Chapter 13 Mean Square Estimation

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30050, R.O.C.

13-1 Introduction

Concern:

- To estimate the random process $\boldsymbol{s}(t)$ in terms of another related process $\boldsymbol{x}(\xi)$ for $a \le \xi \le b$.

Theorem 13-1 The best linear estimator of $\boldsymbol{s}(t)$ in terms of $\{\boldsymbol{x}(\xi) : a \le \xi \le b\}$, which is of the form

$$\hat{\boldsymbol{s}}(t) = \int_a^b h(\alpha, t)\boldsymbol{x}(\alpha)d\alpha$$

and which minimizes the MS error $P_t = E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))^2]$, satisfies

$$R_{sx}(t, s) = \int_a^b h(\alpha, t)R_{xx}(\alpha, s)d\alpha \text{ for } a \le s \le b.$$

Proof:

$$\begin{aligned}
P_t &= E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))^2] \\
&= E[\boldsymbol{s}^2(t)] + \int_a^b \int_a^b h(\alpha, t)h(\beta, t)E[\boldsymbol{x}(\alpha)\boldsymbol{x}(\beta)]d\alpha d\beta - 2\int_a^b h(\alpha, t)E[\boldsymbol{s}(t)\boldsymbol{x}(\alpha)]d\alpha \\
&= R_{ss}(0) + \int_a^b \int_a^b h(\alpha, t)h(\beta, t)R_{xx}(\alpha, \beta)d\alpha d\beta - 2\int_a^b h(\alpha, t)R_{sx}(t, \alpha)d\alpha.
\end{aligned}$$

13-1 Introduction

Under Riemann integrability assumption,

$$\begin{aligned}
\frac{\partial P_t}{\partial h(s,t)} \\
&= \underbrace{\left(\int_{a,\beta\neq s}^{b} h(\beta,t)R_{xx}(s,\beta)d\beta + \int_{a,\alpha\neq s}^{b} h(\alpha,t)R_{xx}(\alpha,s)d\alpha + 2h(s,t)R_{xx}(s,s)\right)}_{\text{conceptually}} - 2R_{sx}(t,s) \\
&= 2\int_{a}^{b} h(\alpha,t)R_{xx}(s,\alpha)d\alpha - 2R_{sx}(t,s).
\end{aligned}$$

$\square$

Remark

- The minimum MS error is given by:

$$\begin{aligned}
P_t &= R_{ss}(0) + \int_a^b h(\alpha,t)\left(\int_a^b h(\beta,t)R_{xx}(\alpha,\beta)d\beta\right)d\alpha - 2\int_a^b h(\alpha,t)R_{sx}(t,\alpha)d\alpha \\
&= R_{ss}(0) + \int_a^b h(\alpha,t)R_{sx}(t,\alpha)d\alpha - 2\int_a^b h(\alpha,t)R_{sx}(t,\alpha)d\alpha \\
&= R_{ss}(0) - \int_a^b h(\alpha,t)R_{sx}(t,\alpha)d\alpha.
\end{aligned}$$

Orthogonality of Optimal MS Estimation

Theorem 13-1′ Following Theorem 13-1, we also have:

$$E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))\boldsymbol{x}(\xi)] = 0 \text{ for } a \le \xi \le b.$$

Proof:

$$\begin{aligned} E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))\boldsymbol{x}(\xi)] &= E[\boldsymbol{s}(t)\boldsymbol{x}(\xi)] - \int_a^b h(\alpha, t) E[\boldsymbol{x}(\alpha)\boldsymbol{x}(\xi)] d\alpha \\ &= R_{sx}(t, \xi) - \int_a^b h(\alpha, t) R_{xx}(\alpha, \xi) d\alpha \\ &= R_{sx}(t, \xi) - R_{sx}(t, \xi) = 0. \end{aligned}$$

$\square$

Orthogonality principle

- Linear estimator $\hat{\boldsymbol{s}}(t)$ that minimizes $E[\langle \boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t), \boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t)\rangle] = E[\|\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t)\|^2]$ should satisfy $E[\langle \boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t), \hat{\boldsymbol{s}}(t)\rangle] = 0$.
- This may not be true for a non-linear estimator! (Note that the linear combination of $\{\boldsymbol{x}(\xi), a \le \xi \le b\}$ spans a hyperplane in an inner product space.)

Terminologies

Terminologies. With $[a, b]$ = data interval,

- If $t \in [a, b]$, the estimate operation of $\hat{\boldsymbol{s}}(t)$ is called *smoothing*.
- If $t > b$ and $\boldsymbol{x}(\xi) = \boldsymbol{s}(\xi)$, $\hat{\boldsymbol{s}}(t)$ is called *forward predictor*.
- If $t < a$ and $\boldsymbol{x}(\xi) = \boldsymbol{s}(\xi)$, $\hat{\boldsymbol{s}}(t)$ is called *backward predictor*.
- If $t \notin [a, b]$ and $\boldsymbol{x}(\xi) \neq \boldsymbol{s}(\xi)$, the estimate operation of $\hat{\boldsymbol{s}}(t)$ is called *filtering and prediction*.

Forward Prediction Under Stationarity

Theorem 13-1 The best linear estimator of $\boldsymbol{s}(t)$ in terms of $\{\boldsymbol{s}(\xi) : a \leq \xi \leq b\}$, which is of the form

$$\hat{\boldsymbol{s}}(t) = \int_a^b h(\alpha, t)\boldsymbol{s}(\alpha)d\alpha$$

and which minimizes the MS error $P = E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))^2]$, satisfies

$$R_{ss}(t, s) = \int_a^b h(\alpha, t)R_{ss}(\alpha, s)d\alpha \text{ for } a \leq s \leq b.$$

In addition,

$$P_t \;=\; R_{ss}(0) - \int_a^b h(\alpha, t)R_{ss}(t, \alpha)d\alpha.$$

- If $\boldsymbol{s}(t)$ is stationary, $a = b$ (i.e., $\xi = a = b$) and $t = a + \lambda$, we have $s = a$ and

$$R_{ss}(\lambda) = h(a, a + \lambda)R_{ss}(0)$$

$$\Rightarrow h(a, a+\lambda) = \frac{R_{ss}(\lambda)}{R_{ss}(0)} \text{ and } \hat{\boldsymbol{s}}(a+\lambda) = \frac{R_{ss}(\lambda)}{R_{ss}(0)}\boldsymbol{s}(a) \text{ and } \boxed{P_{a+\lambda} = R_{ss}(0) - \frac{R_{ss}^2(\lambda)}{R_{ss}(0)}}.$$

Forward Prediction Under Stationarity

Theorem 13-1′

$$E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))\boldsymbol{s}(\xi)] = 0 \text{ for } a \le \xi \le b.$$

If $E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))\boldsymbol{s}(\xi)] = 0$ for $\xi < a$,
then $\hat{\boldsymbol{s}}(t)$ is the best linear predictor of $\boldsymbol{s}(t)$ in terms of $\{\boldsymbol{s}(\xi), \xi \le b\}$,
although it only uses the information of $\{\boldsymbol{s}(\xi), a \le \xi \le b\}$.

If $\dfrac{R_{ss}(v)}{R_{ss}(u)} = \dfrac{R_{ss}(0)}{R_{ss}(u-v)}$ for any $u \ge v$, then for $\xi < a$,

$$\begin{aligned} E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))\boldsymbol{s}(\xi)] &= R_{ss}(t-\xi) - \int_a^b h(\alpha,t)R_{ss}(\alpha-\xi)d\alpha \\ &= R_{ss}(t-\xi)\left(1 - \int_a^b h(\alpha,t)\frac{R_{ss}(\alpha-\xi)}{R_{ss}(t-\xi)}d\alpha\right) \\ &= R_{ss}(t-\xi)\left(1 - \int_a^b h(\alpha,t)\frac{R_{ss}(\alpha)}{R_{ss}(t)}d\alpha\right) \\ &= 0. \end{aligned}$$

If, in addition, $a = b$ in the above case, $\boldsymbol{s}(t)$ is named the *wide-sense Markov of order 1* (i.e., best linear prediction based on **one** point is the best prediction based on the entire past.)

Theorem 13-1 Revisited

Theorem 13-1 The best linear estimator of $\boldsymbol{s}(t)$ in terms of $\{\boldsymbol{x}_i(\xi) : a \leq \xi \leq b\}_{i=1}^k$, which is of the form

$$\hat{\boldsymbol{s}}(t) = \sum_{i=1}^{k} \int_a^b h_i(\alpha, t)\boldsymbol{x}_i(\alpha)d\alpha$$

and which minimizes the MS error $P_t = E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))^2]$, satisfies

$$R_{sx_i}(t, s) = \sum_{\ell=1}^{k} \int_a^b h_i(\alpha, t)R_{x_\ell x_i}(\alpha, s)d\alpha \text{ for } a \leq s \leq b \text{ and } 1 \leq i \leq k.$$

Proof: A different proof is used here. The optimal estimator should satisfy:

$$E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))\boldsymbol{x}_i(\xi)] = 0 \text{ for } a \leq \xi \leq b \text{ and } 1 \leq i \leq k.$$

Hence,

$$R_{sx_i}(t, \xi) = \sum_{\ell=1}^{n} \int_a^b h_\ell(\alpha, t)R_{x_\ell x_i}(\alpha, \xi)d\alpha.$$

□

Theorem 13-1 Revisited

Example. If $\boldsymbol{s}(t)$ is stationary, $a = b$, $t = a + \lambda$, $\boldsymbol{x}_1(t) = \boldsymbol{s}(t)$ and $\boldsymbol{x}_2(t) = \boldsymbol{s}'(t)$, then

$$\begin{cases} R_{ss}(a+\lambda, a) = h_1(a, a+\lambda)R_{ss}(a,a) + h_2(a, a+\lambda)R_{s's}(a,a) \\ \qquad\qquad\qquad\;\; = h_1(a, a+\lambda)R_{ss}(a,a) + h_2(a, a+\lambda)R_{ss'}(a,a) \\ R_{ss'}(a+\lambda, a) = h_1(a, a+\lambda)R_{ss'}(a,a) + h_2(a, a+\lambda)R_{s's'}(a,a) \end{cases}$$

which in turns implies:

$$\begin{cases} R_{ss}(\lambda) &= h_1(a, a+\lambda)R_{ss}(0) - h_2(a, a+\lambda)R'_{ss}(0) \\ -R'_{ss}(\lambda) &= -h_1(a, a+\lambda)R'_{ss}(0) - h_2(a, a+\lambda)R''_{ss}(0) \end{cases}$$

$$R_{ss'}(t_1, t_2) = \frac{\partial R_{ss}(t_1 - t_2)}{\partial t_2} = -R'_{ss}(t_1 - t_2)$$

and

$$R_{s's'}(t_1, t_2) = \frac{\partial R_{ss'}(t_1, t_2)}{\partial t_1} = -R''_{ss}(t_1 - t_2).$$

Theorem 13-1 Revisited

$$\Rightarrow \begin{cases} h_1(a, a+\lambda) &= h_1(\lambda) = \dfrac{R_{ss}(\lambda)R''_{ss}(0) + R'_{ss}(\lambda)R'_{ss}(0)}{R_{ss}(0)R''_{ss}(0) + [R'_{ss}(0)]^2} = \dfrac{R_{ss}(\lambda)}{R_{ss}(0)} \\ h_2(a, a+\lambda) &= h_2(\lambda) = \dfrac{R'_{ss}(\lambda)R_{ss}(0) - R'_{ss}(0)R_{ss}(\lambda)}{R_{ss}(0)R''_{ss}(0) + [R'_{ss}(0)]^2} = \dfrac{R'_{ss}(\lambda)}{R''_{ss}(0)} \end{cases}$$

where it is reasonable to assume that $R'_{ss}(0) = 0$.

$$\Rightarrow \hat{\boldsymbol{s}}(a+\lambda) = \frac{R_{ss}(\lambda)}{R_{ss}(0)}\boldsymbol{s}(a) + \frac{R'_{ss}(\lambda)}{R''_{ss}(0)}\boldsymbol{s}'(a).$$

Theorem 13-1 Revisited

$$
\begin{aligned}
\Rightarrow P_t &= E[(\boldsymbol{s}(a+\lambda) - \hat{\boldsymbol{s}}(a+\lambda))^2] \\
&= E[\boldsymbol{s}^2(a+\lambda)] + \left(\frac{R_{ss}(\lambda)}{R_{ss}(0)}\right)^2 E[\boldsymbol{s}^2(a)] + \left(\frac{R'_{ss}(\lambda)}{R''_{ss}(0)}\right)^2 E[(\boldsymbol{s}'(a))^2] \\
&\quad -2\left(\frac{R_{ss}(\lambda)}{R_{ss}(0)}\right) E[\boldsymbol{s}(a+\lambda)\boldsymbol{s}(a)] - 2\left(\frac{R'_{ss}(\lambda)}{R''_{ss}(0)}\right) E[\boldsymbol{s}(a+\lambda)\boldsymbol{s}'(a)] \\
&\quad +2\left(\frac{R_{ss}(\lambda)}{R_{ss}(0)}\right)\left(\frac{R'_{ss}(\lambda)}{R''_{ss}(0)}\right) E[\boldsymbol{s}(a)\boldsymbol{s}'(a)] \\
&= R_{ss}(0) + \left(\frac{R_{ss}(\lambda)}{R_{ss}(0)}\right)^2 R_{ss}(0) - \left(\frac{R'_{ss}(\lambda)}{R''_{ss}(0)}\right)^2 R''_{ss}(0) \\
&\quad -2\left(\frac{R_{ss}(\lambda)}{R_{ss}(0)}\right) R_{ss}(\lambda) + 2\left(\frac{R'_{ss}(\lambda)}{R''_{ss}(0)}\right) R'_{ss}(\lambda) \cancel{- 2\left(\frac{R_{ss}(\lambda)}{R_{ss}(0)}\right)\left(\frac{R'_{ss}(\lambda)}{R''_{ss}(0)}\right) R'_{ss}(0)} \\
&= R_{ss}(0) - \frac{R^2_{ss}(\lambda)}{R_{ss}(0)} + \frac{(R'_{ss}(\lambda))^2}{R''_{ss}(0)}.
\end{aligned}
$$

Theorem 13-1 The best linear estimator of $\boldsymbol{s}(t)$ in terms of $\{\boldsymbol{x}(\xi) : a \le \xi \le b\}$, which is of the form

$$\hat{\boldsymbol{s}}(t) = \int_a^b h(\alpha, t)\boldsymbol{x}(\alpha)d\alpha$$

and which minimizes the MS error $P_t = E[(\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t))^2]$, satisfies

$$R_{sx}(t, s) = \int_a^b h(\alpha, t)R_{xx}(\alpha, s)d\alpha \text{ for } a \le s \le b.$$

- If $\boldsymbol{s}(t)$ and $\boldsymbol{x}(t)$ are joint stationary and $t = a = b$, we have $s = t$ and

$$R_{sx}(0) = h(t, t)R_{ss}(0)$$

$$\Rightarrow h(t, t) = h = \frac{R_{sx}(0)}{R_{ss}(0)} \text{ and } \hat{\boldsymbol{s}}(t) = \frac{R_{sx}(0)}{R_{ss}(0)}\boldsymbol{x}(t) \text{ and } P_t = R_{ss}(0) - \frac{R_{sx}^2(0)}{R_{ss}(0)}.$$

Interpolation

Concern

- To estimate, in the MS sense, $\boldsymbol{s}(t+\lambda)$ in terms of $\{\boldsymbol{s}(t+kT)\}_{k=-N}^{N}$ with the form

$$\hat{\boldsymbol{s}}(t+\lambda) = \sum_{k=-N}^{N} a_k \boldsymbol{s}(t+kT).$$

By using the **orthogonality principle**:

$$E\left\{\left[\boldsymbol{s}(t+\lambda) - \sum_{k=-N}^{N} a_k \boldsymbol{s}(t+kT)\right] \boldsymbol{s}(t+nT)\right\} = 0,$$

we obtain

$$\sum_{k=-N}^{N} a_k R_{ss}(kT - nT) = R_{ss}(\lambda - nT) \text{ for } -N \le n \le N.$$

In addition,

$$P_t = E\left\{\left[\boldsymbol{s}(t+\lambda) - \sum_{k=-N}^{N} a_k \boldsymbol{s}(t+kT)\right] \boldsymbol{s}(t+\lambda)\right\} = R_{ss}(0) - \sum_{k=-N}^{N} a_k R_{ss}(\lambda - kT).$$

Theorem 10-9 (Stochastic sampling theorem) If $\boldsymbol{s}(t)$ is BL with bandwidth σ, then

$$\boldsymbol{s}(t+\lambda) = \sum_{k=-\infty}^{\infty} \frac{\sin[\sigma(\lambda - kT)]}{\sigma(\lambda - kT)} \boldsymbol{s}(t+kT) \quad \text{(in the MS sense)},$$

where $T = \pi/\sigma$.

Example. Let $S_{ss}(\omega) = 1$ for $|\omega| < \sigma$ and zero, otherwise. Then,

$$R_{ss}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{ss}(\omega)e^{j\omega\tau}d\omega = \frac{1}{2\pi}\int_{-\sigma}^{\sigma} e^{j\omega\tau}d\omega = \frac{\sin(\tau\sigma)}{\pi\tau}.$$

We thus derive for $T\sigma = \pi$ that

$$\sum_{k=-N}^{N} a_k \frac{\sin((kT - nT)\sigma)}{\pi(kT - nT)} = \frac{\sin((\lambda - nT)\sigma)}{\pi(\lambda - nT)} \text{ for } -N \le n \le N$$

$$\Leftrightarrow \sum_{k=-N}^{N} a_k \frac{\sin(\pi(k-n))}{\pi(k-n)} = \frac{T\sin(\sigma(\lambda - nT))}{\pi(\lambda - nT)} \text{ for } -N \le n \le N$$

$$\Leftrightarrow a_n = \frac{\sin(\sigma(\lambda - nT))}{\sigma(\lambda - nT)} \text{ for } -N \le n \le N$$

□

Example for Smoothing

Concern

- To estimate (real WSS) $\boldsymbol{s}(t)$ in terms of $\{\boldsymbol{x}(\xi), -\infty < \xi < \infty\}$ with WSS $\boldsymbol{x}(t) = \boldsymbol{s}(t) + \boldsymbol{v}(t)$.

Using the orthogonality principle:

$$E\left\{\left[\boldsymbol{s}(t) - \int_{-\infty}^{\infty} h(\alpha, t)\boldsymbol{x}(t-\alpha)d\alpha\right]\boldsymbol{x}(t-\xi)\right\} = 0 \text{ for } -\infty < \xi < \infty,$$

or equivalently,

$$R_{sx}(\xi) = \int_{-\infty}^{\infty} h(\alpha, t) R_{xx}(\xi - \alpha) d\alpha$$

This gives that

$$H(\omega; t) = H(\omega) = \frac{S_{sx}(\omega)}{S_{xx}(\omega)}$$

which is named the *noncausal Wiener filter*.

Assume $\boldsymbol{v}(t)$ is zero-mean and is independent of $\boldsymbol{s}(t)$. Then,

$$R_{sx}(\tau) = E[\boldsymbol{s}(t+\tau)(\boldsymbol{s}(t) + \boldsymbol{v}(t))] = R_{ss}(\tau)$$

$$R_{xx}(\tau) = E[(\boldsymbol{s}(t+\tau) + \boldsymbol{v}(t+\tau))(\boldsymbol{s}(t) + \boldsymbol{v}(t))] = R_{ss}(\tau) + R_{vv}(\tau).$$

Example for Smoothing

In such case, the best filter is:

$$H(\omega) = \frac{S_{ss}(\omega)}{S_{ss}(\omega) + S_{vv}(\omega)},$$

which is real and symmetric (because $R_{ss}(\tau)$ is real and symmetric, and $R_{vv}(\tau)$ is real and symmetric). And

$$\begin{aligned}
P_t &= E\left\{\left[\boldsymbol{s}(t) - \int_{-\infty}^{\infty} h(\alpha, t)\boldsymbol{x}(t-\alpha)d\alpha\right]\boldsymbol{s}(t)\right\} \\
&= R_{ss}(0) - \int_{-\infty}^{\infty} h(\alpha)R_{ss}(\alpha)d\alpha \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{ss}(\omega)d\omega - \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)e^{j\omega\alpha}d\omega\right)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{ss}(\omega')e^{j\omega'\alpha}d\omega'\right)d\alpha \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{ss}(\omega)d\omega - \frac{1}{2\pi}\int_{-\infty}^{\infty} H(-\omega')S_{ss}(\omega')d\omega' \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_{ss}(\omega)S_{vv}(\omega)}{S_{ss}(\omega) + S_{vv}(\omega)}d\omega.
\end{aligned}$$

Conclusion: As long as there is no overlap in $S_{ss}(\omega)$ and $S_{vv}(\omega)$, P_t is zero!

The end of Section 13-1 Introduction

13-2 Prediction

Prediction of $\boldsymbol{s}[n]$ in terms of:

- Entire past: $\{\boldsymbol{s}[n-k]\}_{k\geq 1}$
- r-step away past: $\{\boldsymbol{s}[n-k]\}_{k\geq r}$
- Finite past: $\{\boldsymbol{s}[n-k]\}_{r\leq k\leq N}$
- $\ldots$

Assume throughout Section 13-2 that $\boldsymbol{s}[n]$ is stationary.

Prediction Based on Entire Past

- $\hat{\boldsymbol{s}}[n] = \sum_{k=1}^{\infty} h[k,n]\boldsymbol{s}[n-k]$.
- **Orthogonality principle**: For all $m \geq 1$,

$$0 = E[(\boldsymbol{s}[n] - \hat{\boldsymbol{s}}[n])\boldsymbol{s}[n-m]] \;=\; R_{ss}[m] - \sum_{k=1}^{\infty} h[k,n]R_{ss}[m-k]$$

  Hence, again under stationarity assumption, $h[k,n] = h[k]$ is invariant in n.

- Therefore, the best prediction filter satisfies:

$$R_{ss}[m] = \sum_{k=1}^{\infty} h[k]R_{ss}[m-k] \text{ for } m \geq 1.$$

  This is called the *Wiener-Höpf equation* (in digital form).

Property of Error Process

- $\hat{\boldsymbol{s}}[n]$ is the response of the *predictor filter*

$$H[z] = h[1]z^{-1} + h[2]z^{-2} + \cdots + h[k]z^{-k} + \cdots$$

due to the input $\boldsymbol{s}[n]$.

- Hence, the error process defined as $\boldsymbol{e}[n] = \boldsymbol{s}[n] - \hat{\boldsymbol{s}}[n] = \boldsymbol{s}[n] - \sum_{k=1}^{\infty} h[k]\boldsymbol{s}[n-k]$ is the response of the filter

$$\mathtt{E}[z] = 1 - H[z]$$

due to the input $\boldsymbol{s}[n]$.

- *Claim:* The error process $\boldsymbol{e}[n]$ is white.

  Proof:

  - $\boldsymbol{e}[n]$ is orthogonal to $\boldsymbol{s}[n-m]$ for all $m \geq 1$.
  - $\boldsymbol{e}[n-m]$ is a linear combination of $\boldsymbol{s}[n-m-\ell]$ for all $\ell \geq 0$.
  - Hence, $\boldsymbol{e}[n]$ is orthogonal to $\boldsymbol{e}[n-m]$ for all $m \geq 1$, and $R_{ee}[m] = P\delta[m]$, where $P = E[\boldsymbol{e}^2[n]] = E[\boldsymbol{e}[n]\boldsymbol{s}[n]]$ is the minimum MS power. $\square$

Property of Error Process

Theorem 13-2 All zeros of $\mathtt{E}[z]$ satisfy $|z| \le 1$.

Proof: If there exists a z_i such that $\mathtt{E}[z_i] = 0$ and $|z_i| > 1$, then form a new error filter as:

$$\mathtt{E}_0[z] = \mathtt{E}[z]\frac{1 - z^{-1}/z_i^*}{1 - z_i z^{-1}}.$$

Then, by letting $z_i = |z_i|e^{j\theta_i}$ and $\omega' = \omega - \theta_i$, we have:

$$\begin{aligned}
|\mathtt{E}_0[e^{j\omega}]|^2 &= |\mathtt{E}[e^{j\omega}]|^2 \left|\frac{e^{j\omega} - e^{j\theta_i}/|z_i|}{e^{j\omega} - |z_i|e^{j\theta_i}}\right|^2 = |\mathtt{E}[e^{j\omega}]|^2 \left|\frac{e^{j\omega'} - 1/|z_i|}{e^{j\omega'} - |z_i|}\right|^2 \\
&= |\mathtt{E}[e^{j\omega}]|^2 \frac{1 - (2/|z_i|)\cos(\omega') + 1/|z_i|^2}{1 - 2|z_i|\cos(\omega') + |z_i|^2} \\
&= |\mathtt{E}[e^{j\omega}]|^2 \frac{1}{|z_i|^2} < |\mathtt{E}[e^{j\omega}]|^2.
\end{aligned}$$

However,

$$P = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\mathtt{E}[e^{j\omega}]|^2 S_{ss}[\omega] d\omega$$

is the minimum MS error that can be achieved, and $\mathtt{E}_0[z]$ improves the minimum MS error. Thus, the desired contradiction is obtained. $\square$

Solving Wiener-Höpf Equation Under Regularity

The z-transform technique cannot be applied to solve the Wiener-Höpf Equation. If

$$R_{sx}[m] = \sum_{k=1}^{\infty} h[k] R_{xx}[m-k] \text{ for } \textbf{all integer } m,$$

then $H[z] = S_{sx}[z]/S_{xx}[z]$, where

$$S_{xx}[z] = \sum_{k=-\infty}^{\infty} R_{xx}[k] z^{-k} \quad \text{and} \quad S_{sx}[z] = \sum_{k=-\infty}^{\infty} R_{sx}[k] z^{-k}.$$

However,

$$R_{sx}[m] = \sum_{k=1}^{\infty} h[k] R_{xx}[m-k] \textbf{ only} \text{ for } m \geq 1.$$

Solving Wiener-Höpf Equation Under Regularity

Solving Wiener-Höpf equation under the assumption that $\boldsymbol{s}[n]$ is stationary and regular

- A regular process can be represented as the response of a causal finite-energy system due to a unit-power white-noise process $\boldsymbol{i}[n]$. So,

$$\boldsymbol{s}[n] = \sum_{k=0}^{\infty} \mathtt{l}[k]\boldsymbol{i}[n-k].$$

- Then, $\hat{\boldsymbol{s}}[n] = \sum_{k=1}^{\infty} h[k]\boldsymbol{s}[n-k]$ can be written as

$$\hat{\boldsymbol{s}}[n] = \sum_{k=1}^{\infty} g[k]\boldsymbol{i}[n-k],$$

for some $\{g[k]\}_{k=1}^{\infty}$ that minimizes the MS error.

- The orthogonality principle then gives that for all $m \geq 1$,

$$\begin{aligned} 0 &= E\left\{\left(\boldsymbol{s}[n] - \sum_{k=1}^{\infty} g[k]\boldsymbol{i}[n-k]\right)\boldsymbol{i}[n-m]\right\} \\ &= R_{si}[m] - \sum_{k=1}^{\infty} g[k]R_{ii}[m-k] = R_{si}[m] - g[m], \end{aligned}$$

which implies $g[m] = R_{si}[m]$.

Solving Wiener-Höpf Equation Under Regularity

- By regularity,

$$R_{si}[m] = E[\boldsymbol{s}[n]\boldsymbol{i}[n-m]] \ = \sum_{k=0}^{\infty} \mathtt{l}[k]E\{\boldsymbol{i}[n-k]\boldsymbol{i}[n-m]\} = \mathtt{l}[m].$$

This concludes to the first important result:

$$\hat{\boldsymbol{s}}[n] = \sum_{k=1}^{\infty} \mathtt{l}[k]\boldsymbol{i}[n-k]$$

is the best linear predictor for a regular and stationary process

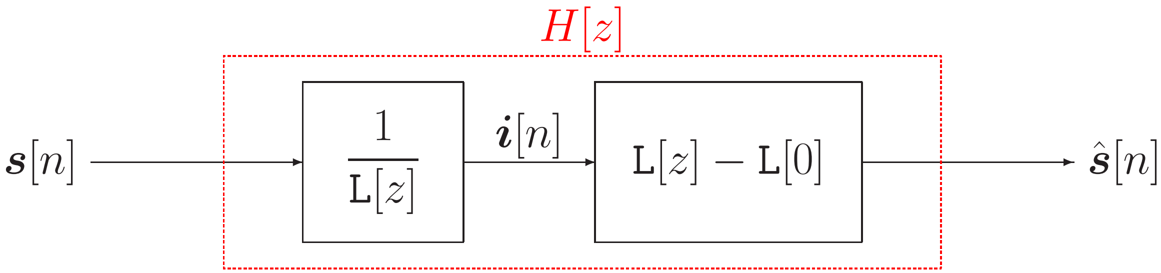
$$\boldsymbol{s}[n] = \sum_{k=0}^{\infty} \mathtt{l}[k]\boldsymbol{i}[n-k] \quad \text{and} \quad P = \mathtt{l}^2[0].$$

- By noting that $\boldsymbol{i}[n]$ is the response of system $1/\mathtt{L}[z]$ due to input $\boldsymbol{s}[n]$, and $\hat{\boldsymbol{s}}[n]$ is the response of system $\mathtt{L}[z] - \mathtt{l}[0]$ due to input $\boldsymbol{i}[n]$, we obtain:

$$H[z] = \frac{1}{\mathtt{L}[z]}(\mathtt{L}[z] - \mathtt{l}[0]) = 1 - \frac{\mathtt{l}[0]}{\mathtt{L}[z]} = 1 - \frac{\lim_{z\uparrow\infty} \mathtt{L}[z]}{\mathtt{L}[z]}.$$

□

Solving Wiener-Höpf Equation Under Regularity



If $S[\omega]$ is a rational spectrum, then $\mathtt{L}[z]$ can be obtained as follows.

- $S[z] = A((z + z^{-1})/2)/B((z + z^{-1})/2)$.
- Then the roots of $S[z]$ are symmetric with respect to the **unit circle**. So, we can separate them into two groups: **Inside** group that consists of all roots with $|z| < 1$, and the **outside** group that consists of all roots with $|z| > 1$.
- Form $\mathtt{L}[z]$ by the ratio of two polynomials with the inside roots of $S[z]$.

Solving Wiener-Höpf Equation Under Regularity

Example 13-3 (Slide 11-16) $S_{ss}[\omega] = \dfrac{5 - 4\cos(\omega)}{10 - 6\cos(\omega)}$

Then, $\mathtt{L}[z] = \dfrac{2z-1}{3z-1}$.

In this case,

$$H[z] = 1 - \frac{\lim_{z\uparrow\infty}\mathtt{L}[z]}{\mathtt{L}[z]} = 1 - \frac{2/3}{\dfrac{2z-1}{3z-1}} = 1 - \frac{2z-2/3}{2z-1} = \frac{-(1/6)z^{-1}}{1-(1/2)z^{-1}}.$$

Consequently,

$$\hat{\boldsymbol{s}}[n] - \frac{1}{2}\hat{\boldsymbol{s}}[n-1] = -\frac{1}{6}\boldsymbol{s}[n-1]$$

or equivalently,

$$\hat{\boldsymbol{s}}[n] = -\frac{1}{6}\boldsymbol{s}[n-1] + \frac{1}{2}\hat{\boldsymbol{s}}[n-1].$$

Kolmogorov-Szego MS Error Formula

Appendix 12A A minimum-phase system $\mathtt{L}[z]$ satisfies

$$\log \mathtt{l}^2[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log |\mathtt{L}[e^{j\omega}]|^2 d\omega.$$

Kolmogorov and Szego noted from the above result and $S_{ss}(\omega) = |\mathtt{L}[e^{j\omega}]|^2$ that

$$P = \mathtt{l}^2[0] = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log |\mathtt{L}[e^{j\omega}]|^2 d\omega\right\} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log S_{ss}[\omega] d\omega\right\}.$$

This is named the *Kolmogorov-Szego MS Error Formula*.

Wide-Sense Markov of Order N

- If $\boldsymbol{s}[n]$ is an autoregressive (AR) process, then (Slide 11-46)

$$\mathtt{L}[z] = \frac{b_0}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}}.$$

- Then,

$$H[z] = 1 - \frac{\lim_{z\uparrow\infty} \mathtt{L}[z]}{\mathtt{L}[z]} = 1 - \frac{b_0}{\dfrac{b_0}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}}} = a_1 z^{-1} + \cdots + a_n z^{-n},$$

which implies

$$\hat{\boldsymbol{s}}[n] = -a_1 \boldsymbol{s}[n-1] - \cdots - a_N \boldsymbol{s}[n-N].$$

- Then, $\boldsymbol{s}[n]$ is called the *wide-sense Markov of order N*.
  - Best linear prediction based on the past N points is the best prediction based on the entire past.

r-Step Predictor

Concern

- To find the best linear estimator of $\boldsymbol{s}[n]$, in the MS sense, in terms of the r-step-away entire past, i.e., $\{\boldsymbol{s}[n-k]\}_{k\geq r}$.

$$\hat{\boldsymbol{s}}[n] = \sum_{k=r}^{\infty} \mathtt{l}[k]\boldsymbol{i}[n-k]$$

is the best linear r-step predictor for a regular and stationary process

$$\boldsymbol{s}[n] = \sum_{k=0}^{\infty} \mathtt{l}[k]\boldsymbol{i}[n-k] \quad \text{and} \quad P = \sum_{k=0}^{r-1} \mathtt{l}^2[k].$$

Proof:

– A regular process can be represented as the response of a causal finite-energy system due to a unit-power white-noise process $\boldsymbol{i}[n]$. So,

$$\boldsymbol{s}[n] = \sum_{k=0}^{\infty} \mathtt{l}[k]\boldsymbol{i}[n-k].$$

r-Step Predictor

– Then, $\hat{\boldsymbol{s}}[n] = \sum_{k=r}^{\infty} h[k]\boldsymbol{s}[n-k]$ can be written as

$$\hat{\boldsymbol{s}}[n] = \sum_{k=r}^{\infty} g[k]\boldsymbol{i}[n-k],$$

for some $\{g[k]\}_{k=1}^{\infty}$ that minimizes the MS error.

– The orthogonality principle then gives that for all $m \geq r$,

$$\begin{aligned} 0 &= E\left\{\left(\boldsymbol{s}[n] - \sum_{k=r}^{\infty} g[k]\boldsymbol{i}[n-k]\right)\boldsymbol{i}[n-m]\right\} \\ &= R_{si}[m] - \sum_{k=r}^{\infty} g[k]R_{ii}[m-k] = R_{si}[m] - g[m], \end{aligned}$$

which implies $g[m] = R_{si}[m]$ for $m \geq r$.

– By regularity,

$$R_{si}[m] = E[\boldsymbol{s}[n]\boldsymbol{i}[n-m]] \;=\; \sum_{k=0}^{\infty} \mathtt{l}[k]E\{\boldsymbol{i}[n-k]\boldsymbol{i}[n-m]\} = \mathtt{l}[m].$$

$\square$

r-Step Predictor

- In addition, it can be derived that

$$H_r[z] = 1 - \frac{1}{\mathtt{L}[z]} \sum_{k=0}^{r-1} \mathtt{l}[k] z^{-k}.$$

Example 13-4 Suppose $R_{ss}[m] = a^{|m|}$ for $0 < a < 1$. Then,

$$\begin{aligned}
S_{ss}[z] &= \sum_{m=-\infty}^{\infty} R_{ss}[m] z^{-m} = \sum_{m=0}^{\infty} a^m (z^{-m} + z^m) - 1 \\
&= \frac{1}{1 - az^{-1}} + \frac{1}{1 - az} - 1 = \frac{1 - a^2}{(1 - az^{-1})(1 - az)}. \\
\Rightarrow \mathtt{L}[z] &= \frac{b}{1 - az^{-1}} = b(1 + az^{-1} + a^2 z^{-2} + \cdots), \quad \text{where } b = \sqrt{1 - a^2}. \\
\Rightarrow H_r[z] &= 1 - \frac{1}{\mathtt{L}[z]} \sum_{k=0}^{r-1} \mathtt{l}[k] z^{-k} \\
&= 1 - \frac{(1 - az^{-1})}{b} \sum_{k=0}^{r-1} b a^k z^{-k} \\
&= 1 - (1 - az^{-1})(1 + az^{-1} + a^2 z^{-2} + \cdots + a^{r-1} z^{-(r-1)}) = a^r z^{-r}
\end{aligned}$$

Analog Wiener-Höph Equation

Concern:

- To linearly estimate the random process $\boldsymbol{s}(t+\lambda)$ in terms of its entire past $\{\boldsymbol{s}(t-\tau), \tau \geq 0\}$ in the MS sense.

Analog Wiener-Höph equation

- *Orthogonality principle*:

$$
\begin{aligned}
& E\left\{\left[\boldsymbol{s}(t+\lambda)-\int_0^{\infty} h(\alpha)\boldsymbol{s}(t-\alpha)d\alpha\right]\boldsymbol{s}(t-\tau)\right\}=0 \text{ for all } \tau \geq 0 \\
\Leftrightarrow \quad & R_{ss}(\lambda+\tau)=\int_0^{\infty} h(\alpha)R_{ss}(\tau-\alpha)d\alpha \text{ for all } \tau \geq 0
\end{aligned}
$$

The solution of (analog) Wiener-Höph equation is named the *causal Wiener filter*.

Solving Wiener-Höpf Equation Under Regularity

Solving Wiener-Höpf equation under the assumption that $\boldsymbol{s}(t)$ is stationary and regular.

- A regular process can be represented as the response of a causal finite-energy system due to a unit-power white-noise process $\boldsymbol{i}(t)$. So,

$$\boldsymbol{s}(t+\lambda) = \int_0^\infty \mathtt{l}(\alpha)\boldsymbol{i}(t+\lambda-\alpha)d\alpha.$$

- Then, $\hat{\boldsymbol{s}}(t+\lambda) = \int_0^\infty h(\alpha)\boldsymbol{s}(t-\alpha)d\alpha$ can be written as

$$\hat{\boldsymbol{s}}(t+\lambda) = \int_0^\infty g(\alpha)\boldsymbol{i}(t-\alpha)d\alpha,$$

  for some $\{g(t)\}_{t\geq 0}$ that minimizes the MS error.

- The orthogonality principle then gives that for all $\tau \geq 0$,

$$\begin{aligned} 0 &= E\left\{\left(\boldsymbol{s}(t+\lambda) - \int_0^\infty g(\alpha)\boldsymbol{i}(t-\alpha)d\alpha\right)\boldsymbol{i}(t-\tau)\right\} \\ &= R_{si}(\lambda+\tau) - \int_0^\infty g(\alpha)R_{ii}(\tau-\alpha)d\alpha, \end{aligned}$$

  which implies $g(\tau) = R_{si}(\lambda+\tau)$.

Solving Wiener-Höpf Equation Under Regularity

- By regularity,

$$R_{si}(\lambda+\tau) = E[\boldsymbol{s}(t)\boldsymbol{i}(t-\lambda-\tau)] \;=\; \int_0^\infty \mathtt{l}(\alpha)E\{\boldsymbol{i}(t-\alpha)\boldsymbol{i}(t-\lambda-\tau)\}d\alpha = \mathtt{l}(\lambda+\tau).$$

This concludes to the first important result:

$$\hat{\boldsymbol{s}}(t+\lambda) = \int_0^\infty \mathtt{l}(\lambda+\alpha)\boldsymbol{i}(t-\alpha)d\alpha = \int_\lambda^\infty \mathtt{l}(\alpha)\boldsymbol{i}(t+\lambda-\alpha)d\alpha$$

is the best linear predictor for a regular and stationary process

$$\boldsymbol{s}(t+\lambda) = \int_0^\infty \mathtt{l}(\alpha)\boldsymbol{i}(t+\lambda-\alpha)d\alpha \text{ and } P = \int_0^\lambda \mathtt{l}^2(\alpha)d\alpha.$$

- By noting that $\boldsymbol{i}(t)$ is the response of system $1/\mathtt{L}(s)$ due to input $\boldsymbol{s}(t)$, and $\hat{\boldsymbol{s}}(t+\lambda)$ is the response of system $\mathtt{l}(\tau+\lambda)\mathbf{1}\{\tau \geq 0\}$ due to input $\boldsymbol{i}(t)$, we obtain:

$$H(\omega) = \frac{1}{\mathtt{L}(\omega)}\int_0^\infty \mathtt{l}(\tau+\lambda)e^{-j\omega\tau}d\tau.$$

□

Solving Wiener-Höpf Equation Under Regularity

Example 13-5 $R_{ss}(\tau) = 2\alpha e^{-\alpha|\tau|}$ with $0 < \alpha < 1$.

$$\Rightarrow \quad S_{ss}(\omega) = \int_{-\infty}^{\infty} 2\alpha e^{-\alpha|\tau|} e^{-j\omega\tau} d\tau = \frac{4\alpha^2}{\alpha^2 + \omega^2}$$

$$\Rightarrow \quad S_{ss}(s) = \left.\frac{4\alpha^2}{\alpha^2 + \omega^2}\right|_{\omega=-js} = \frac{4\alpha^2}{\alpha^2 - s^2} = \frac{2\alpha}{(\alpha + s)} \frac{2\alpha}{(\alpha - s)} = \mathtt{L}(s)\mathtt{L}(-s)$$

$$\Rightarrow \quad \mathtt{L}(s) = \frac{2\alpha}{\alpha + s}$$

$$\Rightarrow \quad \mathtt{L}(\omega) = \frac{2\alpha}{\alpha + j\omega}$$

$$\Rightarrow \quad \mathtt{l}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2\alpha}{\alpha + j\omega} e^{-j\omega\tau} d\omega = 2\alpha e^{-\alpha\tau} \mathbf{1}\{\tau \ge 0\}$$

$$\Rightarrow \quad H(\omega) = \frac{1}{\mathtt{L}(\omega)} \int_{0}^{\infty} \mathtt{l}(\tau + \lambda) e^{-j\omega\tau} d\tau = e^{-\alpha\lambda}$$

$$\Rightarrow \quad h(\tau) = e^{-\alpha\lambda}\delta(\tau)$$

$$\Rightarrow \quad \hat{\boldsymbol{s}}(t + \lambda) = \int_{0}^{\infty} h(\tau)\boldsymbol{s}(t - \tau) d\tau = e^{-\alpha\lambda}\boldsymbol{s}(t).$$

Solving Wiener-Höpf Equation Under Regularity

An alternative way to express Wiener-Höpf equation

- The Wiener-Höpf equation only depends on $R_{ss}(\tau)$; hence, any process with the same autocorrelation function should result in the same predictor.
- (Slide 9-100) Define a process $\boldsymbol{z}(t) = e^{j\boldsymbol{\omega}t}$, where $\boldsymbol{\omega}$ has density $A(\omega)$. Then,

$$R_{zz}(\tau) = E\left\{e^{j\boldsymbol{\omega}(t+\tau)}e^{-j\boldsymbol{\omega}t}\right\} = \int_{-\infty}^{\infty} A(\omega)e^{j\omega\tau}d\tau.$$

  So, $\boldsymbol{z}(t)$ is a process with power spectrum $2\pi A(\omega)$.

- The best-MS linear predictor for $\boldsymbol{z}(t+\lambda)$ in terms of $\{\boldsymbol{z}(t-\tau)\}_{\tau\geq 0}$ is

$$\hat{\boldsymbol{z}}(t+\lambda) = \int_0^{\infty} h(\alpha)\boldsymbol{z}(t-\alpha)d\alpha = \int_0^{\infty} h(\alpha)e^{j\boldsymbol{\omega}(t-\alpha)}d\alpha = e^{j\boldsymbol{\omega}t}H(\boldsymbol{\omega}),$$

  and should satisfy

$$\begin{aligned} & E\left\{(\boldsymbol{z}(t+\lambda) - \hat{\boldsymbol{z}}(t+\lambda))\boldsymbol{z}^*(t-\tau)\right\} = 0 \text{ for } \tau \geq 0 \\ \Leftrightarrow \quad & E\left\{e^{j\boldsymbol{\omega}(\lambda+\tau)} - e^{j\boldsymbol{\omega}\tau}H(\boldsymbol{\omega})\right\} = 0 \text{ for } \tau \geq 0 \\ \Leftrightarrow \quad & \int_{-\infty}^{\infty}[A(\omega)e^{j\omega\lambda}]e^{j\omega\tau}d\omega = \int_{-\infty}^{\infty} A(\omega)H(\omega)e^{j\omega\tau}d\omega \text{ for } \tau \geq 0 \end{aligned}$$

Example 13-5 Revisited. Let's confirm the alternative expression in terms of Example 13-5.

$$S_{zz}(\omega) = 2\pi A(\omega) = \frac{4\alpha^2}{\alpha^2+\omega^2}$$

Then,

$$\int_{-\infty}^{\infty}[A(\omega)e^{j\omega\lambda}]e^{j\omega\tau}d\omega \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{4\alpha^2}{(\alpha^2+\omega^2)}e^{j\omega(\lambda+\tau)}d\omega = 2\alpha e^{-\alpha|\tau+\lambda|}$$

and

$$\begin{aligned}\int_{-\infty}^{\infty}A(\omega)H(\omega)e^{j\omega\tau}d\omega = 2\alpha e^{-\alpha(|\tau|+\lambda)} &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{4\alpha^2}{(\alpha^2+\omega^2)}e^{-\alpha\lambda}e^{j\omega\tau}d\omega \\ &= 2\alpha e^{-\alpha(|\tau|+\lambda)}\end{aligned}$$

$\Rightarrow\; |\tau+\lambda| = |\tau|+\lambda,$ which is valid only for $\tau \geq \max\{0,-\lambda\} = 0$ (since $\lambda > 0$)

Note that it is erroneous to claim $A(\omega)e^{j\omega\lambda} = A(\omega)H(\omega)$ from

$$\int_{-\infty}^{\infty}[A(\omega)e^{j\omega\lambda}]e^{j\omega\tau}d\omega = \int_{-\infty}^{\infty}A(\omega)H(\omega)e^{j\omega\tau}d\tau$$

because the equation holds only for $\tau \geq 0$.

Predictable Processes

Definition (Predictable processes) A process $\boldsymbol{s}[n]$ is *predictable* if it equals its linear predictor, i.e.,

$$\boldsymbol{s}[n] = \sum_{k=1}^{\infty} h[k]\boldsymbol{s}[n-k]$$

and there is no MS prediction error.

Formula for predictable processes.

- Let $\mathtt{E}[z] = 1 - H[z] = 1 - \sum_{k=1}^{\infty} h[k]z^{-k}$. Then, the prediction error equals

$$P = E[(\boldsymbol{s}[n] - \hat{\boldsymbol{s}}[n])\boldsymbol{s}[n]] = R_{ss}[0] - \sum_{k=1}^{\infty} h[k]R_{ss}[k].$$

  Equivalently,

$$P = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\mathtt{E}[e^{j\omega}]|^2 S_{ss}[\omega] d\omega.$$

- For predictable processes, $P = 0$, which indicates from $S_{ss}[\omega] \geq 0$ that

$$S_{xx}[\omega] > 0 \text{ only possibly at those } \omega\text{'s with } \mathtt{E}[e^{j\omega}] = 0.$$

  As $\mathtt{E}[z]$ is a polynomial of z, it follows that for countably many ω_i,

$$S_{ss}[\omega] = 2\pi \sum_i \alpha_i \delta(\omega - \omega_i) \text{ where } \mathtt{E}[e^{j\omega_i}] = 0.$$

Predictable Processes

- This concludes that a process $\boldsymbol{s}[n]$ that is a sum of exponentials:

$$\boldsymbol{s}[n] = \sum_i \boldsymbol{c}_i e^{j\omega_i n} \text{ where } \{\boldsymbol{c}_i\} \text{ uncorrelated and zero-mean}$$

  is predictable, and its prediction filter equals $H[z] = 1 - \mathtt{E}[z]$, where

$$\mathtt{E}[z] = \prod_{i=1}^{m} \left(1 - e^{j\omega_i} z^{-1}\right).$$

FIR Predictors

Concern

- To find the best linear estimator of $\boldsymbol{s}[n]$, in the MS sense, in terms of its N most recent past, i.e., $\{\boldsymbol{s}[n-k]\}_{1\le k\le N}$.
- This is also named the *forward predictor* of order N.

Yule-Walker equations

- By orthogonality principle,
$$E\left\{\left(\boldsymbol{s}[n]-\sum_{k=1}^{N}a_k\boldsymbol{s}[n-k]\right)\boldsymbol{s}[n-m]\right\}=0 \text{ for } 1\le m\le N.$$
This yields
$$R_{ss}[m]-\sum_{k=1}^{N}a_kR_{ss}[m-k]=0 \text{ for } 1\le m\le N$$
or equivalently,
$$\begin{bmatrix}R_{ss}[1]\\R_{ss}[2]\\\vdots\\R_{ss}[N]\end{bmatrix}=\begin{bmatrix}R_{ss}[0] & R_{ss}[-1] & \cdots & R_{ss}[1-N]\\R_{ss}[1] & R_{ss}[0] & \cdots & R_{ss}[2-N]\\\vdots & \vdots & \ddots & \vdots\\R_{ss}[N-1] & R_{ss}[N-2] & \cdots & R_{ss}[0]\end{bmatrix}\begin{bmatrix}a_1\\a_2\\\vdots\\a_N\end{bmatrix}$$

FIR Predictors

- The MS estimate error is equal to:

$$P_N = E\left\{\left(\boldsymbol{s}[n] - \sum_{k=1}^{N} a_k \boldsymbol{s}[n-k]\right)\boldsymbol{s}[n]\right\} = R_{ss}[0] - \sum_{k=1}^{N} a_k R_{ss}[-k].$$

- We can incorporate the above result into the Yule-Walker equations:

$$\begin{aligned}\begin{bmatrix} P_N & 0 & \cdots & 0 \end{bmatrix} &= \begin{bmatrix} 1 & -a_1 & \cdots & -a_N \end{bmatrix} \begin{bmatrix} R_{ss}[0] & R_{ss}[1] & R_{ss}[2] & \cdots & R_{ss}[N] \\ R_{ss}[-1] & R_{ss}[0] & R_{ss}[1] & \cdots & R_{ss}[N-1] \\ R_{ss}[-2] & R_{ss}[-1] & R_{ss}[0] & \cdots & R_{ss}[N-2] \\ \vdots & \vdots & \ddots & \vdots & \\ R_{ss}[-N] & R_{ss}[1-N] & R_{ss}[2-N] & \cdots & R_{ss}[0] \end{bmatrix} \\ &= \begin{bmatrix} 1 & -a_1 & \cdots & -a_N \end{bmatrix} \mathbb{D}_{N+1}\end{aligned}$$

Recall that for a square matrix $\mathbb{D}$:

$$\mathbb{D} \cdot \mathrm{Adj}(\mathbb{D}) = |A|\mathbb{I},$$

where $D_{i,j}$ is the cofactor of element $d_{i,j}$ in $\mathbb{D}$ (specifically, $D_{i,j} = (-1)^{i+j} M_{i,j}$ and $M_{i,j}$ is the determinant of the matrix by removing those elements at the same row and the same column as $d_{i,j}$), and $\mathrm{Adj}(\mathbb{D}) = [D_{i,j}]^T$.

Hence,

$$\begin{bmatrix} P_N & 0 & \cdots & 0 \end{bmatrix} \mathrm{Adj}(\mathbb{D}_{N+1}) = \begin{bmatrix} 1 & -a_1 & \cdots & -a_N \end{bmatrix} |\mathbb{D}_{N+1}|,$$

which implies $P_N|\mathbb{D}_N| = |\mathbb{D}_{N+1}|$.

- As a result, $P_N = \begin{cases} 0, & \text{if for some } k \leq N,\ |\mathbb{D}_k| \neq 0 \text{ and } |\mathbb{D}_{k+1}| = 0 \\ \dfrac{|\mathbb{D}_{N+1}|}{|\mathbb{D}_N|}, & |\mathbb{D}_N| \neq 0. \end{cases}$

Final note of the optimal $\{a_k\}_{k=1}^N$

- The optimal a_1 in a system of order N may be different from that in a system of order $N+1$. This may cause some scalability problem in implementation.

Example. Suppose $R_{ss}[m] = \rho^{|m|}$ for $m = 0, \pm 1$, and zero, otherwise.
Then,
$a_1 = \rho$ and $P_1 = 1 - \rho^2$ when $N = 1$.
$a_1 = \rho/(1-\rho^2)$, $a_2 = -\rho^2/(1-\rho^2)$ and $P_2 = (1-2\rho^2)/(1-\rho^2)$ when $N = 2$.

Implementation Structure of FIR Predictor

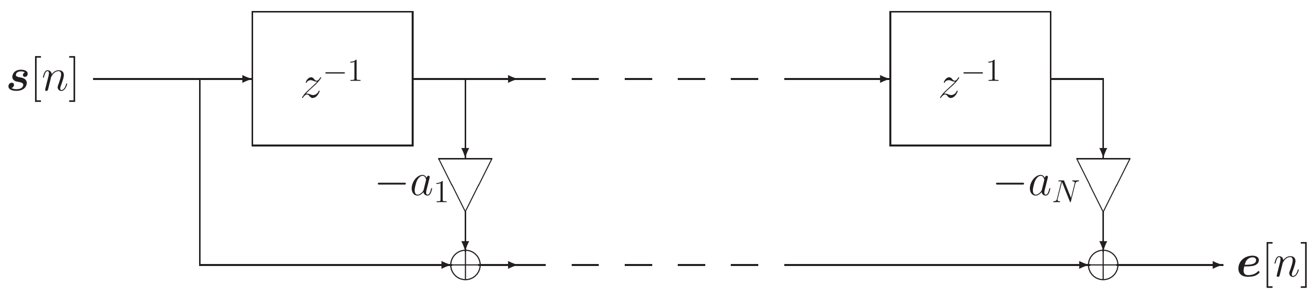
Non-scalable straightforward structure

- The predictor error

$$e[n] = s[n] - \hat{s}[n] = s[n] - \sum_{k=1}^{N} a_k s[n-k]$$

can be obtained by input $s[n]$ to the filter $H[z] = 1 - a_1 z^{-1} - \cdots - a_N z^{-N}$.

- The filter $H[z]$ can be implemented using the ladder structure as follows.

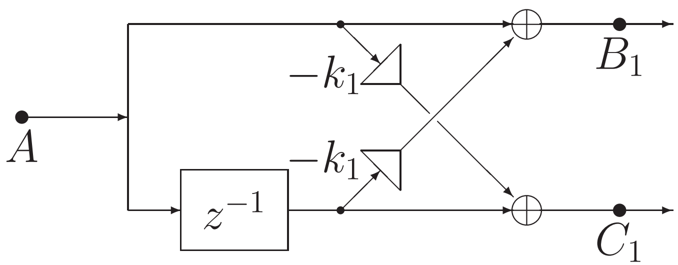


- This structure is not scalable in coefficients $\{a_k\}_{k=1}^N$ (since coefficients $\{a_k\}_{k=1}^N$ are dependent on N).

Implementation Structure of FIR Predictor

Is there a scalable implementation structure?

- Denote the optimal $\{a_k\}_{k=1}^N$ in a system of order N as $\{a_k^{(N)}\}_{k=1}^N$.
- Consider the below lattice structure:



Denote the input at A as $\boldsymbol{s}[n]$.

Denote the outputs at B_1 and C_1 respectively by $\hat{\boldsymbol{e}}_1[n]$ and $\check{\boldsymbol{e}}_1[n]$.

Then,

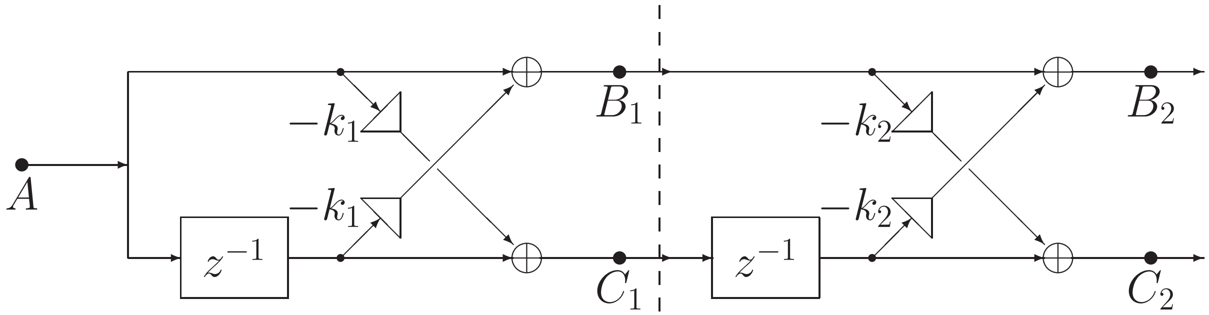
$$\begin{aligned}\hat{\boldsymbol{e}}_1[n] &= \boldsymbol{s}[n] - k_1\boldsymbol{s}[n-1]\\ \check{\boldsymbol{e}}_1[n] &= -k_1\boldsymbol{s}[n] + \boldsymbol{s}[n-1]\end{aligned}$$

So, the filters for output $\hat{\boldsymbol{e}}_1[n]$ and output $\check{\boldsymbol{e}}_1[n]$ are

$$\begin{aligned}\hat{\mathtt{E}}_1[z] &= 1 - k_1 z^{-1}\\ \check{\mathtt{E}}_1[z] &= -k_1 + z^{-1} = z^{-1}\hat{\mathtt{E}}_1[1/z]\end{aligned}$$

Implementation Structure of FIR Predictor

- Consider the below lattice structure:



Denote the input at A as $\boldsymbol{s}[n]$.

Denote the outputs at B_2 and C_2 respectively by $\hat{\boldsymbol{e}}_2[n]$ and $\check{\boldsymbol{e}}_2[n]$.

Then,

$$\begin{aligned}\hat{\boldsymbol{e}}_2[n] &= \hat{\boldsymbol{e}}_1[n] - k_2\check{\boldsymbol{e}}_1[n-1]\\ \check{\boldsymbol{e}}_2[n] &= -k_2\hat{\boldsymbol{e}}_1[n] + \check{\boldsymbol{e}}_1[n-1]\end{aligned}$$

So, the filters for output $\hat{\boldsymbol{e}}_2[n]$ and output $\check{\boldsymbol{e}}_2[n]$ are

$$\begin{aligned}\hat{\mathtt{E}}_2[z] &= \hat{\mathtt{E}}_1[z] - k_2 z^{-1}\check{\mathtt{E}}_1[z]\\ \check{\mathtt{E}}_2[z] &= -k_2\hat{\mathtt{E}}_1[z] + z^{-1}\check{\mathtt{E}}_1[z]\end{aligned}$$

Implementation Structure of FIR Predictor

- Continuing cascading more "lattices," we obtain

$$\begin{aligned}\hat{\boldsymbol{e}}_N[n] &= \hat{\boldsymbol{e}}_{N-1}[n] - k_N \check{\boldsymbol{e}}_{N-1}[n-1] \\ \check{\boldsymbol{e}}_N[n] &= -k_N \hat{\boldsymbol{e}}_{N-1}[n] + \check{\boldsymbol{e}}_{N-1}[n-1]\end{aligned}$$

and

$$\begin{aligned}\hat{\mathtt{E}}_N[z] &= \hat{\mathtt{E}}_{N-1}[z] - k_N z^{-1} \check{\mathtt{E}}_{N-1}[z] \\ \check{\mathtt{E}}_N[z] &= -k_N \hat{\mathtt{E}}_{N-1}[z] + z^{-1} \check{\mathtt{E}}_{N-1}[z]\end{aligned}$$

Then, $\check{\mathtt{E}}_N[z] = z^{-N} \hat{\mathtt{E}}_N[1/z]$.

Proof: Suppose $\check{\mathtt{E}}_{N-1}[z] = z^{-(N-1)} \hat{\mathtt{E}}_{N-1}[1/z]$. Then,

$$\begin{aligned}z^{-N} \hat{\mathtt{E}}_N[1/z] &= z^{-N} \left(\hat{\mathtt{E}}_{N-1}[1/z] - k_N z \check{\mathtt{E}}_{N-1}[1/z] \right) \\ &= z^{-N} \left(z^{N-1} \check{\mathtt{E}}_{N-1}[z] - k_N z (z^{N-1} \hat{\mathtt{E}}_{N-1}[z]) \right) \\ &= -k_N \hat{\mathtt{E}}_{N-1}[z] + z^{-1} \check{\mathtt{E}}_{N-1}[z] \\ &= \check{\mathtt{E}}_N[z].\end{aligned}$$

Implementation Structure of FIR Predictor

- By $\check{\mathtt{E}}_N[z] = z^{-N}\hat{\mathtt{E}}_N[1/z]$, we know that if

$$\hat{\mathtt{E}}_N[z] = 1 - a_1^{(N)} z^{-1} - \cdots - a_N^{(N)} z^{-N},$$

then

$$\check{\mathtt{E}}_N[z] = z^{-N} - a_1^{(N)} z^{-(N-1)} - \cdots - a_N^{(N)}.$$

In summary,

  - $\hat{\boldsymbol{e}}_N[n]$ is the *forward* prediction error for predicting $\boldsymbol{s}[n]$ in terms of its most recent N *pasts*. In other words,

$$\hat{\boldsymbol{e}}_N[n] = \boldsymbol{s}[n] - \hat{\boldsymbol{s}}_N[n] = \boldsymbol{s}[n] - \sum_{k=1}^{N} a_k^{(N)} \boldsymbol{s}[n-k].$$

  - $\check{\boldsymbol{e}}_N[n]$ is the *backward* prediction error for predicting $\boldsymbol{s}[n-N]$ in terms of its most recent N *futures*. In other words,

$$\check{\boldsymbol{e}}_N[n] = \boldsymbol{s}[n-N] - \check{\boldsymbol{s}}_N[n-N] = \boldsymbol{s}[n-N] - \sum_{k=1}^{N} a_k^{(N)} \boldsymbol{s}[n-N+k].$$

Implementation Structure of FIR Predictor

Derivation of k_N

- From

$$\hat{\mathtt{E}}_{N-1}[z] = 1 - a_1^{(N-1)} z^{-1} - \cdots - a_{N-1}^{(N-1)} z^{-(N-1)},$$

and

$$\check{\mathtt{E}}_{N-1}[z] = z^{-N} \hat{\mathtt{E}}_{N-1}[1/z],$$

we derive:

$$\begin{aligned} \hat{\mathtt{E}}_N[z] &= \hat{\mathtt{E}}_{N-1}[z] - k_N z^{-1} \check{\mathtt{E}}_{N-1}[z] \\ &= \left(1 - a_1^{(N-1)} z^{-1} - \cdots - a_{N-1}^{(N-1)} z^{-(N-1)}\right) \\ &\quad -k_N \left(z^{-N} - a_1^{(N-1)} z^{-(N-1)} - \cdots - a_{N-1}^{(N-1)} z^{-1}\right) \\ &= 1 - \left(a_1^{(N-1)} - k_N a_{N-1}^{(N-1)}\right) z^{-1} - \left(a_2^{(N-1)} - k_N a_{N-2}^{(N-1)}\right) z^{-2} - \cdots \\ &\quad - \left(a_{N-1}^{(N-1)} - k_N a_1^{(N-1)}\right) z^{-(N-1)} - k_N z^{-N}. \end{aligned}$$

Comparing termwisely with

$$\hat{\mathtt{E}}_N[z] = 1 - a_1^{(N)} z^{-1} - \cdots - a_N^{(N)} z^{-N},$$

we yield:

$$a_k^{(N)} = a_k^{(N-1)} - {\color{red}k_N} a_{N-k}^{(N-1)} \text{ for } 1 \le k < N \quad \text{and} \quad a_N^{(N)} = {\color{red}k_N}.$$

Implementation Structure of FIR Predictor

- It remains to solve k_N:

$$\begin{bmatrix} P_N & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 1 & -a_1^{(N)} & \cdots & -a_N^{(N)} \end{bmatrix} \begin{bmatrix} R_{ss}[0] & R_{ss}[1] & R_{ss}[2] & \cdots & R_{ss}[N] \\ R_{ss}[-1] & R_{ss}[0] & R_{ss}[1] & \cdots & R_{ss}[N-1] \\ R_{ss}[-2] & R_{ss}[-1] & R_{ss}[0] & \cdots & R_{ss}[N-2] \\ \vdots & \vdots & \ddots & \vdots & \\ R_{ss}[-N] & R_{ss}[1-N] & R_{ss}[2-N] & \cdots & R_{ss}[0] \end{bmatrix}$$

implies

$$0 = R_{ss}[N] - \sum_{k=1}^{N-1} a_k^{(N)} R_{ss}[N-k] - k_N R_{ss}[0]$$

$$\Rightarrow \; 0 = R_{ss}[N] - \sum_{k=1}^{N-1} \left(a_k^{(N-1)} - k_N a_{N-k}^{(N-1)} \right) R_{ss}[N-k] - k_N R_{ss}[0]$$

$$\Rightarrow \; k_N = \frac{R_{ss}[N] - \sum_{k=1}^{N-1} a_k^{(N-1)} R_{ss}[N-k]}{R_{ss}[0] - \sum_{k=1}^{N-1} a_{N-k}^{(N-1)} R_{ss}[N-k]} = \frac{1}{P_{N-1}} \left(R_{ss}[N] - \sum_{k=1}^{N-1} a_k^{(N-1)} R_{ss}[N-k] \right),$$

where the last step follows from the fact that $R_{ss}[N-k] = R_{ss}[k-N]$ (See Slide 13-39).

- The above (blue-colored) formula gives k_N from known P_{N-1} and $\{a_k^{(N-1)}\}_{k=1}^{N-1}$.

Implementation Structure of FIR Predictor

Alternative derivation of k_N in terms of $\hat{\boldsymbol{e}}_N[n] = \hat{\boldsymbol{e}}_{N-1}[n] - k_N \check{\boldsymbol{e}}_{N-1}[n-1]$.

- $\begin{cases} \hat{\boldsymbol{e}}_N[n] = \boldsymbol{s}[n] - \hat{\boldsymbol{s}}_N[n] = \boldsymbol{s}[n] - \sum_{k=1}^{N} a_k^{(N)} \boldsymbol{s}[n-k] \\ \hat{\boldsymbol{e}}_{N-1}[n] = \boldsymbol{s}[n] - \hat{\boldsymbol{s}}_{N-1}[n] = \boldsymbol{s}[n] - \sum_{k=1}^{N-1} a_k^{(N-1)} \boldsymbol{s}[n-k] \end{cases}$

  implies $\quad P_N = E\left[\,\hat{\boldsymbol{e}}_N[n]\,\boldsymbol{s}[n]\,\right] \quad$ and $\quad P_{N-1} = E\left[\,\hat{\boldsymbol{e}}_{N-1}[n]\,\boldsymbol{s}[n]\,\right]$.

- $$\begin{aligned} \check{\boldsymbol{e}}_{N-1}[n-1] &= \boldsymbol{s}[(n-1)-(N-1)] - \sum_{k=1}^{N-1} a_k^{(N-1)} \boldsymbol{s}[(n-1)-(N-1)+k] \\ &= \boldsymbol{s}[n-N] - \sum_{k=1}^{N-1} a_k^{(N-1)} \boldsymbol{s}[n-N+k] \end{aligned}$$

  implies

  $$\begin{aligned} E[\,\check{\boldsymbol{e}}_{N-1}[n-1]\boldsymbol{s}[n]\,] &= E[\boldsymbol{s}[n-N]\boldsymbol{s}[n]] - \sum_{k=1}^{N-1} a_k^{(N-1)} E[\boldsymbol{s}[n-N+k]\boldsymbol{s}[n]] \\ &= R_{ss}[N] - \sum_{k=1}^{N-1} a_k^{(N-1)} S_{ss}[N-k] = k_N P_{N-1}. \end{aligned}$$

- Hence, $P_N = P_{N-1} - k_N(k_N P_{N-1}) = (1 - k_N^2) P_{N-1}$.

Levinson's Algorithm

Concern:

- A recursive algorithm to obtain k_N and MS estimate error P_N.

Levinson's algorithm

- $k_1 = a_1^{(1)} = R_{ss}[1]/R_{ss}[0]$ and $P_1 = (1 - k_1^2)R_{ss}[0]$.
- Assume that $\{a_k^{(N-1)}\}_{k=1}^{N-1}$, k_{N-1} and P_{N-1} are known.

  Then, it can be derived that

$$k_N = \frac{1}{P_{N-1}}\left(R_{ss}[N] - \sum_{k=1}^{N-1} a_k^{(N-1)} R_{ss}[N-k]\right)$$

$$P_N = (1 - k_N^2)P_{N-1}$$

$$a_k^{(N)} = \begin{cases} a_k^{(N-1)} - k_N a_{N-k}^{(N-1)}, & 1 \le k \le N-1 \\ k_N, & k = N \end{cases}$$

Properties of FIR estimator

- $P_1 \geq P_2 \geq \cdots \geq P_N \geq \cdots \geq 0$.
- If $P_N > 0$,

  then $|k_i| < 1$ for $1 \leq i \leq N$,

  and z_i (the root of $\hat{\mathtt{E}}_N[z] = 1 - \sum_{k=1}^{N} a_k^{(N)} z^{-k}$) satisfies $|z_i| < 1$ for $1 \leq i \leq N$.
- If $P_{N-1} > 0$ and $P_N = 0$,

  then $|k_i| < 1$ for $1 \leq i < N$ and $k_N = 1$,

  and $|z_i| = 1$ for $1 \leq i \leq N$,

  which indicates that $\boldsymbol{s}[n]$ is predictable and consists of line spectrum.
- If $P = \lim_{N\to} P_N > 0$,

  then
  $$P = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log(S_{ss}[\omega])d\omega\right\} = \mathtt{l}[0] = \lim_{N\to\infty} \frac{|\mathbb{D}_{N+1}|}{|\mathbb{D}_N|}.$$
- If $P_{M-1} > P_M$ but $P_M (= P_{M+1} = \cdots) = P$,

  then $k_i = 0$ for $i > M$,

  and $\boldsymbol{s}[n]$ is wide-sense Markov of order M.

  $\boxed{\boldsymbol{s}[n] \text{ is autoregressive (AR) if, and only if, it is wide-sense Markov of finite order.}}$

Properties of FIR estimator

Equal predictor of two processes

- Suppose process $\boldsymbol{s}[n]$ and $\bar{\boldsymbol{s}}[n]$ have the same autocorrelation function up to order M.

  Then, the predictors of these two processes of order M are identical because the predictors only depend on the value of $R_{ss}[m]$ for $|m| \leq M$.

  Also, from Levinson's algorithm, we learn that P_M for both processes are the same since $P_M = \prod_{i=1}^{M}(1 - k_i^2)R_{ss}[0]$.

Kalman Innovations

Define the process $\boldsymbol{i}[n]$ as

$$\begin{aligned}
\boldsymbol{i}[n] &\triangleq \frac{\hat{\boldsymbol{e}}_n[n]}{\sqrt{P_n}} \\
&= \frac{1}{\sqrt{P_n}}\left(\boldsymbol{s}[n] - \sum_{k=1}^{n} a_k^{(n)} \boldsymbol{s}[n-k]\right) \\
&= \sum_{k=0}^{n} \gamma_k^{(n)} \boldsymbol{s}[k] \text{ for some } \{\gamma_k^{(n)} = -a_{n-k}^{(n)}/\sqrt{P_n}\}_{k=0}^{n-1} \text{ and } \gamma_n^{(n)} = 1/\sqrt{P_n}.
\end{aligned}$$

By orthogonality principle, $\boldsymbol{i}[n]$ is orthogonal to $\boldsymbol{s}[n-m]$ for $1 \le m \le n$; hence, $\boldsymbol{i}[n]$ is orthogonal to $\boldsymbol{i}[n-m]$ for $1 \le m \le n$, and $E[\boldsymbol{i}^2[n]] = 1$.

Kalman Innovations

In matrix form,

$$\begin{aligned}\begin{bmatrix}\boldsymbol{i}[0] & \boldsymbol{i}[1] & \cdots & \boldsymbol{i}[n]\end{bmatrix} &= \begin{bmatrix}\boldsymbol{s}[0] & \boldsymbol{s}[1] & \cdots & \boldsymbol{s}[n]\end{bmatrix}\begin{bmatrix}\gamma_0^{(0)} & \gamma_0^{(1)} & \cdots & \gamma_0^{(n)} \\ 0 & \gamma_1^{(1)} & \cdots & \gamma_1^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_n^{(n)}\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{s}[0] & \boldsymbol{s}[1] & \cdots & \boldsymbol{s}[n]\end{bmatrix}\boldsymbol{\Gamma}_{n+1}\end{aligned}$$

Remarks

- This is similarly the Gram-Schmidt orthonormalization procedure for $\begin{bmatrix}\boldsymbol{s}[0] & \boldsymbol{s}[1] & \cdots & \boldsymbol{s}[n]\end{bmatrix}$.
- In terminologies, $\boldsymbol{i}[n]$ is called the *Kalman innovations* of $\boldsymbol{s}[n]$, and $\boldsymbol{\Gamma}_{n+1}$ is called the *Kalman whitening filter* of $\boldsymbol{s}[n]$.
- It can then be derived:

$$\begin{aligned}\begin{bmatrix}\boldsymbol{s}[0] & \boldsymbol{s}[1] & \cdots & \boldsymbol{s}[n]\end{bmatrix} &= \begin{bmatrix}\boldsymbol{i}[0] & \boldsymbol{i}[1] & \cdots & \boldsymbol{i}[n]\end{bmatrix}\begin{bmatrix}\ell_0^{(0)} & \ell_0^{(1)} & \cdots & \ell_0^{(n)} \\ 0 & \ell_1^{(1)} & \cdots & \ell_1^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \ell_n^{(n)}\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{i}[0] & \boldsymbol{i}[1] & \cdots & \boldsymbol{i}[n]\end{bmatrix}\mathbb{L}_{n+1}\end{aligned}$$

Kalman Innovations

Then, the covariance matrix of $\boldsymbol{s}[n]$ is given by:

$$\mathbb{R}_{n+1} \triangleq E\left\{ \begin{bmatrix} \boldsymbol{s}[0] \\ \boldsymbol{s}[1] \\ \cdots \\ \boldsymbol{s}[n] \end{bmatrix} \begin{bmatrix} \boldsymbol{s}[0] & \boldsymbol{s}[1] & \cdots & \boldsymbol{s}[n] \end{bmatrix} \right\} = \mathbb{L}_{n+1}^T \mathbb{L}_{n+1}.$$

Therefore,

$$\boldsymbol{\Gamma}_{n+1}^T \mathbb{R}_{n+1} \boldsymbol{\Gamma}_{n+1} = \mathbb{I}_{n+1},$$

where $\mathbb{I}_{n+1}$ is the identity matrix.

The end of Section 13-2 Prediction