Homework Two

EMALCA 2025: High-dimensional Statistics

Due in class on July 3 (Submit in groups of 2 or 3)

1. Let $A_j \in \mathbb{R}^{d\times d}$ be given deterministic matrices, for $j = 1\ldots, N$, and let $z \sim N(0, I_d)$. Set

$$U = \max_{j=1,\ldots,N} \|A_j z\|_2.$$

   a) (**2 points**) For any fixed j show that $\|A_j z\|_2$ is sub-Gaussian and obtain a bound on its sub- Gaussian parameter (or norm).

   b) (**2 points**) Derive an upper bound on U that holds with high probability (for large N) and has logarithmic dependence on N.

   c) (**2 points**) Derive an upper bound on $\mathbb{E}(U)$.

   d) (**2 points**) Someone claims that U is sub-Gaussian. Prove or disprove

   Hint: Using triangle inequality, show that the function $z \to \max_{j=1,\ldots,N} \|A_j z\|_2$ is Lipschitz continuous. Then use one of the inequalities we learned in class.

2. (**2 points**) **Exercise 2.15 from HDS.**

Exercise 2.15 (Concentration and kernel density estimation) Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of random variables drawn from a density f on the real line. A standard estimate of f is the *kernel density estimate*

$$\widehat{f_n}(x) := \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where $K:\mathbb{R} \to [0, \infty)$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(t)\, dt = 1$, and $h > 0$ is a bandwidth parameter. Suppose that we assess the quality of $\widehat{f_n}$ using the L^1-norm $\|\widehat{f_n} - f\|_1 := \int_{-\infty}^{\infty} |\widehat{f_n}(t) - f(t)|\, dt$. Prove that

$$\mathbb{P}[\|\widehat{f_n} - f\|_1 \geq \mathbb{E}[\|\widehat{f_n} - f\|_1] + \delta] \leq e^{-\frac{n\delta^2}{8}}.$$

Hint: Use the bounded difference inequality.