Problem Set 4 CS265, Autumn 2022

Due: Friday 10/28 at 11:59pm on Gradescope

Group members: Rylan Schaeffer

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* $G_n := \{(i, j, k) : i, j, k \in \{0, 1, \dots, n\}\}$ *be embedded?*]
[**HINT:** *A disc of radius* r *has area* πr^2.]

SOLUTION: Consider the n^3 points on the grid in $\mathbb{R}^3$. For each pair of points x, y, note that $||x - y||_2 \geq 1$. For purposes of contradiction, assume the distortion constraints hold for $f(\cdot)$. From the lower distortion constraint, we know that:

$$||f(x) - f(y)||_2 \geq \beta||x - y||_2 \geq \beta > 0$$

since the minimum distance is 1. From this constraint, we can think of each grid point as mapped to $\mathbb{R}^2$, surrounded by a disc with radius $\geq \beta/2$. The area covered by one of these discs is thus $\geq \pi(\beta/2)^2$, and with n^3 disjoint discs, the total area $A(n)$ covered must be:

$$A(n) \geq n^3\pi(\beta/2)^2$$

Since the most efficient packing of disks is in a close-packed structure (see Wikipedia's Sphere Packing page), a lower bound on the radius $R(n)$ of the circle that encompasses the necessary area is at least:

$$R(n) \geq \frac{n^{3/2}\beta}{2}$$

This means that the diameter $D(n)$ is at least:

$$D(n) \geq n^{3/2}\beta$$

This means that there are two points a, b on the grid that have distance at least $n^{3/2}\beta$:

$$n^{3/2}\beta \leq ||f(a) - f(b)||_2$$

Note that in our grid in $\mathbb{R}^3$, the maximum possible distance is $\leq (n^2 + n^2 + n^2)^{1/2} = \sqrt{3}n$ from $(0, 0, 0)$ to (n, n, n). Under the upper bound distortion constraint, this means that $\forall x, y$:

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

This violates the upper bound distortion constraint because there is no constant α such that $\forall \alpha, x, y$:

$$n^{3/2}\beta \leq ||f(x) - f(y)||_2 \leq \alpha\beta\sqrt{3}n$$

Thus, we have our contradiction: it's impossible to satisfy both the upper and lower distortion bounds simultaneously.

2. **(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.*]

SOLUTION: We want to show both a lower bound and an upper bound on the distortion. The upper bound proof is similar to the proof in lecture, since $|d(x, S_{ij}) - d(y, S_{ij})| \leq d(x, y)$ still holds as it depends on the original metric d, not ℓ_p. Specifically:

$$|d(x, S_{ij}) - d(y, S_{ij})| \leq d(x, y) \Rightarrow |d(x, S_{ij}) - d(y, S_{ij})|^P \leq d(x, y)^P$$

And then in ℓ_p:

$$\begin{aligned}
||f(x) - f(y)||_p &= \sqrt[p]{\sum_{ij} |d(x, S_{ij}) - d(y, S_{ij})|^p} \\
&\leq \sqrt[p]{\sum_{ij} d(x, y)^p} \\
&= \sqrt[p]{k\, d(x, y)^p} \\
&= \sqrt[p]{k}\, d(x, y)
\end{aligned}$$

This establishes the upper bound. To establish the lower bound, recall that we showed in lecture that:

$$||f(x) - f(y)||_1 \geq \frac{k}{b \log n} d(x, y)$$

Using Hint 2, we have:

$$||f(x) - f(y)||_p \geq k^{1/p-1}||f(x) - f(y)||_1 \geq k^{1/p-1}\frac{k}{b \log n} d(x, y) = \frac{\sqrt[p]{k}}{b \log n} d(x, y)$$

The two bounds prove the claim:

$$\frac{\sqrt[p]{k}}{b \log n} d(x,y) \le ||f(x) - f(y)||_p \le \sqrt[p]{k}\, d(x,y)$$

where $\alpha = \sqrt[p]{k}$ is the scaling and $O(\log n)$ is the distortion.

3. **(11 pt.) Johnson-Lindenstrauss with ± 1 entries:** In the lecture notes and videos we showed that a matrix of standard Gaussians can be used to get a dimension reducing map with very little distortion. However, a matrix of arbitrary real numbers can be cumbersome to store and compute with. In this problem you'll show that you can get essentially the same guarantees using random matrices with ± 1 entries. Throughout this problem, let A be an $m \times d$ matrix who's entries are independently set to $+1$ with probability $1/2$ and otherwise to -1, and $z \in \mathrm{R}^d$ be an arbitrary unit vector.[1]

   In this problem, you can use the statements from previous subparts even if you do not successfully prove them.

   (a) **(2 pt.)** Show that $\mathbb{E}[\|Az\|_2^2] = m$.

   (b) **(2 pt.)** For $Y \sim N(0,1)$, show that for any even $k \ge 0$, $\mathbb{E}[Y^k] \ge 1$, and for odd $k \ge 0$, $\mathbb{E}[Y^k] = 0$.
   [**HINT:** *There are many solutions to this. Try to find a short one!*]

   (c) **(2 pt.)** Prove that for any independent $X_1, \ldots, X_n$ and independent $Y_1, \ldots, Y_n$, if, for all integers $k \ge 0$ and $i = 1, \ldots, n$,

   $$0 \le \mathbb{E}[(X_i)^k] \le \mathbb{E}[(Y_i)^k]$$

   then for all integers $p \ge 0$,

   $$\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^p\right] \le \mathbb{E}\left[\left(\sum_{i=1}^n Y_i\right)^p\right]$$

   (d) **(4 pt.)** Let B be an $m \times d$ matrix who entries are independently drawn from $N(0,1)$. Prove that, for any $t \ge 0$ and unit vector z, if $\mathbb{E}[e^{t\|Bz\|_2^2}]$ is finite[2], then

   $$\mathbb{E}[e^{t\|Az\|_2^2}] \le \mathbb{E}[e^{t\|Bz\|_2^2}]$$

   [**HINT:** *For any random variable X, $\mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbb{E}[X^k]$*]

   (e) **(1 pt.)** Show that, for any $\epsilon \in (0,1]$,

   $$\Pr[\|Az\|_2^2 \ge m(1+\epsilon)] \le e^{-\Omega(m\epsilon^2)}.$$

[1] You may wonder why the proof from the lecture notes doesn't directly apply to ± 1 entries. This is because, when the entries are drawn from a normal distribution, we can use the rotational invariance of Gaussians to rotate z until it is a standard unit vector. That trick no longer applies if the entries are ± 1.

[2] For the purpose of your solutions, feel free to ignore this "is finite."

If your proof is similar to that of Theorem 1 in lecture notes 8, we encourage you to emphasize only the differences from the proof in the lecture notes rather than copying the entire proof.

(f) **(0 pt.) [Optional: this won't be graded.]** Show that, for any $\epsilon \in (0, 1]$,

$$\Pr[\|Az\|_2^2 \le m(1-\epsilon)] \le e^{-\Omega(m\epsilon^2)}.$$

[**HINT:** *We recommend you first show that for any* independent *and* nonnegative *random variables* $X_1, \ldots, X_m$*, defining* $S = \sum_{i=1}^m X_i$*, the probability* $S \le \mathbb{E}[S] - \Delta$ *is at most* $\exp(-\Omega(\Delta^2 / \sum_{i=1}^m \mathbb{E}[X_i^2]))$*. To do so, use the inequality* $e^{-v} \le 1 - v + v^2/2$ *which holds for any* $v \ge 0$*. Feel free to use the fact that for* $Y \sim N(0, 1)$*,* $\mathbb{E}[Y^4] = 3$*.*]

SOLUTION:

(a) First, note that the first moment of any element of A_{ij} is:

$$\mathbb{E}[A_{ij}] = \frac{1}{2}(-1) + \frac{1}{2}(1) = 0$$

and the second moment is:

$$\mathbb{E}[A_{ij}^2] = \frac{1}{2}(-1)^2 + \frac{1}{2}(1)^2 = 1$$

Using the trace, and linearity of expectation:

$$\mathbb{E}[||Az||_2^2] = \mathbb{E}[Tr[z^T A^T Az]] = Tr[z^T \mathbb{E}[A^T A]z] = Tr[z^T \text{diag}\big(\sum_i^m \mathbb{E}[A_{1i}^2], ..., \sum_i^m \mathbb{E}[A_{di}^2]\big)z]$$

Substituting the second moment and using that $z^T z = 1$:

$$\mathbb{E}[||Az||_2^2] = Tr[z^T \text{diag}(m, ..., m)z] = mTr[z^T z] = m$$

(b) For odd $k \ge 0$: Note that Y is symmetric about 0, meaning $\mathbb{P}[Y = y] = \mathbb{P}[Y = -y]$. Consequently, for any odd k, by symmetry, $\mathbb{E}[Y^k] = \mathbb{E}[(-Y)^k] = -\mathbb{E}[Y^k]$. The only value equal to its own negative is 0, meaning $\mathbb{E}[Y^k] = 0$ for odd $k \ge 0$.

For even $k \ge 0$: We can use moment generating functions. Specifically, for the standard normal, the MGF is:

$$M_Y(t) = e^{t^2/2}$$

And recall that $\mathbb{E}[Y^k] = \frac{d^k}{dt^k} M_Y(t = 0)$. Computing the 0th derivative:

$$\mathbb{E}[Y^0] = 1$$

Computing the 2nd derivative:

$$\mathbb{E}[Y^2] = e^{0^2/2} + (0)^2 e^{t^2/2} = 1$$

For every even k, taking the derivative and applying the chain rule will annihilate some terms with t prefactors, meaning they'll survive when we evaluate at $t = 0$. For instance, $k = 4$:

$$\frac{d^4}{dt^4}M_y(t=0) = e^{t^2/2} + te^{t^2/2} + 2e^{t^2/2} + 2t^2e^{t^2/2} + 3t^2e^{t^2/2} + t^4e^{t^2/2} = 1 + 2 = 3 \geq 1$$

Since nothing will ever produce a negative prefactor and since we are only adding additional terms, and since the first term ($k = 0$) is 1, the sum must be ≥ 1. More specifically, the general rule for even $k \geq 0$ is $\mathbb{E}[Y^k] = k!! \geq 1$.

(c) Suppose $X_1, ..., X_n$ and $Y_i, ..., Y_n$ are independent and suppose that $\forall k \geq 0$ and $i = 1, ..., n$:

$$0 \leq \mathbb{E}[(X_i)^k] \leq \mathbb{E}[(Y_i)^k]$$

Then by the Multinomial theorem, independence and linearity of expectation:

$$\begin{aligned}
\mathbb{E}[(\sum_i X_i)^p] &= \mathbb{E}\Big[\sum_{k_1+...+k_n=p} \binom{p}{k_1, k_2, ..., k_n} \prod_i X_i^{k_i}\Big] \\
&= \sum_{k_1+...+k_n=p} \binom{p}{k_1, k_2, ..., k_n} \prod_i \mathbb{E}[X_i^{k_i}] \\
&\leq \sum_{k_1+...+k_n=p} \binom{p}{k_1, k_2, ..., k_n} \prod_i \mathbb{E}[Y_i^{k_i}] \\
&= \mathbb{E}[(\sum_i Y_i)^p]
\end{aligned}$$

(d) Let A be an $m \times d$ matrix with entries drawn i.i.d. from $\{-1, +1\}$ with probability 1/2 each, and B be an $m \times d$ matrix with entries drawn i.i.d. from $N(0, 1)$. Let z be an arbitrary unit vector and assume: $t \geq 0$ and $\mathbb{E}[\exp(t||Bz||_2^2)] < \infty$. Our goal is to show:

$$\mathbb{E}[e^{t||Az||_2^2}] \leq \mathbb{E}[e^{t||Bz||_2^2}]$$

We want to show the above, and we know from the Taylor series expansion of e that the above will hold if we can show that $\forall k \geq 0$,

$$\mathbb{E}[(||Az||_2^2)^k] \leq \mathbb{E}[(||Bz||_2^2)^k]$$

We can write $||Az||_2^2 = \sum_i^m (A_i z)^2$. From 3c, we know that the above statement will hold if we can show that $\forall k' \geq 0$:

$$0 \leq \mathbb{E}[((A_i z)^2)^{k'}] \leq \mathbb{E}[((B_i z)^2)^{k'}]$$

Note that $A_i z = \sum_j^d A_{ij} z_j$. Re-applying 3c, the above will hold if $\forall k'' \geq 0$:

$$\mathbb{E}[(A_{ij}z_j)^{k''}] \leq \mathbb{E}[(B_{ij}z_j)^{k''}]$$

To show that this holds, we can use our result from 3b. Specifically, we know that for odd k, $\mathbb{E}[A_{ij}^k] = 0$ (by symmetry), and that for even k, $\mathbb{E}[A_{ij}^k] = 1$ (because for even k, $(-1)^k = (1)^k = 1$), and that for odd k, $\mathbb{E}[B_{ij}^k] = 0$, and for even k, $\mathbb{E}[B_{ij}^k] \geq 1$. Thus, for all $k \geq 0$ and $\forall i, j$:

$$\mathbb{E}[A_{ij}^k] \leq \mathbb{E}[B_{ij}^k]$$

Note this also implies that:

$$\mathbb{E}[A_{ij}z_j] \leq \mathbb{E}[B_{ij}z_j]$$

since z_j is some constant, which also implies that:

$$\mathbb{E}[(A_{ij}z_j)^{k''}] \leq \mathbb{E}[(B_{ij}z_j)^{k''}]$$

"Recursing" back up our chain of reasoning, the goal statement holds.

(e) Per Chernoff:

$$\mathbb{P}[||Az||_2^2 \geq m(1+\epsilon)] \leq \mathbb{P}\Big[\Big|||Az||_2^2 - m\Big| \geq m\epsilon] \leq 2e^{-m\epsilon^2/3} = e^{\Omega(-m\epsilon^2)}$$