Lecture 7

Optimal Estimation Theory and Its Applications

Prof. Yan LIANG

liangyan@nwpu.edu.cn

School of Automation,

Northwestern Polytechnical University

Autumn,2019

Outline

1. Linear Minimum Variance Estimation
2. Orthogonality Principle
3. Wiener Filtering
4. Appendix: Stationary Stochastic Process

Linear Minimum Variance Estimation

Motivation

The Minimum Variance estimator requires knowing the conditional PDF $\mathrm{p}\{\mathrm{X}|\mathrm{Z}\}$. It may be difficult in applications.

- the main concern of the estimation is usually to obtain the estimate and its covariance (first two moments);
- to estimate the first two moments of a random vector, compared with reconstruction the PDF, is always computationally efficient ;

The LMV estimator is desirable in applications.

Linear Minimum Variance Estimation

Definition

The LMV estimate of X is the linear/affine function of Z. And among all linear/affine functions of Z, the LMV estimate obtains the minimum variance of estimate error i.e.,

$$E\left\{\left[X-\hat{X}_{\mathrm{LMV}}\right]\left[X-\hat{X}_{\mathrm{LMV}}\right]^{T}\right\} \leq E\left\{\left[X-\hat{X}\right]\left[X-\hat{X}\right]^{T}\right\}$$

with

$$\hat{X}=AZ+B \text{ and } \hat{X}_{\mathrm{LMV}}=A_{\mathrm{LMV}}Z+B_{\mathrm{LMV}}$$

where A and B are free parameter matrix and vector with proper dimensions, respectively while $\mathrm{A}_{\mathrm{LMV}}$ and $\mathrm{B}_{\mathrm{LMV}}$ are the corresponding optimal ones.

Linear Minimum Variance Estimation

Derivation

Let

$$C = -(\mathrm{E}X - B - A\mathrm{E}Z), \mathrm{Var}(X) = E\left((X - \mathrm{E}X)(\bullet)^T\right), \mathrm{Var}(Z) = E\left((Z - \mathrm{E}Z)(\bullet)^T\right)$$

and

$$\mathrm{Cov}(X, Z) = \mathrm{E}\left((X - \mathrm{E}X)(Z - \mathrm{E}Z)^T\right)$$

we have

$$\begin{aligned} J &= \mathrm{E}\left((X - B - AZ)(X - B - AZ)^T\right) = \mathrm{E}\left((X - \mathrm{E}X - C - A(Z - \mathrm{E}Z))(\bullet)^T\right) \\ &= \mathrm{Var}(X) + CC^T + A\mathrm{Var}(Z)A^T - \mathrm{Cov}(X, Z)A^T - A\mathrm{Cov}(Z, X) \end{aligned}$$

The right hand of the above equation plus and minus $\mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z, X)$ is

$$\begin{aligned} J = &\mathrm{Var}(X) - \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z, X) \\ &+ CC^T + \left(A - \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\right)\mathrm{Var}(Z)(\bullet)^T. \end{aligned}$$

Linear Minimum Variance Estimation

Derivation(Cont.)

$$
\begin{aligned}
J = {} & \mathrm{Var}(X) - \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z,X) \\
& + CC^T + \left(A - \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\right)\mathrm{Var}(Z)(\bullet)^T.
\end{aligned}
$$

By the fact that

Semi-definite

$$CC^T \geq 0 \quad \text{and} \quad \left(A - \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\right)\mathrm{Var}(Z)(\bullet)^T \geq 0,$$

we have

$$J \geq \mathrm{Var}(X) - \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z,X)$$

And the equality holds, if and only if

$$C = 0 \quad \text{and} \quad A = \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z).$$

Linear Minimum Variance Estimation

The filter parameters of the LMV estimator are

$$\Rightarrow \begin{cases} C_{\mathrm{LMV}} = 0 \\ A_{\mathrm{LMV}} = \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z) \end{cases}$$

$$\Rightarrow B_{\mathrm{LMV}} = \mathrm{E}X - \mathrm{C}ov(X, Z)\mathrm{Var}^{-1}(Z)\mathrm{E}Z$$

The LMV estimate is

$$\begin{aligned} \hat{X}_{\mathrm{LMV}} &= B_{\mathrm{LMV}} + A_{\mathrm{LMV}}Z \\ &= \mathrm{E}X + \mathrm{C}ov(X, Z)\mathrm{Var}^{-1}(Z)\,(Z - \mathrm{E}Z) \end{aligned}$$

Linear Minimum Variance Estimation

To obtain the LMV estimate

$$\begin{aligned}\hat{X}_{\mathrm{LMV}} &= B_{\mathrm{LMV}} + A_{\mathrm{LMV}} Z \\ &= \mathrm{E}X + \mathrm{C}ov(X, Z)\mathrm{Var}^{-1}(Z)\,(Z - \mathrm{E}Z)\end{aligned}$$

we need prior information:

first-order moments EX and EZ

second-order moments $\mathrm{Cov}(\mathrm{X}, \mathrm{Z})$ and $\mathrm{Var}(\mathrm{Z})$

instead of the conditional probability $p\{X|Z)$ utilized in the Minimum Mean Square Error (MMSE) estimate.

Linear Minimum Variance Estimation

As similar as the previous lesson, the Linear Minimum Variance estimate is equivalent to the Linear Minimum-Mean-Square-Error estimate. In other words, we have the following alternative derivation .

$$\hat{X}_{\mathrm{LMV}} = \arg\min_{\hat{X}} \left\{ \mathrm{E}\left(\left(X - \hat{X}\right)^T \left(X - \hat{X}\right) \right) \right\}$$

$$= \arg\min_{\hat{X}} \left\{ \mathrm{trace}\left\{ \mathrm{E}\left(\left(X - \hat{X}\right) \left(X - \hat{X}\right)^T \right) \right\} \right\}$$

Hint:

trace(A)=trace(A^T) A is a square matrix.

trace(AB)=trace(BA) if AB and BA are square matrices.

trace(A)=A if A is a scalar.

Linear Minimum Variance Estimation

$$J = \mathrm{trace}\left\{\mathrm{E}\left((X - AZ - B)(\bullet)^T\right)\right\}$$
$$\left.\frac{\partial J}{\partial A}\right|_{A=A_{\mathrm{LMV}}} = 2\left(A\mathrm{E}\left(ZZ^T\right) + B(\mathrm{E}Z)^T - \mathrm{E}\left(XZ^T\right)\right)\Big|_{A=A_{\mathrm{LMV}}} = 0$$
$$\left.\frac{\partial J}{\partial B}\right|_{B=B_{\mathrm{LMV}}} = -2\left(B + A\mathrm{E}Z - \mathrm{E}X\right)\big|_{B=B_{\mathrm{LMV}}} = 0$$

$$\Rightarrow \begin{cases} A_{\mathrm{LMV}} = \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z) \\ B_{\mathrm{LMV}} = \mathrm{E}X - \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\mathrm{E}Z \end{cases}$$

Hint: the derivative formulas of the trace with respect to a matrix see the Appendix in Lesson 3.

Linear Minimum Variance Estimation

Consideration

Is the LMV estimate is unbiased?

Linear Minimum Variance Estimation

$$\hat{X}_{\mathrm{LMV}} = \mathrm{E}X + \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\,(Z - \mathrm{E}Z)$$

$$\mathrm{E}\left(X - \hat{X}_{\mathrm{LMV}}\right) = \mathrm{E}X - \mathrm{E}X - \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\,(\mathrm{E}Z - \mathrm{E}Z) = 0$$

Yes. The LMV estimate is unbiased.

Furthermore, if the LMV estimate is shown biased in actual applications, what happens?

The cause is that the a priori parameters, EX and/or EZ, utilized in the estimator design must be different from the actual values, EX and/or EZ.

In other words, system identification is not acceptable.

Linear Minimum Variance Estimation

$$\mathrm{E}\left(X - \hat{X}_{\mathrm{LMV}}\right) = \mathrm{E}X - \mathrm{E}X - \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\left(\mathrm{E}Z - \mathrm{E}Z\right) = 0$$

The condition of the unbiasedness is somewhat mild. Even

$$\mathrm{E}X \neq \mathrm{E}X, \mathrm{E}Z \neq \mathrm{E}Z$$

the unbiasedness will still hold if choosing the suitable parameter

$$Cov\,(X, Z)\,Var^{-1}\,(Z)\,.$$

In other words, in the case that $\mathrm{E}X \neq \mathrm{E}X, \mathrm{E}Z \neq \mathrm{E}Z$ we may consider introducing a self-adaption mechanism in calculating $Cov\,(X, Z)\,, Var^{-1}\,(Z)$, or $Cov\,(X, Z)\,Var^{-1}\,(Z)$ to guarantee the unbiasedness. Such calculation is data-driven, i.e., from the obtained measurements. This idea results in adaptive estimators for dynamic systems.

Linear Minimum Variance Estimation

Consideration

Does introducing stochastic measurements must improve the estimate accuracy in statistics?

Linear Minimum Variance Estimation

In the case that the measurement is not available, we have to estimate X by a constant, i.e.,

$$
\begin{aligned}
J &= \mathrm{E}\left((X-\hat{X})^T(X-\hat{X})\right) \\
&= \mathrm{E}(X^TX) - \hat{X}^T\mathrm{E}X - (\mathrm{E}X)^T\hat{X} + \mathrm{E}(\hat{X}^T\hat{X}) \\
&= \mathrm{E}(X^TX) - 2\hat{X}^T\mathrm{E}X + \hat{X}^T\hat{X}
\end{aligned}
$$

$$
\frac{\partial J}{\partial \hat{X}} = -2\mathrm{E}X + 2\hat{X} = 0
$$

$$
\Rightarrow \hat{X} = \mathrm{E}X
$$

Without measurements, the MV estimate is the priori mean.

$$
\mathrm{E}\left((X-\hat{X})(X-\hat{X})^T\right) = \mathrm{E}\left((X-\mathrm{E}X)(X-(\mathrm{E}X)^T\right) = \mathrm{Var}(X)
$$

Without measurements, the MV estimate covariance is the priori covariance.

Linear Minimum Variance Estimation

When the measurement is available, we have the LMV estimate

$$\hat{X}_{\mathrm{LMV}} = \mathrm{E}X + \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\,(Z - \mathrm{E}Z)$$

and its covariance

$$\begin{aligned}
&\mathrm{E}\left((X-\hat{X})(X-\hat{X})^T\right) = \mathrm{E}\left(\left(\mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\,(Z-\mathrm{E}Z) - (X-\mathrm{E}X)\right)(\bullet)^T\right) \\
&= \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\mathrm{Var}(Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z,X) \\
&\ - 2\mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z,X) + \mathrm{Var}(X) \\
&= \mathrm{Var}(X) - \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z,X) \le \mathrm{Var}(X)
\end{aligned}$$

Conclusion: introducing measurements is helpful to improve estimate accuracy, even the measurement is stochastic.

Hint:the improvement holds in statistics, NOT for a sampling

Linear Minimum Variance Estimation

Does the measurement is always useful in the LMV estimation?

$$
\begin{aligned}
&\mathrm{E}\left((X-\hat{X})(X-\hat{X})^T\right) \\
&= \mathrm{Var}(X) - \mathrm{Cov}(X,Z)\mathrm{Var}^{-1}(Z)\mathrm{Cov}(Z,X)
\end{aligned}
$$

Conclusion: In some situations, introducing measurements is useless in improving estimate accuracy:

(1) Z is linearly independent of X, i.e.,

$$Cov\,(X,Z) = 0$$

(2) Z is corrupted by noises too severely to carry any useful information, i.e.,

$$\mathrm{Var}^{-1}(Z) = 0$$

Linear Minimum Variance Estimation

Example 4.1: Consider X and its measurement Z satisfy the following linear constraint

$$Z = HX + V$$

Given $EX{=}\mathrm{m}_x$,$\mathrm{Var}(\mathrm{X})$ =P_x, $\mathrm{EV} = 0, \mathrm{E}(\mathrm{VV}^{\mathrm{T}}) = \mathrm{R}, \mathrm{E}(\mathrm{XV}^{\mathrm{T}}) = 0$, determine the LMV estimate.

$$\boxed{\hat{X}_{\mathrm{LMV}} = \mathrm{E}X + \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\,(Z - \mathrm{E}Z)}$$

Linear Minimum Variance Estimation

$$\mathrm{E}X = \mathrm{m}_x$$

$$\mathrm{E}Z = \mathrm{E}\,(HX + V) = H\mathrm{E}X + \mathrm{E}V = H\mathrm{m}_x$$

$$\mathrm{Var}(Z) = \mathrm{E}\left((Z - \mathrm{E}Z)\,(\bullet)^T\right) = \mathrm{E}\left((HX - H\mathrm{E}X + V)\,(\bullet)^T\right)$$
$$= H\mathrm{E}\left((X - \mathrm{E}X)(\bullet)^T\right)H^T + \mathrm{E}\left(VV^T\right) = H\mathrm{P}_x H^T + R$$

$$\mathrm{Cov}(X, Z) = \mathrm{E}\left((X - EX)\,(Z - \mathrm{E}Z)^T\right)$$
$$= \mathrm{E}\left((X - \mathrm{E}X)\,(H(X - \mathrm{E}X) + V)^T\right)$$
$$= \mathrm{E}\left((X - \mathrm{E}X)\,(\bullet)^T\right)H^T = \mathrm{P}_x H^T$$

$$\hat{X}_{\mathrm{LMV}} = \mathrm{E}X + \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\,(Z - \mathrm{E}Z)$$
$$= \mathrm{m}_x + \mathrm{P}_x H^T\left(H\mathrm{P}_x H^T + R\right)^{-1}(Z - H\mathrm{m}_x)$$

Linear Minimum Variance Estimation

$$\hat{X}_{\mathrm{LMV}} = [P_x - P_x H^T (H P_x H^T + R)^{-1} H P_x] {P_x}^{-1} m_x + P_x H^T (H P_x H^T + R)^{-1} Z$$

Based on Matrix inversion formula

$$P - PH^T (HPH^T + R)^{-1} HP = (P^{-1} + H^T R^{-1} H)^{-1}$$

$$\begin{aligned}
\hat{X}_{\mathrm{LMV}} &= \left({P_x}^{-1} + H^T R^{-1} H\right)^{-1} \left({P_x}^{-1} m_x + ({P_x}^{-1} + H^T R^{-1} H) P_x H^T (H P_x H^T + R)^{-1} Z\right) \\
&= \left({P_x}^{-1} + H^T R^{-1} H\right)^{-1} \left({P_x}^{-1} m_x + (H^T + H^T R^{-1} H P_x H^T)(H P_x H^T + R)^{-1} Z\right) \\
&= \left({P_x}^{-1} + H^T R^{-1} H\right)^{-1} \left({P_x}^{-1} m_x + H^T R^{-1} (R + H P_x H^T)(H P_x H^T + R)^{-1} Z\right) \\
&= \left({P_x}^{-1} + H^T R^{-1} H\right)^{-1} \left({P_x}^{-1} m_x + H^T R^{-1} Z\right)
\end{aligned}$$

It is easy to prove that $E\left(\left(X - \hat{X}_{\mathrm{LMV}}\right)(\bullet)^T\right) = \left({P_x}^{-1} + H^T R^{-1} H\right)^{-1}$.

Linear Minimum Variance Estimation

If there is no a priori statistical information about X, we have the following LMV estimate with $P_x^{-1} = 0$. From

$$\hat{X}_{LMV} = \left(P_x^{-1} + H^T R^{-1} H\right)\left(P_x^{-1} m_x + H^T R^{-1} Z\right)$$

we have

$$\begin{cases} \hat{X}_{LMV}(Z) = \left(H R^{-1} H\right)^{-1} H^T R^{-1} Z \\ Var\left(\hat{X}_{LMV}\right) = \left(H R^{-1} H\right)^{-1} \end{cases}$$

$\hat{X}_{LSW} = \left(H^T W H\right)^{-1} H^T W Z = R^{-1}$ Markov estimate

Conclusion: Without a priori statistical information about X, the LMV estimate degrades to the Markov estimate.

Orthogonality Principle

Consider random vectors, X and Z with the means being μ_X and μ_Z, respectively. Their cross-covariance is

$$
\begin{aligned}
Cov\left(X, Z\right) &= E\left(\left(X - \mu_X\right)\left(Z - \mu_Z\right)^T\right) \\
&= E\left(XZ^T\right) - \mu_X\mu_Z^T - \mu_X\mu_Z^T + \mu_X\mu_Z^T \\
&= E\left(XZ^T\right) - \mu_X\mu_Z^T
\end{aligned}
$$

we have

$$
\mathrm{E}(XZ^T) = \mathrm{Cov}(X, Z) + \ \mu_X\mu_Z^T
$$

Orthogonality Principle

$$\boxed{\mathrm{E}(XZ^T) = \mathrm{Cov}(X,Z) + \ \mu_X \mu_Z^T}$$

If $\mathrm{Cov}(X,Z) = 0$, i.e.,$\mathrm{E}(XZ^T) = \ \mu_X \mu_Z^T$, then X and Z are called uncorrelated or linearly independent.

If $\mathrm{E}(XZ^T) = 0$, then X and Z are called orthogonal.

If $p(x,z) = p(x)p(z)$, then X and Z are called independent.

Note: Different from deterministic vectors, the random vectors have above definition "uncorrelated" or "orthogonal" based on mathematic expectation. And the independence is defined based on the PDFs.

Orthogonality Principle

The independence of X and Z represents that they has no common information in statistics. In other words, the samples of Z and its any functions are useless in estimate X.

The linear independence of X and Z represents that the samples of Z and its any linear functions are useless in estimate X.

The independence includes the linear independence as a special case. Their equivalence needs other conditions, for example Caussian-distributed of X and Z.

If $\mu_X = 0$ or $\mu_Z = 0$, the orthogonality is equivalent to the linear independence.

Orthogonality Principle

Definition

Consider X and Z are n-dimensional and m-dimensional random vectors with the first two moments, respectively. If

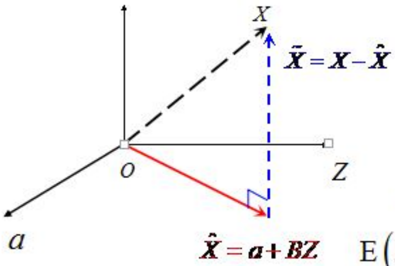
1. $\hat{X} = a + BZ$,where a and B are deterministic —Linear
2. $\mathrm{E}\hat{X} = \mathrm{E}X,$ —Unbiased
3. $\mathrm{E}\left((X - \hat{X})Z^T\right) = 0,$ —Orthogonal

$\hat{X}$ is called the orthogonal projection of X on Z.$\tilde{X} = X - \hat{X}$ is called the orthogonal component projection of X on Z, representing $X \perp Z$.

By the fact that $E\tilde{X} = 0$, we have $E\left(\tilde{X}Z^T\right) = 0$ and $E\tilde{X}EZ^T = 0$

Here the orthogonality is equivalent to the linear independence.

Orthogonality Principle



Orthogonal Project

$$E\tilde{X} = 0$$ --Unbiased

$$E\left(\tilde{X}Z^T\right) = 0$$ --Orthogonal

$$E(\tilde{X}a^T) = E(\tilde{X})a^T = 0 \qquad \tilde{X} \perp a$$

$$E\left(\tilde{X}(BZ)^T\right) = E(\tilde{X}Z^T B^T) = 0 \qquad \tilde{X} \perp BZ$$

$$E\left(\tilde{X}(a+BZ)^T\right) = 0 \qquad \tilde{X} \perp \hat{X}$$

Orthogonality Principle

Is the LMV estimate equivalent to the orthogonal project?

Solution

The first condition (linearity) is satisfied according to the definition.

The second condition (unbiasedness) has been proved in this lesson.

The third condition (orthogonality) is shown as follows

$$
\begin{aligned}
&\mathrm{E}\left((X-\hat{X}_{\mathrm{LMV}})Z^T\right) = \mathrm{E}\left((X-\hat{X}_{\mathrm{LMV}})(Z-\mu_Z)^T\right) \\
&= \mathrm{E}\left(\left((X-\mu_X) - \mathrm{C}ov(X,Z)\mathrm{V}ar^{-1}(Z)(Z-\mu_Z)\right)(Z-\mu_Z)^T\right) \\
&= \mathrm{C}ov(X,Z) - \mathrm{C}ov(X,Z)\mathrm{V}ar^{-1}(Z)\mathrm{V}ar(Z) = 0
\end{aligned}
$$

Yes. They are equivalent.

Orthogonality Principle

What we obtain from the equivalence?

(1) It is more convenient to derive the LMV estimator through determining $EX, EZ, Var\,(Z)$, and $Cov\,(X, Z)$.

$$\hat{X}_{\mathrm{LMV}} = \mathrm{E}X + \mathrm{Cov}(X, Z)\mathrm{Var}^{-1}(Z)\,(Z - \mathrm{E}Z)$$

(2) The orthogonality can be utilized to testify whether or not the resultant estimator is optimal in the LMV sense. If it is not satisfied, the adaptive estimator can be designed to guarantee it.

D. Zhou, et al., Extension of Friedland's separate-bias estimation to randomly time-varying bias for nonlinear systems, IEEE Transactions on Automatic Control.38 (8), 1993,1270-1273.

Q. Xia, et al., Adaptive fading Kalman filter with an application, Automatica.30, 1994, 1333-1338.

Wiener Filtering

$X(t)$ the signal to be estimated

$Z(t)$ the measurement

X and Z are zero-mean and with known cross-correlation

$$E\left\{(X(t) - EX(t))(Z(t-\tau) - EZ(t-\tau))^T\right\} = R_{XZ}(\tau)$$

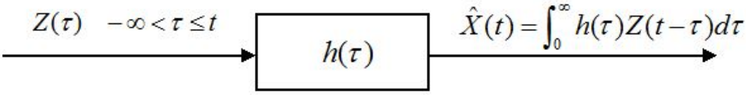
and self-correlation

$$E\left\{(Z(t) - EZ(t))(Z(t-\tau) - EZ(t-\tau))^T\right\} = R_{ZZ}(\tau)$$

Requirement

Design a linear time-invariant estimator/filter with the impulse response $\{\mathrm{h}(\tau), 0 \leq \tau \leq \mathrm{t}\}$ to obtain the LMV estimate of X(t) based on the measurements up to time t,$\{\mathrm{Z}(\tau), 0 \leq \tau \leq \mathrm{t}\}$.

Wiener Filtering



Input—Output relationship of linear filter

$h(\tau)$ impulse response function

$h(\tau) = 0$ for $\forall \tau < 0$ Physically realizable

Wiener Filtering

According to the orthogonality principle, we have

$$E\left\{\left[X(t)-\hat{X}_{\mathrm{LMV}}(t)\right]Z^T(t-\tau)\right\}=0 \quad (0\le\tau\le t)$$

$$\Rightarrow E\left\{\left[X(t)-\int_0^\infty h(\lambda)Z(t-\lambda)d\lambda\right]Z^T(t-\tau)\right\}=0$$

$$\Rightarrow \underline{E\left\{X(t)Z^T(t-\tau)\right\}}=\int_0^\infty h(\lambda)E\underline{\left\{Z(t-\lambda)Z^T(t-\tau)\right\}}d\lambda$$

$$\Rightarrow \underline{R_{XZ}(\tau)}=\int_0^\infty h(\lambda)\underline{R_{ZZ}(\tau-\lambda)}d\lambda$$

Wiener—Hoff Integration Equation

Wiener Filtering

The Wiener—Hoff Integration Equation

$$R_{XZ}(\tau) = \int_0^{\infty} h(\lambda) R_{ZZ}(\tau - \lambda)\, d\lambda$$

needs numerical solution. It is always computation-intensive, especially for the high-dimension case.

The Wiener filter for discrete-time systems is similar.

Consideration: Does the Wiener filter require the linear measurement?

$$Z(\tau) = X(\tau) + V(\tau)$$

Appendix: Stationary Stochastic Process

Strictly stationary

A stationary stochastic process (SSP) X(t) is defined satisfying the following equivalence of the joint PDFs

$$p\left\{X\left(t_1\right), X\left(t_2\right), \ldots, X\left(t_n\right)\right\} = p\left\{X\left(t_1+t_0\right), X\left(t_2+t_0\right), \ldots, X\left(t_n+t_0\right)\right\}$$

for any $n, t_0, t_1, \ldots, t_n$

Widely stationary

A stationary stochastic process (SSP) X(t) is defined satisfying

$$EX\left(t_1\right) = EX\left(t_1+t_0\right), E\left(X\left(t_i\right) X\left(t_j\right)\right) = E\left(X\left(t_i+t_0\right) X\left(t_j+t_0\right)\right)$$

for any t_0, t_1, t_i, t_j

Appendix: Stationary Stochastic Process

Strictly stationary $\not\longrightarrow$ Widely stationary

By the fact that $X(t)$ has the probability density, but may not have the first moments, i.e., mean and variance.

In other words, if the first moments exist, then

Strictly stationary $\longrightarrow$ Widely stationary

Widely stationary $\not\longrightarrow$ Strictly stationary

By the fact that just given the first moments, i.e., mean and variance, it may not be sufficient to obtain the probability density.

In other words, if the first moments is sufficient to determine the PDF, then Widely stationary $\longrightarrow$ Strictly stationary

Appendix: Stationary Stochastic Process

Consider a continuous-time stationary stochastic process (SSP) with the following autocorrelation function

$$R_X(\tau) = E\left(X(t+\tau)X^T(t)\right)$$

The spectral density $\Phi_x(s)$ is defined the value of the following function at the point $s = j\omega$

$$\Phi_X(s) = \int_{-\infty}^{+\infty} R_X(t)e^{-st}dt$$

Appendix: Stationary Stochastic Process

Consider a discrete-time stationary stochastic process with the following autocorrelation function

$$R_X(k) = E\left(X(i+k)X^T(i)\right)$$

The spectral density $\Phi_x(z)$ is defined the value of the following function at the point $z = e^{j\omega}$

$$\Phi_X(s) = \sum_{k=-\infty}^{+\infty} R_X(t) z^{-k}$$

Appendix: Stationary Stochastic Process

If u(t) and y(t) are input and output of a linear system with impulse response h(t) and transfer function H(s), then the correlation functions and spectral densities are related by

$$
\begin{aligned}
R_{YU}(t) &= E\left(y(t+\tau)u^T(\tau)\right) = h(t) * R_U(t) \\
R_Y(t) &= E\left(y(t+\tau)y^T(\tau)\right) = h(t) * R_U(t) * h^T(-t) \\
\Phi_{YU}(s) &= H(s)\Phi_U(s) \\
\Phi_Y(s) &= H(s)\Phi_U(s)H^T(-s)
\end{aligned}
$$

where * represents convolution. In the discrete counterparts, $-s$ is replaced by Z^{-1}.

Appendix: Stationary Stochastic Process

Spectral Factorization Theorem

Let $|\Phi_x(s)| \neq 0$ for almost every s (such a process is said to be of full rank). Then,$\Phi_r(s)$ can be factored uniquely as

$$\Phi_Y(s) = H(s)\Omega H^T(-s)$$

where H(s) is a square, real, rational transfer function with all poles in the open left-half plane; $\lim\limits_{s\to\infty} H(s) = I; H^{-1}(s)$ has all poles in the left-half plane or on the $j\omega$ axis; and $\Omega = \Omega^\tau > 0$ is a real matrix.

In other words, the SSP y(t) can be regarded as the output of the stable system H(s) when the input is white noise with spectral density Ω.

Example 4.2: Discrete Spectral Factorization

The rational (in $\cos w$) spectral density

$$\Phi(w) = \frac{1.25 + \cos w}{1.0625 + 0.5\cos w} = \frac{(e^{jw} + 0.5)(e^{-jw} + 0.5)}{(e^{jw} + 0.25)(e^{-jw} + 0.25)}$$

can be factored in four ways:

$$H_1(z) = \frac{z + 0.5}{z + 0.25} \quad H_2(z) = \frac{z + 0.5}{1 + 0.25z} \quad H_3(z) = \frac{1 + 0.5z}{z + 0.25} \quad H_4(z) = \frac{1 + 0.5z}{1 + 0.25z}$$

Only $H_1(z)$ has no ploes and zeros outside the unit circle. Thus, if the stable system $H_1(z)$ has a white noise input with unit spectral density, the output will have spectral density given by $\Phi(w)$.

Example 4.3: Continuous Spectral Factorization

The rational function of w

$$\Phi(w) = \frac{w^2+4}{w^2+1} = \frac{(jw+2)(-jw+2)}{(jw+1)(-jw+1)}$$

can be factored in four ways; one of which has

$$H(s) = \frac{s+2}{s+1}$$

the minimum phase the system with this transfer function will have output with spectral density $\Phi(w)$ when driven by continuous-time white noise with unit spectral density.