Stochastic Process

- *Stochastic process* (*random process*) is a time evolution of a statistical phenomenon according to probabilistic laws (function of time and not deterministic).
- We study a stochastic process at discrete and uniformly spaced instances of time.
- *Discrete-time series* (or *time series*) or *discrete-time signal* is one particular realization of a discrete-time stochastic process.
- A stochastic process is *strictly stationary* if its statistical properties are invariant over time, i.e., all the joint probability density functions of all times remain the same.

(1) Characterization of stochastic processes

- We consider a discrete-time stochastic process represented by the time series $u[n]$, $u[n-1]$, ..., $u[n-M]$.
- The *mean-value function* is

$$\mu[n] = E\left[u[n]\right] = \sum_{u[n]=-\infty}^{\infty} u[n]\, f_{u[n]}$$

where E is the statistical expectation operator (ensemble average) and $f_{u[n]}$ is the probability mass function of $u[n]$.
- The *autocorrelation function* is

$$r[n,n-k] = E\left[u[n]u^*[n-k]\right], k = 0,\ \pm1,\ \pm2,\ \cdots$$

where * denotes complex conjugation.
- The *autocovariance function* is

$$\begin{aligned} c[n,n-k] &= E\left[\left(u[n]-\mu[n]\right)\left(u[n-k]-\mu[n-k]\right)^*\right], k = 0,\ \pm1,\ \pm2,\ \cdots \\ &= r[n,n-k]-\mu[n]\mu^*[n-k]. \end{aligned}$$

- For *strictly stationary stochastic processes*, all the joint probability density functions at all times remain the same. Therefore,
  (1) $\forall n, \mu[n] = \mu,\ \ r[n,n-k] = r[k],\ \ c[n,n-k] = c[k]$.
  (2) $r[0] = E\left[\left|u[n]\right|^2\right]$ is the *mean-square value* of $u[n]$.
  (3) $c[0] = r[0] - \mu^2 = \sigma_u^2$ is the variance of $u[n]$.
- In addition, if $\forall n, E\left[\left|u[n]\right|^2\right] < \infty$, the process is *wide sense stationary (WSS)* and the above three conditions hold.

(2) Mean ergodic theorem for WSS processes

- The *expectation* is "*ensemble averages*" of a stochastic process across the process.
- The "*time average*" is a long-term sample averages along the process. That is

$$\hat{\mu}[N] = \frac{1}{N}\sum_{n=0}^{N-1} u[n]$$

and $\hat{\mu}(N)$ itself is a random variable. We find that $\hat{\mu}(N)$ is unbiased since

$$E\left[\hat{\mu}[N]\right] = \mu, \forall N .$$

- We say the process $u[n]$ is *mean ergodic in the mean-square error sense* if

$$\lim_{N\to\infty} E\left[\left(\mu - \hat{\mu}[N]\right)^2\right] = \lim_{N\to\infty} \frac{1}{N}\sum_{l=-N+1}^{N-1}\left(1-\frac{|l|}{N}\right)c(l) = 0 ,$$

i.e., the process is asymptotically uncorrelated.
- The process is *correlation ergodic in the mean-square sen*se, if

$$r[k,N] = \frac{1}{N}\sum_{n=0}^{N-1} u[n]u^*[n-k], 0 \le k \le N-1$$

and

$$\lim_{N\to\infty} E\left[\left(r[k]-\hat{r}[k,N]\right)^2\right] = 0 .$$

(3) Correlation matrix

- Let $M \times 1$ *observation vector* $\mathbf{u}[n]$ be

$$\mathbf{u}[n] = \left[u[n], u[n-1], \cdots, u[n-M+1]\right]^T .$$

- We define the *correlation matrix* $\mathbf{R}$ as

$$\begin{aligned}\mathbf{R} &= E\left[\mathbf{u}[n]\mathbf{u}^H[n]\right]\\ &= \begin{bmatrix} E\left[u[n]u^*[n]\right] & E\left[u[n]u^*[n-1]\right] & \cdots & E\left[u[n]u^*[n-M+1]\right] \\ E\left[u[n-1]u^*[n]\right] & E\left[u[n-1]u^*[n-1]\right] & \cdots & E\left[u[n-1]u^*[n-M+1]\right] \\ \vdots & \vdots & \vdots & \vdots \\ E\left[u[n-M+1]u^*[n]\right] & E\left[u[n-M+1]u^*[n-1]\right] & \cdots & E\left[u[n-M+1]u^*[n-M+1]\right] \end{bmatrix}.\end{aligned}$$

Then,

$$\mathbf{R} = \begin{bmatrix} r[0] & r[1] & \cdots & r[M-1] \\ r[-1] & r[0] & \cdots & r[M-2] \\ \vdots & \vdots & \vdots & \vdots \\ r[-M+1] & r[-M+2] & \cdots & r[0] \end{bmatrix}.$$

- The correlation matrix has the following properties.

(1) From the definition of **R**, $\mathbf{R} = \mathbf{R}^H$. Therefore, $r[-k] = r^*[k]$. If $u[n]$ is real $\mathbf{R} = \mathbf{R}^T$. Therefore,

$$\mathbf{R} = \begin{bmatrix} r[0] & r[1] & \cdots & r[M-1] \\ r^*[1] & r[0] & \cdots & r[M-2] \\ \vdots & \vdots & \vdots & \vdots \\ r^*[M-1] & r^*[M-2] & \cdots & r[0] \end{bmatrix}.$$

(2) **R** is *Toeplitz* iff $u[n]$ is WSS.

(3) **R** is always *nonnegative definite* ($\forall \mathbf{x} \neq \mathbf{0}, \mathbf{x}^H \mathbf{R} \mathbf{x} \geq 0$) and almost always *positive definite* ($\forall \mathbf{x} \neq \mathbf{0}, \mathbf{x}^H \mathbf{R} \mathbf{x} > 0$). Therefore, **R** is almost always *nonsingular* and det(**R**) ≠ 0 and $\mathbf{R}^{-1}$ exists.

(4) Let $\mathbf{u}^B[n] = \left[u[n-M+1], u[n-M+1], \cdots, u[n]\right]^T$, then

$$E\left[\mathbf{u}^B[n]\mathbf{u}^{BH}[n]\right] = \mathbf{R}^T.$$

(5) $\mathbf{R}_{M+1} = \begin{bmatrix} r[0] & \mathbf{r}^H \\ \mathbf{r} & \mathbf{R}_M \end{bmatrix} = \begin{bmatrix} \mathbf{R}_M & \mathbf{r}^{B*} \\ \mathbf{r}^{BT} & r[0] \end{bmatrix}$ with $\mathbf{r}^H = \left[r[1], r[2], \cdots, r[M]\right]$ and $\mathbf{r}^{BT} = \left[r[-M], r[-M+1], \cdots, r[-1]\right]$.

Stochastic Models

- *Stochastic model* is a representation of a stochastic process.
- We are interested in a linear process as a LTI system where

$$u[n]+\sum_{k=1}^{M}a_k^*u[n-k]=\sum_{k=0}^{N}b_k^*v[n-k]$$

with $v[n]$, white Gaussian noise as an input with $E\left[v[n]\right]=0$ for all n, $r[0]=\sigma_v^2$, and $r[k]=0$ for all $k\neq 0$.

(1) Autoregressive model (AR) model

- Time series $u[n]$, $u[n\text{-}1]$, ..., $u[n - M]$ represents the realization of an AR of order M if

$$u[n]+\sum_{k=1}^{M}a_k^*u[n-k]=v[n] \quad \text{or} \quad u[n]=\sum_{k=1}^{M}a_k^*u[n-k]+v[n].$$

- With $a_0=1$, $\sum_{k=0}^{M}a_k^*u[n-k]=v[n]$, $H_{AR}(z)=\dfrac{U(z)}{V(z)}=\dfrac{1}{\sum_{k=0}^{M}a_k^*z^{-k}}$. Therefore, the filter is an all-pole filter with infinite duration.

(2) Moving average (MA) model

- Time series $u[n]$, $u[n\text{-}1]$, ..., $u[n - M]$ represents the realization of an MA of order N if

$$u[n]=\sum_{k=0}^{N}b_k^*v[n-k] \quad \text{with} \quad b_0=1.$$

- $H_{MA}(z)=\dfrac{U(z)}{V(z)}=\sum_{k=0}^{N}b_k^*z^{-k}$. Therefore, the filter is an all-zero filter with finite duration.

(3) Autoregressive-moving average (ARMA) model

- Time series $u[n]$, $u[n\text{-}1]$, ..., $u[n - M]$ represents the realization of an ARMA of order (M, N) if

$$u[n]+\sum_{k=1}^{M}a_k^*u[n-k]=\sum_{k=0}^{N}b_k^*v[n-k] \quad \text{with} \quad b_0=1.$$

- $H_{ARMA}(z) = \dfrac{U(z)}{V(z)} = \dfrac{\sum_{k=0}^{N} b_k^* z^{-k}}{\sum_{k=0}^{M} a_k^* z^{-k}}$. Therefore, the filter is a mixed type with infinite duration.

(4) Wold decomposition

- Any stationary discrete-time stochastic process $x[n]$ can be expressed as

$$x[n] = u[n] + s[n]$$

where

(1) $u[n]$ and $s[n]$ are uncorrelated,

(2) $u[n]$ is a general linear process represented by the MA model:

$$u[n] = \sum_{k=0}^{N} b_k^* v[n-k] \quad \text{with} \quad b_0 = 1, \text{ and } \sum_{k=0}^{N} |b_k|^2 < \infty$$

with white-noise process $v[n]$ such that $E\left[v[n]s^*[k]\right] = 0$ for all (n, k),

(3) $s[n]$ is a predictable process, that is, the process can be predicted from its own past with zero prediction variance.

(5) Yule-Walker equation (AR model)

- Unique description of an AR model of order M:

(1) The AR coefficients $a_1, a_2, \cdots, a_M$ and

(2) The variance σ_v^2 of $v[n]$.

- Consider the AR model $\sum_{k=0}^{M} a_k^* u[n-k] = v[n]$ with $a_0 = 1$. Multiply both sides by $u^*[n-l]$ and take expectations. Then,

$$E\left[\sum_{k=0}^{M} a_k^* u[n-k] u^*[n-l]\right] = E\left[v[n]u^*[n-l]\right], \text{ or}$$

$$\sum_{k=0}^{M} a_k^* r[l-k] = 0 \quad \text{for} \quad l > 0.$$

For $l = 1, 2, ..., M$, we have

$$\begin{aligned}
l = 1 \Rightarrow \quad & a_0^* r[1] + a_1^* r[0] + a_2^* r[-1] + \cdots + a_M^* r[1-M] = 0 \\
l = 2 \Rightarrow \quad & a_0^* r[2] + a_1^* r[1] + a_2^* r[0] + \cdots + a_M^* r[2-M] = 0 \\
\vdots \quad & \\
l = M \Rightarrow \quad & a_0^* r[M] + a_1^* r[M-1] + a_2^* r[M-2] + \cdots + a_M^* r[0] = 0
\end{aligned}$$

or

$$\begin{bmatrix} r[0] & r[1] & \cdots & r[M-1] \\ r^*[1] & r[0] & \cdots & r[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ r^*[M-1] & r^*[M-2] & \cdots & r[0] \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} = \begin{bmatrix} r^*[1] \\ r^*[2] \\ \vdots \\ r^*[M] \end{bmatrix}$$

with $w_k = -a_k$. In short,

$$\mathbf{R}\mathbf{w} = \mathbf{r} \quad \text{or} \quad \mathbf{w} = \mathbf{R}^{-1}\mathbf{r}.$$

- Note that $E\left[v[n]u^*[n]\right] = E\left[v[n]v^*[n]\right] = \sigma_v^2$ and $\sigma_v^2 = \sum_{k=0}^{M} a_k r[k]$.

Power Spectral Density

- *Autocorrelation function*: time-domain description of the 2nd order statistics
- *Power spectral density* (or power spectrum or spectrum): frequency-domain description of the 2nd order statistics

(1) Power spectral density (PSD)

- Consider a discrete-time time series of infinite duration;

$$\cdots u[n-M]\cdots u[0]\cdots u[n]u[n+1]\cdots .$$

- For a segment of length N, define $u_N[n]=\begin{cases} u[n], & n=0,1,\cdots,N-1 \\ 0, & \text{otherwise}\end{cases}$. Then,

$$U_N(\omega)=\sum_{n=0}^{N-1}u_N[n]e^{-j\omega n},\ \ U_N^*(\omega)=\sum_{k=0}^{N-1}u_N^*[k]e^{j\omega k}\text{, and}$$

$$\left|U_N(\omega)\right|^2=\sum_{n=0}^{N-1}\sum_{k=0}^{N-1}u_N[n]u_N^*[k]e^{-j\omega(n-k)}.$$

- We take expected values of both sides as

$$E\left[\left|U_N(\omega)\right|^2\right]=\sum_{n=0}^{N-1}\sum_{k=0}^{N-1}E\left[u_N[n]u_N^*[k]\right]e^{-j\omega(n-k)}=\sum_{n=0}^{N-1}\sum_{k=0}^{N-1}r_N[n-k]e^{-j\omega(n-k)}.$$

Since we have $r_N[n-k]=\begin{cases} E\left[u_N[n]u_N^*[k]\right]=r[n-k] & \text{for } 0\le(n,k)\le N-1 \\ 0 & \text{otherwise}\end{cases}$,

$$E\left[\left|U_N(\omega)\right|^2\right]=\sum_{n=0}^{N-1}\sum_{k=0}^{N-1}r[n-k]e^{-j\omega(n-k)}.$$

Let $l=n-k$, $\frac{1}{N}E\left[\left|U_N(\omega)\right|^2\right]=\sum_{l=-N+1}^{N-1}\left(1-\frac{|l|}{N}\right)r[l]e^{-j\omega l}=\sum_{l=-N+1}^{N-1}w_B[l]r[l]e^{-j\omega l}$ where

$w_B[l]$ is the Barlett window. Note that $w_B[l]\to 1$ as $N\to\infty$.

- We define the *periodogram* of the windowed time series $u_N[n]$ as

$$P_N(\omega)=\frac{1}{N}\left|U_N(\omega)\right|^2.$$

- We define the *power spectral density* of a WSS discrete-time stochastic process as

$$S(\omega)=\lim_{N\to\infty}E\left[P_N(\omega)\right]=\sum_{l=-\infty}^{\infty}r[l]e^{-j\omega l}\quad\text{with}\quad r[l]=E\left[u[n]u^*[n-l]\right].$$

(2) Properties of PSD

- $S(\omega)=\sum_{l=-\infty}^{\infty} r[l]e^{-j\omega l},\ -\pi<\omega\le\pi \Leftrightarrow r[l]=\frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)e^{j\omega l}d\omega, l=0,\ \pm1,\ \pm2,\ \cdots$
- $S(\omega+2\pi k)=S(\omega)$ for integer k. $-\pi<\omega\le\pi$ is the Nyquist interval.
- PSD of a discrete-time WSS process is real since $r[-k]=r^*[k]$.
- PSD of a real-valued discrete-time WSS process is even (i.e., symmetric), i.e., $S(\omega)=S(-\omega)$.
- $r[0]=\frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)d\omega$: mean-square value or expected power across 1 Ω resistor
- $S(\omega)\ge 0$ for all ω

(3) Transmission of a WSS process through a LTI system

- For a stable LTI system, $S_o(\omega)=\left|H(e^{j\omega})\right|^2 S(\omega)$.

Eigenanalysis of Correlation Matrix

- Consider an $M \times M$ Hermitian matrix **R** which is a correlation matrix of a WSS process.

(1) Eigenvalue and eigenvector

- If $\mathbf{Rq} = \lambda\mathbf{q}$, λ is an *eigenvalue* and **q** is the corresponding *eigenvector*. That is, **q** is *invariant in direction* from the linear transformation by **R**.
- The solution of the characteristic equation, $\det(\mathbf{R} - \lambda\mathbf{I}) = 0$ provide all λ_i and $\mathbf{q}_i$.

(2) Properties of eigenvalues and eigenvectors of R from WSS process

- $\mathbf{R}^k\mathbf{q} = \lambda^k\mathbf{q}$
- $\{\lambda_i\}_{i=1}^{M}$ are all *real* and *nonnegative*.
- If $\{\lambda_i\}_{i=1}^{M}$ are distinct, $\{\mathbf{q}_i\}_{i=1}^{M}$ are linearly independent.
- If $\{\lambda_i\}_{i=1}^{M}$ are distinct, $\{\mathbf{q}_i\}_{i=1}^{M}$ are orthogonal. $\{\mathbf{q}_i\}_{i=1}^{M}$ is an orthogonal basis of span(**R**).
- If $\{\lambda_i\}_{i=1}^{M}$ are distinct and $\{\mathbf{q}_i\}_{i=1}^{M}$ are normalized so that $\mathbf{q}_j^H\mathbf{q}_j = \delta_{ij}$,

$$\mathbf{Q}^H\mathbf{RQ} = \mathbf{\Lambda} \quad \text{(unitary similarity transform)}$$

  with $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_M]$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_M)$. Since $\mathbf{RQ} = \mathbf{Q\Lambda}$, $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$ or $\mathbf{Q}^{-1} = \mathbf{Q}^H$ (unitary matrix),

$$\mathbf{R} = \mathbf{Q\Lambda Q}^H = \sum_{i=1}^{M} \lambda_i \mathbf{q}_i \mathbf{q}_i^H \quad \text{(spectral theorem or Mercer's theorem).}$$

- $\mathrm{tr}\left(\mathbf{Q}^H\mathbf{RQ}\right) = \mathrm{tr}(\mathbf{RQQ}^H) = \mathrm{tr}(\mathbf{R}) = \mathrm{tr}(\mathbf{\Lambda}) = \sum_{i=1}^{M}\lambda_i$.
- The condition number of **R** is $\chi(\mathbf{R}) = \|\mathbf{R}\| \|\mathbf{R}^{-1}\| = \frac{\lambda_{\max}}{\lambda_{\min}}$. **R** is ill-conditioned if $\chi(\mathbf{R})$ is large.
- $S_{\min} \le \lambda_i \le S_{\max}$ where $S_{\min}$ and $S_{\max}$ are the minimum and maximum of the power spectral density of the process. Therefore, $\chi(\mathbf{R}) = \frac{\lambda_{\max}}{\lambda_{\min}} \le \frac{S_{\max}}{S_{\min}}$.

• If $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_M$, $\lambda_k = \min_{\dim(T)=k} \max_{\substack{\mathbf{x}\in T \\ \mathbf{x}\neq\mathbf{0}}} \dfrac{\mathbf{x}^H \mathbf{R}\mathbf{x}}{\mathbf{x}^H \mathbf{x}}, k = 1, 2, \cdots, M$.

• (<u>*Karhunen-Loeve expansion*</u>) $\mathbf{u}[n] = \sum_{i=1}^{M} c_i[n]\mathbf{q}_i$ with $c_i[n] = \mathbf{q}_i^H \mathbf{u}[n]$ for $i = 1, 2, ..., M$.

Note

$$E\left[c_i[n]\right] = 0 \text{ for } i = 1, 2, ..., M \text{ and } E\left[c_i[n]c_j^*[n]\right] = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases}.$$

Also,

$$\sum_{i=1}^{n} \left|c_i[n]\right|^2 = \left\|\mathbf{u}[n]\right\|^2 \text{ and } E\left[\left|c_i[n]\right|^2\right] = \lambda_i \text{ for } i = 1, 2, ..., M.$$

(3) Low-rank modeling or subspace decomposition

- Assume $\lambda_1 > \lambda_2 > \cdots > \lambda_M$ and $\{\lambda_i\}_{i=p+1}^{M}$ are all small, then we can approximate $\mathbf{u}[n]$ by

$$\hat{\mathbf{u}}[n] = \sum_{i=1}^{p} c_i[n]\mathbf{q}_i, p < M .$$

- span$\{\mathbf{q}_i\}_{i=1}^{p}$ is a <u>*feature space*</u> whereas span$\{\mathbf{q}_i\}_{i=1}^{M}$ is a <u>*data space*</u>.
- The <u>*reconstruction error vector*</u> is

$$\mathbf{e}[n] = \hat{\mathbf{u}}[n] - \mathbf{u}[n] = \sum_{i=p+1}^{M} c_i[n]\mathbf{q}_i .$$

- The <u>*mean-square error*</u> is

$$\varepsilon = E\left[\left\|\mathbf{e}[n]\right\|^2\right] = \sum_{i=p+1}^{M} \lambda_i .$$

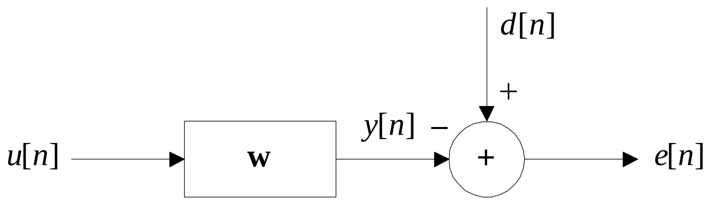
(4) Eigenfilter

- We want to maximize the output signal-to-noise ratio (SNR).
- Eigendecomposition of the correlation matrix $\mathbf{R}$ is performed.
- The eigenvector $\mathbf{q}_{\max}$ which corresponds to the maximal eigenvalue defines the impulse response of the optimal filter.
- This eigenfilter maximize the output SNR for a random signal.
- This corresponds to the matched filter for a deterministic signal.

Wiener Filter

Linear optimal filtering based on known statistics
$u[n]$ and $d[n]$ are jointly WSS

(1) Formulation and solution



Let $\mathbf{w} = \left[w_0 w_1 \cdots w_{M-1}\right]^T \in \mathbf{C}^M$ and $\mathbf{u}[n] = \left[u[n]u[n-1]\cdots u[n-M+1]\right]^T$. Let $y[n] = \mathbf{w}^H\mathbf{u}[n]$ and $e[n] = d[n] - y[n]$. Define

$$J(\mathbf{w}) = E\left\{\left|e[n]\right|^2\right\} = E\left\{\left|d[n] - \mathbf{w}^H\mathbf{u}[n]\right|^2\right\} \text{ and } \mathbf{w}_o = \arg\min_{\mathbf{w}\in\mathbf{C}^M} J(\mathbf{w}).$$

From the <u>*orthogonality principle (OP)*</u>,

$$E\left\{\mathbf{u}[n]e_o^*[n]\right\} = \mathbf{0} \text{ where } e_o[n] = d[n] - \mathbf{w}_o{}^H\mathbf{u}[n].$$

Therefore, $E\left\{\mathbf{u}[n]\left(d[n] - \mathbf{w}_o{}^H\mathbf{u}[n]\right)^H\right\} = E\left\{\mathbf{u}[n]d^*[n]\right\} - E\left\{\mathbf{u}[n]\mathbf{u}^H[n]\right\}\mathbf{w}_o = \mathbf{0}$ and

$$\mathbf{R}\mathbf{w}_o = \mathbf{p}$$

where $\mathbf{p} = E\left\{\mathbf{u}[n]d^*[n]\right\}$ and $\mathbf{R} = E\left\{\mathbf{u}[n]\mathbf{u}^H[n]\right\}$. This equation is called as the <u>*normal equation*</u> or <u>*Wiener-Hopf equation*</u>.

The solution of the normal equation is

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$$

and

$$\hat{d}[n \mid U_n] = y_o[n] = \mathbf{w}_o^H\mathbf{u}[n].$$

Let $\sigma_d^2 = E\left\{\left|d[n]\right|^2\right\}$ and

$$\sigma_{\hat{d}}^2 = E\left\{\left|\hat{d}[n\,|\,U_n]\right|^2\right\} = E\left\{\left|y_o[n]\right|^2\right\} = E\left\{\mathbf{w}_o^H\mathbf{u}[n]\mathbf{u}^H[n]\mathbf{w}_o\right\} = \mathbf{w}_o^H\mathbf{R}\mathbf{w}_o = \mathbf{w}_o^H\mathbf{p} = \mathbf{p}^H\mathbf{w}_o \quad .$$

Since $d[n] = \hat{d}[n\,|\,U_n] + e_o[n]$ and $E\left\{\hat{d}[n\,|\,U_n]e_o^*[n]\right\} = 0$ from OP,

$$J_{\min} = J(\mathbf{w}_o) = E\left\{\left|e_o[n]\right|^2\right\} = \sigma_d^2 - \sigma_{\hat{d}}^2 \quad \text{and} \quad \varepsilon = \frac{J_{\min}}{\sigma_d^2} = 1 - \frac{\sigma_{\hat{d}}^2}{\sigma_d^2}\,.$$

(2) Error-performance surface in canonical coordinate

Let

$$\begin{aligned} J(\mathbf{w}) &= E\left\{\left(d[n] - \mathbf{w}^H\mathbf{u}[n]\right)\left(d[n] - \mathbf{w}^H\mathbf{u}[n]\right)^H\right\} \\ &= \sigma_d^2 - \mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + \mathbf{w}^H\mathbf{R}\mathbf{w} \\ &= \sigma_d^2 + \left(\mathbf{w} - \mathbf{w}_o\right)^H \mathbf{R}\left(\mathbf{w} - \mathbf{w}_o\right) - \mathbf{w}_o^H\mathbf{R}\mathbf{w}_o \\ &= \sigma_d^2 + \left(\mathbf{w} - \mathbf{w}_o\right)^H \mathbf{R}\left(\mathbf{w} - \mathbf{w}_o\right) - \mathbf{p}^H\mathbf{R}^{-1}\mathbf{p} \end{aligned} \quad ,$$

then $J_{\min} = J(\mathbf{w}_o) = \sigma_d^2 - \mathbf{p}^H\mathbf{R}^{-1}\mathbf{p}$. Since $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, $\mathbf{\Lambda} = diag[\lambda_1\lambda_2\cdots\lambda_M]$, and $\mathbf{Q} = [\mathbf{q}_1\mathbf{q}_2\cdots\mathbf{q}_M]$ with $\mathbf{R}\mathbf{q}_i = \lambda_i\mathbf{q}_i$,

$$\begin{aligned} J(\mathbf{w}) &= J_{\min} + \left(\mathbf{w} - \mathbf{w}_o\right)^H \mathbf{R}\left(\mathbf{w} - \mathbf{w}_o\right) \\ &= J_{\min} + \mathbf{v}^H\mathbf{\Lambda}\mathbf{v} \\ &= J_{\min} + \sum_{k=1}^{M}\lambda_k\left|v_k\right|^2 \end{aligned}$$

with the principal axis $\mathbf{v} = \mathbf{Q}^H(\mathbf{w} - \mathbf{w}_o)$.

Linear Prediction

- We assume a discrete-time WSS process.
- Consider a discrete-time time series of infinite duration;

$$\cdots u[n-M]\cdots u[0]\cdots u[n]u[n+1]\cdots.$$

(1) Forward prediction

- Let $U_{n-1}=\mathrm{span}\{u[n-k]\}_{k=1}^{M}$. Also let the <u>*predicted value*</u> of $u[n]$ and the <u>*forward prediction error*</u> be

  $\hat{u}[n|U_{n-1}]=\sum_{k=1}^{M}w_{f,k}^{*}u[n-k]$ and $f_M[n]=u[n]-\hat{u}[n|U_{n-1}]$ with the <u>*tap-weight vector*</u>

$$\mathbf{w}_f=\left[w_{f,1},w_{f,2},\cdots,w_{f,M}\right]^T=\arg\min_{\mathbf{w}_f\in\mathbf{C}^M}E\left[|f_M[n]|^2\right].$$

- Let $\mathbf{u}[n-1]=\left[u[n-1],u[n-2],\cdots,u[n-M]\right]^T$ and $d[n]=u[n]$, then

  $\hat{u}[n|U_{n-1}]=\mathbf{w}_f^H\mathbf{u}[n-1]$. Compare this with Wiener filter.

- From the <u>*normal equation*</u> or <u>*Wiener-Hopf equation*</u>, $\mathbf{R}\mathbf{w}_f=\mathbf{p}$ with

$$\mathbf{R}=E\left\{\mathbf{u}[n-1]\mathbf{u}^H[n-1]\right\}=\begin{bmatrix} r[0] & r[1] & \cdots & r[M-1] \\ r[-1] & r[0] & \cdots & r[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ r[-M+1] & r[-M+2] & \cdots & r[0]\end{bmatrix}\text{ and}$$

$$\mathbf{p}=E\left\{\mathbf{u}[n-1]d^*[n]\right\}=E\left\{\mathbf{u}[n-1]u^*[n]\right\}=\begin{bmatrix} r[-1] \\ r[-2] \\ \vdots \\ r[-M]\end{bmatrix}=\mathbf{r},$$

$$\mathbf{w}_f=\mathbf{R}^{-1}\mathbf{r}.$$

- The <u>*forward prediction error power*</u> is

$$P_M=\min E\left[|f_M[n]|^2\right]=\sigma_d^2-\mathbf{p}^H\mathbf{w}_0=r[0]-\mathbf{r}^H\mathbf{w}_f.$$

- <u>*Connection with AR process*</u>;

  FLP: $f_M[n]=u[n]-\mathbf{w}_f^H\mathbf{u}[n-1]=u[n]-\sum_{k=1}^{M}w_{f,k}^{*}u[n-k]$

AR: $u[n] = v[n] - \sum_{k=1}^{M} a_k^* u[n-k]$ where $v[n]$ is white Gaussian noise.

- In FLP, if $f_M[n]$ becomes white, the prediction using the order M is optimal.
- Augmented Wiener-Hopf equation for FLP;

$$\begin{bmatrix} r[0] & \mathbf{r}^H \\ \mathbf{r} & \mathbf{R} \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{w}_f \end{bmatrix} = \begin{bmatrix} P_M \\ 0 \end{bmatrix}.$$

(2) Backward prediction

- Let $U_n = \text{span}\{u[n-k]\}_{k=0}^{M-1}$. Also let the *predicted value* of $u[n-M]$ and the *backward prediction error* be

$$\hat{u}[n-1|U_n] = \sum_{k=1}^{M} w_{b,k}^* u[n-k+1] \quad \text{and} \quad b_M[n] = u[n-M] - \hat{u}[n-M(U_n]$$

with the *tap-weight vector* $\mathbf{w}_b = \left[w_{b,1}, w_{b,2}, \cdots, w_{b,M}\right]^T = \arg\min_{\mathbf{w}_b \in \mathbf{C}^M} E\left[|b_M[n]|^2\right]$.

- Let $\mathbf{u}[n] = [u[n], u[n-1], \cdots, u[n-M+1]]^T$ and $d[n] = u[n-M]$, then $\hat{u}[n-M|U_n] = \mathbf{w}_b^H \mathbf{u}[n]$. Compare this with Wiener filter.
- From the *normal equation* or *Wiener-Hopf equation*, $\mathbf{R}\mathbf{w}_b = \mathbf{p}$ with

$$\mathbf{R} = E\left\{\mathbf{u}[n]\mathbf{u}^H[n]\right\} = \begin{bmatrix} r[0] & r[1] & \cdots & r[M-1] \\ r[-1] & r[0] & \cdots & r[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ r[-M+1] & r[-M+2] & \cdots & r[0] \end{bmatrix} \text{ and}$$

$$\mathbf{p} = E\left\{\mathbf{u}[n]d^*[n]\right\} = E\left\{\mathbf{u}[n]u^*[n-M]\right\} = \begin{bmatrix} r[M] \\ r[M-1] \\ \vdots \\ r[1] \end{bmatrix} = \mathbf{r}^{B*},$$

$$\mathbf{w}_b = \mathbf{R}^{-1}\mathbf{r}^{B*}.$$

- The *backward prediction error power* is

$$P_M = \min E\left[|b_M[n]|^2\right] = \sigma_d^2 - \mathbf{p}^H \mathbf{w}_0 = r[0] - \mathbf{r}^{BT}\mathbf{w}_b.$$