Finite Metric Spaces and Partitions

Sariel Har-Peled

December 7, 2004

598 - Approximation Algorithms in Geometry

1 Finite Metric Spaces

Definition 1.1 A metric space is a pair $(X,d)$ where $X$ is a set and $d : X \times X \to [0,\infty)$ is a metric, satisfying the following axioms: (i) $d(x,y) = 0$ iff $x = y$, (ii) $d(x,y) = d(y,x)$, and (iii) $d(x,y) + d(y,z) \geq d(x,z)$ (triangle inequality).

For example, $\mathbb{R}^2$ with the regular euclidean distance is a metric space.

It is usually of interest to consider the finite case, where $X$ is an $n$-point set. Then, the function $d$ can be specified by $\binom{n}{2}$ real numbers. Alternatively, one can think about $(X,d)$ is a weighted complete graph, where we specify weights on the edges, and the resulting graph comply with the triangle inequality.

In fact, finite metric spaces rise naturally from graphs. Indeed, let $G$ be a weighted graph defined over $X$, and let $d_G(x,y)$ be the length of the shortest path between $x$ and $y$ in $G$. It is easy to verify that $(X,d_G)$ is a finite metric space. As such if the graph $G$ is sparse, it provides a compact representation to the finite space $(X,d_G)$.

Definition 1.2 Let $(X,d)$ be an $n$-point metric space. We denote by $b(x,r) = \{y \in X \mid d(x,y) < r\}$ the open ball of radius $r$ about $x$.

Underling our discussion of metric spaces are algorithmic applications. The hardness of various computational problems depends heavily on the structure of the finite metric space. Thus, given a finite metric space, and a computational task, it is natural to try to map the given metric space into a new metric where the task at hand becomes easy.

Example 1.3 For example, computing the diameter is not trivial in two dimensions, but is easy in one dimension. Thus, if we could map points in two dimensions into points in one dimension, such that the diameter is preserved, then computing the diameter becomes easy. In fact, this approach yields an efficient approximation algorithm, see Exercise 7.1 below.

Of course, this mapping from one metric space to another, is going to introduce error. We would be interested in minimizing the error introduced by such a mapping.

Definition 1.4 Let $(X,d_X)$ and $(Y,d_Y)$ be metric spaces. A mapping $f : X \to Y$ is called an embedding, and is $C$-Lipschitz if $d_Y(f(x),f(y)) \leq C \cdot d_X(x,y)$ for all $x,y \in X$. The mapping $f$ is called $K$-bi-Lipschitz if there exists a $C > 0$ such that

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


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 $\text{dist}(f)$. The least distortion with which $X$ may be embedded in $Y$ is denoted $c_Y(X)$.

There are several powerful results in this vain, that would be presented:

1. Probabilistic trees - every finite metric can be randomly embedded into a tree such that the “expected” distortion for a specific pair of points is $O(\log n)$.

2. Bourgain embedding - shows that any $n$-point metric space can be embedded into (finite dimensional) metric space with $O(\log n)$ distortion.

3. JL lemma - shows that any $n$-point set in Euclidean space with the regular Euclidean distance can be embedded into $\mathbb{R}^k$ with distortion $(1 + \varepsilon)$, where $k = O(\varepsilon^{-2} \log n)$.

## 2 Partitions

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 $\mathcal{D}_X$.

### 2.1 Constructing the partition

Let $\Delta = 2^u$ be a prescribed parameter, which is the required diameter of the resulting clusters. Choose, uniformly at random, a permutation $\pi$ of $X$ and a random value $\alpha \in [1/4, 1/2]$. Let $R = \alpha \Delta$, and observe that it is uniformly distributed in the interval $[\Delta/4, \Delta/2]$.

The partition is now defined as follows: A point $x \in X$ is assigned to the cluster $C_y$ of $y$, where $y$ is the first point in the permutation in distance $\leq R$ from $x$. Formally,

$$C_y = \left\{ x \in X \mid x \in b(y, R) \text{ and } \pi(y) \leq \pi(z) \text{ for all } z \in X \text{ with } x \in b(z, R) \right\}.$$ 

Let $P = \{C_y\}_{y \in X}$ denote the resulting partition.

### 2.2 Properties

**Lemma 2.1** Let $(X, d)$ be a finite metric space, $\Delta = 2^u$ a prescribed parameter, and let $P$ be the partition of $X$ generated by the above random partition. Then the following holds:

(i) For any $C \in P$, we have $\text{diam}(C) \leq \Delta$.

(ii) Let $x$ be any point of $X$, and $t$ a parameter $\leq \Delta/8$. Then,

$$\Pr[b(x, t) \not\subseteq P(x)] \leq \frac{8t}{\Delta} \ln \frac{b}{a},$$

where $a = |b(x, \Delta/8)|$, $b = |b(x, \Delta)|$. 


Proof: Since $C_y \subseteq b(y, R)$, we have that $\text{diam}(C_y) \leq \Delta$, and thus the first claim holds.

Let $U$ be the set of points of $b(x, \Delta)$, such that $u \in U$ iff $b(u, R) \cap b(x, t) \neq \emptyset$. Arrange the points of $U$ in increasing distance from $x$, and let $w_1, \ldots, w_t$ denote the resulting order, where $b' = |U|$. 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 first point in $\pi$ such that $b(x, t) \cap C_{w_k} \neq \emptyset$, and yet $b(x, t) \not\subseteq C_{w_k}$. Note that if $w_k \notin b(x, \Delta/8)$, then $Pr[\mathcal{E}_k] = 0$ since $b(x, t) \subseteq b(x, \Delta/8) \subseteq b(w_k, \Delta/4) \subseteq b(w_k, R)$.

In particular, $w_1, \ldots, w_a \in b(x, \Delta/8)$ and as such $Pr[\mathcal{E}_1] = \cdots = Pr[\mathcal{E}_a] = 0$. Also, note that if $d(x, w_k) < R - t$ then $b(w_k, R)$ contains $b(x, t)$ and as such $\mathcal{E}_k$ cannot happen. Similarly, if $d(x, w_k) > R + t$ then $b(w_k, R) \cap b(x, t) = \emptyset$ and $\mathcal{E}_k$ cannot happen. As such, if $\mathcal{E}_k$ happen then $R - t \leq d(x, w_k) \leq R + t$. Namely, if $\mathcal{E}_k$ happen then $R \in I_k$. Namely, $Pr[\mathcal{E}_k] = Pr[\mathcal{E}_k \cap (R \in I_k)] = Pr[R \in I_k] \cdot Pr[\mathcal{E}_k | R \in I_k]$. Now, $R$ is uniformly distributed in the interval $[\Delta/4, \Delta/2]$, and $I_k$ is an interval of length $2t$. Thus, $Pr[R \in I_k] \leq 2t/(\Delta/4) = 8t/\Delta$.

Next, to bound $Pr[\mathcal{E}_k | R \in I_k]$, we observe that $w_1, \ldots, w_{k-1}$ are closer to $x$ than $w_k$ and their distance to $b(x, t)$ is smaller than $R$. Thus, if any of them appear before $w_k$ in $\pi$ then $\mathcal{E}_k$ does not happen. Thus, $Pr[\mathcal{E}_k | R \in I_k]$ is bounded by the probability that $w_k$ is the first to appear in $\pi$ out of $w_1, \ldots, w_k$. But this probability is $1/k$, and thus $Pr[\mathcal{E}_k | R \in I_k] \leq 1/k$.

We are now ready for the kill. Indeed,

$$Pr[b(x, t) \not\subseteq P(x)] = \sum_{k=1}^{b'} Pr[\mathcal{E}_k] = \sum_{k=a+1}^{b'} Pr[\mathcal{E}_k] = \sum_{k=a+1}^{b'} Pr[R \in I_k] \cdot Pr[\mathcal{E}_k | R \in I_k]$$

$$\leq \sum_{k=a+1}^{b'} \frac{8t}{\Delta} \cdot \frac{1}{k} \leq \frac{8t}{\Delta} \ln \frac{b'}{a} \leq \frac{8t}{\Delta} \ln \frac{b}{a},$$

since $\sum_{k=a+1}^{b} \frac{1}{k} \leq \int_{a}^{b} \frac{dx}{x} = \ln \frac{b}{a}$ and $b' \leq b$.

\section{Probabilistic embedding into trees}

In this section, given $n$-point finite metric $(X, d)$. we would like to embed it into HST. One can verify that for any embedding into HST, the distortion in the worst case is $\Omega(n)$ (see Exercise 6.2). Thus, we define a randomized algorithm that embed $(X, d)$ into a tree. Let $T$ be the resulting tree, and consider two points $x, y \in X$. Consider the random variable $d_T(x, y)$. The tree $T$ we constructed in such a way that distances never shrink; i.e. $d(x, y) \leq d_T(x, y)$. The \textit{probabilistic distortion} of this embedding is $\max_{x,y} E \left[ \frac{d_T(x, y)}{d(x, y)} \right]$. Somewhat surprisingly, one can find such an embedding with logarithmic probabilistic distortion.

\textbf{Definition 3.1} \textit{Hierarchically well-separated tree} (HST) is a metric space defined on the leaves of a rooted tree $T$. To each vertex $u \in T$ there is associated a label $\Delta_u \geq 0$ such that $\Delta_u = 0$ if and only if $u$ is a leaf of $T$. The labels are such that if a vertex $u$ is a child of a vertex $v$ then $\Delta_u \leq \Delta_v$.

The distance between two leaves $x, y \in T$ is defined as $\Delta_{\text{lca}(x, y)}$, where $\text{lca}(x, y)$ is the least common ancestor of $x$ and $y$ in $T$.

A HST $T$ is a $k$-HST if for a vertex $v \in T$, we have that $\Delta_v \leq \Delta_{\text{p}(v)}/k$.

\textbf{Theorem 3.2} Given $n$-point metric $(X, d)$ one can randomly embed it into a $2$-HST with probabilistic distortion \leq $24 \ln n$.

\textit{Proof:} The construction is recursive. Let $\text{diam}(P)$, and compute a random partition of $X$ with cluster diameter $\text{diam}(P)/2$, using the construction of Section 2.1. We recursively construct
Let $x \in v$. Theorem 4.2 Let $v$

The bounded spread case

by Lemma 2.1. Thus, $u$

and let $\Delta$ be a random partition of $P$, for each cluster, and hang the resulting clusters on the root node $v$. Clearly, the resulting tree is a 2-HST.

For a node $v \in T$, let $X(v)$ be the set of points of $X$ contained in the subtree of $v$.

For the analysis, assume $\text{diam}(P) = 1$, and consider two points $x, y \in X$. We consider a node $v \in T$ to be in level $i$ if $\text{level}(v) = \lceil \log \Delta_v \rceil = i$. The two points $x$ and $y$ correspond to two leaves in $T$, and let $u$ be the least common ancestor of $x$ and $y$ in $t$. We have $d_T(x, y) \leq 2^{\text{level}(v)}$. Furthermore, note that a long a path the levels are strictly monotonically increasing.

In fact, we are going to be conservative, and let $w$ be the first ancestor of $x$, such that $b = b(x, d(x, y))$ is not properly contained in $X(u_1), \ldots, X(u_m)$, where $u_1, \ldots, u_m$ are the children of $w$. Clearly, $\text{level}(w) > \text{level}(v)$. Thus, $d_T(x, y) \leq 2^{\text{level}(w)}$.

Consider the path $\sigma$ from the root of $T$ to $X$, and let $\mathcal{E}_i$ be the event that $b$ is not fully contained in $X(v_i)$, where $v_i$ is the node of $\sigma$ of level $i$ (if such a node exists). Furthermore, let $Y_i$ be the indicator variable which is 1 if $\mathcal{E}_i$ is the first to happen out of the sequence of events $\mathcal{E}_0, \mathcal{E}_{-1}, \ldots$ Clearly, $d_T(x, y) \leq \sum Y_i 2^i$.

Let $t = d(x, y) + j = \lfloor \log d(x, y) \rfloor$, and $n_i = |b(x, 2^i)|$ for $i = 0, \ldots, -\infty$. We have

$$
E[d_T(x, y)] \leq \sum_{i=j}^{0} E[Y_i] 2^i \leq \sum_{i=j}^{0} 2^i \Pr[\mathcal{E}_i \cap \mathcal{E}_{i-1} \cap \cdots \mathcal{E}_0] \leq \sum_{i=j}^{0} 2^i \cdot \frac{8t\ln n_i}{n_i-3},
$$

by Lemma 2.1 Thus,

$$
E[d_T(x, y)] \leq 8t \ln \left( \prod_{i=j}^{0} \frac{n_i}{n_i-3} \right) \leq 8t \ln(n_0 \cdot n_1 \cdot n_2) \leq 24t \ln n.
$$

It thus follows, that the expected distortion for $x$ and $y$ is $\leq 24 \ln n$.

## 4 Embedding any metric space into Euclidean space

**Lemma 4.1** Let $(X, d)$ be a metric, and let $Y \subset X$. Consider the mapping $f : X \to \mathbb{R}$, where $f(x) = \min_{y \in Y} d(x, y)$. Then for any $x, y \in X$, we have $|f(x) - f(y)| \leq d(x, y)$. Namely $f$ is nonexpansive.

**Proof:** Indeed, let $x'$ and $y'$ be the closet points of $Y$ to $x$ and $y$, respectively. Observe that $f(x) = d(x, x') \leq d(x, y') \leq d(x, y) + d(y, y') = d(x, y) + f(y)$ by the triangle inequality. Thus, $f(x) - f(y) \leq d(x, y)$. By symmetry, we have $f(y) - f(x) \leq d(x, y)$. Thus, $|f(x) - f(y)| \leq d(x, y)$.

### 4.1 The bounded spread case

Let $(X, d)$ be a $n$-point metric. The spread of $X$, denoted by $\Phi(X) = \frac{\text{diam}(X)}{\min_{x \neq y \in X} d(x, y)}$.

**Theorem 4.2** Given a $n$-point metric $(X, d)$, with spread $\Phi$, one can embed it into Euclidean space $\mathbb{R}^k$ with distortion WHATEVER, where $k = \cdots$

**Proof:** Assume that $\text{diam}(X) = \Phi$, and let $r_i = 2^{i-2}$, for $i = 1, \ldots, \alpha$, where $\alpha = \lceil \log \Phi \rceil$. Let $P_{i,j}$ be a random partition of $P$ with diameter $r_i$, for $i = 1, \ldots, \alpha$ and $j = 1, \ldots, \beta$, where $\beta = \lceil c \log n \rceil$ and $c$ is a large enough constant to be determined shortly.
For each cluster of $P_{i,j}$ randomly toss a coin, and let $V_{i,j}$ be the all the points of $X$ that belong to clusters in $P_{i,j}$ that got tail in their coin toss. For a point $u \in X$, let $f_{i,j}(x) = d(x, V_{i,j}) = \min_{v \in V_{i,j}} d(x, v)$, for $i = 0, \ldots, m$ and $j = 1, \ldots, \beta$. Let $F : X \to \mathbb{R}^{(m+1)\beta}$ be the embedding, such that $F(x) = (f_{0,1}(x), f_{0,2}(x), \ldots, f_{0,\beta}(x), f_{1,1}(x), f_{0,2}(x), \ldots, f_{1,\beta}(x), \ldots, f_{m,1}(x), f_{m,2}(x), \ldots, f_{m,\beta}(x))$.

Next, consider two points $x, y \in X$, with distance $\phi = d(x, y)$. Let $k$ be an integer such that $r_u \leq \phi/2 \leq r_{u+1}$. Clearly, in any partition of $P_{u,\alpha}, \ldots, P_{u,\beta}$ the points $x$ and $y$ belong to different clusters. Furthermore, with probability half $x \in V_{u,j}$ and $y \in V_{u,j}$, for $j = 1, \ldots, \beta$.

Let $\mathcal{E}_j$ denote the event that $b(x, \rho) \subseteq V_{u,j}$ and $y \notin V_{u,j}$, for $j = 1, \ldots, \beta$, where $\rho = \phi/(16 \ln n)$. By Lemma 2.1, we have $\Pr[b(x, \rho) \notin V_{u,j}] \leq 1/2$. Thus, $\Pr[\mathcal{E}_j] \geq 1/8$. Notice, that if $\mathcal{E}_j$ happens, then $f_{u,j}(x) \geq \rho$ and $f_{u,j}(y) = 0$.

Let $X_j$ be an indicator variable which is 1 if $\mathcal{E}_i$ happens, for $j = 1, \ldots, \beta$. Let $Z = \sum_j X_j$, and we have $\mu = \mathbb{E}[Z] = \mathbb{E}\left[\sum_j X_j\right] \geq \beta/8$. Thus, the probability that only $\beta/16$ of $\mathcal{E}_1, \ldots, \mathcal{E}_\beta$ happens, is $\Pr[Z < (1 - 1/2) \mathbb{E}[Z]]$. By the Chernoff inequality, we have $\Pr[Z < (1 - 1/2) \mathbb{E}[Z]] \leq \exp(-\mu1/(2 \cdot 2^2)) = \exp(-\beta/64) \leq 1/n^{10}$, if we set $c = 640$.

Thus, with high probability

$$\|F(x)F(y)\| \geq \sqrt{\sum_{j=1}^{\beta} (f_{u,j}(x) - f_{u,j}(y))^2} \geq \sqrt{\rho^2 \frac{\beta}{16}} = \sqrt{\frac{\beta \rho}{4}} = \phi \cdot \sqrt{\frac{\beta}{64 \ln n}}.
$$

On the other hand, $|f_{i,j}(x) - f_{i,j}(y)| \leq d(x, y) = \phi \leq 16 \rho \ln n$. Thus,

$$\|F(x)F(y)\| \leq \sqrt{\alpha \beta (16 \rho \ln n)^2} \leq 16 \sqrt{\alpha \beta \rho \ln n} = \sqrt{\alpha \beta} \cdot \phi.$$

Thus, setting $G(x) = F(x)^{\frac{64 \ln n}{\sqrt{\beta}}}$, we get a mapping that maps two points of distance $\phi$ from each other to two points with distance in the range $[\phi, \phi \cdot 64 \sqrt{\alpha \ln n}]$. Namely, $G(\cdot)$ is an embedding with distortion $O(\sqrt{\alpha \ln n}) = O(\sqrt{\ln \Phi \ln n})$.

The probability that $G$ fails on one of the pairs, is smaller than $(1/n^{10}) \cdot \binom{n}{2} < 1/n^8$. In particular, we can check the distortion of $G$ for all $\binom{n}{2}$ pairs, and if any of them fail (i.e., the distortion is too big), we restart the process.

### 4.2 The unbounded spread case

Our next task, is to extend Theorem 4.2 to the case of unbounded spread. Indeed, let $(X, d)$ be an $n$-point metric, such that $\operatorname{diam}(X) \leq 1/2$. Again, we look on the different resolutions $r_1, r_2, \ldots$, where $r_i = 1/2^{i-1}$. For each one of those resolutions $r_i$, we can embed this resolution into $\beta$ coordinates, as done for the bounded case. Then we concatenate the coordinates together.

There are two problems with this approach, first the number of resulting coordinates is infinite, and second is that a pair $x, y$, might be distorted a “lot” because it contributes to all resolutions, not only to its “relevant” resolutions.

Both problems can be overcome with careful tinkering. Indeed, for a resolution $r_i$, we are going to modify the metric, so that it ignores short distances (i.e., distances $\leq r_i/n^2$). Formally, for each resolution $r_i$, let $G_i = (X, \hat{E}_i)$ be the graph where two points $x, y$ are connected if $d(x, y) \leq r_i/n^2$. Consider a connected component $C \subseteq G_i$. For any two points $x, y \in C$, we have $d(x, y) \leq n(r_i/n^2) \leq r_i/n$. Let $X_i$ be the set of connected components of $G_i$, and define the distances between two connected components $C, C' \in X_i$, to be $d_i(C, C') = d(C, C') = \min_{e \in C, c' \in C'} d(e, c')$.

It is easy to verify that $(X_i, d_i)$ is a metric space (see Exercise 6.1). Furthermore, we can naturally embed $(X, d)$ into $(X_i, d_i)$ by mapping a point $x \in X$ to its connected components.
in $X_i$. Essentially $(X_i, d_i)$ is a snapped version of the metric $(X, d)$, with the advantage that $\Phi((X, d_i)) = O(n^2)$. We now embed $X_i$ into $\beta = O(\log n)$ coordinates. Next, for any point of $X$ we embed it into those $\beta$ coordinates, by using the embedding of its connected component in $X_i$. Let $E_i$ be the embedding for resolution $r_i$. Namely, $E_i(x) = (f_{i,1}(x), f_{i,2}(x), \ldots, f_{i,\beta}(x))$, where $f_{i,j} = \min(\|d(x, V_{i,j})\|, 2r_i)$. The resulting embedding is $F(x) = \oplus E_i(x) = (E_1(x), E_2(x), \ldots)$.

Since we slightly modified the definition of $f_{i,j}(\cdot)$, we have to show that $f_{i,j}(\cdot)$ is nonexpansive. Indeed, consider two points $x, y \in X$, and observe that

$$|f_{i,j}(x) - f_{i,j}(y)| \leq |d(x, V_{i,j}) - d(y, V_{i,j})| \leq d(x, y).$$

For a pair $x, y \in X$, and let $\phi = d(x, y)$. To see that $F(\cdot)$ is the required embedding (up to scaling), observe that, by the same argumentation of Theorem 4.2 we have that with high probability

$$\|F(x)F(y)\| \geq \phi \cdot \frac{\sqrt{\beta}}{64 \ln n}.$$

To get an upper bound on this distance, observe that for $i$ such that $r_i > \phi n^2$, we have $E_i(x) = E_i(y)$. Thus,

$$\|F(x)F(y)\|^2 = \sum_i \|E_i(x) - E_i(y)\|^2 = \sum_{i, r_i < \phi n^2} \|E_i(x) - E_i(y)\|^2 \leq \sum_{i, \phi n^2 < r_i < \phi n^2} \|E_i(x) - E_i(y)\|^2 + \sum_{i, r_i < \phi n^2} \|E_i(x) - E_i(y)\|^2 = \beta \phi^2 \log n^4 + \sum_{i, r_i < \phi n^2} (2r_i)^2 \beta \leq 4 \beta \phi^2 \log n^4 + \frac{4 \beta \phi^2}{n^4} \leq 5 \beta \phi^2 \log n.$$

Thus, $\|F(x)F(y)\| \leq \phi \sqrt{5 \beta \log n}$. We conclude, that with high probability, $F(\cdot)$ is an embedding of $X$ into Euclidean space with distortion $(\phi \sqrt{5 \beta \log n})/(\phi \cdot \frac{\sqrt{\beta}}{64 \ln n}) = O(\log^{3/2} n)$.

We still have to handle the infinite number of coordinates problem. However, the above proof shows that we care about a resolution $r_i$ (i.e., it contributes to the estimates in the above proof) only if there is a pair $x, y$ such that $r_i/n^2 \leq d(x, y) \leq r_i n^2$. Thus, for every pair of distances there are $O(\log n)$ relevant resolutions. Thus, there are at most $\eta = O(n^2 \beta \log n) = O(n^2 \log^2 n)$ relevant coordinates, and we can ignore all the other coordinates. Next, consider the affine subspace $h$ that spans $F(P)$. Clearly, it is $n - 1$ dimensional, and consider the projection $G : \mathbb{R}^n \rightarrow \mathbb{R}^{n-1}$ that projects a point to its closest point in $h$. Clearly, $G(F(\cdot))$ is an embedding with the same distortion for $P$, and the target space is of dimension $n - 1$.

Note, that all this process succeeds with high probability. If it fails, we try again. We conclude:

**Theorem 4.3 (Low quality Bourgain theorem.)** Given a $n$-point metric $M$, one can embed it into Euclidean space of dimension $n - 1$, such that the distortion of the embedding is at most $O(\log^{3/2} n)$.

Using the Johnson-Lindenstrauss lemma, the dimension can be further reduced to $O(\log n)$. In fact, being more careful in the proof, it is possible to reduce the dimension to $O(\log n)$ directly.

## 5 Bibliographical notes

The partitions we use are due to Calinescu et al. [CKR01]. The idea of embedding into spanning trees is due to Alon et al. [AKPW95], which showed that one can get a probabilistic distortion of
$2^{O\left(\sqrt{\log n \log \log n}\right)}$. Yair Bartal realized that by allowing trees with additional vertices, one can get a considerably better result. In particular, he showed [Bar96] that probabilistic embedding into trees can be done with polylogarithmic average distortion. He later improved the distortion to $O(\log n \log \log n)$ in [Bar98]. Improving this result was an open question, culminating in the work of Fakcharoenphol et al. [FRT03] which achieve the optimal $O(\log n)$ distortion.

Interestingly, if one does not care about the optimal distortion, one can get similar result (for embedding into probabilistic trees), by first embedding the metric into Euclidean space, then reduce the dimension by the Johnson-Lindenstrauss lemma, and finally, construct an HST by constructing a quadtree over the points. The “trick” is to randomly translate the quadtree. It is easy to verify that this yields $O(\log^4 n)$ distortion. See the survey by Indyk [Ind01] for more details. This random shifting of quadtrees is a powerful technique that was used in getting several result, and it is a crucial ingredient in Arora [Arc98] approximation algorithm for Euclidean TSP.

Our proof of Lemma 2.1 (which is originally from [FRT03]) is taken from [KLMN04]. The proof of Theorem 4.3 is by Gupta [Gup00].

A good exposition of metric spaces is available in Matoušek [Mat02].

6 Exercises

Exercise 6.1 (Partition induced metric.)

(a) Let $(X, d)$ be a metric space, and let $P$ be a partition of $X$. Prove that the metric $(P, d')$ is a metric, where $d'(C, C') = d(C, C') = \min_{x \in C, y \in C'} d(x, y)$ and $C, C' \in P$.

(b) Let $(X, d)$ be a $n$-point metric space, and consider the set $U = \left\{ i \mid 2^i \leq d(x, y) \leq 2^{i+1}, \text{ for } x, y \in X \right\}$. Prove that $|U| = O(n)$. Namely, there are only $n$ different resolutions that “matter” for a finite metric space.

Exercise 6.2 Show a finite metric space, such that for any embedding into HST, the distortion of the embedding is $\Omega(n)$ in the worst case.

7 Exercises - not for submission

Exercise 7.1 (Computing the diameter via embeddings.)

(a) (h:1) Let $\ell$ be a line in the plane, and consider the embedding $f : \mathbb{R}^2 \to \ell$, which be a projection of the plane into $\ell$. Prove that $f$ is 1-Lipschitz, but it is not $K$-bi-Lipschitz for any constant $K$.

(b) (h:3) Prove that one can find a family of projections $F$ of size $O(1/\sqrt{\varepsilon})$, such that for any two points $x, y \in \mathbb{R}^2$, for one of the projections $f \in F$ we have $d(f(x), f(y)) \geq (1 - \varepsilon)d(x, y)$.

(c) (h:1) Given a set $P$ of $n$ in the plane, given a $O(n/\sqrt{\varepsilon})$ time algorithm that outputs two points $x, y \in P$, such that $d(x, y) \geq (1 - \varepsilon)diam(P)$, where $diam(P) = \max_{z,w \in P} d(z, w)$ is the diameter of $P$.

(d) (h:2) Given $P$, show how to extract, in $O(n)$ time, a set $Q \subseteq P$ of size $O(\varepsilon^{-2})$, such that $diam(Q) \geq (1 - \varepsilon/2)diam(P)$. (Hint: Construct a grid of appropriate resolution.)
In particular, give an \((1 - \varepsilon)\)-approximation algorithm to the diameter of \(P\) that works in \(O(n + \varepsilon^{-2.5})\) time. (There are slightly faster approximation algorithms known for approximating the diameter.)

**Acknowledgments**

The presentation in this write-up follows closely the insightful suggestions of Manor Mendel.

**References**


