On the Impossibility of Dimension Reduction in ℓ_1

BO BRINKMAN

Miami University, Oxford, Ohio

AND

MOSES CHARIKAR

Princeton University, Princeton, New Jersey

Abstract. The Johnson–Lindenstrauss lemma shows that any n points in Euclidean space (i.e., $\mathbb{R}^n$ with distances measured under the ℓ_2 norm) may be mapped down to $O((\log n)/\varepsilon^2)$ dimensions such that no pairwise distance is distorted by more than a $(1+\varepsilon)$ factor. Determining whether such dimension reduction is possible in ℓ_1 has been an intriguing open question. We show strong lower bounds for general dimension reduction in ℓ_1. We give an explicit family of n points in ℓ_1 such that any embedding with constant distortion D requires $n^{\Omega(1/D^2)}$ dimensions. This proves that there is no analog of the Johnson–Lindenstrauss lemma for ℓ_1; in fact, embedding with any constant distortion requires $n^{\Omega(1)}$ dimensions. Further, embedding the points into ℓ_1 with $(1+\varepsilon)$ distortion requires $n^{\frac{1}{2}-O(\varepsilon\log(\frac{1}{\varepsilon}))}$ dimensions. Our proof establishes this lower bound for shortest path metrics of series-parallel graphs. We make extensive use of linear programming and duality in devising our bounds. We expect that the tools and techniques we develop will be useful for future investigations of embeddings into ℓ_1.

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Geometrical problems and computations; computations on discrete structures*; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms*

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Metric spaces, embedding, dimension reduction

---

A preliminary version of this article appeared in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science* (Oct.), IEEE Computer Society Press, Los Alamitos, CA, 2003.

B. Brinkman's work was completed while at Princeton University, and was supported by a Frances Upton Fellowship and DOE grant DE-FG02-02ER25540.

M. Charikar's work was supported by National Science Foundation (NSF) ITR grant CCR-0205594, DOE Early Career Principal Investigator award DE-FG02-02ER25540, NSF CAREER award CCR-0237113, and an Alfred P. Sloan Fellowship.

Authors' addresses: B. Brinkman, 230-L Kreger Hall, Miami University, Oxford, OH 45056, e-mail: brinkmwj@muohio.edu; M. Charikar, 35 Olden St., Princeton, NJ 08544, e-mail: moses@princeton.edu.

Journal of the ACM, Vol. 52, No. 5, September 2005, pp. 766–788.

1. *Introduction*

Dimension reduction refers to mapping points in a high-dimensional space to a space with low dimensionality 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 $M_1 = (X_1, \Delta_1)$ and $M_2 = (X_2, \Delta_2)$, a mapping $f : X_1 \to X_2$ is said to be an embedding of M_1 into M_2 with distortion D if

$$\exists r > 0 \text{ such that } \forall u, v \in X_1,\ \frac{r}{D}\Delta_1(u, v) \leq \Delta_2(f(u), f(v)) \leq r\,\Delta_1(u, v).$$

In general, the factor r is intended to allow the embedding to scale distances by some arbitrary fixed factor. (See Matoušek [2002] for details.) In this article, however, we will usually assume that $r = 1$ because we deal only with normed spaces, and these may already be scaled arbitrarily.

The fundamental result in this area is the Johnson–Lindenstrauss lemma [1984], which shows that any set of n points in Euclidean space can be mapped down to $O((\log n)/\varepsilon^2)$ dimensions such that all distances are distorted by at most $(1 + \varepsilon)$. Moreover, such a mapping can be computed with high probability by simply projecting the set of points onto randomly chosen unit vectors.[1]

Metric embeddings have traditionally been studied by functional analysts, and have recently attracted a lot of attention in the theoretical computer science community due to connections to approximation algorithms and the design of efficient algorithms. Dimension reduction techniques using the Johnson–Lindenstrauss lemma and closely related methods have recently found numerous algorithmic applications: for example, approximate searching for nearest neighbors [Indyk and Motwani 1998; Kushilevitz et al. 2000; Indyk 2000a], clustering of high-dimensional point sets [Arora and Kannan 2001; Dasgupta 1999; Ostrovsky and Rabani 2000], streaming computation [Alon et al. 1999; Indyk 2000b] and so on. (See the recent survey by Indyk [2001].)

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, the dimension reduction question for the ℓ_1 norm stands out and has attracted attention in several recent surveys on the subject of metric embeddings. This question is interesting both because of its inherent theoretical appeal as well as its potential algorithmic applications. Indyk [2001], in his tutorial on algorithmic applications of embeddings from FOCS, asks: "*Is there an analog of JL lemma for other norms, especially ℓ_1? This would give a powerful technique for designing approximation algorithms for ℓ_1 norms ...* " Linial [2002], in his article on finite metric spaces at the International Congress of Mathematicians, says the following about the "*mysterious ℓ_1*": "*We know much less about metric embeddings into ℓ_1, and the attempts to understand them give rise to many intriguing open problems $\cdots$ What is the smallest $k = k(n, \varepsilon)$ so that every n-point metric in ℓ_1 can*

[1] The proof of the original result of Johnson and Lindenstrauss was subsequently simplified by a number of later works: Frankl and Maehara [1988], Indyk and Motwani [1998], Dasgupta and Gupta [1999], Arriaga and Vempala [1999] and Achlioptas [2001].

be embedded into ℓ_1^k with distortion $< 1 + \varepsilon$? We know very little at the moment, namely $\Omega(\log n) \leq k \leq O(n \log n)$ for constant $\varepsilon > 0$. The lower bound is trivial and the upper bound is from Schechtman [1987] *and Talagrand* [1990]."

Known Results on Dimension Reduction. Ball [1990] 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 [1997] gives a very good overview of the results in this area, particularly for isometric embeddings into ℓ_1 and 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 [de Reyna and Rodriguez-Piazza 1992; Matoušek 1996].

The only known dimension reduction theorem for ℓ_1 is due to Indyk [2000b]. He showed that there is an embedding from ℓ_1^d to $\ell_1^{d'}$ with $d' = (\log 1/\delta)^{O(1/\varepsilon)}$ such that distances do not increase by more than a factor $(1 + \varepsilon)$ with probability ε and distances do not decrease by more than a factor $(1 - \varepsilon)$ with probability $1 - \delta$. Note, however, that with probability $1 - \varepsilon$, any distance can increase arbitrarily. This result holds for any ℓ_p norm with $p \in [1, 2]$. Kushilevitz et al. [2000] 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.

In a recent paper, Charikar and Sahai [2002] showed that one cannot hope to use *linear* embeddings for obtaining dimension reduction in ℓ_1. In particular they exhibited a set of $O(n)$ points in ℓ_1^n such that any linear embedding into ℓ_1^d incurs distortion $\Omega(\sqrt{n/d})$. They also constructed low dimensional low distortion embeddings for special classes of ℓ_1 embeddable metrics, including tree metrics and shortest path metrics of outer-planar graphs. This work introduced and used the notion of a *stretch-limited* embedding in proving lower bounds. We employ and make significant refinements to this proof technique in order to obtain our new results.

Our Results. We show strong lower bounds for general dimension reduction in ℓ_1. We give an explicit family of n points in ℓ_1 such that any embedding with distortion D requires $n^{\Omega(1/D^2)}$ dimensions. This proves that there is no analog of the Johnson–Lindenstrauss lemma for ℓ_1; in fact, embedding with any constant distortion requires $n^{\Omega(1)}$ dimensions. Further, embedding our set of points into ℓ_1 with $(1 + \varepsilon)$ distortion requires $n^{\frac{1}{2} - O(\varepsilon \log(\frac{1}{\varepsilon}))}$ dimensions. Our proof establishes this lower bound for series-parallel graphs, indicating that the low distortion low dimensional embeddings constructed in Charikar and Sahai [2002] cannot be extended to this class.

Subsequent to our work, Lee and Naor [2004] have provided a more elementary proof of our result.

Organization. We first introduce, in Section 1.1, the notion of *stretch* as a proxy for ℓ_1 dimensionality. We then give an overview of our proof technique in Section 1.2. In Section 2, we introduce a family of series-parallel graphs and establish lower bounds on the number of dimensions required to embed them into ℓ_1 with a specified distortion. In Section 3, we show how the lower bound for the graph family implies a lower bound for dimension reduction in ℓ_1. We conclude in Section 4 with a discussion of open problems and directions for future work.

1.1. STRETCH-LIMITED EMBEDDINGS AS A PROXY FOR DIMENSION. We first describe *stretch-limited embeddings* which were introduced in Charikar and Sahai [2002], though our presentation is somewhat different.

Definition 1.1 *Stretch-Limited Embedding*. An embedding $\sigma = (f_\sigma, \Delta_\sigma)$ of a metric $M = (X, \Delta)$, which consists of a mapping $f_\sigma : X \to \mathbb{R}^t$ *along with* a distance function Δ_σ. We will denote the ith dimension of u under f_σ as $f_\sigma(u)_i$, and we can view each $f_\sigma(\cdot)_i$ as a mapping onto the real line. Weights $\{w_1, w_2, \ldots, w_t\}$ are assigned to each line embedding $f_\sigma(\cdot)_i$ such that $\sum_{i=1}^t w_i = 1$, and Δ_σ is defined to be the weighted average of distances under the $f_\sigma(\cdot)_i$:

$$\Delta_\sigma(f_\sigma(u), f_\sigma(v)) \stackrel{def}{=} \sum_{i=1}^{t} w_i |f_\sigma(u)_i - f_\sigma(v)_i|.$$

The distortion of this embedding is defined to be $D = \inf D'$ such that

$$\exists r > 0,\ \forall u, v \in X,\ \frac{r}{D'}\Delta(u, v) \leq \Delta_\sigma(f_\sigma(u), f_\sigma(v)) \leq r\,\Delta(u, v).$$

If for all points u and v in the original metric space, and $\forall i \in \{1, \ldots, t\}$,

$$|f_\sigma(u)_i - f_\sigma(v)_i| \leq rs \cdot \Delta(u, v),$$

the embedding is said to be stretch s limited.[2]

The $f_\sigma(\cdot)_i$ are simply mappings onto the real line. When the resulting points on the real line are equipped with the normal definition of distance, $\Delta(u, v) = |u - v|$, we call the metric a "line metric." A stretch-s embedding is a convex combination of line metrics where distances in any line metric cannot be more than a factor s larger than distances in the *original* metric. We will assume, without loss of generality, that $r = 1$, implying that a stretch-limited embedding is always a contraction. Scaling a stretch-limited embedding affects neither the distortion nor the stretch incurred.

CLAIM 1. *The existence of a D-distortion embedding of a metric $M = (X, \Delta)$ into ℓ_1^s implies the existence of a D-distortion stretch-limited embedding of M with stretch s.*

PROOF. Consider a mapping g of M into ℓ_1^s. Let $g(\cdot)_i$ denote the ith dimension under g, which we think of as an embedding into a line. Let $\sigma = (f_\sigma, \Delta_\sigma)$ be the stretch limited embedding defined by $f_\sigma(u)_i = sg(u)_i$, and $w_i = \frac{1}{s}$. Then:

$$\begin{aligned}\Delta_\sigma(f_\sigma(u), f_\sigma(v)) &= \sum_{i=1}^{s} \frac{1}{s} |f_\sigma(u)_i - f_\sigma(v)_i| \\ &= \sum_{i=1}^{s} \frac{1}{s} (s|g(u)_i - g(v)_i|) \\ &= \|g(u) - g(v)\|_1.\end{aligned}$$

Since distances are identical under σ and g, their distortions must be equal. □

[2] An alternate definition of stretch is obtained by replacing the stretch condition by: $|f_\sigma(u)_i - f_\sigma(v)_i| \leq rs \cdot \Delta_\sigma(f(u), f(v))$. Note that this alternate definition makes the notion of stretch a function of the resulting metric independent of the original metric. However, the definition of stretch we choose seems easier to work with for proving lower bounds; any lower bounds we prove also hold for this alternate definition of stretch.

CLAIM 2. *The existence of a D-distortion stretch-limited embedding of a metric $M = (X, \Delta)$ with stretch s implies the existence of a $D(1+\varepsilon)$-distortion embedding of M into $\ell_1^{O(sD\log(n)/\varepsilon^2)}$ (for sufficiently small $\varepsilon > 0$).*

PROOF. Consider a stretch-s embedding $\sigma = (f_\sigma, \Delta_\sigma)$ as a probability distribution on line metrics, where each line metric $f_\sigma(\cdot)_i$ has probability w_i. For each of the m dimensions we will take a random line metric $f_\sigma(\cdot)_l$ from this distribution, and let the value of x in this dimension be $f_\sigma(x)_l/m$.

Consider the distance of a particular pair of points u and v in a random $f_\sigma(\cdot)_i$ where i is picked with probability w_i. For convenience of notation we will call this distribution $\mathcal{I}$, and I denotes a random index chosen from this distribution. The expected distance is exactly the distance between u and v in the stretch-limited embedding, which, in turn, is in the range $[\Delta(u, v)/D, \Delta(u, v)]$. The stretch condition imposes a bound on the value of $|f_\sigma(u)_i - f_\sigma(v)_i|$, namely that

$$\forall i : |f_\sigma(u)_i - f_\sigma(v)_i| \leq s\Delta(u, v).$$

Now we can use standard Chernoff–Hoeffding bounds:

$$Pr\left[\left(\sum_{j=1}^{m} |f_\sigma(u)_{I_j} - f_\sigma(v)_{I_j}|\right)\frac{1}{s\Delta(u, v)} \geq (1+\varepsilon)\frac{m}{s\Delta(u, v)}E[|f_\sigma(u)_I - f_\sigma(v)_I|]\right]$$

$$< \left(\frac{\exp(\varepsilon)}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\frac{m\Delta_\sigma(u,v)}{s\Delta(u,v)}}$$

and

$$Pr\left[\left(\sum_{j=1}^{m} |f_\sigma(u)_{I_j} - f_\sigma(v)_{I_j}|\right)\frac{1}{s\Delta(u, v)} \leq (1-\varepsilon)\frac{m}{s\Delta(u, v)}E[|f_\sigma(u)_I - f_\sigma(v)_I|]\right]$$

$$< \left(\frac{\exp(-\varepsilon)}{(1-\varepsilon)^{(1-\varepsilon)}}\right)^{\frac{m\Delta_\sigma(u,v)}{s\Delta(u,v)}},$$

where the I_j are i.i.d. random variables chosen from $\mathcal{I}$. Noticing that $\frac{\Delta_\sigma(u,v)}{\Delta(u,v)} \geq \frac{1}{D}$, we set $m = 8sD\log(n)/\varepsilon^2$. For this choice of m the probability that, for a particular pairwise distance, the average over m samples is not within $(1 \pm \varepsilon)$ of its expectation is at most $1/n^2$ (for $\varepsilon > 0$ small enough). Since there are $\binom{n}{2}$ pairs of points, $P[\text{relative error is at most } \varepsilon] \geq 1/2$. Hence, there exists an embedding in ℓ_1^m with $m = 8sD\log(n)/\varepsilon^2$ and distortion at most $D(1+\varepsilon)$.[3] □

Note that stretch-s embeddings are more general than ℓ_1 embeddings with s dimensions: Not only will we allow them to have arbitrary dimension, but we allow them to use any convex combination of line metrics, not simply an average. By the results above, stretch is a good proxy for dimension in ℓ_1 embeddings.

[3] Note that, if we used the alternate definition of stretch discussed earlier, the claim can be strengthened to guarantee an embedding into $\ell_1^{O(s\log n/\varepsilon^2)}$.

1.2. PROOF OVERVIEW. A technique commonly used to prove lower bounds on the distortion for embedding a metric $M_1 = (X_1, \Delta_1)$ into $M_2 = (X_2, \Delta_2)$ is to consider two (non-negative) linear combinations of distances: $\alpha_1 = \sum \alpha_{ij}\Delta_1(i, j)$ and $\beta_1 = \sum \beta_{ij}\Delta_1(i, j)$ where $\alpha_{ij}, \beta_{ij} \geq 0$ (see Matoušek [2002]). The goal is to pick distances for α and β such that in embedding M_1 into M_2, distances in α tend to expand while distances in β tend to contract. Letting $\alpha_2 = \sum \alpha_{ij}\Delta_2(f(i), f(j))$ and $\beta_2 = \sum \beta_{ij}\Delta_2(f(i), f(j))$, we will then try to prove that $\frac{\alpha_2}{\beta_2} \geq D\frac{\alpha_1}{\beta_1}$ (for some $D > 1$) for all embeddings $f : X_1 \to X_2$. This will establish a lower bound of D on the distortion of embedding M_1 into M_2.

In order to prove lower bounds for dimension reduction in ℓ_1, we adapt this technique. First, low-dimensional embeddings seem tricky to reason about. Instead, we focus on low-stretch embeddings, exploiting the connection between stretch-limited embeddings and embeddings in low dimensions. Our goal will be to prove a lower bound on the stretch s needed to achieve a given distortion D.

Recall that a stretch-limited embedding can be assumed to be nonexpansive without loss of generality. As a result, a stretch-limited embedding with distortion D must satisfy the property that no distance expands and also no distance contracts by more than a factor D. Let us write down these constraints as we would give them for an optimization problem:

$$\forall u, v \in X,\ \Delta_\sigma(f_\sigma(u), f_\sigma(v)) \leq \|u - v\|_1 \text{ (Nonexpansion constraint } NE_{uv}\text{)}$$

and

$$\forall u, v \in X,\ \frac{1}{D}\|u - v\|_1 \leq \Delta_\sigma(f_\sigma(u), f_\sigma(v)) \text{ (Low-distortion constraint } LD_{uv}\text{).}$$

Now consider linear combination of these constraints of the form

$$\forall \gamma_{uv}, \lambda_{uv} \geq 0,\ \sum_{u,v \in X} (\gamma_{uv} NE_{uv} + \lambda_{uv} LD_{uv}).$$

This is a class of linear constraints, and if all of the NE_{uv} and LD_{uv} are satisfied, so is any linear inequality L of this form. Recall that a stretch-limited embedding is a convex combination of line metrics. From the convexity of stretch-limited embeddings, we can see that:

Observation 1.2 If L is satisfied by some stretch-limited embedding σ, then L must be satisfied by at least one of the line embeddings $f_\sigma(\cdot)_i$.

Our task now is to, given some family of series-parallel graph metrics, derive a single inequality L by finding values for the γ_{uv} and λ_{uv}. The intuition is that this L should be hard to satisfy, and we pick the γ_{uv} and λ_{uv} to get the best bound possible. We will then show that any dimension that satisfies L incurs high stretch, and therefore that no stretch-limited embedding with low stretch exists.

Charikar and Sahai [2002] used this technique to prove lower bounds for linear embeddings (though our exposition is a little different). In that case, the restriction to linear embeddings and a careful choice of the inequality on pairwise distances made it possible to prove a lower bound on the stretch s required. How can one prove lower bounds on the stretch for arbitrary (i.e., nonlinear) line embeddings? Our innovation is to express the problem of minimizing stretch so as to satisfy the inequality L as a linear program. In general, finding such an LP formulation might

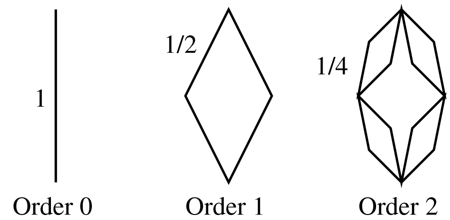

FIG. 1. Recursive diamond graphs of different orders.

be very difficult. However, we are able to obtain an LP that minimizes stretch for a carefully chosen family of points in ℓ_1 and a particular set of linear inequalities on pairwise distances. Having obtained the LP formulation, we consider the dual LP and exhibit a dual feasible solution. This establishes a lower bound on the stretch. Our dual solution was in fact extrapolated from embeddings generated by the CPLEX LP solver for large instances of our LP. This allowed us to discover combinatorial structure in our problem that we could leverage for our proof.

We will temporarily ignore the problem of how to derive the "correct" values of γ_{uv} and λ_{uv}. This is, however, crucial to obtaining a good lower bound on stretch, and we will discuss our solution to this problem in Appendix A. Again, optimization techniques and LP duality come to the rescue. For now, we will present our proof by directly obtaining the single "hard" constraint L.

2. *Series-Parallel Graphs Require High ℓ_1 Dimension*

In order to prove our results, we will focus on one particular family of series-parallel graphs, which we call the recursive diamond graphs. These are the graphs that Newman and Rabinovich [2002] previously used to establish an $\Omega(\sqrt{\log n})$ lower bound for embedding planar graphs into ℓ_2. The order 0 recursive diamond graph is a single edge, with length one. In order to make the order k graph from the order $k-1$ graph, replace each edge of length $1/2^{k-1}$ with a four-edge diamond, with edges of length $1/2^k$ (see Figure 1). This is a family of series-parallel graphs with 4^k edges and $\frac{2}{3}4^k + \frac{4}{3}$ vertices. Furthermore, the work of Gupta et al. [2004] shows that this graph can be embedded into ℓ_1 with constant distortion[4] (with many dimensions).

We will need some terminology in order to talk about the graph (see Figure 2). We will use n to refer to the number of vertices in a given diamond graph, and k to refer to the order (number of levels) of the graph. Each vertex has some k such that it is present in the order k graph, but not in the order $k-1$ graph. We will refer to any vertex as a *level k* vertex if it first appears in the order k graph. When an edge is replaced with a diamond, the two new vertices that are created will be called *siblings*, and we will refer to the pair of siblings as the *diagonal*[5] of this diamond.

[4] The recursive diamond graph is in fact a series-parallel "bundle." Gupta et al. [2004] show that such graphs can be embedded into ℓ_1 with distortion 2. We present an elementary distortion 2 embedding of the recursive diamond graph in Section 3.1.

[5] These are called *anti-edges* in Newman and Rabinovich [2002].

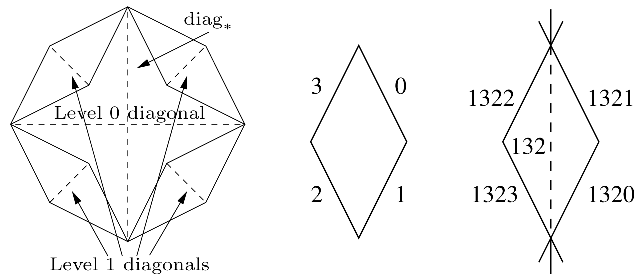

FIG. 2. Diamond graph terminology and labels for edges.

We will say that it is a level k diagonal if the vertices concerned are level k vertices. Finally, there is a natural parent–child relationship between diamonds of different levels: A diamond is a child of the diamond whose edge it replaces. An ancestor of a diamond is defined in the obvious way, and the ancestors of an edge are the diamonds of each order in which the edge participates.

Every edge in the graph is labeled by a string in $\{0, 1, 2, 3\}^k$. A particular diamond has four edges, which we will number 0, 1, 2 and 3 (again, see Figure 2). The label for the ith edge of a diamond is obtained by concatenating i to the label of its parent edge. We label diamonds with the label of the parent edge. Also, we use this same label to label the diagonal edge. For a label x, edge(x) denotes the edge labeled by x and diag(x) denotes the diagonal whose label is x. This leaves the original edge of the graph unlabeled. We will treat it as being "diagonal like" and refer to it as diag$_*$. We will return to the matter of exactly specifying the labeling in a later section: For now it is sufficient to notice that the 0 edge is always opposite the 2 edge and the 1 edge is always opposite the 3 edge.

We may now state our first result.

THEOREM 2.1. *A recursive diamond graph on n vertices requires* $n^{\Omega(1/D^2)}$ *dimensions to embed in* ℓ_1 *with distortion at most a constant* D.

2.1. PROOF. We focus on the n-vertex diamond graph (which has k levels). Consider $\mathcal{E}$, the set of all edges in the graph, and $\mathcal{D}$, the set of all diagonals. We will bound D by showing that the edges tend to expand while the diagonals tend to contract (refer back to the proof overview).

2.1.1. *The D-Distortion Constraint.* We first develop our key constraint on edge and diagonal lengths imposed by D. Recall that we have labeled the edges and diagonals. We refer to the length of the image of edge(x) as e'_x and the length of the image of diag(y) as d_y.[6]

We assumed without loss of generality that our embedding σ is nonexpansive and has distortion at most D. The nonexpansive property of σ implies that

$$-\sum_{x\in\{0,1,2,3\}^k} e'_x/2^k \geq -1.$$

[6]We use e'_x here because we later define a variable e_x, which is the *signed* length of an edge. The reader may safely ignore this distinction for now.

There are 4^k edges of length $\frac{1}{2^k}$, so this says that the average length of an edge must be at most $\frac{1}{2^k}$. The D-distortion property implies that

$$D\left(d_* + \sum_{i=0}^{k-1} \sum_{y\in\{0,1,2,3\}^i} d_y/2^i\right) \geq k+1.$$

At a level i, there are 4^i diagonals of length 2^i. This bound gives a weighted average of the diagonal lengths such that the total contribution of each level to the average is the same, and says that this weighted average cannot be more than a factor D smaller in the embedding than for the original metric. We combine these constraints to get a single constraint (referred to as the distortion constraint) that should be hard to satisfy:[7]

$$\forall\gamma \geq 0,\ D\left(d_* + \sum_{i=0}^{k-1} \sum_{y\in\{0,1,2,3\}^i} d_y/2^i\right) - \gamma \sum_{x\in\{0,1,2,3\}^k} - e'_x/2^k \geq k+1-\gamma.$$

We will eventually optimize γ in order to make this bound as strong as possible.

If this is true for a convex combination of line metrics, then it must be true for at least one of those line metrics. We will show a lower bound on the stretch s which must be incurred by a line metric which satisfies this constraint given values of n, k and D.

Let ρ be a line metric from σ which might satisfy this constraint. In order to simplify notation, let us re-interpret the meaning of e'_x to be the length of edge(x) in ρ and d_x to be the length of diag(x) in ρ. From this point on we will work only in the candidate line ρ, and all distances are in ρ unless otherwise noted.

2.1.2. *Constraints on Edges and Diagonals.* Before we can write down our LP, we will need a few more constraints. There is a very strong relationship between the length of an edge and the lengths of the diagonals of the edge's ancestor diamonds, and this will give us a second set of constraints.

We first precisely specify the labeling scheme for the edges of the recursive diamond graph. The labeling scheme we choose will depend on the particular ρ we are considering, but we will describe how to choose a labeling that satisfies our needs for *any* given ρ.

For each edge(x), we will designate one end point as the *head* and the other as the *tail* (denoted by head(edge(x)) and tail(edge(x)), respectively). Similarly, the end points of diag(x) are labeled as the *top* end point (denoted by top(diag(x))) and the *bottom* end point (denoted by bot(diag(x))). This labeling is done such that $\rho(\text{top}(\text{diag}(x))) \geq \rho(\text{bot}(\text{diag}(x)))$ (ties are broken arbitrarily).[8] We will now derive, for every edge(x) in the graph, an expression for $\rho(\text{head}(\text{edge}(x))) - \rho(\text{tail}(\text{edge}(x)))$ in terms of the lengths of edge(x)'s parent diagonals.

The edges of the diamond connect the end points of the parent edge(x) to the end points of diag(x) (see Figure 3). The edge connecting head(edge(x)) to top(diag(x)) is called the 0-edge. The edge connecting top(diag(x)) to tail(edge(x)) is called the

[7] Again, see Matoušek [2002] for more details.
[8] This choice is what causes the labeling to be dependent on ρ.

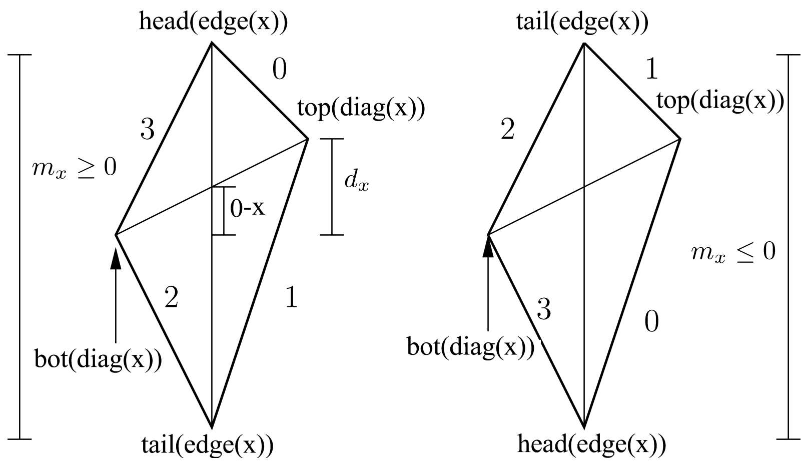

FIG. 3. Single dimension of embedded diamond.

1-edge. The edge connecting bot(diag(x)) to tail(edge(x)) is called the 2-edge. The edge connecting head(edge(x)) to bot(diag(x)) is called the 3-edge. Further, head(edge(x)) is considered the *head* for the 0-edge and the 3-edge; for each of these edges, the end point of diag(x) incident on it is considered the *tail*. tail(edge(x)) is considered the *tail* for the 1-edge and the 2-edge; for each of these edges, the end point of diag(x) incident on it is considered the *head*.

We define the following:

$$\begin{aligned} e_x &= \rho(\text{head}(\text{edge}(x))) - \rho(\text{tail}(\text{edge}(x))) \\ d_x &= \rho(\text{top}(\text{diag}(x))) - \rho(\text{bot}(\text{diag}(x))) \\ o_x &= \frac{\rho(\text{top}(\text{diag}(x))) + \rho(\text{bot}(\text{diag}(x)))}{2} - \frac{\rho(\text{head}(\text{edge}(x))) + \rho(\text{tail}(\text{edge}(x)))}{2}. \end{aligned}$$

Note that these definitions correspond exactly to our earlier definitions: $|e_x| = e'_x$ is the length of edge(x) (we allow e_x to be negative when head(edge(x)) $<$ tail(edge(x))) in ρ and d_x is the length of diag(x) in ρ. We refer to o_x as the *offset* of the diamond labeled x in ρ.

Now, we can calculate e_{x0}, e_{x1}, e_{x2} and e_{x3} in terms of e_x, d_x and o_x as follows:

LEMMA 2.2

$$e_{x0} = \frac{e_x}{2} - \frac{d_x}{2} - o_x \qquad e_{x1} = \frac{e_x}{2} + \frac{d_x}{2} + o_x$$
$$e_{x2} = \frac{e_x}{2} - \frac{d_x}{2} + o_x \qquad e_{x3} = \frac{e_x}{2} + \frac{d_x}{2} - o_x.$$

PROOF. We will show the calculation for e_{x0} only.

$$\begin{aligned} e_{x0} &= \rho(\text{head}(\text{edge}(x0))) - \rho(\text{tail}(\text{edge}(x0))) \\ &= \rho(\text{head}(\text{edge}(x))) - \rho(\text{top}(\text{diag}(x))) \\ &= \frac{\rho(\text{head}(\text{edge}(x))) - \rho(\text{tail}(\text{edge}(x)))}{2} + \frac{\rho(\text{head}(\text{edge}(x))) + \rho(\text{tail}(\text{edge}(x)))}{2} \\ &\quad - \frac{\rho(\text{top}(\text{diag}(x))) - \rho(\text{bot}(\text{diag}(x)))}{2} - \frac{\rho(\text{top}(\text{diag}(x))) + \rho(\text{bot}(\text{diag}(x)))}{2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{e_x}{2} - \frac{d_x}{2} \\
&\quad - \frac{\rho(\mathrm{top}(\mathrm{diag}(x))) + \rho(\mathrm{bot}(\mathrm{diag}(x)))}{2} + \frac{\rho(\mathrm{head}(\mathrm{edge}(x))) + \rho(\mathrm{tail}(\mathrm{edge}(x)))}{2} \\
&= \frac{e_x}{2} - \frac{d_x}{2} - o_x
\end{aligned}
$$

The proofs for e_{x1}, e_{x2} and e_{x3} are similar. □

Using this, one can obtain an expression for e_x in terms of the diagonal lengths and the offsets of the diamonds that are ancestors of edge(x). We use $y \sqsubset x$ to denote that y is a prefix of x, and the empty string is a prefix of every string. We also use $|x|$ to denote the length of the string x.

LEMMA 2.3

$$
e_x = \frac{d_*}{2^{|x|}} + \sum_{y \sqsubset x} S(x_{|y|+1}) \frac{d_y}{2^{|x|-|y|}} + \sum_{y \sqsubset x} T(x_{|y|+1}) \frac{o_y}{2^{|x|-|y|-1}}
$$

where $S(0) = S(2) = -1$, $S(1) = S(3) = +1$ and $T(0) = T(3) = -1$, $T(1) = T(2) = +1$.

PROOF. We prove this by induction on $|x|$.

Base Case. Consider $|x| = 0$. In this case, $e_x = d_*$ and the statement is true.

Inductive Step. Suppose the statement is true for all x such that $|x| = i$. Now consider e_{x0}, where $|x| = i$. From Lemma 2.1.2, $e_{x0} = \frac{e_x}{2} - \frac{d_x}{2} - o_x$. Using the expression we have for e_x from the inductive hypothesis, we get:

$$
\begin{aligned}
e_{x0} &= \frac{1}{2}\left(\frac{d_*}{2^i} + \sum_{y \sqsubset x} \frac{S(x_{|y|+1}) d_y}{2^{i-|y|}} + \sum_{y \sqsubset x} \frac{T(x_{|y|+1}) o_y}{2^{i-|y|-1}}\right) - \frac{d_x}{2} - o_x \\
&= \left(\frac{d_*}{2^{i+1}} + \sum_{y \sqsubset x} \frac{S(x_{|y|+1}) d_y}{2^{i+1-|y|}} + \sum_{y \sqsubset x} \frac{T(x_{|y|+1}) o_y}{2^{i+1-|y|-1}}\right) - \frac{d_x}{2} - o_x \\
&= \frac{d_*}{2^{i+1}} + \sum_{y \sqsubset x0} \frac{S(x_{|y|+1}) d_y}{2^{i+1-|y|}} + \sum_{y \sqsubset x0} \frac{T(x_{|y|+1}) o_y}{2^{i+1-|y|-1}}
\end{aligned}
$$

Thus, the statement holds for e_{x0} as well. Similarly, we can show that the statement holds for e_{x1}, e_{x2} and e_{x3}. By induction, the statement of the lemma is true. □

2.2. GROUPING EDGES AND DIAGONALS. Before we continue, we would like to remove the dependence on o_y. We noticed in experiments that optimal embeddings had all the o_y set to zero, so we expect removing the o_y should not hurt our bounds much. In fact, we will place our edges and diagonals into groups, and write our constraints in terms of the average distances in these groups. The careful choice of our labeling will cause the o_y terms to cancel out.

In particular, we group edges into 2^k groups of 2^k edges each. Groups are identified with labels in $\{0, 1\}^k$. For a group labeled by $z \in \{0, 1\}^k$, edge(x) belongs to the group z if $x(\mathit{mod}\ 2) = z$. Here $x(\mathit{mod}\ 2)$ refers to the label obtained by performing a coordinate-wise *mod* 2 operation. Similarly, diagonals of level i are grouped into

2^i groups, identified with labels in $\{0, 1\}^i$.

$$\overline{e_z} = \frac{1}{2^k} \sum_{\{x : x (mod\ 2) = z\}} |e_x|$$

$$\overline{d_z} = \frac{1}{2^i} \sum_{\{x : x (mod\ 2) = z\}} d_x.$$

In other words, $\overline{e_z}$ and $\overline{d_z}$ are the average lengths of their constituent edges and diagonals.

We can immediately rewrite our distortion constraint in terms of $\overline{e_z}$ and $\overline{d_z}$ without changing anything

$$\forall \gamma \geq 0 \quad D \left(\overline{d_*} + \sum_{i=0}^{k-1} \sum_{y \in \{0,1\}^i} \overline{d_y} \right) - \gamma \sum_{x \in \{0,1\}^k} \overline{e_x} \geq k + 1 - \gamma.$$

CLAIM 3. *For a group label $z \in \{0, 1\}^k$,*

$$\frac{1}{2^k} \sum_{\{x : x (mod\ 2) = z\}} e_x = \frac{d_*}{2^k} + \sum_{y \sqsubset z} S(z_{|y|+1}) \frac{\overline{d_y}}{2^{k-|y|}}$$

PROOF. Using Lemma 2.3, the value of the left-hand side is as follows:

$$\begin{aligned} \text{LHS} &= \frac{1}{2^k} \sum_{\{x : x (mod\ 2) = z\}} \left(\frac{d_*}{2^k} + \sum_{y \sqsubset x} \frac{S(x_{|y|+1}) d_y}{2^{k-|y|}} + \sum_{y \sqsubset x} \frac{T(x_{|y|+1}) o_y}{2^{k-|y|-1}} \right) \\ &= \frac{d_*}{2^k} + E_1(z) + E_2(z), \end{aligned}$$

where

$$E_1(z) = \frac{1}{2^k} \sum_{\{x : x (mod\ 2) = z\}} \sum_{y \sqsubset x} \frac{S(x_{|y|+1}) d_y}{2^{k-|y|}}$$

$$E_2(z) = \frac{1}{2^k} \sum_{\{x : x (mod\ 2) = z\}} \sum_{y \sqsubset x} \frac{T(x_{|y|+1}) o_y}{2^{k-|y|-1}}$$

We now simplify the two expressions $E_1(z)$ and $E_2(z)$. Note that $x, y \in \{0, 1, 2, 3\}^j$ for some j while $y', z \in \{0, 1\}^j$.

$$\begin{aligned} E_1(z) &= \frac{1}{2^k} \sum_{\{x : x (mod\ 2) = z\}} \sum_{y \sqsubset x} \frac{S(x_{|y|+1}) d_y}{2^{k-|y|}} \\ &= \sum_{y' \sqsubset z} \frac{1}{2^{|y'|}} \sum_{\{y : y (mod\ 2) = y'\}} \frac{S(z_{|y|+1}) d_y}{2^{k-|y|}} \\ &= \sum_{y' \sqsubset z} \frac{S(z_{|y'|+1}) \overline{d_{y'}}}{2^{k-|y'|}}, \end{aligned}$$

$$\begin{aligned} E_2(z) &= \frac{1}{2^k} \sum_{\{x:x(mod\,2)=z\}} \sum_{y \sqsubset x} \frac{T(x_{|y|+1}) o_y}{2^{k-|y|-1}} \\ &= \sum_{y' \sqsubset z} \frac{1}{2^{|y'|+1}} \sum_{\{y:y(mod\,2)=y'\}} \frac{(T(z_{|y|+1}) + T(2 + z_{|y|+1})) o_y}{2^{k-|y|-1}} \\ &= 0. \end{aligned}$$

The last equality follows from the fact that $T(i) + T(2+i) = 0$ for $i \in \{0, 1\}$. Substituting the values of $E_1(z)$ and $E_2(z)$ in the expression we derived earlier proves the claim. □

LEMMA 2.4. *For a group label $z \in \{0, 1\}^k$,*

$$\begin{aligned} \overline{e_z} &\geq \frac{d_*}{2^k} + \sum_{y \sqsubset z} S(z_{|y|+1}) \frac{\overline{d_y}}{2^{k-|y|}} \\ \overline{e_z} &\geq -\frac{d_*}{2^k} - \sum_{y \sqsubset x} S(z_{|y|+1}) \frac{\overline{d_y}}{2^{k-|y|}}. \end{aligned}$$

PROOF

$$\begin{aligned} \overline{e_z} &= \frac{1}{2^k} \sum_{\{x:x(mod\,2)=z\}} |e_x| \\ &\geq \frac{1}{2^k} \sum_{\{x:x(mod\,2)=z\}} e_x. \end{aligned}$$

Using Claim 3, we get the first inequality we need to prove. Also,

$$\begin{aligned} \overline{e_z} &= \frac{1}{2^k} \sum_{\{x:x(mod\,2)=z\}} |e_x| \\ &\geq -\frac{1}{2^k} \sum_{\{x:x(mod\,2)=z\}} e_x. \end{aligned}$$

Again, using Claim 3 gives the second inequality.

These inequalities result when we replace the e_x in Claim 3 with the $|e_x|$ from the definition of $\overline{e_z}$.

Linear Program for Minimizing Stretch. We have already derived three of the four constraints that we will use in our linear program. All that remains is to provide a lower bound for stretch.

Consider the stretch incurred by edge(x) in the line ρ. For every edge(x) = (u, v), $s \geq |e_x|/\Delta(u, v) \geq 2^k |e_x|$, where Δ is understood to be the distance function for the original metric space. Since $\max_{\{x:x(mod\,2)=z\}} |e_x| \geq \overline{e_z}$, we conclude that $\forall z \in \{0, 1\}^k \; s \geq 2^k \overline{e_z}$.

Now we are ready to give our linear program (see Table I). Note that we will optimize γ later, but that it is constant with respect to the variables of the LP. We provide the names of the dual variables in brackets for reference. We have carefully derived our constraints so that we can see that the solution to our LP is no larger than the minimum stretch needed to embed the recursive diamond graph into ℓ_1.

TABLE I. THE LINEAR PROGRAM

min s			
$D\left(\overline{d_*}+\sum_{i=0}^{k-1}\sum_{y\in\{0,1\}^i}\overline{d_y}\right)-\gamma\sum_{z\in\{0,1\}^k}\overline{e_z}$		$\geq k+1-\gamma$	$[\mu]$
$\forall z\in\{0,1\}^k$	$\frac{s}{2^k}-\overline{e_z}$	≥ 0	$[p_z]$
$\forall z\in\{0,1\}^k$	$\overline{e_z}+\left(\frac{\overline{d_*}}{2^k}+\sum_{y\sqsubset z}\frac{S(z_{\lvert y\rvert+1})\overline{d_y}}{2^{k-\lvert y\rvert}}\right)$	≥ 0	$[\alpha_z]$
$\forall z\in\{0,1\}^k$	$\overline{e_z}-\left(\frac{\overline{d_*}}{2^k}+\sum_{y\sqsubset z}\frac{S(z_{\lvert y\rvert+1})\overline{d_y}}{2^{k-\lvert y\rvert}}\right)$	≥ 0	$[\beta_z]$

TABLE II. THE DUAL LINEAR PROGRAM

max $(k+1-\gamma)\mu$			
$\forall z\in\{0,1\}^k$	$-\gamma\mu-p_z+\alpha_z+\beta_z$	≤ 0	$[\overline{e_z}]$
	$\sum_{z\in\{0,1\}^k}p_z$	$\leq 2^k$	$[s]$
$\forall y\in\bigcup_{i\in[0,k-1]}\{0,1\}^i$	$D\mu+\sum_{v\in\{0,1\}^{k-\lvert y\rvert}}\frac{S((yv)_{\lvert y\rvert+1})(\alpha_{yv}-\beta_{yv})}{2^{k-\lvert y\rvert}}$	≤ 0	$[\overline{d_y}]$
	$D\mu+\sum_{z\in\{0,1\}^k}\frac{\alpha_z-\beta_z}{2^k}$	≤ 0	$[\overline{d_*}]$

TABLE III. DUAL WITH μ FACTORED OUT

max $(k+1-\gamma)\mu$			
$\forall z\in\{0,1\}^k$	$\mu\left(-\gamma-p_z^*+\alpha_z^*+\beta_z^*\right)$	≤ 0	$[\overline{e_z}]$
	$\mu\left(\sum_{z\in\{0,1\}^k}p_z^*\right)$	$\leq 2^k$	$[s]$
$\forall y\in\bigcup_{i\in[0,k-1]}\{0,1\}^i$	$\mu\left(D+\sum_{v\in\{0,1\}^{k-\lvert y\rvert}}\frac{S((yv)_{\lvert y\rvert+1})(\alpha_{yv}^*-\beta_{yv}^*)}{2^{k-\lvert y\rvert}}\right)$	≤ 0	$[\overline{d_y}]$
	$\mu\left(D+\sum_{z\in\{0,1\}^k}\frac{\alpha_z^*-\beta_z^*}{2^k}\right)$	≤ 0	$[\overline{d_*}]$

Dual Linear Program for the Lower Bound on Stretch. We have formulated an LP minimization problem whose optimum value is a lower bound on the minimum stretch for a D-distortion embedding. In order to prove our lower bound we give the dual of this LP and a feasible solution. We construct the dual in the normal way (see Table II).

Next, we give our solution for this LP. In fact, our solution is very simple. Every variable is just a constant multiple of μ: $p_x = p_x^*\mu$, $\alpha_x = \alpha_x^*\mu$ and $\beta_x = \beta_x^*\mu$. We will specify the values of these constants, and then maximize μ subject to the constraints of the dual in order to get our bound. For these purposes, we can rewrite the dual LP (see Table III).

The Dual Solution. We now give our solution to the dual in Table IV. We use $\|x\|_1$ to denote the number of 1s in the 0-1 string x.

CLAIM 4. *The values of γ, α_z^*, β_z^* and p_z^* in Table* IV *give a feasible solution for our dual LP.*

PROOF. First check the $\overline{e_z}$ constraint, that $(-\gamma - p_z^* + \alpha_z^* + \beta_z^*) \leq 0$ for all z. We break this into three cases.

Case 1. $\|z\|_1 \leq k/2 - k/2D - 1$

$$\begin{aligned} -p_z^* + \alpha_z^* + \beta_z^* &= D(k-1-2\|z\|_1) - 2D(k/2 - k/2D - \|z\|_1) \\ &= \gamma. \end{aligned}$$

TABLE IV. THE DUAL SOLUTION

$$\gamma = k - D$$

$$\alpha_x^* = \begin{cases} D(k-1-2\|x\|_1) & \text{if } \|x\|_1 \le k/2-1 \\ 0 & \text{otherwise} \end{cases}$$

$$\beta_x^* = \begin{cases} D(2\|x\|_1-k+1) & \text{if } \|x\|_1 \ge k/2 \\ 0 & \text{otherwise} \end{cases}$$

$$p_x^* = \begin{cases} 2D(k/2-k/2D-\|x\|_1) & \text{if } \|x\|_1 \le k/2-k/2D-1 \\ 2D(\|x\|_1-k/2-k/2D+1) & \text{if } \|x\|_1 \ge k/2+k/2D \\ 0 & \text{otherwise} \end{cases}$$

Case 2. $\|z\|_1 \ge k/2 + k/2D$

$$\begin{aligned} -p_z^* + \alpha_z^* + \beta_z^* &= D(2\|z\|_1 - k + 1) - 2D(\|z\|_1 - k/2 - k/2D + 1) \\ &= \gamma. \end{aligned}$$

Case 3. $k/2 - k/2D - 1 < \|z\|_1 < k/2 + k/2D$

$$\begin{aligned} p_z^* &= 0 \\ \alpha_z^* &\le D(k - 1 - 2(k/2 - k/2D)) = k - D = \gamma \\ \beta_z^* &\le D(2(k/2 + k/2D - 1) - k + 1) = k - D = \gamma. \end{aligned}$$

Since the ranges where α_x^* and β_x^* are positive do not overlap, this proves that the $\overline{e_z}$ constraint is satisfied.

Now let us skip to the $\overline{d_y}$ constraint. In order to prove this, we will use the following lemma:

LEMMA 2.5

$$\forall i \in [0, k-1],\ y \in \{0,1\}^i,\ v \in \{0,1\}^{k-i-1}$$

$$S((y0v)_{|y|+1})(\alpha_{y0v}^* - \beta_{y0v}^*) + S((y1v)_{|y|+1})(\alpha_{y1v}^* - \beta_{y1v}^*) = -2D.$$

PROOF (OF LEMMA 2.5). $S((y0v)_{|y|+1}) = -1$ and $S((y1v)_{|y|+1}) = +1$, so

$$\begin{aligned} &S((y0v)_{|y|+1})(\alpha_{y0v}^* - \beta_{y0v}^*) + S((y1v)_{|y|+1})(\alpha_{y1v}^* - \beta_{y1v}^*) \\ &\quad = -\alpha_{y0v}^* + \beta_{y0v}^* + \alpha_{y1v}^* - \beta_{y1v}^*. \end{aligned}$$

Since $\|y0v\|_1 + 1 = \|y1v\|_1$, there are three cases.

Case 1. $\|y0v\|_1, \|y1v\|_1 \le k/2 - 1$

$$\begin{aligned} -\alpha_{y0v}^* + \beta_{y0v}^* + \alpha_{y1v}^* - \beta_{y1v}^* &= \\ -\alpha_i^* + \alpha_{i+1}^* &= \\ -D(k-1-2i) + D(k-1-2(i+1)) &= -2D. \end{aligned}$$

Case 2. $\|y0v\|_1, \|y1v\|_1 \ge k/2$

$$\begin{aligned} -\alpha_{y0v}^* + \beta_{y0v}^* + \alpha_{y1v}^* - \beta_{y1v}^* &= \\ \beta_i^* - \beta_{i+1}^* &= \\ D(2i-k+1) - D(2(i+1)-k+1) &= -2D. \end{aligned}$$

Case 3. $\|y0v\|_1 = k/2 - 1$, $\|y1v\|_1 = k/2$

$$\begin{aligned} -\alpha^*_{y0v} + \beta^*_{y0v} + \alpha^*_{y1v} - \beta^*_{y1v} &= \\ -\alpha^*_i - \beta^*_{i+1} &= \\ -D - D &= -2D. \end{aligned}$$

□

Applying Lemma 2.5, we conclude that

$$\begin{aligned} D + \sum_{v\in\{0,1\}^{k-|y|}} \frac{S((yv)_{|y|+1})(\alpha^*_{yv} - \beta^*_{yv})}{2^{k-|y|}} &= \\ D - \sum_{v'\in\{0,1\}^{k-|y|-1}} \frac{2D}{2^{k-|y|}} &= \\ D - \frac{(2^{k-|y|-1})(2D)}{2^{k-|y|}} &= 0. \end{aligned}$$

Hence, the $\overline{d_y}$ constraints are all satisfied: In fact, they are all tight.

The case for the $\overline{d_*}$ constraint is even simpler because the sign for α^*_z is always positive and the sign for β^*_z is always negative. For every x with

$$\|x\|_1 = l \le k/2 - 1,$$

pair x with y such that (x xor y) $= 111\cdots 1$ (in other words, y is the bitwise NOT of x). Note that

$$\|y\|_1 = k - l \ge k/2 + 1,$$

and that

$$\alpha^*_l - \beta^*_{k-l} = D(k - 1 - 2l) - D(2(k-l) - k + 1) = -2D.$$

This accounts for all x except where $\|x\|_1 = k/2$. In this case α_* is 0, so we see that the $\overline{d_*}$ constraint is satisfied.

Finally, we return to the s constraint. Recall that our lower bound will be $(1 + D)\mu$. This constraint is the only one which limits μ, and we will try to make μ as big as we can. Hence, $\mu = 2^k / \sum_z p^*_z$. Let us now bound p^*_z:

$$\begin{aligned} \sum p^*_z &= 2D\left(\sum_{i=0}^{\frac{k}{2}-\frac{k}{2D}-1}\left(\frac{k}{2} - \frac{k}{2D} - i\right)\binom{k}{i} + \sum_{i=\frac{k}{2}+\frac{k}{2D}}^{k}\left(i - \frac{k}{2} - \frac{k}{2D} + 1\right)\binom{k}{i}\right) \\ &= 2D\left(\sum_{i=\frac{k}{2}+\frac{k}{2D}+1}^{k}\left(i - \frac{k}{2} - \frac{k}{2D}\right)\binom{k}{i} + \sum_{i=\frac{k}{2}+\frac{k}{2D}}^{k}\left(i - \frac{k}{2} - \frac{k}{2D} + 1\right)\binom{k}{i}\right) \\ &= 2D\left(\sum_{i=\frac{k}{2}+\frac{k}{2D}}^{k}\left(2\left(i - \frac{k}{2} - \frac{k}{2D}\right) + 1\right)\binom{k}{i}\right) \end{aligned}$$

Now observe that

$$\begin{aligned}\binom{k}{i+1} &= \frac{k-i}{i+1}\binom{k}{i} \leq \frac{k/2-k/2D}{k/2+k/2D}\binom{k}{i}\\ &= \frac{D-1}{D+1}\binom{k}{i},\end{aligned}$$

which implies that

$$\binom{k}{i+t} \leq \left(\frac{D-1}{D+1}\right)^t \binom{k}{i}.$$

It is simple to check that

$$\begin{aligned}\sum_{t=0}^{\infty}(2t+1)r^t &= 2\sum_{t=0}^{\infty}(t+1)r^t - \sum_{t=0}^{\infty} r^t\\ &= \frac{2}{(1-r)^2} - \frac{1}{1-r} = \frac{1+r}{(1-r)^2}.\end{aligned}$$

Substituting $r = (\frac{D-1}{D+1})$, we can see that

$$\begin{aligned}\sum p_z^* &\leq 2D\binom{k}{\frac{k}{2}+\frac{k}{2D}}\sum_{t=0}^{k/2-k/2D}(2t+1)\left(\frac{D-1}{D+1}\right)^t\\ &\leq 2D\binom{k}{k/2+k/2D}\frac{D(D+1)}{2}\\ &= D^2(1+D)\binom{k}{k/2+k/2D}.\end{aligned}$$

$$\begin{aligned}\text{Hence, } \mu &= \frac{2^k}{\sum_z p_z^*} \geq \frac{2^k}{D^2(1+D)\binom{k}{k/2+k/2D}}\\ \mathsf{LP}_{\mathsf{dual}} &= (1+D)\mu \geq \frac{2^k}{D^2\binom{k}{k/2+k/2D}}\end{aligned}$$

Using Stirling's approximation, we get a lower bound of $\Omega((1/D^2)\, 2^{k(1-H(\frac{1}{2}(1+\frac{1}{D})))})$. Note that the number of points $n = \Theta(2^{2k})$. For large constant D, this bound becomes $n^{\Omega(1/D^2)}$. For $D = (1+\varepsilon)$ where ε is small, the bound becomes $n^{\frac{1}{2}-O(\varepsilon\log(1/\varepsilon))}$. This concludes the proof of Theorem 2.1.

We should note that our bounds are presented to make clear that constant distortion embeddings require polynomial dimension. Once $D \approx \sqrt{\log(n)}$ our bound does not say anything interesting. On the other hand, our estimate for $\binom{k}{k/2+k/2D}$ is a strict overestimate, so our bound does indeed hold for any D. We can deduce a few simple corollaries from this.

COROLLARY 2.5. *For any distortion $D = O((\log(n))^{c/2})$ where $c < 1$, at least $2^{\Omega(\log(n)^{1-c})}$ dimensions are required for the recursive diamond graph.*

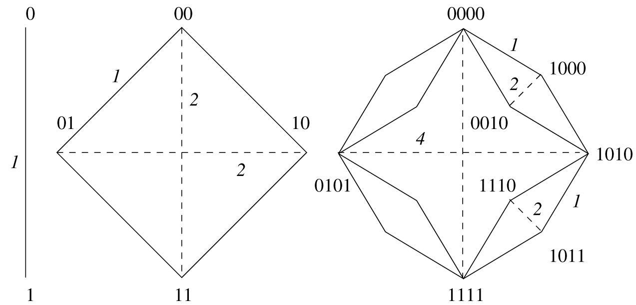

FIG. 4. Using the diamond graph to generate our point set.

In other words, super-logarithmic dimension is needed even when almost $\sqrt{\log(n)}$ distortion is allowed.

On the other hand, the result of Rao [1999] implies a distortion $D = O(\sqrt{\log(n)})$ embedding with dimension $O(\log(n))$ for any planar graph. This, in turn, implies that the dependence on $1/D^2$ in the exponent cannot be improved (say to $1/D$) for planar graphs.

COROLLARY 2.6. *The lower bound on dimension in terms of distortion cannot be improved to $n^{\Omega(1/D^{2-\delta})}$ for any $\delta > 0$ for any family of planar metrics.*

This is easy to see because any bound of this form would imply that embedding planar graphs with distortion $\sqrt{\log(n)}$ would require $2^{\Omega(\log(n)^{\delta/2})}$ dimensions, which is super-logarithmic for any constant $\delta > 0$.

3. *Some ℓ_1 Metrics Require High Dimension*

So far, we have proved that some series-parallel graphs do not admit low distortion, low dimension embeddings. This is in contrast to Gupta et al. [2004] who prove that series-parallel graphs can be embedded into ℓ_1 with constant distortion (with high dimension). We can go one step further and provide a family of point sets native to ℓ_1, which have the same properties as the recursive diamond graph. This gives our final theorem:

THEOREM 3.1. *There are n point metrics in $\ell_1^{O(\sqrt{n})}$, which require $n^{\Omega(1/D^2)}$ dimensions if only D distortion is allowed.*

PROOF. We build our point set with a construction analogous to the construction of the recursive diamond graph (see Figure 4). Let the original edge have end points at 0 and 1. Our "vertices" will be points in $\{0, 1\}^i$ (i.e., vertices of the hamming-cube). To go from level i to level $i + 1$, first double the number of dimensions. The vertices of the parent edge are at the points u and v. Replace them with the points uu (u concatenated with u) and vv. The children will be the points uv and vu. The level-k recursive diamond graph corresponds to a set of $\Theta(4^k)$ points in 2^k dimensions.

CLAIM 5. *Every "edge" in a level k point set has length* 1.

PROOF. We prove this by induction on the level of the point set.

Base Case. The original edge in the point set is between 0 and 1.

Inductive Step. By the inductive hypothesis, the end points u and v of an edge at level i has length 1. The four child edges at level $i+1$ are (uu, uv), (uu, vu), (vv, uv) and (vv, vu). Since $\|u-u\|_1 = \|v-v\|_1 = 0$ and $\|u-v\|_1 = 1$, $\|uu-uv\|_1 = \|uu-vu\|_1 = \|vv-uv\|_1 = \|vv-vu\|_1 = 1$. □

CLAIM 6. *Each diagonal at level i has length 2^{k-i} in the level k point set.*

PROOF. Again, we proceed by induction on the level of the point set.

Base Case. The level 0 diagonal in the level 1 point set is from 01 to 10, which has length 2.

Inductive Step. In the level j graph a level $i < j$ diagonal between points u and v has length 2^{j-i}. In the level $j+1$ graph, these points are replaced with uu and vv. $\|uu-vv\|_1 = 2\|u-v\|_1 = 2^{j+1-i}$. The new diagonals in the level $j+1$ graph are at level j. A given new diagonal with parents u and v has end points uv and vu, and $\|uv-vu\|_1 = 2\|u-v\|_1 = 2 = 2^{j+1-j}$ by Claim 5. □

If we divide all distances by 2^k, this point set has exactly the same "edge lengths" and "diagonal lengths" as the recursive diamond graph. Since our LP constraints only depend on these distances, our lower bound for the recursive diamond graph immediately applies to this point set.

As we noted at the end of Section 2, any distortion $D = (1+\varepsilon)$ (for ε small enough) requires $n^{\frac{1}{2}-O(\varepsilon\log(1/\varepsilon))}$. This is in stark contrast to the case of the Johnson–Lindenstrauss Lemma, where the dependence on ε is only $O(1/\varepsilon^2)$.

3.1. THE POINT SET AS AN EMBEDDING OF THE DIAMOND GRAPH. As a brief aside, we would like to observe that the point set above is in fact an embedding of the diamond graph into ℓ_1 with distortion at most 2 and dimension $\Theta(\sqrt{n})$. We implicitly defined a mapping from points of the diamond graph into points of the Hamming cube by defining the construction of our ℓ_1 point set in analogy to the construction of the diamond graph.

CLAIM 7. *If two points u and v lie on a shortest path from the original top vertex to the original bottom vertex, then the distance between u and v is preserved exactly in this embedding.*

PROOF. Consider every shortest path from the original top vertex (corresponding to $(0, 0, \ldots, 0)$) to the original bottom vertex (now $(1, 1, \ldots, 1)$). By Claim 5, we see that at each step exactly one coordinate switches from 0 to 1 (or vice-versa). Since there are k dimensions and each shortest path from the top to bottom is length exactly k, this implies that at each step along the path exactly one "0" is switched to a "1," and then it is never switched back. This means the length of any path which lies wholly on a single shortest path from the top to the bottom is preserved exactly. □

Let $\diamond_x$ denote the diamond with label x. If u and v do not lie on the same shortest path, there must be some $\diamond_x$ such that u descends (without loss of generality) from the left side of $\diamond_x$ and v from the right.

CLAIM 8. *If u and v descend from opposite sides of a diamond $\diamond_x$ but lie on the same side of diag(x) (the diagonal of $\diamond_x$), then their distance is preserved exactly in the embedding.*

PROOF. Refer to the top vertex of $\diamond_x$ as top($\diamond_x$), and similarly bot($\diamond_x$) is the bottom vertex, left($\diamond_x$) the left vertex and right($\diamond_x$) the right vertex. The vectors for top($\diamond_x$) and bot($\diamond_x$) differ at exactly p positions, where top($\diamond_x$) has $0s$ and bot($\diamond_x$) has $1s$. left($\diamond_x$) and right($\diamond_x$) differ from top($\diamond_x$) and bot($\diamond_x$) in exactly $\frac{p}{2}$ positions (by Claim 7), and from each other in p positions (by Claim 6). Assume (without loss of generality) that u and v are both on the top($\diamond_x$) side of diag(x). Consider any shortest path from left($\diamond_x$) to right($\diamond_x$) that passes through both u and v. Then, at each step from left($\diamond_x$) to top($\diamond_x$) we switch a 1 to a 0, and at each step from top($\diamond_x$) to right($\diamond_x$), we switch a 0 to a 1. But the positions where the 1 to 0 switches occur must be *disjoint* from the 0 to 1 switches, otherwise, there would be a path from left($\diamond_x$) to right($\diamond_x$) shorter than p. Therefore, $\|u - v\|_1 = \|u - \text{top}(\diamond_x)\|_1 + \|\text{top}(\diamond_x) - v\|_1 = \Delta_G(u, \text{top}(\diamond_x)) + \Delta_G(\text{top}(\diamond_x), v)$ (where Δ_G is the distance in the graph), showing that the distance is preserved exactly. $\square$

There is only one remaining case.

CLAIM 9. *If u and v descend from opposite sides of a diamond $\diamond_x$, and they also lie on opposite sides of diag(x), their distance is preserved up to a factor* 2.

PROOF. In this case, $\frac{\Delta_G(\text{top}(\diamond_x),\text{bot}(\diamond_x))}{2} \leq \Delta_G(u, v) \leq \Delta_G(\text{top}(\diamond_x), \text{bot}(\diamond_x))$. Now assume (without loss of generality) that u is below left($\diamond_x$) and v is above right($\diamond_x$). u must have $1s$ at any position where left($\diamond_x$) has $1s$, and v must have $0s$ at any position where right($\diamond_x$) has $0s$ (by Claim 7). Also note that left($\diamond_x$) and right($\diamond_x$) have the same number of $1s$ and $0s$. This means there are $\frac{p}{2} = \frac{\Delta_G(\text{left}(\diamond_x),\text{right}(\diamond_x))}{2} = \frac{\Delta_G(\text{top}(\diamond_x),\text{bot}(\diamond_x))}{2}$ positions where left($\diamond_x$) has $1s$ and right($\diamond_x$) has $0s$. Since these positions cannot change for u and v, $\|u - v\|_1 \geq \frac{\Delta_G(\text{top}(\diamond_x),\text{bot}(\diamond_x))}{2} \geq \frac{\Delta_G(u,v)}{2}$. $\square$

We conclude that this is indeed a distortion 2 embedding for the recursive diamond graph of k levels with dimension $d = 2^k \approx \sqrt{\frac{3}{2}n}$ where n is the number of vertices.

4. *Conclusions*

We have given the first proof that some point sets in ℓ_1 require a polynomial number of dimensions if only constant distortion is allowed. Our results show the following lower bounds on the dimension–distortion tradeoff for ℓ_1: (1) for any distortion D, the number of dimensions $d = n^{\Omega(1/D^2)}$, and (2) for $D = (1 + \varepsilon)$, $d = n^{\frac{1}{2} - O(\varepsilon \log(\frac{1}{\varepsilon}))}$. For distortion $(1 + \varepsilon)$, the best upper bound on the number of dimensions is $O(n \log n)$ [Schechtman 1987; Talagrand 1990]. Our lower bound for very small ε may be weak because the set of n points that we use embeds isometrically into $O(\sqrt{n})$ dimensions. It may be possible to improve this aspect of the lower bound using a different construction.

A more interesting question is whether our lower bound $n^{\Omega(1/D^2)}$ can be improved, say to $n^{\Omega(1/D)}$. This is connected to the following question: How much distortion is required to embed n points in ℓ_1 into ℓ_2? The current upper bound is $O(\log n)$ (by Bourgain [1985]), while the best lower bound is $\Omega(\sqrt{\log n})$ (e.g., the Hamming cube with $\log n$ dimensions). It is known (see Linial et al. [1995]) that n points in ℓ_2 can be embedded into $\ell_1^{O(\log n)}$ with distortion $(1 + \varepsilon)$. If n points in ℓ_1 can be embedded

TABLE V. PRIMAL1: THE LINEAR PROGRAM WITH A COMBINED CONSTRAINT ($\lambda, \lambda_0, \dots \lambda_{k-1}, \gamma$ ARE CONSTANTS)

min s			
$D\left(\lambda\overline{d_*} + \sum_{i=0}^{k-1}\lambda_i \sum_{y\in\{0,1\}^i}\overline{d_y}\right) - \gamma\sum_{z\in\{0,1\}^k}\overline{e_z}$		$\geq \lambda + \sum_{i=0}^{k-1}\lambda_i - \gamma$	$[\mu]$
$\forall z \in \{0,1\}^k$	$\frac{s}{2^k} - \overline{e_z}$	≥ 0	$[p_z]$
$\forall z \in \{0,1\}^k$	$\overline{e_z} + \left(\frac{\overline{d_*}}{2^k} + \sum_{y\sqsubset z}\frac{S(z_{\lvert y\rvert+1})\overline{d_y}}{2^{k-\lvert y\rvert}}\right)$	≥ 0	$[\alpha_z]$
$\forall z \in \{0,1\}^k$	$\overline{e_z} - \left(\frac{\overline{d_*}}{2^k} + \sum_{y\sqsubset z}\frac{S(z_{\lvert y\rvert+1})\overline{d_y}}{2^{k-\lvert y\rvert}}\right)$	≥ 0	$[\beta_z]$

TABLE VI. PRIMAL2: THE LINEAR PROGRAM WITH SEPARATE CONSTRAINTS ($\lambda, \lambda_0, \dots \lambda_{k-1}, \gamma$ ARE DUAL VARIABLES)

min s			
	$\sum_{z\in\{0,1\}^k}\overline{e_z}$	≤ 1	$[\gamma]$
	$D\overline{d_*}$	≥ 1	$[\lambda]$
$\forall i \in \{0,\dots,k-1\}$	$D\sum_{y\in\{0,1\}^i}\overline{d_y}$	≥ 1	$[\lambda_i]$
$\forall z \in \{0,1\}^k$	$\frac{s}{2^k} - \overline{e_z}$	≥ 0	$[p_z]$
$\forall z \in \{0,1\}^k$	$\overline{e_z} + \left(\frac{\overline{d_*}}{2^k} + \sum_{y\sqsubset z}\frac{S(z_{\lvert y\rvert+1})\overline{d_y}}{2^{k-\lvert y\rvert}}\right)$	≥ 0	$[\alpha_z]$
$\forall z \in \{0,1\}^k$	$\overline{e_z} - \left(\frac{\overline{d_*}}{2^k} + \sum_{y\sqsubset z}\frac{S(z_{\lvert y\rvert+1})\overline{d_y}}{2^{k-\lvert y\rvert}}\right)$	≥ 0	$[\beta_z]$

into ℓ_2 with distortion D, this would give ℓ_1 dimension reduction down to $O(\log n)$ dimensions with distortion $D(1+\varepsilon)$. Note that $D = O(\sqrt{\log n})$ corresponds to a dimension-distortion tradeoff of $d = n^{O((\log D)/D^2)}$. We know that the recursive diamond graphs, and in fact all planar graphs, embed in ℓ_2 with distortion $O(\sqrt{\log n})$ by the result of Rao [1999]. If we wish to significantly improve the $n^{\Omega(1/D^2)}$ lower bound, we will need to study nonplanar graphs.

It would be very interesting to devise dimension reduction schemes for ℓ_1 that require $d = n^{f(D)}$ dimensions in order to guarantee distortion at most D for broader classes of metrics. Currently, we know of no nontrivial dimension–distortion tradeoffs for series-parallel graphs, planar graphs, or general finite metric spaces. Our research suggests one possible avenue for progress in this direction. Our lower bounds came from the dual solutions of certain LPs: The primal solutions to these LPs give stretch limited embeddings for the recursive diamond graph. Studying these embeddings may lead to dimension–distortion tradeoffs for series-parallel graphs in ℓ_1.

Appendix

A. *Determining Weights for Individual Constraints via Duality*

Consider the LP we used for minimizing the stretch of a distortion D embedding subject to the derived constraint on edges and diagonal lengths. The LP is presented in Table V. In contrast to the LP we used in the proof in the main text, this LP uses a more general form of the distortion constraint with weights $\lambda, \lambda_0, \dots, \lambda_{k-1}, \gamma$ for the bounds on the average diagonal lengths and average edge lengths. We describe a technique to determine the *optimal* values for these weights.

In order to do this, we consider instead, a closely related LP (see Table VI) in which the combined constraint is replaced by individual constraints for the average edge length and average diagonal lengths at each level. A priori, it is not clear

TABLE VII. DUAL2: THE DUAL LINEAR PROGRAM FOR THE PRIMAL WITH SEPARATE CONSTRAINTS ($\lambda, \lambda_0, \dots \lambda_{k-1}, \gamma$ ARE VARIABLES)

max $\lambda + \sum_{i=0}^{k-1} \lambda_i - \gamma$			
$\forall z \in \{0,1\}^k$	$-\gamma - p_z + \alpha_z + \beta_z$	≤ 0	$[\overline{e_z}]$
	$\sum_{z\in\{0,1\}^k} p_z$	$\le 2^k$	$[s]$
$\forall y \in \bigcup_{i\in[0,k-1]}\{0,1\}^i$	$D\lambda_i + \sum_{v\in\{0,1\}^{k-\lvert y\rvert}} \frac{S((yv)_{\lvert y\rvert+1})(\alpha_{yv}-\beta_{yv})}{2^{k-\lvert y\rvert}}$	≤ 0	$[\overline{d_y}]$
	$D\lambda + \sum_{z\in\{0,1\}^k} \frac{\alpha_z-\beta_z}{2^k}$	≤ 0	$[\overline{d_*}]$

TABLE VIII. DUAL1: THE DUAL LINEAR PROGRAM FOR THE COMBINED CONSTRAINT PRIMAL ($\lambda, \lambda_0, \dots, \lambda_{k-1}, \gamma$ ARE CONSTANTS, μ IS A VARIABLE)

max $\left(\lambda + \sum_{i=0}^{k-1} \lambda_i - \gamma\right)\mu$			
$\forall z \in \{0,1\}^k$	$-\gamma\mu - p_z + \alpha_z + \beta_z$	≤ 0	$[\overline{e_z}]$
	$\sum_{z\in\{0,1\}^k} p_z$	$\le 2^k$	$[s]$
$\forall y \in \bigcup_{i\in[0,k-1]}\{0,1\}^i$	$D\lambda_i\mu + \sum_{v\in\{0,1\}^{k-\lvert y\rvert}} \frac{S((yv)_{\lvert y\rvert+1})(\alpha_{yv}-\beta_{yv})}{2^{k-\lvert y\rvert}}$	≤ 0	$[\overline{d_y}]$
	$D\lambda\mu + \sum_{z\in\{0,1\}^k} \frac{\alpha_z-\beta_z}{2^k}$	≤ 0	$[\overline{d_*}]$

whether the bound on stretch obtained by this LP is a valid lower bound on the stretch for a D distortion embedding. This is because our argument of the validity of the previous LP used the fact that we had a single combined constraint. This allowed us to conclude that there exists a single stretch-limited line embedding that satisfies this constraint. Nevertheless, we will prove that the bound produced by this new LP can be obtained from the previous LP by setting the weights appropriately. (The use of $\lambda, \lambda_0, \dots, \lambda_{k-1}, \gamma$ for weights in Primal1 as well as dual variables for the constraints in Primal2 is deliberate.)

Consider now the dual to Primal2, given in Table VII. We look at any feasible solution to Dual2 and prove that this is a valid lower bound on the value of Primal1. In order to do this, we look at the values of the dual variables $\lambda, \lambda_0, \dots, \lambda_{k-1}, \gamma$ and use them as the values of the weights in Primal1.

Consider the dual to Primal1, given in Table VIII. Note that $\lambda, \lambda_0, \dots, \lambda_{k-1}, \gamma$ are constants whose values are the same as the values of the corresponding variables in a specific feasible solution to Dual2. We claim that there exists a solution to Dual1 whose value is equal to the feasible solution to Dual2. In order to see this, we simply set $\mu = 1$ and use the same values for the rest of the variables as in the feasible solution to Dual2. It is easy to see that all the feasibility constraints are satisfied and the value of the two solutions is identical. This implies that Dual2 actually gives a lower bound on Primal1 for an appropriate setting of weights $\lambda, \lambda_0, \dots, \lambda_{k-1}, \gamma$; moreover these weights are simply the values of dual variables in Dual2. This is, in fact, how we determined the weights used in the LP we presented in the main text.

ACKNOWLEDGMENTS. We would like to thank the anonymous referees for their helpful comments.

REFERENCES

ACHLIOPTAS, D. 2001. Database-friendly random projections. In *Proceedings of 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. ACM, New York.

ALON, N., MATIAS, Y., AND SZEGEDY, M. 1999. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci. 58*, 1, 137–147.

ARORA, S., AND KANNAN, R. 2001. Learning mixtures of arbitrary gaussians. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*. ACM, New York, 247–257.

ARRIAGA, R., AND VEMPALA, S. 1999. An algorithmic theory of learning: Robust concepts and random projection. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 616–623.

BALL, K. 1990. Isometric embedding in l_p-spaces. *Europ. J. Combinat. 11*, 305–311.

BOURGAIN, J. 1985. On Lipschitz embedding of finite metric spaces into Hilbert space. *Isr. J. Math. 52*, 46–52.

CHARIKAR, M., AND SAHAI, A. 2002. Dimension reduction in the ℓ_1 norm. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 551–560.

DASGUPTA, S. 1999. Learning mixtures of Gaussians. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 634–644.

DASGUPTA, S., AND GUPTA, A. 1999. An elementary proof of the Johnson–Lindenstrauss lemma. Tech. Rep. TR-99-06, International Computer Science Institute, UC Berkeley.

DE REYNA, J. A., AND RODRIGUEZ-PIAZZA, L. 1992. Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces. *Isr. J. Math. 79*, 103–111.

DEZA, M., AND LAURENT, M. 1997. *Geometry of Cuts and Metrics*. Springer-Verlag, New York.

FRANKL, P., AND MAEHARA, H. 1988. The Johnson–Lindenstrauss lemma and the sphericity of some graphs. *J. Combinat. Theory, Ser. B 44*, 355–362.

GUPTA, A., NEWMAN, I., RABINOVICH, Y., AND SINCLAIR, A. 2004. Cuts, trees and ℓ_1 embeddings of graphs. *Combinatorica 24*, 2 (Apr.), 233–269.

INDYK, P. 2000a. Dimensionality reduction techniques for proximity problems. In *Proceedings of the 9th Annual ACM/SIAM Symposium on Discrete Algorithms*. ACM, New York, 371–378.

INDYK, P. 2000b. Stable distributions, pseudorandom generators, embeddings and data stream computation. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 189–197.

INDYK, P. 2001. Algorithmic applications of low-distortion embeddings. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 10–33.

INDYK, P., AND MOTWANI, R. 1998. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*. ACM, New York, 604–613.

JOHNSON, W., AND LINDENSTRAUSS, J. 1984. Extensions of Lipschitz mapping into Hilbert space. *Contemp. Math. 26*, 189–206.

KUSHILEVITZ, E., OSTROVSKY, R., AND RABANI, Y. 2000. Efficient search for approximate nearest neighbor in high-dimensional spaces. *SIAM J. Comput. 30*, 2, 457–474.

LEE, J., AND NAOR, A. 2004. Embedding the diamond graph in l_p and dimension reduction in l_1. *Geomet. Funct. Anal. 14*, 4 (Aug.), 745–474.

LINIAL, N. 2002. Finite metric spaces—combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians III*. 573–586.

LINIAL, N., LONDON, E., AND RABINOVICH, Y. 1995. The geometry of graphs and some of its algorithmic applications. *Combinatorica 15*, 215–245.

MATOUŠEK, J. 1996. On the distortion required for embedding finite metric spaces into normed spaces. *Isr. J. Math. 93*, 333–344.

MATOUŠEK, J. 2002. Chapter 15: Embedding finite metric spaces into euclidean spaces. *Lectures on Discrete Geometry*. Graduate Texts in Mathematics, vol. 212. Springer-Verlag.

NEWMAN, I., AND RABINOVICH, Y. 2002. A lower bound on the distortion of embedding planar metrics into Euclidean space. In *Proceedings of the 18th Annual Symposium on Computational Geometry*. 94–96.

OSTROVSKY, R., AND RABANI, Y. 2000. Polynomial time approximation schemes for geometric k-clustering. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 349–358.

RAO, S. 1999. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*. 300–306.

SCHECHTMAN, G. 1987. More on embedding subspaces of L_p in ℓ_r^n. *Composi. Math. 61*, 2, 159–169.

TALAGRAND, M. 1990. Embedding subspaces of L_1 into ℓ_1^n. *Proceedings of the American Mathetmatical Society*, *vol. 108,* 2, 363–369.

RECEIVED APRIL 2004; REVISED MAY 2005; ACCEPTED MAY 2005