Bourgain's Embedding Theorem, Johnson-Lindenstrauss Lemma, and the Sparsest Cut Problem

Richard Ye

June 2023

Abstract

This paper will go over two major theorems in the field of metric dimension reduction, the Bourgain Embedding Theorem and the Johnson-Lindenstrauss Theorem, and their respective proofs. In addition, we will examine some applications of metric dimension reduction, and these theorems in particular, to the Sparsest Cut algorithm problem.

1 Introduction

Metric embeddings and metric dimension reduction are critical tools in computer science and mathematics and can be useful in tackling the curse of dimensionality. These techniques allow for the transformation of higher dimensional metric spaces into lower ones without significantly altering the distances between points.

Two major results in this field that we will examine in this paper are the Bourgain embedding theorem and the Johnson-Lindenstrauss Lemma. The Bourgain embedding theorem provides a method to embed any finite metric space into Hilbert space with minimal distortion. The Johnson-Lindenstrauss Lemma shows that high-dimensional data can be projected into a lower-dimensional space while nearly maintaining pairwise distances.

In addition, we will examine some applications of metric embedding, and these theorems in particular, to the Sparsest Cut algorithm problem, an NP-hard problem in combinatorial optimization.

Definition 1.1: *Bi-Lipschitz Embedding*
We say a metric space (M, d_M) embeds with distortion α, $\alpha \geq 1$, into metric space (N, d_N) if there is a embedding given by a mapping, $f : M \to N$, and

scaling factor τ, $\tau > 0$, such that for every $a, b \in M$,

$$\tau d_M(a,b) \leq d_N(f(a), f(b)) \leq \alpha\tau d_M(a,b)$$

Denote $c_N(M)$ the infimum of $\alpha \in [1, \infty]$ for which α is the distortion of an embedding of (M, d_M) into (N, d_N). When N is a L_p space, then denote $c_{L_p}(M) = c_p(M)$.

Definition 1.2:
$c_2(M)$ is called *Euclidean Distortion*, since it measures how close to being a subset of Euclidean space M is.

2 Bourgain's Embedding Theorem

Theorem 2.1: Bourgain's Embedding Theorem(General Case)

Consider (X, d) be a metric space with n points. Then $c_2(X) \lesssim log(n)$. i.e. $c_2(X) = O(log(n))$

So this theorem implies that the Euclidean distortion of a n-point metric space is $O(log(n))$.

This theorem is particularly interesting because it gives the result that we can represent arbitrary metric space into a nice normed space(such as Euclidean), with only a logarithmic distortion.

Proof:

Outline: For this proof, we will be following the argument of [2]. So the idea is to construct a function, f, for any metric space, X, such that f embeds from X to a space with Euclidean norm, and then show that embedding f has distortion of $O(\log n)$

To start, consider we take a random subset of X, B, by independently adjoining elements in X each with probability $1/2^j$ to get B. Then for every subset $A \subseteq X$, we can denote

$$\mathbf{P}(B = A) = \pi_j(A) = \frac{1}{2^{j \times |A|}}(1 - 1/2^j)^{n-|A|}$$

(The intuition for $\pi_j(A)$ is chopping the metric space up at dyadic scales)
Now consider $k = \lfloor log_2 n \rfloor + 1$. We can define function $f : X \to \mathbf{R}^{2^X}$ (note that $\mathbf{R}^{2^X}$ is equipped with euclidean norm).

$$f(x) = (f(x)_A)_{A \subseteq X}$$

Where $f(x)_A = (\frac{1}{k}\sum_{j=1}^{k} \pi_j(A))^{1/2} d(x, A)$.
(Note $d(x, A)$ is defined as the distance between x and the closest element of A).

Now we just want to show that embedding f has bi-lipschitz distortion of $O(logn)$.

To show the RHS(upper bound), it follows that for any $a, b \in X$,

$$||f(a) - f(b)||_2^2 = \sum_{A \subseteq X} \frac{1}{k} \sum_{j=1}^{k} \pi_j(A)(d(a, A) - d(b, A))^2$$

$$\leq \sum_{A \subseteq X} \sum_{j=1}^{k} \frac{1}{k} \pi_j(A)(d(a, b))^2$$

since $x \to d(x, A)$ is 1-Lipschitz, and since $\{\pi_j(A)\}_{A \subseteq X}$ sums to 1 $\forall j$(i.e. probabilities sum to 1), thus we have

$$= \frac{1}{k} kd(a, b)^2 = d(a, b)^2$$

so we get that

$$||f(a) - f(b)||_2^2 \leq d(a, b)^2$$

and we are done with the upper bound(RHS).

Now for the lower bound(LHS), again take points $a, b \in X$, and now take some $j \in [1, k]$.

Let's define $r_j(x, y)$ as the smallest positive r such that for balls $B(a, r)$ and $B(b, r)$, we have that

$$|B(a, r)| \geq 2^j$$

$$|B(b, r)| \geq 2^j$$

Now we need this following lemma:

Lemma 2.2:
Let us denote

$$\tilde{r}_j(a, b) = min\{r_j(a, b), \frac{1}{3} d(a, b)\}$$

Consider any random subset of X, A_j, that we get with respect to π_j. Then, $\forall a \neq b \in X$, we have that

$$\mathbf{E}_{\pi_j}(d(a, A_j) - d(b, A_j))^2 \gtrsim (\tilde{r}_j(a, b) - \tilde{r}_{j-1}(a, b))^2$$

Proof of Lemma 2.2:

So we want to show(probability with respect to measure π_j) that

$$\mathbf{P}(|d(a, A_j) - d(b, A_j)| \geq \tilde{r}_j(a, b) - \tilde{r}_{j-1}(a, b)) \gtrsim 1$$

Then the lemma follows by the Markov inequality. Assume $\tilde{r}_j(a,b) > \tilde{r}_{j-1}(a,b)$, since otherwise it is trivial. Thus, by definition, it implies that $\tilde{r}_{j-1}(a,b) = r_{j-1}(a,b)$. Now again by definition of $r_j(a,b)$, we know that

$$|B(a,\tilde{r}_{j-1}(a,b))| \geq 2^{j-1} \text{ and } |B(b,\tilde{r}_{j-1}(a,b))| \geq 2^{j-1}$$

Now from this statement, if we consider open balls, we can get that,

$$|B^o(a,r_j(a,b))| > 2^j \text{ and } |B^o(b,r_j(a,b))| > 2^j$$

Note that by definition of $\tilde{r}_j(a,b)$, we know that the open and closed balls, $B^o(a,r_j(a,b))$ and $B(b,\tilde{r}_{j-1}(a,b))$ are disjoint.

Now let us take a random subset A as we illustrated previously.

Notice that if $A \cap B^o(a,r_j(a,b)) = \varnothing$ and $A \cap B(b,\tilde{r}_{j-1}(a,b)) \neq \varnothing$, we have

$$|d(a,A) - d(b,A)| > \tilde{r}_j(a,b) - \tilde{r}_{j-1}(a,b)$$

Now,

$$\mathbf{P}(|d(a,A) - d(b,A)| \geq \tilde{r}_j(a,b) - \tilde{r}_{j-1}(a,b))$$

From what we *noticed* previously, and that the balls are disjoint, we get the previous quantity is

$$\geq \mathbf{P}(A \cap B^o(a,\tilde{r}_j(a,b) = \varnothing) \text{ AND } A \cap B^o(b,\tilde{r}_{j-1}(a,b)) = \varnothing)$$

$$= \mathbf{P}(A \cap B^o(a,\tilde{r}_j(a,b) = \varnothing))\mathbf{P}(A \cap B^o(b,\tilde{r}_{j-1}(a,b)) = \varnothing))$$

Where we used properties of independence for the last equality.

Applying the values from definitions, we get

$$(1 - \frac{1}{2^j})^{|B^o(x,\tilde{r}_j(a,b))|}(1 - (1 - \frac{1}{2^j})^{|B^o(x,\tilde{r}_{j-1}(a,b))|})$$

by definition,

$$\geq (1 - \frac{1}{2^j})^{2^j}(1 - (1 - \frac{1}{2^j})^{2^j}) \gtrsim 1$$

So we have shown that

$$\mathbf{P}(|d(a,A_j) - d(b,A_j)| \geq \tilde{r}_j(a,b) - \tilde{r}_{j-1}(a,b)) \gtrsim 1$$

Now Markov's inequality states that

$$\mathbf{P}(X \geq c) \leq \frac{\mathbf{E}(X)}{c}$$

Thus, we directly get

$$\mathbf{E}(d(a,A_j) - d(b,A_j)) \gtrsim (\tilde{r}_j(a,b) - \tilde{r}_{j-1}(a,b))$$

i.e.

$$\mathbf{E}_{\pi_j}(d(a, A_j) - d(b, A_j))^2 \gtrsim (\tilde{r}_j(a, b) - \tilde{r}_{j-1}(a, b))^2$$

□

Now we can finish the LHS proof of the Bourgain embedding theorem by applying this lemma, so

$$||f(a) - f(b)||_2^2 = \sum_{A \subseteq X} \frac{1}{k} \sum_{j=1}^{k} \pi_j(A)(d(a, A) - d(b, A))^2$$

$$= \frac{1}{k} \sum_{j=1}^{k} \mathbf{E}_{\pi_j}(d(a, A_j) - d(b, A_j))^2$$

by the lemma,

$$\gtrsim \frac{1}{k} \sum_{j=1}^{k} (\tilde{r}_j(a, b) - \tilde{r}_{j-1}(a, b))^2$$

by Jensen's inequality,

$$\geq \frac{1}{k^2} (\sum_{j=1}^{k} (\tilde{r}_j(a, b) - \tilde{r}_{j-1}(a, b)))^2$$

$$= \frac{1}{k^2} \tilde{r}_k(a, b)^2$$

$$\simeq \frac{1}{k^2} d(a, b)^2$$

(Note that $\simeq$ means is equal up to a constant, i.e. $= O(\frac{1}{k^2} d(a, b)^2)$)

Plugging in k we can just rewrite that as:

$$||f(a) - f(b)||_2 \gtrsim \frac{1}{log(n)} d(a, b)$$

And we have proved the LHS.

Thus by definition of bi-lipschitz embedding, we proved that f has bi-lipschitz distortion of at most $O(log(n))$. □.

This is for the case of embedding into L_2. In addition, we have the general case:

Theorem 2.3: For any n-point metric space (X, d) and any $1 \leq p < \infty$, then

$$c_p(X) \lesssim \frac{log(n)}{p}$$

Proof Idea:
We can Use the same argument as proof for Theorem 2.1, however, instead of using $1/2^j$ when defining $\pi_j(A)$, use q^j, with $q \in (0, 1)$, and optimize with respect to q.

3 Johnson-Lindenstrass Lemma

Theorem 3.1: JL Lemma

Let any $0 < \epsilon < 1$, and a n-point set, X, with elements $\in \mathbf{R}^d$, suppose

$$k \geq 4(\epsilon^2/2 - \epsilon^3/3)^{-1} log(n)$$

Then there is an embedding $f : \mathbf{R}^d \to \mathbf{R}^k$ such that for any $a, b \in X$, we have

$$(1-\epsilon)||a-b||^2 \leq ||f(a) - f(b)||^2 \leq (1+\epsilon)||a-b||^2$$

In addition, this map/embedding can be found with a **polynomial runtime** algorithm.

Basically, this lemma states one can find, in polynomial time, an embedding such that a set of points in a higher dimensional space can be mapped to a lower dimensional space with the distances between points nearly preserved(up to bi-lipschitz constant).

Proof:

Outline: For this proof we will be following the argument of [4]. So the idea is that we show that the squared length of a random vector is tightly concentrated about its mean when it is projected to a lower dimensional subspace. In particular, the length of the random vector is not distorted by more than $(1 \pm \epsilon)$, with probability $O(1/n^2)$, which means that there is a $\geq O(1/n)$ probability that the embedding procedure gives us the properties that we want. Then we can just repeat the procedure a polynomial number of times to get the desired constant probability of success of finding the embedding with desired properties.

Consider $X_1, X_2, ..., X_d$ i.i.d. $N(0, 1)$ random variables and $Y = 1/||X||(X_1, ..., X_d) \in R^d$, i.e. Y is d-dimensional unit random vector constructed from X_i's. Let $Z \in R^k$ be the projection of Y onto its first k dimensions, now it is clear that the expected squared norm of Z is $\mu = \mathbf{E}[||Z||^2] = k/d$

Now we have this following lemma:

Lemma 3.2:

If $k < d$, then

1. If $\beta > 1$,

$$\mathbf{P}(L \leq \frac{\beta k}{d}) \leq \beta^{k/2}(1 + \frac{(1-\beta)k}{d-k})^{(d-k)/2} \leq exp(\frac{k}{2}(1 - \beta + ln\beta))$$

2. If $\beta < 1$,

$$\mathbf{P}(L \geq \frac{\beta k}{d}) \leq \beta^{k/2}(1 + \frac{(1-\beta)k}{d-k})^{(d-k)/2} \leq exp(\frac{k}{2}(1 - \beta + ln\beta))$$

This lemma basically states that L is probabilistically tightly concentrated about μ.

Proof of lemma 3.2:
Another way of stating Lemma 3.2 is

$$\mathbf{P}[d(X_1^2 + ... + X_k^2) \leq k\beta(X_1^2 + ... + X_d^2)] \leq \beta^{k/2}(1 + \frac{k(1-\beta)}{d-k})^{(d-k)/2}$$

Let's first focus on part 1,
To show this,

$$\mathbf{P}[d(X_1^2 + ... + X_k^2) \leq k\beta(X_1^2 + ... + X_d^2)] = \mathbf{P}[k\beta(X_1^2 + ... + X_d^2) - d(X_1^2 + ... + X_k^2) \geq 0]$$

Now for $t > 0$, this is equal to

$$= \mathbf{P}[exp(t(k\beta(X_1^2 + ... + X_d^2) - d(X_1^2 + ... + X_k^2))) \geq 1]$$

Applying the Markov inequality,

$$\leq \mathbf{E}[exp(t(k\beta(X_1^2 + ... + X_d^2) - d(X_1^2 + ... + X_k^2)))]$$

Since X_i are i.i.d. $N(0,1)$, and using the easily proved fact that $\mathbf{E}(e^{sX^2}) = 1/\sqrt{1-2s}, -\infty < s < 1/2$ when X is such, we have

$$\mathbf{E}[exp(tk\beta^2 X)]^{d-k}\mathbf{E}[exp(t(k\beta - d)X)^2]^k$$

$$= (1 - 2tk\beta)^{-(d-k)/2}(1 - 2t(k\beta - t))^{-k/2} = g(t)$$

Which we will denote as $g(t)$. (Note: this means we have constraints $tk\beta < 1/2$ and $t(k\beta - d) < 1/2$, i.e. $0 < t < 1/2k\beta$)
Now, to minimize $g(t)$ is the same as maximizing

$$(1 - 2tk\beta)^{(d-k)}(1 - 2t(k\beta - t))^k = f(t)$$

Which we will denote $f(t)$.
Now we can just take first order conditions, we get that the minimum is at

$$t_0 = \frac{1-\beta}{2\beta(d-k\beta)}$$

Since $g(t_0) = 1/\sqrt{f(t_0)}$, thus

$$g(t_0) = f(t_0)^{-1/2} = [(\frac{d-k}{d-k\beta})^{d-k}(\frac{1}{B})^k]^{-1/2}$$

$$= \beta^{k/2}(\frac{d-k\beta}{d-k})^{(d-k)/2}$$

$$= \beta^{k/2}(1 + \frac{k(1-\beta)}{d-k})^{(d-k)/2}$$

Thus we show the inequality for Lemma 3.2(1).

Now similarly for lemma 3.2(2), we get that

$$\mathbf{P}[d(X_1^2+...+X_k^2) \le k\beta(X_1^2+...+X_d^2)] \le (1+2tk\beta)^{-(d-k)/2}(1+2t(k\beta-t))^{-k/2} = g(-t)$$

(Note: we get constraints $0 < t < 1/2(d-k\beta)$)
Denote upper bound $g(-t)$. Clearly, this expression is minimized at $-t_0$, where t_0 is the same as defined in the proof for 3.2(1).

Now similarly, plug in t_0 from above and we get that

$$\mathbf{P}[d(X_1^2+...+X_k^2) \le k\beta(X_1^2+...+X_d^2)] \le \beta^{k/2}(\frac{d-k\beta}{d-k})^{(d-k)/2}$$

$$= \beta^{k/2}(1 + \frac{k(1-\beta)}{d-k})^{(d-k)/2}$$

Thus we have proven the statements for both parts of the lemma 3.2.

$\square$

Now back to the proof of Theorem 3.1,

Clearly, if $d \le k$ then it is trivial, so let's assume $d > k$. Now take k dimensional subspace S at random. Consider v_i' be the projection of point $v_i \in V$ onto S.

Now consider $L = ||v_i' - v_j'||^2$ and $\mu = (k/d)||v_i - v_j||^2$.

Now knowing the lemma, we apply 3.2(1) to get,

$$\mathbf{P}(L \le (1-\epsilon)\mu) \le exp(\frac{k}{2}(1-(1-\epsilon)) + \log(1-\epsilon))$$

And since $\log(1-a) \leq -a - a^2/2$ for any $0 \leq a < 1$, so

$$\leq exp(\frac{k}{2}(\epsilon - (\epsilon + \epsilon^2/2)))$$

$$= exp(-\frac{k\epsilon^2}{4})$$

$$\leq exp(-2\log n) = 1/n^2$$

Now applying 3.2(2),

$$\mathbf{P}(L \geq (1+\epsilon)\mu) \leq exp(\frac{k}{2}(1-(1+\epsilon)) + \log(1+\epsilon))$$

And since $\log(1+a) \leq a - a^2/2 + a^3/3$, so

$$\leq exp(\frac{k}{2}(-\epsilon + (\epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3})))$$

$$= exp(-\frac{k(\epsilon^2/2 - \epsilon^3/3)}{2})$$

$$\leq exp(-2\log n) = \frac{1}{n^2}$$

Now given these conclusions, consider map $f(v_i) = (\sqrt{(d/k)})v'_i$. Now for an arbitrary fixed pair, v_i, v_j, applying the above derivation, we know that the probability that the distortion, $||f(v_i) - f(v_j)||^2/||v_i - v_j||^2$ not in $[1-\epsilon, 1+\epsilon]$ is at most equal to $2/n^2$ i.e.

$$\mathbf{P}(||f(v_i) - (v_j)||^2/||v_i - v_j||^2 \notin [1-\epsilon, 1+\epsilon]) \leq 1/n^2 + 1/n^2 = 2/n^2$$

Thus, the probability that any pair of points has a large distortion is bounded by $\binom{n}{2} \times 2/n^2 = 1 - 1/n$. So the embedding f has the desired properties with probability $\geq 1/n$. Now if we repeat this projection procedure $O(n)$ times, then we can find an embedding that satisfies the desired properties. Thus we have the randomized polynomial run-time algorithm that we wanted.

$\square$.

4 Sparsest Cut Problem

As an application of the previously discussed theorems and the topic of metric embeddings, we will examine the Sparsest Cut algorithms problem.

Informally, the goal of this problem is to partition a weighted undirected graph into two large-as-possible sets while keeping as many total edges as possible(minimizing number of edges removed or ”cut” by the partition).

The context for this problem is:

- There is a weighted, undirected graph, $G = (V, E)$, with set of vertices, V, and set of edges, E(assigned with positive edge weights/capacities).
- For every pair of vertices, $s_i, t_i \in V$, in the graph, they are assigned some non-negative demand value, D_i.

Definition 4.1a : A *cut*, $(S, \bar{S})$, is a partitioning of the nodes, V, of the graph, G, into two sets, S and $\bar{S}$.

Definition 4.1b : The *capacity*, c_e, of an edge, $e \in E$, is the weight that is assigned to that edge.

Definition 4.1c : The *demand*, D_i, between two nodes, s_i, t_i, is a value assigned to every one of the $\binom{|V|}{2}$ pairs of vertices in V.

Definition 4.2 : The *sparsity* of a cut is

$$\phi(S) = \frac{C(S, \bar{S})}{D(S, \bar{S})}$$

Where

$$C(S, \bar{S}) = |E(S, \bar{S})| = \sum_{e\text{'s that cross the cut}} c_e$$

And

$$D(S, \bar{S}) = \sum_{i\text{'s s.t. } s_i, t_i \text{ are separated by the cut}} D_i$$

So the goal of the problem is to find the cut, $(S, \bar{S})$, with the minimum sparsity as defined above.

Let us denote the minimum sparsity, $\min_{S \subseteq V} \Phi(S) = \Phi(S^*) = \Phi^*$.

Note: A special case, what is called the "uniform" case, of this problem, say we have unit demands between all vertices, then our problem can be rewritten as minimizing

$$\min \frac{C(S, \bar{S})}{|S||\bar{S}|}$$

where $|S|, |\bar{S}|$ are just the size of the sets. So we would want to minimize the sum capacity of the edges cut by $(S, \bar{S})$, while maximizing the $|S||\bar{S}|$. i.e. partition the graph such that it minimizes the edges lost, while maximizing the size of the pieces. This gives a rough intuition of what this problem is trying to achieve.

We can note that this is known to be an NP-hard problem.

At a glance, this problem is not related to metrics, so we now want to reformulate this problem to be finding the minimum over what are called cut metrics.

Definition 4.3: For a given cut, $S \subseteq V$, a *Cut metric*, δ_S, associated with S, is

$$\delta_S(u,v) = \left\{ \begin{array}{ll} 0, & \text{if } u,v \in S \text{ or } u,v \in \bar{S} \\ 1, & \text{otherwise} \end{array} \right\}$$

(Note this is not an actual metric, since metrics of different points cannot be 0).

Now, we can try to view any n-point graph as a vector in $\mathbf{R}^{\binom{n}{2}}$, so each coordinate corresponds to a pair of vertices in the the graph(for our purposes, for pairs of vertices with no edges between them can just take value 0).

Notationally, for a metric d from cut S, the corresponding vector is $\bar{d} \in R^{\binom{n}{2}}$. With this, we have

1. $\alpha\bar{d} + (1-\alpha)\bar{d}$, $\alpha \in [0,1]$ is a metric
2. $k\bar{d}$ is a metric.

(i.e. the set of all metrics forms a convex cone $\in \mathbf{R}^{\binom{n}{2}}$).
Now we can rewrite the sparsest cut problem from before as

$$\min_{\text{all cut metrics S}} \frac{\bar{c} \cdot \bar{\delta}_S}{\bar{D} \cdot \bar{\delta}_S}$$

Where δ_S is the cut metric for cut S, and $\bar{\delta}_S$ is its corresponding vector in $\mathbf{R}^{\binom{n}{2}}$. So, for example if vertex i, j is separated by cut S, then their corresponding cooridnate in $\mathbf{R}^{\binom{n}{2}}$ would have value 1, and 0 otherwise. And $\bar{c}$ is a vector in $\mathbf{R}^{\binom{\mathbf{n}}{\mathbf{2}}}$, and $\bar{c}_{ij}$ is just the capacity of the edge between vertex i, j, and $\bar{D}_{ij}$ is similarly just the demand between vertex i, j.

Now denote

$$CUT_n = \{\bar{d} | d = \sum_{S \subseteq V} \alpha_s \delta_s, \alpha \geq 0\}$$

i.e. convex combination sum of all cut metrics, is the positive cone formed by all of the cut metrics.

In addition,

Claim 4.4: CUT_n is in *one-to-one correspondence* to the n-point subsets of $\mathbf{R}^t$ where each subset is equipped with the l_1 norm, and t is the total number of possible cuts in the n-point graph.

In other words:
Any metric in l_1 can be represented by a positive linear combination of cuts.

Proof:

- $\subseteq$ direction: Consider any metric in CUT_n, for every cut, S, $\alpha_S > 0$, we have a dimension, in which has value 0 for $x \in S$ and value α_S for $x \in \bar{S}$. Thus, this shows $CUT_n \subseteq l_1$.
- Other direction: So the idea is to look at each coordinate, Consider taking one dimension, d, from a n-point set in R^n, and order the points in increasing value in dimension d. Say WLOG we get $v_1, ..., v_k$ distinct values. Then we can define $k-1$ cut metrics $S_i = \{x | x_d \leq v_{i+1}\}$, and consider $\alpha_i = v_{i+1} - v_i$. Then along dimension d, $|x_d - y_d| = \sum_{i=1}^{k} \alpha_i \delta_{S_i}$. And we can construct cut metrics in this way for every dimension. i.e. there is a metric in CUT_n for every n-point metric in l_1, so $l_1 \subseteq CUT_n$

Note, for algorithmic implementation, we have the **lemma 4.5:**
Given metric $\mu \in l_1$ of dimension D, then there is a poly(n,d) time procedure that outputs a set of $\alpha_S \geq 0$, $S \subseteq V$ s.t. $\mu = \sum_{S \subseteq V} \alpha_s \delta_s$

Now, since the optimum of the sparsest cut problem has to be achieved at an extreme point, we thus can rewrite our problem as,

$$\Phi* = \min_{d \in CUT_n} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}}$$

And using **Claim 4.4,** we can rewrite as optimizing for $\phi*$ over CUT_n,

$$\Phi* = \min_{d \in l_1} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}}$$

However, since the sparsest cut problem is NP-hard, it is impossible to solve this problem over the metrics in l_1.

Now let's consider a relaxation.

We can relax the problem by changing the domain of d to all metrics. Thus we get

$$\lambda^* := \min_{d \text{ a metric}} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}}$$

It is clear that $\lambda^* \leq \Phi^*$ since $l_1 \subseteq \{\text{set of all metrics}\}$.

Note that now, we can compute λ^* using a linear program(LP):

$$\min c_{ij} d_{ij}$$

subject to the following:

$$d_{ij} \leq d_{ik} + d_{jk}$$

$$D_{ij} d_{ij} = 1$$

$$d_{ij} \geq 0$$

Assuming that the LP finds the metric d that minimizes λ^*, now we need a way to find the cut (S, S') such that $\Phi(S) \approx \Phi^*$.

The idea is to embed the metric we get from the LP into a l_1 metric a way that the distances are not changed too significantly. Then, we can recover a cut metric from the l_1 metric with the same optimizing objective, and we are done, and we get what we want, up to what we lose from the distortion.

First let us define,
Definition 4.6:
The *integrality gap* is the ratio between the optimal value we obtain from solving the LP relaxation and the optimal value of the original problem.

Now, applying the metric embedding knowledge that we discussed in previous sections, we have that:

Theorem 4.7:
For any metric space (V, d) equipped with metric d, $\exists$ a metric $\mu = \mu(d) \in l_1$ such that $\forall x, y \in V$,

$$d(x, y) \leq \mu(x, y) \leq \alpha d(x, y)$$

Then the integrality gap for the sparsest cut linear program(LP) from before is α.

Proof:
Suppose the LP returns metric d, and consider the metric $\mu \in l_1$ such that $d \leq \mu \leq \alpha d$.
Applying this, we have $\bar{c} \cdot \bar{\mu} \leq \alpha \bar{c} \cdot \bar{d}$ and $\bar{D} \cdot \bar{\mu} \geq \bar{D} \cdot \bar{d}$

Then, we have

$$\Phi(\mu) := \frac{\bar{c} \cdot \bar{\mu}}{\bar{D} \cdot \bar{\mu}} \leq \frac{\alpha \bar{c} \cdot \bar{d}}{\bar{D} \cdot \bar{d}} = \alpha \frac{\bar{c} \cdot \bar{d}}{\bar{D} \cdot \bar{d}} = \alpha \lambda * \leq \alpha \Phi *$$

i.e. The LP finds solution with value at most α times the optimal sparsest cut, Φ^*, in other words, integrality gap of the LP is upper bounded by/at most α.

Now, applying the *Bourgain Embedding Theorem* that we examined previously, and a result from Linial et al.[3], we can directly get

Theorem 4.8:
For all metrics d, $\exists$ $\mu \in l_1$ s.t. integrality gap $\alpha = O(\log n)$. Moreover, the number of dimensions needed is at most $O(\log^2 n)$[3].

And following 4.7 and 4.8, we have

Corollary 4.9: The LP for the relaxation of the sparsest cut has integrality gap of $O(\log n)$.

So after all of this, we have shown that the integrality gap between Φ^* and λ^* are small by applying metric embeddings and the theorems examined previously. But, we want to find the sparsest cut, i.e. we want to find a cut, $(S, \bar{S})$ such that $\Phi(S)$ is minimized.

To find this, we need to construct a cut, $(S, \bar{S})$, from $\mu \in l_1$ metric such that,

$$\Phi(S) \leq \Phi(\mu)$$

So we have,

$$\Phi(\mu) = \frac{\bar{c} \cdot \bar{\mu}}{\bar{D} \cdot \bar{\mu}}$$

And by **Claim 4.4**, we know any metric in l_1 can be represented by positive linear combination of cuts, so

$$= \frac{\bar{c} \cdot \sum \alpha_S \delta_S}{\bar{D} \cdot \sum \alpha_S \delta_S}$$

$$= \frac{\sum \alpha_S (\bar{c} \cdot \delta_S)}{\sum \alpha_S (\bar{D} \cdot \delta_S)}$$

$$\geq \min_{S, \alpha_S > 0} \frac{\alpha_S (\bar{c} \cdot \delta_S)}{\alpha_S (\bar{D} \cdot \delta_S)}$$

$$= \min_{S, \alpha_S > 0} \Phi(S)$$

i.e. This means we can just pick the best cut S among the ones with $\alpha_S > 0$ in the cut decomposition of μ.

And now, using **Lemma 4.5** and **Theorem 4.8**, we can find the positive linear combination representation of μ in polynomial runtime, *poly*(n), and with at maximum $O(n \log^2 n)$ cuts.

Finally, these not only show that the integrality gap is small(i.e. **Corollary 4.9**), but also,

Theorem 4.10:
Given the solution to the Linear Program(LP), d, which solves for the minimum LP-value, λ^*, we can efficiently, in polynomial time, find a cut, $(S, \bar{S})$, such that

$$\Phi(S) \leq O(\log n) \times \lambda^*$$

□

In summary, we wanted to solve the sparsest cut problem, which we reformulated using metric spaces as solving

$$\Phi^* = \min_{d \in l_1} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}}$$

Which since is NP-hard, we then relaxed to solving

$$\lambda^* := \min_{d \text{ a metric}} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}}$$

Which can be solved using a Linear program(LP).

We finally showed, using results from the field of metric embeddings, that this relaxation creates an at worst $O(\log n)$ gap.

References

[1] Anupam Gupta, Sparsest Cut and l_1 Embeddings, CMU Lecture Notes, 2008.

[2] Assaf Naor, Metric Embeddings and Lipschitz Extensions, Princeton University Department of Mathematics Lecture Notes, 2015.

[3] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. Combinatorica, 15(2):215–245, 1995.

[4] Sanjoy Dasgupta and Anupam Gupta, An Elementary Proof of a Theorem of Johnson and Lindenstrauss, DOI 10.1002/rsa.10073, 2002.