level cuts in each coordinate.

Proof of Corollary 3.1.2: Let α^* denote the value of the optimum cut, and let α_ϕ be the cut of minimum generalized flux among the $O(\log^3 n)$ elementary cuts resulting from applying Fact 2 above.

$$\begin{aligned}
\alpha^* \leq \alpha_\phi \leq \quad & \leq \min_{S \subseteq V} \frac{\sum_S \mathrm{cap}(\partial S) y_S}{\sum_S \mathrm{dem}(\partial S) y_S} \\
& = \frac{\sum_{i,j} \mathrm{cap}(i,j) \sum_{(i,j) \in \partial S} y_S}{\sum_{i,j} \mathrm{dem}(i,j) \sum_{(i,j) \in \partial S} y_S} \\
& = \frac{\sum_{i,j} \mathrm{cap}(i,j) \cdot d_{ij}}{\sum_{i,j} \mathrm{dem}(i,j) \cdot \frac{d_{ij}}{4H(n)}} \\
& \leq 4H(n) z^*
\end{aligned}$$

Note that the first inequality follows from 3.1.2 and the second follows from the property of the metric embedding 3.1.1. ∎

To summarize the overall approach: We wanted a cut, we relax the condition of wanting a cut metric to that of a general metric, then use the embedding result to get back a cut metric from the general metric. Interestingly enough, though fairly general, this is still the best known approach for many problems.

3.2 Problems on Metrics

These are problems in which the metrics spaces are themselves an object of study. We look at one such problem below.

Problem 3.2.1 *Given a symmetric matrix of positive real numbers $D(i, j)$ (that need not satisfy the triangle inequality), find the optimum embedding A^* of the points (rows/columns of D) on a line, such that $||D - A||_\infty = \min_A \max_{ij} |D(i, j) - A(i, j)|$ is minimized.*

This problem has application to RH (Radiation Hybrid) mapping (see, e.g., the paper of Håstad, Ivansson and Lagergren (1998) for details). We will describe a 3-approximation for this problem.

We define an *x-fixed mapping* A^x as the map that satisfies the following property: $A^x(x) = 0$ and $A^x(v) = D(x, v)$.

Theorem 3.2.2 *The best fixed x-fixed mapping over all x is a 3-approximation.*

Proof: Let ϵ^* denote the perturbation that the best embedding A^* achieves. Let y denote the point which is embedded in the leftmost position by A^*. By a linear translation, we will assume that $A^*(y) = 0$.

Consider the mapping A^y. For all i, j, we are interested in the quantity $||A^y(i) - A^y(j)| - D(i, j)|$. We know that $|A^*(i) - D(y, i)| \leq \epsilon^*$ and $|A^*(j) - D(y, j)| \leq \epsilon^*$. Thus,

$$||A^y(i) - A^y(j)| - D(i, j)| \leq ||A^*(i) - A^*(j)| - D(i, j)| + 2\epsilon^* \leq 3\epsilon^*$$

The last step follows because A^* is an embedding into a line and has distortion ϵ^* in every coordinate including (i, j). ■

The paper of Håstad et al. (1998) also gives a 2-approximation algorithm for the problem by formulating and solving an interesting version of 2-SAT. The paper also shows that this problem is hard to approximate within $2 - \epsilon$ for any $\epsilon > 0$ unless 3-colorable graphs can be colored using $\lceil 4/\epsilon \rceil$ colors, which seems unlikely given that the current best algorithms (due to Blum and Karger (1997)) require $n^{\Theta(1)}$ colors for 3-colorable graphs.

3.3 Matoušek's Result: Embeddings into ℓ_∞

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

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

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

Lemma 3.3.1 *When $\beta_S = 1$, $\forall S \subseteq V$, $|f(x) - f(y)|_\infty \leq d(x, y)$.*

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 3.3.2 (Matoušek (1996)) *Given an arbitrary metric space (X, d) on n points, there is an embedding $\phi : (X, d) \to \ell_\infty^{O(Dn^{2/D} \log n)}$ which has distortion D.*

Corollary 3.3.3 *For the special case of $D = \Theta(\log n)$, this implies that $(X, d) \to \ell_\infty^{O(\log^2 n)}$.*

The above theorem though stated for the ℓ_∞ norm also gives a result about the ℓ_1 norms with another logarithmic loss in the distortion. (See Homework 1.)

Corollary 3.3.4 *There exists an embedding $(X, d) \to \ell_1^{O(\log^2 n)}$ with distortion $O(\log^2 n)$.*

The embedding used to prove Theorem 3.3.2 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 i and j at a distance $d(i, j)$. Suppose we have two disjoint balls, the first $B_r(i)$, which is centered at i and has radius r and second $B_{r+\Delta}(j)$ which is centered at j and has radius $r + \Delta$. Suppose we had a subset $S \subset V$ such that $B_r(i) \cap S \neq \emptyset$ and $B_{r+\Delta}(j) \cap S = \emptyset$. Then, this implies that the points $\phi(i)$ and $\phi(j)$ 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 i, j, there is at least one subset (depending on i, j) in the family that satisfies the property mentioned above with Δ about $d(i, j)/D$.

The idea of the algorithm is the following: Let $\Delta = d(i, j)/D$, let B_i denote a ball of radius $i\Delta$ and define it for $i = 1, 2, \ldots, D/2$. Consider balls $B_1, B_3, \ldots$ centered around i and $B_2, B_4, \ldots$ around j. See Figure 3.3. It is easily seen that there exist t and j such that $|B_t| \geq n^{2(j-1)/D}$ and $B_{t+1} \leq n^{2j/D}$. If we sample at the rate $p^i = n^{-2i/D}$ for $i = 1, \ldots, D/2$, then for the $i = j$, 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 $m = O(n^{2/D} \log n)$ this will guarantee that for each vertex pair (i, j) we get a separation of Δ in some coordinate.

We now describe and analyze the algorithm formally.

3.3.1 Algorithm

1. Pick random subsets S_{ij} in $D/2$ levels of density $j = 1, 2, \ldots, D/2$.

2. For each level, include each node independently with probability $n^{\frac{-2j}{D}} = p^j$ (by definition).

3. Choose $i = 1, \ldots, m$ subsets per level where $m = O(n^{2/D} \log n)$. Return the embedding

$$\phi(x) = \oplus_{j=1}^{D/2} \oplus_{i=1}^{m} d(x, S_{ij})$$

Lemma 3.3.5 *Fix $u, v \in X$. There exists an index $j \in \{1, \ldots, D/2\}$ such that if S_{ij} is chosen*

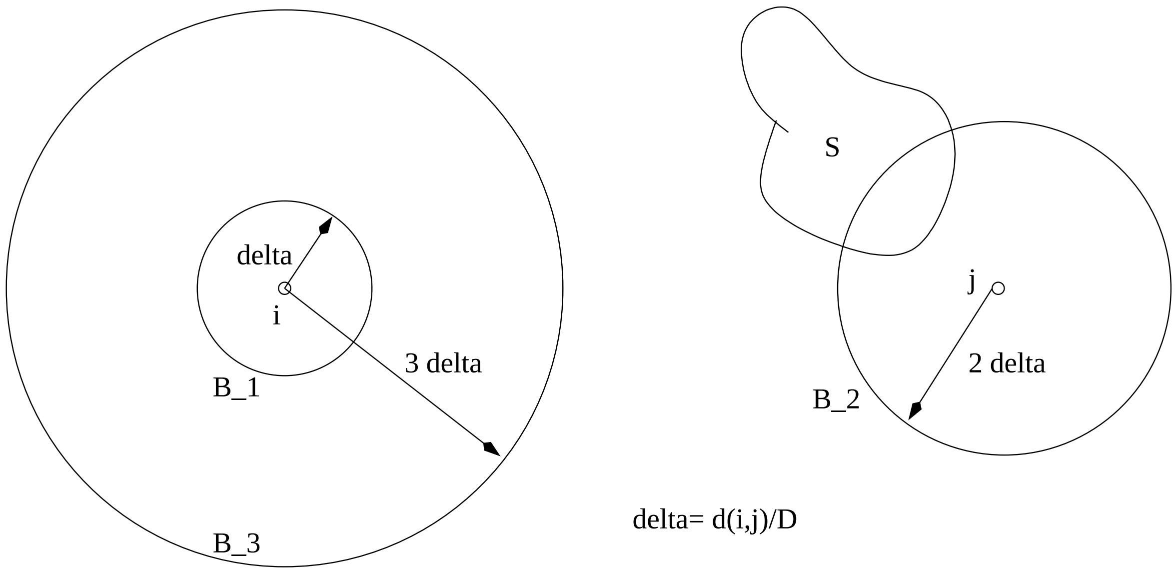

Figure 3.3.1: Figure describing Matoušek's Analysis

randomly as above,

$$P(|d(u, S_{ij}) - d(v, S_{ij})| \geq \frac{d(u,v)}{D}) \geq \frac{p}{12} = \frac{1}{12n^{2/D}} \tag{3.3.3}$$

We first see how the lemma implies Theorem 3.3.2.

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 some pair (u, v) is not well separated is at most $\binom{n}{2} \cdot n^{-2} \leq 1/2$. ■

We now prove Lemma 3.3.5.

Proof: [Lemma 3.3.5] By a simple expansion argument $\exists t \in \{1, \dots, D/2\}$ such that $B_t \geq n^{2(j-1)/D}$ and $B_{t+1} \leq n^{2j/D}$. Note that j is implicitly defined by the t. We consider the event that S_{ij} hits B_t and avoids B_{t+1}.

First observe that since B_t and B_{t+1} are disjoint balls, the events that S_{ij} hits B_t and the event that S_{ij} avoids B_{t+1} are disjoint.

The probability of S_{in} hitting B_t is $1 - P(S_{in} \cap B_t = \emptyset)$ 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 3.3.2 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 $p/12$. ■

References

Aumann, Y. and Rabani, Y. (1998). An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301.

Bhatt, S. N. and Leighton, F. T. (1984). A framework for solving VLSI graph layout problems. *Journal of Computer and System Sciences*, 28(2):300–343.

Blum, A. and Karger, D. (1997). An $O(n^{3/14})$-coloring algorithm for 3-colorable graphs. *Inform. Process. Lett.*, 61(1):49–53.

Håstad, J., Ivansson, L., and Lagergren, J. (1998). Fitting points on the real line and its application to RH mapping. In *Algorithms—ESA '98 (Venice)*, volume 1461 of *Lecture Notes in Comput. Sci.*, pages 465–476. Springer, Berlin.

Linial, N., London, E., and Rabinovich, Y. (1995). The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245. (Preliminary version in *35th FOCS*, 1994).

Matoušek, J. (1996). On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344.