In all problems of this homework, and also in the future homework sets, $C_1, C_2, \ldots$ denote positive absolute constants of your choice. Thus, whenever you see $C_1$, you can replace it any positive constant you like, for example 10 or 100. Usually, you will find it easier to choose big values, so $C_1 = 100$ might be a good choice. Remember that these constants may not depend on anything, so $C_2 = \sqrt{n}$ is not a valid choice, for example.

The first problem is about an arbitrary set of $n$ unit vectors in $\mathbb{R}^n$, i.e. vectors $x_1, \ldots, x_n$ satisfying $\|x_i\|_2 = 1$ for all $i$. Their sum $x_1 + \cdots + x_n$ has norm at most $n$, due to the triangle inequality. The bound $n$ is obviously optimal, and is attained if all vectors $v_i$ are the same. However, the sum can be made much smaller by carefully selecting the signs for the vectors $x_i$.

Problem 1 (Balancing vectors)

Let $x_1, \ldots, x_n$ be an arbitrary set of unit vectors in $\mathbb{R}^n$. Prove that there exist $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$ such that

$$\|\varepsilon_1 x_1 + \cdots + \varepsilon_n x_n\|_2 \leq \sqrt{n}.$$

Hint: make use of a probabilistic method. Choose $\varepsilon_i$ at random and independently, and compute $\mathbb{E}\|\varepsilon_1 x_1 + \cdots + \varepsilon_n x_n\|_2^2$. The result in Problem 5 of HW1 can help.

The rest of the problems are about a standard normal random vector $X$, i.e. a vector $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ whose coordinates $X_i$ are independent standard normal random variables $N(0, 1)$.

In the first problem, we show that the Euclidean norm of $X$ is tightly concentrated about $\sqrt{n}$, namely that $\|X\|_2 = \sqrt{n} + O(1)$ with high probability. Such tight concentration is surprising: the error $O(1)$ is tiny compared to the leading term $\sqrt{n}$.

Problem 2 (Random vectors have norm $\approx \sqrt{n}$)

Let $X$ be a standard normal random vector in $\mathbb{R}^n$, where $n \geq C_1$.

(a) Check that

$$\mathbb{E}\|X\|_2^2 = n \quad \text{and} \quad \text{Var} \left(\|X\|_2^2\right) = 2n.$$

Hint: Expressing the norm squared as a sum, proving the first equation becomes easy. The second equation reduces to computing the expectation of $\|X\|_2^2$. Express this quantity as a double sum. Then use (without proof) the fact that $\mathbb{E} g^4 = 3$ if $g \sim N(0, 1)$.
(b) Conclude that
\[ \|X\|_2^2 - n \leq C_2 \sqrt{n} \]  with probability at least 0.99.

*Hint: use Chebychev’s inequality.*

(c) Deduce that
\[ \frac{1}{2} \sqrt{n} \leq \|X\|_2 \leq 2 \sqrt{n} \]  with probability at least 0.99.

(d) Prove the tighter bound:
\[ \|X\|_2 - \sqrt{n} \leq C_3 \]  with probability at least 0.99.

*Hint: use the identity \(|a - b| = |a^2 - b^2| / |a + b| \) for \( a = \|X\|_2 \) and \( b = \sqrt{n} \). Part (b)
gives an upper bound on \(|a^2 - b^2|\), and \( \|X\|_2 \geq 0 \) yields a lower bound on \(|a + b|\).*

The next problem offers yet another look into the strange worlds in high dimensions. If we choose two random vectors \( X \) and \( Y \) independently at random in \( \mathbb{R}^n \), they happen to be almost orthogonal to each other with high probability (despite their knowing nothing about each other)!

**Problem 3 (Random vectors are almost orthogonal)**

Let \( X \) and \( Y \) be independent standard normal random vectors in \( \mathbb{R}^n \), where \( n \geq C_3 \).

(a) Check that
\[ \mathbb{E} \langle X, Y \rangle^2 = n. \]

(b) Deduce that
\[ \langle X, Y \rangle^2 \leq C_4 n \]  with probability at least 0.99.

(c) Denote by \( \theta \) the angle between the vectors \( X \) and \( Y \), as shown in the figure below. Prove that
\[ \left| \theta - \frac{\pi}{2} \right| \leq \frac{C_5}{\sqrt{n}} \]  with probability at least 0.97.

*Hint: use the formula \( \cos^2(\theta) = \frac{\langle X, Y \rangle^2}{\|X\|_2^2 \|Y\|_2^2} \). Part (b) gives an upper bound on the numerator, and part (c) of Problem 2 yields a lower bound on the denominator. Take the intersection of the three events, each of which happens with probability at least 0.99.*