University of Michigan–Ann Arbor

Department of Electrical Engineering and Computer Science

EECS 498 004 **Advanced Graph Algorithms**, Fall 2021

Lecture 3: Approximating Conductance

September 7, 2021

Instructor: Thatchaphol Saranurak

Scribe: Gary Hoppenworth

1. Overview

Conductance is an important property of graphs that measures how well-connected they are. Expanders are graphs with "large" conductance. (Typically, graphs on n vertices are considered good expanders if they have conductance greater than $1/\text{polylog}\, n$.) Unfortunately, computing conductance exactly is NP-hard in general graphs.

However, there is an efficient $O(\log n)$-approximation of conductance which allows us to check the expansion of a graph (up to some extent). This approximation relies on two ingredients: 1) Bourgain's metric embedding into ℓ_1-metrics and 2) connections between ℓ_1-metrics and cut-metrics. These two tools are used to obtain an approximation via linear programming relaxation. In this lecture, we introduce the necessary background for the above two ingredients and then prove the LP relaxation approximation.

2. Metrics

A metric space (or metric) is a mathematical notion that captures distance. Metric spaces are composed of a set X of points and a distance function d that lets us measure the distance between points.

2.1. Definition. Let X be a set and let d be a function $d : X \times X \mapsto \mathbb{R}_{\geq 0}$. We say that (X, d) is a **metric** if the following hold:

1. $d(x, x) = 0$ for all $x \in X$ [1]
2. $d(x, y) = d(y, x)$ for all $x, y \in X$
3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$

The function d is called the **distance function**.

[1] In mathematics, this condition is stated as $d(x, y) = 0$ if and only if $x = y$. What we call metric in computer science is usually referred to as semi-metric in math.

2.1. Examples of Metrics

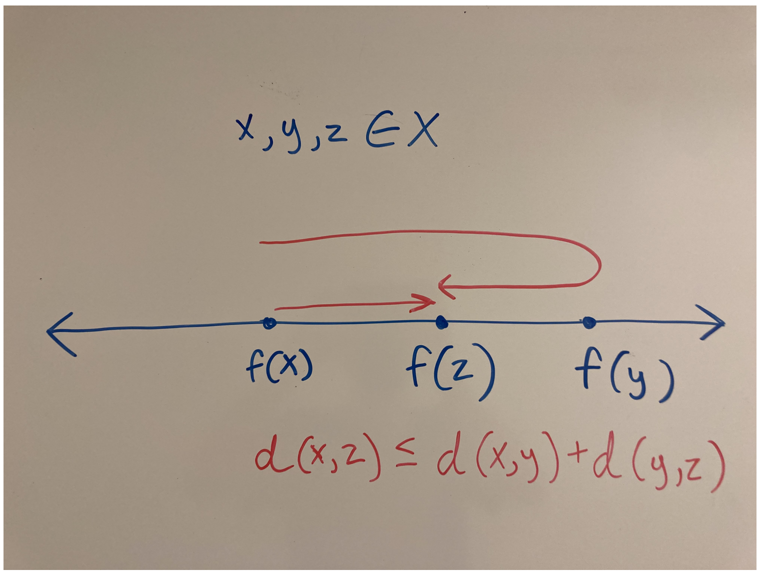
Line Metric: Given a set X and a function $f : X \mapsto \mathbb{R}$, define the distance function as $d(x, y) = |f(x) - f(y)|$.

Proof that this is a metric:

1. $d(x, x) = |f(x) - f(x)| = 0$ for all $x \in X$
2. $d(x, y) = |f(x) - f(y)| = |f(y) - f(x)| = d(y, x)$ for all $x, y \in X$.
3. $d(x, z) = |f(x) - f(z)| = |f(x) - f(y) + f(y) - f(z)| \leq |f(x) - f(y)| + |f(y) - f(z)| = d(x, y) + d(y, z)$ for all $x, y, z \in X$

□

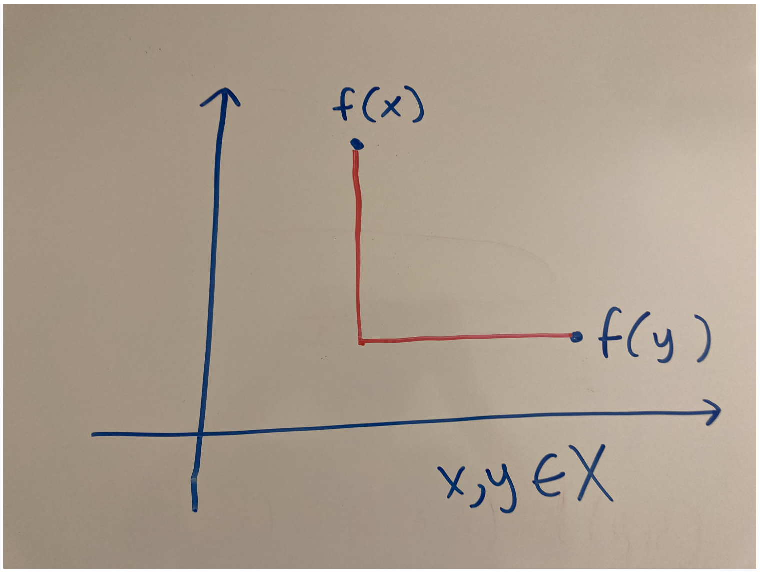
Below is a visualization of a line metric and the proof of statement 3.



ℓ_1-Metric: Given a set X and a function $f : X \mapsto \mathbb{R}^m$, define the distance function as $d(x, y) = \|f(x) - f(y)\|_1 = \sum_i |f_i(x) - f_i(y)|$. (Note when $m = 1$ this is the line metric.)

Proof that this is a metric: The ℓ_1-metric is really a sum of line metrics. This proof is essentially the same as the previous proof. □

Below is a visualization of distances in an ℓ_1-metric.



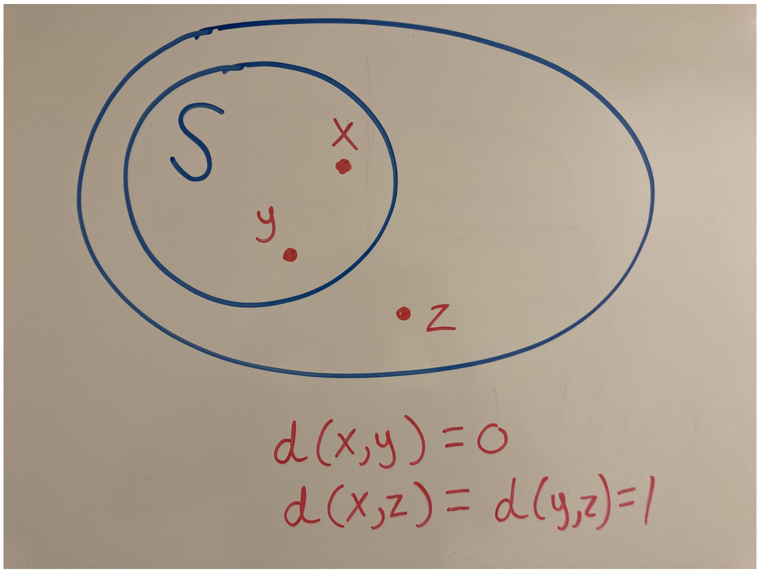
Cut Metric: Given a set X and a set $S \subseteq X$, let $d(x,y) = 0$ if $x, y \in S$ or $x, y \in X \backslash S$. Else, $d(x,y) = 1$.

Proof that this is a metric:

1. $d(x,x) = 0$, since x cannot belong to both S and $X \setminus S$.
2. $d(x,y) = d(y,x)$ since the definition of d does not depend on the order of x and y.
3. If $d(x,z) = 0$, then the inequality immediately holds. Else $d(x,z) = 1$. Assume without loss of generality that $x \in S$ and $z \in X \setminus S$. If $y \in S$, then $d(y,z) = 1$. Otherwise, $y \in X \setminus S$ and $d(x,y) = 1$. Either way, $d(x,y) + d(y,z) \geq 1 = d(x,z)$.

□

Below is a visualization of distances in the cut metric for a set S.



Shortest Path Metric: Given a set X and an undirected graph $G = (X, E)$, define the distance function as $d(x, y) = d_G(x, y)$, where $d_G(x, y)$ is the length of the shortest (x, y)-path in G.

Proof that this is a metric:

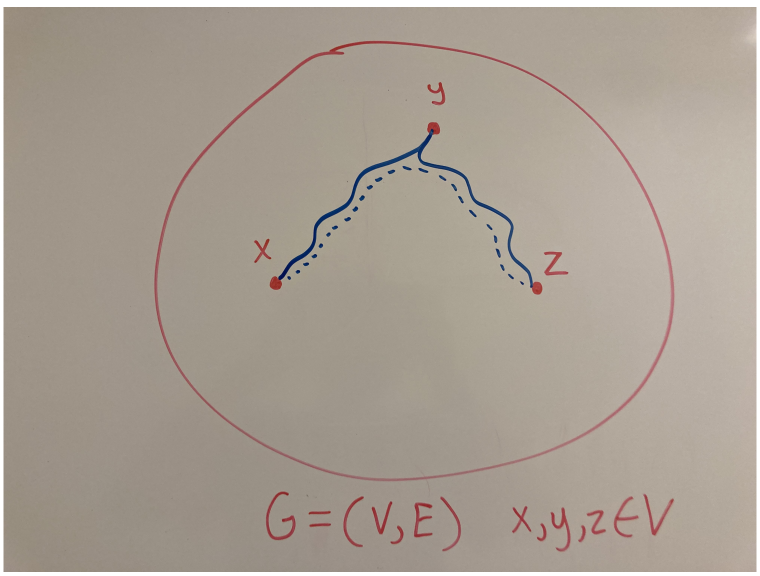
1. $d(x, x) = \text{dist}_G(x, x) = 0$. (The shortest path from a vertex to itself has length zero.)

2. $d(x, y) = \text{dist}_G(x, y) = \text{dist}_G(y, x) = d(y, x)$. (A shortest (x, y)-path has the same length as a shortest (y, x)-path in any undirected graph.)

3. $d(x, z) = \text{dist}_G(x, z) \leq \text{dist}_G(x, y) + \text{dist}_G(y, z) = d(x, y) + d(y, z)$. The inequality follows from the fact that a (x, y)-shortest path can be combined with a (y, z)-shortest path to obtain a (x, z)-path.

□

Below is a visualization of the proof of the triangle inequality (statement 3) for undirected graphs.



3. Metric Embeddings

3.1. Motivation

Computing the distance in some metrics is more expensive than in other metrics. This can be seen by comparing the costs of computing distances in the shortest path metric and the ℓ_1-metric.

Consider a shortest path metric of a graph G on n vertices and m edges. Computing $d(x, y) = \text{dist}_G(x, y)$ where $x, y \in V(G)$ takes $\Omega(m)$ time using Dijkstra's algorithm. By precomputing all distances you can answer queries in $O(1)$ time, but this requires $\Omega(n^2)$ space. These computations can be expensive for large graphs.

On the other hand, for an ℓ_1-metric with dimension k, computing $d(x,y) = \|f(x) - f(y)\|$, where $x, y \in \mathbb{R}^k$, takes $O(k)$ time by simply reading x and y and performing $O(k)$ operations. This computation is cheap for small k.

Because we can compute distances in a low dimension ℓ_1-metric quickly, it would be advantageous to represent any metric (X,d) as an ℓ_1-metric with low dimension. This would allow us to measure distance in the (X,d) metric (perhaps within some distortion factor) using cheaper computations. We call this representation an embedding and define it formally in the next section.

3.2. Formal Definition

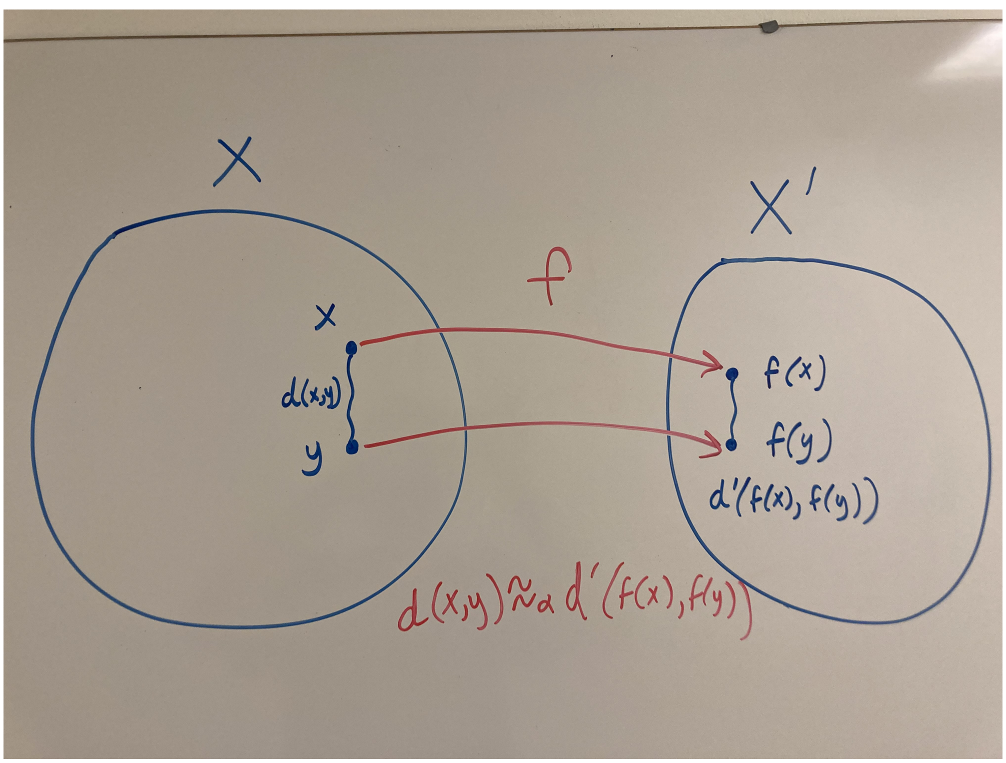
3.1. Definition (Metric Embeddings)**.** Let (X,d) and (X',d') be two metrics. Let $f : X \mapsto X'$ be a mapping such that for any $x, y \in X$

$$d'(f(x), f(y)) \leq d(x,y) \leq \alpha \cdot d'(f(x), f(y))$$

Then f is a **metric embedding**. We say that f **embeds** (X,d) into (X',d') with **distortion** α. This can also be written as

$$(X,d) \overset{\alpha}{\hookrightarrow} (X',d')$$

Below is a visualization of the notion of a metric embedding.



3.2. *Remark.* When $\alpha = 1$, we say that f has **no distortion**, or is **distance-preserving**, or is **isometric**.

3.3. Definition (ℓ_p^k-metrics)**.** We say that a metric (X, d) is a ℓ_p^k-metric ($p \geq 1$) if and only if there is an embedding f that embeds (X, d) into (X', d') with no distortion, and (X', d') is an ℓ_p-metric with dimension k.

3.4. Example. A cut metric (X, d_S) is a line metric (and therefore it is an ℓ_1^1-metric).

Proof: Define the function $1_S : X \mapsto \{0, 1\}$ so that $1_S(x) = 1$ if and only if $x \in S$. Let $d'(x, y) = |1_S(x) - 1_S(y)|$ for all $x, y \in X$. Note that (X, d') is a line metric. It follows from the definition of the cut metric that $d_S(x, y) = |1_S(x) - 1_S(y)| = d'(x, y)$ for all $x, y \in X$. Thus we have an embedding of (X, d_S) to (X, d') with distortion $\alpha = 1$. □

We now describe the two main ingredients that we use to approximate conductance. The first ingredient will show that any metric can be embedded in a ℓ_1^p metric for "small" p with low distortion. The second ingredient will show that any ℓ_1^p metric can be expressed as a conic combination of cut-metrics. Thus cut-metrics are in some sense universal: they capture every metric in some (approximate) sense.

3.3. Ingredient 1: Bourgain's Metric Embedding

A surprising fact is that every n-point metric can be embedded into an ℓ_1-metric of dimension $O(\log^2 n)$ with distortion $O(\log n)$. This is powerful because the ℓ_1-metric we are embedding into has both small dimension and small distortion. The formal statement of this fact is as follows.

3.5. Theorem (Bourgain)**.** *Given an n-point metric (X, d), there is an embedding $f : X \to \mathbb{R}^{O(\log^2 n)}$ such that for any $x, y \in X$,*

$$\left\|f(x) - f(y)\right\|_1 \leq d(x, y) \leq O(\log n) \cdot \left\|f(x) - f(y)\right\|_1$$

Moreover, given d, in $\tilde{O}(n^2)$ time, the mapping f can be computed correctly with probability at least $1 - 1/n$. If (X, d) is a shortest path metric of an m-edge graph, then the running time is $\tilde{O}(m)$ time. (Here $\tilde{O}(\cdot)$ is used to hide polylog *factors.)*

Note that we may ensure that for all $x \in X$, the i-th coordinate $f_i(x) = d(x, y)$ for some $y \in X$, so the coordinate $f_i(x)$ is not *too* big. The original embedding to an ℓ_1-metric was by Bourgain (1985)[2], but the ℓ_1-metric had more dimensions. The bound of $O(\log^2 n)$ dimensions is due to Linial, London, and Rabinovich (1995)[3]. There are many other examples of embedding metrics into more friendly metrics, such as the Johnson-Lindenstrauss dimension reduction and Bartal's tree embedding. In fact, there are entire courses on metric embeddings[4].

We can use Bourgain's embedding to approximate the shortest path distances in a graph G efficiently. For instance, given a graph G on n vertices, Bourgain's theorem allows us to embed the shortest path metric of G into a $\ell_1^{O(\log^2 n)}$-metric with $O(\log n)$ distortion. Then we can label each vertex of G with $O(\log^2 n)$ coordinates so that given vertices $u, v \in V(G)$, we can obtain a $O(\log n)$-approximation of $\mathrm{dist}_G(u, v)$ in $O(\log^2 n)$ time by reading the labels of u and v and

[2]J. Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. Israel J. of Math. 1985.

[3]Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. Combinatorica, 15(2):215–245, 1995. (Preliminary version in 35th FOCS, 1994).

[4]https://home.ttic.edu/ harry/teaching/pdf/lecture1.pdf

performing $O(\log^2 n)$ operations compute the corresponding distance in the $\ell_1^{O(\log^2 n)}$-metric. The total space used is $O(n \log^2 n)$, which is sublinear in the number of edges. However, we will be using Bourgain's embedding in a very different way.

4. Ingredient 2: ℓ_1-Metrics are in the Cut Cone

We saw earlier that a cut metric is a very simple ℓ_1-metric. Now we will see the other direction: every ℓ_1-metric is a combination of cut metrics. To formalize this, we will need some additional notation.

4.1. Observation. *For any n-point metric (X, d), we may think of d as a vector $d \in \mathbb{R}^{\binom{n}{2}}$. Each coordinate in this vector corresponds to a pair of points $x, y \in X$ and has value $d(x, y)$.*

We will refer to the vector $d \in \mathbb{R}^{\binom{n}{2}}$ as an n-point metric, since it encodes everything about (X, d).

4.2. Fact (Closure under addition)**.** *If d_1 and d_2 are n-point metrics, then $d_1 + d_2$ is an n-point metric.*

Proof. We must verify that the three requirements of a metric space hold for the $d_1 + d_2$ metric space, which is defined as $(d_1 + d_2)(x, y) = d_1(x, y) + d_2(x, y)$.

1. $(d_1 + d_2)(x, x) = d_1(x, x) + d_2(x, x) = 0$, where the final equality follows since d_1 and d_2 are metric spaces.
2. $(d_1 + d_2)(x, y) = d_1(x, y) + d_2(x, y) = d_1(y, x) + d_2(y, x) = (d_1 + d_2)(y, x)$, where the final equality follows since d_1 and d_2 are metric spaces.
3. $(d_1 + d_2)(x, z) = d_1(x, z) + d_2(x, z) \leq d_1(x, y) + d_1(y, z) + d_2(x, y) + d_2(y, z) = (d_1 + d_2)(x, y) + (d_1 + d_2)(x, z)$. The inequality follows from the fact that d_1 and d_2 are metric spaces.

□

4.3. Fact (Closure under scalar multiplication)**.** *If d is an n-point metric, then $\alpha \cdot d$ is a metric.*

Proof. We must verify that the three requirements of a metric space hold for the $\alpha \cdot d$ metric space, which is defined as $(\alpha \cdot d)(x, y) = \alpha \cdot d(x, y)$.

$(\alpha \cdot d)(x, y) = \alpha \cdot d(x, y) = 0$, since d is a metric.

$(\alpha \cdot d)(x, y) = \alpha \cdot d(x, y) = \alpha \cdot d(y, x) = (\alpha \cdot d)(y, x)$, since d is a metric.

$(\alpha \cdot d)(x, z) = \alpha \cdot d(x, z) \leq \alpha(d(x, y) + d(y, z)) = (\alpha \cdot d)(x, y) + (\alpha \cdot d)(y, z)$. (The inequality follows from the fact that d is a metric.) □

To summarize, non-negative linear combinations of metrics are metrics.

4.1. The Cone of Cut Metrics

4.4. Notation. Let $\text{CUTCONE}_n = \{d \in \mathbb{R}^{\binom{n}{2}} \mid d = \sum_{S \subseteq V} \alpha_S \cdot d_S \text{ where } \alpha_S \geq 0 \text{ and } d_S \text{ is a cut metric}\}$

CUTCONE_n contains all nonnegative combinations of all cut metrics (i.e. it is a *convex cone* of cut metrics). By Fact 4.2 and Fact 4.3, any $d \in \text{CUTCONE}_n$ is a metric.

4.2. Key Insight: $[0,1]$-Line Metrics are in the Cut Cone

4.5. Definition. We define a $[0,1]$-line metric to be a line metric with distance function $d(x,y) = |f(x) - f(y)|$, where f is a function $f : X \mapsto [0,1]$. That is, f maps only to points in the interval $[0,1]$ on the line.

4.6. Lemma. *Let (X,d) be a $[0,1]$-line metric where $d(x,y) = |f(x) - f(y)|$ for some $f : X \mapsto [0,1]$. Pick a threshold $t \in [0,1]$ uniformly at random. Then define*

$$S_t = \{x \mid f(x) \le t\}$$

Then for every $u, v \in X$,

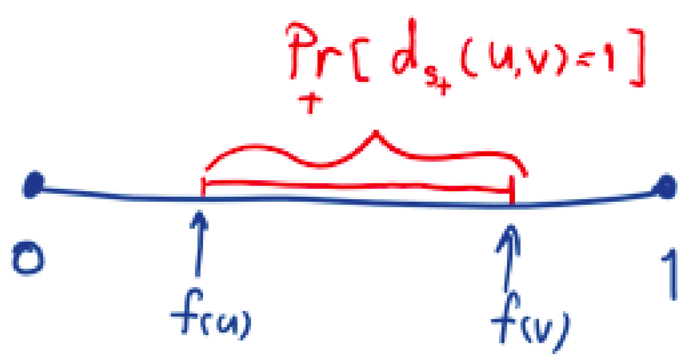
$$\mathbb{E}_t[d_{S_t}(u,v)] = d(u,v).$$

That is, $d = \mathbb{E}_t[d_{S_t}]$ is simply a distribution of cut metrics.

Proof. Note that d_{S_t} is the distance function for the cut metric defined by cut $S_t \subseteq X$. Consider $u, v \in X$, and assume without loss of generality that $f(u) \le f(v)$.

$$\begin{aligned}
\mathbb{E}_t[d_{S_t}(u,v)] &= \Pr_t[d_{S_t}(u,v) = 1] && \text{by the definition of expectation} \\
&= \Pr_t[|S_t \cap \{u,v\}| = 1] && \text{by the definition of the cut metric} \\
&= \Pr_t[f(u) \le t \le f(v)] && \text{by the definition of } S_t \\
&= |f(v) - f(u)| && \text{since } t \text{ is sampled uniformly from } [0,1] \text{ (see figure)} \\
&= d(u,v)
\end{aligned}$$

□



If d is an n−point metric, we can rephrase this result in an equivalent way. Write $X = \{x_1, x_2, \ldots, x_n\}$, where $f(x_i) \le f(x_{i+1})$ (i.e. so the points are ordered lowest to highest). Then for $1 \le i < n$, define $S'_i = \{x_1, \ldots, x_i\}$, and let $\alpha_i = f(x_{i+1}) - f(x_i)$.

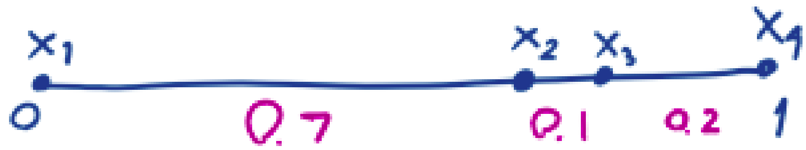
4.7. Observation. *Let d' be the metric*

$$d' = \sum_{i=1}^{n-1} \alpha_i d_{S'_i}$$

then $d' \in$ CUTCONE$_n$.

Proof. Observe that for $1 \le i < n$, we have that $d_{S'_i}$ is a cut metric and $\alpha_i > 0$. Then d' is a nonnegative linear combination of cut metrics and therefore is in $\textsc{CutCone}_n$. □

Example: Suppose our set X was defined as $X = \{x_1, x_2, x_3, x_4\}$ as below.



Then we would define $S'_1 = \{x_1\}$, $S'_2 = \{x_1, x_2\}$, and $S'_3 = \{x_1, x_2, x_3\}$. Additionally, $\alpha_1 = 0.7$, $\alpha_2 = 0.1$, and $\alpha_3 = 0.2$. Then the metric d' would equal $d' = 0.7d_{S'_1} + 0.1d_{S'_2} + 0.2d_{S'_3}$. We will prove that metric d' as defined in Observation 4.7 is identical to the original $[0,1]$-line metric d.

4.8. Claim. $d = d' = \sum_{i=1}^{n-1} \alpha_i d_{S'_i}$

Proof. Observe that for any choice of $t \in [0,1]$, we have that $S_t = S'_i$ for some i. Specifically, $S_t = S'_i$ if and only if $f(x_i) \le t < f(x_{i+1})$. Then we have that:

$$\begin{aligned} d &= \mathbb{E}_t[d_{S_t}] \\ &= \sum_{i=1}^{n-1} \Pr_t[S_t = S'_i] d_{S'_i} \\ &= \sum_{i=1}^{n-1} \Pr_t[f(x_i) \le t < f(x_{i+1})] d_{S'_i} \\ &= \sum_{i=1}^{n-1} \alpha_i d_{S'_i} \\ &= d' \end{aligned}$$

□

Observe that $S'_i \subset S'_{i+1}$ for $1 \le i < n$. Then following corollary is immediate.

4.9. Corollary. *Any $[0,1]$-line metric d of n points is a nonnegative combination of $n-1$* ***nested*** *cut metrics. In particular, $d \in \textsc{CutCone}_n$.*

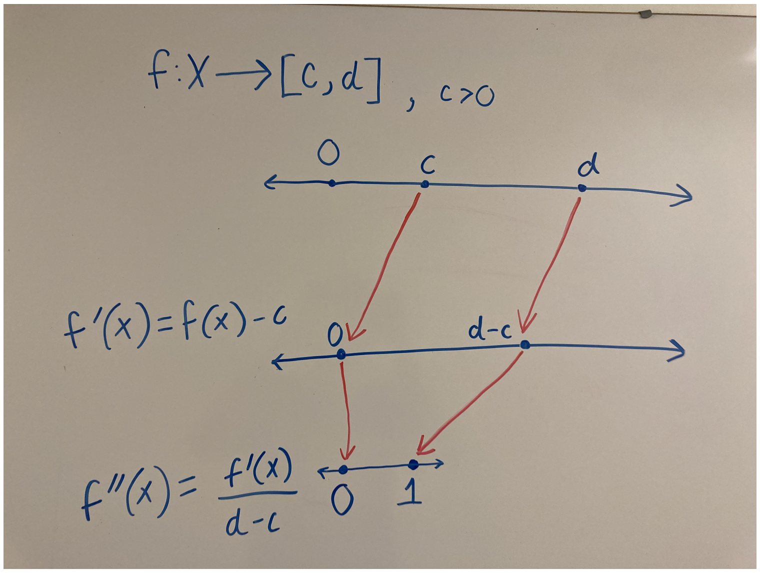
4.3. Generalizing to any Line Metric

Now we extend this result to arbitrary $n-$point line metrics. Consider an $n-$point line metric (X, d) where $d(x,y) = |f(x) - f(y)|$ for some $f : X \mapsto \mathbb{R}$. Without loss of generality, we can assume that $\min_x f(x) = 0$. This is because if $\min_x f(x) = c > 0$, then we may let $f'(x) = f(x) - c$ for all x (translating f), so that $d(x,y) = |f(x) - f(y)| = |f'(x) - f'(y)|$.

It follows that $d(x,y) = |f(x) - f(y)|$ for some function $f : X \mapsto [0, \alpha]$, where $\alpha > 0$. Now if we define $f'(x) = 1/\alpha \cdot f(x)$ for all x (rescaling f), then observe that f' is a function from X to $[0,1]$.

Then the line metric d' defined as $d'(x, y) = |f'(x) - f'(y)|$ is a $[0, 1]$-line metric. Furthermore, $d = \alpha \cdot d'$, so d is a nonnegative combination of a $[0, 1]$-line metrics.

This translation and rescaling can be visualized in the figure below.



Note that since metric d' is a nonnegative combination of $n-1$ nested cut metrics by Corollary 4.9, it follows that metric d is a nonnegative combination of $n-1$ cut metrics as well.

4.10. Corollary. *Any line metric d on n points is a nonnegative combination of $n-1$ nested cut metrics. In particular, $d \in \text{CUTCONE}_n$.*

4.4. Generalizing to any ℓ_1-Metric

Consider any ℓ_1^k-metric (X, d) where $d(x, y) = \|f(x) - f(y)\|_1$ for some $f : X \mapsto \mathbb{R}^k$.

4.11. Observation. *d is a sum of k line metrics.*

Proof. Let $d_i(x, y) = |f_i(x) - f_i(y)|$, where $f_i(x)$ is the ith coordinate of $f(x)$. Note that (X, d_i) is a line metric. Furthermore, by the the definition of an ℓ_1^k-metric, we have that

$$d(x, y) = \|f(x) - f(y)\|_1 = \sum_{i=1}^{k} |f_i(x) - f_i(y)| = \sum_{i=1}^{k} d_i(x, y)$$

Then d is a nonnegative combination of the k line metrics $d_1, \ldots, d_k$. □

Since d is a nonnegative combination of k line metrics, which are in turn each a nonnegative combination of $n-1$ cut metrics by Corollary 4.10, it follows that d is a nonnegative combination of $k(n-1)$ cut metrics.

4.12. Corollary. *Any ℓ_1^k-metric d of n points is a nonnegative combination of $k(n-1)$ cut metrics. In particular, $d \in \text{CUTCONE}_n$.*

Given an ℓ_1-metric and its mapping f, we can efficiently compute its equivalent nonnegative combination of cut metrics specified in Corollary 4.12.

4.13. Notation. Given a mapping f where $d(x,y) = \left\|f(x) - f(y)\right\|_1$, define $\mathcal{S}_f$ as a collection of cuts $S \subseteq X$ such that $d = \sum_{S \in \mathcal{S}_f} \alpha_S \cdot d_S$ where $\alpha_S > 0$ and d_S is a cut metric.

4.14. Corollary. *Given a mapping f of an n-point ℓ_1^k-metric $d(x,y) = \left\|f(x) - f(y)\right\|_1$, we can compute S_f and $\{\alpha_S\}_{S \in \mathcal{S}}$ in polynomial time.*

Proof. In $O(kn)$ time we can obtain the k line metrics $d_1, \ldots, d_k$ of Observation 4.11. Now in $O(kn)$ time we can rescale and translate each of these k line metrics $d_1, \ldots, d_k$ to obtain k $[0,1]$-line metrics $d_1^{01}, \ldots, d_k^{01}$ as in the proof of Corollary 4.10. Finally, for each $[0,1]$-line metric d_i^{01}, we can obtain the corresponding collection of cuts by replicating the procedure in Corollary 4.9. This requires us to sort the n points according to function f_i and obtain $O(n)$ cuts of size $O(n)$ each in $O(n^2)$ time. To compute the coefficients for the collection of cuts it suffices to compute $O(n)$ distances between points, which can be done in $O(n)$ time. Then overall this requires $O(kn^2)$ time. To obtain our final collection $\mathcal{S}_f$ we simply need to undo the scaling we performed to obtain $[0,1]$-line metrics earlier by changing the cut coefficients appropriately for each $[0,1]$-line metric. This can be done in $O(kn)$ time. Then we can compute $\mathcal{S}_f$ and $\{\alpha_S\}_{S \in \mathcal{S}_f}$ in $O(kn^2)$ time. □

4.15. *Remark.* Notice that the bottleneck is actually the size of the output - we can actually compute $\mathcal{S}_f$ in $O(k \cdot n \log n)$ time if the nested cuts in $\mathcal{S}_f$ are represented implicitly.

5. Approximating Conductance via LP Relaxation

5.1. Overview

Our goal is to approximate the conductance $\Phi(G)$ of a graph G. To accomplish this we will first define another notion of expansion $\Phi(G,H)$, which generalizes the notion of conductance (this will make notation easier). Then we will define a linear program (LP) relaxation of $\Phi(G,H)$ denoted by $\mathrm{LR}(G,H)$ such that $\mathrm{LR}(G,H) \leq \Phi(G,H)$. Then $\mathrm{LR}(G,H)$ can be computed in polynomial time since it can be modeled by a LP. We will show that $\mathrm{LR}(G,H)$ is not too small using the ingredients we have seen above.

5.2. A General Notion of Edge Expansion: H-expansion $\Phi(G,H)$

Let $G = (V, E_G, w_G)$ be a weighted undirected graph, and let $H = (V, E_H, w_H)$ be a weighted undirected graph with the same set of vertices as G.

5.1. Definition. The H-expansion of a cut S is defined as

$$\Phi_{G,H}(S) = \frac{w_G(\partial_G S)}{w_H(\partial_H S)}$$

where $w_G(E') = \sum_{e \in E'} w_G(e)$ and $w_H(E') = \sum_{e \in E'} w_H(e)$.

5.2. Definition. The H-expansion of a graph G is defined as [5]

$$\Phi(G,H) = \min_{S: w_H(\partial_H S) \neq 0} \Phi_{G,H}(S)$$

Observe that the H-expansion of a graph G minimizes the H-expansion over all cuts S where this value is well-defined. The lemma below shows that $\Phi(G,H)$ generalizes the $\mathbf{d}$-expansion $\Phi(G,\mathbf{d})$ of a graph.

5.3. Lemma. *Given $G = (V,E)$, let H be a complete weighted graph with $w_H(u,v) = \frac{\boldsymbol{d}(u)\boldsymbol{d}(v)}{\boldsymbol{d}(V)}$ for each u,v. We have $\Phi_{G,H}(S) = \Theta(\Phi_{G,\boldsymbol{d}}(S))$ for all $S \subseteq V$ and so $\Phi(G,H) = \Theta(\Phi(G,\boldsymbol{d}))$.*

Proof. This follows because for any S where $\mathbf{d}(S) \leq \mathbf{d}(V)/2$, we have

$$w_H(\partial_H S) = \sum_{u \in S} \sum_{v \notin S} \frac{\mathbf{d}(u)\mathbf{d}(v)}{\mathbf{d}(V)} = \sum_{u \in S} \mathbf{d}(u) \frac{\mathbf{d}(V \setminus S)}{\mathbf{d}(V)} = \Theta(\mathbf{d}(S))$$

since $\mathbf{d}(V \setminus S) \geq \frac{1}{2}\mathbf{d}(V)$.

It follows that

$$\Phi_{G,H}(S) = \frac{w_G(\partial_G S)}{w_H(\partial_H S)} = \frac{w_G(\partial_G S)}{\Theta(\mathbf{d}(S))} = \Theta(\Phi_{G,\mathbf{d}}(S)).$$

□

Recall that conductance $\Phi(G) = \Phi(G,\mathbf{d})$, where $\mathbf{d}(u) = \deg(u)$. It immediately follows from Lemma 5.3 that we can choose a graph H so that $\Phi(G,H) = \Theta(\Phi(G))$.

5.4. Corollary. *Given $G = (V,E)$, let H be a complete weighted graph where $w_H(u,v) = \frac{\deg(u)\deg(v)}{\mathrm{vol}(V)}$ for each u,v. Then $\Phi(G,H) = \Theta(\Phi(G))$.*

5.3. LP Relaxation of $\Phi(G,H)$

We start with the following observation:

5.5. Observation.

$$w_G(\partial_G S) = \sum_{u,v : |\{u,v\} \cap S| = 1} w_G(u,v) = \sum_{u,v} w_G(u,v) \cdot d_S(u,v)$$

where $d_S : V \times V \mapsto \{0,1\}$ is a cut metric defined by $S \subset V$, and the sum $\sum_{u,v}$ is summing over unordered *pairs of vertices.*

Using this observation, we may rewrite $\Phi(G,H)$ as

$$\Phi(G,H) = \min_{S: w_H(\partial_H S) \neq 0} \frac{w_G(\partial_G S)}{w_H(\partial_H S)} = \min_{d_S : \text{cut metric over } V} \frac{\sum_{u,v} w_G(u,v) \cdot d_S(u,v)}{\sum_{u,v} w_H(u,v) \cdot d_S(u,v)}$$

Then it follows that we may consider computing $\Phi(G,H)$ as an optimization problem minimizing over all cut metrics. Now the crucial step by Leighton and Rao that allows us to obtain a LP approximation is to instead minimize over *all metrics* instead of all cut metrics. We now define this new measure, which we denote $\mathrm{LR}(G,H)$.

[5] $\Phi(G,H)$ is also called the "non-uniform sparsity of G with respect to H".

5.6. Notation.

$$\mathrm{LR}(G,H) = \min_{d:\text{metric over } V} \frac{\sum_{u,v} w_G(u,v)\cdot d(u,v)}{\sum_{u,v} w_H(u,v)\cdot d(u,v)}$$

Now we claim that $\mathrm{LR}(G,H)$ can be modeled as a linear program (LP).

5.7. Claim. $\mathrm{LR}(G,H)$ *can be modeled as a linear program (LP).*

Proof. First, note that by normalizing, we can assume that $\sum_{u,v} w_H(u,v)\cdot d(u,v) = 1$. More concretely, if there exists a metric d' that minimizes $LR(G,H)$, then there exists a metric d'' that can be obtained by scaling d' that minimizes $LR(G,H)$ and further $\sum_{u,v} w_H(u,v)\cdot d''(u,v) = 1$. Minimizing $\sum_{u,v} w_G(u,v)\cdot d(u,v)$ over all metrics d where $\sum_{u,v} w_H(u,v)\cdot d(u,v) = 1$ is captured exactly by the following LP:

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

This follows from the fact that any collection of $\binom{V}{2}$ values $d(u,v)$ that satisfy the above conditions correspond to a $|V|$-point metric (by definition), and so the value of the solution to this LP is exactly $\mathrm{LR}(G,H)$. □

The following is immediate.

5.8. Lemma. *Let λ^* be the optimal value of the above LP. We have $\lambda^* = \mathrm{LR}(G,H)$ and we can compute it in polynomial time.*

Now we need to ensure that $\mathrm{LR}(G,H)$ is a good approximation of $\Phi(G,H)$. It is immediate that $\mathrm{LR}(G,H) \le \Phi(G,H)$ because we consider all metrics including all cut metrics. We will now show that $\mathrm{LR}(G,H)$ cannot be too much smaller than $\Phi(G,H)$.

5.4. Showing that LR(G,H) is a Good Relaxation

We will prove that $\mathrm{LR}(G,H) \ge \frac{1}{\Theta(\log n)}\Phi(G,H)$. In other words, we will show that relaxing cut metrics to all metrics may reduce our desired value by at most a $O(\log n)$ factor. Why might we expect this to be true? At a high level, our ingredients say the following:

1. **Bourgain's embedding:** Any metric (X,d) can be "captured" by an ℓ_1-metric with just $O(\log n)$ distortion.
2. **ℓ_1-metrics are in the cut cone:** any ℓ_1-metric is just a combination of cut metrics.

Taken together, these two facts basically suggest that any metric is captured by a combination of cut metrics within a $O(\log n)$ factor. We will now formally prove our desired inequality.

5.9. Lemma. $\mathrm{LR}(G,H) \ge \frac{1}{\Theta(\log n)}\Phi(G,H)$

Proof. Let d^{OPT} be the optimal metric from the linear program 1. Then we may write $\mathrm{LR}(G,H)$ as

$$\mathrm{LR}(G,H) = \frac{\sum_{u,v} w_G(u,v)\cdot d^{\mathrm{OPT}}(u,v)}{\sum_{u,v} w_H(u,v)\cdot d^{\mathrm{OPT}}(u,v)}$$

Then by applying Bourgain's to d^{OPT}, we get an embedding $f : V \mapsto \mathbb{R}^{O(\log^2 n)}$ where for all $u,v \in V$,

$$\left\|f(u)-f(v)\right\|_1 \le d^{\mathrm{OPT}}(u,v) \le O(\log n)\cdot \left\|f(u)-f(v)\right\|_1$$

Then by applying Corollary 4.14 to f, we have

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

where $\mathcal{S}_f$ is a collection of $O(n\log^2 n)$ cuts. So we have

$$\begin{aligned}
\mathrm{LR}(G,H) &\ge \frac{1}{\Theta(\log n)}\cdot \frac{\sum_{u,v} w_G(u,v)\sum_{S\in\mathcal{S}_f}\alpha_S d_S(u,v)}{\sum_{u,v} w_H(u,v)\sum_{S\in\mathcal{S}_f}\alpha_S d_S(u,v)} \\
&= \frac{1}{\Theta(\log n)}\cdot \frac{\sum_{S\in\mathcal{S}_f}\alpha_S(\sum_{u,v} w_G(u,v)d_S(u,v))}{\sum_{S\in\mathcal{S}_f}\alpha_S(\sum_{u,v} w_H(u,v)d_S(u,v))} \\
&\ge \frac{1}{\Theta(\log n)}\cdot \min_{S\in\mathcal{S}_f}\frac{\sum_{u,v} w_G(u,v)d_S(u,v)}{\sum_{u,v} w_H(u,v)d_S(u,v)} \qquad\qquad \text{as } \frac{\sum_{i=1}^k a_i}{\sum_{i=1}^k b_i} \ge \min_i \frac{a_i}{b_i} \\
&= \frac{1}{\Theta(\log n)}\cdot \min_{S\in\mathcal{S}_f}\frac{w_G(\partial_G S)}{w_H(\partial_H S)} \\
&= \frac{1}{\Theta(\log n)}\cdot \min_{S\in\mathcal{S}_f}\Phi_{G,H}(S) \\
&\ge \frac{1}{\Theta(\log n)}\cdot \Phi(G,H)
\end{aligned}$$

Therefore we can conclude

$$\mathrm{LR}(G,H) \le \Phi(G,H) \le O(\log n)\cdot \mathrm{LR}(G,H)$$

□

The following theorem is immediate.

5.10. Theorem. *There is a randomized algorithm that $O(\log n)$-approximates $\Phi(G,H)$ in polynomial time. The same holds for $\Phi(G)$.*

5.11. *Remark.* In fact, the only slow step in the above algorithm is to obtain d^{OPT}. All other steps take near linear time.

5.12. Exercise. Given an embedding f of the optimal metric d^{OPT} of the LP, show how to compute a cut S such that $\Phi_{G,H}(S) \le O(\log n)\cdot\Phi(G,H)$ in $\tilde{O}(|E(G)|+|E(H)|)$ time. In the case where we want to compute conductance, the running time is just $\tilde{O}(|E(G)|)$.

6. Limitation and Outlook

This approach gives a $O(\log n)$ approximation to $\Phi(G, H)$, which is tight (you will prove this in the homework). We can obtain better approximations via smaller distortion metric embeddings. Bourgain's embedding is very general in that it can embed any metric into an ℓ_1-metric. However, we can obtain better embeddings by considering a more specific set of metrics, like the set of *negative-type* metrics.

6.1. Definition. A *negative-type* metric (X, d) is such that there exists $f : X \mapsto \mathbb{R}^k$ such that $d(x, y) = \|f(x) - f(y)\|_2^2$.

Any n-point negative-type metric can be embedded into an ℓ_1-metric with distortion $O(\sqrt{\log n} \log\log n)$ [6]. This embedding is similar to Bourgain's, but with better distortion on a smaller class of metrics. A relaxation of $\Phi(G, H)$ to all negative-type metrics can be computed using semidefinite programming (SDP). Following the same argument as the one in this lecture (but more complicated), we get an $O(\sqrt{\log n} \log\log n)$-approximation algorithm for $\Phi(G, H)$. This topic is further exmplored in a survey by Naor on connections between expansion and functional analysis[7].

This approach works for a general notion of edge expansion in undirected graphs. Metric embeddings have also been studied in the context of vertex expansion[8]. Later, we will see a more robust framework based on the "cut-matching game" which works for both **vertex expansion** and **directed graphs**.

7. Exercises

The shortest path metric is universal in the following sense:

7.1. Exercise. Show that any (finite) metric (X, d) is a shortest path metric of a weighted graph $G = (X, E, w)$ where $w(u, v) = d(u, v)$.

7.2. Exercise (More examples of metrics)**.** Convince yourself that the following are metrics:

1. **(ℓ_p metric):** $X = \mathbb{R}^m$ and $d(x, y) = \|x - y\|_p = (\sum_i (x_i - y_i)^p)^{1/p}$.
   - For $p = 2$, $\|x - y\|_2 = \sqrt{\sum_i (x_i - y_i)^2}$ is just a Euclidean distance.
   - For $p = \infty$, $\|x - y\|_\infty = \max_i\{|x_i - y_i|\}$
2. **(uniform metric):** Let X be any set and $d(x, y) = 1$ for all $x \neq y$.
3. **(1-2 metric):** Let X be any set and $d(x, y) \in \{1, 2\}$ for all $x \neq y$.
4. **(edit distance):** Let $X = \{0, 1\}^*$ be a set of strings. Let $d(x, y)$ be the edit distance between x and y.

The earlier discussion about ℓ_1-metrics and cut metrics gives the following characterization:

7.3. Exercise. Let $L1_n$ contain all n-point ℓ_1-metrics. Prove that $L1_n = \textsc{CutCone}_n$.

[6] Sanjeev Arora, James Lee, and Assaf Naor. Euclidean distortion and the sparsest cut

[7] Assaf Naor. L1 embeddings of the heisenberg group and fast estimation of graph isoperimetry, 2010.

[8] https://homes.cs.washington.edu/ jrl/papers/pdf/fhl-sicomp.pdf