## 4.3 Embeddings Into Normed Spaces

In this section, we discuss low-distortion embeddings of general metric spaces into $\ell_p^d = (\mathbb{R}^d, \|\cdot\|_p)$, with particular emphasis on the cases $p = 1, 2, \infty$.

Let $(X, \rho)$ be a finite metric space. To specify an embedding $f : X \to \ell_p$, we have to define $d$ coordinate functions $f_i : X \to \mathbb{R}$, $1 \leq i \leq d$.

Our main approach to construct such functions will be to consider distances from (nonempty) subsets. More precisely, let $\emptyset \neq A \subseteq X$. For $x \in X$, we define the distance from $x$ to $A$ as

$$\rho(x, A) := \min_{a \in A} \rho(x, a).$$

In particular, this defines a function $f_A : X \to \mathbb{R}$, $f_A(x) = \rho(x, A)$. As a first application, we consider embeddings into $\ell_\infty$. For future reference, we note that

$$\nu_A(x, y) := |f_A(x) - f_A(y)| \leq \rho(x, y) \tag{4.1}$$

for all $x, y$. We leave the proof as an exercise.

**Proposition 4.8.** *Let $(X, \rho)$ be a metric space with $|X| = n$ points. Then there is an isometry $f : X \to \ell_\infty^n$, i.e., a map such that $\|f(x) - f(y)\|_\infty = \rho(x, y)$ for all $x, y \in X$.*

*Proof.* We index the coordinates of $\mathbb{R}^n$ by the elements $a \in X$ and use the above functions $f_A$ for all singleton sets $A = \{a\}$. Thus, we take the functions $f_a(x) = \rho(x, a)$ (where we simplify notation and write $f_a$ instead of $f_{\{a\}}$) and let $f = (f_a)_{a \in X} : X \to \mathbb{R}^n$.

By (4.1), we have $|f_a(x) - f_a(y)| \leq \rho(x, y)$ for all $x, y, a \in X$. This shows that $\|f(x) - f(y)\|_\infty = \max_A |f_A(x) - f_A(y)| \leq \rho(x, y)$. On the other hand, given $x, y \in X$, consider $a = x$, for instance. We have $f_x(x) = 0$ and $f_x(y) = \rho(x, y)$, i.e., $\|f(x) - f(y)\|_\infty \geq |f_x(x) - f_x(y)| = \rho(x, y)$. $\square$

The goal of this section is to prove the following, more general result concerning embeddings into $\ell_p^d$.

**Theorem 4.9.** *Let $1 \leq p < \infty$. Every $n$-point metric space $(X, \rho)$ can be embedded into some $\ell_p^d$ with distortion at most $O(\log n)$.*

We will spell the proof out in detail for the cases $p = 1$ and $p = 2$. With minor modifications the same proof also works for embeddings into $(\mathbb{R}^d, \|\cdot\|_p)$ for any $p$-norm, $1 \leq p \leq \infty$. The proof is probabilistic, and at first seems to require a target dimension of $d = 2^n$ (where $n = |X|$, as before). Using Chernov bounds, this can be pushed down to $d = O(\log^2 n)$. Moreover, the proof immediately yields a randomized algorithm to construct such embeddings. We will say more about this along the way, while leaving some of the details as exercises.

**Lemma 4.10.** *Let $(X, \rho)$ be a finite metric space, let $A_1, \ldots, A_d \subseteq X$ be nonempty subsets, and let $\nu_i(x, y) = |\rho(x, A_i) - \rho(y, A_i)|$, $1 \leq i \leq d$. Assume that there are real numbers $D \geq 1$ and $\alpha_1, \ldots, \alpha_d \geq 0$ such that $\sum_{i=1}^d \alpha_i = 1$ and*

$$\sum_{i=1}^{d} \alpha_i \nu_i \geq \frac{1}{D} \rho.$$

*Then for every $1 \leq p \leq \infty$, there is an embedding $f : (X, \rho) \to \ell_p^d$ with distortion at most $D$.*

*Proof.* We give the details for $p = 1, 2$. For $1 \leq i \leq d$, let $f_i(x) = \rho(x, A_i)$. We define $f : V \to \mathbb{R}^d$ by

$$f(x) = (\alpha_1 f_1(x), \ldots, \alpha_d f_d(x)).$$

By the second assumption on the $\alpha_i$ and by (4.1),

$$\|f(x) - f(y)\|_1 = \sum_i \alpha_i \underbrace{\nu_i(x, y)}_{\leq \rho(x,y)} \leq \rho(x, y).$$

On the other hand, again by our assumptions on the $\alpha_i$ and the $\nu_i$,

$$\|f(x) - f(y)\|_1 = \sum_{i=1}^{d} \alpha_i |f_i(x) - f_i(y)| = \sum_i \alpha_i \nu_i(x, y) \geq \frac{1}{D} \rho(x, y).$$

In the case $p = 2$, we define

$$f(x) = (\sqrt{\alpha_1} f_1(x), \ldots, \sqrt{\alpha_d} f_d(x)).$$

Then

$$\|f(x) - f(y)\|_2^2 = \sum_{i=1}^{d} \alpha_i \nu_i(x, y)^2 \leq \underbrace{\Big(\sum_i \alpha_i\Big)}_{=1} \rho(x, y)$$

for all $x, y \in X$. Moreover, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\frac{1}{D} \rho(x, y) \leq \sum_{i=1}^{d} \alpha_i \nu_i(x, y) &= \sum_{i=1}^{d} \sqrt{\alpha_i} \Big( \sqrt{\alpha_i} \nu_i(x, y) \Big) \\
&\leq \underbrace{\Big( \sum_{i=1}^{d} \alpha_i \Big)^{1/2}}_{=1} \Big( \sum_{i=1}^{d} \alpha_i \nu_i(x, y)^2 \Big)^{1/2} = \|f(x) - f(y)\|_2
\end{aligned}$$

For arbitrary $1 < p < \infty$, one uses the function $f(x) = (\sqrt[p]{\alpha_1} f_1(x), \ldots, \sqrt[p]{\alpha_d} f_d(x))$ and Hölder's inequality. The details are left as an exercise. $\square$

Here is the main technical lemma that allows us to find the sets $A_i$ that we need for the preceding lemma:

**Lemma 4.11.** *Let $u, v \in X$, $u \neq v$. Then there exist real numbers $\Delta_1, \Delta_2, \ldots, \Delta_q \geq 0$ with $\sum_{j=1}^{q} \Delta_j = \frac{1}{4}\rho(u,v)$, where $q = \lfloor \log_2 n \rfloor + 1$, such that the following holds for all $1 \leq j \leq q$:*

*Let $A_j$ be a random subset of $X$, where we include each point of $V$ independently with probability $2^{-j}$. Then*

$$\Pr[\underbrace{|\rho(u,A_j) - \rho(v,A_j)|}_{\nu_{A_j}(u,v)} \geq \Delta_j] \geq \frac{1}{12}.$$

*Consequently, the expectation of the nonnegative random variable $\nu_{A_j}(u,v)$ satisfies $\mathbf{E}[\nu_{A_j}(u,v)] \geq \frac{1}{12}\Delta_j$.*

*Proof of Lemma 4.11.* Fix $u, v \in X$, $u \neq v$. Set $r_q := \frac{1}{4}\rho(u,v)$. For $0 \leq j \leq q-1$, define $\tilde{r}_j$ to be the smallest radius such that both $|B_{\tilde{r}_j}(u)| \geq 2^j$ and $|B_{\tilde{r}_j}(v)| \geq 2^j$, where $B_r(x) = \{y \in X : \rho(x,y) \leq r\}$. Set $r_j = \min\{r_q, \tilde{r}_j\}$. We will show that the lemma is true with $\Delta_j := r_j - r_{j-1}$.

Fix $j \in \{1, \ldots, q\}$. We may assume that $r_{j-1} = \tilde{r}_{j-1} < r_q$, since otherwise $\Delta_j = 0$ and the conclusion of the lemma holds trivially. Consequently, both closed balls $B_{r_{j-1}}(u)$ and $B_{r_{j-1}}(v)$ contain at least $2^{j-1}$ points. Let $A_j$ be a random subset of $X$ with point probability $2^{-j}$. By definition of $r_j$, at least one of the *open* balls $B^\circ_{r_j}(u) = \{x \in X : \rho(x,u) < r_j\}$ and $B^\circ_{r_j}(v)$ contains less than $2^j$ points, say $|B^\circ_{r_j}(u)| < 2^j$ (this also holds for $j = q$, since $|X| \leq 2^q$). Call $A_j$ *good* if $A_j \cap B_{r_{j-1}}(v) \neq \emptyset$ and $A_j \cap B^\circ_{r_j}(u) = \emptyset$. The former ball has cardinality at least $2^{j-1}$ and the latter at most $2^j$, and the two balls are disjoint. Therefore,

$$\Pr[A_j \cap B_{r_{j-1}}(v) \neq \emptyset] = 1 - (1 - 2^{-j})^{2^{j-1}} \geq 1 - e^{-2^{-j}2^{j-1}} \geq 1 - e^{-1/2} > 1/3.$$

On the other hand,

$$\Pr[A_j \cap B^\circ_{r_j}(u) = \emptyset] \geq (1 - 2^{-j})^{2^j} \geq 1/4,$$

since $0 \leq 2^{-j} \leq 1/2$. The two events are independent, so we may multiply their probabilities and conclude that with probability at least $1/12$, $A_j$ is good. But then $\rho(u, A_j) \geq r_j$ and $\rho(v, A_j) \leq r_{j-1}$, so $\nu_{A_j}(u,v) \geq \Delta_j$. □

*Proof that Lemma 4.11 implies Theorem 4.9.* For each subset $A \subseteq X$, let $\nu_A(u,v) = |\rho(u,A) - \rho(v,A)|$. We want to apply Lemma 4.11. For $1 \leq j \leq q$, let $\pi_j(A) := \Pr[A_j = A]$, where $A_j$ is the random subset with point probability $2^{-j}$ as in the lemma. Then by the lemma, for every pair of distinct points $u, v \in X$,

$$\sum_{A \subseteq X} \pi_j(A) \cdot \nu_A(u,v) = \mathbf{E}[\nu_{A_j}(u,v)] \geq \frac{1}{12}\Delta_j.$$

Summing over all $1 \leq j \leq q$ and exchanging the order of summation, we conclude

$$\sum_{A \subseteq X} \left( \sum_{j=1}^{q} \pi_j(A) \right) \cdot \nu_A(u,v) \geq \frac{1}{12} \sum_{j=1}^{q} \Delta_j = \frac{1}{48} \rho(u,v).$$

Dividing by $q$, we obtain

$$\sum_{A \subseteq X} \underbrace{\left( \frac{1}{q} \sum_{j=1}^{q} \pi_j(A) \right)}_{=:\, \alpha_A} \nu_A(u,v) \geq \frac{1}{48q} \rho(u,v)$$

for all $u, v \in X$. Moreover, for each $j$, we have $\sum_{A \subseteq X} \pi_j(A) = 1$, hence $\sum_{A \subseteq X} \alpha_A = 1$. Thus, by Lemma 4.10, there is an embedding $f : (X, \rho) \to (\mathbb{R}^d, \|\cdot\|_2)$, $d = 2^n$, with distortion at most $48q = O(\log n)$. □

**Remark 4.12.** The proof as it stands yields an embedding dimension of $d = 2^n$. The target dimension can still brought down to $O(\log^2 n)$, using Chernov bounds.

We now describe how this is done. We need the following asymmetric version of the Chernov bound.

**Theorem 3.11 (Asymmetric Chernov bound).** *Let $Y_1, \ldots, Y_m$ be $m$ independent random variables on some probability space that take values in $\{0, 1\}$, and let $Y = Y_1 + \ldots + Y_m$. Then, for all $a > 0$,*

$$\Pr[Y - \mathbf{E}[Y] > a] < e^{-2a^2/m} \quad \text{and} \quad \Pr[Y - \mathbf{E}[Y] < -a] < e^{-2a^2/m}.$$

The proof is very similar to the proof for symmetric $\pm 1$-random variables (where we assumed $\Pr[X_i = +1] = \Pr[X_i = -1] = 1/2$), except that one has to replace the estimate $\frac{e^x + e^{-x}}{2} < e^{x^2/2}$ by the more complicated estimate $pe^{(1-p)x} + (1-p)e^{-px} \leq e^{x^2/8}$, valid for all $p \in [0, 1]$.

**Lemma 4.13.** *Let $(X, \rho)$, $|X| = n$, and $q = \lfloor \log n \rfloor + 1$ be as before. For each $1 \leq j \leq q$, consider $m = C \log n$ independent random subsets $A_j^1, \ldots, A_j^m$ with point probability $2^{-j}$ each, where $C > 0$ is a sufficiently large constant.*

*Then, with probability at least $1 - 1/n$,*

$$\sum_{i=1}^{m} \sum_{j=1}^{q} \frac{1}{mq} \nu_{A_j^i}(u,v) \geq \frac{1}{96q} \rho(u,v)$$

*holds simultaneously for all $u, v \in X$.*

*Proof.* Let the sets $A_j^i$ be as in the lemma. Fix $u, v \in X$, $u \neq v$, let $1 \leq i \leq m$, and consider the random variables

$$Y_j^{u,v,i} := \begin{cases} 1 & \text{if } \nu_{A_j^i}(u,v) \geq \Delta_j \\ 0 & \text{else.} \end{cases}$$

By Lemma 4.11, we have $\mathbf{E}[Y_j^{u,v,i}] = \Pr[Y_j^{u,v,i} = 1] \geq 1/12$ for all $u, v, i, j$. Let $Y_j^{u,v} := \sum_{i=1}^m Y_j^{u,v}$. By linearity of expectation, $\mathbf{E}[Y_j^{u,v}] \geq m/12$. We apply the asymmetric Chernov bound and see that for any particular choice of $u$, $v$, and $j$,

$$\Pr[Y_j^{u,v} \leq m/24] = \Pr[Y_j^{u,v} - \mathbf{E}[Y_j^{u,v}] < -m/24] < e^{-2(m/24)^2/m} = e^{-m/288} < 1/n^4,$$

if we set $m = 4 \cdot 288 \log n$. Consequently,

$$\Pr[Y_j^{u,v} \leq m/24 \text{ for some } u, v, j] < \binom{n}{2}(\log n + 1)/n^4 < 1/n.$$

Otherwise, i.e., if $Y_j^{u,v} \geq m/24$ for all $u, v, j$ (this happens with probability at least $1 - 1/n$), we have $\sum_{i=1}^m \nu_{A_j^i}(u,v) \geq \Delta_j m/24$, i.e.,

$$\sum_{j=1}^q \sum_{i=1}^m \nu_{A_j^i}(u,v) \geq \frac{m}{24} \sum_{j=1}^q \Delta_j = \frac{m}{96}\rho(u,v)$$

for all $u, v$. Dividing both sides by $mq$, we conclude that with probability at least $1 - 1/n$,

$$\sum_{j=1}^q \sum_{i=1}^m \frac{1}{mq} \nu_{A_j^i}(u,v) \geq \frac{1}{96q}\rho(u,v)$$

for all $u, v$. □

It follows from Lemma 4.10 that for $1 \leq p \leq \infty$, there is a map $f : X \to \ell_p^{mq}$ with distortion at most $96q = O(\log n)$ obtained by setting $f_{ij}(x) := \rho(x, A_j^i)$ and $f = (\sqrt[p]{1/mq}) f_{ij})$.

Since we can check the distortion of a given map $(X, \rho) \to (\mathbb{R}^d, \|\cdot\|_p)$ efficiently (in the most simple-minded way, by computing all the pairwise distances in the target space), we immediately obtain a randomized algorithm that computes an embedding with $O(\log n)$-distortion in expected polynomial time (choose the sets $A_j^i$ as above, compute the distortion of the resulting map, and if it is too large, repeat.)

Summarizing, we obtain the following:

**Theorem 4.14.** *If $(X, \rho)$ is an $n$-point metric space and if $1 \leq p \leq \infty$, then there is a mapping $f : (X, \rho) \to (\mathbb{R}^d, \|\cdot\|_p)$ of distortion at most $O(\log n)$, where $d = O(\log^2 n)$.[3] Moreover, there is a randomized algorithm that computes the mapping in expected polynomial time.*

[3]For $p = 2$, the target dimension can be reduced to $O(\log n)$.

## 4.4 Approximating Sparsest Cuts

As an application, we will now discuss how to use Theorem 4.14 to obtain an $O(\log n)$-approximation algorithm for the SPARSESTCUT problem.

Let $G = (V, E)$ be a graph. We recall that a *cut* in $G$ is simply a partition $V = A \cup B$ of the vertex set $V$ into disjoint subsets. A cut is *nontrivial* if both parts are nonempty.

**Definition 4.15.** *The* density *of a cut $V = A \cup B$ is defined as the ratio*

$$\delta(A, B) := \frac{e(A, B)}{|A||B|},$$

*where $e(A, B) := |\{e = \{u, v\} \in E : u \in A$ and $v \in B$, or vice versa$\}|$ is the number of edges connecting $A$ and $B$.*

The SPARSESTCUT problem is the optimization problem of finding a cut of minimal density for a given graph. The problem (and its generalizations to weighted graphs and, more generally, to multicommodity flow networks) has many applications, for instance to VLSI design, see, for instance, [LR99] and the references therein. Even the basic version described above is NP-complete.

**Theorem 4.16.** *There is a randomized algorithm that, given a graph $G = (V, E)$, finds a cut $V = A \cup B$ whose density is at most $O(\log n)$ times the density of a sparsest cut. The expected runtime of the algorithm is polynomial in the size of the graph.*

(We remark that the current record approximation ratio is $O(\sqrt{\log n})$). We recall the notion of a *pseudometric* on a finite set $V$. This is simply a function $\rho : V \times V \to \mathbb{R}$ that satisfies all the axioms of a metric, except that distinct points may have "distance" zero. To be more precise, $\rho$ is a pseudometric if for all $u, v, w \in V$,

1. $\rho(u, v) \geq 0$;

2. $\rho(u, v) = \rho(v, u)$;

3. $\rho(u, w) \leq \rho(u, v) + \rho(v, w)$.

Every cut $V = A \cup B$ defines a pseudometric as follows. We can identify the cut with the function $f_{A,B} : V \to \{0, 1\}$ given by $f_{A,B}(v) = 0$ if $v \in A$ and $f(v) = 1$ if $v \in B$. Then the corresponding *cut pseudometric* is given by $\tau_{A,B}(u, v) := |f_{A,B}(u) - f_{A,B}(v)|$. We can express the density of the cut in terms of this pseudometric. For a subset $F \subseteq \binom{V}{2}$, let $\tau_{A,B}(F) := \sum_{\{u,v\} \in F} \tau_{A,B}(u, v)$. Then the density of the cut equals

$$\delta(A, B) = \frac{\tau_{A,B}(E)}{\tau_{A,B}(\binom{V}{2})}. \tag{4.2}$$

Thus, finding the sparsest cut is equivalent to minimizing the above expression over all cut pseudometrics.

**Step 1: Minimizing the ratio for pseudometrics.** The first step of the approximation algorithm consists of relaxing this problem and minimizing the ratio $\delta(\rho) := \rho(E)/\rho(\binom{V}{2})$ over all pseudometrics $\rho$. Observe that this ratio does not change if we multiply $\rho$ by any constant $C > 0$, i.e., $\delta(\rho) = \delta(C\rho)$. Moreover, the minimum is clearly attained for some pseudometric with $\rho(\binom{V}{2}) > 0$ (else $\delta$ is infinite). Thus, dividing $\rho$ by $\rho(\binom{V}{2})$, if necessary, we may restrict our attention to pseudometrics such that $\rho(\binom{V}{2}) = 1$.

**Observation 4.17.** *Fix a labeling $V = \{v_1, \ldots, v_n\}$ of the elements of $V$. Then there is an obvious 1-1 correspondence (given by $x_{ij} \leftrightarrow \rho(v_i, v_j)$) between pseudometrics $\rho$ on $V$ and $n \times n$ matrices $X = [x_{ij}] \in \mathbb{R}^{n \times n}$ such that for $1 \leq i, j, k \leq n$,*

*(1′)* $x_{ij} \geq 0$

*(2′)* $x_{ij} = x_{ji}$

*(3′)* $x_{ik} \leq x_{ij} + x_{jk}$.

Thus, Step 1 of the approximation algorithm amounts to solving the following linear programming problem:[4]

$$\begin{array}{rl} \text{minimize} & \sum_{i<j,\{i,j\}\in E} x_{ij} \\ \text{over all} & X = [x_{ij}] \in \mathbb{R}^{n\times n} \text{ s.t. (1′) – (3′) hold and} \\ & \sum_{1\leq i<j\leq n} x_{ij} = 1. \end{array} \tag{4.3}$$

**Step 2: Embedding the solution to Step 1 into $\ell_1^{O(\log^2 n)}$ with distortion $O(\log n)$.** Let $\rho_0$ be the pseudometric corresponding to an optimal solution of the linear program (4.3). Note that $\delta(\rho_0)$ is a lower bound for the density of a sparsest cut. Let $f : (V, \rho_0) \to \ell_1^d$ be an embedding with distortion $D = O(\log n)$, where $d = O(\log^2 n)$. Such an embedding exists, and we can find it in randomized polynomial time, by[5] Theorem 4.14.

[4]Recall that linear programming is the problem of maximizing or minimizing a linear function in finitely many variables subject to finitely many linear inequalities. Linear programs can be solved efficiently, i.e., in time that is polynomial in the number of variables, the number of inequalities, and the bitsizes of the coefficients. (In our case, all the coefficients are $0$ or $1$.)

[5]Strictly speaking, Theorem 4.14 guarantees such an embedding for *metrics*, not for pseudometrics. That is not a problem, though. If $\rho_0$ is not a proper metric, greedily pick a subset $W \subseteq V$ such that $\rho_0(u, v) \neq 0$ for all $u, v \in W$ with $u \neq v$. Thus, for every $v \in V$, there exists a unique $g(v) \in W$ such that $\rho_0(v, g(v)) = 0$. If we have low distortion embedding of $f : W \to \ell_1^d$, and we can extend it to an embedding of $V$ by setting $f(v) := f(g(v))$.

Let $\sigma_0(u,v) := \|f(u)-f(v)\|_1$ be the $\ell_1$-pseudometric induced by this embedding. By assumption on $f$, we have $\sigma_0(E) \leq \rho_0(E)$ and $\sigma_o(\binom{V}{2}) \geq \frac{1}{D}\rho_0(\binom{V}{2})$, hence

$$\delta(\sigma_0) \leq D\delta(\rho_0)$$

**Step 3. From $\ell_1$-pseudometrics to cut pseudometrics.** We finish by applying two easy lemmas, whose proofs we leave as exercises:

**Lemma 4.18.** *Let $\sigma$ be a pseudometric induced on an $n$-element set $V$ by a map $f : V \to \ell_1^d$. Then there exist $N \leq d(n-1)$ cut pseudometrics $\tau_1, \ldots, \tau_N$ on $V$ and positive coefficients $\lambda_1, \ldots, \lambda_N > 0$ such that $\sigma = \lambda_1\tau_1 + \ldots + \lambda_N\tau_N$.*

**Lemma 4.19.** *Let $a_1, \ldots, a_N \geq 0$, $b_1, \ldots, b_N > 0$, and $\lambda_1, \ldots, \lambda_N > 0$. Then*

$$\frac{\lambda_1 a_1 + \cdots + \lambda_N a_N}{\lambda_1 b_1 + \cdots + \lambda_N b_N} \geq \min_{1\leq i\leq N} \frac{a_i}{b_i}.$$

Using the first lemma, we write the $\ell_1$-pseudometric $\sigma_0$ as a nonnegative combination

$$\sigma_0 = \lambda_1\tau_1 + \ldots + \lambda_N\tau_N$$

of cut pseudometrics with $N \leq O(n\log^2 n)$ (note that for this step, we do not really need an embedding dimension that is polylogarithmic, any dimension polynomial in $n$ would do).

Let $\tau \in \{\tau_1, \ldots, \tau_N\}$ be the cut pseudometric minimizing the ratio $\delta(\tau) = \frac{\tau(E)}{\tau(\binom{V}{2})}$. It follows from the second lemma that $\delta(\tau) \leq \delta(\sigma_0) \leq D\delta(\rho_0)$, where $D = O(\log n)$, and as we noted above, $\delta(\rho_0)$ is a lower bound for the density of the sparsest cut. Thus, $\tau$ and the corresponding cut $V = A \cup B$ is indeed an $O(\log n)$-approximation of the sparsest cut.