I225E Statistical Signal Processing

12. Linear Model

MAWALIM and UNOKI

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

School of Information Science

JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY
HOKURIKU 1990

Linear model

- Prediction and smoothing
- Non-causal and causal Wiener filter
- (Innovation filter and Laplace transform)
- Linear minimum mean squared error (LMMSE) prediction
- Extension to vector form

1. Signal Processing

Problem: Using stochastic process $X(\xi)$, $(a \le \xi \le b)$, estimate the value of stochastic process $S(t)$ as

$$\hat{S}(t) = \int_a^b h(\alpha) X(\alpha) d\alpha$$

- ***Smoothing***, in case estimation time t is within the data interval ($t \in [a, b]$).
- ***Prediction***, in case estimation time t is outside of the data interval ($t \notin [a, b]$) and $X(t) = S(t)$.
- ***Smoothing and Filtering***, in case estimation time t is outside of the data interval ($t \notin [a, b]$) and $X(t) \neq S(t)$.

Example 1

$S(t)$ is a stationary process.

Predict the future state $S(t+\tau)$ using the past state $S(t)$ as

$$\hat{S}(t+\tau) = aS(t).$$

Find the optimal coefficient a.

By orthogonality principle,

$$E\{\big(S(t+\tau) - aS(t)\big)S(t)\} = 0$$

$$E\{S(t+\tau)S(t)\} - aE\{S(t)S(t)\} = 0$$

$$R(\tau) - aR(0) = 0$$

Hence,

$$a = \frac{R(\tau)}{R(0)}$$

Corresponding mean square error is,

$$\begin{aligned}\text{MSE} &= E\{(S(t+\tau) - aS(t))S(t+\tau)\} \\ &= E\{S(t+\tau)S(t+\tau)\} - aE\{S(t)S(t+\tau)\} \\ &= R(0) - aR(\tau) \\ &= \frac{R^2(0)-R^2(\tau)}{R(0)}\end{aligned}$$

Example 2

Smoothing the present state of $S(t)$ using the present state of another process $X(t)$ as

$$\hat{S}(t) = aX(t).$$

Find the optimal coefficient a.

By orthogonality principle,

$$E\{\big(S(t) - aX(t)\big)X(t)\} = 0$$

$$E\{S(t)X(t)\} - aE\{X(t)X(t)\} = 0$$

$$R_{SX}(0) - aR_{XX}(0) \ = 0$$

Hence,

$$a = \frac{R_{SX}(0)}{R_{XX}(0)}$$

Corresponding mean square error is,

$$\begin{aligned}\mathrm{MSE} &= E\{(S(t) - aX(t))S(t)\} \\ &= E\{S(t)S(t)\} - aE\{X(t)S(t)\} \\ &= R_{SS}(0) - aR_{SX}(0) \\ &= \frac{R_{SS}(0)R_{XX}(0) - R_{SX}^2(0)}{R_{XX}(0)}\end{aligned}$$

2. Smoothing

Problem

Using the entire interval ($t \in [-\infty, \infty]$) of the following process

$$X(t) = S(t) + v(t),$$

estimate the present state of $S(t)$ as

$$\hat{S}(t) = \int_{-\infty}^{\infty} h(\alpha) X(t-\alpha)\, d\alpha.$$

Find the optimal filter $h(\alpha)$.

Solution

By orthogonality principle,

$$S(t) - \hat{S}(t) \perp X(\xi), \text{ for all } \xi \in [-\infty, \infty].$$

Denote $\xi = t - \tau$, For all τ

$$E\left\{\left[S(t) - \int_{-\infty}^{\infty} h(\alpha) X(t-\alpha) d\alpha\right] X(t-\tau)\right\} = 0$$

$$E\{S(t)X(t-\tau)\} - \int_{-\infty}^{\infty} h(\alpha) E\{X(t-\alpha)X(t-\tau)\} d\alpha$$
$$= 0$$

$$R_{SX}(\tau) - \int_{-\infty}^{\infty} h(\alpha) R_{XX}(\tau - \alpha) d\alpha = 0$$

(Wiener-Hopf equation)

Fourier transform of $R_{SX}(\tau) = h(\tau) * R_{XX}(\tau)$ gives

$$S_{SX}(\omega) = H(\omega) S_{XX}(\omega).$$

Hence,

$$\boxed{H(\omega) = \frac{S_{SX}(\omega)}{S_{XX}(\omega)}.}$$

This is called ***Non-causal Wiener Filter***.

Corresponding mean square error is

$$
\begin{aligned}
Bmse &= E\left\{\left(S(t) - \int_{-\infty}^{\infty} h(\alpha)X(t-\alpha)d\alpha\right)S(t)\right\} \\
&= E\{S(t)S(t)\} - \int_{-\infty}^{\infty} h(\alpha)E\{X(t-\alpha)S(t)\}d\alpha \\
&= R_{SS}(0) - \int_{-\infty}^{\infty} h(\alpha)R_{XS}(-\alpha)d\alpha \\
&= R_{SS}(\tau) - \int_{-\infty}^{\infty} h(\alpha)R_{XS}(\tau-\alpha)d\alpha\Big|_{\tau=0} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} [S_{SS}(\omega) - H(\omega)S_{XS}(\omega)]e^{i\omega\tau}d\omega\Big|_{\tau=0} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_{SS}(\omega)S_{XX}(\omega) - S_{SX}^2(\omega)}{S_{XX}(\omega)}d\omega
\end{aligned}
$$

In case that process $S(t)$ is orthogonal to noise $\nu(t)$.

$$R_{S\nu}(\tau) = 0.$$

This leads to

$$\begin{aligned} R_{SX}(\tau) &= R_{SS}(\tau) + R_{S\nu}(\tau) \\ &= R_{SS}(\tau), \\ R_{XX}(\tau) &= R_{SX}(\tau) + R_{\nu X}(\tau) \\ &= R_{SX}(\tau) + R_{\nu S}(\tau) + R_{\nu\nu}(\tau) \\ &= R_{SX}(\tau) + R_{\nu\nu}(\tau). \end{aligned}$$

Therefore,

$$S_{SX}(\omega) = S_{SS}(\omega),$$
$$S_{XX}(\omega) = S_{SS}(\omega) + S_{\nu\nu}(\omega).$$

By substituting the above equations,

$$H(\omega) = \frac{S_{SS}(\omega)}{S_{SS}(\omega)+S_{\nu\nu}(\omega)},$$
$$\mathrm{MSE} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_{SS}(\omega)S_{\nu\nu}(\omega)}{S_{SS}(\omega)+S_{\nu\nu}(\omega)}\,d\omega.$$

3. Prediction by Innovation Filter

- **Problem**:

Using past state of $S(t-\tau)$, $\tau \in [0, \infty]$, estimate the future state of $S(t+\lambda)$ as

$$\hat{S}(t+\lambda) = \int_0^{\infty} h(\alpha) S(t-\alpha) d\alpha$$

Find the optimal filter $h(\alpha)$.

- **Solution**:

By orthogonality principle,

$$S(t+\lambda) - \hat{S}(t+\lambda) \perp S(\xi) \quad \text{for all } \xi \in [-\infty, 0]$$

Denote $\xi = t - \tau$. For positive value of τ ($\tau \geq 0$),

$$E\left\{\left[S(t+\lambda) - \int_0^\infty h(\alpha)S(t-\alpha)d\alpha\right]S(t-\tau)\right\} = 0$$

$$E\{S(t+\lambda)S(t-\tau)\}$$
$$- \int_0^\infty h(\alpha)E\{S(t-\alpha)S(t-\tau)\}d\alpha = 0$$

$$R_{SS}(\tau+\lambda) - \int_0^\infty h(\alpha)R_{SS}(\tau-\alpha)d\alpha = 0$$

Solution $h(\alpha)$ of Wiener-Hopf integral equation gives ***Causal Wiener Filter***.

Because the integration range of τ is one-sided, transformation is not straightforward.

Solution by Innovation Filter

If stochastic process $S(t)$ is regular, it can be represented as output of innovation filter $L(s)$, to which white noise $I(t)$ is applied as input.

$$S(t+\lambda) = \int_0^\infty l(\alpha) I(t+\lambda-\alpha) d\alpha$$

Removing range $[0,\lambda]$, integration from past provides the optimal prediction.

$$\hat{S}(t+\lambda) = \int_\lambda^\infty l(\alpha) I(t+\lambda-\alpha) d\alpha$$
$$= \int_0^\infty l(\beta+\lambda) I(t-\beta) d\beta.$$

Proof

Prediction error

$$S(t+\lambda) - \hat{S}(t+\lambda) = \int_0^{\lambda} l(\alpha) I(t+\lambda-\alpha) d\alpha$$

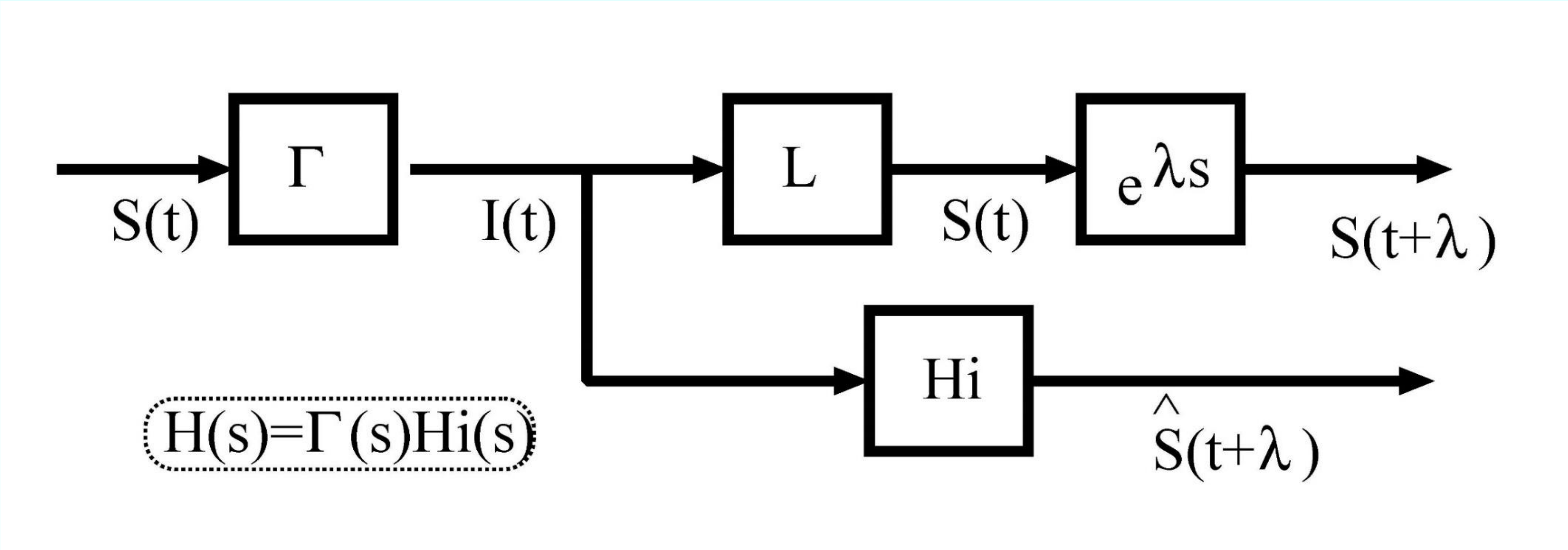
depends only upon interval $t \in (t, t+\lambda)$. They are orthogonal to past $I(t)$ or past $S(t)$.

Construction

Prediction $\hat{S}(t+\lambda)$ is a response of the following filter with white noise input $I(t)$:

$$h_i(t) = l(t+\lambda) U(t),$$

$$H_i(s) = \int_0^{\infty} h_i(t) e^{-st} dt.$$



Therefore, for input $S(t)$, corresponding filter is given by

$$H(s) = \frac{H_i(s)}{L(s)}$$

- Procedure
  1. Factorization: $S_{SS}(s) = L(s)L(-s)$
  2. Find $l(t)$ by inverse Laplace transform of $L(s)$, and construct $h_i(t) = l(t+\lambda)U(t)$.
  3. Find $H_i(s)$ by Laplace transform of $h_i(t)$, and determine $H(s) = H_i(s)/L(s)$.

- Error

$$\mathrm{MSE} = E\left\{\left|\int_0^\lambda l(\alpha)I(t+\lambda-\alpha)d\alpha\right|^2\right\}$$

$$= \int_0^\lambda l^2(\alpha)d\alpha$$

Example

Suppose autocorrelation function of $S(t)$ is given by $R_{SS}(\tau) = 2\alpha e^{-\alpha|\tau|}$. Construct its prediction filter $H(s)$.

1. Spectrum of $S(t)$ is $S_{SS}(s) = \frac{1}{\alpha^2 - s^2}$. By factorization, innovation filter is $L(s) = \frac{1}{\alpha + s}$.
2. By inverse Laplace transform, $l(t) = e^{-\alpha t}U(t)$. This leads to $h_i(t) = l(t + \lambda) = e^{-\alpha\lambda}e^{-\alpha t}U(t)$.
3. By Laplace transform, $H_i(s) = \frac{e^{-\alpha t}}{\alpha + s}$.

Hence,

$$H(s) = \frac{H_i(s)}{L(s)} = e^{-\alpha\lambda}.$$

By inverse Laplace transform, $h(t) = e^{-\alpha\lambda}\delta(t)$;
this leads to $\hat{S}(t+\lambda) = e^{-\alpha\lambda}S(t)$.

$\rightarrow$ For prediction, only present state $S(t)$ is important; past state has no influence.

4. Linear Prediction

- **Problem** Given time series of $X[0], X[1], \cdots, X[N-1]$, predict one-step future state $X[N]$ as

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

Find the optimal coefficients $\boldsymbol{a}$.

- **Solution** Error

$$\begin{aligned} \epsilon &= X[N] - \hat{X}[N] = X[N] + \sum_{n=0}^{N-1} a_n X[n] \\ &= a_0 X[0] + a_1 X[1] + \cdots + a_{N-1} X[N-1] + X[N] \\ &= \sum_{n=0}^{N} a_n X[n] \qquad \text{(where } a_N = 1\text{)} \end{aligned}$$

is orthogonal to data

$$E\{\epsilon X^*[k]\} = \sum_{n=0}^{N} a_n E\{X[n] X^*[k]\} = 0$$

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

Supposing $X[n]$ is stationary,

$$E\{X[n]X^*[k]\} = R(n-k) = r_{n-k} = r_{k-n}^*$$

Therefore,

$$E\{\epsilon X^*[k]\} = \sum_{n=0}^{N} a_n r_{n-k} = 0,$$

$$a_N = 1,\ k = 0,1,\cdots,N-1.$$

$$a_0 r_0 + a_1 r_1 + a_2 r_2 + \cdots + a_{N-1} r_{N-1} + r_N = 0$$

$$a_0 r_1^* + a_1 r_0 \ \ + a_2 r_1 \ \ + \cdots + a_{N-1} r_{N-2} + r_{N-1} = 0$$

$$\vdots$$

$$a_0 r_{N-1}^* + a_1 r_{N-2}^* + a_2 r_{N-3}^* + \cdots + a_{N-1} r_0 \ \ + r_1 = 0$$

Minimum squared error is

$$\begin{aligned}\sigma^2 = E\{|\epsilon|^2\}\\ &= E\{\epsilon(X^*[N] + \textstyle\sum_{n=0}^{N-1} a_n^* X^*[n])\}\\ &= E\{\epsilon X^*[N]\} + \textstyle\sum_{n=0}^{N-1} a_n^* E\{\epsilon X^*[n]\}\\ &= E\{\epsilon X^*[N]\}\\ &= E\{(\textstyle\sum_{n=0}^{N} a_n X[n])X^*[N]\}\\ &= \textstyle\sum_{n=0}^{N} a_n E\{X[n]X^*[N]\}\\ &= \textstyle\sum_{n=0}^{N} a_n r_{N-n}^*\\ &= a_0 r_N^* + a_1 r_{N-1}^* + a_2 r_{N-2}^* + \cdots + a_{N-1} r_1^* + r_0\end{aligned}$$

In matrix representation,

$$\begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_N \\ r_1^* & r_0 & r_1 & \cdots & r_{N-1} \\ r_2^* & r_1^* & r_0 & \cdots & r_{N-2} \\ & & & \vdots & \\ r_{N-1}^* & r_{N-2}^* & \cdots & r_0 & r_1 \\ r_N^* & r_{N-1}^* & \cdots & r_1^* & r_0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{N-1} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \sigma^2 \end{bmatrix}$$

$(N+1) \times (N+1)$ matrix on the left-hand side is called ***Toeplitz*** matrix.

Denoting the Toeplitz matrix as T and its inverse as

$$\boldsymbol{T}^{-1} = \begin{bmatrix} T_{0,0}^{-1} & T_{0,1}^{-1} & \cdots & T_{0,N}^{-1} \\ T_{1,0}^{-1} & T_{1,1}^{-1} & \cdots & T_{1,N}^{-1} \\ & & \vdots & \\ T_{N,0}^{-1} & T_{N,1}^{-1} & \cdots & T_{N,N}^{-1} \end{bmatrix}$$

then $$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \\ 1 \end{bmatrix} = \boldsymbol{T}^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} T_{0,N}^{-1} \\ T_{1,N}^{-1} \\ \vdots \\ T_{N-1,N}^{-1} \\ T_{N,N}^{-1} \end{bmatrix}$$

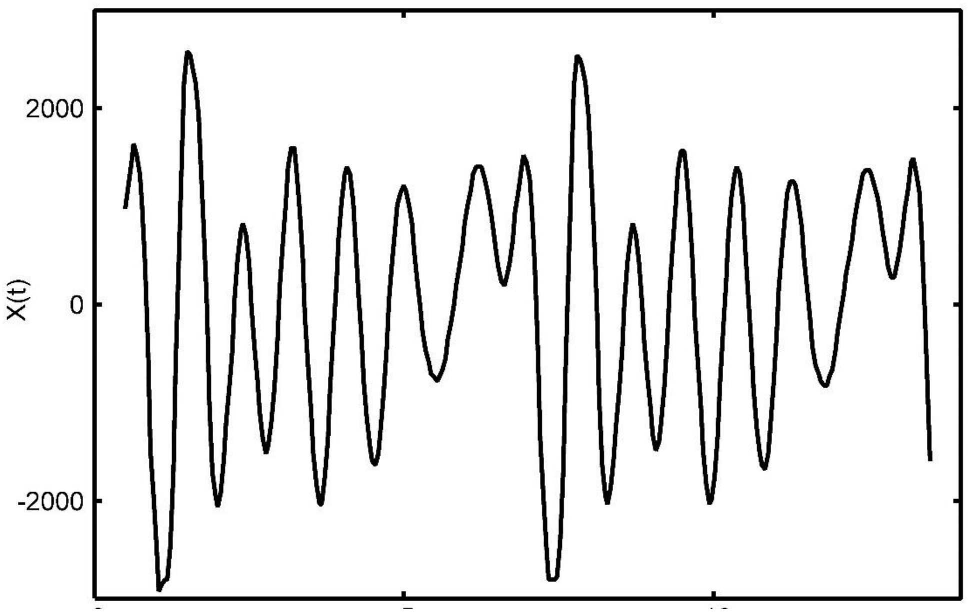
Therefore, coefficients a_i and minimum mean square error are given by

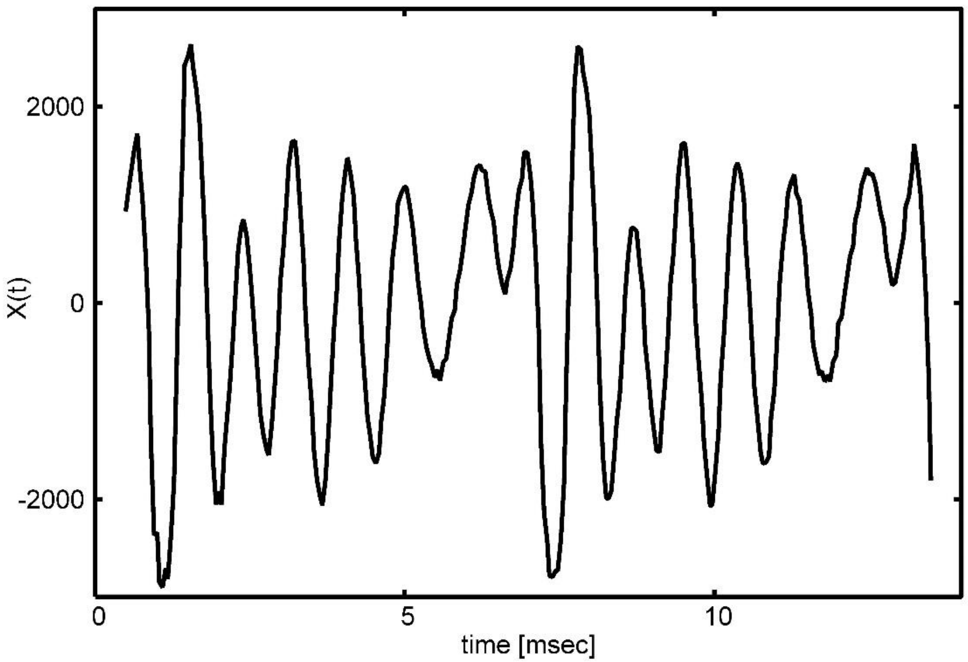
$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix} = \frac{1}{T_{N,N}^{-1}} \begin{bmatrix} T_{0,N}^{-1} \\ T_{1,N}^{-1} \\ \vdots \\ T_{N-1,N}^{-1} \end{bmatrix},$$

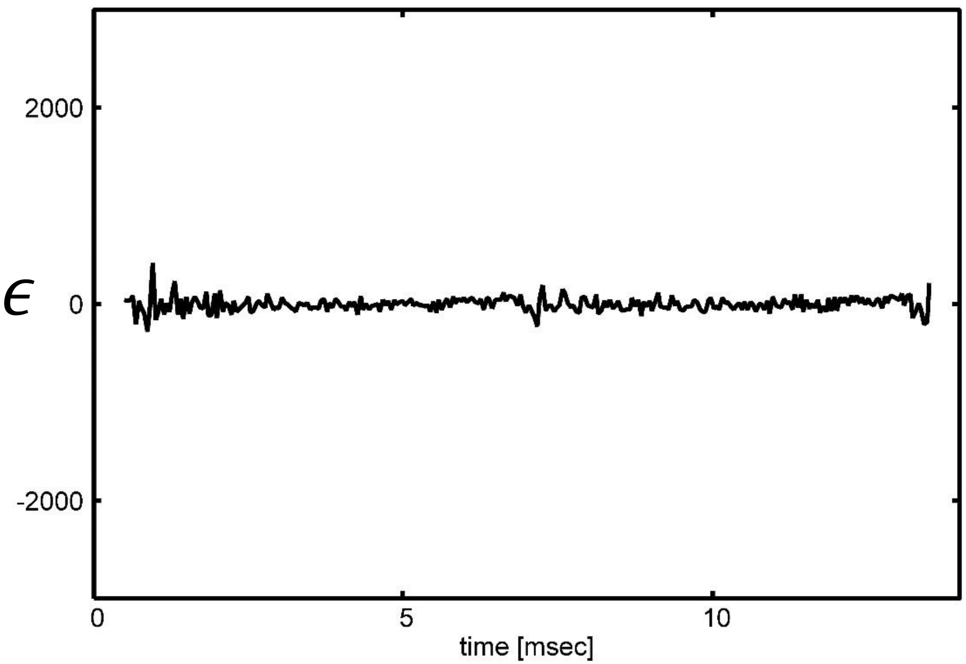
$$\sigma^2 = \frac{1}{T_{N,N}^{-1}}$$

Application of ***Linear prediction*** to speech

(left top: original)
(left bottom: LP)







5. Extension of LMMSE to Vector Form

Each element of random vector $\theta = \{\theta_1, \theta_2, \cdots, \theta_p\}$ is estimated from N data $X[0], X[1], \cdots, X[N-1]$ as

$$\theta_i = \sum_{n=0}^{N-1} a_{in} X[n] + a_{iN} \ (i = 1,2,\cdots,p).$$

From the scalar version of LMMSE,

$$\hat{\theta}_i = E(\theta_i) + C_{\theta_i X} C_{XX}^{-1} (\boldsymbol{X} - E(\boldsymbol{X})),$$

$$Bmse(\hat{\theta}_i) = C_{\theta_i \theta_i} - C_{\theta_i X} C_{XX}^{-1} C_{X \theta_i}.$$

In vector representation,

$$\hat{\theta} = \begin{bmatrix} E(\theta_1) \\ E(\theta_2) \\ \vdots \\ E(\theta_p) \end{bmatrix} + \begin{bmatrix} C_{\theta_1 X} \\ C_{\theta_2 X} \\ \vdots \\ C_{\theta_p X} \end{bmatrix} C_{XX}^{-1}\left(\boldsymbol{X} - E(\boldsymbol{X})\right)$$

$$= E(\theta) + \boldsymbol{C}_{\theta X}\boldsymbol{C}_{XX}^{-1}\left(\boldsymbol{X} - E(\boldsymbol{X})\right).$$

Error matrix: $M_{\hat{\theta}} = E\left(\left(\theta - \hat{\theta}\right)\left(\theta - \hat{\theta}\right)^T\right)$

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

MMSE: $Bmse\left(\hat{\theta}_i\right) = \left[M_{\hat{\theta}}\right]_{ii}$

6. Sequential LMMSE Estimation

From $N-1$ random data $X[1], X[2], \cdots, X[N-1]$, we estimate random vector θ via LMMSE estimation, where the estimate and the error matrix are given by

$$\hat{\theta}[N-1],$$

$$M[N-1] = E\left[(\theta - \hat{\theta}[N-1])\big(\theta - \hat{\theta}[N-1]\big)^T\right].$$

Suppose now that *N*th data $X[N]$ is newly observed, where observation matrix is decomposed as

$$H[N] = \begin{bmatrix} H[N-1] \\ h^T[N] \end{bmatrix} = \begin{bmatrix} (N-1) \times p \\ 1 \times p \end{bmatrix}.$$

Then, the estimate and the error matrix can be updated as follows.

$$\hat{\theta}[n] = \hat{\theta}[n-1]K[n](x[n] - h^T[n]\hat{\theta}[n-1]),$$
$$K[n] = \frac{M[n-1]h[n]}{\sigma_n^2 + h^T[n]M[n-1]h[n]},$$
$$M[n] = (I - K[n]h^T[n])M[n-1].$$

Here, we supposed that the errors $\theta - \hat{\theta}[n]$ are independent and their covariance matrix has diagonal elements of $\mathrm{diag}(\sigma_0^2, \sigma_1^2, \cdots, \sigma_n^2)$.