2. Linear Minimum Mean Squared Error Estimation

2.1. Linear minimum mean squared error estimators

- ***Situation considered:***

  - A random sequence $X(1), \ldots, X(M)$ whose realizations can be observed.

  - A random variable Y which has to be estimated.

  - We seek an estimate of Y with a **linear estimator** of the form:

$$\boxed{\hat{Y} = h_0 + \sum_{m=1}^{M} h_m X(m)}\,.$$

  - A measure of the goodness of $\hat{Y}$ is the mean squared error (MSE):

$$\boxed{\mathbf{E}[(\hat{Y} - Y)^2]}\,.$$

- ***Covariance and variance of random variables:***
  Let U and V denote two random variables with expectation $\mu_U \equiv \mathbf{E}[U]$ and $\mu_V \equiv \mathbf{E}[V]$.

  - The **covariance** of U and V is defined to be:

$$\begin{aligned} \Sigma_{UV} &\equiv \mathbf{E}[(U - \mu_U)(V - \mu_V)] \\ &= \mathbf{E}[UV] - \mu_U \mu_V \end{aligned}$$

  - The **variance** of U is defined to be:

$$\begin{aligned} \sigma_U^2 &\equiv \mathbf{E}[(U - \mu_U)^2] = \Sigma_{UU} \\ &= \mathbf{E}[U^2] - (\mu_U)^2 \end{aligned}$$

Let $\boldsymbol{U} \equiv [U(1), \ldots, U(M)]^T$ and $\boldsymbol{V} \equiv [V(1), \ldots, V(M')]^T$ denote two random vectors.
The covariance matrix of $\boldsymbol{U}$ and $\boldsymbol{V}$ is defined as

$$\Sigma_{\boldsymbol{UV}} \equiv \begin{bmatrix} \Sigma_{U(1)V(1)} & \cdots & \Sigma_{U(1)V(M')} \\ \cdots & \cdots & \cdots \\ \Sigma_{U(M)V(1)} & \cdots & \Sigma_{U(M)V(M')} \end{bmatrix}$$

A direct way to obtain $\Sigma_{\boldsymbol{UV}}$:

$$\begin{aligned} \Sigma_{\boldsymbol{UV}} &= \mathbf{E}[(\boldsymbol{U} - \mu_{\boldsymbol{U}})(\boldsymbol{V} - \mu_{\boldsymbol{V}})^T] \\ &= \mathbf{E}[\boldsymbol{U}\boldsymbol{V}^T] - \mu_{\boldsymbol{U}}(\mu_{\boldsymbol{V}})^T \end{aligned}$$

where

$$\begin{aligned} &\mu_{\boldsymbol{U}} \equiv \mathbf{E}[\boldsymbol{U}] = [\mathbf{E}[U(1)], \ldots, \mathbf{E}[U(M)]]^T \\ &\mu_{\boldsymbol{V}} \equiv \mathbf{E}[\boldsymbol{V}] \end{aligned}$$

Examples: $\boldsymbol{U} = \boldsymbol{X} \equiv [X(1), \ldots, X(M)]^T$ and $\boldsymbol{V} = Y$.
In the sequel we shall frequently make use of the following covariance matrix and vector:

(i) $$\begin{aligned} \Sigma_{\boldsymbol{XX}} &= \mathbf{E}[(\boldsymbol{X} - \mu_{\boldsymbol{X}})(\boldsymbol{X} - \mu_{\boldsymbol{X}})^T] \\ &= \begin{bmatrix} \sigma_{X(1)}^2 & \cdots & \Sigma_{X(1)X(M)} \\ \cdots & \cdots & \cdots \\ \Sigma_{X(M)X(1)} & \cdots & \sigma_{X(M)}^2 \end{bmatrix} \end{aligned}$$

(ii) $$\begin{aligned} \Sigma_{\boldsymbol{X}Y} &= \mathbf{E}[(\boldsymbol{X} - \mu_{\boldsymbol{X}})(Y - \mu_Y)] \\ &= \begin{bmatrix} \Sigma_{X(1)Y} & \cdots & \Sigma_{X(M)Y} \end{bmatrix}^T \end{aligned}$$

- ***Linear minimum mean squared error estimator (LMMSEE)***
  A LMMSEE of Y is a linear estimator, i.e. an estimator of the form

$$\hat{Y} = h_0 + \sum_{m=1}^{M} h_m X(m) \;,$$

  which minimizes the MSE $\mathbf{E}[(\hat{Y} - Y)^2]$.

A linear estimator is entirely determined by the $(M+1)$-dimensional vector $\boldsymbol{h} \equiv [h_0, \ldots, h_M]^T$.

- ***Orthogonality principle:***

Orthogonality principle:
A necessary condition for $\boldsymbol{h} \equiv [h_0, \ldots, h_M]^T$ to be the coefficient vector of the LMMSEE is that its entries fulfils the $(M+1)$ identities:

$$\mathbf{E}[Y - \hat{Y}] = \mathbf{E}\left[Y - \left(h_0 + \sum_{m=1}^{M} h_m X(m)\right)\right] = 0 \qquad (2.1a)$$

$$\mathbf{E}[(Y - \hat{Y})X(j)] = \mathbf{E}\left[\left\{Y - \left(h_0 + \sum_{m=1}^{M} h_m X(m)\right)\right\}X(j)\right] = 0, \qquad (2.1b)$$
$$j = 1, \ldots, M$$

Proof:
Because the coefficient vector of the LMMSEE minimizes $\mathbf{E}[(\hat{Y} - Y)^2]$, its components must satisfy the set of equations:

$$\frac{\partial}{\partial h_j}\mathbf{E}[(\hat{Y} - Y)^2] = 0 \qquad j = 0, \ldots, M.$$

□

Notice that the two expressions in (2.1) can be rewritten as:

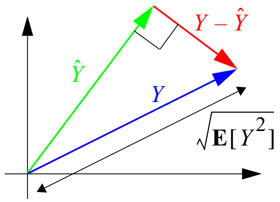
$$\mathbf{E}[Y - \hat{Y}] = 0 \qquad (2.2a)$$

$$\mathbf{E}[(Y - \hat{Y})X(j)] = 0 \qquad j = 1, \ldots, M \qquad (2.2b)$$

Important consequences of the orthogonality principle:

$$\mathbf{E}[(Y - \hat{Y})\hat{Y}] = 0 \qquad (2.3a)$$

$$\mathbf{E}[(Y - \hat{Y})^2] = \mathbf{E}[Y^2] - \mathbf{E}[\hat{Y}^2] = \mathbf{E}[(Y - \hat{Y})Y] \qquad (2.3b)$$



Geometrical interpretation:
Let U and V denote two random variables with finite second moment, i.e. $\mathbf{E}[U^2] < \infty$ and $\mathbf{E}[V^2] < \infty$.
Then, the expectation $\mathbf{E}[UV]$ can be viewed as the **scalar or inner product** of U and V.

Within this interpretation:

- U and V are uncorrelated, i.e. $\mathbf{E}[UV] = 0$ if and only if, they are **orthogonal**,
- $\sqrt{\mathbf{E}[U^2]}$ is the norm (length) of U.

Interpretation of both equations in (2.3):

- (2.3a): the estimation error $Y - \hat{Y}$ and the estimate $\hat{Y}$ are orthogonal.
- (2.3b): results from Pythagoras' Theorem.

- ***Computation of the coefficient vector of the LMMSEE:***
The coefficients of the LMMSEE satisfy the relationships:

$$\mu_Y = h_0 + \sum_{m=1}^{M} h_m \mu_{X(m)} = h_0 + (\boldsymbol{h}^{-})^T \mu_{\boldsymbol{X}}$$

$$\Sigma_{\boldsymbol{X}Y} = \Sigma_{\boldsymbol{XX}}\boldsymbol{h}^{-}$$

where $\boldsymbol{h}^{-} \equiv [h_1, \ldots, h_M]^T$ and $\boldsymbol{X} \equiv [X_1, \ldots, X_M]^T$.

Proof:
Both identities follow by appropriately reformulating relations (2.1a) and (2.1b) and using a matrix notation for the latter one.

□

Thus, provided $(\Sigma_{XX})^{-1}$ exists the coefficients of the LMMSEE are given by:

$$\boxed{\begin{aligned} \boldsymbol{h}^{\bar{}} &= (\Sigma_{XX})^{-1}\Sigma_{XY} && (2.4a) \\ h_0 &= \mu_Y - (\boldsymbol{h}^{\bar{}})^T \mu_{\boldsymbol{X}} = \mu_Y - \Sigma_{XY}{}^T(\Sigma_{XX})^{-1}\mu_{\boldsymbol{X}} && (2.4b) \end{aligned}}$$

- ***Example: Linear prediction of a WSS process***

  Let $Y(n)$ denote a WSS process with

  - zero mean, i.e $\mathbf{E}[Y(n)] = 0$,
  - autocorrelation function $\mathbf{E}[Y(n)Y(n+k)] = R_{YY}(k)$

  We seek the LMMSEE for the present value of $Y(n)$ based on the M past observations $Y(n-1), \ldots, Y(n-M)$ of the process. Hence,

  - $Y = Y(n)$
  - $X(m) = Y(n-m)$, $m = 1, \ldots, M$, i.e.

  $$\boldsymbol{X} = [Y(n-1), \ldots, Y(n-M)]^T$$

  Because $\mu_Y = 0$ and $\mu_{\boldsymbol{X}} = 0$, it follows from $(2.4b)$ that

  $$h_0 = 0$$

  Computation of Σ_{XY} and Σ_{XX}:

  $$-\Sigma_{XY} = [\mathbf{E}[Y(n-1)Y(n)], \ldots, \mathbf{E}[Y(n-M)Y(n)]]^T$$

  $$= [R_{YY}(1), \ldots, R_{YY}(M)]^T$$

  $$-\Sigma_{XX} =$$

  $$= \begin{bmatrix} \mathbf{E}[Y(n-1)^2] & \mathbf{E}[Y(n-1)Y(n-2)] & \ldots & \mathbf{E}[Y(n-1)Y(n-M)] \\ \mathbf{E}[Y(n-2)Y(n-1)] & \mathbf{E}[Y(n-2)^2] & \ldots & \mathbf{E}[Y(n-2)Y(n-M)] \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{E}[Y(n-M)Y(n-1)] & \mathbf{E}[Y(n-M)Y(n-2)] & \ldots & \mathbf{E}[Y(n-M)^2] \end{bmatrix}$$

  $$= \begin{bmatrix} R_{YY}(0) & R_{YY}(1) & R_{YY}(2) & \ldots & R_{YY}(M-1) \\ R_{YY}(1) & R_{YY}(0) & R_{YY}(1) & \ldots & R_{YY}(M-2) \\ R_{YY}(2) & R_{YY}(1) & R_{YY}(0) & \ldots & R_{YY}(M-3) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ R_{YY}(M-1) & R_{YY}(M-2) & R_{YY}(M-3) & \ldots & R_{YY}(0) \end{bmatrix}$$

- ***Residual error using a LMMSEE:***

  The MSE resulting when using a LMMSEE is

  $$\boxed{\mathbf{E}[(\hat{Y} - Y)^2] = \sigma_Y{}^2 - (\boldsymbol{h}^{\bar{}})^T\Sigma_{XY} \qquad (2.5)}$$

  Proof:

  $$\begin{aligned} \mathbf{E}[(Y-\hat{Y})^2] &= \mathbf{E}[(Y-\hat{Y})Y] \\ &= \mathbf{E}[Y^2] - \mathbf{E}[\hat{Y}Y] \end{aligned}$$

  □

2.2. Minimum mean squared error estimators

- ***Conditional expectation:***

  Let U and V denote two random variables.

  The conditional expectation of V given $U = u$ is observed is defined to be

  $$\mathbf{E}[V|u] \equiv \int vp(v|u)dv.$$

  Notice that $\mathbf{E}[V|U]$ is a random variable. In the sequel we shall make use of the following important property of conditional expectations:

  $$\mathbf{E}[\mathbf{E}[V|U]] = \mathbf{E}[V]$$

  Proof:

  □

- ***Minimum mean squared error estimator (MMSEE):***
  The MMSEE of Y based on the observation of $X(1), \ldots, X(M)$ is of the form:

$$\boxed{\widehat{Y}(X(1), \ldots, X(M)) = \mathbf{E}[Y|X(1), \ldots, X(M)]}$$

Hence if $X(1) = x(1), \ldots, X(M) = x(M)$ is observed, then

$$\begin{aligned} \widehat{Y}(x(1), \ldots, x(M)) &= \mathbf{E}[Y|x(1), \ldots, x(M)] \\ &= \int y p(y|x(1), \ldots, x(M)) dy \end{aligned}$$

Proof:

Let $\hat{Y}$ denote an arbitrary estimator. Then,

$$\begin{aligned} \mathbf{E}[(\hat{Y}-Y)^2] &= \mathbf{E}[((\hat{Y}-\widehat{Y})-(Y-\widehat{Y}))^2] \\ &= \mathbf{E}[(\hat{Y}-\widehat{Y})^2] - \underbrace{2\mathbf{E}[(\hat{Y}-\widehat{Y})(Y-\widehat{Y})]}_{=0} + \mathbf{E}[(Y-\widehat{Y})^2] \\ &= \mathbf{E}[(\hat{Y}-\widehat{Y})^2] + \mathbf{E}[(Y-\widehat{Y})^2] \end{aligned}$$

Thus,

$$\mathbf{E}[(\hat{Y}-Y)^2] \geq \mathbf{E}[(Y-\widehat{Y})^2]$$

with equality if, and only if, $\hat{Y} = \widehat{Y}$. We still have to prove that $\mathbf{E}[(\hat{Y}-\widehat{Y})(Y-\widehat{Y})] = 0$.

$\square$

Example: Multivariate Gaussian variables:

$[Y, X(1), \ldots, X(M)]^T \sim N(\mu, \Sigma)$ with

- $\mu \equiv [\mu_Y, \mu_{X(1)}, \ldots, \mu_{X(M)}]^T$
- $\Sigma \equiv \begin{bmatrix} \sigma_Y^2 & (\Sigma_{\boldsymbol{X}Y})^T \\ \Sigma_{\boldsymbol{X}Y} & \Sigma_{\boldsymbol{XX}} \end{bmatrix}$

From Equation (6.22) in [Shanmugan] it follows that

$$\boxed{\widehat{Y} = \mathbf{E}[Y|\boldsymbol{X}] = \mu_Y + (\Sigma_{\boldsymbol{X}Y})^T(\Sigma_{\boldsymbol{XX}})^{-1}(\boldsymbol{X}-\mu_{\boldsymbol{X}})}$$

Bivariate case: $M = 1, X(1) = X$

- $\Sigma_{XX} = \sigma_X^2$,
- $\Sigma_{XY} = \rho\sigma_X\sigma_Y$, where $\rho \equiv \dfrac{\Sigma_{XY}}{\sigma_X\sigma_Y}$ is the correlation coefficient of Y and X.

In this case,

$$\begin{aligned} \widehat{Y} = \mathbf{E}[Y|X] &= \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(X-\mu_X) \\ &= \underbrace{\left(\mu_Y - \frac{\rho\sigma_Y}{\sigma_X}\mu_X\right)}_{h_0} + \underbrace{\left(\frac{\rho\sigma_Y}{\sigma_X}\right)}_{h_1} X \end{aligned}$$

We can observe that $\widehat{Y}$ is linear, i.e. is the LMMSEE $\hat{Y} = \widehat{Y}$ in the bivariate case. This is also true in the general multivariate Gaussian case. In fact,

> $\hat{Y} = \widehat{Y}$ if, and only if, $[Y, X(1), \ldots, X(M)]^T$ is a Gaussian random vector.

2.3. Time-discrete Wiener filters

- ***Problem:***

  Estimation of a WSS random sequence $Y(n)$ based on the observation of another sequence $X(n)$. Without loss of generality we assume that $\mathbf{E}[Y(n)] = \mathbf{E}[X(n)] = 0$.

  The goodness of the estimator $\hat{Y}(n)$ is described by the MSE

$$\mathbf{E}[(\hat{Y}(n) - Y(n))^2].$$

  We distinguish between two cases:

  - **Prediction:**

    $\hat{Y}(n)$ depends on one or several past observations of $X(n)$ only, i.e.

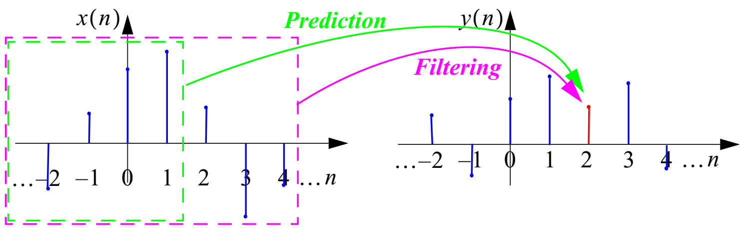
$$\hat{Y}(n) = \hat{Y}(X(n_1), X(n_2), \ldots) \text{ with } n_1, n_2, \ldots < n$$

  - **Filtering:**

    $\hat{Y}(n)$ depends on the present observation and/or one or many future observation(s) of $X(n)$, i.e.

$$\hat{Y}(n) = \hat{Y}(X(n_1), X(n_2), \ldots) \text{ where at least one } n_i \geq n$$

  If all $n_i \leq n$, the filter is **causal** otherwise it is **noncausal**.



Typical application: WSS signal embedded in additive white noise

$$X(n) = Y(n) + W(n) .$$

where,

- $W(n)$ is a white noise sequence,
- $Y(n)$ is a WSS process
- $Y(n)$ and $W(n)$ are uncorrelated.

However, the theoretical treatment is more general as shown below.

2.3.1. Noncausal Wiener filters

- ***Linear Minimum Mean Squared Error Filter***

  We seek a linear filter

$$\hat{Y}(n) = \sum_{m=-\infty}^{\infty} h(m)X(n-m) = h(n)*X(n)$$

  which minimizes the MSE $\mathbf{E}[(\hat{Y}(n) - Y(n))^2]$.

  Such a filter exists. It is called a **Wiener filter** in honour of his inventor.

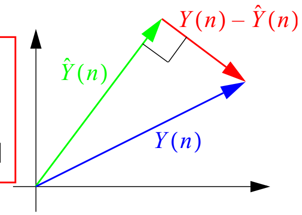
- ***Orthogonality principle (time domain):***

  The coefficients of a Wiener filter satisfy the conditions:

$$\boxed{\begin{aligned}&\mathbf{E}[(Y(n) - \hat{Y}(n))X(n-k)] = \\ &\quad = \mathbf{E}\left[\left(Y(n) - \sum_{m=-\infty}^{\infty} h(m)X(n-m)\right)X(n-k)\right] = 0, \quad k = \ldots, -1, 0, 1, \ldots\end{aligned}}$$

  It follows from these identities (see also (2.3)) that

$$\boxed{\begin{aligned}&\mathbf{E}[(Y(n) - \hat{Y}(n))\hat{Y}(n)] = 0 \\ &\mathbf{E}[(Y(n) - \hat{Y}(n))^2] = \mathbf{E}[Y(n)^2] - \mathbf{E}[\hat{Y}(n)^2] \\ &\qquad = \mathbf{E}[(Y(n) - \hat{Y}(n))Y(n)]\end{aligned}}$$



  With the definitions

$$R_{XX}(k) \equiv \mathbf{E}[X(n)X(n+k)],\ R_{XY}(k) \equiv \mathbf{E}[X(n)Y(n+k)]$$

  we can recast the orthogonality conditions as follows:

$$\boxed{\begin{aligned}&R_{XY}(k) = \sum_{m=-\infty}^{\infty} h(m)R_{XX}(k-m) \qquad k = \ldots, -1, 0, 1, \ldots \\ &R_{XY}(k) = h(k)*R_{XX}(k) \qquad \text{Wiener-Hopf equation}\end{aligned}}$$

- ***Orthogonality principle (frequency domain):***

$$\boxed{S_{XY}(f) = H(f)S_{XX}(f)}$$

where

$$S_{XY}(f) \equiv F\{R_{XY}(k)\}$$

- ***Transfer function of the Wiener filter:***

$$\boxed{H(f) = \frac{S_{XY}(f)}{S_{XX}(f)}}$$

- ***MSE of the Wiener filter (time-domain formulation):***

$$\boxed{\mathbf{E}[(Y(n)-\hat{Y}(n))^2] = \sigma_Y^2 - \sum_{m=-\infty}^{\infty} h(m)R_{XY}(m)}$$

Proof:

$$\mathbf{E}[(Y(n)-\hat{Y}(n))^2] = \mathbf{E}[Y(n)^2] - \mathbf{E}[\hat{Y}(n)Y(n)]$$

$\square$

- ***MSE of the Wiener filter (frequency-domain formulation):***
We can rewrite the above identity as:

$$\mathbf{E}[(Y(n)-\hat{Y}(n))^2] = R_{YY}(0) - \sum_{m=-\infty}^{\infty} h(m)R_{YX}(-m)$$

$\mathbf{E}[(Y(n)-\hat{Y}(n))^2]$ is the value $p(0)$ of the sequence

$$p(k) = R_{YY}(k) - \sum_{m=-\infty}^{\infty} h(m)R_{YX}(k-m)$$

$$= R_{YY}(k) - h(k) * R_{YX}(k)$$

$$\circ\!\!-\!\!\bullet$$

$$P(f) = S_{YY}(f) - H(f)S_{YX}(f) = S_{YY}(f) - \frac{|S_{XY}(f)|^2}{S_{XX}(f)}$$

Hence,

$$\mathbf{E}[(Y(n)-\hat{Y}(n))^2] = p(0) = \int_{-1/2}^{1/2} P(f)df$$

$$\boxed{\mathbf{E}[(Y(n)-\hat{Y}(n))^2] = \int_{-1/2}^{1/2}\left[S_{YY}(f) - \frac{|S_{XY}(f)|^2}{S_{XX}(f)}\right]df}$$

2.3.2. Causal Wiener filters

A. $X(n)$ is a white noise.

We first assume that $X(n)$ is a white noise with unit variance, i.e.
$\mathbf{E}[X(n)X(n+k)] = \delta(k)$.

- ***Derivation of the causal Wiener filter from the noncausal Wiener filter:***
Let us consider the noncausal Wiener filter

$$\hat{Y}(n) = \sum_{m=-\infty}^{\infty} h(m)X(n-m)$$

whose transfer function is given by

$$H(f) = \frac{S_{XY}(f)}{S_{XX}(f)} = S_{XY}(f).$$

Then, the causal Wiener filter $\hat{Y}_c(n)$ results by cancelling the noncausal part of the non-causal Wiener filter:

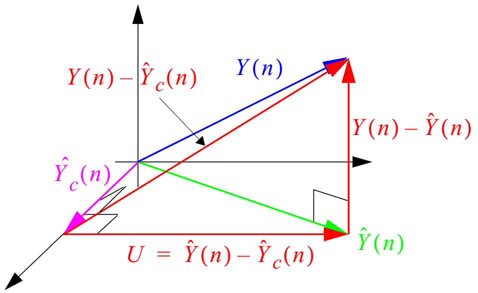
$$\boxed{\hat{Y}_c(n) = \sum_{m=0}^{\infty} h(m)X(n-m)}$$

Sketch of the proof:

$\hat{Y}(n)$ can be written as

$$\hat{Y}(n) = \underbrace{\sum_{m=-\infty}^{-1} h(m)X(n-m)}_{\equiv U} + \underbrace{\sum_{m=0}^{\infty} h(m)X(n-m)}_{V \equiv \hat{Y}_c(n)}$$

Because $X(n)$ is a white noise, the causal part $V = \hat{Y}_c(n)$ and the noncausal part $U = \hat{Y}(n) - \hat{Y}_c(n)$ of $\hat{Y}(n)$ are orthogonal. It follows from this property



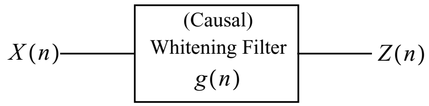
that $\hat{Y}_c(n)$ and $Y(n)$ are orthogonal, i.e. that $\hat{Y}_c(n)$ minimizes the MSE within the class of linear causal estimators.

□

B. $X(n)$ **is an arbitrary WSS process whose spectrum satisfies the Paley-Wiener condition.**

Usually, the above truncation procedure to obtain $\hat{Y}_c(n)$ does not apply because U and V are correlated and therefore not orthogonal in the general case.

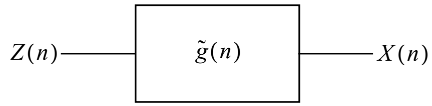
- ***Causal whitening filter:***
  However, we can show (see the Spectral Decomposition Theorem below) that provided $S_{XX}(f)$ satisfies the Paley-Wiener condition (see below) then $X(n)$ can be **converted into an equivalent white noise sequence** $Z(n)$ **with unit variance** by filtering it with an appropriate causal filter $g(n)$,



$\mathbf{E}[Z(n)Z(n+k)] = \delta(k)$

This operation is called **whitening** and the filter $g(n)$ is called a **whitening filter**.

equivalent $\equiv$ there exists another causal filter $\tilde{g}(n)$ so that $X(n) = \tilde{g}(n) * Z(n)$:



Notice that if

$$G(f) \equiv F\{g(n)\}$$

$$\tilde{G}(f) \equiv F\{\tilde{g}(n)\}$$

then

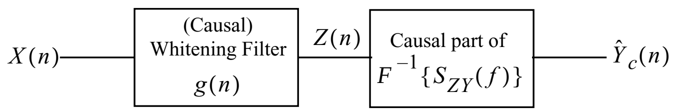
$$|G(f)|^2 = S_{XX}(f)^{-1}$$

$$|\tilde{G}(f)|^2 = S_{XX}(f) \qquad (2.6)$$

We shall see that a whitening filter exists such that

$$\tilde{G}(f) = G(f)^{-1}$$

- ***Causal Wiener filter***
  Making use of the result in Part A, the block diagram of the causal Wiener filter is



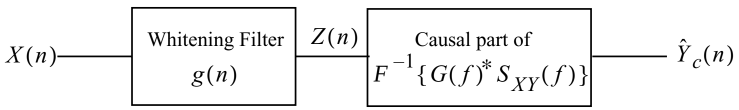
$S_{ZY}(f)$ is obtained from $S_{XY}(f)$ according to

$$\boxed{S_{ZY}(f) = G(f)^* S_{XY}(f)}$$

Proof:

$\square$

Hence, the block diagram of the causal Wiener filter is:



- ***Spectral Decomposition Theorem:***

Let $S_{XX}(f)$ satisfies the so-called **Paley-Wiener condition**:

$$\boxed{\int_{-1/2}^{1/2} \log(S_{XX}(f))df > -\infty}$$

Then $S_{XX}(f)$ can be written as

$$S_{XX}(f) = G(f)^{+}G(f)^{-}$$

with $G(f)^{+}$ and $G(f)^{-}$ satisfying

$$\boxed{|G(f)^{+}|^2 = |G(f)^{-}|^2 = S_{XX}(f)}.$$

Moreover, the sequences

$$\boxed{\begin{aligned} g(n)^{+} &\equiv F^{-1}\{G(f)^{+}\} \\ g(n)^{-} &\equiv F^{-1}\{G(f)^{-}\} \\ g^{-1}(n)^{+} &\equiv F^{-1}\{1/G(f)^{+}\} \\ g^{-1}(n)^{-} &\equiv F^{-1}\{1/G(f)^{-}\} \end{aligned}}$$

satisfy

$$\boxed{\begin{array}{lll} g(n)^{+} = g^{-1}(n)^{+} = 0 & n < 0 & \text{Causal sequences} \\ g(n)^{-} = g^{-1}(n)^{-} = 0 & n > 0 & \text{Anticausal sequences} \end{array}}$$

- ***Whitening filter (cont'd):***

The sought whitening filter used to obtained $Z(n)$ is

$$\boxed{g(n) = g^{-1}(n)^{+}}$$

and

$$\tilde{g}(n) = g(n)^{+}.$$

It can be easily verified that both sequences satisfy the identities in (2.6).

2.3.3. Finite Wiener filters

- ***Finite linear filter:***

$$\hat{Y}(n) = \sum_{m=-M_1}^{M_2} h(m)X(n-m)$$

- ***Wiener-Hopf equation:***

By applying the orthogonality principle we obtain the Wiener-Hopf system of equations:

$$\boxed{\Sigma_{XY} = \Sigma_{XX}\boldsymbol{h}}$$

where

$$\boldsymbol{h} \equiv [h(-M_1), \ldots, h(M_2)]^T$$

and

$$\Sigma_{XY} \equiv [R_{XY}(-M_1), \ldots, R_{XY}(M_2)]^T$$

$$\Sigma_{XX} \equiv$$

$$\equiv \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & R_{XX}(2) & \ldots & R_{XX}(M_1+M_2) \\ R_{XX}(1) & R_{XX}(0) & R_{XX}(1) & \ldots & R_{XX}(M_1+M_2-1) \\ R_{XX}(2) & R_{XX}(1) & R_{XX}(0) & \ldots & R_{XX}(M_1+M_2-2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ R_{XX}(M_1+M_2) & R_{XX}(M_1+M_2-1) & R_{XX}(M_1+M_2-2) & \ldots & R_{XX}(0) \end{bmatrix}$$

Coefficient vector of the finite Wiener filter:

$$\boxed{\boldsymbol{h} = (\Sigma_{XX})^{-1}\Sigma_{XY}}$$

provided Σ_{XX} is invertible.