3.7 Lecture 19 (14/Nov): cont. Bourgain embedding

3.7.1 cont. Bourgain embedding: L_1

We use this notation for open balls:

$$B_r(x) = \{z : d(x,z) < r\}$$

and for closed balls, $\bar{B}_r(x) = \{z : d(x,z) \le r\}$.

Recall that we are now analyzing the embedding for any single pair of points x, y.

Let $\rho_0 = 0$ and, for $t > 0$ define

$$\rho_t = \sup\{r : |B_r(x)| < 2^t \text{ or } |B_r(y)| < 2^t\} \tag{3.17}$$

up to $\hat{t} = \max\{t : \text{RHS} < d(x,y)/2\}$.

It is possible to have $\hat{t} = 0$ (for instance if no other points are near x and y).

Observe that for the *closed* balls $\bar{B}$ we have that for all $t \le \hat{t}$, $|\bar{B}_{\rho_t}(x)| \ge 2^t$ and $|\bar{B}_{\rho_t}(y)| \ge 2^t$. This means in particular that (due to the radius cap at $d(x,y)/2$, which means that y is excluded from these balls around x and vice versa), $\hat{t} < s$.

Set $\rho_{\hat{t}+1} = d(x,y)/2$, which means that it still holds for $t = \hat{t}+1$ that $|B_{\rho_t}(x)| < 2^t$ or $|B_{\rho_t}(y)| < 2^t$, although (in contrast to $t \le \hat{t}$), $\rho_{\hat{t}+1}$ is not the largest radius for which this holds.

Note $\hat{t}+1 \ge 1$. Also, $\rho_{\hat{t}+1} > \rho_{\hat{t}}$ (because the latter was defined to be less than $d(x,y)/2$). But for $t \le \hat{t}$ it is possible to have $\rho_t = \rho_{t-1}$.

$\hat{t}+1$ will be the number of scales used in the analysis of the lower bound for the pair x, y. I.e., we use the sets T_{tj} for $0 \le t \le \hat{t}+1$. Any contribution from higher-t (smaller expected cardinality) sets is "bonus."

Consider any $1 \le t \le \hat{t}+1$.

Lemma 61 *With positive probability (specifically at least $(1 - 1/\sqrt{e})/4$), $|\tau_{t1}(x) - \tau_{t1}(y)| > \rho_t - \rho_{t-1}$.*

Proof: Suppose wlog that $|B_{\rho_t}(x)| < 2^t$. By Eqn (3.17) (with $t-1$), $|\bar{B}_{\rho_{t-1}}(y)| \ge 2^{t-1}$ (and the same for x but we don't need that). If

$$T_{t1} \cap B_{\rho_t}(x) = \emptyset \tag{3.18}$$

and

$$T_{t1} \cap \bar{B}_{\rho_{t-1}}(y) \neq \emptyset \tag{3.19}$$

then

$$\|\tau_{t1}(x) - \tau_{t1}(y)\| > \rho_t - \rho_{t-1}.$$

We wish to show that this conjunction happens with constant probability. (See Fig. 3.9.)

The two events (3.18), (3.19) are independent because T_{t1} is generated by independent sampling, and because, due to the radius cap at $d(x,y)/2$ (and because $\rho_{\hat{t}} < \rho_{\hat{t}+1}$), $B_{\rho_t}(x) \cap \bar{B}_{\rho_{t-1}}(y) = \emptyset$.

First, the x-ball event (3.18):

$$\begin{aligned}
\Pr(T_{t1} \cap B_{\rho_t}(x) = \emptyset) &= (1 - 2^{-t})^{|B_{\rho_t}(x)|} \\
&\ge (1 - 2^{-t})^{2^t} \\
&= (1 - 2^{-t})^{2^t} \\
&\ge 1/4 \quad \text{for } t \ge 1
\end{aligned}$$

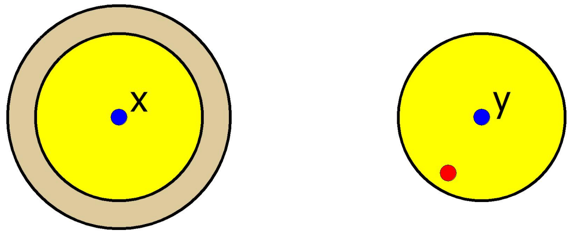

Figure 3.9: Balls $B_{\rho_{t-1}}(x), B_{\rho_t}(x), B_{\rho_{t-1}}(y)$ depicted. Events 3.18 and 3.19 have occurred, because no point has been selected for T_{t1} in the larger-radius (ρ_t) region around x, while some point (marked in red) has been selected for T_{t1} in the smaller-radius (ρ_{t-1}) region around y.

(For large t this is actually about $1/e$.)

Second, the y-ball event (3.19):

$$\begin{aligned}\Pr(T_{t1} \cap \bar{B}_{\rho_{t-1}}(y) \neq \emptyset) &= 1 - (1-2^{-t})^{|B_{\rho_{t-1}}(y)|} \\ &\geq 1 - (1-2^{-t})^{2^{t-1}}\end{aligned}$$

and recalling $1 + x \leq e^x$ for all real x,

$$\geq 1 - e^{-1/2}$$

Consequently, $|\tau_{t1}(x) - \tau_{t1}(y)| > \rho_t - \rho_{t-1}$ with probability at least $(1 - 1/\sqrt{e})/4$. □

Now, let $G_{x,y,t}$ be the "good" event that at least $(1 - 1/\sqrt{e})/8$ of the coordinates at level t, namely $\{\tau_{tj}\}_{j=1}^{s'}$, have

$$|\tau_{tj}(x) - \tau_{tj}(y)| > \rho_t - \rho_{t-1}.$$

If the good event occurs for all t, then for all x, y,

$$\|\tau(x) - \tau(y)\|_1 \geq \frac{1}{s} \frac{(1 - 1/\sqrt{e})}{8} \frac{d(x,y)}{2}.$$

Here the first factor is from the normalization, the second from the definition of good events, and the third from the cap on the ρ_t's.

We can upper bound the probability that a good event $G_{x,y,t}$ fails to happen using Chernoff:

$$\Pr(\neg G_{x,y,t}) \leq e^{-\Omega(s')}.$$

Now taking a union bound over all x, y, t,

$$\Pr(\cup_{x,y,t} \neg G_t) \leq e^{-\Omega(s')} n^2 \lg n < 1/2$$

for a suitable $s' \in \Theta(\log n)$.

To be specific we can use the following version of the Chernoff bound (see problem set 4):

Lemma 62 *Let $F_1, \ldots, F_{s'}$ be independent Bernoulli rvs, each with expectation $\geq \mu$. $\Pr(\sum F_i < (1 - \varepsilon)\mu s') \leq e^{-\varepsilon^2 \mu s'/2}$.*

which permits us (plugging in $\varepsilon = 1/2$) to take $s' = \frac{32\sqrt{e}}{\sqrt{e}-1} \log(n^2 \lg n)$. □

Exercise: Form a Fréchet embedding $X \to \mathbb{R}^n$ by using as T_i's all singleton sets. Argue that this is an isometry of X into $(\mathbb{R}^n, L_\infty)$. Consequently L_∞ is *universal* for finite metrics. (This, I believe, was Fréchet's original result [38].)

3.7.2 Embedding into any L_p, $p \geq 1$

As a matter of fact the above embedding method has distortion just as good into L_p, for any $p \geq 1$. Start by expanding:

$$\|\tau(x) - \tau(y)\|_p = \left(\frac{1}{ss'} \sum_{ij} (\tau_{ij}(x) - \tau_{ij}(y))^p \right)^{1/p} \tag{3.20}$$

We begin with the upper bound, which is unexciting:

$$\begin{aligned}(3.20) &\leq \left(\frac{1}{ss'} \sum_{ij} (d(x,y))^p \right)^{1/p} \\ &= d(x,y).\end{aligned}$$

For the lower bound, we use the power-means inequality. Note that (3.20) is a p'th mean of the quantities $(\tau_{ij}(x) - \tau_{ij}(y))$, ranging over i, j. So from Lemma 15,

$$(3.20) \geq \frac{1}{ss'} \sum_{ij} \left| \tau_{ij}(x) - \tau_{ij}(y) \right| = \|\tau(x) - \tau(y)\|_1$$

so for any τ and any $p > 1$, the L_p distortion of τ is no more than its L_1 distortion. This demonstrates the generalization of Theorem (60) with $L_p^{O(\log^2 n)}$ $(p \geq 1)$ replacing $L_1^{O(\log^2 n)}$.

3.7.3 Aside: Hölder's inequality

Although we already proved the power means inequality directly in Lemma 15, it is worth seeing how it fits into a framework of inequalities. The power means inequality is a comparison between two integrals over a measure space that is also a probability space (i.e., the total measure of the space is 1). Power means follows immediately from an important inequality that holds for *any* measure space (and indeed generalizes the Cauchy-Schwarz inequality), Hölder's inequality:

Lemma 63 (Hölder) *For norms with respect to any fixed measure space, and for* $1/p + 1/q = 1$ *(p and q are "conjugate exponents"),* $\|f\|_p \cdot \|g\|_q \geq \|fg\|_1$.

To see the power means inequality, note that over a probability space, $\|f\|_p$ is simply a p'th mean. Now plug in the function $g = 1$ and Hölder gives you power means.