Math 281C Homework 7 Solutions

1. Define
$$r_S(\lambda,\mu) = \mathbb{E}[\text{soft}_\lambda(y) - \mu]^2,$$
where $y \sim \mathcal{N}(\mu, 1)$. Show that

   (a) $\mu \to r_S(\lambda,\mu)$ is increasing on $[0,\infty)$;

   Solution: It can be calculated that
$$r_S(\lambda,\mu) = \underbrace{\int_{-\infty}^{-\lambda-\mu}(x+\lambda)^2\phi(x)\mathrm{d}x}_{\text{I}} + \underbrace{\mu^2\int_{-\lambda-\mu}^{\lambda-\mu}\phi(x)\mathrm{d}x}_{\text{II}} + \underbrace{\int_{\lambda-\mu}^{\infty}(x-\lambda)^2\phi(x)\mathrm{d}x}_{\text{III}}, \tag{1}$$
and
$$\frac{\partial}{\partial\mu}r_S(\lambda,\mu) = 2\mu(\Phi(\lambda-\mu) - \Phi(-\lambda-\mu)) > 0$$
for $\mu > 0$.

   (b) For all $\lambda > 0$, we have
$$r_S(\lambda, 0) \le \sqrt{\frac{2}{\pi}}\frac{1}{\lambda}e^{-\lambda^2/2};$$

   Solution: When $\mu = 0$, by the symmetric property,
$$\begin{aligned} r_S(\lambda,0) &= 2\int_\lambda^\infty (x-\lambda)^2\phi(x)\mathrm{d}x \\ &= 2\left[\int_\lambda^\infty x^2\phi(x)\mathrm{d}x - 2\lambda\int_\lambda^\infty x\phi(x)\mathrm{d}x + \lambda^2\int_\lambda^\infty \phi(x)\mathrm{d}x\right] \\ &= 2(1+\lambda^2)(1-\Phi(\lambda)) - 2\lambda\phi(\lambda). \end{aligned}$$
Using the Gaussian tail bound from Homework 1
$$1 - \Phi(\lambda) \le \frac{\phi(\lambda)}{\lambda}$$
gives us the desired result.

   (c) When μ approaches $\pm\infty$,
$$\lim_{\mu\to\infty} r_S(\lambda,\mu) = 1 + \lambda^2,$$
and
$$\sup_{\mu\in\mathbb{R}} r_S(\lambda,\mu) = 1 + \lambda^2.$$

   Solution: We only need to show the case for $\mu \to \infty$ by the symmetry. In the equation (1), I $\to 0$ since $(x+\lambda)^2\phi(x)$ is integrable, II $= \mu^2(\Phi(\lambda-\mu) - \Phi(-\lambda-\mu)) \to 0$, and III $\to \mathbb{E}(x-\lambda)^2 = 1 + \lambda^2$ where $x \sim \mathcal{N}(0,1)$. The second result is trivial.

2. Consider the model
$$Y = \theta + \sigma\varepsilon,$$
where $\varepsilon \in \mathbb{R}^n$ consists of independent mean-zero 1-sub-Gaussian components, and assume θ is k-sparse:
$$\|\theta\|_0 = \text{card}\{j \in [n] : \theta_j \neq 0\} \le k.$$
In this question, we investigate the soft-thresholding estimator
$$\widehat{\theta} := \underset{\theta}{\text{argmin}}\, \frac{1}{2}\|\theta - Y\|_2^2 + \lambda\|\theta\|_1.$$

(a) Show that if $\lambda \geq \sigma\|\varepsilon\|_\infty$, then

$$\|\widehat{\theta} - \theta\|_2^2 \leq 4k\lambda^2;$$

Solution: $\widehat{\theta} = \text{soft}_\lambda(Y)$, and when $\lambda \geq \sigma\|\varepsilon\|_\infty$,

$$\begin{aligned}
\|\widehat{\theta} - \theta\|_2^2 &= \sum_{i:\theta_i \neq 0} (\widehat{\theta}_i - \theta_i)^2 + \sum_{i:\theta_i = 0} (\widehat{\theta}_i - \theta_i)^2 \\
&= \sum_{i:\theta_i \neq 0} (\text{soft}_\lambda(Y_i) - \theta_i)^2 + 0 \\
&= \sum_{i:\theta_i \neq 0} (\text{soft}_\lambda(Y_i) - \text{soft}_\lambda(\theta_i) + \text{soft}_\lambda(\theta_i) - \theta_i)^2 \\
&\leq 2 \sum_{i:\theta_i \neq 0} (\text{soft}_\lambda(\theta_i + \sigma\varepsilon_i) - \text{soft}_\lambda(\theta_i))^2 + 2 \sum_{i:\theta_i \neq 0} (\text{soft}_\lambda(\theta_i) - \theta_i)^2 \\
&\leq 2k\lambda^2 + 2k\lambda^2 = 4k\lambda^2.
\end{aligned}$$

In the above derivation, we use the facts that $\text{soft}_\lambda(\cdot)$ is 1-Lipschitz continuous, $|\text{soft}_\lambda(y) - y| \leq \lambda$, $\lambda \geq \sigma\|\varepsilon\|_\infty$, and $\|\theta\|_0 \leq k$.

(b) Show that if

$$\lambda = 2\sqrt{\sigma^2 \log(2n)},$$

then with probability at least $1 - 1/(2n)$,

$$\|\widehat{\theta} - \theta\|_2^2 \leq 16k\sigma^2 \log(2n).$$

Solution: By the sub-Gaussianities of components of ε, we have the tail bound

$$\mathbb{P}(\sigma\|\varepsilon\|_\infty \geq \lambda) \leq n\mathbb{P}(|\sigma\epsilon_i| \geq \lambda) \leq 2n \exp\left(-\frac{\lambda^2}{2\sigma^2} \right) = \frac{1}{2n},$$

which completes the proof.