Online Embeddings

Piotr Indyk[1], Avner Magen[2], Anastasios Sidiropoulos[3], and Anastasios Zouzias[2]

[1] MIT,
indyk@theory.csail.mit.edu
[2] University of Toronto,
{avner,zouzias}@cs.toronto.edu
[3] Toyota Technological Institute,
tasos@tti-c.org

Abstract. We initiate the study of *on-line metric embeddings*. In such an embedding we are given a sequence of n points $X = x_1, \ldots, x_n$ one by one, from a metric space $M = (X, D)$. Our goal is to compute a low-distortion embedding of M into some host space, which has to be constructed in an on-line fashion, so that the image of each x_i depends only on $x_1, \ldots, x_i$. We prove several results translating existing embeddings to the on-line setting, for the case of embedding into ℓ_p spaces, and into distributions over ultrametrics.

Keywords: computational geometry, online algorithms, metric embeddings

1 Introduction

A low-distortion (or bi-Lipschitz) embedding between two metric spaces $M = (X, D)$ and $M' = (X', D')$ is a mapping f such that for any pair of points $p, q \in X$ we have $D(p, q) \leq D'(f(p), f(q)) \leq c \cdot D(p, q)$; the factor c is called the *distortion* of f. In recent years, low-distortion embeddings found numerous applications in computer science [17,12]. This can be, in part, attributed to the fact that embeddings provide a general method for designing approximation algorithms for problems defined over a "hard" metric, by embedding the input into an "easy" metric and solving the problem in there.

For some problems, however, applying this paradigm encounters difficulties. Consider for example the nearest neighbor problem: given a set P of n points in some metric (X, D), the goal is to build a data structure that finds the nearest point in P to a query point $q \in X$. A fundamental theorem of Bourgain [4] shows that it is possible to embed P and the query point q into an "easy" metric space, such as ℓ_2 with distortion $\log n$. This, however, does not translate to an efficient approximation algorithm for the problem for the simple reason that the query point q is *not known* at the preprocessing stage, so it cannot be embedded together with the set P. More specifically, for the approach to work in this scenario we must require that we can *extend* the embeddings $f : P \to \ell_2$ to $g : P \cup \{q\} \to \ell_2$. We note that the aforementioned Bourgain's theorem [4] does not have such an extendability property.

An even more straightforward setting in which the standard notion of embeddings is not quite the right notion comes up in the design of on-line algorithms. Often, the input considered is metric space; at each step the algorithm receives an input point and needs to make decisions about it instantly. In order to use the embedding method, we must require that the embedding would observe the inputs sequentially, so that a point is mapped based only on the *distance information* of the points observed so far. Here is a precise definition of the desired object.

Definition 1. *An on-line embedding of an n-point metric space $M = (X, D)$ where $X = \{x_1, \ldots x_n\}$ into some host metric space M' is a sequence of functions f_k for $k = 1, \ldots, n$ (possibly randomized) such that*

- *f_k depends only on M_k, the restriction of M on $\{x_1, \ldots, x_k\}$.*
- *f_k extends f_{k-1}: for each $x \in \{x_1, \ldots, x_{k-1}\}$, $f_k(x) = f_{k-1}(x)$. If the functions are randomized, the extendability property means that the random bits used for f_{k-1} are a subset of the random bits for f_k, and when these bits between f_{k-1} and f_k coincide the (deterministic) image of $x \in \{x_1, \ldots, x_{k-1}\}$ is the same for these functions.*

The associated distortion of the above $f_1, \ldots, f_n$ is the distortion of f_n. If f_i can be obtained algorithmically, then we say that we have an on-line algorithm for the embedding problem. We also consider on-line embeddings into shortest-path metrics of graphs. In this case, we require that M_k is mapped into a graph G_k, and that every G_k is subgraph of G_{k+1}.

In this work we investigate fundamental embedding questions in the on-line context. Can we hope, for example, to embed a general metric space in Euclidean space in an on-line fashion? Not surprisingly, the use of randomization is almost always essential in the design of such embeddings. It is interesting to relate the above notion to "oblivious embeddings". An embedding is said to be oblivious, if the image of a point does not depend on other points. In the usual (off-line) embeddings, the image of a point may depend on all other points. In this language, on-line embedding is some type of middle-ground between these two types of embeddings. In particular, oblivious embeddings are a special, very restricted case of on-line embedding. Oblivious embeddings play an important role in the design of algorithms, for example in the context of streaming algorithms [11] or in the design of near linear algorithms that rely on embeddings [1]. Indeed, some of our results use oblivious embeddings as a building block, most notably, random projections and construction of random decompositions.

1.1 Results and motivation

Embedding into ℓ_p spaces, and into distributions over ultrametrics. We start our investigation by considering embeddings into ℓ_p spaces, and into distributions over ultrametrics. These target spaces have been studied extensively in the embedding literature.

We observe that Bartal's embedding [2] can be easily modified to work in the on-line setting. We remark that this observation was also made independently by Englert, Räcke, and Westermann [6]. As a consequence, we obtain an on-line analog of Bourgain's theorem [4]. More specifically, we deduce that any n-point metric space with spread[4] Δ can be embedded on-line into ℓ_p with distortion $O((\log \Delta)^{1/p} \log n)$. Similarly, we also obtain an analog of a theorem due to Bartal [2] for embedding into ultrametrics. More precisely, we give an on-line probabilistic embedding of an input metric into a distribution over ultrametrics with distortion $O(\log n \cdot \log \Delta)$.

Doubling metrics. For the special case when the input space is doubling, we give an improved on-line embedding into ultrametrics with distortion $O(\log \Delta)$. We complement this upper bound by exhibiting a distribution $\mathcal{F}$ over doubling metrics (in fact, subsets of $\mathbb{R}^1$) such that any on-line embedding of a metric chosen from $\mathcal{F}$ into ultrametrics has distortion $\Omega(\min\{n, \log \Delta\})$.

Embedding into ℓ_∞. We also consider on-line analogs of another embedding theorem, due to Fréchet, which states that any n-point metric can be embedded into ℓ_∞ with distortion 1. We show that this theorem extends to the on-line setting with the same distortion, albeit larger dimension. By composing our on-line embedding into ℓ_2, with a random projection, we obtain for any $\alpha > \sqrt{2}$, an on-line embedding into ℓ_∞ with distortion $O(\alpha \cdot \log n \sqrt{\log \Delta})$, and dimension $\Omega(\max\{(\log n)^{2/(1-1/e)}, n^{4/(\alpha^2-2)}\})$.

On-line embedding when an (off-line) isometry or near-isometry is possible. Finally, we consider the case of embedding into constant-dimensional ℓ_p spaces. It is well known ([18]) that for any constant dimension there are spaces that require polynomial distortion (e.g. via a simple volume argument). It is therefore natural to study the embedding question for instances that do embed with small distortion. When a metric embeds isometrically into ℓ_2 or ℓ_2^d, it is clear that this isometry can be found on-line. We exhibit a sharp contrast with this simple fact for the case when there is only a near-isometry guaranteed. Using a topological argument, we prove that there exists a distribution $\mathcal{D}$ over metric spaces that $(1+\varepsilon)$-embed into ℓ_2^d, yet any on-line algorithm with input drawn from $\mathcal{D}$ computes an embedding with distortion $n^{\Omega(1/d)}$. In light of our positive results about embedding into ℓ_2 and a result of Matoušek [18], this bound can be shown to be a near-optimal for on-line embeddings.

Remark 1. For simplicity of the exposition, we will assume that n is given to the on-line algorithm in advance. We remark however that with the single exception of embedding into ℓ_∞, all of our algorithms can be modified to work without this knowledge.

[4] The ratio between the largest and the smallest non-zero distances in the metric space.

Input Space	*Host Space*	*Distortion*	*Section*	*Comments*
General	ℓ_p	$O(\log n(\log \Delta)^{1/p})$	2	$p \in [1, \infty]$
General	Ultrametrics	$O(\log n \log \Delta)$	2	
Doubling	ℓ_2	$O(\log \Delta)$	3	
Doubling	Ultrametrics	$O(\log \Delta)$	3	
Doubling	Ultrametrics	$\Omega(\min\{n, \log \Delta\})$	4	
$(1,2)$-metric	ℓ_∞^n	1	5	
General	ℓ_∞	1	5	The input is drawn from a fixed finite set of metrics.
ℓ_2	ℓ_∞^d	D	5.1	$d \approx \Omega(n^{4/(D^2-2)})$
ℓ_∞	ℓ_∞	> 1	[21]	
$(1+\varepsilon)$-embeddable into ℓ_2^d	ℓ_2^d	$\Omega(n^{1/(d-1)})$	6	

Table 1. Summary of results.

1.2 Related work

The notion of low-distortion on-line embeddings is related to the well-studied notion of Lipschitz extensions. A prototypical question in the latter area is: for spaces Y and Z, is it true that for every $X \subset Y$, and every C-Lipschitz[5] mapping $f : X \to Z$ it is possible to extend f to $f' : Y \to Z$ which is C'-Lipschitz, for C' not much greater than C? For many classes of metric spaces the answer to this question is positive (e.g., see the overview in [16]).

One could ask if analogous theorems hold for low-distortion (i.e., bi-Lipschitz) mapping. If so, we could try to construct on-line embeddings by repeatedly constructing bi-Lipschitz extensions to points $p_1, p_2, \ldots$. Unfortunately, bi-Lipschitz extension theorems are more rare, since the constraints are much more stringent.

In the context of the aforementioned work, the on-line embeddings can be viewed as "weak" bi-Lipschitz extension theorems, which hold for only *some* mappings $f : X \to Z$, $X \subset Y$.

1.3 Notation and definitions

For a point $y \in \mathbb{R}^d$, we denote by y_i the i-th coordinate of y. That is, $y = (y_1, \ldots, y_d)$. Similarly, for a function $f : A \to \mathbb{R}^d$, and for $a \in A$, we use the notation $f(a) = (f_1(a), \ldots, f_d(a))$. Also, we denote by ℓ_p the space of sequences with finite p-norm, i.e., $\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p} < \infty$.

Consider a finite metric space (X, D) and let $n = |X|$. For any point $x \in X$ and $r \geq 0$, the ball with radius r around x is defined as $B_X(x, r) = \{z \in X \mid D(x, z) \leq r\}$. We omit the subscript when it is clear from the context. A metric space (X, D) is called Λ-*doubling* if for any $x \in X$, $r \geq 0$ the ball $B(x, r)$ can be covered by Λ balls of radius $r/2$. The *doubling constant* of X is the infimum Λ so that X is Λ-doubling. The *doubling dimension* of X is $\dim(X) = \log_2 \Lambda$. A metric space with $\dim(X) = O(1)$ is called *doubling*. A γ-net for a metric space (X, D) is a set $N \subseteq X$ such that for any $x, y \in N$, $D_X(x, y) \geq \gamma$ and $X \subseteq \cup_{x \in N} B_X(x, \gamma)$. Let $M_1 = (X, D_1)$ and $M_2 = (X, D_2)$ be two metric spaces. We say that M_1 *dominates* M_2 if for every $i, j \in X$, $D_1(i, j) \geq D_2(i, j)$. Let (X, D_1) and (Y, D_2) be two metric space and an embedding $f : X \to Y$. We say that f is *non-expanding* if f doesn't expand distances between every pair $x_1, x_2 \in X$, i.e., $D_2(f(x_1), f(x_2)) \leq D_1(x_1, x_2)$. Similarly, f is *non-contracting* if it doesn't contract pair-wise distances. Also we say that f is α-bi-Lipschitz if there exists $\beta > 0$ such that for every $x_1, x_2 \in X$, $\beta D_1(x_1, x_2) \leq D_2(f(x_1), f(x_2)) \leq \alpha\beta D_1(x_1, x_2)$.

2 Embedding general metrics into ultrametrics and into ℓ_p

In this section we will describe an on-line algorithm for embedding arbitrary metrics into ℓ_p, with distortion $O(\log n \cdot (\log \Delta)^{1/p})$, for any $p \in [1, \infty]$. We also give an on-line probabilistic embedding into a distribution

[5] I.e., a mapping which expands the distances by a factor at most C.

over ultrametrics with distortion $O(\log n \cdot \log \Delta)$. Both algorithms are on-line versions of the algorithm of Bartal [2], for embedding metrics into a distribution of dominating HSTs, with distortion $O(\log^2 n)$. Before we describe the algorithm we need to introduce some notation.

Definition 2 ([2]). *An l-partition of a metric $M = (X, D)$ is a partition $Y_1, \ldots, Y_k$ of X, such that the diameter of each Y_i is at most l.*

For a distribution $\mathcal{F}$ over l-partitions of a metric $M = (X, D)$, and for $u, v \in X$, let $p_{\mathcal{F}}(u, v)$ denote the probability that in an l-partition chosen from $\mathcal{F}$, u and v belong to different clusters.

Definition 3 ([2]). *An (r, ρ, λ)-probabilistic partition of a metric $M = (X, D)$ is a probability distribution $\mathcal{F}$ over $r\rho$-partitions of M, such that for each $u, v \in X$, $p_{\mathcal{F}}(u, v) \leq \lambda \frac{D(u,v)}{r}$. Moreover, $\mathcal{F}$ is ε-forcing if for any $u, v \in X$, with $D(u, v) \leq \varepsilon \cdot r$, we have $p_{\mathcal{F}}(u, v) = 0$.*

We observe that Bartal's algorithm [2] can be interpreted as an on-line algorithm for constructing probabilistic partitions. The input to the problem is a metric $M = (X, D)$, and a parameter r. In the first step of Bartal's algorithm, every edge of length less than r/n is contracted. This step cannot be directly performed in an on-line setting, and this is the reason that the parameters of our probabilistic partition will depend on Δ. More precisely, our partition will be $1/\Delta$-forcing, while the one obtained by Bartal's off-line algorithm is $1/n$-forcing.

The algorithm proceeds as follows. We begin with an empty partition P. At every step j, each $Y_t \in P$ will correspond to a ball of some fixed radius r_t around a point $y_t \in X_j$. Once we have picked y_t, and r_t, they will remain fixed until the end of the algorithm. Assume that we have partitioned all the points $x_1, \ldots, x_{i-1}$, and that we receive x_i. Let $P = \{Y_1, \ldots, Y_k\}$. If $x_i \notin \bigcup_{j \in [k]} B(y_j, r_j)$, then we add a new cluster Y_{k+1} in P, with center $y_{k+1} = x_i$, and we pick the radius $r_{k+1} \in [0, r \log n)$, according to the probability distribution $p(r_{k+1}) = \left(\frac{n}{n-1}\right) \frac{1}{r} e^{-r_{k+1}/r}$. Otherwise, let Y_s be the minimum-index cluster in P, such that $x_i \in B(y_s, r_s)$, and add x_i to Y_s.

By Bartal's analysis on the above procedure, we obtain the following lemma.

Lemma 1. *Let M be a metric, and $r \in [1, \Delta]$. There exists an $1/\Delta$-forcing, $(r, O(\log n), O(1))$-probabilistic partition $\mathcal{F}$ of M, and a randomized on-line algorithm that against any non-adaptive adversary, given M computes a partition P distributed according to $\mathcal{F}$. Moreover, after each step i, the algorithm computes the restriction of P on X_i.*

By the above discussion it follows that for any $r > 0$ we can compute an $(r, O(\log n), O(1))$-probabilistic partition of the input space $M = (X, D)$. It is well known that this implies an embedding into ℓ_p for any $p \in [1, \infty]$. Since the construction is folklore (see e.g. [22,9,8]), we will only give a brief overview, demonstrating that the embedding can be indeed computed in an on-line fashion.

For each $i \in \{1, \ldots, \log \Delta\}$, and for each $j \in \{1, \ldots, O(\log n)\}$ we sample a probabilistic partition $P_{i,j}$ of M with clusters of radius 2^i. Each such cluster corresponds to a subset of a ball of radius 2^i centered at some point of M. For every i, j we compute a mapping $f_{i,j} : X \to \mathbb{R}$ as follows. For each cluster $C \in P_{i,j}$ we chose $s_{i,j} \in \{-1, 1\}$ uniformly at random. Next, for each point $x \in X$ we need to compute its distance $h_{i,j}(x)$ to the "boundary" of the union of all clusters. For every $C \in P_{i,j}$ let $a(C)$, $r(C)$ be the center and radius of C, respectively. We can order the clusters in $P_{i,j} = (C_1, \ldots, C_k)$, so that C_t is created by the on-line algorithm before C_l for every $t < l$. For a point $x \in X$ let $C(x)$ be the cluster containing x. Suppose $C(x) = C_t$. We set $h_{i,j}(x) = \min_{l \in \{1,\ldots,t\}} |r(C_l) - D(x, a(C_l))|$. Note that $h_{i,j}(x)$ can be computed in an on-line fashion. We set $f_{i,j}(x) = s_{i,j} \cdot h_{i,j}(x)$. The resulting embedding is $\varphi(x) = \bigoplus_{i,j} f_{i,j}(x)$. It is now straightforward to verify that with high probability, for all $x, y \in X$ we have $D(x, y) \cdot \Omega((\log n)^{1/p} / \log n) \geq \|\varphi(x) - \varphi(y)\|_p \geq D(x, y) \cdot \Omega((\log n)^{1/p} / \log n)$, implying the following result.

Theorem 1. *There exists an on-line algorithm that for any $p \in [1, \infty]$, against a non-adaptive adversary, computes an embedding of a given metric into $\ell_p^{O(\log n \log \Delta)}$ with distortion $O(\log n \cdot (\log \Delta)^{1/p})$. Note that for $p = \infty$ the distortion is $O(\log n)$.*

Following the analysis of Bartal [2], we also obtain the following result.

Theorem 2. *There exists an on-line algorithm that against a non-adaptive adversary, computes a probabilistic embedding of a given metric into a distribution over ultrametrics with distortion* $O(\log n \cdot \log \Delta)$.

We remark that in the off-line probabilistic embedding into ultrametrics of [2] the distortion is $O(\log^2 n)$. In this bound there is no dependence on Δ due to a preprocessing step that contracts all sufficiently small edges. This step however cannot be implemented in an on-line fashion, so the distortion bound in Theorem 2 is slightly weaker. Interestingly, Theorem 5 implies that removing the dependence on Δ is impossible, unless the distortion becomes polynomially large.

3 Embedding doubling metrics into ultrametrics and into ℓ_2

In this section we give an embedding of doubling metrics into ℓ_2 with distortion $O(\log \Delta)$. We proceed by first giving a probabilistic embedding into ultrametrics. Let $M = (X, D)$ be a doubling metric, with doubling dimension $\lambda = \log_2 \Lambda$.

We begin with an informal description of our approach. Our algorithm proceeds by incrementally constructing an HST[6], and embedding the points of the input space M into its leaves. The algorithm constructs an HST incrementally, embedding X into its leaves. The construction is essentially greedy: assume a good HST was constructed to the points so far, then when a new point p arrives it is necessary to "go down the right branch" of the tree so as to be at a small tree-distance away from points close to p. This is done by letting each internal vertex of the HST of height i correspond to a subset of M of (appropriately randomized) radius about 2^i. When p is too far from the previous centers of the balls it will branch out. The only issue that can arise (and in general, the only reason for randomness) is that while p is too far from the centre of a ball, it is in fact close to some of its members, and so a large expansion may occur when it is not placed in that part of the tree. Randomness allows to deal with this, but when decisions are made online and cannot be changed as in our case, it is not guaranteed to work. What saves the day is the fact that when a metric has bounded doubling dimension the obtained tree has *bounded* degree. This is crucial when bounding the probability of the bad event described above to happen, as at every level of the tree there could be only constant number of possible conflicts, each with low probability.

We now give a formal argument. Let $\delta = \Lambda^3$. Let $T = (V, E)$ be a complete δ-ary tree of depth $\log \Delta$, rooted at a vertex r. For each $v \in V(T)$, let $l(v)$ be the number of edges on the path from r to v in T. We set the length of an edge $\{u, v\} \in E(T)$ to $\Delta \cdot 2^{-\min\{l(u), l(v)\}}$. That is, the length of the edges along a branch from r to a leaf, are $\Delta, \Delta/2, \Delta/4, \ldots, 1$. Fix a left-to-right orientation of the children of each vertex in T. For a vertex $v \in V(T)$, let T_v denote the sub-tree of T rooted at v, and let $c(v)$ denote the left-most leaf of T_v. We refer to the point mapped to $c(v)$ as the *center* of T_v. Let $B(x, r)$ denote the ball centered at x with radius r.

We will describe an on-line embedding f of M into T, against a non-adaptive adversary. We will inductively define mappings $f_1, f_2, \ldots, f_n = f$, with $f_i : \{x_1, \ldots, x_i\} \to V(T)$, such that f_{i+1} is an extension of f_i. We pick a value $\alpha \in [1, 2]$, uniformly at random.

We inductively maintain the following three invariants.

(I1) For any $v \in V(T)$, if a point of X_i is mapped to the subtree T_v, then there is a point of X_i that is mapped to $c(v)$. In other words, the first point of X that is mapped to a subtree T_v has image $c(v)$, and is therefore the center of T_v. Formally, if $f_i(X_i) \cap V(T_v) \neq \emptyset$, then $c(v) \in f_i(X_i)$.

(I2) For any $v \in V(T)$, all the points in X_i that are mapped to T_v are contained inside a ball of radius $\Delta/2^{l(v)-1}$ around the center of T_v in M. Formally, $f_i^{-1}(V(T_v)) \subset B(f_i^{-1}(c(v)), \Delta/2^{l(v)-1})$.

(I3) For any $v \in V(T)$, and for any children $u_1 \neq u_2$ of v, the centers of T_{u_1} and T_{u_2} are at distance at least $\Delta/2^{l(v)+1}$ in M. Formally, $D(f_i^{-1}(c(u_1)), f_i^{-1}(c(u_2))) > \Delta/2^{l(v)+1}$.

[6] See [3, Definition 8] for a definition of HST.

We begin by setting $f_1(x_1) = c(r)$. This choice clearly satisfies invariants (I1)–(I3). Upon receiving a point x_i, we will show how to extend f_{i-1} to f_i. Let $P = p_0, \ldots, p_t$ be the following path in T. We have $p_0 = r$. For each $j \geq 0$, if there exists a child q of p_j such that $V(T_q) \cap f_{i-1}(X_{i-1}) \neq \emptyset$, and $D(f_{i-1}^{-1}(c(q)), x_i) < \alpha \cdot \Delta/2^j$, we set p_{j+1} to be the left-most such child of p_j. Otherwise, we terminate P at p_j.

Claim. There exists a child u of p_t, such that $c(u) \notin f_{i-1}(\{x_1, \ldots, x_{i-1}\})$.

Proof. Suppose that the assertion is not true. Let $y = f_{i-1}^{-1}(p_t)$. Let $v_1, \ldots, v_\delta$ be the children of p_t. By the inductive invariants (I1) and (I2), it follows that for each $i \in [\delta]$, $D(f_{i-1}^{-1}(c(v_i)), y) \leq \Delta/2^{t-1}$. Moreover, by the choice of p_t, $D(y, x_i) \leq \alpha \cdot \Delta/2^{t-1} \leq \Delta/2^{t-2}$. Therefore, the ball of radius $\Delta/2^{t-2}$ around z in M, contains the $\delta + 1 = \Lambda^3 + 1$ points $x_i, f^{-1}(c(v_1)), \ldots, f^{-1}(c(v_\delta))$. However, by the choice of p_t, and by the inductive invariant (I3), it follows that the balls in M of radius $\Delta/2^{t+1}$ around each one of these points are pairwise disjoint, contradicting the fact that the doubling constant of M is Λ.

By Claim 3, we can find a sub-tree rooted at a child q of p_t such that none of the points in X_{i-1} has its image in T_q. We extend f_{i-1} to f_i by setting $f_i(x_i) = c(q)$. It is straight-forward to verify that f_i satisfies the invariants (I1)–(I3). This concludes the description of the embedding. It remains to bound the distortion of f.

Lemma 2. *For any $x, y \in X$, $D_T(f(x), f(y)) \geq \frac{1}{3} D(x, y)$.*

Proof. Let v be the nearest-common ancestor of $f(x)$ and $f(y)$ in T. By invariant (I2) we have $D(x, y) \leq D(x, f^{-1}(c(v))) + D(y, f^{-1}(c(v))) \leq \Delta \cdot 2^{-l(v)+2}$. Moreover, $D_T(f(x), f(y)) = 2 \cdot \Delta \sum_{i=l(v)}^{\log \Delta} 2^{-i} = \Delta \cdot 2^{-l(v)+2} - 2$. The lemma follows since the minimum distance in M is 1.

Lemma 3. *For any $x, y \in X$, $\mathbb{E}[D_T(f(x), f(y))] \leq O(\Lambda^3 \cdot \log \Delta) \cdot D(x, y)$.*

Proof (Proof sketch (full details in the Appendix):). The distance between x and y is about 2^i when they are separated at level i. For this to happen, y must be assigned to a sibling of x at level $i - 1$. The probability of assigning to any particular such sibling is $O(D(x, y)/2^i)$. It is here that we utilize the bounded-degree property. By a union bound over all siblings at this level we get a contribution of $O(\Lambda^3)$ on the expected expansion. Summing up over all $\log \Delta$ levels we get the desired bound.

Theorem 3. *There exists an on-line algorithm that against any non-adaptive adversary, given a metric $M = (X, D)$ of doubling dimension λ, computes a probabilistic embedding of M into a distribution over ultrametrics with distortion $2^{O(\lambda)} \cdot \log \Delta$.*

It is well known that ultrametrics embed isometrically into ℓ_2, and it is easy to see that such an embedding can be computed in an on-line fashion for the HSTs constructed above. We therefore also obtain the following result.

Theorem 4. *There exists an on-line algorithm that against any non-adaptive adversary, given a doubling metric $M = (X, D)$ of doubling dimension λ, computes a probabilistic embedding of M into ℓ_2 with distortion $2^{O(\lambda)} \cdot \log \Delta$.*

Remark 2. In the off-line setting, Krauthgamer et al. [15] have obtained embeddings of doubling metrics into Hilbert space with distortion $O(\sqrt{\log n})$. Their approach however is based on the random partitioning scheme of Calinescu, Karloff, and Rabani [5], and it is not known how to perform this step in the on-line setting.

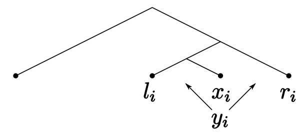

Fig. 1. The evolution of the construction of the ultrametric.

4 Lower bound for probabilistic embeddings into ultrametrics

In this section we present a lower bound for on-line probabilistic embeddings into ultrametrics. Consider the following distribution $\mathcal{F}$ over metric spaces. Each space $M = (X, D)$ in the support of $\mathcal{F}$ is induced by an n-point subset of $\mathbb{R}^1$, with $X = \{x_1, \ldots, x_n\}$, and $D(x_i, x_j) = |x_i - x_j|$. We have $x_1 = 0$, $x_2 = 1$, $x_3 = 1/2$. For each $i \geq 4$, we set $x_i = x_{i-1} + b_i 2^{-i+2}$, where $b_i \in \{-1, 1\}$ is chosen uniformly at random.

It is easy to see that for each $i \geq 3$, there exist points $l_i, r_i \in X$ such that $l_i = x_i - 2^{-i+2}$, $r_i = x_i + 2^{-i+2}$, and $\{x_1, \ldots, x_i\} \cap [l_i, r_i] = \{l_i, x_i, r_i\}$. Moreover, for each $i \in \{3, \ldots, n-1\}$, there uniquely exists $y_i \in \{l_i, r_i\}$, such that $\{x_{i+1}, \ldots, x_n\} \subset [\min\{x_i, y_i\}, \max\{x_i, y_i\}]$.

Claim. Let $M = (X, D)$ be a metric from the support of $\mathcal{F}$. Let f be an embedding of M into an ultrametric $M' = (X, D')$. Then, for each $i \geq 3$, there exists $z_i \in \{l_i, r_i\}$, such that $D'(x_i, z_i) \geq D'(l_i, r_i)$.

Proof. It follows immediately by the fact that M' is an ultrametric, since for any $x_i, l_i, r_i \in X$, $D'(l_i, r_i) \leq \max\{D'(l_i, x_i), D'(x_i, r_i)\}$.

In order to simplify notation, we define for any $i \geq 4$, $\delta_i = D(x_i, y_i)$, and $\delta'_i = D'(x_i, y_i)$.

Claim. Let $M = (X, D)$ be a metric from the support of $\mathcal{F}$. Let f be an on-line embedding of M into an ultrametric $M' = (X, D')$. Then, for any $i \geq 3$, $\mathbf{Pr}[\delta'_i \geq \delta'_{i-1} | \forall j \in \{4, \ldots, i-1\}, \delta'_j \geq \delta'_{j-1}] \geq 1/2$.

Proof. Assume without loss of generality that $z_i = l_i$, since the case $z_i = r_i$ is symmetric. By the construction of M, we have that $\mathbf{Pr}[y_i = z_i | \forall j \in \{4, \ldots, i-1\}, \delta'_j \geq \delta'_{j-1}] = 1/2$. If $y_i = z_i$, then $\delta'_i = D'(x_i, z_i) \geq D'(l_i, r_i) = \delta'_{i-1}$, concluding the proof.

Lemma 4. *Let f be an on-line, non-contracting embedding of M into an ultrametric M'. Then,* $\mathbb{E}[\delta'_{n-1}/\delta_{n-1}] = \Omega(n)$.

Proof. Let $i \geq 4$, and $1 \leq t \leq i-1$. By Claim 4 we have $\mathbf{Pr}[\delta'_i \geq \delta_{i-t}] \geq \mathbf{Pr}[\delta'_i \geq \delta'_{i-t}] \geq \mathbf{Pr}[\forall j \in \{1, \ldots, t\}, \delta'_{i-j+1} \geq \delta'_{i-j}] = \prod_{j=1}^{t} \mathbf{Pr}[\delta'_{i-j+1} \geq \delta'_{i-j} | \forall s \in \{1, \ldots, j-1\}, \delta'_{i-s+1} \geq \delta'_{i-s}] \geq 2^{-t}$. Therefore $\mathbb{E}[\delta'_{n-1}] \geq \sum_{i=3}^{n-1} \delta_i \cdot 2^{-n+i+1} = \sum_{i=3}^{n-1} 2^{-i+2} \cdot 2^{-n+i+1} = \Omega(n \cdot 2^{-n}) = \Omega(n) \cdot \delta_{n-1}$.

Since the aspect ratio (spread) is $\Delta = \Theta(2^n)$, we obtain the following result.

Theorem 5. *There exists a non-adaptive adversary against which any on-line probabilistic embedding into a distribution over ultrametrics has distortion* $\Omega(\min\{n, \log \Delta\})$.

We remark that the above bound is essentially tight, since the input space is a subset of the line, and therefore doubling. By Theorem 3, every doubling metric space probabilistically embeds into ultrametrics with distortion $O(\log \Delta)$.

5 Embedding into ℓ_∞

In the off-line setting, it is well-known that any n-point metric space isometrically embeds into n-dimensional ℓ_∞. Moreover, there is an explicit construction of the embedding due to Fréchet. Let $M = (X, D)$ be an arbitrary metric space. The embedding $f : (X, D) \to \ell_\infty^d$ is simply $f(x_i) = (D(x_i, x_1), D(x_i, x_2), \ldots, D(x_i, x_n))$.

It is clear that the Fréchet embedding does not fit in the on-line setting, since the image of any point x depends on the distances between x and all points of the metric space, in particular the future points.

A similar question regarding the existence of on-line embeddings can be posed: does there exist a bi-Lipschitz extension for *any* embedding into ℓ_∞. The connection with the on-line setting is immediate; it is well-known (see e.g. [16]) that for any metric space $M = (X, D)$, for any $Y \subset X$, and for any a-Lipschitz function $f : Y \to \ell_\infty$, there exists an a-Lipschitz extension $\tilde{f}$ of f, with $\tilde{f} : X \to \ell_\infty$. It seems natural to ask whether this is also true when f and $\tilde{f}$ are required to be a-bi-Lipschitz. Combined with the fact that any metric embeds isometrically into ℓ_∞, this would immediately imply an on-line algorithm for embedding isometrically into ℓ_∞: start with an arbitrary isometry, and extend it at each step to include a new point. Unfortunately, as the next proposition explains, this is not always possible, even for the special case of $(1, 2)$-metrics (the proof appears in the Appendix). We need some new ideas to obtain such an embedding.

Proposition 1. *There exists a finite metric space $M = (X, D)$, $Y \subset X$, and an isometry $f : Y \to \ell_\infty$, such that any extension $\tilde{f} : X \to \ell_\infty$ of f is not an isometry.*

Although it is not possible to extend *any* 1-bi-Lipschitz mapping into ℓ_∞, there exists a *specific* mapping that is extendable, provided that the input space is drawn from a fixed finite family of metrics. We will briefly sketch the proof of this fact, and defer the formal analysis to the Appendix. Consider a metric space M' obtained from M by adding a point p. Suppose that we have an isometry $f : M \to \ell_\infty$. As explained above, f might not be isometrically extendable to M'. The key step is proving that f is always Lipschitz-extendable to M'. We can therefore get an on-line embedding as follows: We maintain a concatenation of embeddings for all metrics in the family of input spaces. When we receive a new point x_i, we isometrically extend all embeddings of spaces that agree with our input on $\{x_1, \ldots, x_i\}$, and Lipschitz-extend the rest.

Theorem 6. *Let $\mathcal{F}$ be a finite collection of n-point metric spaces. There exists an on-line embedding algorithm that given a metric $M \in \mathcal{F}$, computes an isometric embedding of M into ℓ_∞.*

5.1 Low-distortion embeddings into low-dimensional ℓ_∞

In the pursuit of a good embedding of a general metric space into low dimensional ℓ_∞ space we demonstrate the usefulness (and feasibility) of concatenation of two on-line embeddings. In fact one of these embeddings is oblivious, which in particular makes it on-line. Why the concatenation of two on-line embeddings results in yet another on-line embeddings is fairly clear when the embeddings are deterministic; in the case of probabilistic embeddings it suffices to simply concatenate the embeddings in an independent way. In both cases the distortion is the product of the distortions of the individual embeddings. Recall that Section 2 provides us with an on-line embedding of a metric space into Euclidean space. The rest of the section shows that the classical method of projection of points in Euclidean space onto a small number of dimensions supplies low distortion embedding when the host space is taken to be ℓ_∞. To put things in perspective, the classical Johnson-Lindenstrauss lemma [13] considers the case where the image space is equipped with the ℓ_2 norm, and it is well-known that a similar result can be achieved with ℓ_1 as the image space [10, p. 92]. As we will see, ℓ_∞ metric spaces behave quite differently than ℓ_2 and ℓ_1 spaces in this respect, and while a dimension reduction is possible, it is far more limited than the first two spaces.

The main technical ingredient we need is the following concentration result. See also [23] for a similar analysis. The proof is given in the Appendix.

Lemma 5. *Let $u \in \mathbb{R}^n$ be a nonzero vector and let $\alpha > 1$ and $d \geq e^2$. Let y be the normalized projection of u onto d dimensions by a Gaussian matrix as follows: $y = (2/m)Ru$ where R is a $d \times n$ Gaussian random matrix, i.e., a matrix with i.i.d. normal entries and $m = 2\sqrt{\ln d}$. Then $\mathbf{Pr}\left[\|y\|_\infty / \|u\|_2 \leq 1\right] \leq \exp(-\frac{1}{4}\sqrt{d/\ln d})$, and $\mathbf{Pr}\left[\|y\|_\infty / \|u\|_2 \geq \alpha\right] \leq (2/\alpha) d^{1-\alpha^2/2}$.*

With the concentration bound of Lemma 5 it is not hard to derive a good embedding for any n-point set, as is done, say, in the Johnson Lindenstrauss Lemma [13], and we get

Lemma 6. *Let $X \subset \mathbb{R}^n$ an n-point set and let $\alpha > \sqrt{2}$. If $d = \Omega(\max\{(\log n)^{2/(1-1/e)}, n^{4/(\alpha^2-2)}\})$, then the above mapping $f : X \to \ell_\infty^d$ satisfies $\forall x, y \in X$, $\|x-y\|_2 \leq \|f(x)-f(y)\|_\infty \leq \alpha\|x-y\|_2$ with high probability.*

By a straightforward composition of the embeddings in Theorem 1 and Lemma 6, we get

Theorem 7. *There exists an on-line algorithm against any non-adaptive adversary that for any $\alpha > \sqrt{2}$, given a metric $M = (X, D_M)$, computes an embedding of M into ℓ_∞^d with distortion $O(\alpha \cdot \log n \cdot \sqrt{\log \Delta})$ and $d = \Omega(\max\{(\log n)^{2/(1-1/e)}, n^{4/(\alpha^2-2)}\})$.*

Remark 3. The embeddings into ℓ_∞ given in Theorems 1 and 7 are incomparable: the distortion in Theorem 1 is smaller, but the dimension is larger than the one in Theorem 7 for large values of Δ.

6 On-line embedding when an off-line (near-)isometry is possible

It is not hard to see that given an n-point ℓ_2^d metric M, one can compute an online isometric embedding of M into ℓ_2^d. This is simply because there is essentially (up to translations and rotations) a unique isometry, and so keeping extending the isometry online is always possible. However, as soon as we deal with near isometries this uniqueness is lost, and the situation changes dramatically as we next show: even when the input space embeds into ℓ_2^d with distortion $1+\varepsilon$, the best online embedding we can guarantee in general will have distortion that is polynomial in n. We use the following topological lemma from [20].

Lemma 7 ([20]). *Let $\delta < \frac{1}{4}$ and let $f_1, f_2 : S^{d-1} \to \mathbb{R}^d$ be continuous maps satisfying*

- $\|f_i(x) - f_i(y)\|_2 \geq \|x-y\|_2 - \delta$ *for all $x, y \in S^{d-1}$ and all $i \in \{1, 2\}$,*
- $\|f_1(x) - f_2(x)\|_2 \leq \frac{1}{4}$ *for all $x \in S^{d-1}$, and*
- $\Sigma_1 \cap \Sigma_2 = \emptyset$, *where $\Sigma_i = f_i(S^{d-1})$.*

Let U_i denote the unbounded component of $\mathbb{R}^d \setminus \Sigma_i$. Then, either $U_1 \subset U_2$, or $U_2 \subset U_1$.

Theorem 8. *For any $d \geq 2$, for any $\varepsilon > 0$, and for sufficiently large $n > 0$, there exists a distribution $\mathcal{F}$ over n-point metric spaces that embed into ℓ_2^d with distortion $1+\varepsilon$, such that any on-line algorithm on input a metric space chosen from $\mathcal{F}$ outputs an embedding into ℓ_2^d with distortion $\Omega(n^{1/(d-1)})$, and with probability at least $1/2$.*

Proof. Let $\gamma = 1/(\alpha \cdot n^{1/(d-1)})$, where $\alpha > 0$ is a sufficiently large constant. Let S^{d-1} denote the unit $(d-1)$-dimensional sphere, and let X be a γ-net of $(S^{d-1}, \|\cdot\|_2)$. That is, for any $x_1, x_2 \in X$, we have $\|x_1 - x_2\|_2 > \gamma$, and for any $y \in S^{d-1}$ there exists $x \in X$ with $\|x - y\|_2 \leq \gamma$. Such a set X can be always constructed with $|X| \leq O((1/\gamma)^{d-1})$. We will assume for convenience (and without loss of generality) that X contains the point $(1, 0, \ldots, 0)$.

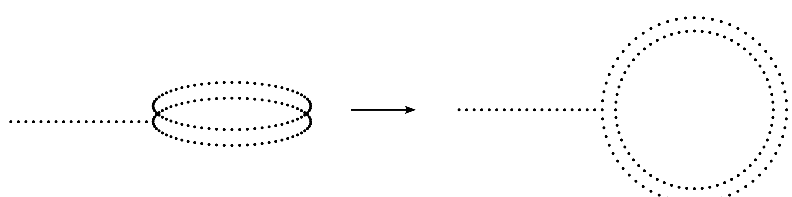

Fig. 2. A realization of Y in $\mathbb{R}^3$ for $d = 2$, and a $(1+\varepsilon)$-embedding into $\mathbb{R}^2$.

Let $t \in \{1, 2\}$. We will define a metric space $M_t = (Y, D_t)$ containing two copies of X, and a discretization of a line segment. Formally, we have $Y = (\{1, 2\} \times X) \cup Z$, where $Z = \{z_1, \ldots, z_{1/\gamma^2}\}$. The distance

$D_t : Y \times Y \to \mathbb{R}_{\geq 0}$ is defined as follows. For each $i \in \{1,2\}$, the set $i \times X$ induces a metrical copy of X, as a subset of $\mathbb{R}^d$. That is, for any $p, q \in X$ we have $D_t((i,p),(i,q)) = \|p-q\|_2$. Moreover, for any $p, q \in X$ we have $D_t((1,p),(2,q)) = \sqrt{\varepsilon^2 + \|p-q\|_2^2}$. The distance D_t induces on Z a line metric. That is, for any $z_j, z_k \in Z$ we have $D_t(z_j, z_k) = |j-k| \cdot \gamma$. For any $z_j \in Z$, and for any $p \in X$ $D_t(z_j,(t,p)) = \|p-(1+j\cdot\gamma, 0, 0, \ldots, 0)\|_2$, and $D_t(z_j,(3-t,p)) = \sqrt{\varepsilon^2 + \|p-(1+j\cdot\gamma,0,0,\ldots,0)\|_2^2}$.

We first argue that for every $t \in \{1,2\}$, the metric space M_t embeds into $\mathbb{R}^d$ with distortion $1+\varepsilon$. To see that, consider the embedding $g : Y \to \mathbb{R}^d$, where for all $p \in X$, $g((t,p)) = (1+\varepsilon)\cdot p$, $g((3-t,p)) = p$, and $g(z_i) = (1+\varepsilon+i\cdot\gamma, 0, \ldots, 0)$. It is straight-forward to check that the distortion of g is $1+\varepsilon$.

We define $\mathcal{F}$ to be the uniform distribution over $\{M_1, M_2\}$. It remains to show that any on-line algorithm, on input a space M_t chosen from $\mathcal{F}$, outputs an embedding with distortion at least $\Omega(n^{1/(d-1)})$, and with probability at least $1/2$. We assume that the on-line algorithm receives first in its input the points in $\{1,2\} \times X$. Let f be the embedding computed by the algorithm. We can assume without loss of generality that f is non-contracting, and that it is c-Lipschitz, for some $c \geq 1$. For any $i \in \{1,2\}$ let f_i be the restriction of f on $\{i\} \times X$. By Kirszbraun's theorem [14], each f_i can be extended to a continuous map $\bar{f}_i : S^{d-1} \to \mathbb{R}^d$, that is also c-Lipschitz. If $c > n^{1/(d-1)}$, then there is nothing to prove, so we may assume $c \leq n^{1/(d-1)}$.

It follows by the analysis in [20] that (i) $\bar{f}_1(S^{d-1}) \cap \bar{f}_2(S^{d-1}) = \emptyset$, (ii) for any $x, y \in S^{d-1}$, $i \in \{1,2\}$, we have $\|f_i(x) - f_i(y)\|_2 \geq \|x-y\|_2 - O(c\cdot\gamma)$, and (iii) for any $x \in S^{d-1}$, we have $\|\bar{f}_1(x) - \bar{f}_2(x)\|_2 = O(c\cdot\gamma)$. Therefore, we can apply Lemma 7 on $\bar{f}_1$ and $\bar{f}_2$. For each $i \in \{1,2\}$, let $U_i = \mathbb{R}^d \setminus \bar{f}_i(S^{d-1})$. It follows that either $U_1 \subset U_2$, or $U_2 \subset U_1$.

Observe that since the algorithm receives first $\{1,2\} \cup X$, and M_t is symmetric on $\{1\} \times X$, $\{2\} \times X$, it follows that $\mathbf{Pr}_{M_t \in \mathcal{F}}[U_{3-t} \subset U_t] = 1/2$. Let $p_0 = (1,0,\ldots,0)$. Observe that $\bar{f}_{3-t}(S^{d-1}) \subseteq B(\bar{f}_{3-t}(p_0), c)$, and on the other hand $\|f(z_{1/\gamma^2}) - \bar{f}_{3-t}(p_0)\|_2 > 1/\gamma > c$. We thus obtain $f(z_{1/\gamma^2}) \in U_{3-t}$. Let $\bar{Z}$ be the polygonal curve in $\mathbb{R}^d$ obtained via affine extension of f on $p_0, z_1, \ldots, z_{1/\gamma^2}$. It follows that $\mathbf{Pr}_{M_t \in \mathcal{F}}[\bar{Z} \cap \bar{f}_{3-t}(S^{d-1}) \neq \emptyset] \geq 1/2$. Therefore, with probability at least $1/2$, two points in Y that are at distance $\Omega(\varepsilon)$ in M_t, are mapped to points that are at distance $O(\gamma c)$. Thus $c = \Omega(1/\gamma) = \Omega(n^{1/(d-1)})$.

Acknowledgements We thank the anonymous referee for pointing out Theorem 4.

References

1. A. Andoni and K. Onak. Approximating Edit Distance in Near-Linear Time. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 199–204, 2009. (Cited on page 2)
2. Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 184–193, 1996. (Cited on pages 2, 4 and 5)
3. Y. Bartal. On approximating arbitrary metrics by tree metrics. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 161–168, 1998. (Cited on page 5)
4. J. Bourgain. On Lipschitz Embedding of Finite Metric Spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985. (Cited on pages 1 and 2)
5. G. Calinescu, H. Karloff, and Y. Rabani. Approximation Algorithms for the 0-extension Problem. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 8–16, 2001. (Cited on page 6)
6. M. Englert, H. Räcke, and M. Westermann. Reordering Buffers for General Metric Spaces. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 556–564, 2007. (Cited on page 2)
7. W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, January 1968. (Cited on page 13)
8. A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded Geometries, Fractals, and Low-Distortion Embeddings. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 534–543, 2003. (Cited on page 4)
9. A. Gupta and R. Ravi. Lecture Notes of Metric Embeddings and Methods, 2003. `http://www.cs.cmu.edu/~anupamg/metrics/lectures/lec10.ps`. (Cited on page 4)
10. P. Indyk. *High-dimensional Computational Geometry*. PhD thesis, Stanford University, 2000. (Cited on page 8)
11. P. Indyk. Stable Distributions, Pseudorandom Generators, Embeddings and Data Stream Computation. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 189–197, 2000. (Cited on page 2)

12. P. Indyk. Tutorial: Algorithmic Applications of Low-distortion Geometric Embeddings. *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 10–33, 2001. (Cited on page 1)
13. W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz Mappings into a Hilbert Space. In Amer. Math. Soc., editor, *In* Conference in Modern Analysis and Probability, pages 189–206, 1984. (Cited on page 8)
14. M. D. Kirszbraun. Uber die zusammenziehenden und lipschitzchen Transformationen. *Fund. Math.*, 22:77–108, 1934. (Cited on page 10)
15. R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured Descent: A new Embedding Method for Finite Metrics. *Geometric Aspects of Functional Analysis*, pages 839–858, 2005. (Cited on page 6)
16. J. R. Lee and A. Naor. Extending Lipschitz Functions via Random Metric Partitions. *Invent. Math.*, 160(1):59–95, 2005. (Cited on pages 3 and 8)
17. N. Linial, E. London, and Y. Rabinovich. The Geometry of Graphs and some of its Algorithmic Applications. *Combinatorica*, 15(2):215–245, 1995. (Cited on page 1)
18. J. Matoušek. Bi-Lipschitz Embeddings into Low-dimensional Euclidean Spaces. *Comment. Math. Univ. Carolinae*, 31:589–600, 1990. (Cited on page 2)
19. J. Matoušek. *Lectures on Discrete Geometry.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002. (Cited on page 12)
20. J. Matoušek and A. Sidiropoulos. Inapproximability for Metric Embeddings into $\mathbb{R}^d$. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 405–413, 2008. (Cited on pages 9 and 10)
21. I. Newman and Y. Rabinovich. Personal communication, 2008. (Cited on page 3)
22. S. Rao. Small Distortion and Volume Preserving Embeddings for Planar and Euclidean Metrics. In *Proceedings of the ACM Symposium on Computational Geometry (SoCG)*, pages 300–306, 1999. (Cited on page 4)
23. G. Schechtman. The Random Version of Dvoretzky's Theorem in ℓ_∞. In *Geometric Aspects of Functional Analysis*, volume 1910, pages 265–270. Springer Berlin / Heidelberg, 2004. (Cited on page 8)

Appendix

Proof of Proposition 1

Proof. Let $X = \{x_1, x_2, x_3, x_4\}$. Let $D(x_2, x_1) = D(x_2, x_3) = D(x_2, x_4) = 1$, and let the distance between any other pair of points be 2. Let $Y = \{x_1, x_2, x_3\}$. Consider an isometry $f : Y \to \ell_\infty$, were for each $i \in \{1, 2, 3\}$, $f_1(x_i) = i$, and for each $j > 1$, $f_j(x_i) = 0$.

Let now $\tilde{f} : X \to \ell_\infty$ be an extension of f. Assume for the sake of contradiction that $\tilde{f}$ is 1-bi-Lipschitz. Then, $\tilde{f}_1(x_4)$ must lie between $\tilde{f}_1(x_1)$ and $\tilde{f}_1(x_2)$, since otherwise either $|\tilde{f}_1(x_4) - \tilde{f}_1(x_1)| > 1$, or $|\tilde{f}_1(x_2) - \tilde{f}_1(x_1)| > 1$. Similarly, $\tilde{f}_1(x_4)$ must also lie between $\tilde{f}_1(x_2)$ and $\tilde{f}_1(x_3)$. Thus, $\tilde{f}_1(x_4) = 2$.

Since $\tilde{f}$ is 1-bi-Lipschitz, it follows that there exists a coordinate j, such that $|\tilde{f}_j(x_4) - \tilde{f}_j(x_1)| = 2$. Since $|\tilde{f}_1(x_4) - \tilde{f}_1(x_1)| = 1$, it follows that $j > 1$. Thus, $|\tilde{f}_j(x_4) - \tilde{f}_j(x_2)| = |\tilde{f}_j(x_4) - \tilde{f}_j(x_1)| = 2 > D(x_4, x_1)$, contradicting the fact that $\tilde{f}$ is 1-bi-Lipschitz.

Proof of Lemma 3

Proof. Let $P^x = p_0^x, \ldots, p_{\log \Delta}^x$, and $P^y = p_0^y, \ldots, p_{\log \Delta}^y$ be paths in T between $p_0^x = p_0^y = r$ and $p_{\log \Delta}^x = f(x)$, $p_{\log \Delta}^y = f(y)$ respectively. Assume that x appears before y in the sequence $x_1, \ldots, x_n$. For any $i \in$

$\{0, \ldots, \log \Delta - 1\}$ we have

$$\begin{aligned}
\mathbf{Pr}[p_{i+1}^x \neq p_{i+1}^y | p_i^x = p_i^y] &\leq \sum_{v \text{ child of } p_i^x} \mathbf{Pr}[p_{i+1}^x = v \text{ and } p_{i+1}^y \neq v | p_i^x = p_i^y] \\
&\leq \sum_{v \text{ child of } p_i^x} \mathbf{Pr}[D(x, f^{-1}(c(v))) \leq \alpha \cdot \Delta / 2^{i+1} < D(y, f^{-1}(c(v)))] \\
&\leq \sum_{v \text{ child of } p_i^x} \mathbf{Pr}\left[\alpha \in \left[\frac{2^{i+1}}{\Delta} D(x, f^{-1}(c(v))), \frac{2^{i+1}}{\Delta} D(y, f^{-1}(c(v)))\right)\right] \\
&\leq \sum_{v \text{ child of } p_i^x} \frac{2^{i+1}}{\Delta} |D(x, f^{-1}(c(v))) - D(y, f^{-1}(c(v)))| \\
&\leq \sum_{v \text{ child of } p_i^x} \frac{2^{i+1}}{\Delta} D(x, y) \\
&= \delta \cdot \frac{2^{i+1}}{\Delta} \cdot D(x, y).
\end{aligned}$$

When the nearest common ancestor (nca) of $f(x)$ and $f(y)$ in T is p_i^x, we have $D_T(f(x), f(y)) = \Delta \cdot 2^{-i+2} - 2$. Therefore

$$\begin{aligned}
\mathbb{E}[D_T(f(x), f(y))] &= \sum_{i=0}^{\log \Delta - 1} \mathbf{Pr}[\text{nca}(f(x), f(y)) = p_i^x] \cdot (\Delta \cdot 2^{-i+2} - 2) \\
&< \sum_{i=0}^{\log \Delta - 1} \mathbf{Pr}[p_i^x = p_i^y \text{ and } p_{i+1}^x \neq p_{i+1}^y] \cdot \Delta \cdot 2^{-i+2} \\
&\leq \sum_{i=0}^{\log \Delta - 1} \delta \cdot \frac{2^{i+1}}{\Delta} \cdot D(x, y) \cdot \Delta \cdot 2^{-i+2} \\
&= O(\Lambda^3 \cdot \log \Delta) \cdot D(x, y)
\end{aligned}$$

6.1 Proof of Theorem 6

We give a formal analysis of the on-line algorithm for embedding isometrically into ℓ_∞. Let $M = (X, D)$ be the input metric, with $X = \{x_1, \ldots, x_n\}$. Let M_i denote the restriction of M on $\{x_1, \ldots, x_i\}$.

The following Lemma is the well-known Helly property of intervals of $\mathbb{R}$ (see e.g. [19]).

Lemma 8 (Helly property of line intervals). *Consider a finite collection of closed real intervals $\Lambda_1, \ldots, \Lambda_k$. If for any $i, j \in [k]$, $\Lambda_i \cap \Lambda_j \neq \emptyset$, then $\bigcap_{i=1}^k \Lambda_i \neq \emptyset$.*

The main idea of the algorithm is as follows. Each coordinate of the resulting embedding is clearly a line metric. The algorithm will produce an embedding that inductively satisfies the property that at each step i, for each possible line metric T dominated by M_i, there exists a coordinate inducing M_i.

The main problem is what to do with the line metrics that are dominated by M_i, but are not dominated by M_{i+1}. As the next lemma explains, we can essentially replace such metrics by line metrics that are dominated by M_{i+1}.

Lemma 9. *Let T be a line metric on $\{x_1, \ldots, x_i\}$ dominated by M_i. Then, T can be extended to a line metric T' on $\{x_1, \ldots, x_{i+1}\}$ dominated by M_{i+1}.*

Proof. Consider $b_1, \ldots b_i \in \mathbb{R}$, such that for any $j,k \in [i]$, $D_T(x_j,x_k) = |b_j - b_k|$. For any $j \in [i]$, define the interval $\Lambda_j = [b_j - D(x_j,x_{i+1}), b_j + D(x_j,x_{i+1})]$. Assume now that there exist $j,k \in [i]$, such that $\Lambda_j \cap \Lambda_k = \emptyset$. Assume without loss of generality that $b_j \leq b_k$. It follows that $b_j + D(x_j,x_{i+1}) < b_k - D(x_k,x_{i+1})$. Since $D(x_j,x_k) = b_k - b_j$, it follows that $D(x_j,x_{i+1}) + D(x_{i+1},x_k) < D(x_j,x_k)$, contradicting the triangle inequality.

We have thus shown that for any $j,k \in [i]$, $\Lambda_j \cap \Lambda_k \neq \emptyset$. By Lemma 8 it follows that there exists $b_{i+1} \in \bigcap_{j=1}^{i} \Lambda_j$. We can now define the extension T' of T by $D_{T'}(x_j,x_k) = |b_j - b_k|$. It remains to verify that T' is dominated by M_{i+1}. Since T is dominated by M_i, it suffices to consider pairs of points x_j, x_{i+1}, for $j \in [i]$. Since $b_{i+1} \in \Lambda_j$, we have $D_{T'}(x_j,x_{i+1}) = |b_j - b_{i+1}| \leq D(x_j,x_{i+1})$, which concludes the proof.

Using Lemma 9, we are now ready to prove the main result of this section. Recall that $\mathcal{F}$ is a finite collection of n-point input metric spaces. Let $\mathcal{A}$ be the set of all distances between pairs of points of the spaces in $\mathcal{F}$. Let $\mathcal{S} = (\Phi_1, \ldots, \Phi_N)$ be a sequence containing all possible line metrics on n points, that have distances in $\mathcal{A}$. For any $i \geq 0$, after receiving the i first points, the algorithm maintains a sequence $\mathcal{S}_i = (\Phi_1^i, \ldots, \Phi_N^i)$ of line metrics, such that each $\Phi_j^i \in \mathcal{S}_i$ is dominated by M_i. Initially, $\mathcal{S}_0 = \mathcal{S}$.

We first show how to compute $\mathcal{S}_i$. Assume that we have correctly computed the sequence $\mathcal{S}_{i-1}$, and we receive x_i. If Φ_j^{i-1} is dominated[7] by M_i, then we set $\Phi_j^i = \Phi_j^{i-1}$. Otherwise, let T be the restriction of Φ_j^i to the first $i-1$ points of M. By Lemma 9 we can extend T to a line metric T' on $\{x_1, \ldots, x_i\}$, dominated by M_i. We pick $T'' \in \mathcal{S}$ such that the restriction of T'' on $\{x_1, \ldots, x_{i+1}\}$ is T', and we set $\Phi_j^i = T''$. Clearly, every $\Phi_j^i \in \mathcal{S}_i$ is dominated by M_i.

Let $f : X \to \mathbb{R}^N$ be the embedding computed by the algorithm. After receiving the i first points, for each $j \in [i]$, $k \in [N]$, the k-th coordinate of $f(x_j)$ is equal to the coordinate of $\Phi_k^i(x_j)$.

Since for each $k \in [N]$, Φ_k^n is dominated by M, it follows that f is non-expanding.

Moreover, observe that $\mathcal{S}_n$ contains all the line metrics $\mathcal{S}$ that are dominated by M. Thus, for each $x_j \in X$, $\mathcal{S}_n$ contains the metric W_j, where for each $t \in [n]$, the coordinate of the image of x_t is $D(x_t, x_j)$. It follows that f is non-contracting, and thus an isometry. This concludes the proof.

Proof of Lemma 5

Proof. As is standard, we will use homogeneity and the rotation invariance of the operator R to assume without loss of generality that $u = (1,0,\ldots,0)^t$. It follows that $\|y\|_\infty = (2/m)\|z\|_\infty$ where $z = (z_1, z_2, \ldots, z_d)^t$ and $z_i \sim \mathcal{N}(0,1)$ for $i = 1, \ldots, d$. Denote by $\varphi(x)$ the probability density function of the normal distribution $\frac{e^{-x^2/2}}{\sqrt{2\pi}}$ and the probability distribution function $\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = \mathbf{Pr}[X \leq x]$ where $X \sim \mathcal{N}(0,1)$. We will use the well-known fact to estimate Φ (see [7, Lemma 2, p. 131])

$$\frac{x^2-1}{x^3}\varphi(x) \leq 1 - \Phi(x) \leq \varphi(x)/x. \tag{1}$$

For the left tail of the lemma, we get that

$$\begin{aligned}
\mathbf{Pr}[\|y\|_\infty \leq 1] = \mathbf{Pr}[\|z\|_\infty \leq m/2] = \mathbf{Pr}[\bigcap_i \{z \mid |z_i| \leq m/2\}] &\leq (\Phi(m/2))^d \\
&= (1-(1-\Phi(m/2)))^d \leq \exp\left(-d(1-\Phi(m/2))\right) \\
&\leq \exp\left(-d\frac{(2m^2-8)\varphi(m/2)}{m^3}\right) \quad (Eqn.\ 1) \\
&\leq \exp\left(-d\frac{e^{-m^2/8}}{m\sqrt{2\pi}}\right) \leq \exp\left(-d\frac{d^{-1/2}}{4\sqrt{\ln d}}\right) \\
&= \exp\left(-\frac{1}{4}\sqrt{d/\ln d}\right).
\end{aligned}$$

[7] Here we mean that an i-point metric dominates an n-point metric, if it dominates its restriction on the first i points.

We turn to bound the right tail.

$$\begin{aligned}
\mathbf{Pr}\left[\|y\|_\infty \geq \alpha\right] &= \mathbf{Pr}\left[\|z\|_\infty \geq m\alpha/2\right] = \mathbf{Pr}\left[\bigcup_i \{z \mid |z_i| \geq m\alpha/2\}\right] \\
&\leq 2d \cdot (1 - \Phi(m\alpha/2)) \\
&\leq 2d \frac{\varphi(m\alpha/2)}{m\alpha/2} = \frac{4d}{m\alpha\sqrt{2\pi}} e^{-m^2\alpha^2/8} \quad (Eqn.\ (1)) \\
&= \frac{2d}{\alpha\sqrt{2\pi \ln d}} d^{-\alpha^2/2} \leq (2/\alpha) d^{1-\alpha^2/2}.
\end{aligned}$$