7.1 Embedding Of / Into Trees

We continue our study with considering tree metric spaces, which due to their algorithmic simplicity, play a central role in a wide range of areas of CS.

**Definition 7.1** (Tree metric). Let $T = (V, E, w)$ be a weighted undirected tree graph with $w: E \rightarrow \mathbb{R}^+$, and let $X \subseteq V$. The function $d: \binom{X}{2} \rightarrow \mathbb{R}^+$ defined by $d(x, y) = d_T(x, y)$ is called a tree metric.

Steiner tree is a spanning tree of the set of vertices $X \subseteq V$ that may contain additional vertices to those in $X$ — these vertices called steiner points. Gupta [3] showed that for any tree the steiner points can be efficiently removed, with distortion 8, and in [2] authors have shown that this is tight (for a complete binary tree).

### 7.1.1 Embedding of Tree Metrics into $\ell_p$ Spaces

We start with the case of $\ell_1$, which has a very useful characterization.

#### 7.1.1.1 Characterization of $\ell_1$: Cut Cone

We define the cone of $\ell_1$ metrics, denoted by $\mathbb{L}_1$.

**Definition 7.2.** For $V = \{1, 2, \ldots, n\}$, let

$$\mathbb{L}_1 := \left\{ d \in \mathbb{R}^{\binom{V}{2}} \mid \exists f: V \rightarrow \ell_1, \text{s.t. } \forall i \neq j \in V, \; d(i, j) = \|f(i) - f(j)\|_1 \right\}.$$

Namely, the $\mathbb{L}_1$ is a set of all $\ell_1$ pseudometric spaces defined over $V$. It can be easily checked that this set is a convex cone \footnote{$C$ is a convex cone if for all scalars $\alpha, \beta \geq 0$, and for all vectors $x, y \in C$, $\alpha \cdot x + \beta \cdot y \in C$} (We also can define $\mathbb{L}_p$ convex cone as all pseudometrics to the power $p$ defined on $V$). Next we show that there is tight relation of the cone $\mathbb{L}_1$ to the cut metrics.

**Definition 7.3** (Cut Metric). Let $V$ be any set, and let $S \subseteq V$. Cut (pseudo)metric on $V$ is a function $\gamma_S: \binom{V}{2} \rightarrow \{0, 1\}$ defined as follows:

$$\gamma_S(u, v) = \begin{cases} 
0, & u, v \in S \text{ or } u \in \bar{S}, \\
1, & \text{otherwise}.
\end{cases}$$

Note that $(V, \gamma_S)$ is indeed a pseudometric space. In addition, note that for all $S \subseteq V$, $\gamma_S \in \mathbb{L}_1$, particularly this is 0/1 vector.

**Theorem 7.1.** Let $V$ be a finite set. Then $d \in \mathbb{L}_1$ iff $d$ is a linear combination, with non-negative coefficients, of cut metrics defined over $V$. 
Proof. \( \Leftarrow \) Assume \( d \) is a linear combination, with non-negative coefficients, of cut metrics. We have to show that there is a map \( f : V \to \ell_1 \) such that for all \( u \neq v \in V \), \( d(u,v) = \| f(u) - f(v) \|_1 \). For all \( u \neq v \in V \), let \( d(u,v) = \sum_{S \subseteq V} \delta_S \cdot \gamma_S(u,v) \), such that \( \delta_S \geq 0 \). For each \( u \in V \), define \( f(u) = (\delta_{S_1} \cdot g_{S_1}(u), \ldots, \delta_{S_k} \cdot g_{S_k}(u)) \), where \( g_S : V \to \{0,1\} \) defined by \( g_S(u) = 0 \), for \( u \in S \), and \( g_S(u) = 1 \), otherwise, for any subset \( S \subseteq V \). Therefore, we have

\[
\| f(u) - f(v) \|_1 = \sum_{S \subseteq V} |\delta_S (g_S(u) - g_S(v))| = \sum_{S \subseteq V} \delta_S (|g_S(u) - g_S(v)|) = \sum_{S \subseteq V} \delta_S \gamma_S(u,v) = d(u,v).
\]

\( \Rightarrow \) Assume \( d \in \ell_1 \). We show that \( d \) is a linear combination of cut metrics, with non-negative coefficients. Let \( f : V \to \ell_1 \) be a mapping such that \( \forall u \neq v \in V \), \( d(u,v) = \| f(u) - f(v) \|_1 \). Note that it is enough to prove the claim for \( f : V \to \ell_1^k \). The generalization to \( \ell_1^k \) is straightforward (by applying the same argument to each coordinate separately). Denote by \( x_1 \leq x_2 \leq \ldots \leq x_n \in \mathbb{R} \) the images of \( f(V) \), and define \( \forall x_i \), \( 1 \leq i \leq n \) the set \( S_{x_i} = \{ v \in V | f(v) \leq x_i \} \subseteq V \). For any \( u \neq v \in V \) consider interval \([f(u), f(v)]\) (w.l.g. \( f(u) \leq f(v) \)). Denote \( x_l = f(u), x_r = f(v) \). Note that \( \forall 1 \leq i < l, u,v \in S_{x_i}, \forall l \leq i < r, u \in S_{x_i}, \forall r \leq i \leq n, u,v \in S_{x_i} \). Therefore, for all \( u \neq v \in V \) we have

\[
d(u,v) = |f(u) - f(v)| = \sum_{i=l}^{r-1} |x_{i+1} - x_i| = \sum_{i=1}^{n-1} |x_{i+1} - x_i| \cdot \gamma_{S_{x_i}}(u,v).
\]

If \( x_l = x_r \), then \( 0 = d(u,v) = \sum_{i=1}^{n-1} |x_{i+1} - x_i| \cdot \gamma_{S_{x_i}}(u,v) \). For the case of \( k \) dimensions, we will obtain a combination of \( kn \) cut metrics.

Now we are ready to prove the following theorem.

**Theorem 7.2.** Every finite tree metric is isometrically embeddable into \( \ell_1 \).

Proof. It is enough to show that any finite tree metric is a linear combination of a cut metrics with non-negative coefficients. Indeed, \( \forall x,y \in X \) it holds that \( d_T(x,y) = \sum_{(u,v) \in E} w(u,v) \cdot \lambda_{(u,v)}(x,y) \), where \( \lambda_{(u,v)} \) is a cut metric obtained by removing edge \((u,v)\) from the tree. Note that this results in embedding into \( \ell_1^{n-1} \), where \( n \) is the number of vertices in the tree.

\[ \square \]

The following is known for \( \ell_p, p \neq 1, \infty \).

**Theorem 7.3 ([H]).** Every tree metric on \( n \) points is embeddable into \( \ell_p \) with distortion \( O(\max(p,2)^{\sqrt{\log(\log(n))}}) \), \( p \neq 1, \infty \).

The bound is tight for \( p = 2 \) (for the complete binary tree [Bourgain]). For the special kind of tree metrics - ultrametrics, there is a stronger result of isometric embedability into any \( \ell_p \)-space.

### 7.1.2 Ultrametrics are in \( \ell_p \), for any \( p \geq 1 \)

**Definition 7.4.** A finite metric space \((X,d_x)\) is an ultrametric space if one of the following equivalent statements holds:

1. \( \forall x,y,z \in X \) it holds that \( d_x(x,y) \leq \max\{d_x(x,z),d_x(z,y)\} \).

2. \((X,d_x)\) is a metric space defined on the leaves of the rooted weighted tree, with non-negative weights, such that each path from the root to a leave has the same weight. The metric is the shortest path metric.

3. \((X,d_x)\) is a metric space defined on the leaves of the rooted tree with the labels on its nodes. For each \( v \in T \) there is a label \( \Delta(v) \geq 0 \) such that \( \Delta(v) = 0 \) iff \( v \) is a leaf, and if \( v \) is a child of \( u \), then \( \Delta(v) \leq \Delta(u) \). The distance is defined by \( \forall x,y \in X, d_x(x,y) = \Delta(lca_T(x,y)) \).

7-2
For a parameter $k \geq 1$, a $k$-HST metric is a special type of an ultrametric:

**Definition 7.5** (k-Hierarchically Separated Tree). For $k \geq 1$, $k$-HST metric is defined by the item (3) in the above definition with the stronger requirement: if $v$ is a child of $u$, then $\Delta(v) \leq \frac{\Delta(u)}{k}$.

Note that every ultrametric is $1$-HST metric.

**Theorem 7.4** (Exercise.). Every ultrametric is embeddable into $k$-HST with distortion $k$.

Next we prove the following basic result:

**Theorem 7.5.** Every finite ultrametric isometrically embeds into $\ell_p$, $\forall 1 \leq p \leq \infty$.

*Proof.* Let $(U, d)$ be an ultrametric on $n$ points, given by the labeled tree representation, as in item (3) of the definition, and let $1 < p < \infty$ (for $p = 1, \infty$ the theorem is correct). Let $\Delta(U) := \text{diam}(U)$, and note that $\Delta(U)$ is the label of the root of the tree representing $U$. We build an isometric $f : U \to \ell_p$ inductively: the embedding maps the points of $U$ onto the sphere of radius $\frac{1}{2^{1/p}} \Delta(U)$, i.e. $\forall x \in U$, $\|f(x)\|_p = \frac{1}{2^{1/p}} \Delta(U)$. For $|U| = 1$, the only node in the tree goes to 0. Assume that any ultrametric $U'$ of size $|U'| < n$ can be isometrically embedded on the sphere of radius $\frac{1}{2^{1/p}} \Delta(U')$. For an $n$ point ultrametric $U$, let $T$ denote its tree representation. Let $T_i, 1 \leq i \leq t$, denote the subtrees rooted at the children nodes of the root of $T$. These subtrees define ultrametric subspaces of $U$, denoted by $U_1, U_2, \ldots, U_t$. Therefore, by induction’s assumption, for each $1 \leq i \leq t$, there is an isometric embedding $f_i : U_i \to \ell_p$, such that $\forall x \in U_i$ it holds that $\|f_i(x)\|_p = \frac{1}{2^{1/p}} \Delta(U_i)$. Define $f : U \to \ell_p$ as follows: $\forall x \in U$, $f(x) = \tilde{f}_1(x) \cdot g_1(x), \ldots, \tilde{f}_t(x) \cdot g_t(x)$ (the dot operation here is the concatenation), where

$$
\tilde{f}_i(x) = \begin{cases} f_i(x) & \text{if } x \in U_i, \\ 0 & \text{otherwise.} \end{cases}
$$

and

$$
g_i(x) = \begin{cases} \left( \frac{\Delta(U)^p - \Delta(U_i)^p}{2} \right)^{\frac{1}{p}} & \text{if } x \in U_i, \\ 0 & \text{otherwise.} \end{cases}
$$

For any $x \in U_i$, we have:

$$
\|f(x)\|^p_p = \|f_i(x)\|^p_p + |g_i(x)|^p = \text{induct. assmp.} = \frac{\Delta(U_i)^p}{2} + \frac{\Delta(U)^p - \Delta(U_i)^p}{2} = \frac{1}{2} \Delta(U)^p.
$$

For $x \neq y \in U_i$, for some $i$, it holds that $\|f(x) - f(y)\|_p = \|f_i(x) - f_i(y)\|_p = \text{induct. assmp.} = d(x, y)$.

For $x \in U_i$ and $y \in U_j$, for $i \neq j$, we have

$$
\|f(x) - f(y)\|^p_p = \|f_i(x)\|^p_p + \|f_j(y)\|^p_p + (g_i(x))^p + (g_j(y))^p = \frac{\Delta(U)^p}{2} + \frac{\Delta(U)^p}{2} = \Delta(U)^p = d(x, y)^p.
$$

Note that the dimension of the above embedding is $O(n)$. And the lower bound $\Omega(n)$ is achieved by the equilateral space. Can it be improved in price of distortion?

**Theorem 7.6** (Bartal, Mendel). Every ultrametric on $n$ points is embeddable into $\ell^d_p$ (for all $1 \leq p \leq \infty$), with distortion $1 + \epsilon$ ($0 < \epsilon \leq 1$) and dimension $d = O\left(\frac{\log n}{\epsilon^2}\right)$.
7.2 Embedding into Tree Metrics

We study embeddings of general metric spaces into tree metrics. First we consider embedding $X$ into minimum spanning tree of its graph representation and estimate the distortion.

Theorem 7.7. Let $G = (V, E, w)$ be a weighted graph with non-negative weights, and let $T$ be an MST of $G$. Then (the identity) embedding of $G$ into $T$ has distortion $n - 1$, where $|V| = n$.

Proof. W.l.o.g. the weights of edges of $G$ that form triangle, satisfy the triangle inequality (otherwise remove the edges which violate the property, without affecting the metric defined by the graph). Note that the embedding is non-contractive. Let $x, y \in V$ be any vertices. We show that $d_T(x, y) \leq (n - 1)d_G(x, y)$.

- $(x, y) \in E$: denote by $e$ an edge of the maximum weight on the path of $T$, between $x$ and $y$. It holds that $w(e) \leq w(x, y)$, as otherwise the tree $\hat{T} = T \setminus \{e\} \cup \{(x, y)\}$ is lighter than $T$. In addition, by our assumption, $w(x, y) = d_G(x, y)$. Thus, $d_T(x, y) \leq (n - 1)d_G(x, y)$.

- $(x, y) \notin E$: denote by $x = u_1, u_2, \ldots, u_k = y$ the shortest path between $x$ and $y$ in $G$. Then, by the triangle inequality (on the tree): $d_T(x, y) \leq d_T(u_1, u_2) + d_T(u_2, u_3) + \ldots + d_T(u_{k-1}, u_k) \leq \text{by first case} \leq (n - 1) \sum_{i=1}^{k-1} w(u_i, u_{i+1}) = (n - 1)d_G(x, y)$.

Remark 7.8. Actually this proves that the distortion of the embedding is bounded by $m - 1$, where $m$ is the length of the largest cycle in the graph.

Can we do better in terms of distortion? The answer is negative. For example, an embedding of an unweighted $n$-point cycle into its MST incurs distortion $n - 1$. Although, it can be shown that there is an embedding of $C_n$ into a weighted Steiner tree, with distortion $< n - 1$. How well can we do?

Theorem 7.9 ([5]). Any embedding of $C_n$ into a tree metric has distortion at least $\frac{2}{3} - 1$.

Proof. Assume $n = 3k$ and consider 3 points $A, B, C$ in $C_n$ such that the distance between each pair is $k$. Let $f: C_n \rightarrow T$ be an embedding of $C_n$ into a tree $T$. Assume without loss of generality that $f$ is non-expansive. The (shortest) paths $f(A) \rightarrow f(B), f(B) \rightarrow f(C)$ and $f(A) \rightarrow f(C)$ in tree $T$ split on the same node $O$ (otherwise there is a cycle in the tree). Consider the path $f(A) \rightarrow f(B)$ in the tree. The node $O$ is somewhere on the path ($O$ might be $f(A), f(B), f(C)$). Denote the points of the $A - B$ chain in the circuit $C_n$ by $x_1 = A, x_2, \ldots, x_{\frac{1}{k}n + 1} = B$. Consider nodes $f(x_1) = f(A), f(x_2), \ldots, f(x_{\frac{1}{k}n + 1}) = f(B)$ on the tree $T$. Note that these nodes do not necessary form path on the tree. There exists (at least one) pair $(x_i, x_{i+1}) \in C_n$ such that paths $O \rightarrow f(x_i)$ and $O \rightarrow f(x_{i+1})$ on the tree split exactly on the node $O$ (otherwise there is a cycle in the tree, since $O$ is lying on the tree path between $f(A)$ and $f(B)$). Denote such a pair by $(x_i, x_{i+1})$. As $f$ is non-expansive we obtain $d_T(f(x_i), f(x_{i+1})) \leq d_{C_n}(x_i, x_{i+1}) = 1$. Therefore at least one of $x_i, x_{i+1}$ is at distance at most $\frac{1}{2}$ from $O$. Denote this node by $z_{AB}$ and its preimage by $x_{AB}$.

The same argument applies for paths $f(A) \rightarrow f(C)$ and for $f(B) \rightarrow f(C)$. Namely, there exist nodes $z_{AC}$ and $z_{BC}$ with distance at most $\frac{1}{2}$ to $O$ and with preimages $x_{AC}$ and $x_{BC}$.

Therefore, we obtain that the distance between any pair of $\{z_{AB}, z_{AC}, z_{BC}\}$ is at most 1. In addition, there at least one pair of $(x_{AB}, x_{AC}, x_{BC})$ such that distance (in $C_n$) between them is at least $\frac{n}{3}$. Thus, $f$ has distortion at least $\frac{n}{3}$. In the case of general $n$ (i.e., $n$ which is not a factor of 3), the bound on distortion is $n/3 - 1$, since we can find a pair on the cycle of distance at least $n/3 - 1$, which would contract to at most 1 on the tree.

\[\square\]

Theorem 7.10 (Har-Peled, Mendel). Every metric space on $n$ points is embeddable into ultrametric with distortion $(n - 1)$.
It can be shown that there exists an $n$-point metric space (a line metric) such that any its embedding into an ultrametric requires distortion at least $n - 1$. In the next lecture we will prove the following:

**Theorem 7.11 (II).** Every metric space on $n$ points is embeddable into an ultrametric, by the non-contractive embedding $f$ with distortion $O(n)$, and with the following bounds on $\ell_q$-distortion:

- for $1 \leq q < 2$, $\ell_q - \text{dist}(f) = O(1)$;
- for $q = 2$, $\ell_q - \text{dist}(f) = O(\sqrt{\log n})$;
- for $2 < q \leq \infty$, $\ell_q - \text{dist}(f) = O(n^{1-2/q})$.

**References**


