Dimension Reduction in the ℓ_1 norm

Moses Charikar
Princeton University
moses@cs.princeton.edu

Amit Sahai
Princeton University
sahai@cs.princeton.edu

Abstract

The Johnson-Lindenstrauss Lemma shows that any set of n points in Euclidean space can be mapped linearly down to $O((\log n)/\epsilon^2)$ dimensions such that all pairwise distances are distorted by at most $1+\epsilon$. We study the following basic question: Does there exist an analogue of the Johnson-Lindenstrauss Lemma for the ℓ_1 norm?

Note that Johnson-Lindenstrauss Lemma gives a linear embedding which is independent of the point set. For the ℓ_1 norm, we show that one cannot hope to use linear embeddings as a dimensionality reduction tool for general point sets, even if the linear embedding is chosen as a function of the given point set. In particular, we construct a set of $O(n)$ points in ℓ_1^n such that any *linear embedding into ℓ_1^d must incur a distortion of $\Omega(\sqrt{n/d})$. This bound is tight up to a $\log n$ factor. We then initiate a systematic study of general classes of ℓ_1 embeddable metrics that admit low dimensional, small distortion embeddings. In particular, we show dimensionality reduction theorems for tree metrics, circular-decomposable metrics, and metrics supported on $K_{2,3}$-free graphs, giving embeddings into $\ell_1^{O(\log^2 n)}$ with constant distortion. Finally, we also present lower bounds on dimension reduction techniques for other ℓ_p norms.*

Our work suggests that the notion of a stretch-limited *embedding, where no distance is stretched by more than a factor d in any dimension, is important to the study of dimension reduction for ℓ_1. We use such stretch limited embeddings as a tool for proving lower bounds for dimension reduction and also as an algorithmic tool for proving positive results.*

1. Introduction

Dimension reduction refers to mappings points in a high dimensional space to a space with low dimensions while approximately preserving some property of the original points. We will be interested in dimension reduction techniques that map ℓ_p^d to $\ell_p^{d'}$ and approximately preserve pairwise distances of points. Given metric spaces (M_1, d_1) and (M_2, d_2), a mapping $f : M_1 \rightarrow M_2$ is said to be an embedding of M_1 into M_2 with distortion D if, for all $x, y \in M_1$,

$$\frac{d_1(x,y)}{D} \leq d_2(f(x), f(y)) \leq d_1(x,y).$$

The fundamental result in this area is the Johnson-Lindenstrauss lemma [18] which shows that any set of n points in Euclidean space can be mapped down to $O((\log n)/\epsilon^2)$ dimensions such that all distances are distorted by at most $1+\epsilon$. Moreover, such a mapping can be computed with high probability by simply projecting the set of points on randomly chosen unit vectors. [1]

Dimensionality reduction techniques using the Johnson-Lindenstrauss lemma and closely related methods have recently found numerous algorithmic applications: e.g. approximate searching for nearest neighbors [17, 19, 14], clustering of high dimensional point sets [8, 27], streaming computation [3, 15] and so on. See the recent survey by Indyk [16] for several applications of dimensionality reduction techniques.

The Johnson-Lindenstrauss lemma has proved to be a particularly useful tool since the ℓ_2 norm is a commonly used norm in various settings. A natural question to ask is whether there exists an analogue of the Johnson-Lindenstrauss lemma for other ℓ_p norms. Surprisingly little is known about this question. In particular, it would be very interesting to address this question for the ℓ_1 norm. (See the survey by Indyk [16] and the recent article by Linial [20] for discussions of this). This is certainly an intriguing theoretical question. Apart from its theoretical interest, the existence of a dimension reduction result for the ℓ_1 norm is motivated by several applications (See Section 3.3 in [16]).

Known results on dimension reduction: Ball [5] studied upper and lower bounds on the minimum dimension required for *isometric* embeddings in ℓ_p, proving linear lower bounds and quadratic upper bounds. The book by Deza and Laurent [10] gives a very good overview of the results in this area, particularly for isometric embeddings into ℓ_1 and

[1] The proof of the original result of Johnson and Lindenstrauss was subsequently simplified by a number of later works: Frankl and Maehara [11], Indyk and Motwani [17], Dasgupta and Gupta [9], and Achlioptas [1].

0272-5428/02 $17.00 © 2002 IEEE

ℓ_2. It is known that dimension reduction is not possible in the ℓ_∞ norm. In general, we need $\Omega(n)$ dimensions to represent a set of n points in ℓ_∞ with any distortion less than 3 [4, 24]. The results of Schechtman and Talagrand [29, 30] show that every n point metric in ℓ_1 can be embedded into $O(n \log n)$ dimensions with distortion at most $1 + \epsilon$.

The only known dimensionality reduction theorem for ℓ_1 is due to Indyk [15]. He showed that there is an embedding from ℓ_1^d to $\ell_1^{d'}$ with $d' = (\log 1/\delta)^{O(1/\epsilon)}$ such that distances do not increase with probability ϵ and distances do not decrease by a factor $(1+\epsilon)$ with probability $1-\delta$. Note, however, that with probability $1-\epsilon$, any distance can increase arbitrarily. In fact this holds for any ℓ_p norm with $p \in [1,2]$. Kushilevitz, Ostrovsky and Rabani [19] showed a dimension reduction result for the Hamming cube of a different flavor: they give low dimensional embeddings that can distinguish between two specified distance thresholds.

Another possible technique to obtain low dimensional ℓ_1 embeddings would be to first embed into ℓ_2, and then map the points to ℓ_1 (since ℓ_2 can be isometrically embedded into ℓ_1). In fact, we can get a mapping into $O(\log n)$ dimensions in ℓ_1 with distortion at most $1+\epsilon$ by simply projecting points onto random unit vectors. Indeed, by Bourgain's theorem [6], one can embed any metric into ℓ_2 with $O(\log n)$ distortion, so this gives a general technique to obtain low-dimensional embeddings into ℓ_1 with logarithmic distortion. However, this approach is in general limited to not do much better: In fact, there are examples of ℓ_1 embeddable metrics on n points which have distortion $\Omega(\sqrt{\log n})$ for any embedding into ℓ_2 [26].

Another aspect of low dimensional embeddings that has been investigated is the tradeoff between dimension and distortion in embedding metric spaces into normed spaces. Arias-de-Reyna and Rodríguez-Piazza [4] and Matousek [24] studied the minimum dimension of a normed space required to D-embed any n point metric, for fixed D. Matousek [23] also studied the minimum distortion required to embed a metric space into a d dimensional normed space, for fixed d. Gupta [12] showed an almost optimal dimension-distortion tradeoff for embedding trees into ℓ_2^d.

However the basic question still remains open: *Does there exist an analogue of the Johnson-Lindenstrauss dimension reduction lemma for the ℓ_1 norm ?* Note that the low dimensional embedding given by the Johnson-Lindenstrauss lemma for the ℓ_2 norm is particularly simple. Firstly, the embedding is linear and secondly, the embedding is oblivious, i.e. the image of each point is computed independently of the other points.

Our results: For the ℓ_1 norm, we show that we cannot hope to use linear embeddings as a dimensionality reduction tool for general point sets (even if the linear embedding is chosen as a function of the given point set). In particular, we construct a set of $O(n)$ points in ℓ_1^n such that any linear embedding into ℓ_1^d must have distortion $\Omega(\sqrt{n/d})$. This is tight within a logarithmic factor, i.e. there is a linear mapping of any set of $poly(n)$ points in ℓ_1^n to $\ell_1^{d \log n}$ with distortion $O(\sqrt{n/d})$. We reduce our analysis of embeddings into d dimensions to analyzing embeddings into arbitrary dimensions where no distance is stretched by more than a factor d in any dimension. We use such stretch limited embeddings as a tool for proving lower bounds for dimension reduction and later, as an algorithmic tool for proving positive results. We believe that understanding the nature of such stretch limited embeddings will ultimately lead to a resolution of the dimensionality reduction question for ℓ_1.

We initiate a systematic study of general classes of ℓ_1 embeddable metrics that admit low dimensional, small distortion embeddings. A natural general class of metrics are metrics *supported* on particular families of graphs. (A metric is supported on a graph if it is the shortest path metric of the graph for some weighting of the graph edges). It is known that tree metrics, metrics supported on outerplanar graphs and more generally, metrics supported on $K_{2,3}$-free graphs can be embedded isometrically into ℓ_1. We prove that tree metrics on n points can be embedded with distortion at most $(1+\epsilon)$ in ℓ_1 with $O(\frac{1}{\epsilon^2}\log^2 n)$ dimensions. For a class of ℓ_1 metrics called circular decomposable metrics (which includes metrics on outerplanar graphs and hence, all tree metrics as well), we show that we can get embeddings into $O(\frac{1}{\epsilon^2}\log^2 n)$ with distortion at most $3+\epsilon$. For any metric supported on a $K_{2,3}$-free graph, we show that the metric can be embedded in $O(\log^2 n)$ dimensions with constant distortion. An interesting feature of our proof is that we show that any metric on a $K_{2,3}$-free graph can be represented by *two* trees such the average distance in the two trees approximates all distances within a constant factor. This structural result may be of independent interest.

Finally, we address the issue of dimensionality reduction for other ℓ_p norms. We show that recent lower bounds for distance approximation in the data stream model [3, 28] imply very strong lower bounds for oblivious dimension reduction techniques for ℓ_p norms. In fact, for any oblivious dimension reduction technique that embeds ℓ_p^d into $\ell_p^{d'}$ and achieves a distortion of c with probability at least $1-\epsilon$ (for some small constant ϵ), this implies a dimension lower bound of $\Omega(d^{1-2/p}/(c^2 \log d))$. Further, we establish lower bounds for arbitrary (i.e. non-oblivious) dimension reduction in ℓ_p norms for $p \geq 4$. Our lower bound on the dimensions required grows exponentially with p (roughly of the form $4^p \log n$).

Organization: In Section 2, we present our lower bound on dimensionality reduction using linear embeddings for ℓ_1. In Section 3, we build a basic algorithmic tool for ℓ_1 dimensionality reduction, showing a dimension reduction theorem for ℓ_1 embeddings where each point has non-zero coordinates only in few dimensions. This immediately implies low dimension embeddings for tree metrics. In Sections 4 and 5, we apply this tool to obtain low dimension embeddings for metrics supported on $K_{2,3}$-free graphs, and circular-decomposable embeddings. Finally, in Section 6, we present lower bounds on dimension reduction for general ℓ_p norms.

2. Lower bound for Linear Embeddings

For a set of $O(n)$ points in ℓ_1^n, we give a lower bound on the distortion that can be achieved by linear embeddings into d dimensions. Consider embeddings from ℓ_1^n to ℓ_1^d that do not increase distances. Distances in the destination space are computed by adding up the distances in each of the d dimensions. For convenience, we pretend that each dimension is scaled by a factor of d and distances are computed by averaging over the d dimensions. Before scaling, each original dimension was a mapping of the points to a line so that no distance was increased. After scaling, each dimension is an embedding of the n points on a line, so that no distance is increased by more than a factor of d. Distances are computed by averaging distances in d such embeddings, i.e. weighing each such scaled dimension by a weight $\frac{1}{d}$.

We can now express (a relaxation of) the problem of embedding a set of points in d dimensions with minimum distortion as a linear program: the LP solution is allowed to average over embeddings onto lines such that no distance is increased by more than d. The relaxation comes from the fact that we do not restrict weights to be in the set $\{0, \frac{1}{d}\}$; instead the weights are real numbers in $[0, 1]$ that add up to 1.

Let σ denote an embedding onto a line that does not increase distances by more than d. We use d_{uv}^σ to denote the distance between u and v in the embedding. d_{uv} denotes the distance between i and j in the original metric. The variable x_σ denotes the fractional weight given to embedding σ. In general, the embeddings σ range over all possible embeddings on lines that do not stretch distances by more then d; in this form the LP can be used to prove bounds on arbitrary d dimensional embediings. To prove a lower bound on the distortion of linear embeddings, we restrict σ to range over linear embeddings. In our formulation, $y = 1/c$, where c is the distortion. We wish to minimize c, i.e. maximize y.

The LP is as follows:

$$\begin{array}{rrcl} & \max\ y & & \\ & \sum_\sigma x_\sigma & \leq & 1 \\ \forall\, uv & \sum_\sigma x_\sigma \cdot d_{uv}^\sigma & \leq & d_{uv} \\ \forall\, uv & y \cdot d_{uv} - \sum_\sigma x_\sigma \cdot d_{uv}^\sigma & \leq & 0 \end{array}$$

The dual is the following:

$$\begin{array}{rrcl} & \min\ z + \sum_{uv} \alpha_{uv} \cdot d_{uv} & & \\ & \sum_{uv} \beta_{uv} \cdot d_{uv} & \geq & 1 \\ \forall\, \sigma & z + \sum_{uv} \alpha_{uv} \cdot d_{uv}^\sigma - \sum_{uv} \beta_{uv} \cdot d_{uv}^\sigma & \geq & 0 \end{array}$$

In other words, we can write the dual as follows:

$$\text{Find } \alpha_{uv}, \beta_{uv} \text{ so as to minimize } \frac{z + \sum_{uv} \alpha_{uv} \cdot d_{uv}}{\sum_{uv} \beta_{uv} \cdot d_{uv}}$$

$$\text{subject to the constraint } \frac{z + \sum_{uv} \alpha_{uv} \cdot d_{uv}^\sigma}{\sum_{uv} \beta_{uv} \cdot d_{uv}^\sigma} \geq 1 \ \forall \sigma \quad (1)$$

Any feasible solution to the dual gives a lower bound on the distortion. Now we give a specific set of $O(n)$ points in ℓ_1^n for which we prove a lower bound. The point set consists of the origin O together with n points $\{P_i\}$ and $m = O(n)$ points $\{Q_i\}$. P_i is simply the ith unit vector, i.e. it has a 1 in the ith coordinate and 0's elsewhere. The Q_i's are chosen to have each coordinate set to either $+1$ or -1 at random, drawn from a pairwise independent distribution, i.e. every pair of coordinates is independent. There exist pairwise independent sample spaces of size $m = O(n)$; we have one Q_i for each sample in the sample space. Note that $d(O, P_i) = 1$ and $d(O, Q_i) = n$.

Consider any linear projection of ℓ_1^n onto a line. It can be represented by a vector $(\gamma_1, \gamma_2, \ldots, \gamma_n)$ so that the projection of $x = (x_1, x_2, \ldots x_n)$ is $\sigma(x) = \sum_i x_i \gamma_i$. We only consider linear projections that do not stretch distances by more than d. Since $\sigma(O) = 0$ and $\sigma(P_i) = \gamma_i$, this means that $|\gamma_i| \leq d$.

We now compute the distance of Q_i from the origin under the projection. $d(O, \sigma(Q_i)) = |\sum_j \gamma_j Q_{ij}|$. Let X be a random variable that measures the distance of $\sigma(Q_i)$ from the origin for a random i. Then,

$$\begin{aligned} \mathbf{E}[X^2] &= \mathbf{E}_i[(\sum_j \gamma_j Q_{ij})^2] \\ &= \mathbf{E}_i[\sum_j \gamma_j^2] + \mathbf{E}_i[\sum_{j_1 \neq j_2} \gamma_{j_1}\gamma_{j_2} Q_{ij_1} Q_{ij_2}] = \sum_j \gamma_j^2 \end{aligned}$$

Here we used the fact that for a randomly chosen i, Q_{ij_1} and Q_{ij_2} are independent. Under the projection, the average distance of Q_i from the origin is $\mathbf{E}[X] \leq \sqrt{\mathbf{E}[X^2]} = \sqrt{\sum_j \gamma_j^2}$.

Now, recall the formulation of the dual (1). We now choose dual variables z, α_{uv} and β_{uv} so as to obtain our bound. We set $z = n$. We set $\alpha_{uv} = 1$ iff u is the origin and v is some P_i, and 0 otherwise. We set $\beta_{uv} = 2\sqrt{\frac{n}{d}}/m$ iff u is the origin and v is some Q_i, and 0 otherwise. Thus, the expression $\sum_{uv} \beta_{uv} \cdot d^{\sigma}_{uv}$ is equal to $2\sqrt{\frac{n}{d}}$ times the average distance of the Q_i's to the origin in the embedding σ.

We first show that this is a feasible dual solution, i.e. for every permissible linear embedding σ, i.e.

$$z + \sum_{uv} \alpha_{uv} \cdot d^{\sigma}_{uv} \geq \sum_{uv} \beta_{uv} \cdot d^{\sigma}_{uv}. \tag{2}$$

Consider the linear embedding given by $(\gamma_1, \gamma_2, \ldots, \gamma_n)$. The LHS of (2) is exactly $n + \sum_j |\gamma_j|$. On the other hand the RHS of (2) is bounded by $2\sqrt{\frac{n}{d}}\sqrt{\sum_j \gamma_j^2} \leq 2\sqrt{\frac{n}{d}}\sqrt{d\sum_j |\gamma_j|} = 2\sqrt{n\sum_j |\gamma_j|}$. The last inequality follows from the fact that $|\gamma_j| \leq d$. Now the fact that LHS $\geq$ RHS follows simply by the fact that the arithmetic mean is $\geq$ the geometric mean.

The value of the dual solution is given by:

$$\frac{z + \sum_{uv} \alpha_{uv} \cdot d_{uv}}{\sum_{uv} \beta_{uv} \cdot d_{uv}} = \frac{n + n}{2\sqrt{\frac{n}{d}} \cdot n} = \sqrt{\frac{d}{n}}$$

Thus y^*, the optimal value of the LP solution for the set of points is at most $\sqrt{d/n}$. Recall that the distortion was $1/y$. Thus the distortion is at least $\sqrt{n/d}$.

Theorem 1 *There exists a set of $O(n)$ points in ℓ_1^n such that any linear embedding into ℓ_1^d has distortion at least $\sqrt{n/d}$.*

This tradeoff between dimension and distortion for linear embeddings is tight upto a logarithmic factor.

Theorem 2 *There exists a linear embedding of any set V of points in ℓ_1^n to $\ell_1^{d'}$ where $d' = O(d \log |V|)$ and the distortion is $O(\sqrt{\frac{n}{d}})$.*

Proof: We break up the n dimensions into d groups of n/d dimensions each. Now thinking of each group of n/d dimensions separately, for the points restricted to the n/d dimensions in a particular group, distortion between the ℓ_1 and ℓ_2 norms is at most $\sqrt{n/d}$. The ℓ_2 norm can be embedded into ℓ_1 with $O(\log |V|)$ dimensions and $O(1)$ distortion. This gives an overall embedding of the ℓ_1 norm on the n/d dimensions in a group to $\ell_1^{O(\log |V|)}$ with distortion $O(\sqrt{n/d})$. Concatenating the embeddings for all d groups, we get the claimed result. ∎

Our result shows that linear embeddings are particularly bad from the point of view of dimension reduction in ℓ_1. Earlier, we remarked that any n-point ℓ_1 metric can be (non-linearly) embedded in $\ell_1^{O(\log n)}$ with distortion $O(\log n)$; significantly better than what can be achieved by linear embeddings. In fact, for the set of $O(n)$ points we constructed for the lower bound in this section, there is a $(1+\epsilon)$-embedding into $\ell_1^{O(\log n)}$.

3 Dimension Reduction for Small-Support Embeddings

We first consider the case where points are embedded in ℓ_1 using a large number d of dimensions, but where each point has non-zero coordinates only in at most k dimensions. In this case, we can embed these points into ℓ_1 with distortion $1+\epsilon$, using only $O((k/\epsilon)^2 \log d)$ dimensions.

We make use of combinatorial *designs* (also called packings). We seek a set system $S_1, \ldots, S_d$ such that:

- $S_i \subset \{1, 2, \ldots, t\}$.
- $|S_i| = \ell$.
- For any two sets S_i, S_j, where $i \neq j$, we have that $|S_i \cap S_j| \leq (\epsilon/2k) \cdot \ell$.

It is well-known and easy to construct such a set system where $t = O((k/\epsilon)^2 \log d)$ (for instance by picking such sets at random). Given this fact, we are ready to prove the following lemma:

Lemma 1 *Let any number of points be embedded in ℓ_1 using d dimensions, where each point has non-zero coordinates only in at most k dimensions. Then these points can be embedded into ℓ_1 with distortion $1+\epsilon$, using $O((k/\epsilon)^2 \log d)$ dimensions.*

Proof: Let $S_1, \ldots, S_d$ be as stated above. We construct an explicit linear embedding of the points into ℓ_1 space using $t = O((k/\epsilon)^2 \log d)$ dimensions.

The embedding is defined as follows: There will be t dimensions. Original dimension i $(1 \leq i \leq d)$ is linearly mapped to the vector v_i, which has value $1/\ell$ in all coordinates from the set S_i, and value 0 elsewhere. Let f denote this linear mapping. We show that f has distortion at most $1+\epsilon$. First, observe that since each v_i has ℓ_1 norm exactly 1, this embedding cannot increase ℓ_1 distances.

Now, note that the distance between any two points p and q is equal to the ℓ_1 norm of the point $p-q$. Furthermore, $p-q$ will have at most $2k$ non-zero coordinates. Without loss of generality, suppose these are coordinates $1, \ldots, 2k$. Consider any such coordinate i. Since $|S_i \cap S_j| \leq (\epsilon/2k)|S_i|$ for all $j \neq i$, there must be at least $(1-\epsilon)\ell$ dimensions only in S_i, and not in any other set S_j for $j \in \{1, \ldots, 2k\} \setminus \{i\}$. Thus, if we throw out all dimensions common to more than one S_i, each set S_i retains at least a $(1-\epsilon)$ fraction of its dimensions. Let f' be the embedding f restricted to these dimensions which belong uniquely to a single S_i. Then we have $|f(p-q)|_1 \geq |f'(p-q)|_1 \geq (1-\epsilon)|p-q|_1$, completing the proof. ∎

Using this result, we can obtain low-dimensional ℓ_1 embeddings for tree metrics as follows. Given a tree T, we pick an arbitrary vertex as the root and construct a caterpillar decomposition [22, 25, 12] of T, which is a partition of the edge set of T into a collection of paths each of which is a subset of a root-leaf path. One can obtain a caterpillar decomposition such that any root leaf path intersects at most $O(\log n)$ paths in the decomposition; further all such intersections are prefixes of the paths in the decomposition. This leads to an embedding of the vertices of T into ℓ_1 with one dimension corresponding to each path in the decomposition. For a vertex v in the tree, the coordinate corresponding to a path P in the decomposition is simply the length of the intersection of the root-v path with P. It is easy to verify that this is an isometric embedding; further each point has at most $O(\log n)$ non-zero coordinates. An application of Lemma 1 yields an embedding into ℓ_1 with $O((\log^3 n)/\epsilon^2)$ dimensions and distortion at most $1+\epsilon$.

Anupam Gupta suggested the following improvement. Starting with the paths in the caterpillar decomposition, we can construct a collection of $O(\log n)$ *spiders* such that the original tree metric is the sum of the spider metrics. (A spider is simply a subdivision of the star). The spiders are constructed as follows: Imagine the original tree with the paths of the caterpillar decomposition laid out on it. As the construction proceeds, certain paths will be contracted to single points and the tree will be modified. The first spider consists of the paths in the decomposition that are incident on the root. We contract all such paths to the root (keeping track of the vertices that coincide with the root). We repeat this process of constructing a spider (from paths indicident on the current root) and contracting paths on the modified tree until we are left with all the vertices at a single point. The properties of the caterpillar decomposition mentioned before imply that this process terminates after constructing $O(\log n)$ spiders. It is easy to see that for any pair of vertices u, v, the distance in the original tree is the sum of distances between u and v in each of the spider metrics constructed. Each spider yields an ℓ_1 embedding for the vertices of the tree with at most one non-zero coordinate per vertex. Using Lemma 1, we obtain a $(1+\epsilon)$-embedding into ℓ_1 with $O((\log n)/\epsilon^2)$ dimensions. Concatenating these embeddings for the $O(\log n)$ spiders, we obtain the following result:

Theorem 3 *Any tree metric on n nodes can be embedded in ℓ_1 with $O((\log^2 n)/\epsilon^2)$ dimensions with distortion $1+\epsilon$.*

4. Dimension Reduction for $K_{2,3}$-free graph metrics via Trees

In this section, we show how to embed any metric supported on a $K_{2,3}$-free graph into ℓ_1 using polylogarithmic dimensions, with constant distortion. (Note we do not optimize the constant, and rather optimize parameters for simplicity of presentation.) We do this by embedding the metric as a convex combination of exactly 2 tree metrics, and then applying the low-dimension ℓ_1 embedding theorem for trees.

One might ask why previously known small-distortion embeddings of metrics into distributions on tree metrics do not immediately yield low dimensional embeddings, using our dimension-reduction result for tree metrics. The reason is illustrated with the example of a simple n-cycle metric, with all edge distances equal to 1. A natural way to embed this metric onto a distribution on trees is to simply pick an edge at random, and remove it. The resulting distribution on trees is easily seen to have small expected distortion. However, it is terrible from the point of view of dimension reduction: Suppose we took any tree in this distribution; because one edge $e=(u,v)$ was deleted, the distance between u and v became $n-1$, whereas before it was 1. To "average out" this distortion to a constant will therefore require $\Omega(n)$ additional dimensions! It is precisely these "high variance" distances that we must avoid to achieve low dimension embeddings with small distortion.

Indeed, the technique of removing edges to create trees is commonly used to obtain embeddings to distributions on trees (and of course is necessary to obtain embeddings to distributions on *spanning* trees), but it is one we must avoid. A previous embedding of $K_{2,3}$-free graphs to distributions on trees by Gupta et al. [13] uses this technique of removing edges and therefore provably induces trees which can stretch distances by linear factors. We must use a different basic technique.

Flattening. We seek a method of embedding cycles to distributions on trees with small expected distortion, but such that no individual tree can ever increase distances by too much – in fact, we will use a method which never increases distances at all. With any such distribution, it is easy to see that it is safe to sample just a few such trees, and this will yield a small convex combination of tree metrics with small distortion. We call the elementary operation we use to do this *flattening* a cycle. Given a cycle C, we will pick some point p in C – where p is not necessarily a vertex of C (p could lie in some arbitrary point between two vertices in C; in this case we could think of creating a new vertex p and placing it in the cycle). Then we will embed C onto the line, by placing p at 0, and every vertex $v \in C$ at $d(p,v) \geq 0$. We call p the *anchor* of the embedding. This can be seen as "flattening" the cycle onto a line. Note that no distances can increase by flattening, though distances could decrease substantially! We will choose our flattenings carefully, to argue that the expected distance between two vertices cannot decrease too much.

Since we seek tree embeddings, it suffices to embed the biconnected components of the $K_{2,3}$-free graph, which can be either K_4 or outerplanar graphs. K_4 clearly has a distortion 3 embedding as a tree, so we need only show how to embed outerplanar graphs by trees.

Recall that any outerplanar graph can be constructed by considering a sequence of paths P_i, and then doing the following: Start with $G_1 = P_1$. At step i, we consider some edge $e_i = (u_i, v_i)$ on the outer face of G_i, and obtain G_{i+1} by either attaching the endpoints of P_i to u_i and v_i (in this case, the length of the path P_i should be at least the length of the edge e_i), or by attaching only one endpoint of P_i to either u_i or v_i. We also use the concept of a *slack* structure [13]. We say that an outerplanar graph has an α-slack structure if it can be built out of paths P_i such that the length of any path P_i which attaches to both endpoints of an edge e_i is at least α times the length of e_i. The following is a simple generalization of a fact from [13] (the proof can be seen easily from [13], and is omitted):

Lemma 2 *For any $\alpha \geq 1$, given an outerplanar graph $G = (V, E)$, there is an outerplanar graph $H = (V, E')$ with an α-slack structure such that $d_G \geq d_H \geq (1/\alpha) d_G$.*

Thus, by incurring distortion at most 9, we assume that the outerplanar graph G we wish to embed has a 9-slack structure. We will build our embedding inductively based on the sequence of paths P_i. We assume that the vertices of G are numbered arbitrarily, and whenever we write an edge $e = (u, v)$, we will write it so that $u < v$.

We will inductively grow exactly 2 trees, denoted L and R. At stage i, we will denote these trees by L_i and R_i. These trees will have the property that for each path P_i, there will be unique paths in L and R which "correspond" to P_i. Note that these corresponding paths will not necessarily be the same as P_i, or even contain all the vertices of P_i. We will maintain the invariant that in each of our trees, the distance between any two nodes will never be larger than their corresponding distance in G. We will denote distance in the L trees by d_L, and in the R trees by d_R. Initially, $L_1 = L_2 = R_1 = R_2 = P_1$.

At stage i, if P_i is to be attached to a single vertex in G_i, then P_i is attached to that vertex in L_i and R_i to yield L_{i+1} and R_{i+1}, respectively. In this case, the new path will correspond to P_i in L and R.

If the endpoints of P_i are to be connected to both endpoints of some edge $e_i = (u_i, v_i)$ in G_i, then we proceed as follows:

For tree L_i, we locate the segment s (it need not be a single edge) in L_i corresponding to $e_i = (u_i, v_i)$. (In the following discussion, it will be helpful to think of u_i as being "on the left" and v_i as being "on the right".) Since distances can only decrease in each of our tree embeddings, the length of s is at most the length of e_i. We view $P_i \cup s$ as a cycle with circumference $8x + 2y$, where y is the length of the segment s (and note, by the 9-slack structure, that $x \geq y$). Now, we consider the point p which is at distance x from v_i and $x + y$ from u_i (we think of p as being "to the right" of v_i). Let Q_L be the flattened embedding on a line using p as the anchor. We then add this line embedding to L_i as follows: Let $L_{i+1} = L_i$ to start with. First, any vertices from P_i that lie between u_i and v_i in the line embedding Q_L are added as new vertices in L_{i+1} with corresponding locations between u_i and v_i in L_{i+1}, maintaining the distances from the Q_L embedding. We then add two new paths to L_{i+1}: The first is attached to u_i, which contains all vertices from P_i which were embedded with values greater than $x + y$ ("to the left" of u_i) in Q_L. This first path will be the path "corresponding" to P_i in L_i. The second new path is attached to v_i, and contains all vertices from P_i which were embedded with values smaller than x ("to the right" of v_i). Note that the first path essentially corresponds to $6x$ of the length of P_i, whereas the second path only contains the vertices in segment of length $2x$ from P_i. Note also that our embedding of Q_L into L_{i+1} is isometric.

Similarly, tree R_i proceeds identically to L_i, except that the anchor p is chosen to be at distance $3x$ from v_i and distance $3x + y$ from u_i, leading to the flattened embedding Q_R. Thus, in R_{i+1} the longer path (containing vertices with embedding values smaller than $3x$ in Q_R) will be attached to v_i rather than u_i; the shorter path (containing vertices with embedding values greater than $3x + y$ in Q_R) will be attached to u_i. Note again that our embedding of Q_R into R_{i+1} is isometric.

This finishes the description of the embedding to the two trees. We can now establish the following theorem about this embedding, whose proof is given in the full version of this paper:

Theorem 4 *Any metric supported on a $K_{2,3}$-free graph can be embedded into 2 trees with constant distortion.*

Combined with Theorem 3, we get:

Corollary 1 *Any metric supported on a $K_{2,3}$-free graph can be embedded into ℓ_1 with $O(\log^2 n)$ dimensions with constant distortion.*

5. Circular Decomposable Metrics

For a set $S \subseteq V$, the *cut metric* δ_S on V is defined by $\delta_S(x, y) = 1$ iff $|S \cap \{x, y\}| = 1$ and 0 otherwise. A metric d on V is ℓ_1 embeddable iff it can be expressed as a positive linear combination of cut metrics, i.e. $d = \sum_{S \subseteq V} \alpha_S \cdot \delta_S$ where $\delta_S \geq 0$ for all S.

Definition 1 *An ℓ_1 embeddable metric d on V is circular decomposable if there exists an ordering of the points in V, $v_1, v_2, \ldots, v_n$, such that $d = \sum_{S \subseteq V} \alpha_S \cdot \delta_S$ where each $S \subseteq V$ in the above summation with $\alpha_S > 0$ is of the form $\{v_i, v_{i+1}, \ldots v_j\}$.*

Equivalently, we can define *circular decomposable* metrics (CDMs) as follows: An ℓ_1 embeddable metric d on V is circular decomposable if there is an ordering of points in V, $v_1, v_2, \ldots, v_n$, such that in the embedding of d into ℓ_1, for every dimension, the ordering of points of V (sorted by coordinate value in that dimension) is of the form $v_i, v_{i+1}, \ldots v_n, v_1, v_2, \ldots v_{i-1}$ for some i. (Note that we will have to choose ascending or descending order and break ties in a particular way to satisfy this property).

In other words, it is possible to arrange the points of V on a circle $v_0, v_1, v_2, \ldots v_n = v_0$, such that in each dimension of the ℓ_1 embedding of d, the permutation of points is the linear ordering obtained by cutting the circle at some point. We will say that any linear ordering of the form $v_i, v_{i+1}, \ldots v_n, v_1, v_2, \ldots v_{i-1}$ is *consistent* with the circular ordering.

Circular decomposable metrics contain the class of metrics supported on outerplanar graphs (see e.g. [7] or the full version of this paper).

In this section, we will show that every CDM can be embedded in $O(\frac{1}{\epsilon^2} \log^2 n)$ dimensions with distortion $3 + \epsilon$. Before we present the details of the embedding, we give a high level overview of the proof.

A CDM can be viewed as a sum of n linear orderings of points, each consistent with a circular ordering of the points of V: $v_0, v_1, v_2, \ldots v_n = v_0$. (Note that in each ordering, the separation between adjacent points in the ordering could be arbitrary). We will arrange the n linear orderings starting with $v_1, v_2, \ldots v_n$, such that each succesive ordering is obtained from the previous one by moving the last element to the top. Without loss of generality, we assume that n is a power of 2. (If it is not a power of 2, we can make multiple copies of points so that the total number of points becomes a power of 2). We will construct $\log(n) - 1$ dimensions corresponding to each original ordering, such that the sum of distances according to the $\log(n) - 1$ dimensions is exactly equal to the distance in the original ordering. In each of these $\log(n) - 1$ dimensions, the ordering of points is exactly the same as the original ordering, except that several points may coincide (i.e. have the same coordinate value). We think of the $\log(n) - 1$ dimensions so constructed as belonging to $\log(n) - 1$ classes $1, \ldots \log(n) - 1$. The dimensions in each class have a special structure which we will elaborate on later. Next, we will combine groups of successive dimensions in each class. We group $n/2^{r+1}$ dimensions in group r, obtaining 2^{r+1} groups. For each group in each class, we produce 2 dimensions which approximate distances in the group to within a factor of 3. We will show that for each class, any point has non-zero coordinates in only a constant number of dimensions. Thus, we can obtain $O(\frac{1}{\epsilon^2} \log n)$ dimensions that approximate distances for all dimensions in a particular class by a factor of $1 + \epsilon$. Now, concatenating these dimensions for each of the $\log n$ classes, we obtain a $3(1 + \epsilon)$ distortion embedding in $O(\frac{1}{\epsilon^2} \log^2 n)$ dimensions.

We now proceed to elaborate on some of the details.

5.1. Breaking into classes

First we show how each original dimension is broken up into $\log(n) - 1$ dimensions. Recall that in each of these, the ordering of the points is exactly the same as the original ordering. The rth such dimension (referred to as the dimension of class r) is of the following form: The first $n\frac{1}{2^{r+1}}$ points all coincide, followed by points in the range $(n\frac{1}{2^{r+1}}, n\frac{1}{2^r}]$. Then the points in the range $(n\frac{1}{2^r}, n(1 - \frac{1}{2^r}]$ all coincide, followed by points in the range $(n(1 - \frac{1}{2^r}), n(1 - \frac{1}{2^{r+1}})]$. Finally, all the points in the range $(n(1 - \frac{1}{2^{r+1}}), n]$ coincide.

Suppose the distance between succesive points $(i, i+1)$ is x in the original ordering. Now, the distance between succesive points $(i, i+1)$ in the dimension of class r is x if $\{i, i+1\} \subseteq [n\frac{1}{2^{r+1}}, n\frac{1}{2^r}]$, or $\{i, i+1\} \subseteq [n(1 - \frac{1}{2^r}), n(1 - \frac{1}{2^{r+1}})]$. If $\{i, i+1\}$ belongs to neither of these 2 intervals, then the separation between points i and $i+1$ is 0 in the constructed dimension of class r (i.e. i and $i+1$ coincide).

Note that the separation between i and $i+1$ is x in precisely one of the constructed dimensions. Thus the distance between successive points is preserved in the embedding into $\log(n)$ dimensions. Since each of the $\log(n) - 1$ dimensions has the same ordering of points as the original ordering, this implies that all pairwise distances are preserved.

5.2. Embeddings for each class

In each class, we group successive orderings together and obtain an embedding for the entire group. As mentioned earlier, for the embeddings of class r, we group $n/2^{r+1}$ succesive orderings together so as to obtain 2^{r+1} groups.

We now examine the structure of the orderings in each group of class r in greater detail. Note that in each embedding, the top and bottom $1/2^{r+1}$ fraction of points coincide and the middle $1 - \frac{1}{2^{r-1}}$ fraction of points also coincide. Henceforth, we will assume that the coordinate for the middle $1 - \frac{1}{2^{r-1}}$ fraction of points is 0. Consider the first ordering in the group; for convenience, we renumber points so that the first ordering is of the form $1, 2, \ldots, n$. Note that successive orderings are obtained by moving the last point over to the top. Thus, as we move from the first ordering in the group to the last ordering, the last $n/2^{r+1}$ points are moved up to the top one by one.

Notice that of the middle points in the range $(n\frac{1}{2^r}, n(1 - \frac{1}{2^r})]$ that have coordinate 0 in the first ordering, the points in the range $(n\frac{1}{2^r}, n(1 - \frac{3}{2^{r+1}})]$ continue to have coordinate

0 in all the orderings. This will be a useful fact that we will exploit later.

In the next section, we will describe the embedding for the entire group. We first summarize the properties of this embedding and assuming these properties, show how this implies that circular decomposable metrics can be embedded in low dimensions.

The embedding of each group in class r will satisfy the following properties:

1. The embedding will consist of 2 dimensions.
2. Distances in the embedding will not increase and will decrease by at most a factor of 3.
3. All points in the range $(n\frac{1}{2^r}, n(1-\frac{3}{2^{r+1}})]$ (where positions are w.r.t. the first ordering in the group) will be placed at the origin (i.e. have coordinate 0) in the 2 dimensions.

These properties are established in the full version of the paper.

Lemma 3 *For any fixed class r, each point has non-zero cordinates in at most a constant number of dimensions constructed by the embedding for groups in class r.*

Proof: By property 3 of the embeddings constructed, a point will have non-zero coordinates in the embedding for a group in class r only if its position is in the range $(0, n\frac{1}{2^r}]$ or in the range $(n(1-\frac{3}{2^{r+1}}), n]$ (where positions are specified w.r.t. the first ordering in the group).

Note that each group in class r consists of 2^{r+1} consecutive orderings. The top $1/2^r$ fraction of the elements in the ith such group are the elements in the range $\left(n\frac{i-1}{2^{r+1}}, n\frac{i+1}{2^{r+1}}\right]$ in the first ordering in the first group. (Here, position $n+x$ is interpreted as position x). Similarly, the bottom $3/2^r$ elements in the ith group are the elements in the range $\left(n(1+\frac{i-4}{2^{r+1}}), n(1+\frac{i-1}{2^{r+1}})\right]$ in the first ordering in the first group. Note that the first interval is of length $2n/2^{r+1}$ and shifts by an amount $n/2^{r+1}$ with each successive group; the second interval is of length $3n/2^{r+1}$ and shifts by an amount $n/2^{r+1}$ with each succesive group. Hence for any fixed point, the number of groups i for which the point belongs to the first or second interval for the group is at most $2+3=5$.

For all other groups, the coordinate for the point will be 0. Hence the point has non-zero coordinates in at most $5*2=10$ dimensions in the constructed embeddings for groups of class r. ■

Since the embeddings for groups class r satisfy the property that every point has at most a constant number of nonzero coordinates, it follows from Lemma 1 that we can construct an embedding in $O(\frac{1}{\epsilon^2}\log n)$ dimensions for all the groups in class r such that distances are preserved within a factor of $1+\epsilon$. Since the original embeddings constructed distorted distances by a factor of 3, the low dimensional embeddings distort distances by at most $3+\epsilon$. Concatenating these low dimensional embeddings for each class, we get the following result:

Theorem 5 *Every circular decomposable metric on n points can be embedded in $O(\frac{1}{\epsilon^2}\log^2 n)$ dimensions such that distances are distorted by at most a factor of* $3+\epsilon$.

5.3. Low distortion embeddings for each group

We now outline the promised low distortion embedding for each group. In order to do this, we describe an embedding for a slightly more general setting:

Suppose we have points $P=\{p_0,\dots p_\ell\}$ and $Q=\{q_0=p_0, q_1, \dots q_{m-1}, q_m=p_\ell\}$. Further, suppose we have a set of ℓ dimensions where the points are ordered thus: In the kth dimension, the ordering of the points is $p_{k-1},\dots p_0=q_0, q_1,\dots q_{m-1}, q_m=p_\ell, p_{\ell-1},\dots p_k$. Further, the points $p_{k-1},\dots p_0$ coincide in the kth dimension (i.e. have the same coordinate). This also holds for the points $p_\ell,\dots p_k$ in the kth dimension.

Note that each group as decribed previously is of this form. For a group in class r, P is the set of last $\frac{1}{2^{r+1}}$ fraction of the points in the first ordering in the group and Q is the first $1-\frac{1}{2^{r+1}}$ fraction of points in the first ordering in the group. The special structure of the points in P and Q can be exploited to construct an embedding in 2 dimensions so that all pairwise distances are not increased and decreased by at most a factor of 3. Further this embedding satisfies the properties we assumed and used in the previous section. The basic idea is to associate each P_i and Q_i with their average position in the ℓ dimensions. If we construct a single dimension with these coordinates, the distances amongst the P_i's and the distances amongst the Q_i's are preserved. However, the distance between P_i and Q_j could decrease a lot. To remedy this, we construct two dimensions which are slight distortions of the single dimension we described. We associate a perturbation Δ_i with each P_i. In the first dimension, each P_i has coordinate Δ_i less than its average position and in the second dimension, each P_i has coordinate Δ_i more than its average position. These values can be chosen so that distances are distorted by at most a factor of 3. The details appear in the full version of this paper.

6. Lower bounds for dimension reduction in ℓ_p norms

We now prove lower bounds on the dimensionality reduction we may hope to achieve for general ℓ_p-norms.

6.1. Oblivious dimension reduction

Saks and Sun [28] obtain lower bounds for the communication complexity of approximating the ℓ_∞ norm of two d-dimensional vectors. Roughly speaking, their setup is as follows: two players, one having vector x and the other y wish to communicate to compute $||x-y||_\infty$ approximately. Saks and Sun give lower bounds on the communication complexity for doing this. Their bound comes from considering the problem of distinguishing between the cases when the distance is $\leq d$ on the one hand and $\geq d^{1+\delta}$ on the other (δ is a given parameter). Their proof draws pairs (x, y) from a certain distribution on $[n]^d$ where $n = cd^{1+\delta}$. They prove a lower bound of $\Omega(d^{1-2\delta})$ bits on the one round communication complexity of this problem. Note that any dimension reduction technique for ℓ_∞^d can be viewed as a communication protocol for this problem, with the amount of communication required = the number of dimensions times the number of bits required to represent each coordinate. In fact, since all distances are in a range polynomial in d, without loss of generality, we can assume that in each dimension coordinates are represented by at most $O(\log d)$ bits. (This can be done by rounding each coordinate to the nearest multiple of $1/d^3$ say). Then, this implies a lower bound of $\Omega(d^{1-2\delta}/\log d)$ dimensions for oblivious dimension reduction in ℓ_∞^d with distortion d^δ (with success probability larger than a certain constant). Since $||x-y||_\infty \leq ||x-y||_p \leq d^{1/p}||x-y||_\infty$, any oblivious scheme for embedding ℓ_p^d into low dimensions with distortion d^δ will be en embedding scheme for ℓ_∞^d with distortion $d^{1/p+\delta}$. But this requires $\Omega(d^{1-2/p-2\delta}/\log p)$ dimensions. Setting $c = d^\delta$ (the distortion), the lower bound on the number of dimensions is $\Omega(d^{1-2/p}/(c^2 \log p))$.

6.2. Non-oblivious dimension reduction

We will approach the following question: What is the minimum number of dimensions d such that every set X of $O(n)$ points in ℓ_p^n can be $(1+\epsilon)$ embedded in ℓ_p^d? We will prove a lower bound on the number of dimensions required for $p \geq 4$.

For $1 \leq i \leq n$, let A_i be the point in ℓ_p^n whose ith coordinate is 1 and all other coordinates are 0 and let B_i be the point in ℓ_p^n whose ith coordinate is -1 and all other coordinates are 0. The pairwise distances between the points are as follows:

$$\begin{aligned} d(A_i, B_i) &= 2 \quad 1 \leq i \leq n \\ d(A_i, A_j) &= 2^{1/p} \quad 1 \leq i < j \leq n \\ d(B_i, B_j) &= 2^{1/p} \quad 1 \leq i < j \leq n \end{aligned}$$

We will prove that the set $X = \{A_1, B_1, A_2, B_2, \ldots, A_n, B_n\}$ cannot be $(1+\epsilon)$ embedded in very low dimensions. Our proof proceeds by showing that from any $(1+\epsilon)$ distortion embedding of X in ℓ_p^d, we can construct n unit vectors in the positive orthant of $\Re^d$ such that all pairwise dot products are at most ϵ'. (The details appear in the full version of this paper.) Here $\epsilon' = 4\left(\frac{(1+\epsilon)^p - 1}{2^p - 4}\right)$.

A result of Alon [2] gives a lower bound on the number of dimensions needed to have n almost orthogonal unit vectors (i.e. unit vectors in $\Re^d$ such that all dot products are small in magnitude). This then gives us a lower bound for d in terms of n and ϵ': $d = \Omega\left(\frac{\log n}{\epsilon'^2 \log(1/\epsilon')}\right)$. Note that Alon's result only holds for $\epsilon' = \Omega(1/\sqrt{n})$. In order to satisfy this, $p \leq c \log n$ for a small constant c. Thus, we obtain:

Theorem 6 *For $4 \leq p \leq c \log n$, there exists a set of $2n$ points in ℓ_p^n such that any $(1+\epsilon)$ distortion embedding of ℓ_p^n to ℓ_p^d requires the number of dimensions d to be at least*

$$\begin{cases} \Omega\left(\dfrac{4^p \log n}{p(1+\epsilon)^{2p}}\right) & \epsilon \geq \frac{1}{p} \\ \Omega\left(\dfrac{4^p \log n}{\epsilon^2 p^2 (p + \log(1/\epsilon))}\right) & \epsilon \leq \frac{1}{p} \end{cases}$$

7. Conclusions

Our work throws open a number of interesting questions. One question is to determine whether it is possible to obtain an oblivious dimension reduction technique for ℓ_1. We strongly believe that this is not possible. Can we obtain oblivious embeddings with better guarantees than those for linear embeddings ? Another question is to characterize the class of graphs such that metrics supported on them have low dimensional low distortion ℓ_1 embeddings. Can we prove a super logarithmic lower bound for the number of dimensions required to embed tree metrics in ℓ_1 with distortion $(1+\epsilon)$? Can we obtain $(1+\epsilon)$ distortion embeddings with polylogarithmic dimensions for circular decomposable metrics and metrics supported on $K_{2,3}$ free graphs ? It would be very interesting to show an n point metric which is isometrically embeddable in ℓ_1 and yet does not admit a $1+\epsilon$ distortion embedding into ℓ_1 with polylogarithmic dimensions. On the positive side, it would be very interesting to show that every set of n points in ℓ_1 can be embedded in polylogarithmic dimensions with constant distortion. Our results for low dimensional embeddings of restricted ℓ_1 metrics give some hope that this may be possible.

Acknowledgements

We are grateful to Anupam Gupta for valuable discussions and allowing us to include his improved result on low dimensional embeddings of tree metrics. We thank Yair Bartal for pointing out the reference [30].

References

[1] D. Achlioptas. Database-friendly random projections. In *Proc. 20th PODS*, 2001.

[2] N. Alon. Problems and Results in Extremal Combinatorics, Part I. manuscript.

[3] N. Alon, Y. Matias, and M. Szegedy. The Space Complexity of Approximating the Frequency Moments. *JCSS* 58(1): 137-147, 1999

[4] J. Arias-de Reyna and L. Rodriguez-Piazza. Finite metric spaces needing high dimension for lipschitz embeddings in banach spaces. *Israel Journal of Mathematics*, 79:103–111, 1992.

[5] K. Ball. Isometric embedding in l_p-spaces. *European Journal of Combinatorics*, 11:305–311, 1990.

[6] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52:46–52, 1985.

[7] V. Chepoi and B. Fichet. A note on circular decomposable metrics. Geometriae Dedicata, 69:237-240, 1998.

[8] S. Dasgupta. Learning Mixtures of Gaussians. In *Proc. 40th FOCS*, pp. 634-644, 1999.

[9] S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. Technical Report TR-99-06, International Computer Science Institute, Berkeley, CA, 1999.

[10] M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Springer, 1997.

[11] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Comb. Theory, Ser. B*, 44:355–362, 1988.

[12] A. Gupta. Embedding tree metrics into low dimensional Euclidean spaces. *Discrete Computational Geometry*, 24:105-116, 2000.

[13] A. Gupta, I. Newman, Y. Rabinovich and A. Sinclair. Cuts, Trees and l_1-Embeddings of Graphs. In *Proc. 40th FOCS*, pp. 399-409, 1999.

[14] P. Indyk. Dimensionality reduction techniques for proximity problems. In *Proc. 9th SODA*, pp. 371-378, 2000.

[15] P. Indyk. Stable Distributions, Pseudorandom Generators, Embeddings and Data Stream Computation. In *Proc. 41st FOCS*, pp. 189-197, 2000.

[16] P. Indyk. Algorithmic applications of low-distortion embeddings. In *Proc. 42nd FOCS*, pp. 10-33, 2001.

[17] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proc. 30th STOC* pp. 604–613, 1998.

[18] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mapping into Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[19] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient Search for Approximate Nearest Neighbor in High Dimensional Spaces. *SIAM J. Computing*, 30(2): 457-474, 2000.

[20] N. Linial. Finite Metric Spaces - Combinatorics, Geometry and Algorithms. In *Proc. International Congress of Mathematicians*, 2002.

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

[22] N. Linial, A. Magen, and M. Saks. Low distortion Euclidean embedding of trees. *Israel J. Math*, 106:339-348, 1998.

[23] J. Matousek. Note on bi-Lipschitz embeddings into low-dimensional Euclidean spaces. *Comment. Math. Univ. Carolinae*, 31:589-600, 1990.

[24] J. Matousek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics*, 93:333–344, 1996.

[25] J. Matousek. On embedding trees into uniformly convex Banach spaces. *Israel Journal of Mathematics*, 114:221-237, 1999.

[26] J. Matousek. Embedding finite metric spaces into Euclidean spaces, chapter in *Lectures on discrete geometry*. Graduate Texts in Mathematics, Vol. 212, Springer Verlag, 2002.

[27] R. Ostrovsky and Y. Rabani Polynomial Time Approximation Schemes for Geometric k-Clustering. In *Proc. 41st FOCS*, pp. 349-358, 2000.

[28] M. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. To appear in *Proc. 34th STOC*, 2002.

[29] G. Schechtman. More on embedding subspaces of L_p in ℓ_r^n. *Compositio Math.*, 61(2):159-169, 1987.

[30] M. Talagrand. Embedding subspaces of L_1 into ℓ_1^n. *Proc. Amer. Math. Soc.*, 108(2):363-369, 1990.