University of Michigan–Ann Arbor

Department of Electrical Engineering and Computer Science

EECS 498 004 **Advanced Graph Algorithms**, Fall 2021

Lecture 4: Bourgain's Theorem

September 9, 2021

Instructor: Thatchaphol Saranurak

Scribe: Cheng-Hao Fu

1. Briefly Motivate and State the Theorem

In general, an arbitrary distance metric (X, d) can be difficult to work with directly. As seen in the algorithm for computing conductance, it is far easier to work with ℓ_1 metric. In this pursuit, finding embeddings of (X, d) into adequately small-dimension $\mathbb{R}^n$ quickly is a gold-standard of computation.

Bourgain's Theorem gives us a guarantee that we can find a low distortion embedding into low-dimension Euclidean space:

1.1. Theorem (Bourgain)**.** *Given an n-point metric (X, d), there is an embedding $f : X \to \mathbb{R}^{O(\log^2 n)}$ such that, for any $x, y \in X$,*

$$\|f(x) - f(y)\|_1 \leq d(x, y) \leq O(\log n) \cdot \|f(x) - f(y)\|_1.$$

Moreover, given d, in $\tilde{O}(n^2)$ time, the mapping f can be computed correctly with probability at least $1-1/n$. If (X, d) is a shortest path metric of an m-edge graph, then the running time is $\tilde{O}(m)$ time.

Note that our proof will show an embedding f where

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

And after scaling $\bar{f} = \frac{f}{(\log n)^2}$, $\bar{f}$ satisfies the desired inequality in the theorem.

The usefulness of this theorem is clear. For example, in the case of graphs with n vertices, the theorem implies the existence of an assignment of $O(\log^2 n)$ numbers (representing the dimension of Euclidean space) to each node that allows distance calculation on the graph to be done by comparing the labels element-wise with a small amount of distortion. This is much better than the $O(n^2)$ guarantee that Dijkstra's gives on calculating Single Source Shortest Path.

2. The Algorithm

The algorithm itself is quite simple. We first define a notion of the distance of any point to a particular set of points $A \subseteq X$ as follows:

2.1. Definition. The distance between a point x and a set of points A with respect to the distance metric is:

$$d(x, A) := \min_{a \in A} d(x, a).$$

We then define A_{ij} for $i = 1, \ldots, i_{\max} = 10000 \log n$ and for $j = 1, \ldots, j_{\max} = \lceil \log n \rceil$ by letting $A_{ij} \subseteq X$ include each element of X with probability $1/2^j$. We sample all A_{ij} independently.

We then define our embedding as follows:

$$f : X \to \mathbb{R}^{i_{\max} j_{\max}} \text{ where } f_{ij}(x) = d(x, A_{ij})$$

We note that every element in the space now has an associated vector of labels that is of length $O(\log^2 n)$, representing its location in $\mathbb{R}^{i_{\max} j_{\max}}$.

It is then clear that if this embedding achieves the desired distortion, along with the runtime, we will be done.

3. Proving Runtime

3.1. Claim. *Given d, the embedding f can be computed in $\tilde{O}(n^2)$ time. If (X, d) is a shortest path metric defined by an m-graph, then f can be computed in $\tilde{O}(m)$ time.*

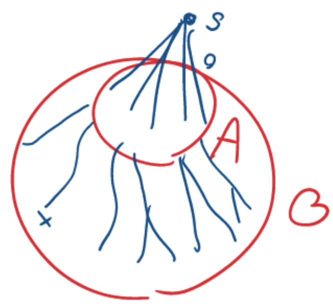
Proof of Claim. First of all, it takes $\tilde{O}(n)$ time to sample all A_{ij}. To compute each entry of $f(x)$

for all $x \in X$, when (X, d) is a shortest path metric, instead of fixing a point in X and compute the distance between itself and all possible A_{ij}, we switch gear by fixing a A_{ij} and computing its distance between all $x \in X$. To do this, for a fixed A_{ij}, we assign a source s that connects to each point in A_{ij} with distance 0, and running Dijkstra's (which takes $\tilde{O}(m)$ time) to compute the shortest path between any point x and s, which is exactly the distance between x and A_{ij}. Since there are $O(\log n^2)$ A_{ij}s that we have to do this with, computing the embedding will take $\tilde{O}(m)$ time, as desired.

When (X, d) is a general metric, the key observation is that we can treat any general metric as the shortest path metric on the weighted complete graph $G_d(X, E, \kappa)$ where $\forall \{x, y\} \in E, \kappa(\{x, y\}) = d(x, y)$. In this way, we could use the similar strategy as we described above to compute f in $\tilde{O}(n^2)$ time since G_d has $O(n^2)$ edges. It remains to show that d is indeed the shortest path metric on G_d. $\forall \{x, y\} \subseteq V$, on one hand, $\mathrm{dist}_{G_d}(x, y) \leq d(x, y)$ since there exists the path $< x, y >$ with weight $d(x, y)$ between x and y; on the other hand, $\mathrm{dist}_{G_d}(x, y) \geq d(x, y)$ since d satisfies triangle inequality whence by an inductive argument we can see that $\forall P \in \mathcal{P}_{x,y}, d(P) \geq d(x,y)$, and therefore the distance of the shortest $(x,y)-$path is also greater than $d(x,y)$.

□

For an illustration of the process:



4. Analysis of Algorithm

Now, the goal is to prove $\Omega(\log n) \cdot d(x,y) \leq \|f(x) - f(y)\|_1 \leq O(\log^2 n) \cdot d(x,y)$. We begin with the upper bound.

4.1. Upper Bound

4.1. Claim. $\|f(x) - f(y)\|_1 \leq O(\log^2 n) \cdot d(x,y)$.

Proof of Claim. We start by breaking down the definition of $\|f(x) - f(y)\|_1$. It is simply the sum of the magnitudes of the element-wise differences of $f(x)$ and $f(y)$. We can then use the definition of f to arrive at:

$$\|f(x) - f(y)\|_1 = \sum_{ij} |f_{ij}(x) - f_{ij}(y)| = \sum_{ij} |d(x, A_{ij}) - d(y, A_{ij})|$$

We know that there are only $O(\log^2 n)$ elements in the final summation, which is exactly the multiplier of $d(x,y)$ in the desired statement. This tells us that proving $|d(x,A_{ij}) - d(y,A_{ij})| \leq d(x,y)$ is sufficient for all i,j.

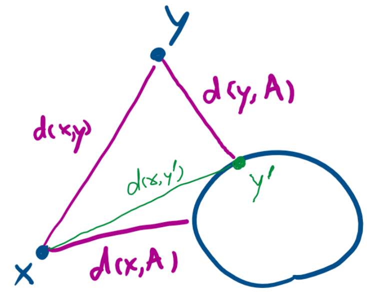
Without loss of generality, assume that $d(x,A_{ij}) \geq d(y,A_{ij})$. It suffices to prove that $d(x,A_{ij}) \leq d(x,y) + d(y,A_{ij})$

Let y' be such that $d(y,A_{ij}) = d(y,y')$.

By definition, we have that $d(x,A_{ij}) \leq d(x,y')$. Then, we get $d(x,y') \leq d(x,y) + d(y,y') = d(x,y) + d(y,A_{ij})$ by triangle inequality. Note that this actually works for any distance metric (X,d) with $A \subseteq X$.

□

Visually:



4.2. Lower Bound

4.2. Claim. $\|f(x) - f(y)\|_1 \geq \Omega(\log n) \cdot d(x, y)$.

Proof of Claim. We begin by defining a couple of useful objects.

4.3. Definition. For any point s and radius r, define $B(s, r) = \{z | d(z, s) \leq r\}$ to be the ball radius r around s. Similarly, $B^o(s, r) = \{z | d(z, s) < r\}$ will be referred to as the open ball around s with radius r.

We then fix $x, y \in X$, and define a couple more things.

Let $r_0 = 0$. For $j = 1, \ldots, j_{\max} = \lceil \log n \rceil$, r_j be the minimum r where both $|B(x, r)|, |B(y, r)| \geq 2^j$.

Define $\Delta_j = r_j - r_{j-1}$. Think of this as the amount that you have to grow the ball in order for the ball to encompass the next magnitude in size. Notice that, by telescoping, $\sum_{j=1}^{j_{\max}} \Delta_j = r_{j_{\max}}$.

We start off by assuming $r_{j_{\max}} < d(x, y)/4$ for now. Combinatorially, this means that the radius that we are considering is significantly less than the distance between them. This, of course, implies that $B(x, r_{j_{\max}})$ and $B(y, r_{j_{\max}})$ are disjoint. Note that there is no reason to expect that this assumption will hold. We will relax the assumption later in the proof.

Without loss of generality, assume that $|B(x, r_j)| \geq |B(y, r_j)|$. This means that y was the point that required r_j to be as large as it was, in fact, we can say then that $|B(y, r_j)| = 2^j$. If it is the case that multiple distances that are the same cause this assumption to be false, we can perturb each of the distances by a small constant ϵ, which will not cause any significant distortion, and give us the equality we are looking for. Thus, we can say that, $|B^o(y, r_j)| < 2^j$. Also, trivially $|B(x, r_{j-1})| \geq 2^{j-1}$.

4.4. Definition. We say a set A_{ij} is *good* if $A_{ij} \cap B(x, r_{j-1}) \neq \emptyset$ and $A_{ij} \cap B(y, r_{j-1}) = \emptyset$

We use the term *good* because it will help us definitively add to the minimum distance between the two points. In fact, we can observe that it will add at minimum Δ_j to $\|f(x) - f(y)\|_1$. This is because $d(y, A_{ij}) \geq r_j$ but $d(x, A_{ij}) \leq r_{j-1}$.

This is extremely powerful. To see why, let us see what happens when all A_{ij} are good. Then we would have

$$\|f(x)-f(y)\| = \sum_{ij} |d(x, A_{ij}) - d(y, A_{ij})| \geq \sum_{ij} \Delta_j = i_{\max} \cdot r_{j_{\max}} = \Omega(\log n) \cdot d(x,y)$$

The first equality comes from the definition of f. The second inequality comes from the justification provided above, and the rest is by definition (Noting that our assumption has $r_{j_{\max}} \approx d(x,y)/4$). This is exactly the result we are looking for, but unfortunately not all A_{ij} are good. But it should be clear that since we are aiming for $\Omega(\log n)$, a constant fraction of A_{ij} being good for each j will be sufficient for the bound. This is what we will next attempt to prove.

4.5. Proposition. *For every i, j, A_{ij} is good with probability at least $1/100$.*

Proof of Proposition. First, we calculate a lower bound the probability that $A_{ij} \cap B^o(y, r_j) = \emptyset$. Because of the way that we picked all of the points in A_{ij}, we have that the probability that every element in the open ball is avoided by A_{ij} is:

$$(1 - 1/2^j)^{|B^o(y,r_j)|} > (1 - 1/2^j)^{2^j} \geq 1/10.$$

The second inequality can be verified via graphing.

Similarly, $A_{ij} \cap B(x, r_{j-1}) = \emptyset$ occurs with probability:

$$(1 - 1/2^j)^{|B(x,r_{j-1})|} \leq (1 - 1/2^j)^{2^{j-1}} \leq 9/10.$$

Thus, $A_{ij} \cap B(x, r_{j-1}) \neq \emptyset$ occurs also with probability at least $1/10$. Because of our assumption that these two balls have an empty intersection, these two events are independent, and thus we can multiply their probabilities together to get a lower bound on the probability that A_{ij} is good:

$$\Pr[A_{ij} \cap B(x, r_{j-1}) \neq \emptyset \text{ and } A_{ij} \cap B^o(y, r_j) = \emptyset] \geq 1/10 \times 1/10 = 1/100$$

□

From the above, for each j, the expected total number of good sets A_{ij} over all i is at least $i_{\max}/100$.

However, before we proceed with the final part of the proof, we need to show that this occurs with high enough probability to get the high probability argument that we also want. This motivates the need for the following claim:

4.6. Claim. *For each j, the total number of good A_{ij} over all i is at least $i_{\max}/200$ with probability at least $1 - 1/n^{10}$.*

Proof of Claim. We will simplify this to the question of proving an upper bound on the probability of $i_{max}/200$ or fewer successes in i_{max} trials of Bernoulli random variables with success probability $1/100$. Let us define the random variables $X_1, .., X_{i_{max}}$ to represent the trials, and $Y = \sum_i X_i$.

Note that with Chernoff Bound, we have:

$$\Pr[Y \leq i_{max}(1/100 - 1/200)] \leq e^{-(1/8)(1/100)i_{max}}$$

Plugging in i_{max} with $10000 \log n$:

$$\Pr[Y \leq i_{max}/200] \leq e^{-(1/8)(1/100)10000 \log n} \leq 1/n^{10}$$

□

This gives us the following corollary:

4.7. Corollary. *For probability at least* $1 - 1/n^{10}$, $\|f(x) - f(y)\| \geq \frac{i_{\max}}{200} \cdot r_{j_{\max}} = \Omega(\log n) \cdot d(x, y)$.

Proof. This follows pretty much exactly the same as the argument used at the top of page 5. The only difference is that we can only sum over the good sets, which makes us lose a constant factor, which is still good enough for us.

$$\|f(x) - f(y)\| \geq \sum_{ij: A_{ij} \text{ is good}} |d(x, A_{ij}) - d(y, A_{ij})| \geq \sum_{ij: A_{ij} \text{ is good}} \Delta_j \geq \frac{i_{\max}}{200} \cdot r_{j_{\max}}$$

The last inequality holds with the probability desired, and so we are done. □

This corollary gives us the tools to prove the lower bound, however, we have to keep in mind that we are still operating under the assumption that $r_{j_{\max}} < d(x, y)/4$. In order to lift this assumption, we redefine $r_j' = \min\{r_j, d(x, y)/4\}$ and work with r_j' instead. By noticing that r_j' is a lower bound on r_j, the above analysis can still go through.

Now that we have addressed this point, we can finish the proof:

Since each distance $d(x, y)$ fails in being mapped to within the bounds with $1/n^{10}$ probability, we can use union bound to see that over all pairs, the probability of failure is less than or equal to:

$$\binom{n}{2}/n^{10} \leq 1/n^8$$

This allows us to say: For probability at least $1 - 1/n^8$, $\|f(x) - f(y)\| = \Omega(\log n) \cdot d(x, y)$ for all $x, y \in X$, which proves the original claim. □

Thus completes the proof of Bourgain's Theorem.

5. Examples

To motivate the way we have built the framework for the theorem, we will begin with a very basic attempt at embedding a distance into a ℓ_1 metric, and justifying why our approach needs to be as complicated as it is. This will be primarily done by testing ideas on shortest path metrics on various types of graphs. In addition to the discussion section, a large part of this section will be supported by the lecture notes of Luca Trevisan of UC-Berkeley[1].

5.1. Cycles

A natural precursor to the method illustrated in Bourgain's Theorem picks a point uniformly at random, and takes the embedding as labeling any point a with some constant factor times the distance between a and that point.

$$f(v) = k \times d(v, a)$$

Distances are then, naturally, the differences between these labels. The fact that this is a metric follows trivially. This is effective on cycles. We start by proving the following fact:

5.1. Claim. *Let a be a random node.* $E[|d(x, a) - d(y, a)|] = \Theta(d(x, y))$.

Proof of Claim. By the same argument used in the proof for 4.1, it is clear that $E[|d(x, a) - d(y, a)|] = O(d(x, y))$. Thus, we just have to justify that the expectation is not too small.

Consider the expected value of the distance when a is on the path between x and y. It is more or less uniformly distributed between 0 and $d(x, y)$ (depending on parity), with the low when the random point is the midpoint, and the high being when the point is one of the endpoints. This occurs with probability approximately equal to $d(x, y)/n$

Thus:

$$E[|d(x, a) - d(y, a)|] \approx d(x, y)^2/(2n) + \alpha$$

Where α represents the contribution in the other case. We also notice that the embedded distance is ranged from $d(x, y)$ to 0, with the high occurring whenever the paths representing x to a and y to a overlap, and the low happening whenever the random point is exactly the same distance to both x, y. This means that α contributes roughly (slightly more than) $d(x, y)/2$ on average per point. This clearly tells us that $E[|d(x, a) - d(y, a)|] = \Omega(d(x, y))$.

Thus finishes the proof. $\square$

In much the same fashion as classic Bourgain's, we can then use a logarithmic number of f as described above, to get the desired probability for the accuracy of our bound.

[1] https://lucatrevisan.github.io/teaching/expanders2016/lecture10.pdf

5.1.1. Not Quite Enough

As is clear upon reading the proofs, this method is much easier. However, we will next see that it is not an effective way to embed an arbitrary metric.

Suppose that we have a metric that has distances between any two points be either 1 or 2. It is 2 if and only if z is one of the endpoints. All other distances are 1.

Looking at $E[|d(r,z) - d(r,v)|]$, reveals:

$$E[|d(r,z) - d(r,v)|]1 + 2/n$$

.
This means that the distortion has the potential to be linear, which is a problem. There are many steps between this simple argument with the proof discussed above, but hopefully this gives sufficient information to inform a little bit of the background of Bourgain's.