Bourgain's Theorem via Padded Decompositions[1]

- *Bourgain's Theorem.* In the last lecture, we saw how the generalized/non-uniform sparsest cut can be solved if we could find metric embeddings of a general metric into $\mathcal{L}_1$ with low distortion. In particular, the following theorem of Bourgain (stylized to capture distortion with respect to S) immediately implies a $O(\log k)$-approximation for the general sparsest cut problem.

  Theorem 1 (Bourgain's Theorem, the Terminal Version)**.** Given any metric space (V, d) and a set $S \subseteq V$ of size at most k, there is a mapping $\psi : V \to \mathbb{R}^{O(\log^2 k)}$ such that with high probability, we have that for any pair of vertices u and v, $||\psi(u) - \psi(v)||_1 \leq d(u, v)$ and for any pair $u, v \in S$, $d(u, v) \leq O(\log k)||\psi(u) - \psi(v)||_1$.

- In this note we give a sketch of a proof. In particular, we focus on the case of $S = V$, that is the case of all pairs. Furthermore, we only prove an "expectation" result rather than a "with high probability" result. More precisely, we describe a randomized algorithm which produces a $\phi : V \to \mathbb{R}^h$ such that for any two points u and v we have $\|\phi(u) - \phi(v)\|_1 \geq d(u, v)$ but $\mathbf{Exp}[\|\phi(u) - \phi(v)\|_1] \leq O(\log n) \cdot d(u, v)$. The "with high probability" statement can be obtained by "repeating, averaging, and concatenating" and applying standard deviation inequalities like the Chernoff bound. We leave this as an exercise. The ψ in the theorem is obtained by defining $\psi(u) := \frac{\phi(u)}{C \log n}$ for a sufficiently large C.

  We describe a proof which uses the random permutation idea that we saw in the randomized multicut algorithm. The key definition is that of **padded decompositions**.

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

  a. The (weak) diameter of each $V_i \in \Pi$ is at most Δ.
  b. For any vertex u and radius r, $\mathbf{Pr}_\Pi[B(u, r)$ *is shattered by* $\Pi] \leq \beta(u) \cdot \frac{4r}{\Delta}$

  Here $\beta : V \to \mathbb{R}_{\geq 0}$ is a function mapping a non-negative real to u, and could depend on Δ. The weak diameter of a subset S is $\max_{u,v \in S} d(u, v)$. The set $B(u, r) := \{v : d(u, v) \leq r\}$ is the ball of radius r around u, and it is shattered by a partition Π if at least two parts Π_i and Π_j have non-trivial intersection with the ball. Finally, a padded decomposition distribution is said to be efficient if it can be efficiently sampled from.

- ***Padded Decompositions and Embedding into ℓ_1.*** We now describe how padded decompositions imply embeddings in a fairly natural way. Let $D := \max_{u,v \in V} d(u, v)$. Our (randomized) mapping ϕ will be a concatenation of these $\lceil \log_2 D \rceil$ different ϕ_t's.

[1]*Lecture notes by Deeparnab Chakrabarty. Last modified : 18th Mar, 2022*
These have not gone through scrutiny and may contain errors. If you find any, or have any other comments, please email me at deeparnab@dartmouth.edu. Highly appreciated!

procedure RANDOMIZED EMBEDDING(V, d):
  for $t = 0$ to $\lceil \log_2 D \rceil$ **do**: ▷ $D := \max_{u,v \in V} d(u,v)$.
    Sample $\Pi_t := (V_1, \ldots, V_{d_t})$ from a $(\beta_t, 2^t)$-padded decomposition distribution. ▷ *β_t's will be defined later*
    Define $\phi_t(u)$ as a d_t-dimensional vector corresponding to the different parts:

$$\phi_t(u)[i] = \begin{cases} 2^t & \text{if } u \in V_i \\ 0 & \text{otherwise} \end{cases}$$

    ▷ *If u and v are in different parts of Π_t, then $\|\phi_t(u) - \phi_t(v)\|_1 = 2^{t+1}$, else it is 0*
  Let ϕ be a *concatenation* of these $\lceil \log_2 D \rceil$ different ϕ_t vectors. So, the dimension of ϕ is $\sum_t d_t$.

Claim 1. For any two points u and v and any t, $\mathbf{Exp}[||\phi_t(u) - \phi_t(v)||_1] \leq \beta_t(u) \cdot 8d(u,v)$. Furthermore, if $t < \log_2 d(u,v)$, then $||\phi_t(u) - \phi_t(v)||_1 = 2^{t+1}$ with probability 1.

Proof. u and v are in different parts of Π_t is equivalent to the event that the ball $B(u, d(u,v))$ is shattered by Π_t. By the definition of padded decompositions, the probability of this is at most $4\beta_t(u)d(u,v)/2^t$. Therefore, $\mathbf{Exp}[||\phi_t(u) - \phi_t(v)||_1] \leq \frac{4\beta_t(u)}{2^t} \cdot 2^{t+1}$, and thus the first assertion of the claim follows. Furthermore, if $t < \log_2 d(u,v)$ implying $d(u,v) > 2^t$, then from the fact that the diameter of every part is $\leq 2^t$ one gets that u and v cannot be in the same part. And so, $||\phi_t(u) - \phi_t(v)||_1 = 2^{t+1}$ with probability 1. □

By the second assertion in Claim 1, we get

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

By the first assertion in Claim 1, we get

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

In sum, we get an embedding of d into ℓ_1 with distortion depending on the β-parameter of the padded decomposition. In the next bullet point, we show how to obtain a padded decomposition with the following parameters.

Theorem 2. For any metric space (V, d) and parameter t, there exists a $(\beta_t, 2^t)$ padded decompisition with

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

If we substitute this in (2), we get

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

Now note that the summations telescope to $\leq 24 \ln n \cdot d(u,v)$. And this completes the proof sketch of Theorem 1.

- ***Padded Decomposition Distributions.*** We now describe a randomized algorithm which generates samples from a $(\beta_t(u), 2^t)$-padded decomposition with $\beta_t(u) \leq 2\ln\left(\frac{|B(u,2^t)|}{|B(u,2^{t-3})|}\right)$.

```
1: procedure PADDED DECOMPOSTION(t): ▷ Return a padded decomposition as asserted in Theorem 2
2:     Sample a random permutation σ of the points in V.
3:     Sample R ∈ [2^{t−2}, 2^{t−1}] uniformly at random.
4:     Define V_i := {v : d(i,v) ≤ R} \ ⋃_{j≤_σ i} V_j.
```

It is clear that, by design, the diameter of every V_i is at most $2R \leq 2^t$. What is more interesting is to prove that for any $u \in V$ and any r, the probability $B(u,r)$ is shattered is at most $\beta_t(u) \cdot \frac{4r}{2^t}$. Let B denote this ball $B(u,r)$. First, observe that if $r > 2^{t-3}$, then $\frac{4r}{2^t} > 1$ and so the shattering claim holds vacuously. Therefore, henceforth we assume $r \leq 2^{t-3}$.

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 $R \in [2^{t-2}, 2^{t-1}]$, we get that i must lie in the set $X := B(u, 2^{t-1} + r) \setminus B(u, 2^{t-2} - r)$. Furthermore, in the random permutation σ, i must appear before any vertex $j \in B(u, 2^{t-2} - r)$ otherwise i won't be the *first* vertex to shatter the ball (either someone else would have shattered, or j would've gobbled the whole ball $B(u,r)$.) Finally, note that if i can non-trivially intersect B, then any $j \in X$ with $d(j,B) \leq d(i,B)$ can non-trivially intersect B. Therefore, if i were the first in σ to shatter B, it better be that all $j \in X$ with $d(j,B) \leq d(i,B)$ come *after* i in σ.

$$\begin{aligned}
\mathbf{Pr}[B(u,r) \quad \text{shattered}] = & \quad \mathbf{Pr}_{R,\sigma}[\exists i \in X : V_i \text{ is the first in } \sigma \text{ to shatter } B(u,r)] \\
\leq & \quad \sum_{i \in X} \mathbf{Pr}_{R,\sigma}[V_i \text{ is the first in } \sigma \text{ to shatter } B(u,r)] \\
\leq & \quad \sum_{i \in X} \mathbf{Pr}_{R,\sigma}[R \in [d(u,i) \pm r] \textbf{ and } \mathcal{E}_i]
\end{aligned}$$

where $\mathcal{E}_i$ is the event that all vertices $j \in B(u, 2^{t-1} + r) \leq_\sigma i$ satisfy (a) $j \notin B(u, 2^{t-2} - r)$ and (b) $d(j,B) > d(i,B)$. As explained above, if $\mathcal{E}_i$ doesn't occur then i cannot be the first vertex to shatter B. Note that $\mathcal{E}_i$ is independent of $R \in [d(u,i) \pm r]$. And therefore,

$$\mathbf{Pr}[B(u,r) \quad \text{shattered}] \leq \sum_{i \in X} \mathbf{Pr}_R[R \in [d(u,i) \pm r] \cdot \mathbf{Pr}[\mathcal{E}_i] \quad \leq \frac{4r}{2^t} \cdot \sum_{i \in X} \mathbf{Pr}[\mathcal{E}_i]$$

If we sort the points in $B(u, 2^{t-1} + r)$ in increasing order of distance from u, then $\mathbf{Pr}[\mathcal{E}_i]$ is $\frac{1}{i}$, and i ranges precisely from $|B(u, 2^{t-2} - r)|$ to $|B(u, 2^{t+1} + r)|$ since that is where the points in X lie. This harmonic sum is indeed bounded by

$$\ln\left(\frac{|B(u, 2^{t-1} + r)|}{|B(u, 2^{t-2} - r)|}\right) \leq \ln\left(\frac{|B(u, 2^t)|}{|B(u, 2^{t-3})|}\right)$$

since $r \leq 2^{t-3}$. This ends the sketch of the proof of Theorem 2.

Notes

Bourgain's theorem on metric embeddings is from the paper [2]. The terminal version as stated in Theorem 1 is first stated in the paper [5] by Linial, London, and Rabinovich, and also in the paper [1] by Aumann and Rabani. The proof above is inspired from the paper [4] by Fakcharoenphol, Rao, and Talwar, which itself is inspired from the paper [3] by Calinescu, Karloff, and Rabani.

References

[1] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM Journal on Computing (SICOMP)*, 27(1):291–301, 1998. Prelim. Version, in IPCO 1995.

[2] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1):46–52, 1985.

[3] G. Calinescu, H. Karloff, and Y. Rabani. An Improved Approximation Algorithm for multiway cut. *J. Comput. System Sci.*, 60, 2000.

[4] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proc., ACM Symposium on Theory of Computing (STOC)*, pages 448–455, 2003.

[5] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–246, 1995. Prelim. Version. in FOCS 1994.