Lecture 10.5: Bourgain's Theorem via Padded Decompositions.

20th Mar, 2015

We want to establish the following theorem

Theorem 1. *Given any metric d over the vertices V, and given k pairs (s_i, t_i), we wish to find a mapping $\phi : V \to \Re^K$ such that*

1. $||\phi(u) - \phi(v)||_1 \leq d(u, v)$ *for all pairs u and v.*
2. $d(s_i, t_i) \leq O(\log k)||\phi(s_i) - \phi(t_i)||_1$ *for all i.*

This was essentially proved by Bourgain. His proof was for the all pairs case (so $k = n^2$), and later London-Linial-Rabinovich and Aumann-Rabani extended it to general k. Below, we do not give their proof. But a different proof essentially due to Fakcharoenphol-Rao-Talwar and Calinescu-Karloff-Rabani. At the heart is the concept of **padded decompositions**.

Definition 1. *Given a metric d over V, a (β, Δ)-***padded decomposition** *of (V, d) is a* distribution *over partitions $(V_1, \ldots, V_T)$ with the following two properties*

1. *The (weak) diameter of each V_i is at most Δ.*
2. *For any vertex v,* $\mathbf{Pr}[B(u, r)$ is shattered by the partition$] \leq \beta \cdot \frac{4r}{\Delta}$

The weak diameter of a subset S is $\max_{u,v \in S} d(u, v)$, $B(u, r) := \{v : d(u, v) \leq r\}$ is the ball of radius r around u, and it's shattered by a partition if at least two parts have non-trivial intersection with it. Finally, a padded decomposition is said to be efficient if it can be efficiently sampled from.

In general, the above β is allowed to be a **function** parametrized by Δ which takes a vertex u as input. For the time being let's keep β to be fixed.

Let us first start with a connection to the low diameter decomposition lemma done in class last time. Consider a (β, Δ)-padded decomposition. As in the low-diameter-lemma, the diameter of each part is at most Δ (or $2R$ in the last class's notation). Let us now argue about the cross edges. Given any edge (u, v), the probability that (u, v) is a cross edge is at most the probability $B(u, d(u, v))$ is shattered. Therefore, the expected cost of the cross edges is at most $\frac{4\beta}{\Delta} \sum_{e \in E} c_e d_e = \frac{4\beta L}{D}$. Thus, for the **uniform** sparsest cut problem, we get a randomized algorithm with expected sparsity $O(\beta)$ times the LP.

Padded Decompositions and Embedding into ℓ_1. We now describe how padded decompositions imply embeddings in a fairly natural way. Our mapping ϕ will be a scaling of a direct sum of $\log D$ different ϕ_t's where D is $\max_{u,v} d(u, v)$ and t runs from 1 to $\log D$. Each ϕ_t is an embedding defined as follows.

1. Sample a partition from the $(\beta, 2^t)$-padded decompostion. Let T be the number of parts.
2. $\phi_t(u)$ is a T-dimensional vector corresponding to the different parts: it is 2^t corresponding to the part which contains u, and 0 otherwise.

Claim 1. *For any two points u and v, $||\phi_t(u) - \phi_t(v)||_1 = 2^{t+1}$ for all $t < \log_2 d(u,v) - 1$.*

Proof. Immediately follows from the fact that the diameter of every part is $\leq 2^t$ and if $t < \log_2 d(u,v) - 1$, then u and v cannot be in the same part, and so $||\phi_t(u) - \phi_t(v)||_1 = 2^{t+1}$. □

Claim 2. *For any two points u and v, we have* $\mathbf{Exp}[||\phi_t(u) - \phi_t(v)||_1] \leq \beta \cdot 8d(u,v)$.

Proof. The probability u and v are in different parts is at most $4\beta d(u,v)/2^t$, and therefore the claim follows. □

Now let us consider the embedding ϕ which takes the direct sum of all the ϕ_t's. By Claim 1, we get

$$\text{For any } u,v, \quad ||\phi(u) - \phi(v)||_1 \geq \sum_{t=0}^{\log_2 d(u,v)-2} 2^{t+1} \geq \frac{d(u,v)}{4} \tag{1}$$

By Claim 2, we get

$$\text{For any } u,v, \quad \mathbf{Exp}[||\phi(u) - \phi(v)||_1] \leq 8d(u,v) \sum_{t=0}^{\log_2 D} \beta_t(u) \tag{2}$$

Note that we have moved to the functional version of β which takes the diameter $\Delta = 2^t$ as a paramerter, and u as the input. If β were just a 'scalar', so to speak, then we would get $8\beta \log_2 D d(u,v)$ in the RHS. In sum, we get an embedding of d into ℓ_1 with distortion depending on the β-parameter of the padded decomposition. Next, we get a good padded decomposition for a general metric.

Good Padded Decompositions. We now describe a good padded decomposition by describing a randomized algorithm which generates samples from this.

1. Sample a random permutation σ of the points in V.
2. Sample $R \in [\Delta/4, \Delta/2]$ uniformly at random.
3. Define $V_i := \{v : d(i,v) \leq R\} \setminus \bigcup_{j \leq_\sigma i} V_j$.

It is clear that the diameter of every V_i is at most Δ; indeed it is at most $2R$. The next theorem shows it is a good padded decomposition. Let $V(u)$ be the V_i that contains u.

Theorem 2. *For any point u,* $\mathbf{Pr}[B(u,r) \not\subseteq V(u)] \leq \frac{8r}{\Delta} \cdot \log\left(\frac{|B(u,\frac{\Delta}{2}+r)|}{|B(u,\frac{\Delta}{4}-r)|}\right)$

Before we prove the theorem, note that the RHS is non-trivial only for $r \leq \Delta/8$. Therefore, we can upper bound the RHS by $\frac{8r}{\Delta} \log\left(\frac{|B(u,\Delta)|}{|B(u,\frac{\Delta}{8})|}\right)$, and so in (2) we can substitute

$$\beta_t(u) = \log\left(\frac{|B(u,2^t)|}{|B(u,2^{t-3})|}\right)$$

which implies that the (2) translates to

$$\text{For any } u, v, \quad \mathbf{Exp}[||\phi(u) - \phi(v)||_1] \leq 8d(u,v) \sum_{t=0}^{\log_2 D} \log\left(\frac{|B(u,2^t)|}{|B(u,2^{t-3})|}\right) \leq 24 \log n \cdot d(u,v)$$

This proves the $O(\log n)$-embedding of any metric onto ℓ_1. We have now all the ingredients for Bourgain's theorem as mentioned in the first para – how will you get the $O(\log k)$? Note that we do not need an upper bound of (1) for all pairs but only the k-pairs (s_i, t_i). We leave this as an exercise, and proceed to prove the theorem above.

Proof of Theorem 2. Let B denote the ball $B(u,r)$. Let us consider a vertex i such that V_i is the first in σ-order to shatter $B(u,r)$. For this to occur, we must have $d(u,i) - r \leq R$ and $R \leq d(u,i) + r$: the former since V_i intersects $B(u,r)$ and the latter since it doesn't contain all of it. Since $R \in [\Delta/4, \Delta/2]$, we get that i must lie in the set $X := B(u, \Delta/2 + r) \setminus B(u, \Delta/4 - r)$.

Furthermore, in the random permutation σ, i must appear before any vertex in $B(u, \Delta/4 - r)$ should appear before i. Now we can make a similar calculation as done for the multicut problem.

$$\begin{aligned}
\mathbf{Pr}[B(u,r) \not\subseteq V(u)] = & \quad \mathbf{Pr}_{R,\sigma}[\exists i \in X : V_i \text{ is the first in } \sigma \text{ to shatter } B(u,r)] \\
\leq & \quad \sum_{i \in X} \mathbf{Pr}_{R,\sigma}[V_i \text{ is the first in } \sigma \text{ to shatter } B(u,r)] \\
\leq & \quad \sum_{i \in X} \mathbf{Pr}_{R,\sigma}[R \in [d(u,i) \pm r] \textbf{ and } \mathcal{E}_i]
\end{aligned}$$

where $\mathcal{E}_i$ is the event that all vertices $j \leq_\sigma i$ s.t. $j \in B(u, \Delta/2 + r)$ have $d(j,B) > d(i,B)$ and no vertex in $B(u, \Delta/4 - r)$ should appear before i. Note that $\mathcal{E}_i$ doesn't occur then i is not the first vertex to shatter $B(u,r)$. As in the multicut proof, we get a harmonic sum which starts at $\frac{1}{|B(u,\Delta/2+r)|}$ and ends at $\frac{1}{|X|+1}$. This proves the theorem. □