Randomized Algorithms 2017A – Problem Set 2

Robert Krauthgamer and Moni Naor

Due: Jan. 12, 2017

1. Analyze the construction below of a stretch 3 distance oracle for a graph $G = (V, E)$ with edge weights $w : E \to \mathbb{R}_+$ (strictly speaking, it is not a distance oracle because it's query time is not fast enough.) Analyze its accuracy (stretch bound) and its storage requirement (it should almost match the one shown in class). Explain whether your bounds for accuracy and storage hold in the worst-case, in expectation, or with high probability.

   Preprocess(G): Choose $L \subset V$ as a random set of $l = \sqrt{n}\,\mathrm{polylog}\, n$ "landmark" vertices (with or without repetitions). For every vertex $v \in V$, store its distances (i) to the $\sqrt{n}$ vertices closest to it, denoted $B_v \subset V$; and (ii) to all the landmark vertices L.

   Query(u,v): If $u \in B_v$, i.e., u is among the $\sqrt{n}$ closest to v, report the distance. Otherwise, report $\min_{w \in L}[d(u, w) + d(w, v)]$.

   Hint: in the "otherwise" case, show that $L \cap B_v \neq \emptyset$.

2. Let $S \in \mathbb{R}^{k \times d}$ be a super-sparse sampling matrix, i.e., each row has a single nonzero entry of value $\sqrt{d/k}$ in a uniformly random location. Prove that for every $\lambda > 0$ and $\varepsilon \in (0, 1)$,

$$\forall y \in \mathbb{R}^d, \|y\|_\infty \leq \lambda \|y\|_2, \qquad \Pr_S \left[\|Sy\|_2^2 \notin (1 \pm \varepsilon)\|y\|_2^2 \right] \leq 2e^{-2\varepsilon^2 k/(d^2\lambda^4)}.$$

3. Given as input n points $x_1, \ldots, x_n \in [m]^d$ for $m = d = n/10$, show how to determine, within $1 + \varepsilon$ approximation, the radius of the point set under ℓ_2-distance, defined as $r = \min_{i \in [n]} \max_{j \in [n]} \|x_i - x_j\|_2$.

   You algorithm should be faster than the naive computation that runs in time $O(n^2 d)$, which in our case is $O(n^3)$.

Extra credit:

4. Show that our the JL Lemma (or just our main technical lemma) extends to a random matrix G whose entries are iid from a distribution that has mean 0, variance 1, and satisfies a sub-Gaussian tail bound $\mathbb{E}[e^{tX}] \leq e^{Ct^2}$ for some constant $C > 0$.

   Hint: Use the following trick. Introduce a standard Gaussian Z independent of X, then it is known that $\mathbb{E}[e^{tZ}] = e^{t^2/2}$, and thus

$$\forall t > 0, \qquad \mathbb{E}_X[e^{tX^2}] = \mathbb{E}_X[e^{(\sqrt{2tX^2})^2/2}] = \mathbb{E}_X \mathbb{E}_Z[e^{\sqrt{2tX^2}Z}] = \mathbb{E}_Z \mathbb{E}_X[e^{\sqrt{2tZ^2}X}] \leq \mathbb{E}_Z[e^{2CtZ^2}],$$

   and the last term evaluates, as discussed in class, to $1/\sqrt{1 - 4Ct}$.