CS 369M: Metric Embeddings and Algorithmic Applications Autumn 2018-2019

Homework 1

Due: Friday 10/19 at 11:59 PM

1. (a) Prove that every ℓ_2 metric on n points embeds isometrically into ℓ_2 with $n-1$ dimensions.

   (b) Prove that every ℓ_1 metric on n points embeds isometrically into ℓ_1 with $\binom{n}{2}$ dimensions.

   Hint: Use Carathéodory's theorem: If a point $x \neq 0$ lies in the convex cone generated by a set $P \subseteq \mathbb{R}^d$, then x can be expressed as a non-negative linear combination of at most d points in P. How can you represent an ℓ_1 metric on n points?

2. Show that every embedding of the shortest path metric of a cycle graph of length n into the line $\mathbb{R}^1$ with the metric $d(x,y) = |x-y|$ has distortion at least $\Omega(n)$.

   Hint: Consider three vertices on the cycle separated by distances $n/3$.

3. For all finite $p \geq 1$, show that the mapping given in the proof of Bourgain's theorem is an embedding into ℓ_p with distortion $O\left(\frac{\log n}{p}\right)$.

   Hint: Use Hölder's inequality.

4. Prove that the integrality gap of the LP for generalized sparsest cut is exactly equal to the worst-case distortion needed to map n-point metrics into ℓ_1.

   Hint: Use LP duality.

5. In this problem, you will show that linear projections perform very poorly for dimension reduction in ℓ_1 (which is in contrast to ℓ_2). More specifically, we will specify here an explicit set of $O(n)$ points in ℓ_1^n, and show that any linear embedding of that point set (with ℓ_1) into ℓ_1^d incurs distortion at least $\sqrt{\frac{n}{d}}$.

   The point set consists of the origin O, the n standard basis vectors P_i (where P_i has 1 in the ith coordinate and 0's elsewhere), and $m = O(n)$ points Q_i with the following property: For every pair of coordinates j_1, j_2 and pair of values in $(x_1, x_2) \in \{1, -1\}^2$, exactly $m/4$ of the points Q_i have the coordinates j_1 and j_2 set to x_1 and x_2, respectively. (Such set of points Q_i can constructed from the support of a pairwise independent distribution.)

   Without loss of generality, consider a linear map $f : \ell_1^n \to \ell_1^d$ that is non-expanding with distortion α, i.e. $\forall x, y \in \{O, P_1, \ldots, P_n, Q_1, \ldots, Q_m\}$,

   $$\frac{1}{\alpha} \|x-y\|_1 \leq \|f(x) - f(y)\|_1 \leq \|x-y\|_1 .$$

   Our goal in this problem is to show that $\alpha \geq \sqrt{\frac{n}{d}}$. W.l.o.g., f maps the origin in $\mathbb{R}^n$ to the origin in $\mathbb{R}^d$.

Let $\sigma_1, \ldots, \sigma_d : \mathbb{R}^n \to \mathbb{R}^1$ such that $f = (\sigma_1, \ldots, \sigma_d)$. Consider $\sigma \in \{\sigma_1, \ldots, \sigma_d\}$ with $\sigma(x_1, \ldots, x_n) = \sum_{j=1}^{n} \gamma_j x_j$.

(a) Prove that $|\gamma_j| \leq 1$.

(b) Prove that $\frac{1}{m} \sum_{i=1}^{m} |\sigma(Q_i)| \leq \sqrt{\sum_{j=1}^{n} \gamma_j^2} \leq \sqrt{\sum_{j=1}^{n} |\gamma_j|}$.

(c) Show that $\frac{n}{d} + \sum_{j=1}^{n} |\sigma(P_j)| \geq 2\sqrt{\frac{n}{d}} \cdot \frac{1}{m} \sum_{i=1}^{m} |\sigma(Q_i)|$,

and conclude that $n + \sum_{j=1}^{n} \|f(P_j)\|_1 \geq 2\sqrt{\frac{n}{d}} \cdot \frac{1}{m} \sum_{i=1}^{m} \|f(Q_i)\|_1$.

(d) Show that the distortion $\alpha \geq \sqrt{\frac{n}{d}}$.