1 Notation

Given a metric space \((X, d)\) and \(S \subset X\), the distance from \(x \in X\) to \(S\) equals
\[
d(x, S) = \inf_{s \in S} d(x, s).
\]
The distance between two sets \(S_1, S_2 \subset X\) equals
\[
d(S_1, S_2) = \inf_{s_1 \in S_1, s_2 \in S_2} d(s_1, s_2).
\]

Exercise 1. Show that distances between sets do not necessarily satisfy the triangle inequality. That is, it is possible that
\[
d(S_1, S_2) + d(S_2, S_3) > d(S_1, S_3)
\]
for some sets \(S_1, S_2\) and \(S_3\).

Exercise 2. Prove that
\[
d(x, y) \geq d(S, x) - d(S, y)
\]
and thus \(d(x, y) \geq |d(S, x) - d(S, y)|\).

Proof. Fix \(\varepsilon > 0\). Let \(y' \in S\) be such that \(d(y', y) \leq d(S, y) + \varepsilon\) (if \(S\) is a finite set, there is \(y' \in S\) s.t. \(d(y, y') = d(S, y)\)). Then
\[
d(x, S) \leq d(x, y') \leq d(x, y) + d(y, y') \leq d(x, y) + d(S, y) + \varepsilon.
\]
We proved that \(d(x, S) \leq d(x, y) + d(S, y) + \varepsilon\) for every \(\varepsilon > 0\). Therefore,
\[
d(x, S) \leq d(x, y) + d(S, y).
\]

Definition 1.1. Let \((X, d)\) be a metric space, \(x_0 \in X\) and \(r > 0\). The (closed) ball of radius \(r\) around \(x_0\) is
\[
B_r(x_0) = \text{Ball}_r(x_0) = \{x : d(x, x_0) \leq r\}.
\]
2 Metric Embeddings of Normed Spaces

Consider two normed spaces \((U, \| \cdot \|_U)\) and \((V, \| \cdot \|_V)\). Let \(f\) be a linear map between \(U\) and \(V\). What is the Lipschitz norm of \(f\)? It is equal

\[
\sup_{x, y \in U, x \neq y} \frac{\|f(x) - f(y)\|_V}{\|x - y\|_U} \quad \text{by linearity of } f
\]

Definition 2.1. The operator norm of \(f\) is

\[
\|f\| \equiv \|f\|_{U \to V} = \sup_{z \in U, z \neq 0} \frac{\|f(z)\|_V}{\|z\|_U}.
\]

The above computation shows that the Lipschitz norm of a linear operator equals its operator norm.

Let \(U\) and \(V\) be two \(d\)-dimensional normed spaces. The Banach-Mazur distance between them is

\[
d_{BM}(U, V) = \min_{\varphi : U \to V} \|\varphi\|_{\varphi^{-1}};
\]

where the minimum is over non-degenerate linear maps \(\varphi : U \to V\).

Exercise 3. Consider two normed spaces \(U\) and \(V\). Let \(B_U\) and \(B_V\) be their unit balls. Prove that there exists a linear map \(\varphi\) such that \(B_V \subseteq \varphi(B_U) \subseteq \alpha B_V\) for some \(\alpha\) then \(d_{BM}(U, V) \leq \alpha\).

The Banach-Mazur distance is a distance in the following sense.

Claim 2.2. The Banach–Mazur distance satisfies the following properties.

- \(d_{BM}(U, U) = 1\)
- \(d_{BM}(U, V) \geq 1\)
- \(d_{BM}(U, V) \cdot d_{BM}(V, W) \geq d_{BM}(U, W)\)

Theorem 2.3. \(d_{BM}(\ell_p^d, \ell_2^d) = d^{1/p - 1/2}\)

Proof. First we observe that \(d_{BM}(\ell_p^d, \ell_2^d) \leq d^{1/p - 1/2}\). Indeed, let us consider the identity map between \(\ell_p^d\) and \(\ell_2^d\) and upper bound its distortion. If \(p \in [1, 2]\), we have \(\|a\|_2 \leq \|a\|_p \leq d^{1/p - 1/2} \|a\|_2\). Thus the identity map from \((\mathbb{R}^d, \| \cdot \|_p)\) to \((\mathbb{R}^d, \| \cdot \|_2)\) has distortion at most \(d^{1/p - 1/2}\). Similarly, if \(p \in [2, \infty]\), we have \(\|a\|_p \leq \|a\|_2 \leq d^{1/2 - 1/p} \|a\|_p\). Thus the identity map from \((\mathbb{R}^d, \| \cdot \|_p)\) to \((\mathbb{R}^d, \| \cdot \|_2)\) has distortion at most \(d^{1/2 - 1/p}\).
Discussion  

Now we need to prove that every linear map \( \varphi : \ell_p^d \to \ell_2^d \) has distortion at least \( d^{1/2 - 1/p} \). Consider the hypercube \( C = \{-1, 1\}^d \subseteq \ell_p^d \). We will prove that even restricted to \( C \), \( \varphi \) has distortion at least \( d^{1/2 - 1/p} \). To gain some intuition, assume that \( p = 1 \) and \( \varphi = id \). How does \( \varphi \) distort the distances between the vertices of the hypercube?

- \( \varphi = id \) preserves the lengths of the edges of \( C \): if \( u, v \in C \) differ in exactly one coordinate then \( \| \varphi(u) - \varphi(v) \|_2 = \| u - v \|_1 = 2 \). Therefore, \( \| \varphi \| \geq \frac{\| \varphi(u) - \varphi(v) \|_2}{\| u - v \|_1} \geq 1 \).

- \( \varphi \) contracts the diagonals of \( C \) by a factor of \( \sqrt{d} \): for \( u \in C \) and \( u' = -u \), we have \( \| u - u' \|_1 = 2d \) and \( \| \varphi(u) - \varphi(u') \|_2 = 2\sqrt{d} \). Therefore, \( \| \varphi^{-1} \| \geq \frac{\| u - u' \|_1}{\| \varphi(u) - \varphi(u') \|_2} \geq \sqrt{d} \).

We see that the distortion of \( \varphi \) is at least \( \| \varphi \| \cdot \| \varphi^{-1} \| \geq 1 \cdot \sqrt{d} = \sqrt{d} \).

Now consider an arbitrary non-degenerate linear map \( \varphi \) and arbitrary \( p \in [1, \infty] \). The example above suggests that we should examine how \( \varphi \) distorts edges and diagonals of \( C \). However, it is not sufficient to look at a single edge or single diagonal. Instead, we compute how \( \varphi \) distorts edges and diagonals on average. First, we look at the edges. Choose a random coordinate \( i \in \{1, \ldots, d\} \) uniformly at random. Then independently choose a random vertex \( u \) of \( C \) uniformly at random. Let \( v \in C \) be the vertex that differs from \( u \) in coordinate \( i \). Then \( u - v = 2e_i \) or \( u - v = -2e_i \). We have \( \| \varphi(u) - \varphi(v) \|_2 \leq \| \varphi \| \cdot \| u - v \|_p \) (always). Therefore,

\[
\| \varphi \|^2 \geq \mathbb{E} \left[ \frac{\| \varphi(u) - \varphi(v) \|_2^2}{\| u - v \|_p^2} \right] = \mathbb{E} \left[ \frac{\| 2e_i \|_2^2}{\| 2e_i \|_p^2} \right] = \mathbb{E} \left[ \frac{\| \varphi(e_i) \|_2^2}{\| \varphi(e_i) \|_p^2} \right] = \frac{1}{d} \sum_{j=1}^{d} \| \varphi(e_j) \|_2^2  
\]

(1)

Similarly, \( \| u - v \|_p \leq \| \varphi^{-1} \| \cdot \| \varphi(u) - \varphi(v) \|_2 \) and thus \( \mathbb{E} \left[ \| u - v \|_p^2 \right] \leq \| \varphi^{-1} \|^2 \cdot \mathbb{E} \left[ \| \varphi(u) - \varphi(v) \|_2^2 \right] \). We have,

\[
\| \varphi^{-1} \|^2 \geq \mathbb{E} \left[ \| u - v \|_p^2 \right] / \mathbb{E} \left[ \| \varphi(u) - \varphi(v) \|_2^2 \right] = \mathbb{E} \left[ \frac{\| 2e_i \|_2^2}{\| 2e_i \|_p^2} \right] = \mathbb{E} \left[ \frac{\| \varphi(e_i) \|_2^2}{\| \varphi(e_i) \|_p^2} \right] = \frac{1}{d} \sum_{j=1}^{d} \| \varphi(e_j) \|_2^2  
\]

(2)

Now let \( u \) be a random vertex of \( C \) and \( u' = -u \). Note that all coordinates \( u_1, \ldots, u_d \) of \( u \) are i.i.d. Bernoulli \( \{\pm 1\} \) random variables. Also, \( u = \sum_{j=1}^{d} u_j e_j \) and therefore \( \varphi(u) = \sum_{j=1}^{d} u_j \varphi(e_j) \). We write,

\[
\mathbb{E} \left[ \| \varphi(u) \|_2^2 \right] = \mathbb{E} \left[ \| \sum_{j=1}^{d} u_j \varphi(e_j) \|_2^2 \right] = \mathbb{E} \left[ \sum_{1 \leq j, j' \leq d} \langle u_j \varphi(e_j), u_{j'} \varphi(e_{j'}) \rangle \right] \]

\[
= \sum_{1 \leq j, j' \leq d} \mathbb{E} [u_j u_{j'}] \cdot \langle \varphi(e_j), \varphi(e_{j'}) \rangle.  
\]

Since all random variable \( u_1, \ldots, u_d \) are independent, \( \mathbb{E} [u_j] = 0 \), and \( u_j^2 = 1 \) (always), we have

\[
\mathbb{E} [u_j u_{j'}] = \begin{cases} 
1, & \text{if } j = j' \\
0, & \text{otherwise}
\end{cases} 
\]
We conclude that
\[ E \left[ \| \varphi(u) \|_2^2 \right] = \sum_{j=1}^{d} \| \varphi(e_j) \|_2^2. \]

As above, we have
\[ \| \varphi \|^2 \geq E \left[ \| \varphi(u) - \varphi(u') \|_p^2 \right] = \sum_{j=1}^{d} \| \varphi(e_j) \|_2^2. \]

Similarly,
\[ \| \varphi^{-1} \|^2 \geq \frac{d^{2/p}}{\sum_{j=1}^{d} \| \varphi(e_j) \|_2^2}. \]  (4)

If \( p \in [1, 2] \), multiplying inequalities (1) and (4), we get \( \| \varphi \|^2 \| \varphi^{-1} \|^2 \geq \frac{d^{2/p}}{d^{2/p}} = \frac{d^{2/p}}{d^{2/p}} \).

Thus, the distortion of \( \varphi \) is at least \( d^{1/p-1/2} \), as required. If \( p \in [2, \infty] \), multiplying inequalities (2) and (3), we get \( \| \varphi \|^2 \| \varphi^{-1} \|^2 \geq \frac{d^{1/p}}{d^{2/p}} = \frac{d^{1/p}}{d^{2/p}} \). Thus, the distortion of \( \varphi \) is at least \( d^{1/2-1/p} \), as required.

\[ \square \]

Using Claim 2.2, we get the following corollary from Theorem 2.3.

**Corollary 2.4.** We have,
- \( d_{BM}(\ell_p^d, \ell_q^d) = d^{1/p-1/q} \) if \( p, q \in [1, 2] \)
- \( d_{BM}(\ell_p^d, \ell_q^d) = d^{1/p-1/q} \) if \( p, q \in [2, \infty] \)

**Fact 2.5.** \( d_{BM}(\ell_p^d, \ell_q^d) = \Theta(\max(1/p-1/2, 1/2-1/q)) \) if \( 1 \leq p \leq 2 \leq q \leq \infty \).

One can ask if there is a non-linear bijection between \( \ell_p^d \) and \( \ell_q^d \) with a smaller distortion. The answer is negative. We omit the details here. However, one way to prove this is as follows. Consider a non-linear map \( \varphi : \ell_p^d \to \ell_p^d \) with distortion \( D \). Note that \( \varphi \) is Lipschitz (as otherwise, it would have an infinite distortion). By Rademacher’s theorem, every Lipschitz map from \( \mathbb{R}^d \) to \( \mathbb{R}^d \) is differentiable almost everywhere. Let \( x \) be any point where \( \varphi \) is differentiable. Consider the differential of \( d_x \varphi \) of \( \varphi \) at \( x \). It is not hard to verify that linear map \( \psi = d_x \varphi : \ell_p^d \to \ell_p^d \) has distortion at most \( D \).

**Fact 2.6** (John Ellipsoid or Löwner–John Ellipsoid). For every convex centrally-symmetric set \( S \subset \mathbb{R}^d \) that contains a neighborhood of the origin, there exists an ellipsoid \( \mathcal{E} \) centered at the origin such that \( \mathcal{E} \subseteq S \subseteq \sqrt{d} \cdot \mathcal{E} \). Specifically, one may choose (a) \( \mathcal{E} \) to be the maximum volume ellipsoid inside \( S \) or (b) \( \sqrt{d} \cdot \mathcal{E} \) to be the minimum volume ellipsoid containing \( S \).

Equivalently, let \( \| \cdot \| \) be an arbitrary norm in \( \mathbb{R}^d \). Then \( d_{BM}(\mathbb{R}^d, \| \cdot \|), \ell_2^d) \leq \sqrt{d} \).
3 Bourgain’s Theorem

Definition 3.1. Let $X$ be a finite metric space and $p \geq 1$. Suppose that $Z \neq \emptyset$ is a random subset of $X$ (chosen according to some probability distribution). For every $u \in X$, define random variable $\xi_u = d(u, Z) = \min_{z \in Z} d(u, z)$. Consider the map $f$ from $X$ to the space of random variables $L_p(\Omega, \mu)$ that sends $u$ to $\xi_u$ (where $\Omega$ is the probability space and $\mu$ is the probability measure on $\Omega$). We say that $f$ is a Fréchet embedding.

Lemma 3.2. Every Fréchet embedding $f$ is non-expanding. That is, $\|f\|_{Lip} \leq 1$.

Proof. Consider a Fréchet embedding that sends $u$ to $\xi_u = d(u, Z)$. For every $u, v \in X$, we have

$$\|\xi_u - \xi_v\|_p = (\mathbb{E} [d(u, Z) - d(v, Z)]^p)^{1/p} \leq (\mathbb{E} [d(u, v)^p])^{1/p} = d(u, v).$$

Remark 3.3. If $X$ is infinite, then the random variable $\xi_u = d(u, Z)$ does not necessarily belong to $L_p(\Omega, \mu)$ (its $p$-norm might be infinite). However, we can define $\bar{\xi}_u$ as $\bar{\xi}_u = d(u, Z) - d(x_0, Z)$, where $x_0$ is some point in $X$. Then the proof of Lemma 3.2 shows that $\|\bar{\xi}_u\|_p \leq d(u, x_0) < \infty$ and the map $f: u \mapsto \bar{\xi}_u$ is non-expanding.

Theorem 3.4 (Bourgain’s Theorem). Every metric space $X$ on $n$ points embeds into $L_p(X, \mu)$ with distortion $O(\log n)$ (for every $p \geq 1$). That is, $c_p(X) = O(\log n)$.

Proof. Let $l = \lceil \log_2 n \rceil + 1$. Construct a random set $Z$ as follows.

- Choose $s$ uniformly at random from $\{1, \ldots, l\}$.
- Initially, let $Z = \emptyset$.
- Add every point of $X$ to $Z$ with probability $1/2^s$, independently.

Now let $f$ be the Fréchet embedding that maps $u \in X$ to random variable $\xi_u = d(Z, u)$. By Lemma 3.2, $f$ is non-expanding. We are going to prove that for every $u$ and $v$,

$$\|f(u) - f(v)\|_p \geq \frac{c}{l} \cdot d(u, v),$$

for some absolute constant $c$. Note that it is sufficient to prove this statement for $p = 1$, since by Lyapunov’s inequality $\|f(u) - f(v)\|_p \geq \|f(u) - f(v)\|_1$.

Consider two points $u$ and $v$. Let $\Delta = d(u, v)/2$. Define interval $I_Z$ as follows: $I_Z = [d(u, Z), d(v, Z)]$ if $d(u, Z) \leq d(v, Z)$, and $I_Z = [d(v, Z), d(u, Z)]$ if $d(v, Z) < d(u, Z)$. That is, $I_Z$ is the interval between $d(u, Z)$ and $d(v, Z)$. Denote the length of $I_Z$ by $|I_Z|$. Let $1_{I_Z}$ be the indicator function of $I_Z$. Write,

$$|d(u, Z) - d(v, Z)| = |I_Z| = \int_{I_Z} 1 \, dt = \int_0^\infty 1_{I_Z}(t) \, dt.$$
Then,
\[
\|f(u) - f(v)\|_1 = \mathbb{E}[\|d(u, Z) - d(v, Z)\|] = \mathbb{E}\left[\int_0^\infty 1_{I_Z}(t) dt\right]
\]
(by Fubini's theorem) \[
= \int_0^\infty \mathbb{E}[1_{I_Z}(t)] dt = \int_0^\infty \Pr(t \in I_Z) dt \geq \int_0^\Delta \Pr(t \in I_Z) dt.
\]

We now prove that \(\Pr(t \in I_Z) \geq \frac{\Omega(1)}{t}\) if \(t \in (0, \Delta)\). That will imply that \(\|f(u) - f(v)\|_1 \geq \frac{\Omega(1)}{t} \cdot \Delta = \frac{\Omega(1)}{t} \cdot d(u, v)\).

Fix \(t \in (0, \Delta)\). Consider balls \(B_t(u)\) and \(B_t(v)\). They are disjoint since \(2t < 2\Delta = d(u, v)\). Assume without loss of generality that \(|B_t(u)| \leq |B_t(v)|\). Denote \(m = |B_t(u)|\). Let \(s_0 = \lfloor \log_2 m \rfloor + 1\). Then \(m < 2^{s_0} \leq 2m\). Let \(E_u\) be the event that \(d(u, Z) > t\), and \(E_v\) be the event that \(d(v, Z) \leq t\). We have,
\[
\Pr(t \in I_Z) = \Pr(d(u, Z) \leq t \leq d(v, Z) \text{ or } d(v, Z) \leq t \leq d(u, Z)) \geq \Pr(d(v, Z) \leq t < d(u, Z)) = \Pr(E_u \text{ and } E_v).
\]

Event \(E_v\) occurs if and only if there is a point in \(Z\) at distance at most \(t\) from \(v\); that is, when \(B_t(v) \cap Z \neq \emptyset\). Event \(E_u\) occurs if and only if \(B_t(u) \cap Z = \emptyset\).

Consider the event \(s = s_0\). It happens with probability \(1/l\). Conditioned on this event, events \(E_u\) and \(E_v\) are independent (since \(B_t(u)\) and \(B_t(v)\) are disjoint) and
\[
\Pr(E_u|s = s_0) = \prod_{w \in B_t(u)} \Pr(w \notin Z|s = s_0) = \prod_{w \in B_t(u)} \left(1 - \frac{1}{2^{s_0}}\right) = \left(1 - \frac{1}{2^{s_0}}\right)^m \geq \frac{1}{e}.
\]
\[
\Pr(E_v|s = s_0) = 1 - \prod_{w \in B_t(v)} \Pr(w \notin Z|s = s_0) = 1 - \prod_{w \in B_t(v)} \left(1 - \frac{1}{2^{s_0}}\right) \geq 1 - \left(1 - \frac{1}{2^{s_0}}\right)^m \\
\geq 1 - \frac{1}{e^{1/2}}.
\]

We get
\[
\Pr(t \in I_Z) \geq \Pr(E_u \text{ and } E_v) \geq \Pr(s = s_0) \Pr(E_u \text{ and } E_v|s = s_0) \\
\geq \frac{1}{l} \Pr(E_u|s = s_0) \Pr(E_v|s = s_0) \geq \Omega\left(\frac{1}{l}\right).
\]

\[\square\]

**Exercise 4.** The set \(Z\) might be equal to \(\emptyset\) in our proof, then random variables \(\xi_u = d(u, Z)\) are not well defined. Show how to fix this problem.

**Proof.** There are many ways to fix this problem. For instance, we can add an extra point \(x_\infty\) to the metric space \(X\), and define \(d(u, x_\infty) = 2 \text{diam}(X)\), where \(\text{diam}(X) = \max_{u,v \in X} d(u, v)\). Then construct the set \(Z\) as before, except that always add \(x_\infty\) to \(Z\). Thus we ensure that \(Z \neq \emptyset\). In other words, we can define \(\xi_u\) as before if \(Z \neq \emptyset\), and \(\xi_u = 2 \text{diam}(X)\) if \(Z = \emptyset\). The rest of the proof goes through without any other changes. \[\square\]
The proof of Bourgain’s theorem provides an efficient randomized procedure for generating set $Z$. As presented here, this procedure gives an embedding only in $L_p(\Omega, \mu)$ and not in $\ell^N_p$. We already know that if a set of $n$ points embeds in $L_p(\Omega, \mu)$ with distortion $D$ then it embeds in $\ell_p^\binom{n}{2}$ with distortion $D$. However, in fact, we need only $N = O((\log n)^2)$ dimensions: for every value of $s \in \{1, \ldots, l\}$ we make $\Theta(\log n)$ samples of the set $Z$. Then the total number of samples equals $\Theta((\log n)^2)$. Using the Chernoff bound, it is easy to show that the distortion of the obtained embedding is $O(\log n)$ w.h.p.

**Fact 3.5** (Matoušek). Let $D_{n,p}$ be the smallest number $D$ such that every metric space on $n$ points embeds in $\ell_p$ with distortion at most $D_{n,p}$. Then

$$D_{n,p} = \Theta \left( \frac{\log n}{p} \right).$$