Problem Set 4 CS265, Fall 2020

Due: October 16 (Friday) at 23:59 (Pacific Time)

Please follow the homework policies on the course website.

---

1. **(4 pt.)** Prove that $(\mathbb{R}^3, \ell_2)$ cannot be embedded into $(\mathbb{R}^2, \ell_2)$ with bounded distortion. In other words, there are no functions $f : \mathbb{R}^3 \to \mathbb{R}^2$ and constants $\alpha, \beta > 0$ such that the following inequality holds for all $x, y \in \mathbb{R}^3$:

$$\beta\|x - y\|_2 \leq \|f(x) - f(y)\|_2 \leq \alpha\beta\|x - y\|_2.$$

[**HINT:** *Try a proof by contradiction. How should the grid* $G_n := \{(i, j, k) : i, j, k \in \{0, 1, \ldots, n\}\}$ *be embedded?*]
[**HINT:** *A disc of radius* r *has area* πr^2*.*]

2. **(10 pt.)** The *equilateral dimension* of a metric space is the maximum number of points in the space that are all at the same distance from each other. In this problem, we will determine the equilateral dimension of the d-dimensional Euclidean space $\mathcal{E}^d = (\mathbb{R}^d, \ell_2)$, and also explore a relaxed version of the equilateral dimension where the pairwise distances are only required to be *approximately* the same.

   (a) **(2 pt.)** Let X be a set of $d+1$ points in $\mathcal{E}^m$ for some arbitrary positive integer m. Show that (X, ℓ_2) can be isometrically embedded into $\mathcal{E}^d$.
   [**HINT:** *Suppose* $X = \{x_0, \ldots, x_d\}$*. Consider the linear subspace spanned by the vectors* $x_i - x_0$ *for* $i = 1, \ldots, d$*. How large can the dimension of the subspace be?*]

   (b) **(2 pt.)** For all positive integers d, use Part (a) to show that there exist $d + 1$ points in $\mathcal{E}^d$ that are all at distance 1 from each other. This shows that the equilateral dimension of $\mathcal{E}^d$ is at least $d + 1$.

   (c) **(3 pt.)** Show that there exists a constant $c > 0$ such that for all positive integers d, one can find at least 2^{cd} points in $\mathcal{E}^d$ that are all at distances between 1 and 1.1 from each other.
   [**HINT:** *Try applying the JohnsonLindenstrauss lemma.*]

   (d) **(0 pt.) [Optional: this won't be graded.]** For all positive integers d, show that one cannot find $d + 2$ points in $\mathcal{E}^d$ that are all at distance 1 from each other. Together with Part (b), this shows that the equilateral dimension of $\mathcal{E}^d$ is exactly $d + 1$.
   [**HINT:** *Prove by contradiction. Suppose there are* $d + 2$ *such points* $x_0, x_1, \ldots, x_{d+1}$*. Consider the vectors* $v_i = x_i - x_0$ *and the matrix* $A = [v_1, \cdots, v_{d+1}]$ *(with the* v_i *as columns) of size* $d \times (d + 1)$*. Explicitly compute the inner products* $v_i^\mathsf{T} v_j$ *and the matrix* $A^\mathsf{T} A$*. Use the fact that* $\operatorname{rank}(A^\mathsf{T} A) = \operatorname{rank}(A)$ *to derive a contradiction.*[1]]

   (e) **(3 pt.)** Show that there exists a constant $C > 0$ such that for all positive integers d, one cannot find more than 2^{Cd} points in $\mathcal{E}^d$ that are all at distances between 1 and 1.1 from each other.
   [**HINT:** *Try using a similar strategy to problem 1. You can use the fact that the volume*

[1] This proof strategy can be used to show that the Johnson-Lindenstrauss lemma is "nearly tight". See Problem 4.

of a d-dimensional ball of radius r scales as γr^d for some constant γ that depends on d but not r. It may be helpful if you understand the proof of Lemma 3 in lecture notes #9, but that is not required for working on this problem.]

3. **(4 pt.)** We showed that Bourgain's embedding allows us to embed an arbitrary metric space (X, d) with $|X| = n$ into $(\mathbb{R}^k, \ell_1)$ with target dimension k being $O((\log n)^2)$ and distortion being $O(\log n)$. Moreover, the embedding can be computed efficiently using a randomized algorithm. Prove that the exact same embedding computed by the randomized algorithm also achieves $O(\log n)$ distortion with high probability when the target metric is ℓ_p for $p > 1$. We encourage you to emphasize only the differences from the proof in the lecture notes rather than copying the entire proof.
[**HINT:** *Let $f : X \to \mathbb{R}^k$ denote the relevant embedding. For any two points $x, y \in X$, we showed that $\|f(x) - f(y)\|_1 \leq k \cdot d(x, y)$. Can we say something similar about $\|f(x) - f(y)\|_p$?*]
[**HINT:** *For any two points $a, b \in \mathbb{R}^k$ and $p > 1$, it holds that $\|a - b\|_p \geq k^{(1/p)-1}\|a - b\|_1$. This is a special case of Hölder's inequality.*]

4. **(0 pt.)** **[Optional: this won't be graded.]** Prove that the Johnson-Lindenstrauss lemma is "nearly tight".[2] Specifically, for a sufficiently small positive constant c, show that for every positive integer $n > 100$ and any $\epsilon \in [1/\sqrt{n}, 1/10)$, one can find a positive integer d together with a set X of n points in $\mathbb{R}^d$ such that for any positive integer $m < \frac{c \ln n}{\epsilon^2 \ln(1/\epsilon)}$, there is no way to embed (X, ℓ_2) into $(\mathbb{R}^m, \ell_2)$ with distortion $1 + \epsilon$. You can use the following theorem in linear algebra without proof:

   Theorem 1 (Alon '03)**.** *There is a small positive constant c with the following property. Let $A = (a_{ij})$ be an n by n real matrix with $a_{ii} = 1$ for all i and $|a_{ij}| \leq \epsilon$ for all $i \neq j$. If $n > 4$ and $\epsilon \in [1/\sqrt{n}, 1/2)$, then*
   $$\operatorname{rank}(A) \geq \frac{c \ln n}{\epsilon^2 \ln(1/\epsilon)}.$$

   [**HINT:** *Try using a similar proof strategy as Problem 2(d).*]

[2] For improved results, see paper by Larsen and Nelson: Optimality of the Johnson-Lindenstrauss Lemma. (Link to the FOCS 2017 version: `https://ieeexplore.ieee.org/abstract/document/8104096`).