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? Try to pin down the intuition that the embedding of the grid would need to have lots of points fairly close together—within a smallish circle—but each point should not be too close to any other point, and then derive a contradiction from the fact that there just isn't enough area to fit all those points without some being too close....*]

SOLUTION: Suppose that there exists an embedding f such that

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

Note that for any distinct $x, y \in G_n$, we have

$$1 \leq \|x-y\|_2 \leq n\sqrt{3}$$

and therefore we have

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

This implies that we can draw a disc of radius $\frac{\beta}{2}$ around each $f(x)$ without any overlap. In addition, because for all $x \in G_n$ we have

$$\|f((0,0,0))-f(x)\|_2 \leq \alpha\beta n\sqrt{3}$$

all these disks are contained in the disc of radius $\alpha\beta n\sqrt{3} + \frac{\beta}{2}$ centered at $f((0,0,0))$. This large disc has area

$$\pi(\alpha\beta n\sqrt{3} + \frac{\beta}{2})^2$$

There are $(n+1)^3$ non-overlapping small discs, each with area $\pi(\beta/2)^2$, contained in the large disc. Therefore, we must have

$$(n+1)^3\pi(\beta/2)^2 \leq \pi(\alpha\beta n\sqrt{3} + \frac{\beta}{2})^2$$

However, since the left hand is $\Theta(n^3)$ and the right side is $\Theta(n^2)$, this does not hold for large n, and so such an embedding does not exist.

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_2. [This actually holds for any ℓ_p metric for any $p \geq 1$, but this problem just asks you to prove it for ℓ_2]. 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 $\|f(x) - f(y)\|_2$?*]
[**HINT:** *For any two points $a, b \in \mathbb{R}^k$ it holds that $\|a - b\|_2 \geq \frac{1}{\sqrt{k}}\|a - b\|_1$. This is a special case of Hölder's inequality.*]

SOLUTION: We showed in the lecture notes that

$$|d(x, S_{i,j}) - d(y, S_{i,j})| \leq d(x, y)$$

Plugging this into the ℓ_p norm gives

$$\begin{aligned}\|f(x) - f(y)\|_p =& (\sum_{i,j}(d(x, S_{i,j}) - d(y, S_{i,j}))^p)^{\frac{1}{p}} \\ \leq& (\sum_{i,j}(d(x, y))^p)^{\frac{1}{p}} \\ =& k^{\frac{1}{p}} d(x, y)\end{aligned}$$

We now prove that with high probability,

$$\|f(x) - f(y)\|_p \geq \frac{k^{\frac{1}{p}}}{b \cdot \log n} d(x, y)$$

We construct the sets $S_{i,j}$ and choose c in the same way, so that

$$\|f(x) - f(y)\|_1 \geq \frac{k}{2^6 \log n} d(x, y)/3$$

Then, we have

$$\begin{aligned}\|f(x) - f(y)\|_p \geq& k^{\frac{1}{p}-1}\|f(x) - f(y)\|_1 \\ \geq& k^{\frac{1}{p}-1}\frac{k}{2^6 \log n} d(x, y)/3 \\ =& \frac{k^{\frac{1}{p}}}{3 \cdot 2^6 \log n} d(x, y)\end{aligned}$$

which completes the proof.

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 $k \geq 0$, $\mathbb{E}[Y^k] \geq 1$, and for odd $k \geq 0$, $\mathbb{E}[Y^k] = 0$.
[**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 $k \geq 0$ and $i = 1, \ldots, n$,

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

then for all integers $p \geq 0$,

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

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

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

[**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 $\epsilon \in (0,1]$,

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

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 $\epsilon \in (0,1]$,

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

[**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 $S \leq \mathbb{E}[S] - \Delta$ is at most $\exp(-\Omega(\Delta^2/\sum_{i=1}^{m}\mathbb{E}[X_i^2]))$. To do so, use the inequality $e^{-v} \leq 1 - v + v^2/2$ which holds for any $v \geq 0$. Feel free to use the fact that for $Y \sim N(0,1)$, $\mathbb{E}[Y^4] = 3$.*]

[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."

SOLUTION:

(a) By linearity of expectation,

$$\mathbb{E}[\|Az\|_2^2] = \mathbb{E}\left[\sum_{i=1}^{m}((Az)_i)^2\right] = \sum_{i=1}^{m}\mathbb{E}\left[((Az)_i)^2\right]$$

Since $(Az)_i$ for $i = 1, \ldots, m$ all have the same distribution, it is sufficient for us to show that $\mathbb{E}[((Az)_1)^2] = 1$. The distribution of $(Az)_1$ is just that of $\sigma \cdot z$ where $\sigma \in \{\pm 1\}^d$ has every element chosen uniformly and independently from $\{\pm 1\}$. We compute,

$$\begin{aligned}
\mathbb{E}[((Az)_1)^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{d}\sigma_i z_i\right)^2\right] \\
&= \sum_{i=1}^{d}\mathbb{E}\left[\sigma_j^2 z_i^2\right] + \sum_{i\neq j}^{d}\mathbb{E}\left[\sigma_i\sigma_j z_i z_j\right] \\
&= \sum_{i=1}^{d} z_i^2 = 1
\end{aligned}$$

where $\mathbb{E}[\sigma_i\sigma_j] = \mathbb{E}[\sigma_i]\mathbb{E}[\sigma_j] = 0$ is by independence of σ_i, σ_j.

(b) For odd k, since Y^k is symmetric, as long as $\mathbb{E}[Y^k]$ is finite, it must be zero. This is because, for $f(t)$ the pdf of Y, the integrals $\int_0^\infty f(t)t^k dt$ and $\int_{-\infty}^0 f(t')(t')^k dt'$ cancel out whenever $t' = -t$. For a normal distribution, the tails decay proportional to $t^k \cdot e^{-t^2/2} = e^{-\Omega(t^2)}$, which is fast enough for those integrals to converge, and so $\mathbb{E}[Y^k] = 0$.

Now for the even case. For $k = 0$, for any random variable Y, $\mathbb{E}[Y^0] = 1$. For $k = 2$, since $N(0,1)$ has mean 0 and variance 1, $\mathbb{E}[Y^2] = 1$. For even $k \geq 2$, we apply Jensen's inequality,

$$\mathbb{E}[Y^k] = \mathbb{E}[(Y^2)^{k/2}] \geq \mathbb{E}[(Y^2)]^{k/2} = 1^{k/2} = 1.$$

Another approach is to use the fact (which we proved in lecture) that the moment generating function of a standard Gaussian is $M(t) = e^{\frac{1}{2}t^2}$. This means that

$$\sum_{k=0}^{\infty}\frac{t^k}{k!}\mathbb{E}[Y^k] = \mathbb{E}[e^{tY}] = e^{\frac{1}{2}t^2} = \sum_{n=0}^{\infty}\frac{t^{2n}}{2^n n!}$$

where the last equality uses the Taylor series $e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}$ substituting $x = t^2/2$. Comparing terms in the leftmost and rightmost sums, we see that $\mathbb{E}[Y^k] = 0$ if k is odd and $\mathbb{E}[Y^k] = \frac{k!}{2^{k/2}(k/2)!} = (k-1)(k-3)\cdots 3 \cdot 1 \geq 1$ when k is even.

(c) By linearity of expectation, it is sufficient to prove that

$$\sum_{i_1,\ldots,i_p=1}^{n}\mathbb{E}\left[X_{i_1}\cdots X_{i_p}\right] \leq \sum_{i_1,\ldots,i_p=1}^{n}\mathbb{E}\left[Y_{i_1}\cdots Y_{i_p}\right].$$

For any fixed setting of $i_1, \ldots, i_p$, using independence of $X_1, \ldots, X_n$, there are integers $k_1, \ldots, k_n \geq 0$ such that

$$\mathbb{E}\left[X_{i_1} \cdots X_{i_p}\right] = \prod_{j=1}^{n} \mathbb{E}[X_j^{k_j}],$$

namely, k_j counts the number of $i_1, \ldots, i_p$ that are equal to j. The same decomposition holds for the Y variables. Using the assumptions of the problem, $\prod_{j=1}^{n} \mathbb{E}[X_j^{k_j}] \leq \prod_{j=1}^{n} \mathbb{E}[Y_j^{k_j}]$, and so the desired inequality holds.

(d) We can decompose $\|Az\|_2^2 = \sum_{i\in[m]} (Az)_i^2$. Each term of that sum is independent and identically distributed according to $\sigma \cdot z$ where $\sigma \in \{\pm 1\}^d$ has every element chosen uniformly and independently from $\{\pm 1\}$. Therefore,

$$\mathbb{E}[e^{t\|Az\|_2^2}] = \mathbb{E}\left[e^{t(\sigma\cdot z)^2}\right]^m.$$

We expand that moment generating function using the Taylor series for e^x and linearity of expectation to obtain,

$$\mathbb{E}[e^{t\|Az\|_2^2}] = \left(\sum_{p=0}^{\infty} \frac{t^p \cdot \mathbb{E}[(\sigma \cdot z)^{2p}]}{p!}\right)^m$$

The same decomposition holds for B, where $\alpha \sim N(0,1)^d$,

$$\mathbb{E}[e^{t\|Bz\|_2^2}] = \left(\sum_{p=0}^{\infty} \frac{t^p \cdot \mathbb{E}[(\alpha \cdot z)^{2p}]}{p!}\right)^m$$

In both equations, the value inside the $(\cdot)^m$ is nonnegative, so it is sufficient to prove that each element of the sum satisfies the desired inequality. That is, we wish to prove $\mathbb{E}[(\sigma \cdot z)^{2p}] \leq \mathbb{E}[(\alpha \cdot z)^{2p}]$ for integers $p \geq 0$. To do so, we apply part (c). For each $i \in [d]$, let $X_i = \sigma_i z_i$ and $Y_i = \alpha_i z_i$. Then, for any even k

$$\mathbb{E}[X_i^k] = \mathbb{E}[\sigma_i^k] z_i^k = z_i^k.$$

By part (b), $\mathbb{E}[Y_i^k] \geq z_i^k$ for even k. For odd k both $\mathbb{E}[X_i^k] = \mathbb{E}[Y_i^k] = 0$. The desired result follows from part (c).

(e) We bound

$$\begin{aligned}
\Pr[\|Az\|_2^2 \geq m(1+\epsilon)] &\leq \Pr\left[e^{t\|Az\|_2^2} \geq e^{tm(1+\epsilon)}\right] \\
&\leq \frac{\mathbb{E}\left[e^{t\|Az\|_2^2}\right]}{e^{tm(1+\epsilon)}} && \text{(Markov's inequality)} \\
&\leq \frac{\mathbb{E}\left[e^{t\|Bz\|_2^2}\right]}{e^{tm(1+\epsilon)}}. && \text{(part (d))}
\end{aligned}$$

We continue as we did in the lecture notes, setting $t = \epsilon/4$.

(f) First, we prove the inequality in the hint. Let $X_1, \dots, X_m$ be independent nonnegative random variables and S their sum. For any $t \geq 0$,

$$\begin{aligned}
\mathbb{E}\left[e^{-tS}\right] &= \prod_{i\in[m]} \mathbb{E}\left[e^{-tX_i}\right] \\
&\leq \prod_{i\in[m]} \left(1 - t\mathbb{E}[X_i] + \frac{t^2}{2}\mathbb{E}[X_i^2]\right) && (e^{-v} \leq 1 - v + v^2/2 \text{ for } v \geq 0) \\
&\leq \prod_{i\in[m]} e^{\left(-t\mathbb{E}[X_i] + \frac{t^2}{2}\mathbb{E}[X_i^2]\right)} && (1 + v \leq e^v \text{ for all } v \in \mathbb{R}) \\
&= \exp\left(-t\mathbb{E}[S] + \frac{t^2}{2}\sum_{i\in[m]} \mathbb{E}[X_i^2]\right).
\end{aligned}$$

Next, we apply Markov's inequality. For $t \geq 0$

$$\begin{aligned}
\Pr[S \leq \mathbb{E}[S] - \Delta] &= \Pr\left[e^{-tS} \geq e^{-t(\mathbb{E}[S]-\Delta)}\right] \\
&\leq \frac{\mathbb{E}\left[e^{-tS}\right]}{e^{-t(\mathbb{E}[S]-\Delta)}} \\
&\leq \exp\left(-t\mathbb{E}[S] + \frac{t^2}{2}\sum_{i\in[m]} \mathbb{E}[X_i^2] + t(\mathbb{E}[S] - \Delta)\right) \\
&= \exp\left(\frac{t^2}{2}\sum_{i\in[m]} \mathbb{E}[X_i^2] - t\Delta\right)
\end{aligned}$$

Setting $t = \frac{\Delta}{\sum_{i\in[m]} \mathbb{E}[X_i^2]}$ gives

$$\Pr[S \leq \mathbb{E}[S] - \Delta] \leq \exp\left(-\frac{\Delta^2}{2\sum_{i\in[m]} \mathbb{E}[X_i^2]}\right).$$

Next, we will apply that inequality to prove the desired result. Let $X_1, \dots, X_m$ be the elements of Az, so that $\|Az\|_2^2 = \sum_{i\in[m]} X_i^2$. In part (a), we proved that $\mathbb{E}[\|Az\|_2^2] = m$. Applying in the inequality we just proved,

$$\Pr[[\|Az\|_2^2] \leq m - m\epsilon] \leq \exp\left(-\frac{\epsilon^2 m^2}{2\sum_{i\in[m]} \mathbb{E}[X_i^4]}\right).$$

Each X_i is identically distributed, so it's enough to prove that $\mathbb{E}[X_i^4] = O(1)$. Let $Y_1, \dots, Y_m$ be the elements of Bz, where B is defined as in part (d). In the lecture notes, we proved that Y_i is distributed as a normal with mean 0 and variance 1, which means that $\mathbb{E}[Y_i^4] = 3$. By the same argument made in part(d), $\mathbb{E}[X_i^4] \le \mathbb{E}[Y_i^4] = 3$. We conclude that

$$\Pr[[\|Az\|_2^2] \le m(1-\epsilon)] \le \exp\left(-\frac{\epsilon^2 m^2}{6m}\right) = e^{-\frac{m\epsilon^2}{6}}.$$