Lecture 9: Memoryless random variables and low-diameter partitions CSE 525, Fall 2016
Instructor: James R. Lee

1 Random tree embeddings

Let (X, d) be a metric space on n points. Many problems in computation involve data points equipped with a natural distance. Metric spaces also arise as the solutions to linear programming relaxations of combinatorial problems. Often, it is useful to embed a metric space into a "simpler" one while changing the distances as little as possible. This is beneficial if an algorithmic problem can be solved more easily on the simple space.

A prime example of a simple metric space is a *tree metric.* Let T be a graph-theoretic with vertices $V(T)$ and edge set $E(T)$. We will assume that the tree is equipped some nonnegative length $\ell(e)$ on each edge $e \in E(T)$. There is a canonical shortest-path metric which we denote d_T.

A natural goal is to try and map our space (X, d) into a tree metric so that we preserve distances multiplicatively. In other words, we would look for a map $f : X \to V(T)$ so that

$$d(x, y) \leqslant d_T(f(x), f(y)) \leqslant D \cdot d(x, y) \qquad \forall x, y \in X\,,$$

and the *distortion* D is as small as possible. Unfortunately, this doesn't work so well. For instance, if (X, d) is the shortest-path metric on an unweighted n-cycle, one can show that any such mapping must have $D \geqslant \Omega(n)$. (Getting this argument right is actually a little tricky, but it is true.)

On the other hand, if we allow ourselves to use a *random* embedding, then we can approximate distance well in expectation. Consider again the shortest-path metric on an n-cycle C_n. Let T be the random (unweighted) tree that results from a single uniformly random edge of C_n. It's easy to see that for every $x, y \in V(C_n)$, we have

$$d_{C_n}(x, y) \leqslant \mathbb{E}\left[d_T(x, y)\right] \leqslant 2 \cdot d_{C_n}(x, y)\,.$$

The pair with the largest distortion is an edge $\{x, y\}$ of C_n; the expected length of $\{x, y\}$ in T is $(1 - \frac{1}{n})1 + \frac{1}{n}(n-1) \leqslant 2$.

Non-contracting tree embeddings embeddings. We now formalize the goal of embedding into a random tree. We say that (X, d) *admits a random tree embedding with distortion D* if there exists a random tree metric T and a random map $F : X \to V(T)$ that satisfies the following two properties:

1. **Non-contracting.**

   With probability one, for every $x, y \in X$, we have $d_T(F(x), F(y)) \geqslant d(x, y)$.

2. **Non-expanding in expectation.**

   For all $x, y \in X$,
   $$\mathbb{E}[d_T(F(x), F(y))] \leqslant D \cdot d(x, y)\,.$$

There are many scenarios in approximation algorithms and online algorithms where such embeddings can be used to reduce solving a problem in the general case to solving it on a tree by losing a factor of D in the approximation ratio (see Homework #3 for an example). It's also the case that such mappings can be useful for preconditioning diagonally dominant linear systems; in this case, one usually loses a factor of D (or $D^{O(1)}$) in the running time. In a sequence of works, Bartal showed that one can achieve $D = O(\log n \log\log n)$. The optimal bound was obtained a few years later.

Theorem 1.1 (Fakcharoenphol-Rao-Talwar 2003)**.** *Every n-point metric space (X, d) admits a random tree embedding with distortion $O(\log n)$.*

In Homework #3, you will prove that the theorem holds with $D = O((\log n)^2)$. We now discuss the basic primitive one needs.

2 Random low-diameter partitions

Given a parameter $\Delta > 0$ (the diameter bound), our goal is to construct a *random* partition $X = C_1 \cup C_2 \cup \cdots \cup C_k$ of X into sets where $\mathrm{diam}(C_i) \leqslant \Delta$ for every $i = 1, \ldots, k$. Here, $\mathrm{diam}(S) = \max_{x,y \in S} d(x, y)$ denotes the maximum distance in a subset $S \subseteq X$.

Of course, this is easy (we could simply decompose X into sets of singletons). We will also require that for every $x, y \in X$, we have

$$\mathbb{P}[x \text{ and } y \text{ are separated in } P] \leqslant \frac{d(x, y)}{\Delta} \cdot \alpha\,. \tag{2.1}$$

We say that a partition $P = \{C_1, C_2, \ldots, C_k\}$ separates x and y if $x \in C_i$, $y \in C_j$ and $i \neq j$. Our goal will be to prove that such random partitions always exist if we take $\alpha = 8 \ln n$.

Exercise: Prove that if $X = \mathbb{R}$ and $d(x, y) = |x - y|$ then we can find such a random partition with $\alpha = 1$. This demonstrates why the scaling in (2.1) is natural.

2.1 Memoryless random variables

Geometric random variables. Write $X \sim \mathrm{Geom}(p)$ to denote the fact that X is a geometric random variable with mean $1/p$. Recall that X is the number of independent coin flips needed to get heads when the coin comes up heads with probability p. It is easy to see that for every $k \geqslant 1$, we have $\mathbb{P}[X = k] = (1 - p)^{k-1} p$. One can also check that

$$\begin{aligned} \mathbb{P}[X \geqslant k] &= (1 - p)^{k-1} \qquad \forall k \geqslant 1 \\ \mathbb{P}[X = k \mid X \geqslant j] &= (1 - p)^{k-j} p \qquad \forall 1 \leqslant j \leqslant k\,. \end{aligned} \tag{2.2}$$

The last property expresses that a geometric random variable is "memoryless" in the sense that the distribution of $X \mid \{X > j\}$ is the same as the distribution of $X + j$.

Exponential random variables. There is also a continuous memoryless distribution: The exponential distribution. If $X \sim \mathrm{Exp}(\mu)$ is exponential with mean μ, then it has density $\frac{1}{\mu} e^{-t\mu}$, and the property that the distribution of $X \mid \{X > \lambda\}$ is the same that of $X + \lambda$.

2.2 The partitioning algorithm

Consider again a metric space (X, d) and a parameter $\Delta > 0$. Also recall our goal of producing a random partition whose sets have diameter at most Δ and such that (2.1) holds for $\alpha = 8 \ln n$. We may assume that $\Delta > 4 \ln n$, else we are done.

For simplicity, let us assume that $d(x, y) \in \{1, 2, \ldots, n\}$. This will make the analysis slightly easier without sacrificing any of the essential details. Recall also that for $x \in X$ and $r \geqslant 0$, the ball of radius r around x is defined by

$$B(x, r) = \{y \in X : d(x, y) \leqslant r\}\,.$$

Order the vertices $X = \{x_1, \ldots, x_n\}$ arbitrarily. We produce the following random partition. For each $i = 1, 2, \ldots, n$, we choose an independent random variable $R_i \sim \text{Geom}(\frac{4 \ln n}{\Delta})$, and set

$$C_i = B(x_i, R_i) \setminus \bigcup_{j<i} C_j .$$

Thus the ith set if $B(x_i, R_i)$ but we remove the points that have already been clustered. Our partition is $P = \{C_1, C_2, \ldots, C_n\}$.

Note that, as stated, the algorithm could potentially output a set of diameter bigger than Δ: There's even a chance that $R_1 > \Delta$. If $\max_i R_i > \Delta/2$, we will output the partition $P^* = \{\{x\} : x \in X\}$ into singleton clusters. That ensures that the diameter of our sets are always bounded, and we will show this eventuality happens only with very small probability. Let's use $\mathcal{E}$ to denote the event that $\max_i R_i > \Delta/2$.

Now fix $x, y \in X$. Let $\mathcal{E}_{x,y}$ be the event that x and y end up in different sets of the partition P. (We are ignoring P^* for now.) We are interested in proving an upper bound on $\mathbb{P}(\mathcal{E}_{x,y})$. In order to do this, it's helpful to think about the process as "growing" balls around $x_1, x_2, \ldots$ in order until the whole space is partitioned. The growing is because the random radii are geometrically distributed, thus we can think about each center x_i flipping coins until one comes up heads and at each step incrementing the radius by one if the coin comes up tails.

With this picture in mind, it's intuitive that we only need to start getting worried about $\mathcal{E}_{x,y}$ occurring when some ball $B(x_i, R)$ "reaches" one of x or y. Until then, there are lots of growing balls that die out before they ever see one of x or y.

Let's make this intuition precise. Let $\mathcal{Z}_i$ denote the event that $\{x, y\} \cap C_i \neq \emptyset$ and $\{x, y\} \cap C_j = \emptyset$ for $j < i$. In other words, C_i is the first set that contains one of x or y (and it possibly contains both). Then

$$\mathbb{P}(\mathcal{E}_{x,y}) = \sum_{i=1}^{n} \mathbb{P}[\mathcal{Z}_i] \cdot \mathbb{P}[\mathcal{E}_{x,y} \mid \mathcal{Z}_i] .$$

So if we can show that $\mathbb{P}[\mathcal{E}_{x,y} \mid \mathcal{Z}_i] \leqslant p$ it will imply that $\mathbb{P}[\mathcal{E}_{x,y}] \leqslant p$.

Fix i. Let us suppose that $d(x, x_i) \leqslant d(y, x_i)$ (otherwise interchange the roles of x and y). Next, note that $\mathcal{Z}_i \implies \{R_i \geqslant d(x_i, x)\}$. Thus conditioned on $\mathcal{Z}_i$, we have the following: If $R_i \in [d(x_i, x), d(x_i, y))$ then $\mathcal{E}_{x,y}$ occurs, and otherwise $R_i \geqslant d(x_i, y)$ and $\mathcal{E}_{x,y}$ does not occur.

The memoryless property of the geometric distribution means that $R_i - d(x_i, x) \mid \{R_i \geqslant d(x_i, x)\}$ again has law $\text{Geom}(\frac{4 \ln n}{\Delta})$. We conclude that

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{x,y} \mid \mathcal{Z}_i] &\leqslant \mathbb{P}[R_i \in [d(x_i, x), d(y_i, x)) \mid R_i \geqslant d(x_i, x)] \\ &= \mathbb{P}[R_i < d(y_i, x) \mid R_i \geqslant d(x_i, x)] \\ &= \mathbb{P}[X < d(y_i, x) - d(x_i, x)] , \end{aligned}$$

where $X \sim \text{Geom}(\frac{4 \ln n}{\Delta})$.

Using (2.2) and the fact that $d(y, x_i) - d(x, x_i) \leqslant d(x, y)$, we have

$$\mathbb{P}[X < d(y_i, x) - d(x_i, x)] \leqslant \mathbb{P}[X < d(x, y)] = 1 - \mathbb{P}[X \geqslant d(x, y)] = 1 - \left(1 - \frac{4 \ln n}{\Delta}\right)^{d(x,y)-1}$$

Computation yields

$$1 - \left(1 - \frac{4 \ln n}{\Delta}\right)^{d(x,y)-1} \leqslant 1 - \left(1 - \frac{4 \ln n}{\Delta}\right)^{d(x,y)} \leqslant 1 - \left(1 - d(x, y)\frac{4 \ln n}{\Delta}\right) = \frac{d(x, y)}{\Delta} 4 \ln n ,$$

where we have used the fact that $(1-\varepsilon)^k \geqslant 1 - \varepsilon k$ for all $\varepsilon \in [0,1]$ and $k \geqslant 1$.

Thus $\mathbb{P}[\mathbb{E}_{x,y}] \leqslant \frac{d(x,y)}{\Delta} 4 \ln n$. We are almost done, but recall that we sometimes output the partition P^* instead of P. Thus

$$\mathbb{P}[x \text{ and } y \text{ are separated}] \leqslant \mathbb{P}[\mathcal{E}] + \frac{d(x,y)}{\Delta} 4 \ln n \,.$$

Using a union bound along with (2.2), we have

$$\mathbb{P}[\mathcal{E}] \leqslant n \cdot \mathbb{P}[R_1 > \Delta/2] \leqslant n \cdot \left(1 - \frac{4 \ln n}{\Delta}\right)^{\Delta/2} \leqslant n \cdot e^{-2 \ln n} = \frac{1}{n} \,.$$

Here, we have used the fact that $\left(1 - \frac{1}{k}\right)^k \leqslant \frac{1}{e}$ for $k \geqslant 1$. We conclude that

$$\mathbb{P}[x \text{ and } y \text{ are separated}] \leqslant \frac{1}{n} + \frac{d(x,y)}{\Delta} 4 \ln n \leqslant \frac{d(x,y)}{\Delta} 8 \ln n \,,$$

using our assumption that $\frac{d(x,y)}{\Delta} \geqslant \frac{1}{n}$ (since $d(x,y) \in \{1, 2, \ldots, n\}$).