STAT 607: HW3

- Type and submit your code in a PLAIN text format file. Do not use Microsoft Word .doc format or .rtf format or .pdf format. The file name should match your name, and make sure the questions are easily identified in the file. Submit your code file online, and hand in your written answers in class, by the due date.

Let $\mathcal{M}_p$ denote the space of all symmetric matrices of $\mathbb{R}^{p\times p}$, and $\mathcal{M}_p^+$ its subset of positive definite elements. For a matrix θ, $\lambda_{\mathsf{min}}(\theta)$ denotes the smallest eigenvalue of θ.

Let $S \in \mathbb{R}^{p\times p}$ a symmetric semi-positive definite matrix (but not necessarily positive definite). We consider the problem of computing

$$\hat{\theta} = \mathsf{Argmin}_{\theta\in\mathcal{M}_p^+} \left[-\log\det(\theta) + \mathsf{Tr}(\theta S) + \lambda\|\theta\|_1\right],$$

for a given constant $\lambda > 0$, and $\|\theta\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^p \sum_{j=1}^p |\theta_{ij}|$.

1. Show that $\hat{\theta}$ exists and is unique. The optimality condition satisfied by $\hat{\theta}$ implies that there exists a matrix $Z(\hat{\theta}) \in \mathbb{R}^{p\times p}$, with $(Z(\hat{\theta}))_{ij} \in [-1, 1]$ such that
$$S - \hat{\theta}^{-1} + \lambda Z(\hat{\theta}) = 0.$$
Use this to show that
$$\lambda_{\mathsf{min}}(\hat{\theta}) \geq \epsilon_\star \stackrel{\text{def}}{=} \frac{1}{\|S\|_2 + \lambda p}.$$

2. For a symmetric matrix $M \in \mathbb{R}^{p\times p}$, and $\gamma > 0$, define
$$\mathsf{Prox}_\gamma(M) \stackrel{\text{def}}{=} \mathsf{Argmin}_{u\in\mathcal{M}_p} \left[\lambda\|u\|_1 + \frac{1}{2\gamma}\|u - M\|_{\mathsf{F}}^2\right].$$
Show that
$$(\mathsf{Prox}_\gamma(M))_{ij} = \begin{cases} M_{ij} - \gamma\lambda & \text{if } M_{ij} > \gamma\lambda \\ M_{ij} + \gamma\lambda & \text{if } M_{ij} < -\gamma\lambda \\ 0 & \text{otherwise} \end{cases}$$

3. Consider the following iterative algorithm for computing $\hat{\theta}$. Take $\theta_0 \in \mathcal{M}_p^+$ such that $\lambda_{\mathsf{min}}(\theta_0) \geq \epsilon_\star$, and repeat for $k = 0, 1, \ldots,$

$$\theta_{k+1} = \mathsf{Prox}_{\gamma\lambda}\left(\theta_k - \gamma\left(S - \theta_k^{-1}\right)\right). \tag{1}$$

   Using the optimality condition attached to the definition of the Prox function to show that if $0 < \gamma \leq \epsilon_\star^2$, then whenever $\lambda_{\mathsf{min}}(\theta_k) \geq \epsilon_\star$, then we also have $\lambda_{\mathsf{min}}(\theta_{k+1}) \geq \epsilon_\star$. In other words, if the step-size $\gamma > 0$ satisfies $\gamma \leq \epsilon_\star^2$, then positive definiteness is never loss in the iterations (1). Furthermore it can be shown (but we will do this here) that in that case $\lim_k \|\theta_k - \hat{\theta}\|_{\mathsf{F}} = 0$.

4. Using the Cholesky factorization, implement algorithm (1). Do a simulation study to assess empirically how the computational cost of the algorithm scales with the dimension p.