

Lecture 22

Prof. Nick Harvey

University of British Columbia

For many problems in computer science, there is a natural notion of “distance”. For example, perhaps the input data consists of real vectors for which it makes sense to measure their distance via the usual Euclidean distance. But in many cases, it makes sense to measure distances using a different “metric” that is not Euclidean distance.

There are many sophisticated algorithms for manipulating data involving these general metrics. One common paradigm for designing such algorithms is the familiar divide-and-conquer approach. Today we will discuss the fundamental algorithmic tool of **partitioning metric spaces** for the purpose of designing such divide-and-conquer algorithms.

This topic might seem a bit abstract and unmotivated. The next lecture will build on today’s ideas and present some algorithms whose usefulness is more easily appreciable.

1 Metrics

A **metric** is a set of points X together with a “distance function” $d : X \times X \rightarrow \mathbb{R}$ such that

- “Non-negativity”: $d(x, y) \geq 0$ for all $x, y \in X$.
- $d(x, x) = 0$ for all $x \in X$ (but we allow $d(x, y) = 0$ for $x \neq y$).
- “Symmetry”: $d(x, y) = d(y, x)$ for all $x, y \in X$.
- “The triangle inequality”: $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

In some scenarios this would be called a “semimetric” or “pseudometric”; a “metric” would additionally require that $d(x, y) = 0 \iff x = y$.

Here are some standard examples of metrics with which you should be familiar.

- *The Euclidean (or L_2) Metric*: $X = \mathbb{R}^n$ and $d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$.
- *The Manhattan (or L_1) Metric*: $X = \mathbb{R}^n$ and $d(x, y) = \sum_i |x_i - y_i|$.
- *Shortest Path Metrics*: Let $G = (V, E)$ be a graph with non-negative lengths associated with the edges. The set of points is $X = V$. The distance function is defined by letting $d(x, y)$ be the length of the shortest path between x and y .

Notice that in the first two examples the set X is infinite, but in the last example X is finite. When X is finite we call the pair (X, d) a **finite metric space**. We can also obtain finite metric spaces by restricting the first two examples to a finite subset of \mathbb{R}^n and keeping the same distance functions. In computer science we are often only interested in finite metric spaces because the input data to the problem is finite.

2 Lipschitz Random Partitions

Let (X, d) be a metric space. Let $P = \{P_1, P_2, \dots\}$ be a **partition** of X , i.e., the P_i 's are pairwise disjoint and their union is X . The P_i 's are called the **parts**. Let us define the following notation: for $x \in X$, let $P(x)$ be the unique part P_i that contains x .

The **diameter** of a part P_i is $\max\{d(x, y) : x, y \in P_i\}$. We say that the partition P is **Δ -bounded** if every P_i has diameter at most Δ .

As you will see soon, it will be very useful for us to choose a partition of (X, d) *randomly*. We say that a random partition \mathcal{P} is Δ -bounded if every possible partition that can occur as a realization of the random \mathcal{P} is a Δ -bounded partition.

Our goal is that points that are “close” to each other should have good probability of ending up in the same part. Formally, let \mathcal{P} be a randomly chosen partition. We say that \mathcal{P} is **L -Lipschitz** if

$$\Pr[\mathcal{P}(x) \neq \mathcal{P}(y)] \leq L \cdot d(x, y) \quad \forall x, y \in X.$$

So if x and y are close, they have a smaller probability of being assigned to different parts of the partition. Note that this definition is not “scale-invariant”, in the sense that we need to double L if we halve all the distances.

Combining the Δ -bounded and L -Lipschitz concepts is very interesting. Let us illustrate this with an example.

2.1 Example

Consider the “line metric”, where $X = \{1, \dots, n\}$, n is odd, and $d(x, y) = |x - y|$. The diameter of this metric is clearly $n - 1$. Consider the partition P where $P_1 = \{1, \dots, \lfloor n/2 \rfloor\}$ and $P_2 = \{\lceil n/2 \rceil, \dots, n\}$. This partition is Δ -bounded for $\Delta = \lfloor n/2 \rfloor - 1$. Does it capture our goal that “close points should end up in the same part”?

In some sense, yes. The only two consecutive points that ended up in different parts are the points $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$, so *most* pairs of consecutive points did end up in the same part. But if we modify our metric slightly, this is no longer true. Consider making n copies of both of the points $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$, keeping the copies at the same location in the metric. (This is valid because we’re really looking at semimetrics: we allow $d(x, y) = 0$ for different points x and y .) After this change, a *constant fraction* of the consecutive points now ended up in different parts!

So, even in this simple example of a line metric, if we allow multiple copies of (equivalently, non-negative weights on) each point, it becomes much less clear how to choose a partition for which most close points end up in the same part.

Choosing a *random partition* makes life much easier. Pick an index $k \in \{n/3, \dots, 2n/3\}$ uniformly at random, then set $P_1 = \{1, \dots, k\}$ and $P_2 = \{k + 1, \dots, n\}$. Let \mathcal{P} be the resulting random partition. These partitions always have diameter at most $2n/3$ (even if we made multiple copies of points). So \mathcal{P} is Δ -bounded with $\Delta = 2n/3$.

Now consider any two consecutive points i and $i + 1$. They end up in different parts of the partition only if $k = i$, which happens with probability at most $3/n$. Thus $\Pr[\mathcal{P}(i) \neq \mathcal{P}(i + 1)] \leq 3/n$. More generally

$$\Pr[\mathcal{P}(x) \neq \mathcal{P}(y)] \leq \frac{3}{n} \cdot d(x, y).$$

So \mathcal{P} is a L -Lipschitz partition with $L = 3/n$. The key point is: this holds *regardless* of how many

copies of the points we make. So this *same* random partition \mathcal{P} works under any scheme of copying (i.e., weighting) the points of this metric.

2.2 The General Theorem

The previous example achieves our “gold standard” of a random partition: $L = O(1/\Delta)$. We can think of this as meaning that the probability of adjacent points ending up in different parts is roughly the *inverse* of the diameter of those parts. Our main theorem is that, by increasing L by a logarithmic factor, we can obtain a similar partition of *any* finite metric.

Theorem 1 *Let (X, d) be a metric with $|X| = n$. For every $\Delta > 0$, there is a Δ -bounded, L -Lipschitz random partition \mathcal{P} of X with $L = O(\log(n)/\Delta)$.*

This theorem is optimal: for any n there are metrics on n points for which every Δ -bounded, L -Lipschitz partition has $L = \Omega(\log(n)/\Delta)$.

Theorem 1 is a corollary of the following more general theorem. The statement is a bit messy, but the mess will be important in proving the main result of the next lecture. Define the partial Harmonic sum $H(a, b) = \sum_{i=a+1}^b 1/i$. Let $B(x, r) = \{y \in X : d(x, y) \leq r\}$ be the ball of radius r around x .

Theorem 2 *Let (X, d) be a metric with $|X| = n$. For every $\Delta > 0$, there is Δ -bounded random partition \mathcal{P} of X with*

$$\Pr[B(x, r) \not\subseteq \mathcal{P}(x)] \leq \frac{8r}{\Delta} \cdot H(|B(x, \Delta/4 - r)|, |B(x, \Delta/2 + r)|) \quad \forall x \in X, \forall r > 0. \quad (1)$$

PROOF:(of Theorem 1) Let \mathcal{P} be the random partition from Theorem 2. Consider any $x, y \in X$ and let $r = d(x, y)$. Note that if $\mathcal{P}(x) \neq \mathcal{P}(y)$ then $B(x, r) \not\subseteq \mathcal{P}(x)$. Thus

$$\begin{aligned} \Pr[\mathcal{P}(x) \neq \mathcal{P}(y)] &\leq (8r/\Delta) \cdot H(|B(x, \Delta/4 - r)|, |B(x, \Delta/2 + r)|) \\ &\leq (8r/\Delta) \cdot H(0, n) \\ &= O(\log(n)/\Delta) \cdot d(x, y), \end{aligned}$$

since $H(0, n) = \sum_{i=1}^{n+1} 1/i = O(\log n)$. \square

2.3 Proof of Theorem 2

We start off by presenting the algorithm that generates the random partition \mathcal{P} . As is often the case with the algorithms we have seen in this class, the algorithm is very short, yet extremely clever and subtle.

- Pick $\alpha \in (1/4, 1/2]$ uniformly at random.
- Pick a bijection (i.e., ordering) $\pi : \{1, \dots, n\} \rightarrow X$ uniformly at random.
- For $i = 1, \dots, n$
 - Set $P_i = B(\pi(i), \alpha\Delta) \setminus \cup_{j=1}^{i-1} P_j$.
- Output the random partition $\mathcal{P} = \{P_1, \dots, P_n\}$.

Remark. Note that α is not an arbitrary constant; it is random. Its role is analogous to the random choice of k in our previous example.

To prove Theorem 2, we first need to check that the algorithm indeed outputs a partition. By definition, each P_i is disjoint from all earlier P_j with $j < i$, so the P_i 's are pairwise disjoint. Next, each point $\pi(i)$ is either contained in P_i or some earlier P_j , so the union of the P_i 's is X .

Next we should check that this partition is Δ -bounded. That is also easy: since $P_i \subseteq B(\pi(i), \alpha\Delta)$, the diameter of P_i is at most $\alpha\Delta < \Delta$.

The difficult step is proving (1), which we do next time. Here is some vague intuition as to why it might be true. Condition (1) asks us to show that, with good probability, the ball $B(x, r)$ is not chopped into pieces by the partition. But the parts of the partition are themselves balls of radius at least $\Delta/4$ (minus all previous parts of the partition). So as long as $r \ll \Delta/4$, we might be optimistic that the ball $B(x, r)$ does not get chopped up.

2.4 Example

Let X be the following eight points in the plane, with the usual Euclidean distance. Let π be the identity ordering: $\pi(i) = i \forall i$. The algorithm generates the following partition.

