MATH 5473 HW3

1. (a) Let $f(x) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left\{-\frac{1}{2}(X_i-\mu)^T \Sigma^{-1} (X_i-\mu)\right\}$

$$\log f(x) = \sum_{i=1}^{n} \left[-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(X_i-\mu)^T \Sigma^{-1} (X_i-\mu)\right]$$

$$\begin{aligned} l_n &= \mathrm{trace}(\log f(x)) \\ &= \mathrm{trace}\left(-\frac{np}{2}\log(2\pi)\right) - \frac{n}{2}\log\det(\Sigma) - \frac{1}{2}\sum_{i=1}^{n} \mathrm{trace}\left[(X_i-\mu)^T \Sigma^{-1} (X_i-\mu)\right] \\ &= -\frac{n}{2}\mathrm{trace}(\Sigma^{-1} S_n) - \frac{n}{2}\log\det(\Sigma) + C. \end{aligned}$$

here $S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)^T$.

(b)
$$\begin{aligned} f(X+\Delta) &= \mathrm{trace}\left[A(X+\Delta)^{-1}\right] \\ &= \mathrm{trace}\left[AX^{-1}(I+\Delta X^{-1})^{-1}\right] \\ &\approx \mathrm{trace}\left[AX^{-1}(I-\Delta X^{-1})\right] \\ &= \mathrm{trace}(AX^{-1}) - \mathrm{trace}(A'X^{-1}\Delta X^{-1}) \\ &= \mathrm{trace}(AX^{-1}) - \mathrm{trace}(X^{-1}A'X^{-1}\Delta) \\ &= f(X) - \mathrm{trace}(X^{-1}A'X^{-1}\Delta) \end{aligned}$$

(c)
$$\begin{aligned} g(X+\Delta) &= \log\det(X+\Delta) \\ &= \log\det X + \log\det(I + X^{-\frac{1}{2}}\Delta X^{-\frac{1}{2}}) \\ &= \log\det X + \sum_{i=1}^{n}\log(1+\lambda_i), \end{aligned}$$

here λ_i is the i-th eigenvalue of $X^{-\frac{1}{2}}\Delta X^{\frac{1}{2}}$.

For small Δ, $\log(1+\lambda_i) \approx \lambda_i$.

$$g(X+\Delta) \approx \log\det(X) + \sum_{i=1}^{n} \lambda_i$$

$$= \log\det(X) + \mathrm{trace}(X^{-\frac{1}{2}} \Delta X^{-\frac{1}{2}})$$

$$= \log\det(X) + \mathrm{trace}(X^{-1}\Delta)$$

(d) By (a), $\ell(X) = -\frac{n}{2}\,\mathrm{trace}(\Sigma^{-1} S_n) - \frac{n}{2}\log\det(\Sigma) + C.$

$$\Rightarrow \quad 0 = \frac{\partial \ell}{\partial \Sigma} = -\frac{n}{2}(-\Sigma^{-1} S_n \Sigma^{-1}) - \frac{n}{2}\Sigma^{-1}$$

$$\Rightarrow \quad \hat{\Sigma} = S_n.$$

2.(a) Let $R(\mu) = \frac{1}{2}\|y-\mu\|_2^2 + \frac{\lambda}{2}\|\mu\|_2^2 = \frac{1}{2}(y^T y - 2y^T\mu + \mu^T\mu) + \frac{\lambda}{2}\mu^T\mu$

$$\Rightarrow 0 = \frac{\partial R}{\partial \mu} = -\frac{1}{2}y + \frac{1+\lambda}{2}\hat{\mu}$$

$$\Rightarrow \hat{\mu}_i = \frac{1}{1+\lambda} y_i$$

(b) Let $L(\mu) = \frac{1}{2}\|y-\mu\|_2^2 + \lambda\|\mu\|_1$

And $\partial_\mu L = \mu y + \lambda\,\mathrm{sign}(\mu)$.

$$0 \in \partial_\mu L\big|_{\mu=\mu^*} \Rightarrow \mu^* = \begin{cases} y-\lambda & y>\lambda \\ 0 & -\lambda \le y \le \lambda \\ y+\lambda & y<-\lambda \end{cases}$$

Thus, $\hat{\mu}_i = \mathrm{sign}(y_i)(|y_i|-\lambda)_+$.

(c) Let $F(\mu) = \|y-\mu\|_2^2 + \lambda^2\|\mu\|_0$

$$F(\mu_i) = \begin{cases} (y_i-\mu_i)^2 + \lambda^2 & \mu_i \neq 0 \\ y_i^2 & \mu_i = 0 \end{cases}$$

$$\Rightarrow \hat{\mu}(y) = y_i\, I(|y_i|>\lambda).$$

If $\hat{\mu}(y) = (1-g(y))\, y$,

$\Rightarrow$ $g(y) = I(|y| \le \lambda)$, $g(y)$ is not absolutely continuous.

Thus $g(x)$ is not weakly differentiable.

(d). Let $g(y) = -\frac{\alpha}{\|y\|^2}$ $\Rightarrow$ $\hat{\mu} = (1+g)y$.

And $\mathbb{E}\|\hat{\mu}(y) - \mu\|^2 = \mathbb{E}\, U_\alpha(y)$. $0 = \frac{\partial U_\alpha(y)}{\partial \alpha}$,

$\Rightarrow$ $\alpha^* = \arg\min_\alpha U(y) = p-2$.

And $\|y\|^2 \sim \chi^2(\|\mu\|^2, p)$. $N \sim \text{Possion}(\frac{1}{2}\|\mu\|^2)$

$$\mathbb{E}\left(\frac{1}{\|Y\|^2}\right) = \mathbb{E}\left(\frac{1}{p+2N-2}\right) \ge \frac{1}{p+\|\mu\|^2-2},$$

For $p>3$, $\text{Risk}(\mu) < p - \frac{(p-2)^2}{p-2+\|\mu\|^2} < p$.

(e). The above four estimates are all shrinkage rules.

3. For any matrix C and orthogonal matrix P,

$P^T C P$ is admissible iff Cy is admissible.

For given C, we have orthogonal matrix P and diagonal matrix $D = \text{diag}\{d_1, d_2, \cdots\}$. s.t $P^T D P = C$.

Then Dy is admissible if one or two $d_i = 1$ and others $0 \le d_i < 1$.

Pf. Let $d_1 = d_2 = 1$, $0 \le d_i < 1$, $i \ge 3$.

(contradiction) There exists an estimate $H(y) = (h_1(y), h_2(y), \cdots)$.

s.t. $\rho(\theta, H(y)) \le \rho(\theta, Dy)$

$$\sum_{i=1}^{p}\int_{E_p}(h_i(y)-\theta_i)^2\,d\psi \le 2+\sum_{i=3}^{p}\left[d_i^2+\theta_i^2(d_i-1)^2\right]$$

$$\Rightarrow \int_{E_p}\left[(h_1(y)-\theta_1)^2+(h_2(y)-\theta_2)^2\right]d\psi \le 2$$

$\Rightarrow$ Dy is admissible.

Besides. for generenerally $D=\mathrm{diag}(d_i)$, let

$d_i \ge 1 \Rightarrow \tilde{d}_i = 1$, $d_i<0 \Rightarrow \tilde{d}_i = -d_i$.

$d_i = 1 \Rightarrow \tilde{d}_i = \left(1-\frac{k-2}{\sum_{i=1}^{p} y_i^2}\right)$.

s.t $\tilde{D}=\mathrm{diag}(\tilde{d}_i)$. and $\tilde{d}_i=1$ for at most two i. while the others satisfy $0\le \tilde{d}_i<1$.

By the above proof, we know that $\tilde{D}y$ is admissible

$\Rightarrow$ Dy is admissible

$\Rightarrow$ Cy is admissible.