SOLVE ANY FOUR OF THE FOLLOWING FIVE PROBLEMS:

**Problem 1 (25pts).** Let $L : \mathbb{R}^d \to \mathbb{R}$ be a smooth loss function. Recall that $L$ satisfies the restricted strong convexity property at $\theta^* \in \mathbb{R}^d$ if

$$
(\nabla L(\theta^* + \Delta) - \nabla L(\theta^*))^T \Delta \geq \begin{cases} 
\alpha_1 \|\Delta\|^2_2 - \frac{\log d}{n} \|\Delta\|^2_1 & \text{if } \|\Delta\|_2 \leq 1, \\
\alpha_2 \|\Delta\|^2_2 - \tau_2 \sqrt{\frac{\log d}{n} \|\Delta\|_1} & \text{if } \|\Delta\|_2 \geq 1
\end{cases}
$$

for some constants $\alpha_1, \alpha_2 > 0$ and $\tau_1, \tau_2 \geq 0$. Suppose that (i) $L$ is convex, (ii) $\|\Delta\|_1 \leq 2R$, (iii) $n \geq 4R^2 \tau_2 \log d$, and (iv) condition (a) holds. Show that condition (b) also holds with $\alpha_2 = \alpha_1$ and $\tau_2 = 1$.

**Problem 2 (25pts).** Let $g_1, \ldots, g_n \sim \mathcal{N}(0, 1)$ be standard real Gaussian random variables that are not necessarily independent. Show that

$$
\mathbb{E}\left[\max_{1 \leq i \leq n} |g_i|\right] \leq \sqrt{2 \log(2n)}.
$$

*(Hint: For any $\beta > 0$, we have)*

$$
\mathbb{E}\left[\max_{1 \leq i \leq n} |g_i|\right] = \frac{1}{\beta} \mathbb{E}\left[\log \exp\left(\beta \max_{1 \leq i \leq n} |g_i|\right)\right] \leq \frac{1}{\beta} \mathbb{E}\left[\log \sum_{i=1}^{n} \exp(\beta |g_i|)\right].
$$

*Now, choose $\beta$ to optimize the above bound.)*

**Problem 3 (25pts).** Let $X \in \mathbb{R}^{m \times n}$ be a rank-$r$ $m \times n$ matrix with $1 \leq r \leq m \leq n$. Suppose that

$$
X = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T
$$

is an SVD of $X$, where $\Sigma = \text{Diag}(\sigma_1(X), \ldots, \sigma_r(X)) \in \mathbb{R}^{r \times r}$ is a diagonal matrix containing the non-zero singular values of $X$. Let $\|X\|_* = \sum_{i=1}^{r} \sigma_i(X)$ denote the nuclear norm of $X$. Show that

$$
\partial \|X\|_* = \left\{ U \begin{bmatrix} I_r & 0 \\ 0 & W \end{bmatrix} V^T : \|W\| \leq 1 \right\},
$$

where $I_r$ is the $r \times r$ identity matrix and $\|W\|$ is the spectral norm (largest singular value) of $W$.

**Problem 4 (25pts).** Let $A$ be an $m \times n$ matrix whose entries are i.i.d. standard real Gaussian random variables and $\mathbb{S}^{n-1} = \{ x \in \mathbb{R}^n : \|x\|_2 = 1 \}$ denote the unit $n$-sphere. Consider the Gaussian processes

$$
X_{u,v} = u^T Av, \quad Y_{u,v} = g^T u + h^T v,
$$

where $u \in \mathbb{S}^{m-1}, v \in \mathbb{S}^{n-1}$, and $g \sim \mathcal{N}(0, I_m), h \sim \mathcal{N}(0, I_n)$ are two independent standard Gaussian random vectors.
(a) \( \text{(15pts).} \) Show that for any \( u, u' \in S^{m-1} \) and \( v, v' \in S^{n-1} \), we have

\[
E \left[ |X_{u,v} - X_{u',v'}|^2 \right] \leq E \left[ |Y_{u,v} - Y_{u',v'}|^2 \right].
\]

(b) \( \text{(10pts).} \) Recall that the largest singular value of \( A \) is given by

\[
\sigma_{\text{max}}(A) = \sup_{u \in S^{m-1}, v \in S^{n-1}} u^T A v.
\]

Show that for any \( A, B \in \mathbb{R}^{m \times n} \), we have

\[
|\sigma_{\text{max}}(A) - \sigma_{\text{max}}(B)| \leq \|A - B\|_F.
\]

(Remarks: When combined with Slepian’s inequality (a special case of Gordon’s inequality introduced in class), the result in (a) implies that)

\[
E[\sigma_{\text{max}}(A)] = E \left[ \sup_{u \in S^{m-1}, v \in S^{n-1}} u^T A v \right] \leq E \left[ \sup_{u \in S^{m-1}} g^T u \right] + E \left[ \sup_{v \in S^{n-1}} h^T v \right]
= E[\|g\|_2] + E[\|h\|_2] \leq \sqrt{m} + \sqrt{n}.
\]

This can then be combined with the result in (b) to obtain a bound on the tail probability

\[
Pr \left[ \sigma_{\text{max}}(A) \geq E[\sigma_{\text{max}}(A)] + t \right]
\]

(\text{you do not need to derive such a bound}.)

**Problem 5 (25pts).** Another SDP relaxation of the community detection problem under the stochastic block model is the following:

maximize \( \langle A, X \rangle - \mu \langle E, X \rangle \)

subject to \( \text{diag}(X) = e, \; X e \geq m e, \)

\[ X \succeq 0, \quad X \geq 0. \] (SDR)

Here, \( \mu > 0 \) is a given parameter, \( A, X \) are \( n \times n \) symmetric matrices, \( e \) is the \( n \)-dimensional vector of all ones, and \( E = ee^T \) is the \( n \times n \) matrix of all ones.

(a) \( \text{(10pts).} \) Derive the dual of (SDR).

(b) \( \text{(15pts).} \) Does strong duality hold between (SDR) and the dual derived in (a)? Explain.