PRINCETON U. SP'02 COS 598B: ALGORITHMS AND COMPLEXITY

Lecture 11: Embedding ℓ_1 into ℓ_2 with distortion $O(\sqrt{\log n}\log\log n)$.

Lecturer: *Sanjeev Arora* Scribe:*Daniel J. Peng*

Today we sketch the proof of the following theorem. Many details are omitted.

THEOREM 1 (ARORA, LEE, NAOR 2005)
Every n-point subset of ℓ_1 embeds into ℓ_2 with distortion $O(\sqrt{\log n}\log\log n)$.

Actually, the paper proves a more general result: that every ℓ_2^2 space embeds into ℓ_2 with the stated distortion. (Thus as a consequence, ALN conclude that ℓ_2^2 embeds into ℓ_1 with the stated distortion, which gives a $O(\sqrt{\log n}\log\log n)$-approximation for nonuniform SPARSEST CUT. It is conceivable that this particular result could be greatly improved, since the best lowerbound is less than $\log\log n$.) In fact the overview we give applies equally well to this more general statement.

The proof is related to but more complicated than our proof of Bourgain's theorem using the measured descent technique. In particular, we define a sequence of $O(\log n\log\log n)$ zero sets W_t and then the embedding $x \to (d(x, W_t))_t$.

In defining the zero sets we use the ARV Structure Theorem (improved by Lee) which we saw in context of SPARSEST CUT. The Structure Theorem is used via the following strengthening, which appears in a paper that gave a weaker version of Theorem 1 with distortion $O(\log^{3/4} n)$.

LEMMA 2 (CHAWLA, GUPTA, RACKE, 2005)
Let (X,d) be an ℓ_2^2 metric space of n unit vectors. Then for any scale s, there is a distribution on zero sets Y_s such that for all x,y with $d(x,y) > 2^{s+1}$, we have

$$\mathbf{Pr}\left[x \in Y_s \text{ and } d(y, Y_s) > \frac{2^s}{\sqrt{\lg n}}\right] \geq \frac{1}{8}.$$

(A proof of Lemma 2 is also sketched in the ALN paper. It uses an iterative reweighting of the set of points.)

Observe that the ARV Structure Theorem is some average-case version of this when (a) the largest distance is Δ and the average internode distance is $\Omega(\Delta)$ (b) the scale s in question is such that $2^s = \Omega(\Delta)$. To see this, just take the well-separated sets S, T given by the Structure Theorem and designate S as the zero set. Then $\Omega(n)$ of the points have distance at least $\Delta/\sqrt{\log n}$ to S.
NOTATION: In our proof of Bourgain's theorem we were cavalier with what we meant by growth ratio "GR." Today we will be even more cavalier: we use the loose notation

$$GR \approx \frac{|B(x, O(2^m \lg n))|}{\left|B\left(x, \Omega\left(\frac{2^m}{\lg n}\right)\right)\right|}.$$

Recall that such looseness in notation doesn't hurt much. Last time we explained as SUBTLETY 2 that if we let GR denote $\frac{|B(x,2^{m+c})|}{|B(x,2^{m-c'})|}$ then a simple "shifting" idea in the gluing allows us to compensate for this loose notation while increasing the number of coordinates by $O(c + c')$, which hurts the distortion bound by $\sqrt{c+c'}$. Today $c+c'$ is $\log\log n$, and indeed this is one of the two places in the proof why Theorem 1 seems to need the extra "$\log\log n$" factor.

Note that the statement of Lemma 2 is reminiscent of the FRT "padded decomposition" used in proving in Bourgain's theorem. However, the analogous form would have been:

$$\mathbf{Pr}\left[x \in Y_s \text{ and } d(y, Y_s) > \frac{2^s}{\sqrt{\lg GR}}\right] \geq \frac{1}{8},$$

where GR is the growth ratio. Proving this kind of statement is an open problem, but the ALN proof of Theorem 1 gets around this open problem by using a more clever gluing of the scales, and a random sampling idea.

Now we describe these ideas. The embedding will allocate a designated set of coordinates for each possible value of $\log\log GR$. Notice, the number of such values is $\log\log n$, and so the effect on the distortion calculation will also be a $\log\log n$ factor. The big advantage is that this will allow us to assume in the proof that we've already "guessed" the value of GR at x.

As in the proof of Bourgain's theorem, for each distance scale s, let P_s be the randomized partition (of the ℓ_1 space) obtained using the FRT theorem about padded decompositions. Recall that these satisfy the property that (i) each block in P_s has diameter at most 2^{s+1}. Thus for each x, y satisfying $d(x,y) > 2^{s+1}$, x, y are in different blocks of P_s (namely, $P_s(x') \neq P_s(y')$) (ii) The probability is at least 1/2 that $B(x, 2^s/10\log n) \subseteq P_s$.

Suppose we're interested in the distance from point x to point y in the final embedding. Let $d(x,y) \approx 2^m$ in the ℓ_1 space. Suppose the growth ratio GR at x has been "guessed" correctly.

Consider P_s, where $2^s \approx 10 \cdot 2^m \lg n$. The chance is at least 1/2 that $B(x, 2^{m+1})$ is contained entirely within $P_s(x)$. If this happens, then x and y are in the same block since $d(x,y) \approx 2^m$. Now take a random sample of size GR from this block and apply Lemma 2 to this sampled set. The set has size GR, so the distortion term "$\sqrt{\log n}$" in the lemma statement becomes $\sqrt{\log GR}$, and this is the effect we were looking for. (In other words, if two points in the sampled set have distance 2^m then with probability at least 1/8, the first is in the zero-set, and the other has distance at least $2^m/\sqrt{\log GR}$ to the zero-set.) The complication of course is that neither x nor y, the two points we are interested in, may be in the sampled set, and so we need some other way to reason about what happens to x, y. One helpful observation in this regard is that $GR = |B(x, O(2^m \log n))| \,/\, |B(x, 2^m/\log n)|$, and that the block $P_s(x) \subseteq B(x, O(2^m \log n))$. Thus $|B(x, 2^m/\log n)| \geq \frac{1}{GR}|P_s(x)|$, and when we take a random sample of size GR from the block $P_s(x)$ then whp *some* point in $B(x, 2^m/\log n)$ makes it into this sample. Thus instead of x, the proof tries to reason about this nearby point. (One also has to keep in mind what happens around y.)

So here is the first attempt at defining the embedding formally. As usual, we define a sequence of zero sets, and the final embedding has a coordinate for each zero set. For $t = 1, 2, \ldots, \log n$, the function $k(z,t)$ means the same as in our proof of Bourgain's theorem. Let $\widetilde{P_{s,G}}$ be the blocks of partition P_s after they have been sampled down to be of size G.

FIRST ATTEMPT: For each $t = 1, 2, \ldots, \log n$, for each G (the guessed value of GR) define the zero set:

$W_{t,G}$ = the set of z that lie in the zero set obtained by applying Lemma 2 to $\widetilde{P_{K(z,t),G}}$.

The above intuition suggested that if $d(x,y) \approx 2^m$ then

$$\mathbf{E}[(d(x, W_{t,G}) - d(y, W_{t,G}))^2] \geq \Omega\left(\frac{2^m}{\sqrt{G}}\right)^2,$$

whenever $k(x,t) \approx m$. (Here $\mathbf{E}[\cdot]$ is the expectation over all randomness used in all the steps.) As we noted in our proof of Bourgain's theorem, $K(x,t) \approx m$ for $\log GR$ values of t. Thus

$$\sum_t \mathbf{E}\left[(d(x,W_{t,G}) - d(y,W_{t,G}))^2\right] \geq \log GR \times \Omega\left(\frac{2^m}{\sqrt{\log G}}\right)^2 \approx \Omega(2^{2m}),$$

provided G is guessed correctly. Notice that this calculation only assumes that $\log G \approx \log GR$, so we conclude that actually we only needed to guess $\log GR$ up to $O(1)$ factor (as opposed to guessing GR upto $O(1)$ factor). This means that the number of values of t, G is at most $O(\log n)$, and thus

$$\sum_t (d(x,W_{t,G}) - d(y,W_{t,G}))^2 \leq O(\log n) 2^{2m}.$$

This would give a distortion of $O(\sqrt{\log n})$. Of course, to make everything formal about this sloppy reasoning, we need to introduce shifting as in the proof of Bourgain's theorem, which introduces an additional $\log\log n$ factor.

The above sketch mentioned the subtleties that need to be addressed, and details can be found in the ALN writeup. In particular, the definiton of the zero sets is much more complicated, and the proof needs to analyse many cases.The proof of Lemma 2 is also sketched in ALN.