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

Problem 1

1. Let $\{X_n\}_{n=1}^N$ be a 1-D Gaussian random process such that

$$\mathcal{E}_n = X_n - \sum_{i=n-P}^{n-1} h_{n-i} X_i$$

results in $\mathcal{E}_n$ being a sequence of i.i.d. $N(0, \sigma^2)$ random variables for $n = 1, \cdots, N$, and assume that $X_n = 0$ for $n \leq 0$. Compute the ML estimates for the prediction filter h_n and the prediction variance σ^2

solution
Step 1: As proved in Chapter 2 Problem 14, 15:
for X, Y with $\mu_x = 0, \mu_y = 0$, if $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ are jointly Gaussian random vectors
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 $A = R_{xy}R_y^{-1}, \quad C = R_x - R_{xy}R_y^{-1}R_{xy}^T$

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

Note that $R_x - R_{xy}R_y^{-1}R_{xy}^T$ is positive definite, when X and Y are linear independent.

Step 2: As proved in Chapter 2 Problem 8:
The best estimator $\hat{x} = T(y)$ for x when cost function $C(x, \hat{x}) = |x - \hat{x}|^2$ (Mean Square Error) is, by minimizing the functional: the Bayes' Risk $\mathcal{J}(T) = \mathbb{E}[C(x, \hat{x})]$

$$\delta\mathcal{J}(T) = \delta\mathbb{E}[C(x, \hat{x})] = \delta\mathbb{E}[|x - \hat{x}|^2] = 0 \Longrightarrow \hat{x} = T(y) = \mathbb{E}[X|y]$$

Thus, Minimal Mean Square Error (MMSE) estimator it $\hat{X} = T(Y) = \mathbb{E}[X|Y]$

Step 3: Short proof for Property 2.3 and Property 2.4 in Chapter 2 of textbook MBIP
Property 2.3: for random variable vector X, Y, Z, where $f(\cdot)$ is PDF of random variables

$$\int\left[\int \vec{x}\cdot f(\vec{x} \mid \vec{y}, \vec{z})\mathrm{d}\vec{x}\right] f(\vec{z} \mid \vec{y})\mathrm{d}\vec{z} = \int\int \vec{x}\cdot f(\vec{x}, \vec{z} \mid \vec{y})\mathrm{d}\vec{z}\mathrm{d}\vec{x} = \int \vec{x}\cdot\left[\int f(\vec{x}, \vec{z} \mid \vec{y})\mathrm{d}\vec{z}\right]\mathrm{d}\vec{x} = \int \vec{x}\cdot f(\vec{x} \mid \vec{y})\mathrm{d}\vec{x}$$

It hold for any $Y = \vec{y}$, so

$$\mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Y] = \mathbb{E}[X \mid Y]$$

Property 2.4: for random variable vector X, Y, Z, and any function $g(X)$ for X

$$\int \vec{z} \cdot g(\vec{x}) \cdot f(\vec{z} \mid \vec{x}, \vec{y}) \mathrm{d}\vec{z} = \left[\int \vec{z} \cdot f(\vec{z} \mid \vec{x}, \vec{y}) \mathrm{d}\vec{z} \right] \cdot g(\vec{x})$$

It holds for any $X = \vec{x}, Y = \vec{y}$, so

$$\mathbb{E}[g(X)Z \mid X, Y] = g(X)\mathbb{E}[Z \mid X, Y]$$

Set $Z = \{1\}$ with PDF $f(z \mid \vec{x}, \vec{y}) = f(\vec{z}) = \delta(\vec{z} - 1)$

$$\int z \cdot g(\vec{x}) \cdot \delta(z - 1) \mathrm{d}z = \left[\int z \cdot \delta(z - 1) \mathrm{d}z \right] \cdot g(\vec{x}) = g(\vec{x})$$

So, we have

$$\mathbb{E}[g(X) \mid X, Y] = g(X)$$

Set $\{g_k(X)\}_{k=1}^{p} = X \in \mathbb{R}^p$, and $Z = \{1\}$ with PDF $f(z \mid \vec{x}, \vec{y}) = f(\vec{z}) = \delta(\vec{z} - 1)$

$$\int z \cdot \vec{x} \cdot \delta(z - 1) \mathrm{d}z = \left[\int z \cdot \delta(z - 1) \mathrm{d}z \right] \cdot \vec{x} = \vec{x}$$

So, we have

$$\mathbb{E}[X \mid X, Y] = X$$

Especially, consider $X \to Y, Y \to Z, \mathbb{E}[X \mid Y] = g(Y)$

$$\mathbb{E}[\mathbb{E}[X \mid Y] \mid Y, Z] = \mathbb{E}[g(Y) \mid Y, Z] = g(Y) = \mathbb{E}[X \mid Y]$$

It is equivalent to below, consider random variable $W = \{1\}$ with $f(w \mid y, z) = f(w) = \delta(w - 1)$

$$\int \left[\int \vec{x} \cdot f(x \mid y) \mathrm{d}\vec{x} \right] f(w \mid y, z) \mathrm{d}w = \left[\int \vec{x} \cdot f(x \mid y) \mathrm{d}\vec{x} \right]$$

Step 4: As proved in hw 1, Chapter 2 problem 4.
Let X be a jointly Gaussian random vector, and let $A \in \mathbb{R}^{M \times N}$ be a rank M matrix. Then prove that the vector $Y = AX$ is also jointly Gaussian.
Since $M = \mathrm{rank}(A) \le \min(M, N)$, we have $M \le N$
Since X is a jointly Gaussian random vector, the PDF $f_X(x)$ of X is given by

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

Where the mean vector $\mu \equiv \mathbb{E}[X]$, and symmetric positive-definite covariance $R = R^T \equiv \mathbb{E}\left[(X - \mu)(X - \mu)^T\right]$

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

So, we prove that the vector $Y = AX$ is also jointly Gaussian.

$$Y \sim N(A\mu, ARA^T)$$

Step 5: As proved in hw 1, Chapter 2 problem 5.
Let $X \sim N(0, R)$ where R is a $p \times p$ symmetric positive-definite matrix.

a) Prove that if for all $i \neq j, \mathbb{E}[X_i X_j] = 0$ (X_i and X_j are components of X), then X_i and X_j are pair-wise independent.
b) Prove that if for all i, j, X_i and X_j are uncorrelated (X_i and X_j are components of X), then the components of X are jointly independent.

a) Because all i, j, X_i and X_j are uncorrelated, we have $R_{ij} = 0$ for $i \neq j$, denote $\mathbb{E}\left[X_k^2\right] = \sigma_k^2$ for $k \in \{1, \cdots, p\}$

$$R = \operatorname{diag}\left(\sigma_1^2, \cdots, \sigma_p^2\right), \quad R^{-1} = \operatorname{diag}\left(\frac{1}{\sigma_1^2}, \cdots, \frac{1}{\sigma_p^2}\right), \quad |R|^{-1/2} = \prod_{k=1}^{p} \frac{1}{\sigma_k}$$

Thus, for the PDF of X: $f_X(x) = \prod_{k=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{x_k^2}{2\sigma_k^2}\right\}$. For $k \in \{1, \cdots, p\}$, we have

$$f_X(x_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{x_k^2}{2\sigma_k^2}\right\} \cdot \prod_{k'=1, k' \neq k}^{p} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_{k'}} \exp\left\{-\frac{x_{k'}^2}{2\sigma_{k'}^2}\right\} \mathrm{d}x_{k'} = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{x_k^2}{2\sigma_k^2}\right\}$$

For $X_i, X_j, i \neq j$, we have

$$f_X(x_i, x_j) = \frac{1}{2\pi\sigma_i\sigma_j} \exp\left\{-\frac{x_i^2}{2\sigma_i^2} - \frac{x_j^2}{2\sigma_j^2}\right\} = f_X(x_i) \cdot f_X(x_j)$$

So, we prove that if for all $i \neq j, \mathbb{E}[X_i X_j] = 0$ (i.e., X_i and X_j are uncorrelated), then X_i and X_j are pair-wise independent.

b) Similarly, we conclude

$$f_X(x_1, \cdots, x_p) = \prod_{k=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{x_k^2}{2\sigma_k^2}\right\} = \prod_{k=1}^{p} f_X(x_k)$$

In the end, we prove that if for all i, j, X_i and X_j are uncorrelated, then the components of X are jointly independent.

Step 6.a: Combine Step 1,2, notice that for random process $\{X_n\}_{n=1}^N$ with zero-mean $\mathbb{E}[X_n] = 0$, we know that X_n and $\{X_i \text{ for } i < n\}$ are jointly Gaussian random vectors.
Define the **causal predictor** as MMSE estimator $\hat{X}_n \equiv T(\{X_i \text{ for } i < n\}) = \mathbb{E}[X_n | \{X_i \text{ for } i < n\}]$ based on past $\{X_i \text{ for } i < n\}$. Here is the proof of Property 3.1 in textbook MBIP

$$\begin{aligned}\hat{X}_n &\equiv \mathbb{E}[X_n | \{X_i \text{ for } i < n\}] \\ &= \mathbb{E}\left[X_n \cdot (X_1, \cdots, X_{n-1})\right] \cdot \mathbb{E}\left[(X_1, \cdots, X_{n-1})^T \cdot (X_1, \cdots, X_{n-1})\right]^{-1} \cdot (X_1, \cdots, X_{n-1})^T \\ &= (h_{n,1}, h_{n,2} \cdots, h_{n,n-1}) \cdot (X_1, \cdots, X_{n-1})^T\end{aligned}$$

Where the coefficients of past is given by
$(h_{n,1}, h_{n,2} \cdots, h_{n,n-1}) = \mathbb{E}\left[X_n \cdot (X_1, \cdots, X_{n-1})\right] \cdot \mathbb{E}\left[(X_1, \cdots, X_{n-1})^T \cdot (X_1, \cdots, X_{n-1})\right]^{-1}$
Define the error as $\mathcal{E}_n \equiv X_n - \hat{X}_n = X_n - \mathbb{E}[X_n | \{X_i \text{ for } i < n\}]$

$$\begin{aligned}&\mathbb{E}\left[\mathcal{E}_n^2 \mid \{X_i \text{ for } i < n\}\right] = \mathbb{E}\left[(X_n - \hat{X}_n)^2 \mid \{X_i \text{ for } i < n\}\right] \\ &= \mathbb{E}\left[X_n^2\right] - \mathbb{E}\left[X_n \cdot (X_1, \cdots, X_{n-1})\right] \cdot \mathbb{E}\left[(X_1, \cdots, X_{n-1})^T \cdot (X_1, \cdots, X_{n-1})\right]^{-1} \cdot \mathbb{E}\left[X_n \cdot (X_1, \cdots, X_{n-1})\right]^T \\ &= \mathbb{E}\left[X_n^2\right] - (h_{n,1}, h_{n,2} \cdots, h_{n,n-1}) \mathbb{E}\left[(X_1, \cdots, X_{n-1})^T \cdot (X_1, \cdots, X_{n-1})\right] (h_{n,1}, h_{n,2} \cdots, h_{n,n-1})^T \\ &= \mathbb{E}\left[X_n^2\right] - \mathbb{E}\left[\hat{X}_n^2\right]\end{aligned}$$

Here is the proof of Property 3.2 in textbook MBIP. For $X_i, i < n$, we have

$$\begin{aligned}
\mathbb{E}[X_i \hat{X}_n] &= \mathbb{E}[X_i \mathbb{E}[X_n | \{X_i \text{ for } i < n\}]] = \int x_i \left[\int x_n \cdot f(x_n \mid x_1, \cdots, x_{n-1}) \mathrm{d}x_n \right] f(x_1, \cdots, x_{n-1}) \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \\
&= \int x_i \cdot x_n \cdot f(x_1, \cdots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \mathrm{d}x_n = \int x_i \cdot x_n \cdot f(x_i, x_n) \mathrm{d}x_i \mathrm{d}x_n \\
&= \mathbb{E}[X_i X_n]
\end{aligned}$$

$$\mathbb{E}[X_i \mathcal{E}_n] = \mathbb{E}[X_i X_n] - \mathbb{E}[X_i \hat{X}_n] = 0$$

Moreover, for $\mathcal{E}_i, i < n$, we define $F_n \equiv \{X_i \text{ for } i < n\}$, then we have

$$\begin{aligned}
\mathbb{E}[\mathcal{E}_i \mathcal{E}_n] &= \mathbb{E}[(X_i - \mathbb{E}[X_i \mid F_i]) (X_n - \mathbb{E}[X_n \mid F_n])] \\
&= \int \int \left(x_i - \int x_i \cdot f(x_i \mid \vec{f_i}) \mathrm{d}x_i \right) \left(x_n - \int x_n \cdot f(x_n \mid \vec{f_n}) \mathrm{d}x_n \right) f(x_n, \vec{f_n}) \mathrm{d}x_n \mathrm{d}\vec{f_n} \\
&= \int \left(x_i - \int x_i \cdot f(x_i \mid \vec{f_i}) \mathrm{d}x_i \right) \left[\int \left(x_n - \int x_n \cdot f(x_n \mid \vec{f_n}) \mathrm{d}x_n \right) f(x_n \mid \vec{f_n}) \mathrm{d}x_n \right] f(\vec{f_n}) \mathrm{d}\vec{f_n} \\
&= \int \left(x_i - \int x_i \cdot f(x_i \mid \vec{f_i}) \mathrm{d}x_i \right) \left[\int x_n \cdot f(x_n \mid \vec{f_n}) \mathrm{d}x_n - \int x_n \cdot f(x_n \mid \vec{f_n}) \mathrm{d}x_n \right] f(\vec{f_n}) \mathrm{d}\vec{f_n} \\
&= \int \left(x_i - \int x_i \cdot f(x_i \mid \vec{f_i}) \mathrm{d}x_i \right) \cdot 0 \cdot f(\vec{f_n}) \mathrm{d}\vec{f_n} \\
&= 0
\end{aligned}$$

Since $\{X_n\}_{n=1}^N$ is Gaussian random process with zero-mean $\mathbb{E}[X_n] = \mu = 0$, we have the jointly Gaussian PDF for $X \equiv (X_1, \cdots, X_N)^T$, where $R \equiv \mathbb{E}[XX^T]$

$$f_X(x) = \frac{1}{(2\pi)^{N/2}} |R|^{-1/2} \exp\left\{ -\frac{1}{2} x^T R^{-1} x \right\} \Leftrightarrow X \sim N(0, R)$$

With Step 4, we may define $\mathcal{E} \equiv (\mathcal{E}_1, \cdots, \mathcal{E}_N)^T$, and have

$$\mathcal{E} = (I - H)X = AX, \quad H \equiv \begin{pmatrix} 0 & 0 & \cdots & & 0 \\ h_{2,1} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h_{N-1,1} & h_{N-1,2} & \cdots & 0 & 0 \\ h_{N,1} & h_{N,2} & \cdots & h_{N,N-1} & 0 \end{pmatrix}$$

The corresponding distribution PDF of $\mathcal{E}$ is

$$\begin{aligned}
f_{\mathcal{E}}(e) &= \frac{1}{(2\pi)^{N/2}} \left| (I - H) R (I - H)^T \right|^{-1/2} \exp\left\{ -\frac{1}{2} e^T \left((I - H) R (I - H)^T \right)^{-1} e \right\} \\
&\Leftrightarrow \mathcal{E} \sim N\left(0, (I - H) R (I - H)^T\right)
\end{aligned}$$

Notice $\mathbb{E}[\mathcal{E}_i \mathcal{E}_n] = 0$ for $i \neq n$, we know that $\Lambda \equiv \mathbb{E}[\mathcal{E}\mathcal{E}^T]$ is a diagonal matrix

$$\Lambda \equiv \mathbb{E}[\mathcal{E}\mathcal{E}^T] = (I - H) R (I - H)^T = \mathrm{diag}\left(\sigma_1^2, \cdots, \sigma_N^2\right)$$

$$f_{\mathcal{E}}(e) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{ -\frac{e_n^2}{2\sigma_n^2} \right\} \Leftrightarrow \mathcal{E} \sim N(0, \Lambda)$$

Where $\sigma_n^2 \equiv \mathbb{E}[\mathcal{E}_n^2] = \mathbb{E}\left[\mathbb{E}[\mathcal{E}_n^2 \mid \{X_i \text{ for } i < n\}]\right] = \mathbb{E}\left[\mathbb{E}\left[X_n^2\right] - \mathbb{E}\left[\hat{X}_n^2\right]\right] = \mathbb{E}\left[X_n^2\right] - \mathbb{E}\left[\hat{X}_n^2\right]$

Consider coefficients for $\hat{X}_n, \hat{X}_{n+1}$, denote $F_n = (X_1, \cdots, X_{n-1})^T, F_{n+1} = (X_1, \cdots, X_n)^T = (F_n^T, X_n)^T$

$$(h_{n,1}, h_{n,2} \cdots, h_{n,n-1})^T = \mathbb{E}\left[(X_1, \cdots, X_{n-1})^T \cdot (X_1, \cdots, X_{n-1})\right]^{-1} \cdot \mathbb{E}\left[X_n \cdot (X_1, \cdots, X_{n-1})\right]^T$$

$$(h_{n+1,1}, h_{n+1,2} \cdots, h_{n,n-1})^T = \mathbb{E}\left[(X_1, \cdots, X_n)^T \cdot (X_1, \cdots, X_n)\right]^{-1} \cdot \mathbb{E}\left[X_{n+1} \cdot (X_1, \cdots, X_n)\right]^T$$

Step 6.b: If random process $\{X_n\}_{n=0}^N$ with zero-mean $\mathbb{E}[X_n] = 0$, we know that X_n and $\{X_{n-P}, \cdots, X_{n-1}\}$ are jointly Gaussian random vectors.

Furthermore, we can assume $\{X_n\}_{n=0}^N$ is **wide-sense stationary**

$$\mathbb{E}\left[X_n\right] = \mu = 0, \quad \mathbb{E}\left[X_n X_i\right] = \mathbb{E}\left[X_n X_i\right] - \mu^2 = R(|n-i|) \quad \forall n, i \in \{1, \cdots, N\}$$

As proved in Chapter Problem 19, all Gaussian wide-sense stationary random processes are:

a) strict-sense stationary
b) reversible

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

It holds for $m \geq 1, k \geq 0$. Thus, we may set $m \to n - P \geq 1, m + k \to n \geq P + 1, k \to P$

$$\begin{aligned}\text{Cov}[(X_{n-P}, \cdots, X_n)^T (X_{n-P}, \cdots, X_n)] &= \mathbb{E}[(X_{n-P}, \cdots, X_n)^T (X_{n-P}, \cdots, X_n)] \\ &= \begin{pmatrix} R(0) & R(1) & \cdots & R(P) \\ R(1) & R(0) & \cdots & R(P-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(P) & R(P-1) & \cdots & R(0) \end{pmatrix} = R^{(P)}\end{aligned}$$

$$f_{X_{n-P}, \cdots, X_n}(x_{n-P}, \cdots, x_n) = \frac{\left|R^{(P)}\right|^{-1/2}}{(2\pi)^{P/2}} \exp\left\{-\frac{1}{2}(x_{n-P}, \cdots, x_n)^T R^{(P)^{-1}} (x_{n-P}, \cdots, x_n)\right\}$$

$$f_{X_{n-P}, \cdots, X_n}(x_{n-P}, \cdots, x_n) = f_{X_{n-P}, \cdots, X_n}(x_n, \cdots, x_{n-P})$$

Then we only use the past P samples $\{X_{n-P}, \cdots, X_{n-1}\}$ to predict current value X_n

$$\begin{aligned}\hat{X}_n &\equiv \mathbb{E}[X_n | \{X_{n-P}, \cdots, X_{n-1}\}] \\ &= \mathbb{E}\left[X_n \cdot (X_{n-P}, \cdots, X_{n-1})\right] \cdot \mathbb{E}\left[(X_{n-P}, \cdots, X_{n-1})^T \cdot (X_{n-P}, \cdots, X_{n-1})\right]^{-1} \cdot (X_{n-P}, \cdots, X_{n-1})^T \\ &= (h_{n,n-P}, \cdots, h_{n,n-1}) \cdot (X_{n-P}, \cdots, X_{n-1})^T \\ &= (h_P, \cdots, h_1) \cdot (X_{n-P}, \cdots, X_{n-1})^T\end{aligned}$$

Where the coefficients of past is given by

$(h_{n,n-P}, \cdots, h_{n,n-1}) = \mathbb{E}\left[X_n \cdot (X_{n-P}, \cdots, X_{n-1})\right] \cdot \mathbb{E}\left[(X_{n-P}, \cdots, X_{n-1})^T \cdot (X_{n-P}, \cdots, X_{n-1})\right]^{-1} = [R(P), \cdots, R(1)] \cdot \{R^{(P-1)}\}^{-1} = (h_P, \cdots, h_1)$

Define the error as $\mathcal{E}_n \equiv X_n - \hat{X}_n = X_n - \mathbb{E}[X_n | \{X_{n-P}, \cdots, X_{n-1}\}]$