CS 683: Advanced Algorithms	**Bourgain's Theorem**
22 February 2001	**Scribes: Gavin Hurley and Davina Kunvipusilkul**

1 Chernoff Bound

To prove Bourgain's Theorem, we will use the following probabilistic fact.

Lemma 1.1 (Chernoff Bound) *Let $X_1, X_2, \ldots, X_r$ be independent 0/1-valued random variables, $p_i = \mathbb{P}[X_i = 1] = \mathbb{E}X_i$, $\mu = \sum_{i=1}^r p_i = \mathbb{E}[\sum_{i=1}^r X_i]$. Then for any $\delta > 0$, $\mathbb{P}[\sum_{i=1}^r X_i < (1-\delta)\mu] < e^{-\frac{1}{2}\delta^2\mu}$.*

2 Bourgain's Theorem

Theorem 2.1 (Bourgain 1985) *Let (V, d) be a metric space on n points. Then (V, d) can be embedded into $\Re^l$ with the L_1 metric with distortion $O(\log n)$ (with a randomized polynomial-time algorithm).*

3 Proof of Bourgain's Theorem

We need to pick coordinates to embed $x \in V$.

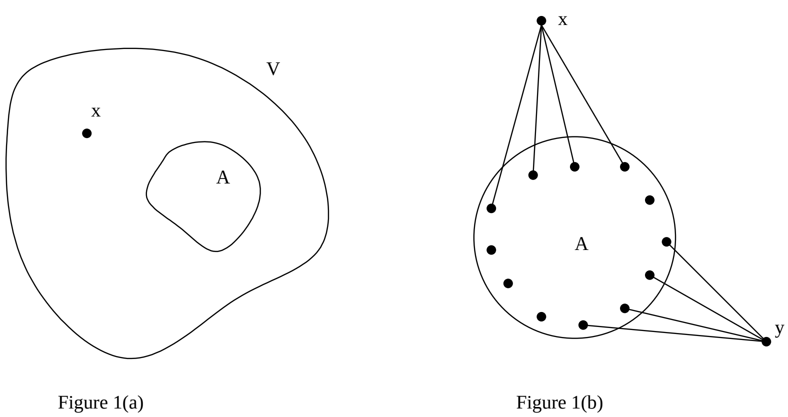

Figure 1:

Let $d(x, A) = \min_{z \in A} d(x, z)$ for $x \in V$ and $A \subset V$.
We will pick a collection of sets $A_1, \ldots, A_l \subseteq V$ and map $x \hookrightarrow [d(x, A_1), d(x, A_2), \ldots, d(x, A_l)]$.

Intuitively, close points get similar coordinates. However, do far points get very different coordinates? In other words, can we have different points with the same distance such as in Figure 1(b)?

Figure 2 illustrates what we hope for. That is, the coordinates of y are $\leq r$ and the coordinates of x are $> r'$. Thus, A contributes $\geq r' - r > 0$ to distance from x to y in L_1.

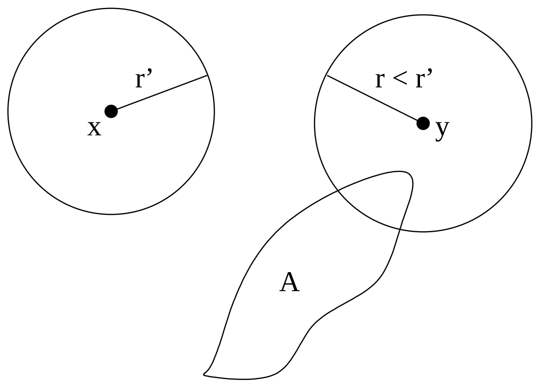

Figure 2:

Problem: We cannot pick a single radius.

Suppose a ball around y has b points and we pick a random set of a points.

$$\mathbb{P}[A \text{ misses ball}] = (\frac{n-b}{n})^a = (1 - \frac{b}{n})^a$$

If $a = \frac{n}{b}$, this is

$$\frac{1}{4} \leq (1 - \frac{b}{n})^{\frac{n}{b}} \leq \frac{1}{e} \text{ (if } b \leq \frac{1}{2}n)$$

Question: How do we pick the ball sizes?
For $i = 0, 1, 2, \ldots, p-1$

Choose q random multi-sets of size 2^i: $A_1^1, A_2^i, \ldots, A_q^i$.

We have $O(\log^2 n)$ sets total and $O(\log^2 n)$ dimensions.

The map will be $x \hookrightarrow < d(x, A_j^i) : i, j >= f(x)$.

First, (easy) upper bound

$$\begin{aligned}
\| f(x) - f(y) \|_1 &= \sum_{i,j} |d(x, A_j^i) - d(y, A_j^i)| \\
&\leq \sum_{i,j} d(x, y) \\
&= q \cdot p \cdot d(x, y) \\
&= O(\log^2 n) \, d(x, y)
\end{aligned}$$

We now need to show that $\| f(x) - f(y) \|_1 \geq \Omega(\log n) \cdot d(x,y)$

Note: Different size classes contribute different amount.

Definition 3.1

$$\begin{aligned} B(x,r) &= \{z : d(x,z) \leq r\} \\ B^o(x,r) &= \{z : d(x,z) < r\} \end{aligned}$$

We now focus on a pair x, y, and lower bound the distance in L_1.

$$\rho_t = \inf\{r : |B(x,r)| \geq 2^t, |B(y,r)| \geq 2^t\}$$

(We set $\rho_0 = 0$, and stop at $\rho_m =$ smallest ρ_t that is $\geq \frac{1}{4}d(x,y)$. We truncate ρ_m to be equal to $\frac{1}{4}d(x,y)$.)

Let us make two basic observations:

1. $B(x,\rho_t) \cap B(y,\rho_{t'}) = \emptyset \;\; \forall t,t' \leq m$ as $\rho_t, \rho_{t'} \leq \frac{1}{4}d(x,y)$.

2. For $0 \leq t \leq m$, at least one of $|B^o(x,\rho_t)| < 2^t$ or $|B^o(y,\rho_t)| < 2^t$ is true. To see this, note from the definition of ρ_t that for $r < \rho_t$ at least one of $|B(x,r)| \geq 2^t$ or $|B(x,r)| \geq 2^t$ was not satisfied.

Pick a size class 2^i, where i satisfies $2^i = \frac{n}{2^t}$; we are only interested in i such that ρ_t is defined. What then is the contribution to the distance between two points x and y associated with the co-ordinates for $A_1^i, A_2^i, \ldots, A_q^i$? We will show that it is

$$\Omega(\log n)(\rho_t - \rho_{t-1}).$$

Note that when we show this we will have completed the proof as then

$$\begin{aligned} ||f(x) - f(y)||_1 &\geq \sum_{t \text{ such that } \rho_t \text{ defined}} \Omega(\log n)(\rho_t - \rho_{t-1}) \\ &= \Omega(\log n)(\rho_m - \rho_0) \\ &= \Omega(\log n)\left[\frac{1}{4}d(x,y)\right] \end{aligned}$$

as required.

For our two points x and y, consider the sets $B^o(x,\rho_t)$ and $B^o(y,\rho_t)$. One of these will have $< 2^t$ points in it; without loss of generality we assume that it is $B^o(x,\rho_t)$. Consider also $B(y,\rho_{t-1})$, a set which has $\geq 2^{t-1}$ points. If a set A_j^i hits $B(y,\rho_{t-1})$ and misses $B^o(x,\rho_t)$ then it will contribute "a lot" to $||f(x) - f(y)||_1$, in a sense to be made more formal later.

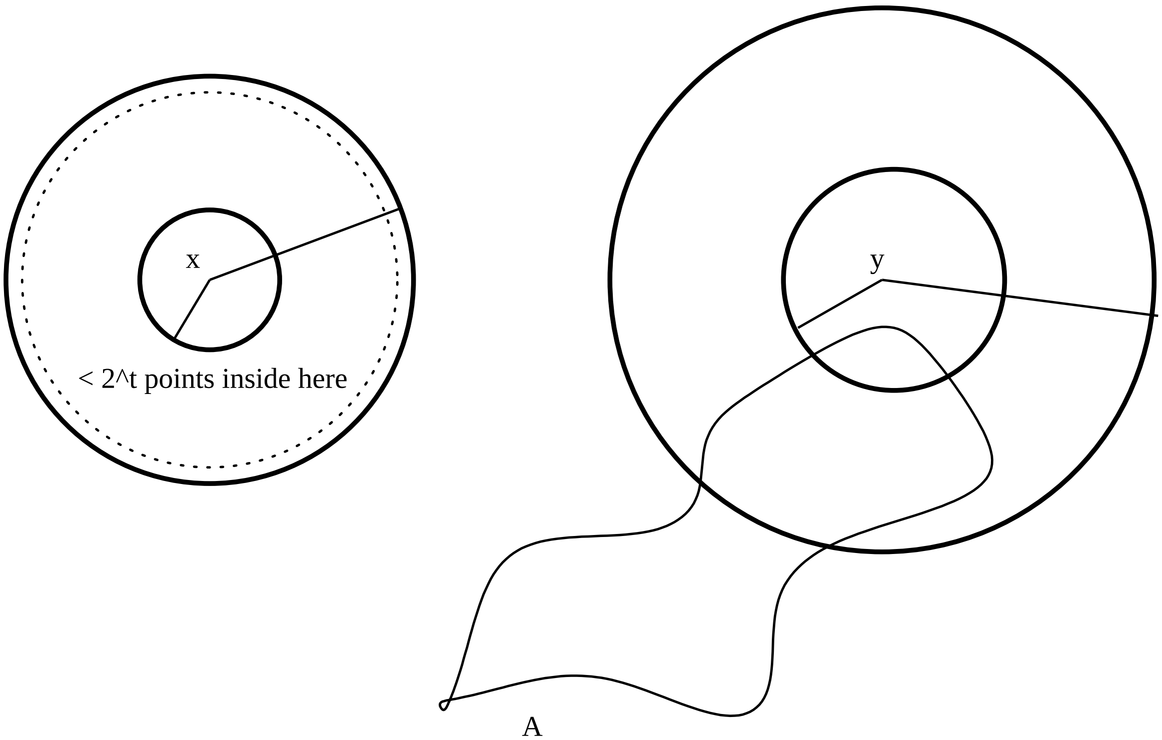

Figure 3: A is the type of set we want.

Recalling that the sets A_j^i are made up by chosing uniformly at random from the n points in V with replacement, we see that

$$\begin{aligned} \mathbb{P}(A_j^i \cap B^o(x, \rho_t) = \emptyset) &\geq \left[1 - \frac{2^t}{n}\right]^{\frac{n}{2^t}} \\ &\geq \frac{1}{4} \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}(A_j^i \cap B(y, \rho_{t-1}) \neq \emptyset) &\geq 1 - \left[1 - \frac{2^{t-1}}{n}\right]^{\frac{n}{2^{t-1}}\frac{1}{2}} \\ &\geq 1 - e^{-\frac{1}{2}}. \end{aligned}$$

Further we note without proof that there exists $\epsilon > 0$ (independent of n, V, etc) such that

$$\mathbb{P}(A_j^i \cap B^o(x, \rho_t) = \emptyset \text{ and } A_j^i \cap B(y, \rho_{t-1}) \neq \emptyset) \geq \epsilon.$$

Thus, every set A_j^i has this property with probability $\geq \epsilon$ and so the expected number of such A_j^i is $\mu \geq \epsilon q$. Taking $\delta = \frac{1}{2}$ and applying the Chernoff bound discussed earlier we see that

$$\begin{aligned}\mathbb{P}(< \frac{1}{2}\epsilon q \text{ of the } A_j^i\text{'s have this property}) &< e^{-\frac{1}{2}(\frac{1}{2})^2(\epsilon q)} \\ &= e^{-\frac{1}{8}(\epsilon q)}\end{aligned}$$

So, choose $q = \frac{24}{\epsilon}\ln n$ to get

$$\mathbb{P}(< \frac{1}{2}\epsilon q \text{ of the } A_j^i\text{'s have this property}) < \frac{1}{n^3}$$

Now, take the union bound over all $\binom{n}{2}$ choices of x and y and all $\log n$ size classes to see that with high probability all size classes have this property for all x, y pairs. Now, if this high probability event happens

$$\begin{aligned}\sum_{j=1}^{q} \left|d(x, A_j^i) - d(y, A_j^i)\right| &\geq \frac{1}{2}\epsilon q(\rho_t - \rho_{t-1}) \\ &= \Omega(\log n)(\rho_t - \rho_{t-1})\end{aligned}$$

as desired and we have thus proved the upper bound.

Thus

$$\Omega(\log n)d(x,y) \leq ||f(x) - f(y)||_1 \leq O(\log^2 n)d(x,y)$$

and so our embedding f does indeed have distortion $O(\log n)$.

4 Special cases

The theorem does give a tight bound, and the embedding is the best possible in the worst case, as we saw in Lecture 4 that for bounded degree expanders

$$\frac{\eta}{\lambda^*} = \Theta(\log n).$$

However, if we consider shortest-path metrics in certain restricted families of graphs (e.g. planar graphs) we can get a better distortion bound γ_n which means that for Multicommodity Flow on these types of graphs

$$\frac{\eta}{\lambda^*} = O(\gamma_n)$$

For example, trees are L_1-embeddable with distortion 1. For planar graphs, a result of Rao shows that an improvement to $O(\sqrt{\log n})$ can be made, although it remains an open question whether this can be further improved to $O(1)$. Gupta et. al. have given an embedding with $O(1)$ distortion for series-parallel graphs, which are a special type of planar graph that is constructed in the following manner: Start with two nodes. At each step you can either add an edge between two existing nodes, or place a node along an existing edge.

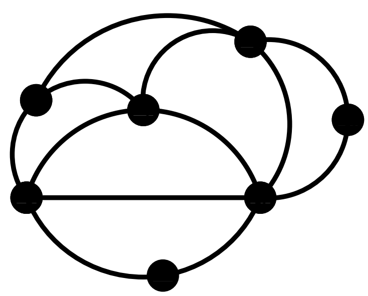

Figure 4: An example of a series-parallel graph.