Measured descent: A new embedding method for finite metrics

Robert Krauthgamer* James R. Lee† Manor Mendel‡ Assaf Naor§

April 20, 2004

Abstract

We devise a new embedding technique, which we call *measured descent*, based on decomposing a metric space locally, at varying speeds, according to the density of some probability measure. This provides a refined and unified framework for the two primary methods of constructing Fréchet embeddings for finite metrics, due to [Bourgain, 1985] and [Rao, 1999]. We prove that any n-point metric space (X, d) embeds in Hilbert space with distortion $O(\sqrt{\alpha_X \cdot \log n})$, where α_X is a geometric estimate on the decomposability of X. As an immediate corollary, we obtain an $O(\sqrt{(\log \lambda_X) \log n})$ distortion embedding, where λ_X is the doubling constant of X. Since $\lambda_X \le n$, this result recovers Bourgain's theorem, but when the metric X is, in a sense, "low-dimensional," improved bounds are achieved.

Our embeddings are volume-respecting for subsets of arbitrary size. One consequence is the existence of $(k, O(\log n))$ volume-respecting embeddings for all $1 \le k \le n$, which is the best possible, and answers positively a question posed in [Feige, 1998]. Our techniques are also used to answer positively a question of Y. Rabinovich, showing that any weighted n-point planar graph embeds in $\ell_\infty^{O(\log n)}$ with $O(1)$ distortion. The $O(\log n)$ bound on the dimension is optimal, and improves upon the previously known bound of $O((\log n)^2)$.

1 Introduction

The theory of low-distortion embeddings of finite metric spaces into normed spaces has attracted a lot of attention in recent decades, due to its intrinsic geometric appeal, as well as its applications in Computer Science. A major driving force in this research area has been the quest for analogies between the theory of finite metric spaces and the local theory of Banach spaces. While being very successful, this point of view did not always result in satisfactory metric analogues of basic theorems from the theory of finite-dimensional normed spaces. An example of this is Bourgain's embedding theorem [Bou85], the forefather of modern embedding theory, which states that every n-point metric space embeds into a Euclidean space with distortion $O(\log n)$. This upper bound on the distortion is known to be optimal [LLR95]. Taking the point of view that $\log n$ is a substitute for the dimension of an n-point metric space (see [Bou85]; this approach is clearly natural when applied to a net in the unit ball of some normed space), an analogue of John's theorem [Joh48]

*IBM Almaden Research Center, San Jose, CA 95120, USA. Email: robi@almaden.ibm.com

†Computer Science Division, University of California, Berkeley, CA 94720. Supported by NSF grant CCR-0121555 and an NSF Graduate Research Fellowship. Email: jrl@cs.berkeley.edu

‡Seibel Center for Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. Email: mendelma@uiuc.edu

§Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA. Email: anaor@microsoft.com

would assert that n-point metrics embed into Hilbert space with distortion $O(\sqrt{\log n})$. As this is not the case, the present work is devoted to a more refined analysis of the Euclidean distortion of finite metrics, and in particular to the role of a metric notion of dimension.

We introduce a new embedding method, called *measured descent*, which unifies and refines the known methods of Bourgain [Bou85] and Rao [Rao99] for constructing Fréchet-type embeddings (i.e. embeddings where each coordinate is proportional to the distance from some subset of the metric space). Our method yields an embedding of any n-point metric space X into ℓ_2 with distortion $O(\sqrt{\alpha_X \log n})$, where α_X is a geometric estimate on the decomposability of X (see Definition 1.3 for details). As $\alpha_X \le O(\log n)$, we obtain a refinement of Bourgain's theorem, and when α_X is small (which includes several important families of metrics) improved distortion bounds are achieved. This technique easily generalizes to produce embeddings which preserve higher dimensional structures (i.e. not just distances between pairs of points). For instance, our embeddings can be made volume-respecting in the sense of Feige (see Section 1.2), and hence we obtain optimal volume-respecting embeddings for arbitrary n-point spaces.

Applications. In recent years, metric embedding has become a frequently used algorithmic tool. For example, embeddings into normed spaces have found applications to approximating the sparsest cut of a graph [LLR95, AR98, ARV04] and the bandwidth of a graph [Fei00, DV01], and to distance labeling schemes (see e.g. [Ind01, Sec. 2.2]). The embeddings introduced in this paper refine our knowledge on these problems, and in some cases improve the known algorithmic results. For instance, they immediately imply an improved approximate max-flow/min-cut theorem (and algorithm) for graphs excluding a fixed minor, an improved algorithm for approximating the bandwidth of graphs whose metric has a small doubling constant, and so forth.

1.1 Notation

Let (X, d) be an n-point metric space. We denote by $B(x, r) = \{y \in X : d(x, y) < r\}$ the open ball of radius r about x. For a subset $S \subseteq X$, we write $d(x, S) = \min_{y \in S} d(x, y)$, and define $\mathrm{diam}(S) = \max_{x,y \in S} d(x, y)$. We recall that the *doubling constant* of X, denoted λ_X, is the least value λ such that every ball in X can be covered by λ balls of half the radius [Lar67, Ass83, Luu98, Hei01]. We say that a measure μ on X is *non-degenerate* if $\mu(x) > 0$ for all $x \in X$. For a non-degenerate measure μ on X define $\Phi(\mu) = \max_{x \in X} \mu(X)/\mu(x)$ to be the *aspect ratio* of μ.

Let (X, d_X) and (Y, d_Y) be metric spaces. A mapping $f : X \to Y$ is called C-Lipschitz if $d_Y(f(x), f(y)) \le C \cdot d_X(x, y)$ for all $x, y \in X$. The mapping f is called K-bi-Lipschitz if there exists a $C > 0$ such that

$$CK^{-1} \cdot d_X(x, y) \le d_Y(f(x), f(y)) \le C \cdot d_X(x, y),$$

for all $x, y \in X$. The least K for which f is K-bi-Lipschitz is called the distortion of f, and is denoted $\mathrm{dist}(f)$. The least distortion with which X may be embedded in Y is denoted $c_Y(X)$. When $Y = L_p$ we use the notation $c_Y(\cdot) = c_p(\cdot)$. Finally, the parameter $c_2(X)$ is called the Euclidean distortion of X.

Metric Decomposition. Let (X, d) be a finite metric space. Given a partition $P = \{C_1, \ldots, C_m\}$ of X, we refer to the sets C_i as *clusters*. We write $\mathcal{P}_X$ for the set of all partitions of X. For $x \in X$ and a partition $P \in \mathcal{P}_X$ we denote by $P(x)$ the unique cluster of P containing x. Finally, the set of all probability distributions on $\mathcal{P}_X$ is denoted $\mathcal{D}_X$.

Definition 1.1 (Padded decomposition). *A (stochastic) decomposition of a finite metric space* (X, d) *is a distribution* $\Pr \in \mathcal{D}_X$ *over partitions of* X*. Given* $\Delta > 0$ *and* $\varepsilon : X \to (0, 1]$*, a* Δ*-bounded* ε*-padded decomposition is one which satisfies the following two conditions.*

1. *For all* $P \in \text{supp}(\Pr)$*, for all* $C \in P$*,* $\text{diam}(C) \leq \Delta$*.*
2. *For all* $x \in X$*,* $\Pr[B(x, \varepsilon(x)\Delta) \nsubseteq P(x)] \leq \frac{1}{2}$*.*

We will actually need a collection of such decompositions, with the diameter bound $\Delta > 0$ ranging over all integral powers of 2 (of course the value 2 is arbitrary).

Definition 1.2 (Decomposition bundle). *Given a function* $\varepsilon : X \times \mathbb{Z} \to (0, 1]$*, an* ε*-padded decomposition bundle on* X *is a function* $\beta : \mathbb{Z} \to \mathcal{D}_X$*, where for every* $u \in \mathbb{Z}$*,* $\beta(u)$ *is a* 2^u*-bounded* $\varepsilon(\cdot, u)$*-padded stochastic decomposition of* X*.*

Finally, we associate to every finite metric space an important "decomposability" parameter α_X. (See [LN04] for relationships to another notions of decomposability.)

Definition 1.3 (Modulus of padded decomposability). *The modulus of padded decomposability of a finite metric space* (X, d) *is defined as*

$$\alpha_X = \inf\{\alpha : \textit{there exists an } \varepsilon\textit{-padded decomposition bundle on } X \textit{ with } \varepsilon(x, u) \equiv 1/\alpha\}.$$

It is known that $\alpha_X = O(\log n)$ [LS93, Bar96], and furthermore $\alpha_X = O(\log \lambda_X)$ [GKL03]. Additionally, if X is the shortest-path metric on an n-point constant-degree expander, then $\alpha_X = \Omega(\log n)$ [Bar96]. For every metric space X induced by an edge-weighted graph which excludes $K_{r,r}$ as a minor, it is shown in the sequence of papers [KPR93, Rao99, FT03] that $\alpha_X = O(r^2)$.

Volume-respecting embeddings. We recall the notion of *volume-respecting embeddings*, which was introduced by Feige [Fei00] as a tool in his study of the graph bandwidth problem. Let $S \subseteq X$ be a k-point subset of X. We define its volume by

$$\text{vol}(S) = \sup\{\text{vol}_{k-1}(\text{conv}(f(S))) : \ f : S \to L_2 \text{ is 1-Lipschitz}\},$$

where for $A \subseteq L_2$, $\text{conv}(A)$ denotes its convex hull and the $(k-1)$-dimensional volume above is computed with respect to the Euclidean structure induced by L_2. A mapping $f : X \to L_2$ is called (k, η)-volume-respecting if it is 1-Lipschitz and for every k-point subset $S \subset X$,

$$\left[\frac{\text{vol}(S)}{\text{vol}_{k-1}(\text{conv}(f(S)))}\right]^{\frac{1}{k-1}} \leq \eta.$$

It is easy to see that a 1-Lipschitz map $f : X \to L_2$ has distortion D if and only if it is $(2, D)$-volume-respecting. Thus the volume-respecting property is a generalization of distortion to larger subsets.

1.2 Results

The following theorem refines Bourgain's result in terms of the decomposability parameter.

Theorem 1.4 (Padded embedding theorem). *For every n-point metric space (X, d), and every $1 \le p \le \infty$,*

$$c_p(X) \le O(\alpha_X^{1-1/p}(\log n)^{1/p}). \tag{1}$$

The proof appears in Sections 1.3 and 2. Since $\alpha_X = O(\log \lambda_X)$, it implies in particular that $c_2(X) \le O\big(\sqrt{(\log \lambda_X) \cdot \log |X|}\big)$, for any metric space X. This refines Bourgain's embedding theorem [Bou85], and improves upon previous embeddings of doubling metrics [GKL03]. It is tight for $\lambda_X = O(1)$ [Laa02, LP01, GKL03], and $\lambda_X = n^{\Omega(1)}$ [LLR95]. The question of whether this bound is tight up to a constant factor for the range $\lambda_X \in \{c_1, \dots, |X|^{c_2}\}$, where $c_1 \in \mathbb{N}$, $0 < c_2 < 1$ are some constants, is an interesting open problem.

For $1 \le p < 2$, the bound $O(\sqrt{\alpha_X \log n})$ is better than (1), and thus, in these cases, it makes sense to construct the embedding first into L_2.

A more careful analysis of the proof of Theorem 1.4 yields the following result, proved in Section 2.2, which answers a question posed by Feige in [Fei00].

Theorem 1.5 (Optimal volume-respecting embeddings). *Every n-point metric space X admits an embedding into L_2 which is $(k, O(\sqrt{\alpha_X \log n}))$ volume-respecting for every $2 \le k \le n$.*

Since $\alpha_X = O(\log n)$, this provides $(k, O(\log n))$-volume-respecting embeddings for every $2 \le k \le n$. This is optimal; a matching lower bound is given in [KLM04] for all $k < n^{1/3}$. We note that the previous best bounds were due to Feige [Fei00], who showed that a variant of Bourgain's embedding achieves distortion $O(\sqrt{\log n} \cdot \sqrt{\log n + k \log k})$ (note that this is $\Omega(\sqrt{n})$ for large values of k), and to Rao who showed that $O((\log n)^{3/2})$ volume distortion is achievable for all $1 \le k \le n$ (this follow indirectly from [Rao99], and was first observed in [Gup00]).

This also improves the dependence on r in Rao's volume-respecting embeddings of $K_{r,r}$-excluded metrics, from $(k, O(r^2\sqrt{\log n}))$, due to [Rao99, FT03], to $(k, O(r\sqrt{\log n}))$.[1] As a corollary, we obtain an improved $O(r\sqrt{\log n})$- approximate max-flow/min-cut algorithm for graphs which exclude $K_{r,r}$ as a minor.

ℓ_∞ embeddings. It is not difficult to see that every n-point metric space (X, d) embeds isometrically into ℓ_∞^n via the map $y \mapsto \{d(x, y)\}_{x \in X}$. And for some spaces, like the shortest-path metrics on expanders, or on the $\log n$-dimensional hypercube (see, e.g. [LMN04]), it is known that $n^{\Omega(1)}$ dimensions are required to obtain any map with $O(1)$ distortion. On the other hand, a simple variant of Rao's embedding shows that every planar metric $O(1)$-embeds into $\ell_\infty^{O((\log n)^2)}$. Thus the dimension required to embed a family of metrics into ℓ_∞ with low distortion is a certain measure of the family's complexity (see [Mat02a]).

In Section 3 we use a refinement of measured descent to prove the following theorem, which answers positively a question posed by Y. Rabinovich [Mat02b], and improves Rao's result to obtain the optimal bound.

Theorem 1.6. *Let X be an n-point edge-weighted planar graph, equipped with the shortest path metric. Then X embeds into $\ell_\infty^{O(\log n)}$ with $O(1)$ distortion.*

The $O(\log n)$ bound on the dimension is clearly optimal (by simple volume arguments). Furthermore, this result is stronger than the $O(\sqrt{\log n})$ distortion bound on Euclidean embeddings of planar metrics, due to Rao [Rao99]. The embedding is produced by "derandomizing" both the

[1] This bound is tight for the path even for $k = 3$, see [DV01, KLM04].

decomposition bundle of [Rao99, KPR93] and the proof of measured descent (applied to this special decomposition bundle).

1.3 Outline of Techniques

The following lemma is based on a decomposition of [CKR01], with the improved analysis of [FHRT03, FRT03]. The extension to general measures was observed in [LN03]. Since this lemma is central to our techniques, its proof is presented in Section 2 for completeness.

Lemma 1.7. *Let (X, d) be a finite metric space and let μ be any non-degenerate measure on X. Then there exists an $\varepsilon(x, u)$-padded decomposition bundle on X where*

$$\varepsilon(x, u) = \left[16 + 16 \log \frac{\mu(B(x, 2^u))}{\mu(B(x, 2^{u-3}))}\right]^{-1}. \tag{2}$$

Remark 1.1. *In [VK87] it was shown that X admits a doubling measure, i.e. a non-degenerate measure μ such that for every $x \in X$ and every $r > 0$ we have $\frac{\mu(B(x,2r))}{\mu(B(x,r))} = \lambda_X^{O(1)}$. We thus recover the fact, first proved in [GKL03], that for every metric space X, $\alpha_X = O(\log \lambda_X)$. In particular, for every d-dimensional normed space Y, $\alpha_Y = O(d)$. In [CCG$^+$98], it is argued that $\alpha_Y = \Omega(d)$ when $Y = \ell_1^d$. The same lower bound was shown to hold for every d dimensional normed space Y in [LN04].*

The main embedding lemma. Let (X, d) be a finite metric space, and for $\varepsilon : X \times \mathbb{Z} \to \mathbb{R}$ define for all $x, y \in X$,

$$\delta_\varepsilon(x, y) = \min \left\{ \varepsilon(x, u) : \ u \in \mathbb{Z} \ \text{ and } \ \frac{d(x, y)}{32} \le 2^u \le \frac{d(x, y)}{2} \right\}.$$

Given a non-degenerate measure μ on X denote for $x, y \in X$:

$$V_\mu(x, y) = \max \left\{ \log \frac{\mu(B(x, 2d(x, y)))}{\mu(B(x, d(x, y)/512))}, \log \frac{\mu(B(y, 2d(x, y)))}{\mu(B(y, d(x, y)/512))} \right\}. \tag{3}$$

In what follows we use the standard notation c_{00} for the space of all finite sequences of real numbers. The following result is the main embedding lemma of this paper.

Lemma 1.8 (Main embedding lemma). *Let X be an n-point metric space, μ a non-degenerate measure on X, and $\beta : \mathbb{Z} \to \mathcal{D}_X$ an $\varepsilon(x, u)$-padded decomposition bundle on X. Then there exists a map $\varphi : X \to c_{00}$ such that for every $1 \le p \le \infty$ and for all distinct $x, y \in X$,*

$$[V_\mu(x, y)]^{1/p} \cdot \min\{\delta_\varepsilon(x, y), \delta_\varepsilon(y, x)\} \le \frac{||\varphi(x) - \varphi(y)||_p}{d(x, y)} \le C \left[\log \Phi(\mu)\right]^{1/p}.$$

Here C is a universal constant.

An informal description. Lemma 1.8 is proved in Section 2.1; here we sketch the main ideas. For simplicity, assume that μ is the counting measure, i.e. $\mu(S) = |S|$ for $S \subseteq X$. We also assume, in the informal discussion that follows, that the ratio of the largest to smallest distance in all n-point metric spaces considered is at most n. It follows that the number of values of $i \in \mathbb{Z}$ for which there is a pair x, y with $d(x,y) \in [2^i, 2^{i+1})$ is only $O(\log n)$.

In [Rao99], it is shown that the shortest-path metric on an unweighted n-point planar graph always admits a distortion $O(\sqrt{\log n})$ embedding into a Euclidean space. For each scale $2^i, i \in \mathbb{Z}$, an embedding is constructed by first partitioning the space into pieces of diameter at most 2^i, and then by mapping each point to its distance to the boundary of the partition (such a map is, necessarily, 1-Lipschitz). The partitioning is done randomly, and using a structural theorem of [KPR93], it is shown that the expected distance of a point to the boundary is $\Omega(2^i)$. By concatenating together these random maps for each relevant scale (i.e. each relevant value of $2^i, i \in \mathbb{Z}$—there are only $O(\log n)$ of them), it is not difficult to see that a distortion $O(\sqrt{\log n})$ map is obtained. Note here that the concatenation is uniform—equal weight is placed on each scale.

An extension of Rao's technique to general metrics requires a different method of random partitioning, and a significant loss is incurred. Even the optimal scheme can only ensure that (in the worst case) the expected distance from a point to the boundary of the 2^i-partition is $\Omega(2^i/\log n)$. Thus the resulting embedding incurs distortion $O((\log n)^{3/2})$, which is far from optimal.

Our work starts with the observation (made in [FRT03], based on the work of [CKR01]) that there is a randomized partitioning scheme where, at scale 2^i, the expected distance from a point x to the boundary is

$$\Omega(2^i) \cdot \left(1 + \log \frac{|B(x, 2^i)|}{|B(x, \epsilon 2^i)|}\right)^{-1}$$

(for some fixed $\epsilon < 1$ which the reader may ignore; see Lemma 1.7 for details). In the worst case, the logarithm may have value $\Omega(\log n)$, but notice that a point x cannot exhibit the worst case behavior at every scale; indeed, $\sum_{i \in \mathbb{Z}} \log \frac{|B(x,2^i)|}{|B(x,\epsilon 2^i)|} = O(\log n)$. This suggests a non-uniform concatenation of scales, where more "weight" is given to scales where the ratio is large.

Indeed, it seems prudent that for a point $x \in X$, we assign weight $\log \frac{|B(x,2^i)|}{|B(x,\epsilon 2^i)|}$ to scale i, to counterbalance the associated loss in the partitioning scheme (the actual calculation is done near line (7) of Section 2.1). Notice that the total weight used is only $O(\log n)$.

The problem which presents itself is that the required weights for distinct points $x, y \in X$ at a given scale 2^i could be drastically different. The main technical contribution of Section 2 is to overcome this hurdle. It is not difficult to see that the local mass at scale 2^i obeys a certain smoothness property. Intuitively, this is a manifestation of the trivial fact that

$$|B(y, 2^i - d(x,y))| \le |B(x, 2^i)| \le |B(y, 2^i + d(x,y)|.$$

(The actual property we use is contained in Claim 2.1). Thus instead of constructing a global partition according to one scale, it is possible to decompose the space "locally" at varying speeds, according to the local mass ratio. This allows the appropriate weightings to be indirectly applied.

To get a feel for this, the reader may consider the following process. Assume that $\mathrm{diam}(X) = 1$. In Rao's embedding, one may think of the space as being decomposed at a uniform speed, as follows: First one starts with the trivial partition defined as a single set which contains the whole space. Then, we refine this partition so that each piece S in the refinement has $\mathrm{diam}(S) \le \frac{1}{2}$. This refinement process is continued so that at time t, each piece has $\mathrm{diam}(S) \le 2^{-t}$. In contrast, our embedding proceeds as follows. At time t, each piece of the current partition should have mass about $2^{\log n - t}$. Hence at time $t = 0$, there is a single piece of mass n, i.e. the whole set X. As t increases, the mass of the pieces shrinks, but the diameters of the corresponding pieces decrease at a non-uniform rate, according to the "local mass ratio." The effect is that for a point $x \in X$ and a scale 2^i with a relatively high mass ratio, our embedding devotes proportionally more time to "working on" x at that scale.

The actual decomposition process is random, and a reasonable amount of delicateness is needed to maintain the proper correspondence between the mass and diameter of pieces (since the partitioning scheme used, i.e. Lemma 1.7, is defined only in terms of diameter). In particular, we do not actually maintain a partition at time t, but instead a sort of fuzzy superposition of local partitions.

The payoff. Using Lemma 1.8 we are in a position to prove Theorem 1.4. We start with the following simple observation, which bounds $V_\mu(x, y)$ from below, and which will be used several times in what follows.

Lemma 1.9. *Let μ be any non-degenerate measure on X and $x, y \in X$, $x \neq y$. Then*

$$\max\left\{\frac{\mu(B(x, 2d(x, y)))}{\mu(B(x, d(x, y)/2))}, \frac{\mu(B(y, 2d(x, y)))}{\mu(B(y, d(x, y)/2))}\right\} \geq 2.$$

Proof. Assume without loss that $\mu(B(x, 2d(x, y))) \leq \mu(B(y, 2d(x, y)))$. Noticing that the two balls $B(x, d(x, y)/2)$ and $B(y, d(x, y)/2)$ are disjoint, and that both are contained in $B(x, 2d(x, y))$; the proof follows. □

Proof of Theorem 1.4. Fix $p \in [1, \infty]$ and let $\mu = |\cdot|$ be the counting measure on X. Let $\varepsilon(x, u)$ be as in (2), and observe that in this case for all $x, y \in X$ we have $\delta_\varepsilon(x, y) \geq [16 + 16V_\mu(x, y)]^{-1}$. Applying Lemma 1.8 to the decomposition bundle of Lemma 1.7 we get a mapping $\varphi_1 : X \to L_p$ such that for all $x, y \in X$,

$$\frac{[V_\mu(x, y)]^{1/p}}{16 + 16V_\mu(x, y)} \leq \frac{||\varphi_1(x) - \varphi_1(y)||_p}{d(x, y)} \leq C(\log n)^{1/p}.$$

On the other hand, Lemma 1.8 applied to the decomposition bundle ensured by the definition of α_X yields a mapping $\varphi_2 : X \to L_p$ for which

$$\frac{[V_\mu(x, y)]^{1/p}}{\alpha_X} \leq \frac{||\varphi_2(x) - \varphi_2(y)||_p}{d(x, y)} \leq C(\log n)^{1/p}.$$

Finally, for $\varphi = \varphi_1 \oplus \varphi_2$ we have

$$\frac{||\varphi(x) - \varphi(y)||_p^p}{d(x, y)^p} \geq \frac{V_\mu(x, y)}{[16 + 16V_\mu(x, y)]^p} + \frac{V_\mu(x, y)}{\alpha_X^p} \geq \Omega\left(\frac{1}{\alpha_X^{p-1}}\right),$$

where we have used the fact that Lemma 1.9 implies that $V_\mu(x, y) \geq \Omega(1)$. □

2 Measured descent

In this section, we prove the main embedding lemma and exhibit the existence of optimal volume-respecting embeddings. We use decomposition bundles to construct random subsets of X, the distances from which are used as coordinates of an embedding into c_{00}. As the diameters of the decompositions become smaller, our embedding "zooms in" on the increasingly finer structure of the space. Our approach is heavily based on the existence of good decomposition bundles; we thus start by proving Lemma 1.7, which is essentially contained in [FRT03].

Proof of Lemma 1.7. By approximate the values $\{\mu(x)\}_{x\in X}$ by rational numbers and duplicating points, it is straightforward to verify that it is enough to prove the required result for the counting measure on X, i.e. when $\mu(S) = |S|$.

Let $\Delta = 2^u$ for some $u \in \mathbb{Z}$. We now describe the distribution $\beta(u)$. Choose, uniformly at random, a permutation π of X and a value $\alpha \in [\frac{1}{4}, \frac{1}{2}]$. For every point $y \in X$, define a cluster

$$C_y = \{x \in X : x \in B(y, \alpha\Delta) \text{ and } \pi(y) \le \pi(z) \text{ for all } z \in X \text{ with } x \in B(z, \alpha\Delta)\}.$$

In words, a point $x \in X$ is assigned to C_y where y is the minimal point according to π that is within distance $\alpha\Delta$ from x.

Clearly the set $P = \{C_y\}_{y\in X}$ constitutes a partition of X. Furthermore, $C_y \subseteq B(y, \alpha\Delta)$, thus $\operatorname{diam}(C_y) \le \Delta$, so requirement (1) in Definition 1.1 is satisfied for every partition P arising from this process. It remains to prove requirement (2).

Fix a point $x \in X$ and some value $t \le \Delta/8$. Let $a = |B(x, \Delta/8)|$, $b = |B(x, \Delta)|$, and arrange the points $w_1, \dots, w_b \in B(x, \Delta)$ in increasing distance from x. Let $I_k = [d(x, w_k) - t, d(x, w_k) + t]$ and write $\mathcal{E}_k$ for the event that w_k is the minimal element according to π such that $B(x, t) \cap C_{w_k} \neq \emptyset$, and yet $B(x, t) \nsubseteq C_{w_k}$. Note that if $w_k \in B(x, \Delta/8)$, then $\Pr[\mathcal{E}_k] = 0$ since $B(x, t) \subseteq B(x, \Delta/8) \subseteq B(w_k, \Delta/4) \subseteq B(w_k, \alpha\Delta)$. It follows that

$$\begin{aligned} \Pr[B(x,t) \nsubseteq P(x)] = \sum_{k=a+1}^{b} \Pr[\mathcal{E}_k] &= \sum_{k=a+1}^{b} \Pr[\alpha\Delta \in I_k] \cdot \Pr[\mathcal{E}_k \mid \alpha\Delta \in I_k] \\ &\le \sum_{k=a+1}^{b} \frac{2t}{\Delta/4} \cdot \frac{1}{k} \le \frac{8t}{\Delta} \left(1 + \log \frac{b}{a}\right). \end{aligned}$$

Setting $t = \varepsilon(x, u)\Delta \le \Delta/8$, where $\varepsilon(x, u)$ is as in (2), the righthand side is at most $\frac{1}{2}$, proving requirement (2) in Definition 1.1. □

2.1 Proof of main embedding lemma

We first introduce some notation. Define two intervals of integers $I, J \subseteq \mathbb{Z}$ by

$$I = \{-6, -5, -4, -3, -2, -1, 0, 1, 2, 3\} \quad \text{and} \quad J = \{0, 1, \dots, \lceil \log_2 \Phi(\mu) \rceil\}.$$

For $t > 0$ write $\kappa(x, t) = \max\{\kappa \in \mathbb{Z} : \mu(B(x, 2^\kappa)) < 2^t\}$. For each $u \in \mathbb{Z}$ let P_u be chosen according to the distribution $\beta(u)$. Additionally, for $u \in \mathbb{Z}$ let $\{\sigma_u(C) : C \subseteq X\}$ be i.i.d. symmetric $\{0,1\}$-valued Bernoulli random variables. We assume throughout the ensuing argument that the random variables $\{\sigma_u(C) :\ C \subseteq X,\ u \in \mathbb{Z}\}$, $\{P_u :\ u \in \mathbb{Z}\}$ are mutually independent. For every $t \in J$ and $i \in I$ define a random subset $W_t^i \subseteq X$ by

$$W_t^i = \{x \in X :\ \sigma_{\kappa(x,t)-i}(P_{\kappa(x,t)-i}(x)) = 0\}.$$

Our random embedding $f : X \to c_{00}$ is defined by $f(x) = (d(x, W_t^i) :\ i \in I,\ t \in J)$. In the sequel, we assume that $p < \infty$; the case $p = \infty$ follows similarly. Since each of the coordinates of f is Lipschitz with constant 1, we have for all $x, y \in X$,

$$\|f(x) - f(y)\|_p^p \le |I| \cdot |J| \cdot d(x, y)^p \le 50[\log \Phi(\mu)]\, d(x, y)^p. \tag{4}$$

The proof will be complete once we show that for all $x, y \in X$

$$\mathbb{E}\|f(x)-f(y)\|_p^p \geq [\Omega(\min\{\delta_\varepsilon(x,y),\delta_\varepsilon(y,x)\}\cdot d(x,y))]^p \cdot V_\mu(x,y). \tag{5}$$

Indeed, denote by $(T, \Pr)$ the probability space on which the above random variables are defined, and consider the space $L_p(T, c_{00})$, i.e. the space of all c_{00} valued random variables ζ on T equipped with the L_p norm $\|\zeta\|_p = (\mathbb{E}\|\zeta\|_p^p)^{1/p}$. Equations (4) and (5) show that the mapping $x \mapsto f(x)$ is the required embedding of X into $L_p(T, c_{00})$. Observe that all the distributions are actually finitely supported, since X is finite, so that this can still be viewed as an embedding into c_{00}. See Remark 2.2 below for more details.

To prove (5) fix $x, y \in X$, $x \neq y$. Without loss of generality we may assume that the maximum in (3) is attained by the first term, namely, $\frac{\mu(B(x,2d(x,y)))}{\mu(B(x,d(x,y)/512))} \geq \frac{\mu(B(y,2d(x,y)))}{\mu(B(y,d(x,y)/512))}$. Using Lemma 1.9, it immediately follows that

$$\frac{\mu(B(x,2d(x,y)))}{\mu(B(x,d(x,y)/512))} \geq 2. \tag{6}$$

Setting $R = \frac{1}{4}d(x,y)$, denote for $i \in \mathbb{Z}$, $s_i = \log_2 \mu(B(x, 2^i R))$. We next extend some immediate bounds on $\kappa(x,t)$ (in terms of R) to any nearby point $z \in B(x, R/256)$.

Claim 2.1. *For $i \in I$ and all $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$, every $z \in B(x, R/256)$ satisfies $\frac{R}{8} \leq 2^{\kappa(z,t)-i} < \frac{5R}{4}$.*

Proof. By definition, $\mu\left(B(z, 2^{\kappa(z,t)})\right) < 2^t \leq \mu\left(B(z, 2^{\kappa(z,t)+1})\right)$. For the upper bound, we have

$$\mu\left(B(x, 2^{\kappa(z,t)} - R/256)\right) \leq \mu\left(B(z, 2^{\kappa(z,t)})\right) < 2^t \leq 2^{s_i} = \mu\left(B(x, 2^i R)\right),$$

implying that $2^{\kappa(z,t)} - \frac{R}{256} < 2^i R$, which yields $2^{\kappa(z,t)-i} < \frac{5R}{4}$. For the lower bound, we have

$$\mu\left(B(x, 2^{\kappa(z,t)+1} + R/256)\right) \geq \mu\left(B(z, 2^{\kappa(z,t)+1})\right) \geq 2^t \geq 2^{s_{i-1}} = \mu\left(B(x, 2^{i-1} R)\right).$$

We conclude that $2^{\kappa(z,t)+1} + \frac{R}{256} \geq 2^{i-1}R$, which implies that $\frac{R}{8} \leq 2^{\kappa(z,t)-i}$. □

Consider the following events

1. $\mathcal{E}_1 = \left\{d(x, X \setminus P_u(x)) \geq \delta_\varepsilon(x,y)\frac{R}{8} \text{ for all } u \in \mathbb{Z} \text{ with } 2^u \in [R/8, 5R/4]\right\}$,
2. $\mathcal{E}_2 = \{\sigma_u(P_u(x)) = 1 \text{ for all } u \in \mathbb{Z} \text{ with } 2^u \in [R/8, 5R/4]\}$,
3. $\mathcal{E}_3 = \{\sigma_u(P_u(x)) = 0 \text{ for all } u \in \mathbb{Z} \text{ with } 2^u \in [R/8, 5R/4]\}$,
4. $\mathcal{E}_{i,t}^{\text{big}} = \left\{d(y, W_t^i) \geq \frac{1}{512}\delta_\varepsilon(x,y)R\right\}$,
5. $\mathcal{E}_{i,t}^{\text{small}} = \left\{d(y, W_t^i) < \frac{1}{512}\delta_\varepsilon(x,y)R\right\}$.

The basic properties of these events are described in the following claim.

Claim 2.2. *The following assertions hold true:*

(a). $\Pr[\mathcal{E}_1], \Pr[\mathcal{E}_2], \Pr[\mathcal{E}_3] \geq 2^{-4}$.

(b). *For all $i \in I$ and $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$, the event $\mathcal{E}_3$ is independent of $\mathcal{E}_{i,t}^{\text{big}}$.*

(c). *For all $i \in I$ and $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$, the event $\mathcal{E}_2$ is independent of $\mathcal{E}_1 \cap \mathcal{E}_{i,t}^{\text{small}}$.*

(d). *For all $i \in I$ and $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$, if the event $\mathcal{E}_3$ occurs then $x \in W_i^t$.*

(e). *For all $i \in I$ and $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$, if the event $\mathcal{E}_1 \cap \mathcal{E}_2$ occurs then $d(x, W_i^t) \geq \frac{1}{256}\delta_\varepsilon(x,y)R$.*

Proof. For the first assertion, fix u such that $2^u \in [R/8, 5R/4]$. Since $\delta_\varepsilon(x,y) \leq \varepsilon(x,u)$, and P_u is chosen from $\beta(u)$ which is $\varepsilon(x,u)$-padded, $\Pr[d(x, X \setminus P_u(x)) \geq \delta_\varepsilon(x,y)2^u] \geq \frac{1}{2}$. In addition, $\Pr[\sigma_u(P_u(x)) = 1] = \Pr[\sigma_u(P_u(x)) = 0] = 1/2$. Furthermore, the number of relevant values of u in each of the events $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$ is at most four for each event, and the outcomes for different values of u are mutually independent. This implies assertion (a).

To prove the second and third assertions note that for $2^u \in [R/8, 5R/4]$, we always have $\operatorname{diam}(P_u(x)) \leq \frac{5R}{4} < \frac{1}{2}\, d(x,y)$, and thus $P_u(x) \neq P_u(y)$. Furthermore, for every $z \in B(y, \frac{1}{512}\delta_\varepsilon(x,y)R)$, we always have $d(x,z) \geq 3R > \operatorname{diam}(P_u(x)) + \operatorname{diam}(p_u(z))$, thus $P_u(x) \neq P_u(z)$, and the choices of $\sigma_u(P_u(x))$ and $\sigma_u(P_u(z))$ are independent. It follows that the value of $\sigma_u(P_u(x))$ is independent of the data determining whether $d(y, W_t^i) < \frac{1}{512}\delta_\varepsilon(x,y)R$, which proves (b). Assertion (c) follows similarly observing that $\sigma_u(P_u(x))$ is independent also of the value of $d(x, X \setminus P_u(x))$.

To prove the last two assertions fix $i \in I$ and $t \in \mathbb{Z} \cap [s_{i-1}, s_i]$. An application of Claim 2.1 to $z = x$ shows that $2^{\kappa(x,t)-i} \in [R/8, 5R/4]$. Now, by the construction of W_t^i, if $\mathcal{E}_3$ occurs then $x \in W_t^i$; this proves assertion (d). Finally, fix any $z \in B\left(x, \frac{1}{256}\delta_\varepsilon(x,y)R\right)$. Since $z \in B(x, R/256)$, Claim 2.1 implies that $2^{\kappa(z,t)-i} \in [R/8, 5R/4]$. The event $\mathcal{E}_1$ implies that for $u = \kappa(z,t) - i$, $P_u(x) = P_u(z)$, and thus $\sigma_{\kappa(z,t)-i}(P_{\kappa(z,t)-i}(z)) = \sigma_{\kappa(z,t)-i}(P_{\kappa(z,t)-i}(x))$. Now $\mathcal{E}_2$ implies that the latter quantity is 1, and hence $z \notin W_t^i$. Assertion (e) follows. □

We can now conclude the proof of Lemma 1.8. Fix $i \in I$ and $t \in [s_{i-1}, s_i]$. By assertions (d) and (e), if either of the (disjoint) events $\mathcal{E}_3 \cap \mathcal{E}_{i,t}^{\text{big}}$ and $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_{i,t}^{\text{small}}$ occurs then $|d(x, W_t^i) - d(y, W_t^i)| \geq \frac{1}{512}\delta_\varepsilon(x,y)R$. The probability of this is $\Pr[\mathcal{E}_3] \cdot \Pr[\mathcal{E}_{i,t}^{\text{big}}] + \Pr[\mathcal{E}_2] \cdot \Pr[\mathcal{E}_1 \cap \mathcal{E}_{i,t}^{\text{small}}] \geq 2^{-4} \Pr[\mathcal{E}_1] = \Omega(1)$, where we have used assertions (a), (b) and (c), and the fact that $\mathcal{E}_{i,t}^{\text{big}} \cup (\mathcal{E}_1 \cap \mathcal{E}_{i,t}^{\text{small}}) \supseteq \mathcal{E}_1$. It follows that $\mathbb{E}|d(x, W_t^i) - d(y, W_t^i)|^p = [\Omega(\delta_\varepsilon(x,y) \cdot d(x,y))]^p$, and hence

$$
\begin{aligned}
\mathbb{E}\|f(x) - f(y)\|_p^p &\geq \sum_{i \in I} \sum_{t \in \mathbb{Z} \cap [s_{i-1}, s_i]} \mathbb{E}|d(x, W_t^i) - d(y, W_t^i)|^p \\
&\geq [\Omega(\delta_\varepsilon(x,y) \cdot d(x,y))]^p \sum_{i=-6}^{3} |Z \cap [s_{i-1}, s_i]| \\
&\geq [\Omega(\delta_\varepsilon(x,y) \cdot d(x,y))]^p \cdot \frac{s_3 - s_{-7}}{2} \qquad (7) \\
&\geq [\Omega(\delta_\varepsilon(x,y) \cdot d(x,y))]^p \cdot V_\mu(x,y),
\end{aligned}
$$

where in (7) we used the fact that (6) implies that $s_3 - s_{-7} \geq 1$.

This completes the proof of Lemma 1.8. □

Remark 2.1. *The above proof actually yields an embedding φ, such that for all $x, y \in X$ satisfying (6),*

$$
\left[\log \frac{\mu(B(x, 2d(x,y)))}{\mu(B(x, d(x,y)/512))}\right]^{1/p} \cdot \delta_\varepsilon(x,y) \leq \frac{||\varphi(x) - \varphi(y)||_p}{d(x,y)} \leq C\left[\log \Phi(\mu)\right]^{1/p}.
$$

Remark 2.2. *If in the above proof we use sampling and a standard Chernoff bound instead of taking expectations, we can ensure that the embedding takes values in* $\mathbb{R}^k$*, where* $k = O[(\log n)\log\Phi(\mu)]$*. (This is because the lower bound on* $\mathbb{E}|d(x, W_t^i) - d(y, W_t^i)|^p$ *relies on an event that happens with constant probability, similar to [LLR95].) In particular, when* μ *is the counting measure on* X *we get that* $k = O[(\log n)^2]$*. It would be interesting to improve this bound to* $k = O(\log n)$ *(if* $p = 2$ *then this follows from the Johnson-Lindenstrauss dimension reduction lemma [JL84]).*

2.2 Optimal volume-respecting embeddings

Here we prove Theorem 1.5. Let $f : X \to \ell_2$ be the embedding constructed in the previous section and $g = f/\sqrt{50\log\Phi(\mu)}$, so that g is 1-Lipschitz. For concreteness, we denote by $(T, \Pr)$ the probability space over which the random embedding g is defined.

Lemma 2.3. *Fix a subset* $Y \subseteq X$*,* $x \in X \setminus Y$ *and let* $y_0 \in Y$ *satisfy* $d(x, Y) = d(x, y_0)$*. Let* Z *be any* ℓ_2 *valued random variable on* $(T, \Pr)$ *which is measurable with respect to the sigma algebra generated by the random variables* $\{g(y)\}_{y\in Y}$*. Then*

$$\frac{\sqrt{\mathbb{E}||Z - g(x)||_2^2}}{d(x, Y)} \geq C \cdot \delta_\varepsilon(x, y_0) \cdot \sqrt{\frac{V_\mu(x, y_0)}{\log\Phi(\mu)}},$$

where C *is a universal constant.*

We will apply Lemma 2.3 to random variables of the form $Z = \sum_{y\in Y} c_y g(y)$, where the c_y's are scalars. However, the same statement holds for Z's which are arbitrary functions of the variables $\{g(y)\}_{y\in Y}$.

To finish that proof of Theorem 1.5 consider the Hilbert space $H = L_2(T, \ell_2)$, i.e. the space of all square integrable ℓ_2 valued random variables ζ on T equipped with the Hilbertian norm $\|\zeta\|_2 = \sqrt{\mathbb{E}\|\zeta\|_2^2}$. Defining $G : X \to H$ via $G(x) = g(x)$, Lemma 2.3 implies that for every $Y \subseteq X$ and $x \in X \setminus Y$,

$$\frac{d_H(G(x), \mathrm{span}(\{G(y)\}_{y\in Y}))}{d(x, Y)} \geq C \cdot \delta_\varepsilon(x, y_0)\sqrt{\frac{V_\mu(x, y_0)}{\log\Phi(\mu)}}.$$

We now argue as in the proof of Theorem 1.4. Let H_1, H_2 be Hilbert spaces and $G_1 : X \to H_1$, $G_2 : X \to H_2$ be two 1-Lipschitz mappings satisfying for every $Y \subseteq X$ and $x \in X \setminus Y$,

$$\frac{d_{H_1}(G_1(x), \mathrm{span}(\{G_1(y)\}_{y\in Y}))}{d(x, Y)} \geq \frac{C}{16 + 16V_\mu(x, y_0)}\sqrt{\frac{V_\mu(x, y_0)}{\log n}},$$

and

$$\frac{d_{H_2}(G_2(x), \mathrm{span}(\{G_2(y)\}_{y\in Y}))}{d(x, Y)} \geq \frac{C}{\alpha_X}\sqrt{\frac{V_\mu(x, y_0)}{\log n}},$$

where $d(x, y_0) = d(x, Y)$ and μ is the counting measure on X. Denoting $H = H_1 \oplus H_2$ and $G = \frac{1}{\sqrt{2}}(G_1 \oplus G_2)$, the same argument as in the proof of Theorem 1.4 implies that

$$\frac{d_H(G(x), \mathrm{span}(\{G(y)\}_{y\in Y}))}{d(x, Y)} \geq \Omega\left(\frac{1}{\sqrt{\alpha_X \log n}}\right).$$

Now, Feige's argument (see Section 5.3 in [Fei00]) yields the required result.

It remains to prove Lemma 2.3. We use the notation of Section 2.1. Denote $R = \frac{1}{4}d(x, y_0)$ and write $Z = (Z_t^i : i \in I,\ t \in J)$. Consider the events $\widetilde{\mathcal{E}}_{i,t}^{\text{big}} = \left\{Z_t^i \geq \frac{1}{512}\delta_\varepsilon(x, y_0)R\right\}$ and $\widetilde{\mathcal{E}}_{i,t}^{\text{small}} = \left\{Z_t^i < \frac{1}{512}\delta_\varepsilon(x, y_0)R\right\}$. Arguing as in Section 2.1, it is enough to check that $\mathcal{E}_3$ is independent of $\widetilde{\mathcal{E}}_{i,t}^{\text{big}}$ and that $\mathcal{E}_2$ is independent of $\mathcal{E}_1 \cap \widetilde{\mathcal{E}}_{i,t}^{\text{small}}$. Observe that the proof of assertions (b) and (c) in Claim 2.2 uses only the fact that $d(x, y) \geq 4R$ (when considering $z \in B(y, \frac{1}{512}\delta_\varepsilon(x, y)R))$, and this now holds for all $y \in Y$. Since we assume that Z_t^i is measurable with respect to $\{d(y, W_t^i)\}_{y \in Y}$, the required independence follows. □

Remark 2.3. *Note that, by general dimension reduction techniques which preserve distance to affine hulls [Mag02], the dimension of the above embedding can be reduced to $O(k \log n)$ while maintaining the volume-respecting property for k-point subsets.*

3 Low-dimensional embeddings of planar metrics

In this section we refine the ideas of the previous section and prove Theorem 1.6. We say that a metric (X, d) is planar (resp. excludes $K_{s,s}$ as a minor) if there exists a graph $G = (X, E)$ with positive edge weights, such that G is planar (resp. does not admit the complete bipartite graph $K_{s,s}$ as a minor) and $d(\cdot, \cdot)$ is the shortest path metric on a subset of G. We shall obtain optimal low-dimensional embeddings of planar metrics into ℓ_∞ by proving the following more general result.

Theorem 3.1. *Let (X, d) be an n-point metric space that excludes $K_{s,s}$ as a minor. Then X embeds into $\ell_\infty^{O(3^s (\log s) \log n)}$ with distortion $O(s^2)$.*

We will need three lemmas. The first one exhibits a family of decompositions with respect to a diameter bound $\Delta > 0$; it follows easily from [KPR93], with improved constants due to [FT03]. Note that in contrast to Definition 1.1 (and also to Rao's embedding [Rao99]), we require that x and y are padded simultaneously.

Lemma 3.2. *There exists a constant c such that for every metric space (X, d) that excludes $K_{s,s}$ as a minor, and for every $\Delta > 0$, there exists a set of $k = 3^s$ partitions $P_1, \ldots, P_k$ of X, such that*

1. *For every $C \in P_i$, $\operatorname{diam}(C) < \Delta$.*
2. *For every pair $x, y \in X$, there exists an i such that for $T = cs^2$,*

$$B(x, \Delta/T) \subseteq P_i(x) \textbf{ and } B(y, \Delta/T) \subseteq P_i(y).$$

Proof. Fix a edge-weighted graph G that does not admit $K_{s,s}$ as a minor and whose shortest path distance is $d(\cdot, \cdot)$. Fix also some $x_0 \in X$ and $\delta > 0$. For $i \in \{0, 1, 2\}$ and $j \in \{0\} \cup \mathbb{N}$ define:

$$A_j^i = \left\{x \in X : 9(j-1) + 3i \leq \frac{d(x, x_0)}{\delta} < 9j + 3i\right\}.$$

For every i, $P^i = \{A_j^i\}_{j \geq 0}$ clearly forms a partition of X. Let us say that a subset $S \subseteq X$ cuts a subset $S' \subseteq X$ if $S \cap S' \neq \emptyset$ and $S' \not\subseteq S$. Observe that for every $x \in X$ at most one of the sets $\{A_j^i : i = 0, 1, 2\ ; j = 0, 1, 2, \ldots\}$ cuts $B(x, \delta)$, as otherwise there exist $z_1, z_2 \in B(x, \delta)$ for which

$d(z_1, z_2) \geq d(x_0, z_1) - d(x_0, z_2) \geq 3\delta$. Thus, for every $x, y \in X$, for one of the partitions P^0, P^1, P^2 both $B(x, \delta)$ and $B(y, \delta)$ are contained in one of its clusters. For each cluster C of the partitions P^0, P^1, P^2, consider the subgraph of G induced on the points of C, partition C into its connected components, and apply the above process again to each such connected component. Continuing this way a total of s times, we end up with 3^s partitions, and in at least one of them, neither $B(x, \delta)$ nor $B(y, \delta)$ is cut. The results of [KPR93, FT03] show there exists a constant $c > 0$ such that the diameter of each cluster in the resulting partitions is at most $cs^2\delta$, and the lemma follows by setting $\delta = \Delta/(cs^2)$. □

We next consider a collection of such decompositions, with diameter bounds $\Delta > 0$ that are proportional to the integral powers of $4T$. Furthermore, we need these decompositions to be nested.

Lemma 3.3. *Let (X, d) be a metric space that excludes $K_{s,s}$ as a minor, and let $T = O(s^2)$ be as in Lemma 3.2. Then for every $a > 0$ there exists $k = 3^s$ families of partitions of X, $\{P_u^i\}_{u \in \mathbb{Z}}$, $i = 1, \dots, k$ with the following properties:*

1. *For each i the partitions $\{P_u^i\}_{u \in \mathbb{Z}}$ are nested, i.e. P_{u-1}^i is a refinement of P_u^i for all u.*
2. *For each i, every $C \in P_u^i$ satisfies $\text{diam}(C) < a(4T)^u$.*
3. *For each $u \in \mathbb{Z}$ and every pair $x, y \in X$, there exists an i such that,*

$$B(x, a(4T)^u/(2T)) \subseteq P_u^i(x) \textbf{ and } B(y, a(4T)^u/(2T)) \subseteq P_u^i(y).$$

Proof. Let $P_1, \dots, P_k$ be partitions as in Lemma 3.2 with $\Delta = a(4T)^u$ and let $Q_1, \dots, Q_k$ be partitions as in Lemma 3.2 with $\Delta = a(4T)^{u-1}$. Fix j and $C \in P_j$, let $S_C = \{A \in Q_j : A \cap C \neq \emptyset, \text{ but } A \not\subseteq C\}$, and replace every $C \in P_j$ by the sets $A \in S_C$ and the set $C' = C \setminus \bigcup_{A \in S_C} A$. Continuing this process we replace the partition P_j by a new partition P_j' such that Q_j is a refinement of P_j'. Note that we do not alter Q_j. Since $\text{diam}(A) \leq a(4T)^{u-1}$, we have that if $C \in P_j$ and $B(x, a(4T)^u/T) \subseteq C$, then $B(x, 2a(4T)^{u-1}) \subseteq C'$. Continuing this process inductively we obtain the required families of nested partitions. □

We next use a nested sequence of partitions $\{P_u\}_{u \in \mathbb{Z}}$ to form a mapping $\psi : X \to \mathbb{R}^{O(\log |X|)}$.

Lemma 3.4. *Let $\{P_u\}_{u \in \mathbb{Z}}$ be a sequence of partitions of X that is nested (i.e. P_{u-1} is a refinement of P_u), and let $m \geq 0$ and $D \geq 2$ be such that for all $C \in P_u$, $\text{diam}(C) < 2^m D^u$. Assume further that $P_{u_1} = \{X\}$, $P_{u_2} = \{\{x\} : x \in X\}$. Then for all $u_2 \leq u \leq u_1$ and all $A \in P_u$ there exists a mapping $\psi : A \to \mathbb{R}^{2\lceil \log_2 |A| \rceil}$ that satisfies:*

(a). *For every $x \in A$ and every $1 \leq j \leq 2\lceil \log_2 |A| \rceil$ there exists $u' < u$ for which $|\psi(x)_j| = \min\{d(x, X \setminus P_{u'}(x)), 2^m D^{u'}\}$,*

(b). *For all $x, y \in A$, $||\psi(x) - \psi(y)||_\infty \leq 2\, d(x, y)$,*

(c). *If $x, y \in A$ are such that for some $u' \leq u$, $d(x, y) \in [2^m D^{u'-1}, 2^{m+1} D^{u'-1})$ and there exists a cluster $C \in P_{u'}$ for which $x, y \in C$, $B(x, 2^{m+1} D^{u'-2}) \subseteq P_{u'-1}(x)$ and $B(y, 2^{m+1} D^{u'-2}) \subseteq P_{u'-1}(y)$, then $||\psi(x) - \psi(y)||_\infty \geq \frac{d(x,y)}{2D}$.*

Proof. Proceed by induction on u. The statement is vacuous for $u = u_2$, so we assume it holds for u and construct the required mapping for $u + 1$. Fix $A \in P_{u+1}$ and assume that $H = \{A_1, \dots, A_r\} \subseteq$

P_u is a partition of A. By induction there are mappings $\psi_i : A_i \to \mathbb{R}^{2\lceil \log_2 |A_i| \rceil}$ satisfying (a)-(c) above (with respect to A_i and u).

For $h \in \mathbb{N}$ denote $\mathcal{C}_h = \{A_i \in H : 2^{h-1} < |A_i| \le 2^h\}$. We claim that for every $i = 1, \dots, r$ there is a choice of a string of signs $\sigma^i \in \{-1,1\}^{2\lceil \log_2 |A| \rceil - 2\lceil \log_2 |A_i| \rceil}$ such that for all h and for all distinct $A_i, A_j \in \mathcal{C}_h$, $\sigma^i \neq \sigma^j$. Indeed, fix h; If $h \ge \log_2 |A|$ then for $A_i \in \mathcal{C}_h$, $|A_i| > 2^{h-1} \ge |A|/2$; thus $|\mathcal{C}_h| = 1$ and there is nothing to prove. So, assume that $2^h < |A|$ and note that $|\mathcal{C}_h| \le |A|/2^{h-1}$. Hence, the required strings of signs exist provided $2^{2\lceil \log_2 |A| \rceil - 2h} \ge |A|/2^{h-1}$, which is true since $2^h < |A|$.

Now, for every $i = 1, \dots, r$ define a mapping $\zeta_i : A_i \to \mathbb{R}^{2\lceil \log_2 |A| \rceil - 2\lceil \log_2 |A_i| \rceil}$ by

$$\zeta_i(x) = \min\{d(x, X \setminus A_i), 2^m D^u\} \cdot \sigma^i.$$

Finally, define the mapping $\psi : X \to \mathbb{R}^{2\lceil \log_2 |A| \rceil}$ by $\psi|_{A_i} = \psi_i \oplus \zeta_i$. Requirement (a) holds for ψ by construction. To prove requirement (b), i.e. that ψ is 2-Lipschitz, fix $x, y \in A$. If for some i, both $x, y \in A_i$ then by the inductive hypothesis ψ_i is 2-Lipschitz, and clearly ζ_i is 1-Lipschitz, so $\|\psi(x) - \psi(y)\|_\infty \le 2d(x,y)$. Otherwise, fix a coordinate $1 \le j \le 2\lceil \log_2 |A| \rceil$ and use (a) to take $u' \le u$ such that $|\psi(x)_j| = d(x, X \setminus P_{u'}(x))$; since $y \notin P_{u'}(x)$, this is at most $d(x,y)$. It similarly follows that $|\psi(y)_j| \le d(x,y)$, and hence $|\psi(x)_j - \psi(y)_j| \le 2d(x,y)$.

To prove that requirement (c) holds for ψ, take $x, y \in A$ and $u' \le u+1$ such that $d(x,y) \in [2^m D^{u'-1}, 2^{m+1} D^{u'-1})$ and there exists a cluster $C \in P_{u'}$ for which $x, y \in C$, $B(x, 2^{m+1} D^{u'-2}) \subseteq P_{u'-1}(x)$ and $B(y, 2^{m+1} D^{u'-2}) \subseteq P_{u'-1}(y)$. The case $u' \le u$ follows by induction, so assume that $u' = u+1$. Let $i, j \in \{1, \dots, r\}$ be such that $x \in A_i$, $y \in A_j$; then $i \neq j$, since $\operatorname{diam}(A_i) < 2^m D^u \le d(x,y)$. Assume first $\lceil \log_2 |A_i| \rceil \neq \lceil \log_2 |A_j| \rceil$, and without loss of generality suppose $\lceil \log_2 |A_i| \rceil < \lceil \log_2 |A_j| \rceil$; then there is a coordinate $\ell = 2\lceil \log_2 |A_i| \rceil + 1$ for which

$$|\psi(x)_\ell| = |\zeta_i(x)_1| = \min\{d(x, X \setminus A_i), 2^m D^u\},$$

and, for some $u'' < u$,

$$|\psi(y)_\ell| = |\psi_j(y)_\ell| = \min\{d(y, X \setminus P_{u''}(y)), 2^m D^{u''}\}.$$

It follows that $|\psi(x)_\ell| \ge 2^{m+1} D^{u-1}$ (since we assumed $B(x, 2^{m+1} D^{u-1}) \subseteq A_i$), and that $|\psi(y)_\ell| \le 2^m D^{u-1}$, and therefore

$$|\psi(x)_\ell - \psi(y)_\ell| \ge 2^m D^{u-1} \ge \frac{d(x,y)}{2D}.$$

It remains to deal with the case $\lceil \log_2 |A_i| \rceil = \lceil \log_2 |A_j| \rceil$. By our choice of sign sequences, in this case there is an index ℓ for which $\sigma^i_\ell \neq \sigma^j_\ell$, and thus, for $\ell' = \ell + 2\lceil \log_2 |A_i| \rceil$, $|\psi(x)_{\ell'} - \psi(y)_{\ell'}| = |\psi(x)_{\ell'}| + |\psi(y)_{\ell'}|$. Since we assumed $B(x, 2^{m+1} D^{u-1}) \subseteq A_i$ and $B(y, 2^{m+1} D^{u-1}) \subseteq A_j$, we get

$$|\psi(x)_{\ell'} - \psi(y)_{\ell'}| \ge 2^{m+2} D^{u-1} \ge \frac{2d(x,y)}{D}.$$

□

Finally, we prove the main result of this section by a concatenating several of the above maps ψ.

Proof of Theorem 3.1. For each $m \in \{0, 1, \ldots, \lceil \log_2(4cs^2) \rceil\}$ set $a = 2^m$, apply Lemma 3.3 to obtain 3^s families of nested partitions $\{P^m_{1,u}\}_{u\in\mathbb{Z}}, \ldots, \{P^m_{3^s,u}\}_{u\in\mathbb{Z}}$ that satisfy the conclusion of Lemma 3.3 with $T = cs^2$. For every $i = 1, \ldots, 3^s$, let ψ^m_i be the mapping that Lemma 3.4 yields for $\{P^m_{i,u}\}_{u\in\mathbb{Z}}$ when setting $A = X$ and $D = 4cs^2$. Consider the map $\Psi = \oplus_{m,i}\psi^m_i$, which takes values in $\ell_\infty^{O(3^s(\log s)\log n)}$. Clearly Ψ is 2-Lipschitz. Moreover, for every $x, y \in X$ there is $m \in \{0, 1, \ldots, \lceil \log_2(4cs^2) \rceil\}$ and $u \in \mathbb{Z}$ such that $d(x, y) \in [2^m D^u, 2^{m+1} D^u)$. By Lemma 3.3, there is $i \in \{1, \ldots, 3^s\}$ for which $B(x, 2^{m+1}D^{u-1}) \subseteq P_u(x)$ and $B(y, 2^{m+1}D^{u-1}) \subseteq P_u(y)$; it then follows using Lemma 3.4 that

$$\|\Psi(x) - \Psi(y)\|_\infty \geq \|\psi^m_i(x) - \psi^m_i(y)\|_\infty = \Omega(d(x, y)/s^2),$$

as required. □

Acknowledgments

The third author is grateful to Y. Bartal for many discussions on related issues during a preliminary stage of this work.

References

[AFH+04] A. Archer, J. Fakcharoenphol, C. Harrelson, R. Krauthgamer, K. Talwar, and E. Tardos. Approximate classification via earthmover metrics. In *15th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1072–1080, January 2004.

[AR98] Yonatan Aumann and Yuval Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301 (electronic), 1998.

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *36th Annual Symposium on the Theory of Computing*, 2004. To appear.

[Ass83] Patrice Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[Bar96] Yair Bartal. Probabilistic approximations of metric space and its algorithmic application. In *37th Annual Symposium on Foundations of Computer Science*, pages 183–193, October 1996.

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[CCG+98] M. Charikar, C. Chekuri, A. Goel, S. Guha, and S. Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science*, 1998.

[CKR01] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 8–16, Philadelphia, PA, 2001. SIAM.

[DV01] J. Dunagan and S. Vempala. On Euclidean embeddings and bandwidth minimization. In *Randomization, approximation, and combinatorial optimization*, pages 229–240. Springer, 2001.

[Fei00] Uriel Feige. Approximating the bandwidth via volume respecting embeddings. *J. Comput. System Sci.*, 60(3):510–539, 2000.

[FHRT03] Jittat Fakcharoenphol, Chris Harrelson, Satish Rao, and Kunal Talwar. An improved approximation algorithm for the 0-extension problem. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 257–265, New York, 2003. ACM.

[FRT03] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 448–455, 2003.

[FT03] J. Fakcharoenphol and K. Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *Proceedings of 6th Workshop on Approximation, Randomization, and Combinatorial Optimization*, volume 2764 of *Lecture Notes in Computer Science*, pages 36–46. Springer, 2003.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543, 2003.

[Gup00] A. Gupta. Embeddings of finite metrics. Ph.D. thesis, University of California, Berkeley, 2000.

[Hei01] Juha Heinonen. *Lectures on analysis on metric spaces.* Universitext. Springer-Verlag, New York, 2001.

[Ind01] Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Annual Symposium on Foundations of Computer Science*, pages 10–33. IEEE Computer Society, 2001.

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[Joh48] Fritz John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[KLM04] Robert Krauthgamer, Nathan Linial, and Avner Magen. Metric embeddings–beyond one-dimensional distortion. *Discrete Comput. Geom.*, 31(3):339–356, 2004.

[KPR93] Philip N. Klein, Serge A. Plotkin, and Satish Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 682–690, 1993.

[Laa02] Tomi J. Laakso. Plane with A_∞-weighted metric not bi-Lipschitz embeddable to $\mathbb{R}^N$. *Bull. London Math. Soc.*, 34(6):667–676, 2002.

[Lar67] D. G. Larman. A new theory of dimension. *Proc. London Math. Soc.*, 17:178–192, 1967.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[LMN04] J. R. Lee, M. Mendel, and A. Naor. Metric structures in L_1: Dimension, snowflakes, and average distortion. In *6th Intern. Symp. of Latin American Theoretical Informatics*, 2004.

[LN03] James R. Lee and Assaf Naor. Extending Lipschitz functions via random metric partitions, 2003.

[LN04] James R. Lee and Assaf Naor. Metric decomposition, smooth measures, and clustering. *Preprint*, 2004.

[LP01] Urs Lang and Conrad Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.

[LS93] Nathan Linial and Michael Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993.

[Luu98] Jouni Luukkainen. Assouad dimension: antifractal metrization, porous sets, and homogeneous measures. *J. Korean Math. Soc.*, 35(1):23–76, 1998.

[Mag02] A. Magen. Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications. In *Randomization, approximation, and combinatorial optimization (RANDOM)*, Cambdrige, MA, 2002. Springer-Verlag.

[Mat02a] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[Mat02b] Jiri Matoušek. Open problems on embeddings of finite metric spaces. *Available at http://kam.mff.cuni.cz/∼matousek/haifaop.ps*, 2002.

[Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.

[VK87] A. L. Vol′berg and S. V. Konyagin. On measures with the doubling condition. *Izv. Akad. Nauk SSSR Ser. Mat.*, 51(3):666–675, 1987.