Lecture 2: Geometric Embeddings (continued)

Lecturer: *Sanjeev Arora* Scribe: *Michael Dinitz*

1 Focus on the ℓ_1 norm

In the last lecture we defined metric spaces, normed spaces, and considered the distortion resulting from certain embeddings. In particular, we proved that l_1 norms cannot always be embedded isometrically into l_2 by considering a specific four-point l_1 norm and showing that it requires at least $\sqrt{2}$ distortion.

Today's lecture further explores the ℓ_1 norm. We see a couple of interesting examples of ℓ_1 spaces. We try to understand the distortion required to embed ℓ_1 into ℓ_2. We also see that this apparently simple norm ("Manhattan distance") is computationally very interesting. We will explore this further in future lectures.

2 Lowerbound for embedding into ℓ_2.

The simplest example of an n-point ℓ_1 metric is the k-dimensional hypercube $\{-1,1\}^k$, assuming $n = 2^k$. Here the ℓ_1 distance between two points is $2 \times$ number of coordinates they differ on.

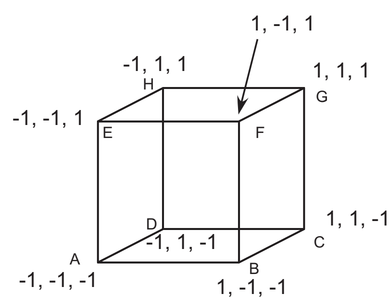

Figure 1: The hypercube in three dimensions

Today we show that this simple example of ℓ_1 requires large distortion for embedding into ℓ_2.

Theorem 1 (Enflo 1969)
Every embedding of the hypercube $\{-1,1\}^k$ into l_2 has distortion at least $\sqrt{k}$. (Thus denoting the number of points by $n = 2^k$, the distortion is $\sqrt{\log n}$.)

Proof: Recall the proof technique that we used previously: we come up with two weight functions, $w_1 : X \times X \to \mathbf{R}^+ \cup \{0\}$ and $w_2 : X \times X \to \mathbf{R}^+ \cup \{0\}$, such that for all embeddings f,

$$\frac{\sum w_1(x,y) d(x,y)^2}{\sum w_2(x,y) d(x,y)^2} \tag{1}$$

is different by a factor of k from

$$\frac{\sum w_1(x,y)\|f(x)-f(y)\|^2}{\sum w_2(x,y)\|f(x)-f(y)\|^2} \tag{2}$$

Here w_1 will put a weight 1 on all diametrically opposite pairs, and w_2 will put a weight 1 on all edges (i.e., adjacent pairs). Then (1) has value

$$\frac{2^k/2 \cdot k^2}{2^{k-1} \cdot k \cdot 1} = k.$$

We show by induction on k that for every embedding f, the value of (2) is at most 1, which will prove the theorem.

The base case $k = 2$ is exactly the problem that we considered last time. Assuming truth for $k-1$ we prove it for k. A dimension k hypercube can be thought of as two $k-1$ dimension hypercubes that have the corresponding points attached. By the quadrilateral inequality, the sum of the squares of the diagonals (where a diagonal refers to the distance between a point i and $-i$) of the k dimensional hypercube is less than or equal to the sum of the squares of the diagonals of the $k-1$ dimensional hypercube plus the sum of the squares of the edges between corresponding points of the two $k-1$ dimensional hypercube. In Figure 1, $k = 3$ and the inequality we are using is $AG^2 + CE^2 \leq AC^2 + EG^2 + AE^2 + GC^2$. Note that AC, EG are diagonals of $k-1$-dimensional hypercubes. So equation (2) becomes (in shorthand)

$$\frac{\sum_{\text{diagonals from } k-1} + \sum_{\text{new adjacencies}}}{\sum_{\text{adjacencies from } k-1} + \sum_{\text{new adjacencies}}} \tag{3}$$

By the inductive hypothesis, $\sum_{\text{diagonals from } k-1} \leq \sum_{\text{adjacencies from } k-1}$, so equation (2) is at most 1. $\square$

3 Understanding the l_1 norm

We saw in the last lecture that we can test a metric to see if it come from an l_2 space in polynomial time. This is more difficult in l_1, though, and is in fact **NP**-hard. The l_1 norm, while simple (anyone can understand the concept of Manhattan distance), seems to actually be more complicated than the l_2 norm. While we will not give a formal proof of this, we will show some of the intuition behind it.

3.1 Tree metrics

One class of metrics are the *tree metrics*, which are metrics that come from the shortest path metric on a weighted tree.

Theorem 2
Every tree metric embeds isometrically into l_1

Proof: We do this by using induction on the number of vertices of the tree. The base case, when there is only one vertex, is obvious. For the inductive step, we assume that all trees with fewer that k vertices can be embedded isometrically into l_1. Let T be a tree with k vertices, and let i and j be any two adjacent vertices with d_{ij} the weight of the edge between them. If we remove the edge $\{i,j\}$ then we have two smaller trees T_1 and T_2 with $i \in T_1$ and $j \in T_2$, so by the inductive hypothesis each of them can be embedded into l_1. They may require different dimensions, though, so say that T_1 is embedded in $\mathbf{R}^m$ and T_2 is embedded in $\mathbf{R}^k$ with the l_1 norm. We assign each vertex to an element of $\mathbf{R}^{m+k+1}$. Let v be a vertex of T. If v is in T_1, then its first m coordinates are the coordinates assigned to v by the embedding of T_1 into $\mathbf{R}^m$, the next coordinate is 0, and the last k coordinates are the coordinates assigned to j by the embedding of T_2 into $\mathbf{R}^k$. Similarly, if v is in T_2, then the first m coordinates are the m coordinates assigned to i by the embedding of T_1 into $\mathbf{R}^m$, the next coordinate is d_{ij}, and the last k coordinates are the coordinates assigned to v by the embedding of T_2 into $\mathbf{R}^k$. For any two vertices u and v, if they are both in T_1 or both in T_2 then the distance between them is clearly equal to the distance between them in the embedding of their subtree, which we know is the same as in the tree metric. If $u \in T_1$ and $v \in T_2$ then the distance between them in the l_1 norm is clearly equal to the distance from u to i plus the distance from i to j (which we know is d_{ij} by the way we assigned the vectors) plus the distance from j to v. Since each of the subtrees embedded isometrically, this is also the distance between them in the tree metric. $\square$

3.2 A characterization of ℓ_1: Cone of cut metrics

In this section we think of an n-point metric space as a subset of $\mathbf{R}^{\binom{n}{2}}$, since the space can be completely described by describing all $\binom{n}{2}$ pairwise distances.

A *convex cone* in $\mathbf{R}^k$ is a subset $S \subseteq \mathbf{R}^k$ where (i) if $x_1, x_2 \in S$ then $\lambda_1 \cdot x_1 + \lambda \cdot x_2 \in S$ for all $\lambda_1, \lambda_2 \geq 0$.

Note that the set of n-point l_1 metrics form a convex cone. (If d_1, d_2 are two finite ℓ_1 metrics on an n-point set, and if they are realizable in l, m dimensionions respectively then $\lambda_1 d_1 + \lambda_2 d_2$ is realizable in $l + m$ dimensions, where the first l dimensions correspond to a copy of d_1 scaled by λ_1 and the last m dimensions correspond to a copy of d_1 scaled by λ_2.)

Definition 1 *A* cut metric *is a subset (i.e. a cut) $S \subseteq [n]$ with $d_S(i,j) = 1$ if i and j are not on the same side of the cut and $d_S(i,j) = 0$ if they are.*

Note that a cut metric is not an actual metric but a pseudo-metric, since $d_S(i,j) = 0$ does not imply that $i = j$. The *cone of cut pseudo-metrics* consists of any $d \in \mathbf{R}^{\binom{n}{2}}$ such that d is expressible as $d = \sum_{S \subseteq [n]} \alpha_S d_S$, where $\alpha_S \geq 0$. Now we see that the two cones defined above are exactly the same. (As a sanity check you may wish to express both the hypercube metric and a tree metric as combination of cut metrics.)

Theorem 3
The set of all l_1 metrics is the same as the cone of cut metrics.

Proof: To show that any element d of the cut cut cone is an l_1 metric, we assign elements to vectors in $\mathbf{R}^{2^n}$, where each coordinate corresponds to a cut. Then if the element is in the ith cut we make its ith coordinate equal to α_S, otherwise we set it to 0. Clearly the l_1 distance between the vectors corresponding to elements i and j is equal to $d(i,j)$, so d is an l_1 metric.

Now we show that any l_1 metric d is expressible as an element of the cut cone. First take a realization of the metric as l_1 distances. It suffices to express each coordinate as a member of the cut cone, since the sum of elements of the cut cone is also in the cut cone. So consider the l_1 metric in $\mathbf{R}^1$ corresponding to each coordinate. This is simply a weighted tree (in fact just a path) and thus a tree metric, and therefore an element of the cut cone. □

This gives some intuition for why l_1 is complicated, since an l_1 metric could be a combination of an exponential number of cuts. Cut metrics and the cone of cut metrics are related to a certain class of problems known as cut problems. The most well-known problem in this class is the MINIMUM-CUT PROBLEM, which is the problem of determining, given a graph $G = (V, E)$, the cut $S \subset V$ that minimizes $\left|E(S, \bar{S})\right|$. This is obviously in $\mathbf{P}$ since there are many efficient algorithms for it (e.g. Ford-Fulkerson). Another cut problem is the MIN RATIO CUT or SPARSEST CUT problem, which is the problem of finding the cut $S \subset V$ with $|S| \leq |V|/2$ that minimizes $\frac{\left|E(S,\bar{S})\right|}{|S|}$. This problem is $\mathbf{NP}$-hard, as are most of the cut problems. These problems are obviously related to cut metrics, and the min-cut problem can be restated as finding the cut S that minimizes $\sum_{\{i,j\}\in E} d_S(i,j)$. For the min ratio cut, we use the formula

$$\frac{\sum_{\{i,j\}\in E} d_S(i,j)}{\sum_{i<j} d_S(i,j)} = \frac{\left|E(S,\bar{S})\right|}{|S|\left|\bar{S}\right|}. \tag{4}$$

Note that $\frac{|V|}{2} \leq \left|\bar{S}\right| \leq |V|$, so if we multiply by $|V|$ then we have the min ratio cut problem up to a factor of 2. Thus the min ratio cut problem is simply to find the cut S with $|S| \leq |V|/2$ that minimizes equation (4). By rescaling, we can think of this as minimizing the numerator of the LHS, subject to the denominator of the RHS. One of the consequences of the ellipsoid algorithm for convex optimization is that testing for membership in a convex set is equivalent to optimizing a linear function over the set. Therefore if we can test to see if a metric is an l_1 metric then we can optimize over the set of l_1 metrics. Since the set of l_1 metrics is the same as the cone of cut metrics, we can optimize over the cone of cut metrics and thus solve all manners of cut problems (including min ratio cut). Therefore deciding if a metric is an l_1 metric is $\mathbf{NP}$-hard. Those who wish to formalize the above intuition should try to reduce from MAX-CUT. (Or see a paper of Karzanov.)