CS-774 Spectral Graph Theory Sep 11, 2023

Lecture 4: An Unexpected Journey: A foray into discrete geometry and back

Lecturer: Akash Kumar *Scribe: Akash Kumar*

1 Sidelights from Geometry

In this lecture, we will discuss how ideas from geometry were used by algorithm designers towards getting approximation algorithms for finding low conductance cuts/sparsest cuts in graphs. This story has two big parts. The first part is a retelling of these ideas from discrete geometry. The second part reveals how algorithm designers combined these geometric insights with ideas from continuous optimization to obtain approximation algorithm for sparsest cut. Both of these parts are intricate (and beautiful). Thus, overall, today's lecture will have several moving parts. You are encouraged to read these notes in breaks. Alright, with that out of the way, let's start with the first part which tells the story of a celebrated result in geometry: the so-called Bourgain's embedding theorem ([Bourgain]). Our lecture borrows *really, really heavily* from these excellent notes/references of [Saranurak, Trevisan, VW].

2 Discrete Geometry: Motivation and Preliminaries

In algorithm design, you frequently come across situations where imposing some geometrical structure on your problem is extremely helpful towards an improved understanding of the problem which sometimes also helps you come up with solutions to the said problems. Let us motivate this via some concrete examples. Later, we will do some formal setup to develop a mathematical language capable of dealing with the complexity of the examples we wish to understand.

2.1 Motivation

Imagine you are given a graph whose vertices are all the major airports in India which you think of as points sitting inside a unit square. The edge set connects all pairs of vertices – it is a complete graph. Suppose for each vertex v, you are given the (x, y) coordinates of v. Finding shortest paths now is a mere triviality – just an application of the classic distance formula. In general, the picture to keep in mind is the following: you are given a graph $G = (V, E)$ with positive edge weights. Finding s-t shortest paths in such graphs can be an expensive operation which might take $\Omega(m)$ time. Geometric ideas come into their own when I can give you an *embedding* which reveals the coordinates (x_v, y_v) to which a vertex v is mapped. If the embedding has some nice properties, you can show that this geometric picture gives good approximate answers and that too fairly quickly! Of course, you might have to embed the vertices in a space with dimesionality larger than two, but as long as the number of dimensions remains small, this geometric technique does fairly well in terms of recovering approximate distances which are within a small multiplicative factor of the true distances. Indeed, you will not be surprised to recall that these geometric ideas are fairly routine

for various classification tasks in machine learning where you embed dataitems from your dataset as vectors in some suitable vector space. The niceness of the embedding often then results in niceness of the classification task you were after. With this motivation, let us take a deeper dive.

2.2 Preliminaries on Discrete Geometry: The notion of a Metric Space

A metric space (or a metric) is a mathematical notion of distance. Let us see the formal definition.

Definition 2.1. *Let X be a set and let $d\colon X \times X \to \mathbb{R}_{\geq 0}$. We call the pair (X, d) a metric space if the following conditions all hold*

- $d(x, x) = 0$ *for all* $x \in X$.
- *Symmetry:* $d(x, y) = d(y, x)$ *for all* $x, y \in X$.
- *Triangle Inequality:* $d(x, z) \leq d(x, y) + d(y, z)$ *for all* $x, y, z \in X$.

The function $d(.,.)$ above is called the distance function on X. It is usually convenient to think of a metric as a vector $d \in \mathbb{R}^{\binom{n}{2}}$. Let us now see a few examples of metric spaces which help parse the definition better.

2.2.1 Examples of Metric Spaces

- **The Line Metric:** Given a set X and a function $f\colon X \to \mathbb{R}$, define $d(x, y) = |f(x) - f(y)|$. Verify that the pair (X, d) obtained this way is a legitimate metric.
- **The Cut Metric:** Given a set X and a boolean function $f\colon X \to \{0, 1\}$, define

$$d(x, y) = \begin{cases} 1 \text{ if } f(x) \neq f(y) \\ 0 \text{ Otherwise} \end{cases}.$$

  Again, you know the drill. Verify that (X, d) is a legitimate metric. An equivalent way to define the cut metric (which better illuminates the name) is to take a set $S \subseteq X$ instead of the boolean function f and define $d(x, y) = 1$ iff the pair x, y is separated by the cut $(S, \overline{S})$. Otherwise, you define $d(x, y) = 0$.
- **The ℓ_1 metric:** Given a set X and a function $f\colon X \to \mathbb{R}^m$, define $d(x, y) = \|f(x) - f(y)\|_1$. Again, verify that (X, d) is a legitimate metric. It is helpful to note that this is just a sum of line metrics you saw in the above example.
- **Graph Metric/Shortest Path Metric:** Given a set X and an undirected graph $G = (X, E)$, define the distance $d(x, y) = d_G(x, y)$ where $d_G(x, y)$ denotes the length of the shortest path between x and y. Verify that this is a legitimate metric.

The point of these examples is to realize that the behavior of metric spaces can range from being pretty benign to pretty wild. The first two examples present arguably simpler metric spaces as

opposed to the last two examples which appear to be pretty wild metric spaces. In particular, let us see how wild graph metrics can get. To this end, we note that any finite metric space can be simulated by a graph metric which means graph metrics is the most general class of metrics[1].

Often, we would like to understand a sophisticated metric in terms of a benign, easier-to-understand metric. You are invited to show a very simple example of this phenomenon via the following exercise.

Problem 2.2. *Show that any cut metric (X, d_S) which is given by $S \subseteq X$ is a line metric in disguise (and thus, it is also an ℓ_1 metric).*

Next, we show the converse. This essentially says that the cut metrics also exhibit a rich behavior in the following sense. We will show (in Theorem 2.8) that any ℓ_1 metric can be represented as a positive linear combination (or a *conic* combination) of cut metrics. It would be good to start with the following definition.

Definition 2.3. ***Cut-Cone*** *We define cut-cone to be the set of all metrics $d \in \mathbb{R}^{\binom{n}{2}}$. That is, we say*

$$\mathsf{CutCone} = \{d \in \mathbb{R}^{\binom{n}{2}} : d = \sum_{S \subseteq [n]} \alpha_S d_S \text{ where } \alpha_S \geq 0 \text{ and } d_S \text{ is a cut-metric}\}.$$

You should VERIFY that each $d \in \mathsf{CutCone}$ is a bonafide metric. With that verification done, we are ready to show something rather cool.

2.2.2 $[0,1]$-valued-line-metrics lie in the CutCone

A line metric is said to be a $[0,1]$-valued-line-metric if it is a line metric where all the distances in the metric lie in the unit interval. That is, it is a line metric (X, d) where you can find a function $f : X \to [0,1]$ such that for all $x, y \in X$ it holds that $d(x,y) = |f(x) - f(y)|$.

Lemma 2.4. *Let (X, d) denote a finite $[0,1]$-valued-line-metric where $d(x,y) = |f(x) - f(y)|$ for some $f : X \to [0,1]$. Pick a threshold $t \sim [0,1]$ uniformly at random. Define*

$$S_t = \{x : f(x) > t\}.$$

Then, for every $u, v \in X$ it holds that

$$\mathbf{E}[d_{S_t}(u,v)] = d(u,v).$$

In other words, $d = \mathbf{E}[d_{S_t}]$ is a distribution over cut-metrics.

Proof of Lemma 2.4. Note that d_{S_t} is the distance function for the cut metric you end up with the set $S_t \subseteq X$ in your hand. Fix a pair $u, v \in X$ and suppose, w.log, $f(u) \leq f(v)$. Note

[1] Indeed, given any metric space (X, d) you can simulate distances between all pairs of vertices trivially by a complete graph where you set the weight on an edge (i, j) to equal $d(i, j)$. Also, if there exists an edge e such that $d_{G \setminus e} = d$ for all pairs of vertices, you can remove such an edge. You can keep removing edges till you arrive at a minimal subgraph removing edges from which induces a metric different from d

$$\begin{aligned}\mathbf{E}[d_{S_t}(u,v)] &= \Pr_t\left(d_{S_t}(u,v)=1\right)\\ &= \Pr_t\left(S_t \ni v \text{ and } S_t \not\ni u\right)\\ &= \Pr_t\left(f(u)\le t\le f(v)\right)\\ &= |f(v)-f(u)|\\ &= d(u,v)\end{aligned}$$

The first step merely follows by the definition of expectation. The second step is a consequence of the definition of the cut-metric. The next step is immediate from the definition of the set S_t. ■

Remark 2.5.

- You are invited to show that a $[0,1]$-valued-line-metric d as above lies in the CutCone. Evidently, you want to express d as a conic combination of only the sweep cuts. What should the coefficients, α_t of the sweep sets $\{S_t\}_t$ be?
- If the proof of Cheeger's inequality from lecture #1 is fresh in your mind, then the above lemma should give you a pause. You might recall that the Cheeger Algorithm arranged the entries of a non-negative vector (obtained by a truncation) and then took sweep cuts on it. I told you that this mysterious step of sorting your vector was done for a deep reason. Lemma 2.4 reveals there is a deep connection between embeddings and cuts. So, in particular, if you are able to obtain a nice embedding of your vector on the unit interval, Lemma 2.4 hints that the solution to your cut problem is lurking behind in the corner.

With Lemma 2.4 in our hand, you can make quick work of the following claim.

Claim 2.6. *Consider a line metric (X,d) where $d(x,y)=|f(x)-f(y)|$ for some*

$$f\colon X\to\mathbb{R}_{\ge 0} \text{ where } \min_{x\in X} f(x)=0.$$

Then d can be written as a conic combination of cut metrics.

Proof. Obvious by rescaling. ■

Claim 2.7. *Consider a line metric (X,d) where $d(x,y)=|f(x)-f(y)|$ for some*

$$f\colon X\to\mathbb{R}.$$

Then d can be written as a conic combination of cut metrics.

Proof. Translate and rescale. ■

Finally, we show that the conic combination of cut metrics can simulate any ℓ_1 metric over $\mathbb{R}^k$.

Theorem 2.8. *Consider a ℓ_1 metric on $\mathbb{R}^k$ (X, d) where $d(x, y) = \|f(x) - f(y)\|_1$ for some*

$$f : X \to \mathbb{R}^k.$$

Then d can be written as a conic combination of cut metrics.

Proof of Theorem 2.8. Use Claim 2.7 coordinate-wise. ■

Apriori, one might think that an ℓ_1 metric is a combination of a huge number of cuts. The following exercise invites you to upperbound the number of cuts a n point ℓ_1 metric over $\mathbb{R}^k$ could be a linear combination of.

Problem 2.9. *Let (X, d) be a ℓ_1 metric space. That is, there exists a mapping $f : X \to \mathbb{R}^k$ such that $d(x, y) = \|f(x) - f(y)\|_1$. Show that there exists a collection of $k \cdot n$ many cuts $(S_t, \overline{S_t})$ and non-negative numbers α_t's such that*

$$d = \sum_{t=1}^{t=k\cdot n} \alpha_t \cdot d_{S_t}.$$

Alright, we have seen a rich interplay between the first three metrics in the list we considered above. We discovered that if we allow for conic combinations, all these metrics are kind of equivalent to each other. Dare we hope for more? Can we show that the wildest example in the list above – the graph metric – is also as benign as other metrics in the list? If we settle for approximate answers, the answer is YES as Bourgain showed in his famous theorem. We cover that theorem in the next section.

3 Bourgain's Theorem

To understand in what approximate sense can one say a graph metric (that is to say *any metric really*) can be simulated as a simpler ℓ_1 metric, we first start with a definition.

Definition 3.1. ***[Distortion]*** *Let (X_1, d_1) and (X_2, d_2) be metric spaces. An embedding $f : X_1 \to X_2$ has distortion at most $\alpha \geq 1$ if there exists an $r > 0$ such that for all $x, y \in X_1$,*

$$r \cdot d_1(x, y) \leq d_2(f(x), f(y)) \leq \alpha \cdot r \cdot d_1(x, y).$$

We often call the space X_2 the host space as it contains embedded inside it, a copy of the space X_1 up to some disortion.

Remark 3.2. Usually, we will consider host spaces which are geometrically easier to think about. Indeed, in this class, the host spaces of interest to us are only ℓ_1 and ℓ_2 (over $\mathbb{R}^k$). Since, the host spaces of interest are not merely ordinary sets but sets with vector space structure, you can now appreciate the definition of distortion better. Suppose the metric (X_1, d_1) embeds with distortion at most α in a $\mathbb{R}^k$ equipped with (say) the ℓ_1 metric. The reason you have two terms sandwiching $d_1(x, y)$ up to factor α just says that the notion of distortion is invariant upto scaling. You can very well, for convenience, pick $r = 1$ and say instead that there exists a mapping f' such that

$$\forall x, y \in X_1,\ d_1(x, y) \leq \|f'(x), f'(y)\|_1 \leq \alpha d_1(x, y).$$

Note that such mappings never contract distances.

I state without proof the following theorem. I will only say the key is to use a semidefinite program.

Theorem 3.3. *For any metric space (X, d) we can determine in polynomial time the minimum α such that (X, d) embeds into ℓ_2 (over $\mathbb{R}^k$ for some $k \in \mathbb{N}$) with distortion at most α.*

Proof. Use a semidefinite program. For more details, see [Arora05]. ■

Alright, so ℓ_2 spaces are nice hosts even from the computational viewpoint (which seeks to find embeddings in ℓ_2 with least distortion). Computationally speaking, ℓ_1 spaces are not that nice. However, this is not too unexpected. After all, thanks to our nice workout in Theorem 2.8, we know the geometry of ℓ_1 embeddings is closely related to combinatorics of cut metrics. And cut metrics are closely related to cut problems in graphs. As we will see, the ℓ_1 metric inherits its computational nastiness from the cut-metric. We elaborate on this in the following remark.

Remark 3.4. *We consider some classic cut problems and show how to cast them in a language of metric embeddings.*

1. ***Metric Version of the* Min-Cut *Problem:*** One of the easiest cut problems, the so called Min-Cut problem seeks to find $S \subseteq V$ which minimizes $|E(S, \overline{S})|$. As you know, this is solvable in time polynomial in the size of the graph.

2. ***Metric Version of the* SparsestCut *problem:*** On the other hand, the poster child problem for this class, the SparsestCut problem is NP-hard. As you know, this problem seeks to find $S \subseteq V, |S| \leq |V|/2$ which minimizes $\frac{|E(S,\overline{S})}{|S|}$. Both these problems are related to cut metrics. In particular, the Min-Cut problem is equivalent to minimizing $\sum_{(i,j)\in E} d_S(i,j)$. For the SparsestCut problem, you can instead minimize

$$\frac{n}{d} \cdot \frac{\sum_{(i,j)\in E} d_S(i,j)}{\sum_{i<j} d_S(i,j)} = \frac{n}{d} \cdot \frac{|E(S,\overline{S})|}{|S| \cdot |\overline{S}|}.$$

   Up to a constant factor, minimizing the latter expression minimizes conductance/SparsestCut. In 1985, [Bourgain] showed that even though finding the least distortion embedding in ℓ_1 is computationally difficult, you can still find embeddings in ℓ_1 with not terribly large distortion of *any n-point metric space* into ℓ_1.

Theorem 3.5. *Given any finite metric space (X, d) with $|X| = n$, there exists an embedding of (X, d) into $(\mathbb{R}^k, \ell_1)$ where:*

- $k = O(\log^2 n)$.
- *The distortion of the embedding is at most $O(\log n)$.*

This is an essential result in the theory of metric embeddings. Coolly enough, the proof is algorithmic and it is a nice example of what feats you can achieve with randomization. So the embedding does not bloat up the dimension too much. Moreover, you don't suffer too much distortion. Later in a homework, you will improve the dimensionality to $O(\log n)$. I note that you cannot however improve the distortion incurred. First, let me present the (randomized) algorithm which constructs this embedding which has the desired properties with high probability.

`BourgainEmbedding`(X, d)

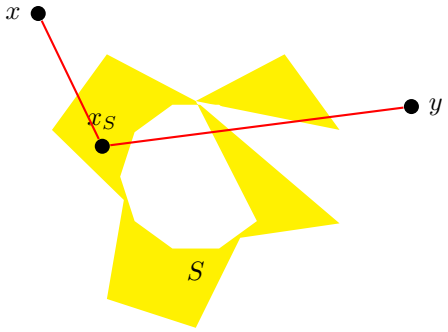
1. For $i = 1, 2, \ldots \log n$.
   (a) For $j = 1, 2, \ldots, c \log n$.
   (b) Assign $S_{i,j} = \emptyset$.
      i. Add each element in X to $S_{i,j}$ with probability 2^{-i}.
2. For $x \in X$
   (a) Define $f(x) = \left(d(x, S_{1,1}), d(x, S_{1,2}), \ldots, d(x, S_{\log n, c \cdot \log n}) \right)^T$

The last line of the above algorithm considers the distance between $x \in X$ and a set $S \subseteq X$. This is defined as the distance of x to the point $x_S \in S$ closest to x, that is $d(x, S) = \min_{y \in S} d(x, y)$. To get more intuition about what the algorithm is doing, fix some $x \in X$ and let us consider some specific iterations for $i \in \{1, 2, \log n\}$. Note that when $i = 1$, **line 1.(b).i** of the algorithm constructs (random) sets $(S_{1,1}, S_{1,2}, \ldots S_{1,c \log n})$. Each and every one of these $c \log n$ sets contains $n/2$ points in expectation. At the end, you get $c \log n$ distances from x to all of these sets in the first iteration. With $i = 2$, you again have $c \cdot \log n$ sets each of which is now less dense and contains around $n/4$ points. Finally, you reach the $\log n$-th iteration. Again, you get $c \log n$ sets each of which has expected size of just 1 point. And yet again, you record all the $c \log n$ distances from x to all of these random sets. Now that the algorithm is (hopefully) clear, let us ask why is this a good idea. The key to this is enshrined in the following lemma.

Lemma 3.6. *The map $f_S(x) = d(x, S)$ is non-expanding, that is for all*

$$x, y \in X, \quad |f(x) - f(y)| \leq d(x, S).$$

Proof of Lemma 3.6. The following picture accompanies this proof.



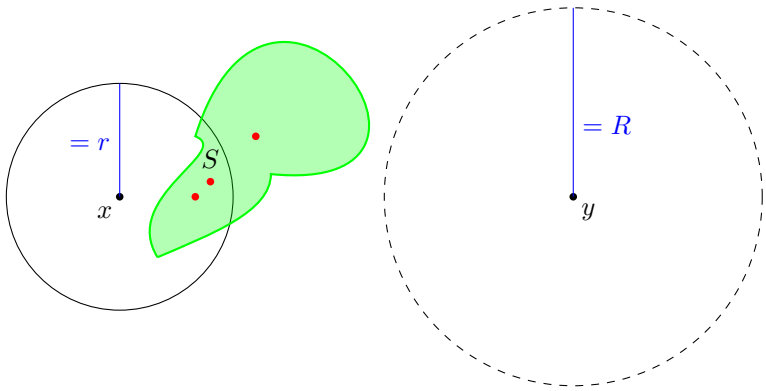
We show the set S in yellow. We fix a pair of points x, y. Suppose $d(x, S) = d(x, x_S)$. Also, observe $d(y, S) \leq d(y, x_S)$. Now, note

$$\begin{aligned} f(y) - f(x) &= d(y, S) - d(x, S) \\ &\leq d(y, x_S) - d(x, x_S) \\ &\leq d(x, y) \qquad \text{By triangle inequality} \end{aligned}$$

On switching the roles of x and y in the above argument, we readily see that $f(x) - f(y) \leq d(x, y)$ as well which settles the lemma. ■

3.1 An intuitive discussion

Now that we have shown this embedding is non-expanding, towards showing that it incurs small distortion, it suffices to show that *with high probability*, this embedding is not-too-shrinking either. So, in particular if we show that for most random sets $S_{i,j}$'s constructed by the algorithm, it holds that $|f(x)-f(y)| \geq \Delta d(x, y)$, for some appropriate value of Δ, then in decent number of coordinates of this embedding you will see something pretty close to the actual distance. In turn, this would mean that the embedding has small distortion. Towards this end, let us fix a pair of points $x, y \in X$ and explore what conditions we want from a random S so that we indeed get an (easy) lowerbound on $|f(x) - f(y)| = |d(x, S) - d(y, S)|$. One such condition is captured in the following picture.



In this picture, the set S (drawn in green) intersects a closed ball of radius r around x denoted

$$B(r, x) = \{z \colon d(z, x) \leq r\}$$

and it is disjoint from the open ball of radius $R > r$ around y which is denoted

$$B^-(y, R) = \{z \colon d(z, y) < R\}.$$

In this case, note that you have $|d(y, S) - d(x, S)| \geq R - r$ which gives us the not-too-shrinking property we wanted. Let us write down the lesson from this intuitive discussion formally via the following claim.

Claim 3.7. *Fix two arbitrary points $x, y \in X$, an index $1 \leq i \leq \log n$ and radii $r < R$. Let S denote some set picked by* `BourgainEmbedding` *algorithm in the i-th iteration in the outer loop. Assume*

$$|B(x, r)| \geq 2^i, \quad |B^-(y, R)| \leq 2^{i+1}, \quad B(x, r) \cap B^-(y, R) = \emptyset.$$

Then with probability at least $1/32$*, we have* $|d(y,S) - d(x,S)| \geq (R-r)$*. This implies that*

$$\mathbf{E}_S\Big[|d(y,S) - d(x,S)|\Big] \geq \Omega(1) \cdot (R-r).$$

Proof of Claim 3.7. If you are a bit paged out, let me page you back in by recalling for you that the only randomness in this problem is the randomness in the choice of the random set S. Alright, so let's begin by identifying the events of interest to us that are motivated by the preceding discussion. Consider the events

- $\mathcal{E}_1 := S \cap B(x,r) \neq \emptyset$, and
- $\mathcal{E}_2 := S \cap B^-(y,R) = \emptyset$, and
- $\mathcal{E} = \mathcal{E}_1 \wedge \mathcal{E}_2$.

We are interested in lower bounding $\Pr(\mathcal{E})$. Bourgain has set this up very nicely because the events $\mathcal{E}_1$ and $\mathcal{E}_2$ are now independent(!). This follows because each point independently opts to be in the set S with probability 2^{-i} and the two sets $B(x,r)$ and $B^-(y,R)$ are disjoint. Thus,

$$\Pr(\mathcal{E}) = \Pr(\mathcal{E}_1) \times \Pr(\mathcal{E}_2).$$

Now, we will just lower bound both of these probabilities. Write

$$\Pr(\mathcal{E}_1) = \Pr\big(B(x,r) \cap S \neq \emptyset\big) = 1 - \left(1 - \frac{1}{2^i}\right)^{|B(x,r)|} \geq 1 - \left(1 - \frac{1}{2^i}\right)^{2^i} \geq 1/4.$$

And,

$$\Pr(\mathcal{E}_2) = \Pr\big(B^-(y,r) \cap S = \emptyset\big) = \left(1 - \frac{1}{2^i}\right)^{|B^-(y,R)|} \geq \left(1 - \frac{1}{2^i}\right)^{2^{i+1}} \geq 1/8.$$

This gives $\Pr(\mathcal{E}) \geq 1/32$ as desired.

■

All that remains is to find the exact radii r and R which gives us the desired number of points in each ball. In other words, we want the set S to have an appropriate density relative to $B(x,r)$ and $B^-(y,R)$. This is done by slowly increasing the radii and then applying the claim.

Overall, this suggests the following strategy. We will show that there are a whole bunch of random sets at level "i" which contribute $\Delta \cdot d(x,y)$ to the distance between points $f(x)$ and $f(y)$. This means a lot of the coordinates in the difference $\|f(x) - f(y)\|_1$ roughly contribute the correct value. This is achieved by arguing that for each i, we pick enough random j's and by showing that this good situation occurs enough of the times.

3.2 Wrapping up with a formal proof

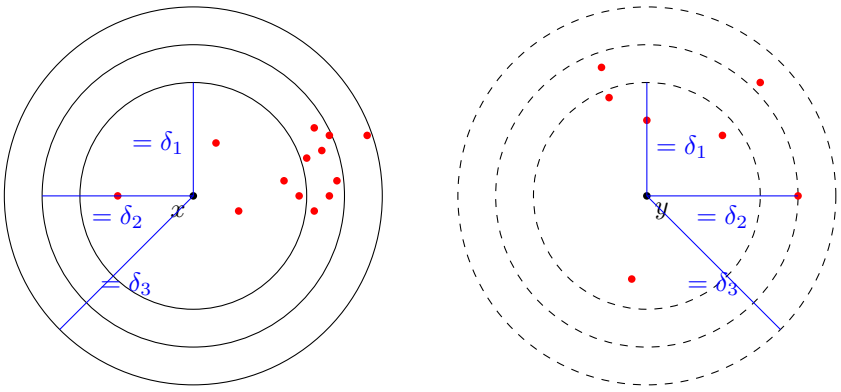
Lemma 3.8. *Let $k = c\log^2 n$. Let f be the embedding obtained by* `BourgainEmbedding`(X, d). *Then with probability at least $1 - 1/n$, for all $x, y \in X$ we have*

$$\frac{k \cdot d(x,y)}{O(\log n)} \leq \|f(x) - f(y)\|_1 \leq k \cdot d(x,y).$$

Note that Lemma 3.8 immediately implies Theorem 3.5. So, we focus on proving this result.

Proof of Lemma 3.8. We have two inequalities. The right-most inequality follows immediately thanks to Lemma 3.6 (why?). Now, let us show the inequality on the left. Fix some $i \leq \log n$ and fix a pair of points x, y. We will show that with probability at least $1 - 1/n^3$, $\frac{k \cdot d(x,y)}{O(\log n)} \leq \|f(x) - f(y)\|_1$. Then, by a union bound over all pairs of points in X, levels, the result will follow.

Recall, from our discussion in §3.1 that for a fixed $i \leq \log n$, we want S to hit a ball of some radius around x and to remain disjoint from another ball of some radius around y. We want the absolute difference between these radii to be comparable to some quantity that is appropriate for the pair of events x, y and the index i (up to a little shrinking). To this end, we choose a bunch of different radii in the following sequence: $0 = \delta_0 < \delta_1 < \delta_2 < \ldots < \delta_t$. Here, the radius δ_i is chosen to be the smallest value (at most $d(x,y)/3$ so that $B(x, \delta_i)$ and $B(y, \delta_i)$ both contain at least 2^i points from the ambient space X. We keep defining new δ_i's till we hit the largest index t such that $\delta_t < d(x,y)/3$ and we define $\delta_{t+1} = d(x,y)/3$. For an illustration, see the following picture.



In the remaining discussion, it will be helpful to fix some $i \leq \log n$ and analyze the situation for this fixed i. We will show that for a fixed pair $x, y \in X$ and a fixed i, with probability at least $1 - 1/n^4$ it holds that

$$\sum_{j \in c\log n} |d(x, S_{i,j}) - d(y, S_{i,j})| \geq \Omega(logn) \cdot (\delta_{i+1} - \delta_i).$$

To this end, recall that Claim 3.7 shows that the expected contribution to the distance between $f(x)$ and $f(y)$ with respect to a set S chosen in the i-th iteration is at least $\Omega(1) \cdot (\delta_{i+1} - \delta_i)$.

By a Chernoff bound it holds that a $1/2^6$ fraction of the $c\log n$ sets considered in line 1.(a) of `BourgainEmbedding`(X, d) contribute at least $\Omega(1)(\delta_{i+1} - \delta_i)$ to ℓ_1 distance between $f(x)$ and $f(y)$.

By a union bound over all the $\log n$ choices for i, with probability at least $1 - \log n/n^4 \geq 1 - 1/n^3$, for a fixed pair x, y you get

$$\|f(x) - f(y)\|_1 = \sum_{i,j} |d(x, S_{i,j}) - d(y, S_{i,j})| \geq \frac{c \cdot \log n}{2^6} \sum_{i=1}^{t} (\delta_{i+1} - \delta_i) = \frac{c \log n}{2^6} \cdot \delta_t = \frac{k}{2^6 \log n} \cdot \frac{d(x,y)}{3}.$$

And this finishes the proof.

■

4 Repurposing Bourgain for Cut Problems

Finally, we come to the last section in today's lecture. This section chronicles how computer scientists combined the ideas of Bourgain with ideas from continuous optimization toward solving interesting cut problems on graphs. I have been sloppy so far because I interchangeably switched between saying sparsest cut and the cut with the smallest conductance. Let me fix this by formally defining the sparsest cut problem. Given a graph $G = (V, E)$, for a set S you define its sparsity as

$$\mathsf{SC}(S) = \frac{|E(S, \overline{S})|}{|S| \cdot |\overline{S}|}$$

Note that unlike conductance, this quantity is symmetric in the sense that $\mathsf{SC}(S) = \mathsf{SC}(\overline{S})$. You define the sparsity of a graph G as follows:

$$\mathsf{SC}(G) = \min_{\emptyset \neq S \subset V} \mathsf{SC}(S).$$

To see why you might care about the sparsest cut objective, note that it is closely related with the problem of finding the smallest conductance cut. Indeed, note that the expression $n/d \cdot \mathsf{SC}(G)$ is at most twice the conductance of the graph (why?). You might recall that we encountered this quantity in lecture #1 as well. Anyway, the punchline is minimizing over cuts with small sparsity allows you to recover cuts with small conductance and sparsity is a bit better behaved notion thanks to the symmetry it enjoys.

Towards finding good approximation algorithms for the sparsest cut problem, in a seminal work, Leighton and Rao started from the metric version of the sparsest cut problem (see **Item 2** in Remark 3.4). Letting S denote the set minimizing sparsity, we first rewrite the sparsest cut objective as follows

$$\mathsf{SC}(G) = \frac{\sum_{(u,v) \in E} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\sum_{u,v \in V} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|} \tag{4.1}$$

In the cut metric language, you can write $d_S(u, v) = |\mathbf{1}_S(u) - \mathbf{1}_S(v)|$. This gives

$$\mathsf{SC}(G) = \frac{\sum_{(u,v)\in E} d_S(u,v)}{\sum_{u,v\in V} d_S(u,v)} \tag{4.2}$$

Leighton and Rao considered a relaxation of this objective which we present next.

4.1 The Leighton-Rao relaxation

Define

$$\mathsf{LR}(G) = \min_{d\text{: metric over } V} \frac{\sum_{(u,v)\in E} d(u,v)}{\sum_{u,v\in V} d(u,v)}$$

Note that $\mathsf{LR}(G)$ is a lowerbound on $\mathsf{SC}(G)$ as $\mathsf{LR}(G)$ takes a minimum over all metrics (including the cut metrics). Next up, as Leighton and Rao observed, you can compute $\mathsf{LR}(G)$ using a linear program.

Claim 4.1. *$\mathsf{LR}(G)$ can be modeled as a linear program.*

Proof of Claim 4.1. The only non-linearity you see in the $\mathsf{LR}(G)$ objective is the denominator. The constraints just insist that d better be a legitimate metric – and they are all linear. We can get rid of the non-linearity in the $\mathsf{LR}(G)$ objective by adding a constraint that

$$\sum_{u,v\in V} d(u,v) = 1.$$

That is to say you consider the following linear program.

$$\begin{aligned} \text{minimize} \quad & \sum_{(u,v)\in E} d(u,v) \\ \text{subject to} \quad & \sum_{u,v} d(u,v) = 1 \\ & d(u,v) \le d(u,w) + d(w,v) \qquad \forall u,v,w \\ & d(u,v) \ge 0 \qquad\qquad\qquad\quad \forall \{u,v\} \in \binom{V}{2} \end{aligned} \tag{4.3}$$

To see this exactly captures $\mathsf{LR}(G)$, let d' denote the metric which achieves this minimum. Note that there exists another metric d'' which is obtained by rescaling d' so that the $\sum_{u,v\in V'} d'(u,v) = 1$. This finishes the proof. ■

The following lemma is immediate from the theory of linear programming.

Lemma 4.2. *Let λ^* denote the optimal value of the LP considered in (Equation 4.3). Then $\lambda^* = \mathsf{LR}(G)$. Moreover, λ^* can be computed in polynomial time.*

Now, we prepare for the final punchline. We will show that $\mathsf{SC}(G) \le O(\log n) \cdot \mathsf{LR}(G)$. Together with $\mathsf{LR}(G) \le \mathsf{SC}(G)$, this means the LP considered in (Equation 4.3) returns a good approximation

to the sparsest cut. The beating heart of Leighton-Rao's approach lies a blackbox application of Bourgain's theorem. This says that any metric (V, d) can be captured by ℓ_1 with at most a $O(\log n)$ distortion. Next, you recall that ℓ_1 metrics are a conic combination of cut-metrics and you use this connection to find a good cut. Alright, let us put this plan in motion.

Lemma 4.3. $SC(G) \leq O(\log n) \cdot LR(G)$.

*Proof of L.*et d^* denote an optimal metric returned by the linear program in (Equation 4.3). This means we have

$$\mathsf{LR}(G) = \frac{\sum_{(u,v)\in E} d^*(u,v)}{\sum_{u,v\in V} d^*(u,v)}. \tag{4.4}$$

Bourgain blackbox!

Using Bourgain, let us embed (V, d^*) into $(\mathbb{R}^k, \ell_1)$ where $k = O(\log^2 n)$ with a (non-shrinking) Bourgain mapping given by $f\colon V \to \mathbb{R}^k$. Thus, for all pairs $u, v \in V$, we have

$$\|f(u) - f(v)\|_1 \leq d^*(u,v) \leq O(\log n)\|f(u) - f(v)\|_1.$$

Consider the numerator on the RHS of (Equation 4.4). We note $\sum_{(u,v)\in E} d^*(u,v) \geq \sum_{(u,v)\in E} \|f(u) - f(v)\|_1$ by the non-shrinking property of the mapping. For the denominator, note that $\sum_{u,v\in V} d^*(u,v) \leq \sum_{u,v\in V} O(\log n)\|f(u) - f(v)\|_1$.

Overall, this gives

$$\mathsf{LR}(G) = \frac{\sum_{(u,v)\in E} d^*(u,v)}{\sum_{u,v\in V} d^*(u,v)} \geq \frac{1}{O(\log n)} \cdot \frac{\sum_{(u,v)\in E} \|f(u) - f(v)\|_1}{\sum_{u,v\in V} \|f(u) - f(v)\|_1}. \tag{4.5}$$

At this point, we recall that an ℓ_1 metric is a conic combination of cut-metrics. This means, for all pairs of vertices $u, v \in V$, you have

$$\|f(u) - f(v)\|_1 = \sum_{S\in\mathcal{S}_f} \alpha_S d_S(u,v)$$

where $\mathcal{S}_f$ is a collection of $n \cdot k = O(n \log^2 n)$ many cuts. Starting from (Equation 4.5), we have

$$
\begin{aligned}
\mathsf{LR}(G) &\geq \frac{1}{O(\log n)} \cdot \frac{\sum_{(u,v)\in E} \|f(u)-f(v)\|_1}{\sum_{u,v\in V} \|f(u)-f(v)\|_1} \\
&= \frac{1}{O(\log n)} \cdot \frac{\sum_{(u,v)\in E} \sum_{S\in\mathcal{S}_f} \alpha_S d_S(u,v)}{\sum_{u,v\in V} \sum_{S\in\mathcal{S}_f} \alpha_S d_S(u,v)} \\
&= \frac{1}{O(\log n)} \cdot \frac{\sum_{S\in\mathcal{S}_f} \alpha_S \sum_{(u,v)\in E} d_S(u,v)}{\sum_{S\in\mathcal{S}_f} \alpha_S \sum_{u,v\in V} d_S(u,v)} && \text{Switch sums in numerator and denominator} \\
&\geq \frac{1}{O(\log n)} \cdot \min_{S\in\mathcal{S}_f} \frac{\sum_{(u,v)\in E} d_S(u,v)}{\sum_{u,v\in V} d_S(u,v)} && \text{Because } \frac{\sum_{i=1}^k a_i}{\sum_{i=1}^k b_i} \geq \min_{i\in[k]} \frac{a_i}{b_i} \\
&= \frac{1}{O(\log n)} \cdot \mathsf{SC}(G)
\end{aligned}
$$

This shows that indeed $\mathsf{LR}(G)$ is a good approximation to $\mathsf{SC}(G)$ up to logarithmic factors. Note that this approach is in fact algorithmic. Indeed, the algorithm just considers all the cuts in the collection $\mathcal{S}_f$ and returns the one with smallest sparsity.

■

References

[Bourgain] Jean Bourgain, On Lipschitz embedding of finite metric spaces in Hilbert space *Israel Journal of Mathematics*, 1985

[Trevisan] Luca Trevisan, Handout 09 and Handout 10, https://lucatrevisan.github.io/expanders2016/, 2016

[Saranurak] Thatchaphol Saranurak Lectures 2.1 and 2.2, https://sites.google.com/site/thsaranurak/teaching/Expander?authuser=0, 2021

[Arora05] Sanjeev Arora, Lecture 01, https://www.cs.princeton.edu/courses/archive/spring05/cos598B/lecture1.pdf, 2005

[VW] Gregory Valiant and Mary Wooters Lecture 07, https://web.stanford.edu/class/cs265/Lectures/Lecture7/l7.pdf, 2022