l_1 embeddings and sparsest cut

We prove Bourgain's theorem.

Theorem(Bourgain): Any n point finite metric can be embedded into $R^{O(\log^2 n)}$ with $O(\log n)$ distortion. Moreover there is is a randomized polynomial time algorithm to obtain the embedding.

We will then recap how the algorithm also leads to the promised $O(\log k)$ approximation for sparsest cut

Notation

Given a metric space (V, d) we need to map V into R^h so that distances are preserved

For $S \subseteq V$, let $d(u, S)$ denote the distance of u from S, that is $\min_{v \in S} d(u, v)$

Recall that an l_1 embedding is essentially a positive sum of cut-metrics

Therefore the idea is to pick some sets S and use them to define the embeddings

Random sets of different sizes turn out be useful

Algorithm: basic version

Let $h = \log n + 1$ (assume it is integer wlog)

For $i = 1$ to h do

S_i = random set with each $u \in V$ picked with probability $p_i = 1/2^{i+1}$

endfor

For each $u \in V$

$f(u) = (d(u,S_1)/h, d(u, S_2)/h, \ldots, d(u, S_h)/h)$

Note that above gives an embedding into $R^{O(\log n)}$

This will not suffice and we will modify it slightly to obtain the final embedding algorithm

Analysis

We focus on a pair u,v and show the following:

$|f(u) - f(v)|_1 \leq d(u,v)$ and hence it is a *contraction*

and
$\text{Expect}[|f(u) - f(v)|_1] \geq c\ d(u,v)/h$ for some constant c

Thus the distance is preserved to within an $O(\log n)$ factor in expectation

For high probability we need to repeat algorithm $\Theta(\log n)$ times and this will be the final algorithm

Analysis

Note that $d(u, S) - d(v, S) \leq d(u,v)$ for any S by triangle inequality

Hence $|f(u) - f(v)|_1 \leq \sum_{i=1}^{h} d(u,v)/h \leq d(u,v)$

Thus the embedding is a contraction

The interesting part is when $|d(u,S) - d(v, S)|$ is large

Let $Ball(u, r) = \{ a \mid d(u,a) \leq r\}$ be the closed ball around u of radius r

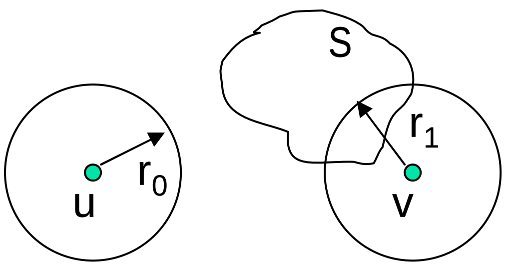
and $Ball'(u,r) = \{a \mid d(u,a) < r\}$ be the open ball

Analysis

The interesting part is when $|d(u,S) - d(v, S)|$ is large

Lemma: Let $r_0, r_1 \leq d(u,v)/2$, let $A = Ball'(u, r_0)$ and $B = Ball(v, r_1)$. If S Å A = ∅ and S Å B ≠ ∅ then $d(u,S) - d(v, S) \geq r_0 - r_1$

proof is easy from picture



Analysis

Fix a pair u,v

The crucial definition for the analysis is the following
$\rho_t = \min_r |Ball(u, r)| \geq 2^t$ ***and*** $|Ball(v,r)| \geq 2^t$

Let $l = \max_t \rho_t < d(u,v)/2$
For $1 \leq t \leq l+1$, let $X_t = |d(u, S_t) - d(v, S_t)|/h$ be the distance contribution of the random set S_t

Main lemma: There exists constant c s.t for $1 \leq t \leq l$
$Expect[X_t] \geq c\ (\rho_t - \rho_{t-1})/h$ and $Expect[X_{l+1}] \geq c\ (d(u,v)/2 - \rho_l)/h$

Analysis

Main lemma: There exists constant c s.t for $1 \leq t \leq l$

$\text{Expect}[X_t] \geq c\ (\rho_t - \rho_{t-1})/h$ and $\text{Expect}[X_{l+1}] \geq c\ (d(u,v)/2 - \rho_l)/h$

From above, we see that

$\text{Expect}[|f(u) - f(v)|_1] = \sum_{t=1}^{h} \text{Expect}[X_t]$

$\geq c/h\ ((\rho_1 - \rho_0) + (\rho_2 - \rho_1) + \ldots + (d(u,v)/2 - \rho_l)$

$\geq\ c\ d(u,v)/2h$

Analysis

Main lemma: There exists constant c s.t for $1 \leq t \leq l$ $\text{Expect}[X_t] \geq c\,(\rho_t - \rho_{t-1})/h$ and $\text{Expect}[X_{l+1}] \geq c\,(d(u,v)/2 - \rho_l)/h$

Consider $t \leq l$

From the definition of ρ_t, either u or v must have the property that $\text{Ball}'(u, \rho_t)$ contains $< 2^t$ points

Wlog assume that $\text{Ball}'(u, \rho_t)$ has $< 2^t$ points

Let $A = \text{Ball}'(u, \rho_t)$ and $B = \text{Ball}(v, \rho_{t-1})$

Note that $\text{Ball}(v, \rho_{t-1})$ has $\geq 2^{t-1}$ points

Observe that A and B are disjoint since $t \leq l$

Analysis

Let $A = \text{Ball'}(u, \rho_t)$ and $B = \text{Ball}(v, \rho_{t-1})$
$|A| < 2^{-t}$ and $|B| \geq 2^{t-1}$

If S Å $A = \emptyset$ and S Å $B \neq \emptyset$ then $d(u, S) - d(v, S) \geq \rho_t - \rho_{t-1}$

Recall that $X_t = |d(u, S_t) - d(v, S_t)|/h$
Therefore $\text{Expect}[X_t] \geq c\ (\rho_t - \rho_{t-1})/h$ where
$c \geq \Pr[S_t$ Å $A = \emptyset$ *and* S_t Å $B \neq \emptyset]$

Since A, B are disjoint
$\Pr[S_t$ Å $A = \emptyset$ *and* S_t Å $B \neq \emptyset] = \Pr[S_t$ Å $A = \emptyset]\ \ \Pr[S_t$ Å $B \neq \emptyset]$

Analysis

Pr[S_t Å A = ∅ *and* S_t Å B ≠ ∅] = Pr[S_t Å A = ∅] Pr[S_t Å B ≠ ∅]

Note that S_t is a random set with each $a \in V$ chosen in S_t independently with probability $p_t = 1/2^{t+1}$

Therefore
Pr[S_t Å A = ∅] $\geq (1 - p_t)^{|A|} \geq 1 - p_t|A| \geq 1 - 2^t/2^{t+1} \geq 1/2$

and
Pr[S_t Å B ≠ ∅] = 1 - Pr[S_t Å B = ∅] = $1 - (1-p_t)^{|B|}$
$\geq 1 - (1-1/2^{t+1})^{2^{t-1}} \geq 1 - e^{1/4}$

Thus Exepct[X_t] $\geq c\ (\rho_t - \rho_{t-1})/h$ where $c \geq (1-e^{1/4})/2$

Analysis

For $t = l+1$ the analysis is essentially the same
Since $\rho_{l+1} \geq d(u,v)/2$ either $Ball'(u,d(u,v)/2)$ or

$Ball'(v, d(u,v)/2)$ has less than 2^{l+1} points

Wlog assume that $|Ball'(u, d(u,v)/2)| < 2^{l+1}$

Set $A = Ball'(u, d(u,v)/2)$ and $B = Ball(v, \rho_l)$
Note that $|B| \geq 2^l$

Now using similar analysis as before we have
$E[X_{l+1}] \geq c(d(u,v)/2 - \rho_l)/h$

Modified algorithm

The analysis shows that for any particular pair u,v
Expec[$|f(u) - f(v)|_1$] $\geq$ c d(u,v)/h
and $|f(u) - f(v))_1 \leq$ d(u,v)

To ensure that all pairs u,v have good probability of having their distance preserved we need to repeat the algorithm several times independently

We describe the algorithm formally

Modified algorithm

Let $h = \log n + 1$ (assume it is integer wlog)

Let $N = 4\log n$

For $i = 1$ to h do

for $j = 1$ to N do

S^j_i = random set with each $u \in V$ picked with probability $p_i = 1/2^{i+1}$

endfor

For each $u \in V$

$f(u) = (d(u,S_1)/hN,\ldots,d(u,S^1_h)/hN,d(u,S^2_1)/hN,\ldots,d(u,S^2_h)/hN,\ldots,d(u,S^N_1)/hN,\ldots, d(u,S^N_h)/hN)$

essentially one coordinate per set chosen for a total of hN coordinates

Analysis

As before we can say that $|f(u) - f(v)|_1 \leq d(u,v)$

Fix pair u,v

Let $Y_j = \sum_{i=1}^{h} |d(u,S^j_i) - d(v, S^j_i)|/hN$

From previous analysis we can say that

$\text{Expect}\ [Y_j] \geq c\ d(u,v)/(hN)$

Therefore

$\text{Expect}[|f(u) - f(v)|_1] = \text{Expect}[\sum_{j=1}^{N} Y_j] \geq c\ d(u,v)/h$

Note that $Y_1, Y_2, \ldots, Y_N$ are *independent* random variables

Since we sum independent random variables, each of which behaves well, we can apply Chernoff bounds (see book) to say that with high probability, that is at least $(1 - 1/n^3)$

$|f(u) - f(v)|_1 \geq c\ d(u,v)/4h$

Analysis

Thus, with probability at least $(1 - 1/n^3)$
$|f(u) - f(v)|_1 \geq c\ d(u,v)/4h$

Therefore with probability at least $1-1/n$
$|f(u) - f(v)|_1 \geq c\ d(u,v)/4h$ for *all pairs* u,v

(Why?)

Therefore with high probability we have an $O(\log n)$ distortion embedding into $hN = O(\log^2 n)$ dimensional l_1 space

Back to sparsest cut

We previously showed that the integrality gap of the LP for sparsest cut is at most $\alpha(n)$ where $\alpha(n)$ is the distortion for embedding a finite metric into l_1. This did not immediately give rise to a polynomial time algorithm to round the LP. Here we show that Bourgain's embedding results in a randomized polynomial time algorithm.

First, we observe that Bourgain's algorithm is a randomized algorithm that can easily be implemented in polynomial time and succeds with high probability.

Sparsest cut

Recall that the LP for sparsest cut gives a metric d on the vertices V s.t

$\beta = \sum_{uv \in E} c(uv)\ d(uv) / \sum_{i=1}^{k} dem(i)\ d(s_i t_i)$

We apply Bourgain's embedding to d^* to obtain an l_1 metric d' on V in $R^{O(\log^2 n)}$

As we argued before we have

$\alpha(n)\ \beta \geq \sum_{uv \in E} c(uv)\ d'(uv) / \sum_{i=1}^{k} dem(i)\ d'(s_i t_i)$

Sparsest cut

We saw earlier that d' can be written as $\sum_S \lambda(S) d_S$ where $\lambda: 2^V \to \mathcal{R}^+$

The proof shows that the number of cuts S with $\lambda(S) > 0$ is at most nh if d' is in R^h and further these can be computed easily in poly time from d'

Since $h = O(\log^2 n)$ we obtain $O(n\log^2 n)$ cuts in the support of λ

As we saw before we can write

$\sum_{uv} c(uv)\, d'(uv) / \sum_i dem(i)\, d'(s_i t_i) = \sum_S \lambda(S)\, c(\delta(S)) / \sum_S \lambda(S)\, dem(\delta(S))$

and therefore there exists a cut S^* s.t that $\lambda(S^*) > 0$ and

$c(\delta(S^*))/dem(\delta(S^*)) \leq \beta\, \alpha(n)$

Since we have $O(n\log^2 n)$ explicit cuts S with $\lambda(S) > 0$ we can simply check all of them and pick the one with the minimum sparsity which is guaranteed by above to have sparsity at most $\beta\, \alpha(n)$

Sparsest cut

Since $\beta = OPT_{LP}$ we obtain an $\alpha(n)$ approximation. Since $\alpha(n) = O(\log n)$ we obtain an $O(\log n)$ approximation

The approximation ratio can be improved to $O(\log k)$ by noticing an additional property of Bourgain's embedding

Since d' is a contraction we have that

$\sum_{uv} c(uv)\ d(uv) \geq \sum_{uv} c(uv)\ d'(uv)$

Therefore, to obtain a ratio α we need to have that

$\sum_i dem(i)\ d(s_i t_i) \geq \alpha\ (\sum_i dem(i)\ d'(s_i t_i))$

Therefore it is sufficient to preserve the distances $d(s_i t_i)$ to within a factor of α

Sparsest cut

Therefore it is sufficient to preserve the distances $d(s_i t_i)$, $1 \leq i \leq k$ to within a factor of α

There are only k such distances. It is relatively easy modify the analysis to obtain such an embedding with $O(\log k)$ distortion for the distances $d(s_i t_i)$.

Instead of choosing $\log^2 n$ random sets where each random set was from the whole vertex set V, we choose $\log^2 k$ sets where each set is from T where T is the set of terminals $\{s_1,t_1,s_2,t_2,\ldots,s_k,t_k\}$. The analysis works on the distances induced on T. For non-terminals uv the distances don't increase and that is sufficient. This leads to the desired $O(\log k)$ (randomized) approximation algorithm

The algorithm can be derandomized but the details are involved and not particularly illuminating

Lower bound

We obtained an $O(\log k)$ approximation for sparset cut which also showed that the flow-cut gap is $O(\log k)$

Can this be improved?

We show that there are examples where the flow-cut gap is $\Omega(\log k)$. In particular we show this for $k=\Theta(n^2)$ which leads to an $\Omega(\log n)$ lower bound on the flow-cut gap

Note that this also shows that Bourgain's theorem is tight. That is, there are n point metrics that require $\Omega(\log n)$ distortion for embedding into l_1

(Why?)

Lower bound

The example is via constant degree expanders which we used for showing the gap for multicut problem as well.

Let G be a 3-regular (each node has degree 3) expander (for each S, $|S| \leq |V|/2$, $|\delta_G(S)| \geq |V|/2$)

Consider the uniform sparsest cut problem on G, that is, each (unordered) pair of vertices uv is a commodity and hence $k = n(n-1)/2$. Demand for each pair is 1

For any S, $|S| \leq |V|/2$ $\text{sparsity}(S) = |\delta(S)|/(|S||V \setminus S|)$

Lower bound

For any S, $|S| \leq |V|/2$ $sparsity(S) = |\delta(S)|/(|S||V \setminus S|)$

Since G is an expander, $|\delta(S)| \geq |S|$ and hence $sparsity(S) \geq 1/|V \setminus S| \geq 2/n$

Therefore *min sparsity* $\geq 2/n$

We wish to show that $OPT_{LP} = O(1/(n \log n))$ which would prove the desired gap

Consider setting $d_e = 1/\log n$ for each edge e of G

$d(uv)$ is then simply the shortest path distance with these edge weights

Lower bound

Consider setting $d_e = 1/\log n$ for each edge e of G

$d(uv)$ is then simply the shortest path distance between u and v with these edge weights

$OPT_{LP} \leq \sum_e d_e / \sum_{uv} d(uv)$

Since G has maximum degree 3, for each u there are at least $n/2$ vertices v such that the shortest path length in G is at least of $\log n/6$. Therefore there are $\Omega(n^2)$ pairs uv such that $d(uv) \geq \log n/6 \cdot 1/\log n \geq 1/6$

Hence $\sum_{uv} d(uv) = \Omega(n^2)$

However $\sum_e d_e \leq 3n/2 \log n$ since total number of edges in G is $3n/2$

Thus $OPT_{LP} = O(1/(n \log n))$

Lower bound

In particular this also shows that the shortest path metric induced by the edges of an expander is not embeddable into l_1 with distortion better than $\Omega(\log n)$ (do you see why?)

Another way to see that $OPT_{LP} = O(1/(n \log n))$ is via duality. Note that $OPT_{LP} = \lambda^*$ where λ^* is the maximum concurrent flow for each commodity.

We observed that the length of the shortest path in G is at least $\log n/6$ for $\Omega(n^2)$ pairs. Thus any flow for such a pair uses paths of lenght at least $\log n/6$. Thus the total capacity needed to route λ^* flow for each pair is $\Omega(\lambda^* n^2 \log n)$. However the total number of edges in the graph is only $3n/2$ and hence $\lambda^* = O(1/(n \log n))$