Assignment 1

Random Matrix Theory in Data Science and Statistics

(EN.553.796, Fall 2026)

Assigned: September 10, 2026   Due: September 22, 2026, 9:00AM

Solve all three problems. Each problem is worth an equal amount towards your grade.

Submit solutions in LaTeX. Write in complete sentences. Include and justify all steps of your arguments, but avoid writing excessive discussion that is not contributing to your solution. You are welcome to include images if you think that will help explain your solutions.

Problem 1 (Subgaussianity)**.** Recall that we call a random variable *centered* if its expectation is zero. A centered random scalar $X \in \mathbb{R}$ is called σ^2*-subgaussian* if $\mathbb{E}[\exp(\lambda X)] \le \exp(\sigma^2\lambda^2/2)$ for all $\lambda \in \mathbb{R}$. A centered random vector $\boldsymbol{x} \in \mathbb{R}^d$ is called the same if, for all $\boldsymbol{u} \in \mathbb{R}^d$ with $\|\boldsymbol{u}\| = 1$, $\langle \boldsymbol{u}, \boldsymbol{x} \rangle$ is a σ^2-subgaussian random scalar.

1. Show that if X is a centered random scalar that is σ^2-subgaussian, then aX is $a^2\sigma^2$-subgaussian, and that if $X_1, \dots, X_n$ are independent and each X_i is σ_i^2-subgaussian, then $\sum_{i=1}^n X_i$ is $(\sum_{i=1}^n \sigma_i^2)$-subgaussian. Conclude that a random vector with independent entries that are each centered and σ^2-subgaussian is itself σ^2-subgaussian.

2. Show that the following three conditions on a centered random scalar X are equivalent:

   (a) There exists $\sigma^2 > 0$ such that X is σ^2-subgaussian.

   (b) There exists $c > 0$ such that, for all $t > 0$, $\mathbb{P}[|X| \ge t] \le 2\exp(-ct^2)$.

   (c) There exists $\lambda > 0$ such that $\mathbb{E}[\exp(\lambda X^2)] < \infty$.

3. For X a centered subgaussian scalar, define

   $$\|X\|_{\mathsf{sg}} := \inf\{t > 0 : \mathbb{E}[\exp(X^2/t^2)] \le 2\}.$$

   Note that this is well-defined by Part 2. Show that $\|\cdot\|_{\mathsf{sg}}$ is a norm on the space of centered subgaussian random variables (over a given probability space, identifying random variables that are equal almost surely).

   (HINT: Prove and use that $(X+Y)^2 \le (1+a)X^2 + (1+\frac{1}{a})Y^2$ for any $a > 0$ (you might be more familiar with the case $a = 1$), together with Hölder's inequality.)

Problem 2 (Discretization)**.** This problem gives some additional applications of the discretization technique we saw recently in class.

1. Suppose that $\boldsymbol{x} \in \mathbb{R}^d$ is a centered σ^2-subgaussian random vector. Show that there are constants $c, C > 0$ such that

   $$\mathbb{P}[\|\boldsymbol{x}\| > C\sigma\sqrt{d}] \le \exp(-cd).$$

   That is, a high-dimensional $O(1)$-subgaussian vector has norm $O(\sqrt{d})$ with high probability.

   (**Hint:** Use ϵ-nets and adapt our discretization of the operator norm from lecture to apply instead to the Euclidean norm $\|\boldsymbol{x}\|$ of a vector.)

2. Let $\boldsymbol{X} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$ be a centered random symmetric matrix. Write $\mathsf{vec}(\boldsymbol{X}) \in \mathbb{R}^{d(d+1)/2}$ for the (symmetric) *vectorization* of $\boldsymbol{X}$, the vector formed by concatenating the entries on and above the diagonal of $\boldsymbol{X}$ (choose any order of entries you like). Show that there are constants $c, C > 0$ such that, if $\mathsf{vec}(\boldsymbol{X})$ is a σ^2-subgaussian random vector, then

   $$\mathbb{P}[\|\boldsymbol{X}\| > C\sigma\sqrt{d}] \le \exp(-cd).$$

   In particular, this implies that a high-dimensional random symmetric matrix with independent $O(1)$-subgaussian entries has operator norm $O(\sqrt{d})$ with high probability.

   (**Hint:** Use ϵ-nets again and view $\boldsymbol{v}^\top \boldsymbol{X}\boldsymbol{v}$ as $\langle \mathsf{vec}(\boldsymbol{X}), \boldsymbol{u}\rangle$ for a suitable $\boldsymbol{u} \in \mathbb{R}^{d(d+1)/2}$.)

Problem 3 (Gaussian combinatorics)**.** The Gaussian measure is the most important one in probability theory, if not all of mathematics. Here you will derive some of its algebraic properties that we will use in class and future assignments.

1. Let $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ for some $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}_{\mathsf{sym}}$ with $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$ (i.e., $\boldsymbol{\Sigma}$ is positive semidefinite). Prove that, for any smooth function $f : \mathbb{R}^d \to \mathbb{R}$ with $|f(\boldsymbol{x})| + \|\nabla f(\boldsymbol{x})\| \le C(1 + \|\boldsymbol{x}\|)^K$ for some $C, K > 0$ and all $\boldsymbol{x} \in \mathbb{R}^d$, we have

   $$\mathbb{E}[g_i f(\boldsymbol{g})] = \sum_{j=1}^{d} \Sigma_{ij}\mathbb{E}[\partial_j f(\boldsymbol{g})] = (\boldsymbol{\Sigma}\,\mathbb{E}[\nabla f(\boldsymbol{g})])_i \tag{1}$$

   where $\partial_i f$ is the partial derivative with respect to the ith argument and ∇ is the gradient (the second equality is just by the definition of gradient).

   (**Hint:** Integrate by parts. You might also find it useful to first treat the case $\boldsymbol{\Sigma} = \boldsymbol{I}_d$, and then to observe that $\boldsymbol{g}$ and $\boldsymbol{\Sigma}^{1/2}\boldsymbol{h}$ for $\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ have the same law.)

2. Let $g \sim \mathcal{N}(0, 1)$ (a Gaussian scalar, not a vector). Prove that, for $k \ge 1$, $\mathbb{E}[g^{2k-1}] = 0$ and $\mathbb{E}[g^{2k}] = \prod_{i=1}^{k}(2i - 1)$.

3. Let $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ as in Part 1, and let $1 \leq i_1 < \cdots < i_k \leq d$. Let $\mathcal{M}$ be the set of all *matchings* of the set $I = \{i_1, \ldots, i_k\}$: a matching is a set of disjoint pairs $\{i_a, i_b\}$ whose union is I. For example, the three matchings of $\{1, 2, 3, 4\}$ are $\{\{1, 2\}, \{3, 4\}\}$, $\{\{1, 3\}, \{2, 4\}\}$, and $\{\{1, 4\}, \{2, 3\}\}$. Prove that

$$\mathbb{E}\left[\prod_{a=1}^{k} g_{i_a}\right] = \sum_{M \in \mathcal{M}} \prod_{\{a,b\} \in M} \Sigma_{ab}. \tag{2}$$

(For example, one case of the claim is that $\mathbb{E}[g_1 g_2 g_3 g_4] = \Sigma_{12}\Sigma_{34} + \Sigma_{13}\Sigma_{24} + \Sigma_{14}\Sigma_{23}$.) Generalize this to allow for repetitions among the $i_1, \ldots, i_k$. Try to be precise. Explain why Part 2 is a special case of this latter generalization.

(**Hint:** Induction.)