# I225E Statistical Signal Processing

## 11. Bayesian Estimation

MAWALIM and UNOKI

candylim@jaist.ac.jp and unoki@jaist.ac.jp

School of Information Science

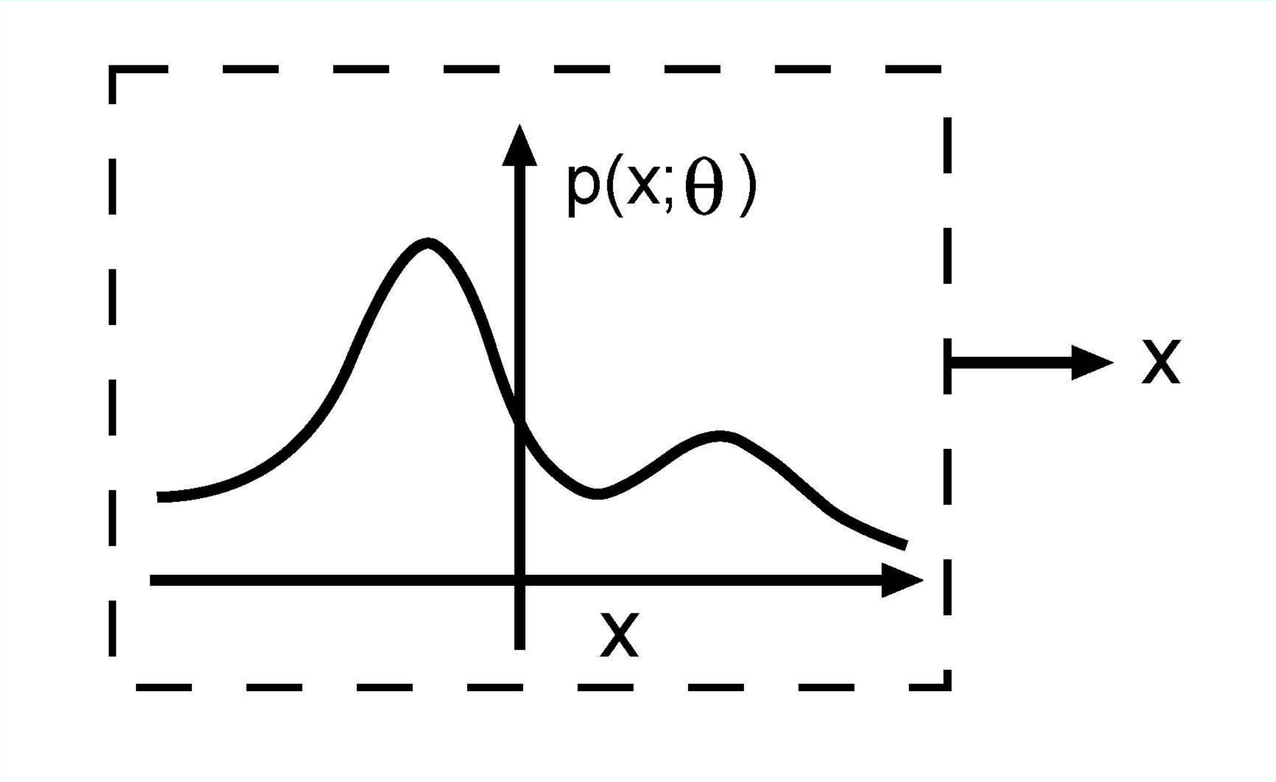
JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY,
HOKURIKU 1990

# Bayesian estimation

- Classical and Bayesian estimation
- Linear minimum mean square-error (MMSE) estimation
- Orthogonality principle
- Yule-Walker equation

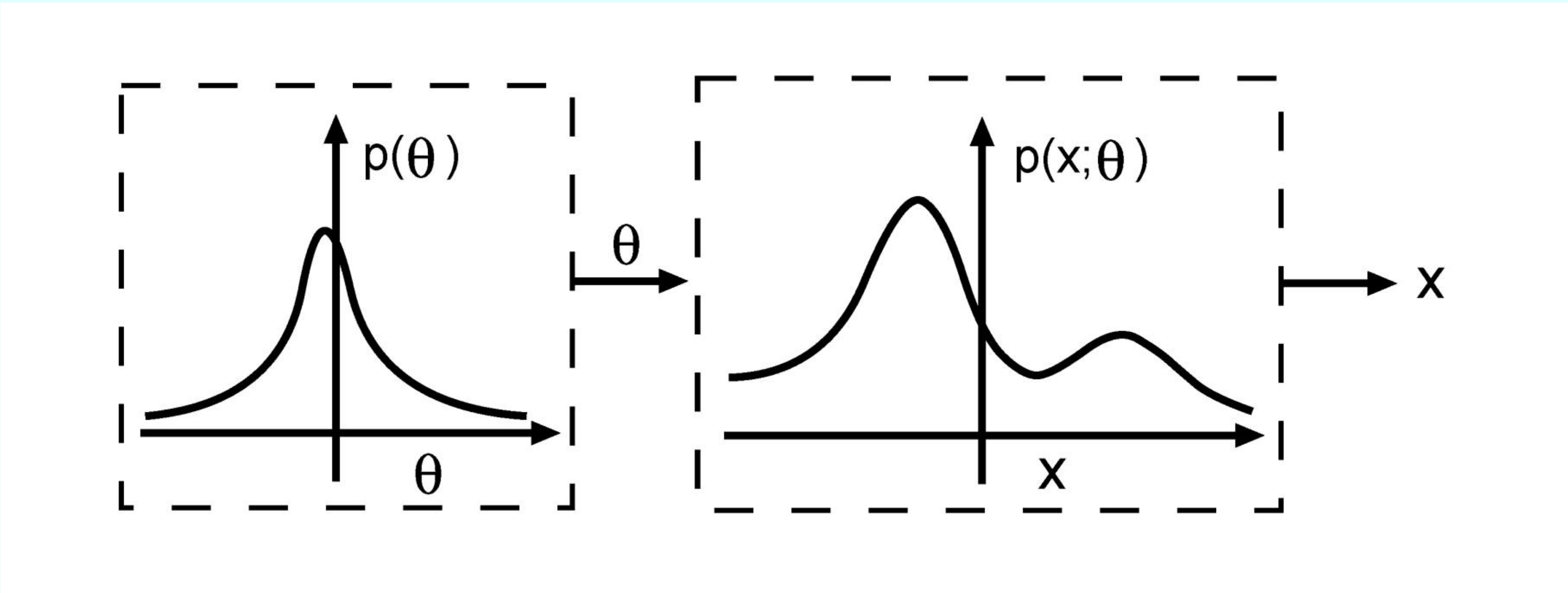
# 1. Introduction

In classical estimation, the unknown parameter $\theta$ is viewed as ***nonrandom***. Statistical model is specified entirely by conditional probability $p(x,\theta)$.

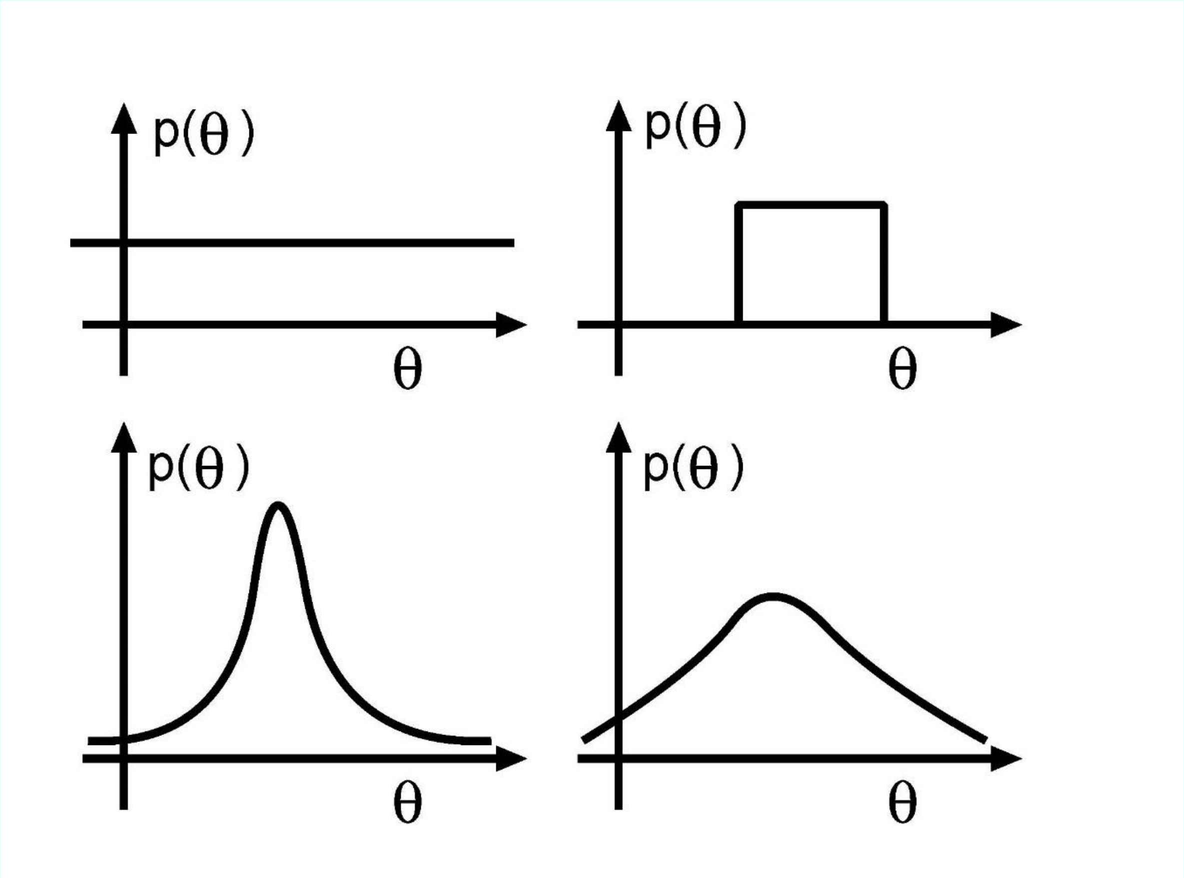


# 1. Introduction

In Baysian estimation, the unknown parameter $\theta$ is viewed as ***random***. Statistical model is specified in terms of conditional probability $p(x|\theta)$ and a prior distribution $p(\theta)$ on $\theta$.



The prior $p(\theta)$ can be used as a “prior knowledge.”



- By Bayes' rule, we may express the posterior distribution of $\theta$ given $x$ as

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}$$

- Definition of Mean Squared Error $E\left\{\left(\theta - \hat{\theta}\right)^2\right\}$
- Classical ($\theta$ is nonrandom):

$$MSE\left(\hat{\theta}\right) = \int \left(\hat{\theta} - \theta\right)^2 p(x, \theta)dx$$

- Bayesian ($\theta$ is random):

$$Bmse\left(\hat{\theta}\right) = \iint \left(\theta - \hat{\theta}\right)^2 p(x, \theta)dxd\theta$$

## Minimum Mean Squared Error (MMSE)

From $p(x,\theta) = p(\theta|x)p(x)$,

$$Bmse(\hat{\theta}) = \int \left[ \int \left(\theta - \hat{\theta}\right)^2 p(\theta|x)p(x)d\theta \right] dx$$

To minimize the inside of $[\cdot]$,

$$\frac{\partial}{\partial\hat{\theta}} \int \left(\theta - \hat{\theta}\right)^2 p(\theta|x)d\theta = \int \frac{\partial}{\partial\hat{\theta}} \left(\theta - \hat{\theta}\right)^2 p(\theta|x)d\theta$$

$$= \int -2\left(\theta - \hat{\theta}\right)p(\theta|x)d\theta$$

$$= -2\int \theta p(\theta|x)d\theta + 2\hat{\theta}\int p(\theta|x)d\theta = 0$$

Hence, $\hat{\theta} = \int \theta p(\theta|x)d\theta = E(\theta|x)$.

The best estimation is given by expectation value of $\theta$ under condition that $x$ is observed.

# 2. Linear MMSE Estimation

With respect to linear estimator

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n X[n] + a_N$$

Minimize Mean Squared Error

$$Bmse(\hat{\theta}) = E_{X,\theta}\left[\left(\theta - \hat{\theta}\right)^2\right]$$

$$\frac{\partial Bmse(\hat{\theta})}{\partial a_N} = \frac{\partial}{\partial a_N} E[(\theta - \sum_{n=0}^{N-1} a_n X[n] - a_N)^2]$$

$$= -2E[\theta - \sum_{n=0}^{N-1} a_n X[n] - a_N] = 0$$

Substituting $a_N = E[\theta] - \sum_{n=0}^{N-1} a_n E[X[n]]$,

$$Bmse\left(\hat{\theta}\right)$$

$$= E[\{\sum_{n=0}^{N-1} a_n (X[n] - E[X[n]]) - (\theta - E[\theta])\}^2]$$

$$= E[\{\boldsymbol{a}^T (\boldsymbol{X} - E[\boldsymbol{X}]) - (\theta - E[\theta])\}^2]$$

$$= E[\boldsymbol{a}^T (\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T \boldsymbol{a}]$$

$$- E[\boldsymbol{a}^T (\boldsymbol{X} - E[\boldsymbol{X}])(\theta - E[\theta])]$$

$$- E[(\theta - E[\theta])(\boldsymbol{X} - E[\boldsymbol{X}])^T \boldsymbol{a}]$$

$$+ E[(\theta - E[\theta])^2]$$

$$= \boldsymbol{a}^T E[\,(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T]\boldsymbol{a}$$

$$- \boldsymbol{a}^T E[(\boldsymbol{X} - E[\boldsymbol{X}])(\theta - E[\theta])]$$

$$- E[(\theta - E[\theta])(\boldsymbol{X} - E[\boldsymbol{X}])^T]\boldsymbol{a}$$

$$+ E[(\theta - E[\theta])^2]$$

$$= \boldsymbol{a}^T \boldsymbol{C}_{XX} \boldsymbol{a} - \boldsymbol{a}^T \boldsymbol{C}_{X\theta} - \boldsymbol{C}_{\theta X} \boldsymbol{a} + \boldsymbol{C}_{\theta\theta}$$

Here,

$$\boldsymbol{X} = [X[0], X[1], \cdots, X[N-1]]^T,$$

$$\boldsymbol{a} = [a_0, a_1, \cdots, a_{N-1}]^T,$$

covariance matrices are

$$\boldsymbol{C}_{XX} = E\left\{\left(\boldsymbol{X} - E(\boldsymbol{X})\right)(\boldsymbol{X} - E(\boldsymbol{X}))^T\right\} \in \Re^{N\times N},$$

$$\boldsymbol{C}_{\theta\theta} = E\left\{\left(\theta - E(\theta)\right)(\theta - E(\theta))^T\right\} \in \Re^{1\times 1}$$

$$\boldsymbol{C}_{\theta X} = E\left\{\left(\theta - E(\theta)\right)(\boldsymbol{X} - E(\boldsymbol{X}))^T\right\} = \boldsymbol{C}_{X\theta}^T \in \Re^{1\times N}$$

$$\frac{\partial Bmse}{\partial \boldsymbol{a}} = 2\boldsymbol{C}_{XX}\boldsymbol{a} - 2\boldsymbol{C}_{X\theta} = 0$$

Hence,

$$\boldsymbol{a} = \boldsymbol{C}_{XX}^{-1}\boldsymbol{C}_{X\theta}$$

$$\begin{aligned}
\hat{\theta} &= \boldsymbol{a}^T\boldsymbol{X} + a_N \\
&= \boldsymbol{C}_{X\theta}^T\boldsymbol{C}_{XX}^{-1}\boldsymbol{X} + E(\theta) - \boldsymbol{C}_{X\theta}^T\boldsymbol{C}_{XX}^{-1}E(\boldsymbol{X}) \\
&= E(\theta) + \boldsymbol{C}_{\theta X}\boldsymbol{C}_{XX}^{-1}\big(\boldsymbol{X} - E(\boldsymbol{X})\big)
\end{aligned}$$

(in case of $E(X) = 0, E(\theta) = 0$)

$$= \boldsymbol{C}_{\theta X}\boldsymbol{C}_{XX}^{-1}\boldsymbol{X}$$

Corresponding minimum mean square error is

$$
\begin{aligned}
Bmse(\hat{\theta}) &= \boldsymbol{a}^T \boldsymbol{C}_{XX} \boldsymbol{a} - \boldsymbol{a}^T \boldsymbol{C}_{X\theta} - \boldsymbol{C}_{\theta X} \boldsymbol{a} + \boldsymbol{C}_{\theta\theta} \\
&= \boldsymbol{C}_{X\theta}^T \boldsymbol{C}_{XX}^{-1} \boldsymbol{C}_{XX} \boldsymbol{C}_{XX}^{-1} \boldsymbol{C}_{X\theta} - \boldsymbol{C}_{X\theta}^T \boldsymbol{C}_{XX}^{-1} \boldsymbol{C}_{X\theta} \\
&\qquad - \boldsymbol{C}_{\theta X} \boldsymbol{C}_{XX}^{-1} \boldsymbol{C}_{X\theta} + \boldsymbol{C}_{\boldsymbol{\theta\theta}} \\
&= \boldsymbol{C}_{\theta\theta} - \boldsymbol{C}_{\theta X} \boldsymbol{C}_{XX}^{-1} \boldsymbol{C}_{X\theta}.
\end{aligned}
$$

# Example

Consider the following random variable,

$$X[n] = \theta + W[n], \quad (n = 0,1,\cdots,N-1)$$

$W[n]$ is independent Gaussian with mean 0 and variance $\sigma^2$, DC component $\theta$ obeys uniform distribution $U(-\theta_0,\theta_0)$. $\theta$ and $W[n]$ are independent. Find the minimum mean square error estimation for $\theta$.

From $E(\theta) = 0, E(X[n]) = 0$ (Namely, $E(\boldsymbol{X}) = 0$).

$$\boldsymbol{C}_{XX} = E(\boldsymbol{X}\boldsymbol{X}^T) = E[(\theta\mathbf{1} + \boldsymbol{W})(\theta\mathbf{1} + W)^T]$$

$$= E(\theta^2)\mathbf{1}\mathbf{1}^T + \sigma^2\boldsymbol{I}$$

$$\boldsymbol{C}_{\theta X} = E(\theta\boldsymbol{X}^T) = E[\theta(\theta\mathbf{1} + \boldsymbol{W})^T] = E(\theta^2)\mathbf{1}^T$$

where $\mathbf{1} = (1,1,\cdots,1)^T$

Therefore ,

$$\hat{\theta} = \boldsymbol{C}_{\theta X}\boldsymbol{C}_{XX}^{-1}\boldsymbol{X}$$

$$= \sigma_\theta^2 \mathbf{1}^T\left(\sigma_\theta^2 \mathbf{1}\mathbf{1}^T + \sigma^2 \boldsymbol{I}\right)^{-1}\boldsymbol{X} \qquad (\leftarrow \sigma_\theta^2 = E(\theta^2))$$

$$= \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma^2/N}\overline{\boldsymbol{X}}$$ (From Woodbury's identity)

$$\left(\boldsymbol{I} + \frac{\sigma_\theta^2}{\sigma^2}\mathbf{1}\mathbf{1}^T\right)^{-1} = \boldsymbol{I} + \frac{\frac{\sigma_\theta^2}{\sigma^2}\mathbf{1}\mathbf{1}^T}{1 + N + \frac{\sigma_\theta^2}{\sigma^2}}$$

$$= \frac{\frac{\theta_0^2}{3}}{\frac{\theta_0^2}{3} + \sigma^2/N}\overline{\boldsymbol{X}} \qquad (\leftarrow \sigma_\theta^2 = \frac{\theta_0^2}{3})$$

# 3. Orthogonality Principle

**Problem:** With respect to linear estimator $\hat{\theta} = \sum_{n=0}^{N-1} a_n X[n]$, error is given by

$$\epsilon = \theta - \hat{\theta} = \theta - \sum_{n=0}^{N-1} a_n X[n].$$

Assuming $E(X) = 0, E(\theta) = 0$, then $a_N = 0$.

The minimum mean square error is

$$Bmse\left(\hat{\theta}\right) = E\{\epsilon^2\} = E\{(\theta - \sum_{n=0}^{N-1} a_n X[n])^2\}$$

# Orthogonality Principle

If coefficients $a_n$ are selected in such a way that error $\epsilon$ is orthogonal to random variables $X[0], X[1], \cdots, X[N-1]$., i.e.

$$E\{\epsilon X[n]\} = E\left\{\left(\theta - \sum_{i=0}^{N-1} a_i X[i]\right) X[n]\right\} = 0$$

$$(n = 0,1,\cdots,N-1),$$

the mean square error $Bmse$ is minimal.

[Proof] Local minimum condition of $Bmse$ (with respect to $a_n$) is

$$\frac{\partial Bmse}{\partial a_n} = E\left\{2\left(\theta - \sum_{i=0}^{N-1} a_i X[i]\right)(-X[n])\right\} = 0$$

$$(n = 0,1,\cdots,N-1).$$

# Determination of *a* by orthogonality principle

Orthogonality principle

$$E\left\{\left(\theta - \sum_{j=0}^{N-1} a_j X[j]\right) X[n]\right\} = 0$$

can be decomposed as

$$E\{\theta X[n]\} = \sum_{j=0}^{N-1} a_j E\{X[j]X[n]\}$$

Denoting $[C_{XX}]_{jn} = E\{X[j]X[n]\}$, $[C_{\theta X}]_n = E\{\theta X[n]\}$,

$$\begin{aligned}
[C_{\theta X}]_0 &= [C_{XX}]_{00} a_0 + [C_{XX}]_{01} a_1 + \cdots + [C_{XX}]_{0N-1}\, a_{N-1} \\
[C_{\theta X}]_1 &= [C_{XX}]_{10} a_0 + [C_{XX}]_{11} a_1 + \cdots + [C_{XX}]_{1N-1}\, a_{N-1} \\
&\vdots \\
[C_{\theta X}]_{N-1} &= [C_{XX}]_{N-10} a_0 + [C_{XX}]_{N-11} a_1 \\
&\quad + \cdots + [C_{XX}]_{N-1N-1} a_{N-1}
\end{aligned}$$

These are called Yule-Walker equation.

If $X[0], X[1], \cdots, X[N-1]$ are linearly independent, inverse matrix $\boldsymbol{C}_{XX}^{-1}$ exists (of $\boldsymbol{C}_{XX}$ is invertible), $\boldsymbol{a}$ is obtained as,

$$\boldsymbol{a} = \boldsymbol{C}_{XX}^{-1} \boldsymbol{C}_{\theta X}$$

# Mean Square Error

From orthogonality principle

$$E\{\epsilon X[n]\} = E\left\{\left(\theta - \hat{\theta}\right)X[n]\right\} = 0,$$

for arbitrary $c_n$ $(n = 0,1,\cdots,N-1)$

$$E\left\{\left(\theta - \hat{\theta}\right)(c_0 X[0] + c_1 X[1] + \cdots + c_{N-1} X[N-1])\right\} = 0$$

Substituting $c_i = a_i$,

$$E\left\{\left(\theta - \hat{\theta}\right)\hat{\theta}\right\} = 0.$$

Therefore,

$$
\begin{aligned}
Bmse(\hat{\theta}) = E\left\{\left|\theta - \hat{\theta}\right|^2\right\} &= E\{(\theta - \hat{\theta})(\theta - \hat{\theta})\} \\
&= E\{\theta\theta - \theta\hat{\theta} - \hat{\theta}\theta + \hat{\theta}\hat{\theta}\} \\
&= E\{\theta\theta - \hat{\theta}\theta - (\theta - \hat{\theta})\hat{\theta}\} \\
&= E\{\theta\theta - \hat{\theta}\theta\} \\
&= E\{(\theta - \hat{\theta})\theta\} = E\{\theta\theta\} - E\{\textstyle\sum_{n=0}^{N-1} a_n X[n]\,\theta\} \\
&= E\{\theta\theta\} - \textstyle\sum_{n=0}^{N-1} a_n E\{X[n]\theta\} \\
&= C_{\theta\theta} - \\
&\quad (a_0[\boldsymbol{C}_{X\theta}]_0 + a_1[\boldsymbol{C}_{X\theta}]_1 + \cdots + a_{N-1}[\boldsymbol{C}_{X\theta}]_{N-1}) \\
&= C_{\theta\theta} - a\boldsymbol{C}_{X\theta} = \boldsymbol{C}_{\theta\theta} - \boldsymbol{C}_{\theta X}\boldsymbol{C}_{XX}^{-1}\boldsymbol{C}_{X\theta}
\end{aligned}
$$