Zhankun Luo (ECE Ph.D. student)
luo333@purdue.edu

Problem 6

6. Let $X \sim N(0, R)$ where R is a $p \times p$ symmetric positive-definite matrix. Further define the precision matrix, $B = R^{-1}$, and use the notation

$$B = \begin{pmatrix} 1/\sigma^2 & A \\ A^t & C \end{pmatrix}$$

where $A \in \mathbb{R}^{1\times(p-1)}$ and $C \in \mathbb{R}^{(p-1)\times(p-1)}$

a) Calculate the marginal density of X_1, the first component of X, given the components of the matrix R.
b) Calculate the conditional density of X_1 given all the remaining components, $Y = [X_2, \cdots, X_p]^T$.
c) What is the conditional mean and covariance of X_1 given Y ?

solution
a) We have proved that in Problem 5: let $X \sim N(\mu, R)$ be a jointly Gaussian random vector, and let $K \in \mathbb{R}^{M\times p}$ be a rank M matrix. Then the vector $Y = KX \sim N(K\mu, KRK^T)$ is also jointly Gaussian. The corresponding PDF for X and Y are

$$f_X(x) = \frac{1}{(2\pi)^{p/2}}|R|^{-1/2}\exp\left\{-\frac{1}{2}(x-\mu)^T R^{-1}(x-\mu)\right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{M/2}}|KRK^T|^{-1/2}\exp\left\{-\frac{1}{2}(y-K\mu)^T (KRK^T)^{-1}(y-K\mu)\right\}$$

Here we know that $X \sim N(0, R)$, so $\mu = 0$ and $Y = KX \sim N(0, KRK^T)$

$$f_X(x) = \frac{1}{(2\pi)^{p/2}}|R|^{-1/2}\exp\left\{-\frac{1}{2}x^T R^{-1}x\right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{M/2}}|KRK^T|^{-1/2}\exp\left\{-\frac{1}{2}y^T (KRK^T)^{-1}y\right\}$$

We may set $K = [1, 0, \cdots, 0] \in \mathbb{R}^{1\times p}$, then $X_1 = KX$, thus $KRK^T = R_{11} = \mathbb{E}[X_1^2]$

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi R_{11}}}\exp\left\{-\frac{x_1^2}{2R_{11}}\right\}$$

b) We may set $K \in \mathbb{R}^{(p-1)\times p}$ as

$$A = \begin{pmatrix} 0_{1\times(p-1)} \\ I_{(p-1)\times(p-1)} \end{pmatrix}$$

$Y = [X_2, \cdots, X_p]^T = KX$, $r \equiv [R_{21}, \cdots, R_{p1}]^T = [R_{12}, \cdots, R_{1p}]^T$ and $R' \equiv KRK^T$ is given by

$$R' \equiv KRK^T = \begin{pmatrix} R_{22} & \cdots & R_{2p} \\ \vdots & \ddots & \vdots \\ R_{p2} & \cdots & R_{pp} \end{pmatrix}, \quad R = \begin{pmatrix} R_{11} & r^T \\ r & R' \end{pmatrix}$$

Consider the adjugate matrix, it is clear that for the inverse matrix B of R. We have that

$$\frac{1}{\sigma^2} = (-1)^{1+1}\frac{|R'|}{|R|} = \frac{|R'|}{|R|} \tag{1}$$

$$a_k = (-1)^{k+1}\frac{(-1)^{k-2}\cdot|R'_{(k-1)}|}{|R|} = -\frac{|R'_{(k-1)}|}{|R|} \quad k\in\{2,\cdots,p\}$$

Where $A=[a_2,\cdots,a_p]$, $R'=[r'_2,\cdots,r'_p]$ and $R'_{(k-1)}=[r'_2,\cdots,r'_{k-1},r,r'_{k+1},\cdots,r'_p]$ whose $(k-1)$th column vector is replaced with column vector r
Since the solution of $R'x=R'[x_1,\cdots,x_{p-1}]=r$ is given by $x_{k-1}=\frac{|R'_{(k-1)}|}{|R'|},\ k\in\{2,\cdots,p\}$

$$A^T = -\frac{|R'|(R')^{-1}r}{|R|} = -\frac{|R'|}{|R|}(R')^{-1}r \tag{2}$$

Moreover, we can derive the relationship between $|R|$ and $|R'|$ by expanding $|R|$ with first row

$$|R| = R_{11}|R'| + \sum_{k=2}^{p}(-1)^{k+1}R_{1k}\cdot(-1)^{k-2}|R'_{(k-1)}| = R_{11}|R'| - r^T(R')^{-1}r|R'|$$

$$\Rightarrow \frac{|R'|}{|R|} = \frac{1}{R_{11}-r^T(R')^{-1}r} \tag{3}$$

Furthermore, we can verify the equation below if $\mathbf{A},\mathbf{D}$ are invertible

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I} & -\mathbf{B}\mathbf{D}^{-1}\\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}\end{bmatrix}$$

Select $\mathbf{A}=R_{11},\mathbf{B}=r^T,\mathbf{C}=r,\mathbf{D}=R'$
Since $R_{11}\neq 0$ and R' is positive definite, both R_{11},R' are invertible

$$R^{-1} = \begin{pmatrix}\frac{1}{\sigma^2} & A\\ A^T & C\end{pmatrix} = \begin{pmatrix}\frac{1}{R_{11}-r^T(R')^{-1}r} & \vec{0}^T\\ \vec{0} & \left(R'-\frac{rr^T}{R_{11}}\right)^{-1}\end{pmatrix}\begin{pmatrix}I & -r^T(R')^{-1}\\ -\frac{r}{R_{11}} & I\end{pmatrix}$$

$$\Rightarrow C = \left(R'-\frac{rr^T}{R_{11}}\right)^{-1} \tag{4}$$

With **Sherman–Morrison formula**

$$\left(\mathbf{A}+uv^T\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}uv^T\mathbf{A}^{-1}}{1+v^T\mathbf{A}^{-1}u}$$

Select $\mathbf{A}=R', u=-\frac{r}{\sqrt{R_{11}}}, v=\frac{r}{\sqrt{R_{11}}}$, notice $(R')^{-T}=(R')^{-1}$ and (1), (2), (3), (4)

$$C = (R')^{-1} + \frac{(R')^{-1}rr^T(R')^{-1}}{R_{11}-r^T(R')^{-1}r} = (R')^{-1} + \frac{|R'|}{|R|}(R')^{-1}rr^T(R')^{-T} = (R')^{-1}+\sigma^2A^TA \tag{5}$$

With $Y = [X_2, \cdots, X_p]^T$ and (1),(5)

$$
\begin{aligned}
f_X(x) &= \frac{1}{(2\pi)^{p/2}}|R|^{-1/2}\exp\left\{-\frac{1}{2}\left[\frac{x_1^2}{\sigma^2} + 2x_1Ay + y^T\left[(R')^{-1} + \sigma^2A^TA\right]y\right]\right\} \\
&= \frac{1}{(2\pi)^{p/2}}|R|^{-1/2}\exp\left\{-\frac{(x_1+\sigma^2Ay)^2}{2\sigma^2}\right\}\cdot\exp\left\{-\frac{1}{2}y^T(R')^{-1}y\right\} \\
f_Y(y) &= \frac{1}{(2\pi)^{(p-1)/2}}|R'|^{-1/2}\exp\left\{-\frac{1}{2}y^T(R')^{-1}y\right\} \\
f_{X_1|Y}(x_1 \mid y) &= \frac{f_X(x)}{f_Y(y)} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x_1+\sigma^2Ay)^2}{2\sigma^2}\right\}
\end{aligned}
$$

c) Treat $Y = [X_2, \cdots, X_p]^T$ as a constant, $X_1 \sim N(-\sigma^2Ay, \sigma^2)$, the conditional mean is given by

$$\mathbb{E}[X_1|Y] = -\sigma^2Ay$$

and the conditional variance is given by

$$\mathrm{Var}[X_1|Y] = \sigma^2$$

Problem 7

7. Let $X \sim N(0, R)$ where R is a $p \times p$ symmetric positive-definite matrix with an eigen decomposition of the form $R = E\Lambda E^T$

a) Calculate the covariance of $\tilde{X} = E^T X$, and show that the components of $\tilde{X}$ are jointly independent Gaussian random variables. (Hint: Use the result of problem 5 above.)
b) Show that if $Y = E\Lambda^{\frac{1}{2}}W$ where $W \sim N(0, I)$, then $Y \sim N(0, R)$. How can this result be of practical value?

solution

a) We have proved that in Problem 5: let $X \sim N(\mu, R)$ be a jointly Gaussian random vector, and let $K \in \mathbb{R}^{M\times p}$ be a rank M matrix. Then the vector $Y = KX \sim N(K\mu, KRK^T)$ is also jointly Gaussian. The corresponding PDF for X and Y are

$$f_X(x) = \frac{1}{(2\pi)^{p/2}}|R|^{-1/2}\exp\left\{-\frac{1}{2}(x-\mu)^T R^{-1}(x-\mu)\right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{M/2}}|KRK^T|^{-1/2}\exp\left\{-\frac{1}{2}(y-K\mu)^T (KRK^T)^{-1}(y-K\mu)\right\}$$

Here we know that $X \sim N(0, R)$, where $\mu = 0$ and $Y = KX \sim N(0, KRK^T)$

$$f_X(x) = \frac{1}{(2\pi)^{p/2}}|R|^{-1/2}\exp\left\{-\frac{1}{2}x^T R^{-1}x\right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{M/2}}|KRK^T|^{-1/2}\exp\left\{-\frac{1}{2}y^T (KRK^T)^{-1}y\right\}$$

We may set $K = E^T \in \mathbb{R}^{p\times p}$, then $\tilde{X} = KX = E^T X$, thus $KRK^T = E^T RE = E^T E\Lambda E^T E = \Lambda$

$$f_{\tilde{X}}(\tilde{x}) = \frac{1}{(2\pi)^{p/2}}|\Lambda|^{-1/2}\exp\left\{-\frac{1}{2}\tilde{x}^T\Lambda^{-1}\tilde{x}\right\} = \prod_{k=1}^{p}\frac{1}{\sqrt{2\pi}\sigma_k}\exp\left\{-\frac{\tilde{x}_k^2}{2\sigma_k^2}\right\}$$

Where $\Lambda = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_p^2)$, use the conclusion in Problem 5: $\tilde{X}_k = [\underbrace{0, \cdots, 0}_{(k-1)\ 0}, 1, \underbrace{0, \cdots, 0}_{(p-k)\ 0}] \cdot X$

Notice $[\underbrace{0, \cdots, 0}_{(k-1)\ 0}, 1, \underbrace{0, \cdots, 0}_{(p-k)\ 0}]\Lambda[\underbrace{0, \cdots, 0}_{(k-1)\ 0}, 1, \underbrace{0, \cdots, 0}_{(p-k)\ 0}]^T = \sigma_k^2$

$$f_{\tilde{X}_k}(\tilde{x}_k) = \frac{1}{\sqrt{2\pi}\sigma_k}\exp\left\{-\frac{\tilde{x}_k^2}{2\sigma_k^2}\right\}$$

In the end, we show that the components of $\tilde{X}$ are jointly independent Gaussian random variables

$$f_{\tilde{X}}(\tilde{x}) = \prod_{k=1}^{p}\frac{1}{\sqrt{2\pi}\sigma_k}\exp\left\{-\frac{\tilde{x}_k^2}{2\sigma_k^2}\right\} = \prod_{k=1}^{p} f_{\tilde{X}_k}(\tilde{x}_k)$$

The covariance of $\tilde{X} = E^T X$ is

$$\mathbb{E}[\tilde{X}\tilde{X}^T] - \mathbb{E}[\tilde{X}]\mathbb{E}[\tilde{X}^T] = \Lambda = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_p^2)$$

b) In problem 5, we know that $X \sim N(0, R)$, where $\mu = 0$ and $Y = KX \sim N(0, KRK^T)$

$$f_X(x) = \frac{1}{(2\pi)^{p/2}}|R|^{-1/2}\exp\left\{-\frac{1}{2}x^T R^{-1}x\right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{M/2}} |KRK^T|^{-1/2} \exp\left\{-\frac{1}{2} y^T (KRK^T)^{-1} y\right\}$$

We replace $W \to X, E\Lambda^{\frac{1}{2}} \to K, R \to I$, we know that $W \sim N(0, I)$ and $Y = (E\Lambda^{\frac{1}{2}})W \sim N(0, R)$, since $(E\Lambda^{\frac{1}{2}})I(E\Lambda^{\frac{1}{2}})^T = E\Lambda E^T = R$

$$f_W(w) = \frac{1}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2} w^T w\right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{p/2}} |R|^{-1/2} \exp\left\{-\frac{1}{2} y^T R^{-1} y\right\}$$

So, we derive the method to generate vector $Y \sim N(0, R)$ step by step:

(1) generate p samples $W_1, \cdots, W_p \overset{iid}{\sim} N(0, 1^2)$ with the pseudo ramdom generator
(2) combine $W_1, \cdots, W_p$ to form a vector $W = [W_1, \cdots, W_p]^T$
(3) compute Y with a linear transformation $Y = (E\Lambda^{\frac{1}{2}})W$

Problem 8

8. For each of the following cost functions, find expressions for the minimum risk Bayesian estimator, and show that it minimizes the risk over all estimators.

a) $C(x,\hat{x}) = |x-\hat{x}|^2$ minimum MSE (MMSE)
b) $C(x,\hat{x}) = |x-\hat{x}|$
c) $C(x,\hat{x}) = 1-\delta(x-\hat{x})$ maximum a posteriori (MAP)

solution
For the best estimator $\hat{x} = T(y)$, we may assume other estimator $\hat{x} = T(y)+\epsilon h(y)$, where $h(y)$ is an arbitrary function, $\epsilon \in \mathbb{R}$. We can computer the Bayes' Risk $\mathcal{J}(T+\epsilon h)$

$$\begin{aligned}\mathcal{J}(T+\epsilon h) = \mathbb{E}[C(x,\hat{x})] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C(x,T(y)+\epsilon h(y)) f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} C(x,T(y)+\epsilon h(y)) f_{X|Y}(x\mid y)\mathrm{d}x\right] f_Y(y)\mathrm{d}y\end{aligned}$$

Since $\hat{x} = T(y)$ has the minimal $\mathcal{J}(T+\epsilon h)$ for any function $h(y)$ at $\epsilon = 0$

$$0 = \frac{\mathrm{d}\mathcal{J}(T+\epsilon h)}{\mathrm{d}\epsilon}\bigg|_{\epsilon=0} = \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} \frac{\partial C(x,T(y)+\epsilon h(y))}{\partial \epsilon}\bigg|_{\epsilon=0} f_{X|Y}(x\mid y)\mathrm{d}x\right] f_Y(y)\mathrm{d}y$$

Especially, when $C(x,\hat{x}) = C(x-\hat{x})$ is a function of $x-\hat{x}$

$$\frac{\partial C(x,T(y)+\epsilon h(y))}{\partial \epsilon}\bigg|_{\epsilon=0} = C'(x-T(y))\cdot h(y)$$

Then, it holds for all any arbitrary function $h(y)$

$$0 = \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} C'(x-T(y)) f_{X|Y}(x\mid y)\mathrm{d}x\right] f_Y(y)h(y)\mathrm{d}y$$

So, the kernel must always be 0

$$\int_{-\infty}^{+\infty} C'(x-T(y)) f_{X|Y}(x\mid y)\mathrm{d}x = 0$$

a) When $C(x-\hat{x}) = |x-\hat{x}|^2$

$$C'(x-T(y)) = 2(x-T(y))$$

In the end, we have

$$\int_{-\infty}^{+\infty} 2(x-T(y)) f_{X|Y}(x\mid y)\mathrm{d}x = 0$$

$$T(y) = \frac{\int_{-\infty}^{+\infty} x f_{X|Y}(x\mid y)\mathrm{d}x}{\int_{-\infty}^{+\infty} f_{X|Y}(x\mid y)\mathrm{d}x} = \int_{-\infty}^{+\infty} x f_{X|Y}(x\mid y)\mathrm{d}x$$

Thus $T(Y)$ is the expectation $\mathbb{E}[X|Y]$ of X given Y

b) When $C(x-\hat{x}) = |x-\hat{x}|$

$$C'(x-T(y)) = \mathrm{sgn}(x-T(y))$$

$$\int_{T(y)}^{+\infty} f_{X|Y}(x\mid y)\mathrm{d}x - \int_{-\infty}^{T(y)} f_{X|Y}(x\mid y)\mathrm{d}x = 0$$

Thus $T(Y)$ is the conditional median of X given Y

c) When $C(x-\hat{x}) = 1-\delta(x-\hat{x})$

$$C'(x-T(y)) = -\delta'(x-T(y))$$

In the end, we have

$$\int_{-\infty}^{+\infty} -\delta'(x-T(y)) f_{X|Y}(x \mid y)\mathrm{d}x = -\delta(x-T(y)) f_{X|Y}(x \mid y)\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \delta(x-T(y)) \frac{\partial f_{X|Y}(x \mid y)}{\partial x}\mathrm{d}x = 0$$

Thus

$$\frac{\partial f_{X|Y}(x \mid y)}{\partial x}\Big|_{x=T(y)} = 0$$

It leads to

$$T(Y) = \operatorname*{argmin}_{x} f_{X|Y}(x \mid Y)$$

Problem 9

9. Let $\{Y_k\}_{k=1}^n$ be i.i.d. Bernoulli random variables with distribution

$$P\{Y_k = 1\} = \theta$$
$$P\{Y_k = 0\} = 1 - \theta$$

Compute the ML estimate of θ

solution

The pmf of Y_k can be written as $P(y_k \mid \theta) = \theta^{y_k}(1-\theta)^{1-y_k}$. Since $\{Y_k\}_{k=1}^n$ are i.i.d.

$$P(y_1, \cdots, y_n \mid \theta) = \prod_{k=1}^{n} \theta^{y_k}(1-\theta)^{1-y_k} = \theta^{\sum_{k=1}^n y_k} \cdot (1-\theta)^{n-\sum_{k=1}^n y_k}$$

Differentiate the log likelihood function at $\theta = \hat{\theta}$

$$\left.\frac{\mathrm{d}\log\left(P(y_1, \cdots, y_n \mid \theta)\right)}{\mathrm{d}\theta}\right|_{\theta=\hat{\theta}} = \frac{\left(\sum_{k=1}^n y_k\right)}{\hat{\theta}} - \frac{n - \sum_{k=1}^n y_k}{1-\hat{\theta}} = 0$$

We obtain the ML estimate $\hat{\theta}$ for parameter θ

$$\hat{\theta} = \frac{\sum_{k=1}^n y_k}{n}$$

Since, $\left.\frac{\mathrm{d}^2 P(y_1, \cdots, y_n \mid \theta)}{\mathrm{d}\theta^2}\right|_{\theta=\frac{\sum_{k=1}^n y_k}{n}} < 0$, we verify that $\hat{\theta} = \frac{\sum_{k=1}^n y_k}{n}$ is MLE

Problem 10

10. Let $\{X_i\}_{i=1}^n$ be i.i.d. random variables with distribution

$$P\{X_i = k\} = \pi_k, \quad s.t. \sum_{i=1}^m \pi_i = 1$$

Compute the MLE of the parameter vector $\theta = [\pi_1, \cdots, \pi_m]^T$ (Hint: You may the use ML the estimate method of of the Lagrange parameter multipliers to calculate the solution to the constrained optimization.)

solution

Since the pmf with parameter θ is $P(x_i \mid \theta) = \prod_{k=1}^m \pi_k^{\delta(x_i - k)}$

$$P(x \mid \theta) = \prod_{i=1}^n P(x_i \mid \theta) = \prod_{i=1}^n \prod_{k=1}^m \pi_k^{\delta(x_i-k)} = \prod_{k=1}^m \pi_k^{\sum_{i=1}^n \delta(x_i-k)}$$

The corresponding log likelihood function is

$$\log\left(P(x \mid \theta)\right) = \sum_{k=1}^m \left[\sum_{i=1}^n \delta(x_i - k)\right] \log(\pi_k)$$

To maximize the log likelihood function subject to the constraint $\sum_{i=1}^m \pi_i - 1 = 0$, we introduce a Lagrange function $L(\theta, \lambda)$ and a multiplier λ

$$L(\theta, \lambda) \equiv \log\left(P(x \mid \theta)\right) - \lambda\left(\sum_{k=1}^m \pi_k - 1\right)$$

For $\pi_k,\ k \in \{1, \cdots, m\}$

$$\frac{\partial L}{\partial \pi_k} = \frac{\sum_{i=1}^n \delta(x_i - k)}{\pi_k} - \lambda = 0 \Rightarrow \pi_k = \frac{\sum_{i=1}^n \delta(x_i - k)}{\lambda}$$

To sum up k from 1 to m, we notice that $\sum_{k=1}^m \delta(x_i - k) = 1$

$$1 = \sum_{k=1}^m \pi_k = \frac{\sum_{k=1}^m \sum_{i=1}^n \delta(x_i - k)}{\lambda} = \frac{\sum_{i=1}^n \sum_{k=1}^m \delta(x_i - k)}{\lambda} = \frac{\sum_{i=1}^n 1}{\lambda} = \frac{n}{\lambda}$$

Thus, we have the MLE for $\theta = [\pi_1, \cdots, \pi_m]^T$

$$\lambda = n, \quad \pi_k = \frac{\sum_{i=1}^n \delta(x_i - k)}{n} \ k \in \{1, \cdots, m\}$$

Problem 11

11. Let $X_1, \cdots, X_n$ be i.i.d. random variables with distribution $N(\mu, \sigma^2)$. Calculate the ML estimate of the parameter vector $\theta = (\mu, \sigma^2)$

solution

Since the pdf with parameter θ is $f_{X_i}(x_i \mid \theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x_i-\mu)^2}{2\sigma^2}\}$

$$f_X(x \mid \theta) = \prod_{i=1}^{n} f_{X_i}(x_i \mid \theta) = \frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left\{\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} x_i^2 - 2\left(\sum_{i=1}^{n} x_i\right)\mu + n\mu^2\right]\right\}$$

The corresponding log likelihood function is

$$\log\left(f_X(x \mid \theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} x_i^2 - 2\left(\sum_{i=1}^{n} x_i\right)\mu + n\mu^2\right]$$

Differentiate the likelihood function with parameters μ, σ^2 at $\theta = \hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$

$$\frac{\partial \log\left(f_X(x \mid \theta)\right)}{\partial \mu}\Big|_{\theta=\hat{\theta}} = -\frac{1}{2\hat{\sigma}^2}\left[-2\left(\sum_{i=1}^{n} x_i\right) + 2n\hat{\mu}\right] = 0$$

$$\frac{\partial \log\left(f_X(x \mid \theta)\right)}{\partial [\sigma^2]}\Big|_{\theta=\hat{\theta}} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2[\hat{\sigma}^2]^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = 0$$

In the end, we derive the MLE for $\theta = (\mu, \sigma^2)$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu})^2}{n}$$

Problem 12

12. Let $X_1, \cdots, X_n$ be i.i.d. random variables with distribution $N(\mu, R)$, where $\mu \in \mathbb{R}^p$ and $R \in \mathbb{R}^{p\times p}$ is a symmetric positive-definite matrix, and let $X = [X_1, \cdots, X_n]$ be the $p \times n$ matrix containing all the random vectors. Let $\theta = (\mu, R)$ denote the parameter vector for the distribution.

a) Derive the expressions for the probability density of $p(x|\theta)$ with the forms given in equations (2.6) and (2.10). (Hint: Use the trace property of equation (2.7).)
b) Compute the joint ML estimate of μ and R

solution
Since the pdf of given parameter θ is

$$\begin{aligned} f_{X_i}(x_i \mid \theta) &= \frac{1}{(2\pi)^{p/2}}|R|^{-1/2} \exp\left\{-\frac{1}{2}(x_i-\mu)^T R^{-1}(x_i-\mu)\right\} \\ &= \frac{1}{(2\pi)^{p/2}}|R|^{-1/2} \exp\left\{-\frac{1}{2}\text{tr}\left[(x_i-\mu)^T R^{-1}(x_i-\mu)\right]\right\} \\ &= \frac{1}{(2\pi)^{p/2}}|R|^{-1/2} \exp\left\{-\frac{1}{2}\text{tr}\left[R^{-1}(x_i-\mu)(x_i-\mu)^T\right]\right\} \end{aligned}$$

a) Since $X_1, \cdots, X_n$ are i.i.d.

$$\begin{aligned} f_X(x \mid \theta) = \prod_{i=1}^n f_{X_i}(x_i \mid \theta) &= \frac{1}{(2\pi)^{np/2}}|R|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[R^{-1}\sum_{i=1}^n (x_i-\mu)(x_i-\mu)^T\right]\right\} \\ &= \frac{1}{(2\pi)^{np/2}}|R|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[R^{-1}(x-\mu\cdot 1_{1\times n})(x-\mu\cdot 1_{1\times n})^T\right]\right\} \\ &= \frac{1}{(2\pi)^{np/2}}|R|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[R^{-1}xx^T\right] + \text{tr}\left[R^{-1}x\cdot 1_{n\times 1}\mu^T\right] - \frac{n}{2}\text{tr}\left[R^{-1}\mu\mu^T\right]]\right\} \\ &= \frac{1}{(2\pi)^{np/2}}|R|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[R^{-1}xx^T\right] + \mu^T R^{-1}x\cdot 1_{n\times 1} - \frac{n}{2}\mu^T R^{-1}\mu\right\} \end{aligned}$$

With substitution $b = \sum_{i=1}^n x_i = x\cdot 1_{n\times 1}$ and $S = \sum_{i=1}^n x_i x_i^T = xx^T$

$$f_X(x \mid \theta) = \frac{1}{(2\pi)^{np/2}}|R|^{-n/2} \exp\left\{-\frac{1}{2}\text{tr}\left[R^{-1}S\right] + \mu^T R^{-1}b - \frac{n}{2}\mu^T R^{-1}\mu\right\}$$

b) The corresponding joint log likelihood function is

$$\log\left(f_X(x \mid \theta)\right) = -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log|R^{-1}| - \frac{1}{2}\text{tr}\left[R^{-1}S\right] + \mu^T R^{-1}b - \frac{n}{2}\mu^T R^{-1}\mu$$

Differentiate $\log\left(f_X(x \mid \theta)\right)$ with parameters$\theta = (\mu, R)$ at $\theta = \hat{\theta} = (\hat{\mu}, \hat{R})$, notice $S = S^T = xx^T$

$$\left.\frac{\partial \log\left(f_X(x \mid \theta)\right)}{\partial \mu}\right|_{\theta=\hat{\theta}} = R^{-1}b - nR^{-1}\hat{\mu} = 0 \Longrightarrow \hat{\mu} = \frac{b}{n} = \frac{x\cdot 1_{n\times 1}}{n}$$

$$\left.\frac{\partial \log\left(f_X(x \mid \theta)\right)}{\partial [R^{-1}]}\right|_{\theta=\hat{\theta}} = \frac{n}{2}\hat{R} - \frac{1}{2}S^T + \hat{\mu}b^T - \frac{n}{2}\hat{\mu}\hat{\mu}^T = 0 \Longrightarrow \hat{R} = \frac{S - 2\hat{\mu}b^T + n\hat{\mu}\hat{\mu}^T}{n} = \frac{(x-\hat{\mu}\cdot 1_{1\times n})(x-\hat{\mu}\cdot 1_{1\times n})}{n}$$

In the end, we have

$$\hat{\mu} = \frac{x\cdot 1_{n\times 1}}{n} = \frac{\sum_{i=1}^n x_i}{n}, \quad \hat{R} = \frac{(x-\hat{\mu}\cdot 1_{1\times n})(x-\hat{\mu}\cdot 1_{1\times n})^T}{n} = \frac{\sum_{i=1}^n (x_i-\hat{\mu})(x_i-\hat{\mu})^T}{n}$$

Problem 13

13. Let X and W be independent Gaussian random vectors of dimension p such that $X \sim N(0, R_x)$ and $W \sim N(0, R_w)$, and let θ be a deterministic vector of dimension p

a) First assume that $Y = \theta + W$, and calculate the ML estimate of θ given Y
b) For the next parts, assume that $Y = X + W$, and calculate an expression for $p_{X|Y}(x \mid y)$, the conditional density of X given Y
c) Calculate the MMSE estimate of X when $Y = X + W$
d) Calculate an expression for the conditional variance of X given Y

solution
a) the ML Estimate $\hat{\theta}$ of θ is Y

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} f_Y(Y \mid \theta) \overset{\theta=Y-W}{=} Y - \underset{W}{\operatorname{argmax}} f_W(W \mid \theta) \left|\frac{\mathrm{d}W}{\mathrm{d}Y}\right| = Y - \underset{W}{\operatorname{argmax}} f_W(W)$$

$$= Y - \underset{W}{\operatorname{argmax}} \frac{1}{(2\pi)^{p/2}} |R_w|^{-1/2} \exp\left\{-\frac{1}{2} W^T R_w^{-1} W\right\} = Y$$

b) Think about the pdf relationship between Y, X, W

$$f_{Y,X,W}(y,x,w) = f_{Y,X}(y,x)\delta(y-x-w) = f_{Y,W}(y,w)\delta(y-x-w) = f_{X,W}(x,w)\delta(y-x-w)$$

So, we always have

$$f_{Y,X}(y,x) = f_{Y,W}(y,w) = f_{X,W}(x,w) \quad \text{s.t. } y-x-w=0$$

Thus, notice X, W are independent, and $X \sim N(0, R_x),\ W \sim N(0, R_w)$

$$\begin{aligned}
f_{Y,X}(y,x) &= f_{X,W}(x,w) = f_X(x) f_W(w) \quad \text{s.t. } w = y - x \\
&= \frac{1}{(2\pi)^{p/2}} |R_x|^{-1/2} \exp\left\{-\frac{1}{2} x^T R_x^{-1} x\right\} \cdot \frac{1}{(2\pi)^{p/2}} |R_w|^{-1/2} \exp\left\{-\frac{1}{2} w^T R_w^{-1} w\right\} \\
&= \frac{1}{(2\pi)^p} |R_x R_w|^{-1/2} \exp\left\{-\frac{1}{2} x^T R_x^{-1} x - \frac{1}{2}(y-x)^T R_w^{-1}(y-x)\right\} \\
&= \frac{1}{(2\pi)^p} |R_x R_w|^{-1/2} \exp\left\{-\frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} R_x^{-1} + R_w^{-1} & -R_w^{-1} \\ -R_w^{-1} & R_w^{-1} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}\right\}
\end{aligned}$$

Notice the equation below if $\mathbf{A}, \mathbf{D}$ are invertible

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix}$$

Let $\mathbf{A} = R_x^{-1} + R_w^{-1}, \mathbf{B} = \mathbf{C} = -R_w^{-1}, \mathbf{D} = R_w^{-1}$
notice $\left(R_x^{-1} + R_w^{-1}\right)^{-1} = R_x(R_x + R_w)^{-1} R_w = R_w(R_x + R_w)^{-1} R_x$

$$(R_w^{-1} - (R_x + R_w)^{-1} R_x R_w^{-1})^{-1} = ((R_x + R_w)^{-1}[R_x + R_w - R_x] R_w^{-1})^{-1} = R_x + R_w$$

$$\begin{aligned}
\begin{pmatrix} R_x^{-1} + R_w^{-1} & -R_w^{-1} \\ -R_w^{-1} & R_w^{-1} \end{pmatrix}^{-1} &= \begin{pmatrix} (R_x^{-1} + R_w^{-1} - R_w^{-1} R_w R_w^{-1})^{-1} & 0 \\ 0 & \left(R_w^{-1} - R_w^{-1}\left(R_x^{-1} + R_w^{-1}\right)^{-1} R_w^{-1}\right)^{-1} \end{pmatrix} \\
&\cdot \begin{pmatrix} I & R_w^{-1} R_w \\ R_w^{-1}\left(R_x^{-1} + R_w^{-1}\right)^{-1} & I \end{pmatrix} \\
&= \begin{pmatrix} R_x & 0 \\ 0 & R_x + R_w \end{pmatrix} \cdot \begin{pmatrix} I & I \\ (R_x + R_w)^{-1} R_x & I \end{pmatrix} \\
&= \begin{pmatrix} R_x & R_x \\ R_x & R_x + R_w \end{pmatrix}
\end{aligned}$$

With **Schur complement** if A, D are invertible, we have

$$\begin{aligned}
\begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}
\end{aligned}$$

Thus, for the determinant, we have

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \det(A - BD^{-1}C)\det(D) = \det(A)\det(D - CA^{-1}B)$$

Thus, we have

$$\begin{vmatrix} R_x & R_x \\ R_x & R_x + R_w \end{vmatrix} = |R_x| \cdot |R_x + R_w - R_x(R_x)^{-1} R_x| = |R_x R_w|$$

So, we conclude that $(X, Y) \sim N\left(0, \begin{pmatrix} R_x & R_x \\ R_x & R_x + R_w \end{pmatrix}\right)$

$$f_{X,Y}(x, y) = f_{Y,X}(y, x) = \frac{1}{(2\pi)^p} \begin{vmatrix} R_x & R_x \\ R_x & R_x + R_w \end{vmatrix}^{-1/2} \exp\left\{ -\frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} R_x & R_x \\ R_x & R_x + R_w \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} \right\}$$

We have proved that in Problem 5: let $X \sim N(\mu, R)$ be a jointly Gaussian random vector, and let $K \in \mathbb{R}^{M \times p}$ be a rank M matrix. Then the vector $Y = KX \sim N(K\mu, KRK^T)$ is also jointly Gaussian. The corresponding PDF for X and Y are

$$f_X(x) = \frac{1}{(2\pi)^{p/2}} |R|^{-1/2} \exp\left\{ -\frac{1}{2}(x - \mu)^T R^{-1} (x - \mu) \right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{M/2}} |KRK^T|^{-1/2} \exp\left\{ -\frac{1}{2}(y - K\mu)^T (KRK^T)^{-1} (y - K\mu) \right\}$$

Here we know that $X \sim N(0, R)$, so $\mu = 0$ and $Y = KX \sim N(0, KRK^T)$

$$f_X(x) = \frac{1}{(2\pi)^{p/2}} |R|^{-1/2} \exp\left\{ -\frac{1}{2} x^T R^{-1} x \right\}$$

$$f_Y(y) = \frac{1}{(2\pi)^{M/2}} |KRK^T|^{-1/2} \exp\left\{ -\frac{1}{2} y^T (KRK^T)^{-1} y \right\}$$

Let $K = [\underbrace{0, \cdots, 0}_{p\ 0}, \underbrace{1, \cdots, 1}_{p\ 1}] \in \mathbb{R}^{1\times 2p}$, then $Y = K[X;Y]$, $K \begin{pmatrix} R_x & R_x \\ R_x & R_x + R_w \end{pmatrix} K^T = R_x + R_w$

So, $Y \sim (0, R_x + R_w)$

$$f_Y(y) = \frac{1}{(2\pi)^{p/2}} |R_x + R_w|^{-1/2} \exp\left\{ -\frac{1}{2} y^T (R_x + R_w)^{-1} y \right\}$$

Thus, notice $(R_x + R_w)^{-1} = R_w^{-1} - (R_x + R_w)^{-1} R_x R_w^{-1} = R_w^{-1} - R_w^{-1}(R_x^{-1} + R_w^{-1})^{-1} R_w^{-1}$

Moreover, we have $R_x - R_x(R_x + R_w)^{-1} R_x = (R_x^{-1} + R_w^{-1})^{-1}$

$$\begin{vmatrix} R_x & R_x \\ R_x & R_x + R_w \end{vmatrix} = |R_x - R_x(R_x + R_w)^{-1} R_x| \cdot |R_x + R_w| = |(R_x^{-1} + R_w^{-1})^{-1}| \cdot |R_x + R_w|$$

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$= \frac{1}{(2\pi)^{p/2}} \left(\frac{\begin{vmatrix} R_x & R_x \\ R_x & R_x + R_w \end{vmatrix}}{|R_x + R_w|} \right)^{-1/2} \exp\left\{ -\frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} R_x^{-1} + R_w^{-1} & -R_w^{-1} \\ -R_w^{-1} & R_w^{-1}(R_x^{-1} + R_w^{-1})^{-1} R_w^{-1} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\}$$

$$= \frac{|(R_x^{-1} + R_w^{-1})^{-1}|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} I \\ -R_w^{-1}(R_x^{-1} + R_w^{-1})^{-1} \end{pmatrix} (R_x^{-1} + R_w^{-1}) \begin{pmatrix} I \\ -R_w^{-1}(R_x^{-1} + R_w^{-1})^{-1} \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} \right\}$$

$$= \frac{|(R_x^{-1} + R_w^{-1})^{-1}|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{ -\frac{1}{2} (x - x_0)^T (R_x^{-1} + R_w^{-1})(x - x_0) \right\}, \quad x_0 \equiv (R_x^{-1} + R_w^{-1})^{-1} R_w^{-1} y$$

Eventually, we obtain

$$(X \mid Y) \sim N\left((R_x^{-1} + R_w^{-1})^{-1} R_w^{-1} Y, \quad (R_x^{-1} + R_w^{-1})^{-1} \right)$$

c) Now let's compute the MMSE estimate of X when $Y = X + W$

We proved that the MMSE estimate of X when Y is given in Problem 8 a) is

$$T(Y)_{MMSE} = \underset{\hat{x} \equiv T(y)}{\operatorname{argmin}} \mathbb{E}[C(x, \hat{x})] \Big|_{C(x,\hat{x}) = |x - \hat{x}|^2} = \mathbb{E}[X \mid Y]$$

So, we know that

$$\hat{X}_{MMSE} = T(Y)_{MMSE} = \mathbb{E}[X \mid Y] = (R_x^{-1} + R_w^{-1})^{-1} R_w^{-1} Y$$

d) The conditional variance of X given Y is

$$\operatorname{Cov}[X \mid Y] = (R_x^{-1} + R_w^{-1})^{-1}$$

Problem 14

14. Show that if X and Y are jointly Gaussian random vectors, then the conditional distribution of X given Y is also Gaussian.

solution

Let consider the cases with $\mu_x = 0, \mu_y = 0$ first

Denote covariance by $R_x = \mathbb{E}[XX^T], R_{xy} = \mathbb{E}[XY^T], R_y = \mathbb{E}[YY^T]$, where $X \in \mathbb{R}^{m\times 1}, Y \in \mathbb{R}^{n\times 1}$

$$f_{X,Y}(x,y) = \frac{1}{(2\pi)^{\frac{m+n}{2}}} \begin{vmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{vmatrix}^{-1/2} \exp\left\{ -\frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} \right\}$$

Notice the equation below if $\mathbf{A}, \mathbf{D}$ are invertible

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix}$$

Let $\mathbf{A} = R_x, \mathbf{B} = R_{xy}, \mathbf{C} = R_{xy}^T, \mathbf{D} = R_y$

$$\begin{aligned} \begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}^{-1} &= \begin{pmatrix} (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} & 0 \\ 0 & (R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} \end{pmatrix} \cdot \begin{pmatrix} I & -R_{xy}R_y^{-1} \\ -R_{xy}^T R^{-1} & I \end{pmatrix} \\ &= \begin{pmatrix} (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} & -(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1} \\ -(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} R_{xy}^T R_x^{-1} & (R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} \end{pmatrix} \end{aligned}$$

Consider **Woodbury matrix identity**

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1}$$

Let $A = R_y,\ C = R_x^{-1},\ U = -R_{xy}^T, V = R_{xy}$, and since the matrix is symmetrical, we have

$$(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} = R_y^{-1} + R_y^{-1} R_{xy}^T (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1}$$

$$-(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} R_{xy}^T R_x^{-1} = -\left[(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1}\right]^T$$

Rewrite the inverse matrix as

$$\begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}^{-1} = \begin{pmatrix} (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} & -(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1} \\ -\left[(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1}\right]^T & R_y^{-1} + R_y^{-1} R_{xy}^T (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1} \end{pmatrix}$$

Moreover, we have

$$\begin{aligned} &\begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & R_y^{-1} \end{pmatrix} \\ &= \begin{pmatrix} (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} & -(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1} \\ -\left[(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1}\right]^T & R_y^{-1} R_{xy}^T (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} R_{xy} R_y^{-1} \end{pmatrix} \\ &= \begin{pmatrix} I \\ -\left[R_{xy}R_y^{-1}\right]^T \end{pmatrix} \cdot (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1} \begin{pmatrix} I \\ -\left[R_{xy}R_y^{-1}\right]^T \end{pmatrix}^T \end{aligned} \tag{1}$$

With **Schur complement** if A, D are invertible, we have

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \end{aligned}$$

Thus, for the determinant, we have

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \det(A - BD^{-1}C)\det(D) = \det(A)\det(D - CA^{-1}B)$$

Let $A = R_x,\ B = R_{xy},\ C = R_{xy}^T,\ D = R_y$, we have

$$\begin{vmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{vmatrix} = |R_x - R_{xy}R_y^{-1}R_{xy}^T| \cdot |R_y| \tag{2}$$

Since the pdf of Y is

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{n}{2}}}|R_y|^{-1/2}\exp\left\{-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} 0 & 0 \\ 0 & R_y^{-1} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right\}$$

With (1), (2), we conclude that

$$f_{X|Y}(x \mid y) = \frac{1}{(2\pi)^{\frac{m}{2}}}\left(\frac{\begin{vmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{vmatrix}}{|R_y|}\right)^{-1/2} \exp\left\{-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \left[\begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & R_y^{-1} \end{pmatrix}\right]\begin{pmatrix} x \\ y \end{pmatrix}\right\}$$

$$= \frac{|R_x - R_{xy}R_y^{-1}R_{xy}^T|^{-1/2}}{(2\pi)^{\frac{m}{2}}}\exp\left\{-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} I \\ -\left[R_{xy}R_y^{-1}\right]^T \end{pmatrix} \cdot (R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1}\begin{pmatrix} I \\ -\left[R_{xy}R_y^{-1}\right]^T \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix}\right\}$$

$$= \frac{|R_x - R_{xy}R_y^{-1}R_{xy}^T|^{-1/2}}{(2\pi)^{\frac{m}{2}}}\exp\left\{-\frac{1}{2}(x - x_0)^T(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1}(x - x_0)\right\}, \quad x_0 \equiv R_{xy}R_y^{-1}y$$

So, we prove that $(X \mid Y)$ with zero means $\mu_x = 0, \mu_y = 0$ follows Gaussian distribution

$$(X \mid Y) \sim N\left(R_{xy}R_y^{-1}Y, \quad R_x - R_{xy}R_y^{-1}R_{xy}^T\right)$$

Then let's introduce μ_x, μ_y now, replace x, y with $x \leftarrow x - \mu_x,\ y \leftarrow y - \mu_y$ in the pdf
We obtain the following equation, where $x_0 \equiv R_{xy}R_y^{-1}(y - \mu_y)$

$$f_{X|Y}(x \mid y) = \frac{|R_x - R_{xy}R_y^{-1}R_{xy}^T|^{-1/2}}{(2\pi)^{\frac{m}{2}}}\exp\left\{-\frac{1}{2}(x - \mu_x - x_0)^T(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1}(x - \mu_x - x_0)\right\}$$

$(X \mid Y)$ follows Gaussian distribution

$$(X \mid Y) \sim N\left(\mu_x + R_{xy}R_y^{-1}(Y - \mu_y), \quad R_x - R_{xy}R_y^{-1}R_{xy}^T\right)$$

Problem 15

15. Show that if $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ are jointly Gaussian random vectors, then $\mathbb{E}[X \mid Y] = AY$ and

$$\mathbb{E}\left[(X - \mathbb{E}[X \mid Y])(X - \mathbb{E}[X \mid Y])^T \mid Y\right] = C$$

Where $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{m\times n}$ is positive definite

solution

As proved in Problem 14, for X, Y with $\mu_x = 0, \mu_y = 0$

Denote covariance by $R_x = \mathbb{E}[XX^T], R_{xy} = \mathbb{E}[XY^T], R_y = \mathbb{E}[YY^T]$, where $X \in \mathbb{R}^{m\times 1}, Y \in \mathbb{R}^{n\times 1}$

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^{\frac{m+n}{2}}} \begin{vmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{vmatrix}^{-1/2} \exp\left\{-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix}\right\}$$

$(X \mid Y)$ with zero means $\mu_x = 0, \mu_y = 0$ follows Gaussian distribution, where $x_0 \equiv R_{xy}R_y^{-1}y$

$$(X \mid Y) \sim N\left(R_{xy}R_y^{-1}Y, \quad R_x - R_{xy}R_y^{-1}R_{xy}^T\right)$$

$$f_{X|Y}(x \mid y) = \frac{|R_x - R_{xy}R_y^{-1}R_{xy}^T|^{-1/2}}{(2\pi)^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}(x - x_0)^T(R_x - R_{xy}R_y^{-1}R_{xy}^T)^{-1}(x - x_0)\right\}$$

So, we know the follows, where $A = R_{xy}R_y^{-1}, \quad C = R_x - R_{xy}R_y^{-1}R_{xy}^T$

$$\mathbb{E}[X \mid Y] = AY = R_{xy}R_y^{-1}Y$$
$$\mathbb{E}\left[(X - \mathbb{E}[X \mid Y])(X - \mathbb{E}[X \mid Y])^T \mid Y\right] = C = R_x - R_{xy}R_y^{-1}R_{xy}^T$$

Let's prove C is positive definite now! With **Schur complement** if A, D are invertible, we have

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$
$$= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

Let $A = R_x,\ B = R_{xy},\ C = R_{xy}^T,\ D = R_y$, we have

$$\begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix} = \begin{pmatrix} I & R_{xy}R_y^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} R_x - R_{xy}R_y^{-1}R_{xy}^T & 0 \\ 0 & R_y \end{pmatrix}\begin{pmatrix} I & 0 \\ R_y^{-1}R_{xy}^T & I \end{pmatrix}$$
$$= \begin{pmatrix} I & 0 \\ R_y^{-1}R_{xy}^T & I \end{pmatrix}^T\begin{pmatrix} R_x - R_{xy}R_y^{-1}R_{xy}^T & 0 \\ 0 & R_y \end{pmatrix}\begin{pmatrix} I & 0 \\ R_y^{-1}R_{xy}^T & I \end{pmatrix}$$

Because $\begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}$ is positive definite, for $\forall z = (z_1, \cdots, z_{m+n})^T \in \mathbb{R}^{(m+n)\times 1}$

$$z^T\begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}z = z^T\mathbb{E}\left[\begin{pmatrix} X \\ Y \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}^T\right]z = \mathbb{E}\left[z^T\begin{pmatrix} X \\ Y \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}^T z\right] = \mathbb{E}\left[\left(z^T \cdot \begin{pmatrix} X \\ Y \end{pmatrix}\right)^2\right] > 0$$

Since $\begin{pmatrix} I & 0 \\ R_y^{-1}R_{xy}^T & I \end{pmatrix}$ is full rank, do transform $z' = \begin{pmatrix} I & 0 \\ R_y^{-1}R_{xy}^T & I \end{pmatrix}z$, space of z' is also $\mathbb{R}^{(m+n)\times 1}$

$$(z')^T\begin{pmatrix} R_x - R_{xy}R_y^{-1}R_{xy}^T & 0 \\ 0 & R_y \end{pmatrix}z' = \mathbb{E}\left[\left(z^T \cdot \begin{pmatrix} X \\ Y \end{pmatrix}\right)^2\right] > 0$$

It holds for any $z' \in \mathbb{R}^{(m+n)\times 1}$, we prove that $C = R_x - R_{xy}R_y^{-1}R_{xy}^T$ is positive definite

Problem 16

16. Let $Y \in \mathbb{R}^m$ and $X \in \mathbb{R}^n$ be zero-mean jointly Gaussian random vectors. Then define the following notation for this problem. Let $p(y,x)$ and $p(y|x)$ be the joint and conditional density of Y given X. Let B be the joint positive-definite precision matrix (i.e., inverse covariance matrix) given by $B^{-1} = \mathbb{E}[ZZ^T]$ where $Z = \begin{pmatrix} Y \\ X \end{pmatrix}$ Furthermore, let C, D, and E be the matrix blocks that form B, so that

$$B = \begin{pmatrix} C & D \\ D^T & E \end{pmatrix}$$

where $C \in \mathbb{R}^{m\times m}$, $D \in \mathbb{R}^{m\times n}$, and $E \in \mathbb{R}^{n\times n}$. Finally, define the matrix A so that $AX = \mathbb{E}[Y \mid X]$, and define the matrix

$$\Lambda^{-1} = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X])(Y - \mathbb{E}[Y \mid X])^T \mid X\right]$$

a) Write out an expression for $p(y,x)$ in terms of B
b) Write out an expression for $p(y \mid x)$ in terms of A and Λ
c) Derive an expression for Λ in terms of C, D, and E
d) Derive an expression for A in terms of C, D, and E

solution

As proved in Problem 14, 15, we can obtain these conclusions by exchange $X \leftrightarrow Y$

$$B = \begin{pmatrix} R_y & R_{xy}^T \\ R_{xy} & R_x \end{pmatrix}^{-1} = \begin{pmatrix} (R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} & -(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} R_{xy}^T R_x^{-1} \\ -\left[(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} R_{xy}^T R_x^{-1}\right]^T & R_x^{-1} + R_x^{-1} R_{xy}(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} R_{xy}^T R_x^{-1} \end{pmatrix}$$

$$C = (R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1}$$
$$D = -(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} R_{xy}^T R_x^{-1}$$
$$E = R_x^{-1} + R_x^{-1} R_{xy}(R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1} R_{xy}^T R_x^{-1}$$

Denote covariance by $R_x = \mathbb{E}[XX^T], R_{xy} = \mathbb{E}[XY^T], R_y = \mathbb{E}[YY^T]$, where $X \in \mathbb{R}^{n\times 1}, Y \in \mathbb{R}^{m\times 1}$

$$f_{Y,X}(y,x) = \frac{1}{(2\pi)^{\frac{m+n}{2}}} \begin{vmatrix} R_y & R_{xy}^T \\ R_{xy} & R_x \end{vmatrix}^{-1/2} \exp\left\{-\frac{1}{2}\begin{pmatrix} y \\ x \end{pmatrix}^T \begin{pmatrix} R_y & R_{xy}^T \\ R_{xy} & R_x \end{pmatrix}^{-1} \begin{pmatrix} y \\ x \end{pmatrix}\right\}$$

$(X \mid Y)$ with zero means $\mu_x = 0, \mu_y = 0$ follows Gaussian distribution, where $y_0 \equiv R_{xy}^T R_x^{-1} x$

$$(Y \mid X) \sim N\left(R_{xy}^T R_x^{-1} X, \quad R_y - R_{xy}^T R_x^{-1} R_{xy}\right)$$

$$f_{Y|X}(y \mid x) = \frac{|R_y - R_{xy}^T R_x^{-1} R_{xy}|^{-1/2}}{(2\pi)^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}(y - y_0)^T (R_y - R_{xy}^T R_x^{-1} R_{xy})^{-1}(y - y_0)\right\}$$

So, for A, Λ, we have

$$A = R_{xy}^T R_x^{-1}$$
$$\Lambda = \left(R_y - R_{xy}^T R_x^{-1} R_{xy}\right)^{-1}$$

a) Write out an expression for $p(y,x)$ in terms of B

$$f_{Y,X}(y,x) = \frac{1}{(2\pi)^{\frac{m+n}{2}}} |B|^{1/2} \exp\left\{-\frac{1}{2}\begin{pmatrix} y \\ x \end{pmatrix}^T B \begin{pmatrix} y \\ x \end{pmatrix}\right\}$$

b) Write out an expression for $p(y \mid x)$ in terms of A and Λ

$$f_{Y|X}(y \mid x) = \frac{|\Lambda|^{1/2}}{(2\pi)^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}(y - Ax)^T \Lambda (y - Ax)\right\}$$

c) Derive an expression for Λ in terms of C, D, and E

$$\Lambda = C$$

d) Derive an expression for A in terms of C, D, and E

$$A = -C^{-1} D$$

Problem 17

17. Let Y and X be random variables, and let Y_{MAP} and Y_{MMSE} be the MAP and MMSE estimates respectively of Y given X. Pick distributions for Y and X so that the MAP estimator is very "poor", but the MMSE estimator is "good"

solution

Consider $X = Y + W$, and Y, W are mutually independent, where W follows the pdf

$$f_W(w) = \frac{1-\epsilon}{\sqrt{2\pi}} \exp\left\{-\frac{w^2}{2}\right\} + \frac{\epsilon}{\sqrt{2\pi}\epsilon} \exp\left\{-\frac{(w-\frac{1}{\epsilon})^2}{2\epsilon^2}\right\}$$

Set $Y \sim N(0, 1^2)$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\}$$

As proved in Problem 13 b), we can compute the pdf of $(Y \mid X)$

$$\begin{aligned} f_{Y|X}(y \mid x) = (1-\epsilon)&\frac{\left|\left(1^{-1}+1^{-1}\right)^{-1}\right|^{-1/2}}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\left(1^{-1}+1^{-1}\right)(y-y_0)^2\right\} \\ &+ (\epsilon)\cdot\frac{\left|\left(1^{-1}+\frac{1}{\epsilon^2}\right)^{-1}\right|^{-1/2}}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\left(1^{-1}+\frac{1}{\epsilon^2}\right)(y-y_1)^2\right\} \\ &\text{where}\quad y_0 \equiv \left(1^{-1}+1^{-1}\right)^{-1}1^{-1}x = \frac{x}{2},\quad y_1 \equiv \frac{1}{\epsilon} + \left(1^{-1}+\frac{1}{\epsilon^2}\right)^{-1}\frac{1}{\epsilon^2}x = \frac{1}{\epsilon} + \frac{x}{1+\epsilon^2}, \end{aligned}$$

Let $\epsilon \to 0_+$

$\text{MAP} = \underset{y}{\operatorname{argmin}}\, f_{Y|X}(y \mid X) = \frac{1}{\epsilon} + \frac{X}{1+\epsilon^2}$ is close to infinity $\to +\infty$

$\text{MMSE} = \mathbb{E}[Y \mid X] = (1-\epsilon)\frac{X}{2} + \epsilon[\frac{1}{\epsilon} + \frac{X}{1+\epsilon^2}]$ is close to $\to \frac{X}{2} + 1$

Problem 18

18. Prove that two zero-mean discrete-time Gaussian random processes have the same distribution, if and only if they have the same time autocovariance function.

solution

Because it is Gaussian random process, write down the pdf, where $\mathbb{E}[(X_m, \cdots, X_{m+k}) = 0_{1\times(k+1)}$ $(X_m, \cdots, X_{m+k})$ and $(Y_m, \cdots, Y_{m+k})$ have the same distribution for all $\forall k, m \in \mathbb{Z}^*$

$$f_{X_m,\cdots,X_{m+k}}(c_0, \cdots, c_k) = \frac{|\text{Cov}[(X_m, \cdots, X_{m+k})^T(X_m, \cdots, X_{m+k})]|^{-1/2}}{(2\pi)^{k/2}}$$
$$\exp\Big\{ -\frac{1}{2}\left[(c_0, \cdots, c_k)^T - \mathbb{E}[(X_m, \cdots, X_{m+k})^T\right]$$
$$\cdot \left[\text{Cov}[(X_m, \cdots, X_{m+k})^T(X_m, \cdots, X_{m+k})]\right]^{-1}\left[(c_0, \cdots, c_k) - \mathbb{E}[(X_m, \cdots, X_{m+k})]\right]\Big\}$$
$$= \frac{|\text{Cov}[(X_m, \cdots, X_{m+k})^T(X_m, \cdots, X_{m+k})]|^{-1/2}}{(2\pi)^{k/2}}$$
$$\exp\left\{-\frac{1}{2}\left[(c_0, \cdots, c_k)^T\right]\left[\text{Cov}[(X_m, \cdots, X_{m+k})^T(X_m, \cdots, X_{m+k})]\right]^{-1}\left[(c_0, \cdots, c_k)\right]\right\}$$
$$= f_{Y_m,\cdots,Y_{m+k}}(c_0, \cdots, c_k)$$
$$= \frac{|\text{Cov}[(Y_m, \cdots, Y_{m+k})^T(Y_m, \cdots, Y_{m+k})]|^{-1/2}}{(2\pi)^{k/2}}$$
$$\exp\left\{-\frac{1}{2}\left[(c_0, \cdots, c_k)^T\right]\left[\text{Cov}[(Y_m, \cdots, Y_{m+k})^T(X_m, \cdots, X_{m+k})]\right]^{-1}\left[(c_0, \cdots, c_k)\right]\right\}$$

Which is equivalent to

$$|\text{Cov}[(X_m, \cdots, X_{m+k})^T(X_m, \cdots, X_{m+k})]| = |\text{Cov}[(Y_m, \cdots, Y_{m+k})^T(Y_m, \cdots, Y_{m+k})]|$$
$$\text{Cov}[(X_m, \cdots, X_{m+k})^T(X_m, \cdots, X_{m+k})]^{-1} = \text{Cov}[(Y_m, \cdots, Y_{m+k})^T(Y_m, \cdots, Y_{m+k})]^{-1}$$

That is equivalent to, it holds for all $\forall k, m \in \mathbb{Z}^*$

$$\text{Cov}[(X_m, \cdots, X_{m+k})^T(X_m, \cdots, X_{m+k})] = \text{Cov}[(Y_m, \cdots, Y_{m+k})^T(Y_m, \cdots, Y_{m+k})]$$

Problem 19

19. Prove all Gaussian wide-sense stationary random processes are:

a) strict-sense stationary
b) reversible

solution

a) **strict-sense stationary**
Definition: for any fixed $k \geq 0$, all $\forall m \in \mathbb{Z}$, it doesn't change for any fixed $(c_1, \cdots, c_k) \in \mathbb{R}^{k\times 1}$

$$F_{X_m,\cdots,X_{m+k}}(c_0, \cdots, c_k) = F_k(c_0, \cdots, c_k)$$

We only have to prove the follows for pdf

$$f_{X_m,\cdots,X_{m+k}}(c_0, \cdots, c_k) = f_k(c_0, \cdots, c_k)$$

Start proof! Freeze k, we know that for wide-sense stationary random processes

$$\mathbb{E}[X_{m+i}] = \mu, \quad \mathbb{E}[X_{m+i}X_{m+j}] - \mu^2 = R(|i-j|) \quad \forall i, j \in \{0, \cdots, k\}$$

$$\begin{aligned}
R^{(m,k)} &\equiv \mathrm{Cov}[(X_m, \cdots, X_{m+k})^T (X_m, \cdots, X_{m+k})] \\
&= \mathbb{E}[(X_m, \cdots, X_{m+k})^T (X_m, \cdots, X_{m+k})] - \mathbb{E}[(X_m, \cdots, X_{m+k})]^T \mathbb{E}[(X_m, \cdots, X_{m+k})] \\
&= \mathbb{E}[(X_m, \cdots, X_{m+k})^T (X_m, \cdots, X_{m+k})] - \mu^2 \cdot 1_{k\times k} \\
&= \begin{pmatrix} R(0) & R(1) & \cdots & R(k) \\ R(1) & R(0) & \cdots & R(k-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(k) & R(k-1) & \cdots & R(0) \end{pmatrix} = R^{(k)}
\end{aligned}$$

Because it is Gaussian random process, we can write down the pdf

$$\begin{aligned}
&f_{X_m,\cdots,X_{m+k}}(c_0, \cdots, c_k) = \frac{|\mathrm{Cov}[(X_m, \cdots, X_{m+k})^T (X_m, \cdots, X_{m+k})]|^{-1/2}}{(2\pi)^{k/2}} \\
&\exp\Big\{ -\frac{1}{2} \big[(c_0, \cdots, c_k)^T - \mathbb{E}[(X_m, \cdots, X_{m+k})^T]\big] \\
&\cdot \big[\mathrm{Cov}[(X_m, \cdots, X_{m+k})^T (X_m, \cdots, X_{m+k})\big]^{-1} \big[(c_0, \cdots, c_k) - \mathbb{E}[(X_m, \cdots, X_{m+k})]\big] \Big\} \\
&= \frac{|R^{(k)}|^{-1/2}}{(2\pi)^{k/2}} \exp\left\{ -\frac{1}{2} \left[(c_0, \cdots, c_k)^T - \mu \cdot 1_{1\times(k+1)}\right] \left[R^{(k)}\right]^{-1} \left[(c_0, \cdots, c_k) - \mu \cdot 1_{(k+1)\times 1}\right] \right\} \\
&= f_k(c_0, \cdots, c_k)
\end{aligned}$$

pdf $f_{X_m,\cdots,X_{m+k}}(c_1, \cdots, c_k) = f_k(c_0, \cdots, c_k)$ is NOT function of m, $\forall k \in \mathbb{Z}^*, \forall (c_0, \cdots, c_k) \in \mathbb{R}^{1\times k}$

---

b) **reversible**
We need to prove that for $\forall k \in \mathbb{Z}^*, \forall (c_1, \cdots, c_k) \in \mathbb{R}^{1\times k}$

$$f_{X_m,\cdots,X_{m+k}}(c_0, \cdots, c_k) = f_{X_m,\cdots,X_{m+k}}(c_k, \cdots, c_0)$$

Actually,

$$
\begin{aligned}
\operatorname{Cov}[(X_m, \cdots, X_{m+k})^T (X_m, \cdots, X_{m+k})] &= \operatorname{Cov}[(X_{m+k}, \cdots, X_m)^T (X_{m+k}, \cdots, X_m)] \\
&= \begin{pmatrix} R(0) & R(1) & \cdots & R(k) \\ R(1) & R(0) & \cdots & R(k-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(k) & R(k-1) & \cdots & R(0) \end{pmatrix} = R^{(k)}
\end{aligned}
$$

$$
f_{X_m, \cdots, X_{m+k}}(c_0, \cdots, c_k) = f_{X_m, \cdots, X_{m+k}}(c_k, \cdots, c_0) = f_k(c_0, \cdots, c_k) = f_k(c_k, \cdots, c_0)
$$

Problem 20

20. Construct an example of a strict-sense stationary random process that is not reversible.

solution
Let's assume random process $X_n, n \in \mathbb{Z}$ iid, and follows

$$P(X_n = 1) = p, \quad P(X_n = 0) = 1 - p, \quad p \in (0,1)$$

Think about random Process Y_n as a "register" with 2 bits to store information of X_n

$$Y_n = (X_{n-1}X_n)_2 = 2^1 \cdot X_{n-1} + 2^0 \cdot X_n$$

Proof of **strict-sense stationary**
We can compute the Probability for any $[Y_m, \cdots, Y_{m+k}]^T$ where $m, k \in \mathbb{Z}, k \geq 0$

$$\begin{aligned}
&P_{m,m+k}\left((c_{-1}c_0)_2, \cdots, (c_{k-1}c_k)_2\right) \equiv P\{Y_m = (c_{-1}c_0)_2, \cdots, Y_{m+k} = (c_{k-1}c_k)_2\} \\
&= P\{X_{m-1} = c_{-1}, X_m = c_0, \cdots, X_{m+k} = c_k\} \\
&= \prod_{k'=-1}^{k} P\{X_{m+k'} = c_{k'}\} \\
&= p^{\sum_{k'=-1}^{k} \delta(c_{k'})} \cdot (1-p)^{\sum_{k'=-1}^{k} \delta(c_{k'}-1)} \\
&= \prod_{k'=-1}^{k} P\{X_{k'} = c_{k'}\} \\
&= P\{X_{-1} = c_{-1}, X_0 = c_0, \cdots, X_k = c_k\} \\
&= P\{Y_0 = (c_{-1}c_0)_2, \cdots, Y_k = (c_{k-1}c_k)_2\} \\
&\equiv P_{0,k}\left((c_{-1}c_0)_2, \cdots, (c_{k-1}c_k)_2\right)
\end{aligned}$$

Where $c_{k'} \in \{1, 0\}, k' = \{-1, 0, \cdots, k\}$. Other than that, when $[Y_m, \cdots, Y_{m+k}]^T \neq [(c_{-1}c_0)_2, \cdots, (c_{k-1}c_k)_2]^T$, we have $P_{m,m+k} = P_{0,k} = 0$
Thus, we prove that for $\forall m, k \in \mathbb{Z}, k \geq 0$, and $\forall [t_0, \cdots, t_k]^T \in \mathbb{R}^{(k+1)\times 1}$

$$\begin{aligned}
&F_{m,m+k}(t_0, t_1, \cdots, t_k) = F_{0,k}(t_0, t_1, \cdots, t_k) \\
&= \sum_{t'_0=-\infty}^{t_0} \cdots \sum_{t'_k=-\infty}^{t_k} P_{m,m+k}(t'_0, \cdots, t'_k) = \sum_{t'_0=-\infty}^{t_0} \cdots \sum_{t'_k=-\infty}^{t_k} P_{0,k}(t'_0, \cdots, t'_k)
\end{aligned}$$

Proof of **NOT reversible**

Set $m = 0, k = 2$ and $[c_{-1}, c_0, c_1] = [1, 1, 0]^T$

$$
\begin{aligned}
&P_{0,k}\left((c_{-1}c_0)_2, (c_0c_1)_2\right) \equiv P\{Y_0 = (c_{-1}c_0)_2, Y_1 = (c_0c_1)_2\} \\
&= P\{X_{-1} = c_{-1}, X_0 = c_0, \cdots, X_1 = c_1\} \\
&= \prod_{k'=-1}^{1} P\{X_{k'} = c_{k'}\} \\
&= p^2(1-p) \\
&\neq 0 \\
&= P\{X_{-1} = 1, X_0 = 0, X_0 = 1, X_1 = 1\} \quad X_0 = 0, X = 1 \text{ is Impossible} \\
&= P\{X_{-1} = c_0, X_0 = c_1, X_0 = c_{-1}, X_1 = c_0\} \\
&= P\{Y_0 = (c_0c_1)_2, Y_1 = (c_{-1}c_0)_2\} \\
&\equiv P_{0,k}\left((c_0c_1)_2, (c_{-1}c_0)_2\right)
\end{aligned}
$$

So, it is **NOT** possible to have

$$F_{0,k}\left(t_0, t_1, \cdots, t_k\right) = F_{m,m+k}\left(t_k, t_{k-1}, \cdots, t_0\right)$$

for $\forall m, k \in \mathbb{Z}, k \geq 0$, and $\forall [t_0, \cdots, t_k]^T \in \mathbb{R}^{(k+1)\times 1}$