Basic Properties of Metric and Normed Spaces

Computational and Metric Geometry

Instructor: Yury Makarychev

The second part of this course is about metric geometry. We will study metric spaces, low distortion metric embeddings, dimension reduction transforms, and other topics. We will discuss numerous applications of metric techniques in computer science.

1 Definitions and Examples

1.1 Metric and Normed Spaces

Definition 1.1. *A metric space is a pair (X, d), where X is a set and d is a function from $X \times X$ to $\mathbb{R}$ such that the following conditions hold for every $x, y, z \in X$.*

1. *Non-negativity:* $d(x, y) \geq 0$.
2. *Symmetry:* $d(x, y) = d(y, x)$.
3. *Triangle inequality:* $d(x, y) + d(y, z) \geq d(x, z)$.
4. $d(x, y) = 0$ *if and only if* $x = y$.

Elements of X are called points of the metric space, and d is called a metric or distance function on X.

Exercise 1. *Prove that condition 1 follows from conditions 2-4.*

Occasionally, spaces that we consider will not satisfy condition 4. We will call such spaces *semi-metric spaces.*

Definition 1.2. *A space (X, d) is a semi-metric space if it satisfies conditions 1-3 and 4′:*

4′. *if* $x = y$ *then* $d(x, y) = 0$.

Examples. Here are several examples of metric spaces.

1. **Euclidean Space.** Space $\mathbb{R}^d$ equipped with the Euclidean distance $d(x, y) = \|x - y\|_2$.

2. **Uniform Metric.** Let X be an arbitrary non-empty set. Define a distance function $d(x, y)$ on X by $d(x, y) = 1$ if $x \neq y$ and $d(x, x) = 0$. The space (X, d) is called a *uniform or discrete metric space.*

3. **Shortest Path Metric on Graphs.** Let $G = (V, E, l)$ be a graph with positive edge lengths $l(e)$. Let $d(u, v)$ be the length of the shortest path between u and v. Then (V, d) is *the shortest path metric* on G.

4. **Tree Metrics.** A very important family of graph metrics is the family of *tree metrics.* A tree metric is the shortest path metric on a tree T.

5. **Cut Semi-metric.** Let V be a set of vertices and $S \subset V$ be a proper subset of V. Cut semi-metric δ_S is defined by $\delta_S(x, y) = 1$ if $x \in S$ and $y \notin S$, or $x \notin S$ and $y \in S$; and $\delta_S(x, y) = 0$, otherwise. In general, the space (X, d) is not a metric since $d(x, y) = 0$ for some $x \neq y$. Nevertheless, $\delta_S(x, y)$ is often called a cut metric.

Definition 1.3. *A normed space is a pair $(V, \|\cdot\|)$, where V is a linear space (vector space) and $\|\cdot\| : V \to \mathbb{R}$ is a norm on V such that the following conditions hold for every $x, y \in V$.*

1. $\|x\| > 0$ *if* $x \neq 0$.
2. $\|x\| = 0$ *if and only if* $x = 0$.
3. $\|\alpha x\| = |\alpha| \cdot \|x\|$ *for every* $\alpha \in \mathbb{R}$.
4. $\|x + y\| \leq \|x\| + \|y\|$ *(convexity).*

Every normed space $(V, \|\cdot\|)$ is a metric space with metric $d(x, y) = \|x - y\|$ on V.

Definition 1.4. *We say that a sequence of points x_i in a metric space is a Cauchy sequence if*

$$\lim_{i \to \infty} \sup_{j \geq i} d(x_i, x_j) = 0.$$

A metric space is complete if every Cauchy sequence has a limit. A Banach space is a complete normed space.

Remark 1.5. *Every finite dimensional normed space is a Banach space. However, an infinite dimensional normed space may or may not be a Banach space. That said, all spaces we discuss in this course will be Banach spaces. Further, for every normed (metric) space V there exists a Banach (complete) space V' that contains it such that V is dense in V'. Here is an example of a non-complete normed space. Let V be the space of infinite sequences $a(1), a(2), \ldots, a(n), \ldots$ in which only a finite number of terms $a(i)$ are non-zero. Define $\|a\| = \sum_{i=1}^{\infty} |a(i)|$. Then $(V, \|\cdot\|)$ is a normed space but it is not complete, and thus $(V, \|\cdot\|)$ is not a Banach space. To see that, define a sequence a_i of elements in V as follows: $a_i(n) = 1/2^n$ if $n \leq i$ and $a_i(n) = 0$, otherwise. Then a_i is a Cauchy sequence but it has no limit in V. Space ℓ_1, which we will define in the next section, is the completion of $(V, \|\cdot\|)$.*

1.2 Lebesgue Spaces $L_p(X, \mu)$

In this section, we define Lebesgue spaces, a very important class of Banach spaces. Let (X, μ) be a measure space. We consider the set of measurable real valued functions on X. For $p \geq 1$, we define the the p–norm of a function f by

$$\|f\|_p = \left(\int_X |f(x)|^p d\mu(x)\right)^{1/p}.$$

If the integral above is infinite (diverges), we write $\|f\|_p = \infty$. Similarly, we define

$$\|f\|_\infty = \sup |f(x)|.$$

Now we define the Lebesgue space $L_p(X, \mu)$ (for $1 \leq p \leq \infty$):

$$L_p(X, \mu) = \{f : f \text{ is measurable w.r.t. measure } \mu;\ \|f\|_p < \infty\}.$$

Caveat: The norm $\|f\|_p$ can be equal to 0 for a function $f \in L_p(X, \mu)$, which is not identically equal to 0. So formally $L_p(X, \mu)$ (as defined above) is not a normed space. The standard way to resolve this problem is to identify functions that differ only on a set of measure 0. The norm $\|\cdot\|_\infty$ is usually defined as

$$\|f\|_\infty = \operatorname{ess\,sup}_{x \in X} |f(x)| = \inf\{\sup_{x \in X} |\tilde{f}(x)| : \tilde{f}(x) = f(x) \text{ almost everywhere}\}.$$

Examples. Consider several examples of L_p–spaces.

1. Space ℓ_p. Let $X = \mathbb{N}$, and μ be the counting measure; i.e. $\mu(S) = |S|$ for $S \subset \mathbb{N}$. The elements of ℓ_p are infinite sequences of real numbers $a = (a_1, a_2, \dots)$ (which we identify with maps from $\mathbb{N}$ to $\mathbb{R}$) s.t. $\|a\|_p < \infty$. The $p-$norm of a sequence $a = (a_1, a_2, \dots)$ equals
$$\|a\|_p = \left(\sum_{i=1}^{\infty} |a|^p\right)^{1/p}.$$

2. Space ℓ_p^d. Let $X = \{1, \dots, d\}$, and μ be again the counting measure; i.e. $\mu(S) = |S|$ for $S \subset \mathbb{N}$. The elements of ℓ_p^d are d-tuples of real numbers $a = (a_1, a_2, \dots, a_d) \in \mathbb{R}^d$. The $p-$norm of a vector $a = (a_1, a_2, \dots, a_d)$ equals
$$\|a\|_p = \left(\sum_{i=1}^{d} |a|^p\right)^{1/p}.$$

3. Space $L_p[a, b]$. Let $X = [a, b]$, and μ be the standard measure on $\mathbb{R}$. The elements of $L_p[a, b]$ are measurable functions $f : [a, b] \to \mathbb{R}$ with $\|f\|_p < \infty$. The p–norm of a function f equals
$$\|f\|_p = \left(\int_a^b |f(x)|^p dx\right)^{1/p}.$$

Lemma 1.6. *For every $1 \le p < q \le \infty$, we have $\ell_p \subset \ell_q$ and $L_q[0,1] \subset L_p[0,1]$. Both inclusions are proper.*

Proof. We consider the case when $q < \infty$. Let $a \in \ell_p$. Let $I = \{i : |a_i| \ge 1\}$. Note that I is a finite set, as otherwise we would have that $\|a\|_p^p \ge \sum_{i\in I} |a_i|^p = \infty$. For every $i \notin I$, we have $|a_i|^q < |a_i|^p$. Therefore,

$$\|a\|_q^q = \sum_{i\in I} |a_i|^q + \sum_{i\notin I} |a_i|^q \le \sum_{i\in I} |a_i|^q + \sum_{i\notin I} |a_i|^p \le \sum_{i\in I} |a_i|^q + \|a\|_p^p < \infty.$$

We conclude that $a \in \ell_q$.

Now let $f \in L_q[0,1]$. Let $I = \{x : |f(x)| \le 1\}$. Note that $|f|^p < |f|^q$ when $x \notin I$, and $\int_I |f(x)|^q\,dx \le \int_I 1\,dx \le 1$. Therefore,

$$\|f\|_p^p = \int_0^1 |f(x)|^p\,dx = \int_I |f(x)|^p\,dx + \int_{[0,1]\setminus I} |f(x)|^p\,dx \le 1 + \int_{[0,1]\setminus I} |f(x)|^q\,dx \le 1 + \|f\|_q^q < \infty.$$

We get that $f \in L_p[0,1]$. $\square$

Exercise 2. *Prove the statement of Lemma 1.6 for $q = \infty$.*

Exercise 3. *Let (X, μ) be a measure space with $\mu(X) < \infty$, and $1 \le p < q \le \infty$. Prove that $L_q(X,\mu) \subset L_p(X,\mu)$. Show that on the other hand $L_q(\mathbb{R}) \not\subset L_p(\mathbb{R})$.*

1.3 Dual Space

Consider a normed space $(V, \|\cdot\|)$ and the space V^* of continuos linear functionals $\phi : V \to \mathbb{R}$ on V. That is, ϕ is the set of linear maps on V such that $\sup_{u\neq 0} \frac{\phi(u)}{\|u\|} < \infty$. Define a norm $\|\cdot\|^*$ on V^* as follows

$$\|\varphi\|^* = \sup_{u\neq 0} \frac{|\phi(u)|}{\|u\|}.$$

$(V^*, \|\cdot\|^*)$ is a Banach space.

Let $p, q \in (1, \infty)$ s.t. $1/p + 1/q = 1$. Later in this course we will show that the dual of ℓ_p is ℓ_q and vice versa. Similarly the dual of $L_p(X,\mu)$ is $L_q(X,\mu)$ and vice versa. The duals of ℓ_1 and $L_1(X,\mu)$ are ℓ_∞ and $L_\infty(X,\mu)$. However, ℓ_1 is not the dual of ℓ_∞, and, in general, $L_1(X,\mu)$ is not the dual of $L_\infty(X,\mu)$. That said, ℓ_1^d and ℓ_∞^d are duals of each other.

We say that V is reflexive if $V = V^{**}$ (or more precisely V^{**} is isometrically isomorphic to V). As we pointed out above, ℓ_p and $L_p(X,\mu)$ are reflexive spaces for $p \in (1,\infty)$. However, ℓ_1, ℓ_∞, $L_1(\mathbb{R})$, and $L_\infty(\mathbb{R})$ are not. Importantly, all finite dimensional spaces are reflexive.

1.4 Unit Balls

We define the unit ball of a normed space $(v, \|\cdot\|)$ as follows: $B = \{v \in V : \|v\| \le 1\}$. Note that B is a closed convex set. Further, $0 \in B$ and B is centrally symmetric; that is, if $u \in B$ then $-u \in B$.

Let V be a finite dimensional space and S be a centrally symmetric closed convex body. Further, assume that some neighborhood of 0 lies in S. Define $\|\cdot\|$ as follows: $\|u\| = \min\{\alpha > 0 : u/\alpha \in S\}$ for $u \neq 0$ and $\|0\| = 0$. Then $(V, \|\cdot\|)$ is a normed space; further, S is the unit ball of V.

We see that in finite dimensions, there is a one-to-one correspondence between norms and their unit balls.

Exercise 4. *Consider Euclidean space V. We identify the dual space V^* with V in the standard way: for $u \in V$, $u(v) = \langle u, v\rangle$. Let $\|\cdot\|$ be an arbitrary norm on V and $\|\cdot\|^*$ be the dual norm. Prove that the unit balls B and B_* of $\|\cdot\|$ and $\|\cdot\|^*$, respectively, are polar sets of each other.*

2 Lyapunov's, Hölder's, and Interpolation Inequalities

In this section, we prove a few inequalities that we will need later.

Theorem 2.1 (Lyapunov's inequality)**.** *Let $1 \leq p < q = \infty$. For every random variable α with finite q-th moment, we have $\|\alpha\|_p \leq \|\alpha\|_q$.*

Proof. The statement is obvious for $q = \infty$ since $|\alpha| < \|\alpha\|_\infty$ almost surely. Let us assume that $q < \infty$. Let $f(x) = x^{q/p}$ for $x \geq 0$. Note that $f(x)$ is a convex function. Let $\beta = |\alpha|^p$ (β is a random variable). We have

$$\|\alpha\|_q^q = \mathbb{E}\left[|\alpha|^q\right] = \mathbb{E}\left[|\beta|^{q/p}\right] = \mathbb{E}\left[f(|\beta|)\right] \overset{\text{by Jensen's Inequality}}{\geq} f(\mathbb{E}\left[|\beta|\right]) = \left(\mathbb{E}\left[|\alpha^p|\right]\right)^{q/p}.$$

We conclude that $\|\alpha\|_q \geq \|\alpha\|_p$ as required. □

Now, we state Hölder's Inequality. The inequality essentially states that $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual norms when $1/p + 1/q = 1$.

Theorem 2.2 (Hölder's Inequality)**.** *Assume that $1/p + 1/q = 1$. Then for every $a, b \in \mathbb{R}^d$.*

$$\langle a, b\rangle \leq \|a\|_p \cdot \|b\|_q$$

Proof. Fix some $b \neq 0$. Consider function $f(a) = \langle a, b\rangle$ on the manifold $M = \{a : \|a\|_p^p = 1\}$. Since M is compact, f attains its maximum on M at some point a. Then $\operatorname{grad} f$ is orthogonal to the tangent space to M at a. Since M is the level set of function $g(a) = \|a\|_p^p = \sum |a_i|^p$, $b = \operatorname{grad} f$ is colinear with $\operatorname{grad} g = (p|a_1|^{p-1} \operatorname{sgn} a_1, \ldots, pa_d^{p-1} \operatorname{sgn} a_d)$. That is, for some $t > 0$, $|b_i| = t|a_i|^{p-1}$ for all i. Therefore, $|a_i||b_i| = t|a_i|^p$. Then

$$\sum_i a_i b_i \leq \sum_i |a_i||b_i| = t \sum_i |a_i|^p = t.$$

On the other hand,

$$\|b\|_q^q = \sum_i |b_i|^q = \sum_i t^q |a_i|^{(p-1)q} = t^q \sum_i |a_i|^p = t^q.$$

We conclude that $t = \|b\|_q$ and $\langle a, b\rangle \le t = \|a\|_p \|b\|_q$. We proved Hölder's inequality for vectors a with $\|a\|_p = 1$. The general case follows for the homogeneity of the inequality. $\square$

Exercise 5. *For a given $a \in \mathbb{R}^d$, define b as follows: $b_i = |a_i|^{p/q} \operatorname{sgn} a_i$. Show that $\langle a, b\rangle = \|a\|_p \|b\|_q$. Conclude that*

$$\|b\|_q = \|b\|_p^* \equiv \sup_{b \neq 0} \frac{\langle a, b\rangle}{\|a\|_p}.$$

Theorem 2.3 (Interpolation Inequality)**.** *Let $1 \le p < r < q \le \infty$. Define $\hat{p} = 1/p$, $\hat{q} = 1/q$, $\hat{r} = 1/r$,*

$$\alpha = \frac{\hat{r} - \hat{q}}{\hat{p} - \hat{q}} \qquad \text{and} \qquad \beta = \frac{\hat{p} - \hat{r}}{\hat{p} - \hat{q}}.$$

$$\|a\|_r \le \|a\|_p^\alpha \cdot \|a\|_q^\beta$$

for every $a \in \mathbb{R}^d$.

Proof. Note that $\alpha + \beta = 1$ and $\hat{r} = \alpha\hat{p} + \beta\hat{q}$ (that is, $\hat{r}$ is a convex combination of $\hat{p}$ and $\hat{q}$ with weights α and β). Let $p' = \frac{p}{\alpha r}$ and $q' = \frac{q}{\beta r}$. Then $1/p' + 1/q' = r \cdot (\alpha\hat{p}) + r \cdot (\beta\hat{q}) = r\hat{r} = 1$

$$\begin{aligned}
\|a\|_r^r = \sum_{i=1}^d |a_i|^r = \sum_{i=1}^d |a_i|^{\alpha r} \cdot |a_i|^{\beta r} &\overset{\text{Hölder}}{\le} \left(\sum_{i=1}^d (|a_i|^{\alpha r})^{p'}\right)^{1/p'} \cdot \left(\sum_{i=1}^d \left(|a_i|^{\beta r}\right)^{q'}\right)^{1/q'} \\
&= \left(\sum_{i=1}^d |a_i|^p\right)^{1/p'} \cdot \left(\sum_{i=1}^d |a_i|^q\right)^{1/q'} = \|a\|_p^{p/p'} \cdot \|a\|_q^{q/q'} = \|a\|_p^{\alpha r} \cdot \|a\|_q^{\beta r}
\end{aligned}$$

Therefore,

$$\|a\|_r \le \|a\|_p^\alpha \cdot \|a\|_q^\beta$$

$\square$

Corollary 2.4. *Let $1 \le p < r \le \infty$. For every $a \in \mathbb{R}^d$, we have*

$$\|a\|_r \le \|a\|_p \le d^{1/r - 1/p} \|a\|_r.$$

Proof. We apply the interpolation inequality with $q = \infty$. Then $\hat{q} = 0$ and thus $\alpha = p/r$, $\beta = 1 - p/r$. We have

$$\|a\|_r \le \|a\|_p^{p/r} \|a\|_\infty^{1-p/r} \le \|a\|_p^{p/r} \|a\|_p^{1-p/r} = \|a\|_p.$$

On the other hand, let ξ be a random coordinate of a chosen uniformly at random. Then,

$$\frac{\|a\|_p}{d^{1/p}} = \left(\sum_{i=1}^d \frac{|a_i|^p}{d}\right)^{1/p} \equiv \|\xi\|_p \overset{\text{Lyapunov's Ineq.}}{\ge} \|\xi\|_r = \left(\sum_{i=1}^d \frac{|a_i|^r}{d}\right)^{1/r} = \frac{\|a\|_r}{d^{1/r}}.$$

Therefore, $\|a\|_p \le d^{1/r - 1/p} \|a\|_r$.

$\square$

3 Metric Embeddings

Consider two metric spaces (X, d_X), (Y, d_Y) and a map $f : X \to Y$. We say that $f : X \to Y$ is a Lipschitz map if there is a number C such that

$$d_Y(f(x_1), f(x_2)) \leq C d_X(x_1, x_2) \text{ for all } x_1, x_2, \in X.$$

The Lipschitz constant $\|f\|_{Lip}$ of f is the minimum C such that this inequality holds.

We say that a bijective map $\varphi : X \to Y$ is an *isometry* if for every $x_1, x_2 \in X$, $d_Y(\varphi(x_1), \varphi(x_2)) = d_X(x_1, x_2)$. We say that an injective map $\varphi : X \to Y$ is an isometric embedding if φ is an isometry between X and $\varphi(X)$ (the image of X under φ).

The distortion of a map $f : X \to Y$ equals $\|f\|_{Lip} \cdot \|f^{-1}\|_{Lip}$ where f^{-1} is the inverse map from $f(X)$ to X.

Exercise 6. *Prove that f has distortion at most D if and only if there is a number $c > 0$ such that*

$$c \cdot d_X(x_1, x_2) \leq d_Y(x_1, x_2) \leq c\, D \cdot d_X(f(x_1), f(x_2)) \textit{ for every } x_1, x_2 \in X.$$

Exercise 7. *Prove the following statements.*

1. *An isometric embedding has distortion 1.*
2. *Let f be a Lipschitz map from X to Y and g be a Lipschitz map from Y to Z then $h = g \circ f$ is a Lipschitz map from X to Z and $\|h\|_{Lip} \leq \|f\|_{Lip} \cdot \|g\|_{Lip}$.*
3. *Let f be an embedding of X into Y and g be an embedding of Y into Z then the distortion of $h = g \circ f$ is at most the product of distortions of f and g.*

4 Embeddings into L_p spaces

Theorem 4.1. *Every finite metric subspace (X, d) embeds isometrically into ℓ_∞^n for $n = |X|$.*

Proof. Denote the elements of X by $x_1, x_2, \ldots, x_n$. Now define the embedding $\varphi : X \to \ell_\infty^n$ as follows

$$\varphi : x \mapsto (d(x_1, x), d(x_2, x), \ldots, d(x_n, x)).$$

We claim that φ is an isometric embedding. That is,

$$\|\varphi(x_i) - \varphi(x_j)\|_\infty = d(x_i, x_j).$$

First, we prove that $\|\varphi(x_i) - \varphi(x_j)\|_\infty \leq d(x_i, x_j)$. We need to show that all coordinates of the vector $\varphi(x_i) - \varphi(x_j)$ are bounded by $d(x_i, x_j)$ in the absolute value. Indeed, the k-th coordinate of $\varphi(x_i) - \varphi(x_j)$ equals $d(x_k, x_i) - d(x_k, x_j)$. From the triangle inequalities for x_i, x_j and x_k, it follows that $|d(x_k, x_i) - d(x_k, x_j)| \leq d(x_i, x_j)$. Now, we verify that

$\|\varphi(x_i) - \varphi(x_j)\|_\infty \geq d(x_i, x_j)$. Note that $\|\varphi(x_i) - \varphi(x_j)\|_\infty \geq |d(x_k, x_i) - d(x_k, x_j)|$ (the absolute value of the k-th coordinate) for every k. In particular, this inequality holds for $k = i$. That is,

$$\|\varphi(x_i) - \varphi(x_j)\|_\infty \geq |d(x_i, x_i) - d(x_i, x_j)| = d(x_i, x_j).$$

□

Theorem 4.2. *Let $p \in [1, \infty)$. Metric space ℓ_2^d (Euclidean d-dimensional space) embeds isometrically into $L_p(X, \mu)$ for some space X.*

Proof. Let $X = \ell_2^d$ and $\mu = \gamma$ be the Gaussian measure on $X = \ell_2^d$ (μ is the probability measure on X with density $e^{-\|x\|^2/2} \Big/ (2\pi)^{d/2}$). Recall that the elements of $L_p(\ell_2^d, \gamma)$ are functions on ℓ_2^d. Let $M = \left(\int_{\ell_2^d} |x_1|^p \, d\gamma(x)\right)^{1/p}$. We construct an embedding φ that maps every $v \in \ell_2^d$ to a function f_v defined as follows:

$$f_v(x) = \frac{\langle v, x \rangle}{M}.$$

We prove that the embedding is an isometry. Consider two vectors u and v. Let $w = u - v$, and $e = w/\|w\|_2$. We have,

$$\begin{aligned}
\|\varphi(u) - \varphi(v)\|_p^p &= \int_{\ell_2^d} \left| \frac{\langle u, x \rangle}{M} - \frac{\langle v, x \rangle}{M} \right|^p d\gamma(x) = \frac{1}{M^p} \int_{\ell_2^d} |\langle u - v, x \rangle|^p \, d\gamma(x) \\
&= \frac{1}{M^p} \int_{\ell_2^d} |\langle \|w\| \, e, x \rangle|^p d\gamma(x) = \frac{1}{M^p} \|w\|^p \int_{\ell_2^d} |\langle e, x \rangle|^p d\gamma(x)
\end{aligned}$$

Consider a coordinate frame in which the x_1-axis is parallel to the vector e (i.e. vector e has coordinates $(1, 0, \dots, 0)$). Then $|\langle e, x \rangle| = |x_1|$. We get

$$\|\varphi(u) - \varphi(v)\|_p = \frac{\|w\|_2}{M} \left(\int_{\ell_2^d} |x_1|^p d\mu(x) \right)^{1/p} = \|w\|_2 = \|u - v\|_2.$$

We proved that the map φ is an isometry. □

We showed that every finite subset S of ℓ_2^d embeds isometrically into space $L_p(X, \mu)$. Can we embed S into a "simpler" space ℓ_p^N (for some N)? We will see that all spaces $L_p(X, \mu)$ (of sufficiently large dimension) have essentially the same finite metric subspaces. Therefore, if a metric space embeds into some $L_p(X, \mu)$, then it also embeds into ℓ_p^N for some N.

Theorem 4.3. *Let S be a finite subset of $L_p(Z, \mu)$, $n = |S|$, and $N = \binom{n}{2} + 1$. Then S isometrically embeds into ℓ_p^N.*

Proof. Consider the linear space $\mathcal{M}$ of all symmetric $n \times n$ matrices with zeros on the diagonal. The dimension of $\mathcal{M}$ is $\binom{n}{2}$. Consider a map $f : \mathbb{R}^n \to \mathcal{M}$ defined as follows. The map f sends a vector $u \in \mathbb{R}^n$ to the matrix $A = (a_{ij})$ with entries $a_{ij} = |u_i - u_j|^p$. Clearly, $f(v) \in \mathcal{M}$ for every $v \in \mathbb{R}^n$. Let $B = f(\mathbb{R}^n) \equiv \{f(u) : u \in \mathbb{R}^n\}$ and $\mathcal{C} = \mathrm{conv}(B)$.

For every metric space (S, d) on a set $S = \{s_1, s_2, \dots, s_n\}$, we define a matrix F^S by $F^S_{ij} = d(s_i, s_j)^p$. The theorem follows from the following two lemmas.

Lemma 4.4. *Suppose that $S = \{s_1, \dots, s_n\} \subset L_p(Z, \mu)$ then $F^S \in \mathcal{C}$.*

Proof. Recall that each element s_i is a function from Z to $\mathbb{R}$. Let $\sigma(z) = (s_1(z), \dots, s_n(z))$ for every $z \in Z$. We have,

$$F^S_{ij} = d(s_i, s_j)^p = \int_Z |s_i(z) - s_j(z)|^p d\mu(z) = \int_Z f_{ij}(\sigma(z)) d\mu(z).$$

Therefore, $F^S = \int_Z f(\sigma(z)) d\mu(z)$. Since $f(\sigma(z)) \in B \subset \mathcal{C}$ for every $z \in Z$, we conclude that $F^S \in \mathcal{C}$. $\square$

Lemma 4.5. *Consider a finite metric space $S = \{s_1, \dots, s_n\}$. Suppose that $F^S \in \mathcal{C}$. Then S embeds into ℓ_p^N, where $N = \binom{n}{2} + 1$.*

Proof. By the Carathéodory theorem, every point in the cone $\mathcal{C}$ can be expressed as a sum of at most $\dim \mathcal{M} + 1 = N$ points in B. In particular, we can write

$$F^S = \sum_{k=1}^{N} b^{(k)} = \sum_{k=1}^{N} f(x^{(k)}),$$

for some $b^{(1)}, \dots, b^{(N)} \in B$ and some $x^{(k)} \in f^{-1}(b^{(k)})$ (x_i is a preimage of $b^{(k)}$). By the definition of F^S, we have

$$d(s_i, s_j)^p = F^S_{ij} = \sum_{k=1}^{N} f_{ij}(x^{(k)}) = \sum_{k=1}^{N} |x_i^{(k)} - x_j^{(k)}|^p. \tag{1}$$

We define the embedding φ of S to ℓ_p^N:

$$\varphi(s_i) = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(N)}).$$

Note that equation (1) says that $d(s_i, s_j)^p = \|\varphi(s_i) - \varphi(s_j)\|_p^p$, and therefore $d(s_i, s_j) = \|\varphi(s_i) - \varphi(s_j)\|_p$. We conclude that φ is an isometric embedding. $\square$

$\square$

Corollary 4.6. *Suppose that S is a subset of ℓ_2^d. Then S isometrically embeds into ℓ_p^N, where $N = \binom{|S|}{2} + 1$.*

Exercise 8. *In our proof, we used the Carathéodory theorem for arbitrary convex sets: every point in the convex hull of $S \subset \mathbb{R}^d$ is a convex combination of at most $d+1$ points from S. Show that if the convex hull $\mathrm{conv}(S)$ of S is a cone, then every point $\mathrm{conv}(S)$ is a linear combination, with positive coefficients, of at most d points in S. Conclude that in the statement of Theorem 4.3 we can replace $N = \binom{n}{2} + 1$ with $N = \binom{n}{2}$.*

Definition 4.7. *Let $c_p(X)$ be the least distortion*[1] *with which a finite metric space (X, d) embeds into ℓ_p.*

Theorem 4.8. *For every finite metric space X and every $p \in [1, \infty]$, we have $1 = c_\infty(X) \leq c_p(X) \leq c_2(X)$.*

Proof. The inequality $1 = c_\infty(X) \leq c_p(X)$ follows from Theorem 4.1. Let f be an embedding of X into $\ell_2(X)$ with distortion $c_2(X)$. By Corollary 4.6, there is an isometric embedding g of $f(X) \subset \ell_2$ into ℓ_p. Then map $g \circ f$ is an embedding of X into ℓ_p with distortion at most $c_2(X)$. We conclude that $c_p(X) \leq c_2(X)$. □

[1] A simple compactness argument shows that there is an embedding with the least possible distortion.