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

  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
  
  $$r[n,n-k] = E[u[n]u^*[n-k]], k = 0, \pm 1, \pm 2, \cdots$$

  where $*$ denotes complex conjugation.
- The *autocovariance function* is
  
  $$c[n,n-k] = E[(u[n] - \mu[n])(u[n-k] - \mu[n-k])^*], k = 0, \pm 1, \pm 2, \cdots$$

  $$= r[n,n-k] - \mu[n] \mu^*[n-k].$$

- 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. $r[0] = E\left[|u[n]|^2\right]$ 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 $\forall n, E\left[|u[n]|^2\right] < \infty$, 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. 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

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

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

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

i.e., the process is asymptotically uncorrelated.

- The process is **correlation ergodic in the mean-square sense**, if

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

and

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

(3) Correlation matrix

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

$$
u[n] = \begin{bmatrix} u[n], u(n-1), \cdots, u(n-M+1) \end{bmatrix}^T.
$$

- We define the **correlation matrix** $R$ as

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

Then,
The correlation matrix has the following properties.

1. From the definition of $R$, $R = R^H$. Therefore, $r[-k] = r^*[k]$. If $u[n]$ is real $R = R^T$. Therefore, 
   $$
   R = \begin{bmatrix}
   r[0] & r[1] & \cdots & r[M-1] \\
   r[-1] & r[0] & \cdots & r[M-2] \\
   \vdots & \vdots & \ddots & \vdots \\
   r[-M+1] & r[-M+2] & \cdots & r[0]
   \end{bmatrix}.
   $$

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

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

4. Let $u^H[n] = [u[n-M+1], u[n-M+1], \ldots, u[n]]^T$, then 
   $$
   E[u^H[n]u^{HH}[n]] = R^T.
   $$

5. $R_{M+1} = \begin{bmatrix}
   r[0] \\
   r \\
   R_M
   \end{bmatrix} = \begin{bmatrix}
   R_M & r^{BT} \\
   r^{H} & r[0]
   \end{bmatrix}$ with $r^H = [r[1], r[2], \ldots, r[M]]$ and $r^{BT} = [r[-M], r[-M+1], \ldots, r[-1]]$. 
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 $E[v[n]] = 0$ for all $n$, $r[0] = \sigma_v^2$, and $r[k] = 0$ for all $k \neq 0$.

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

(2) **Moving average (MA) model**
- Time series $u[n]$, $u[n-1]$, ..., $u[n-M]$ represents the realization of an MA of order $N$ if

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

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

(3) **Autoregressive-moving average (ARMA) model**
- Time series $u[n]$, $u[n-1]$, ..., $u[n-M]$ represents the realization of an ARMA of order $(M, N)$ if

$$u[n] + \sum_{k=1}^{M} a_k^* u[n-k] = \sum_{k=0}^{N} b_k^* v[n-k] \quad \text{with} \quad b_0 = 1.$$
- \[ H_{ARMA}(z) = \frac{U(z)}{V(z)} = \frac{\sum_{k=0}^{N} b_k z^{-k}}{\sum_{k=0}^{M} a_k z^{-k}} \]. 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:

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

with white-noise process \( v[n] \) such that \( E[v[n]s'[k]] = 0 \) 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 \( \{a_1, a_2, \ldots, a_M\} \) and
2. The variance \( \sigma_v^2 \) 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,

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

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

For \( l = 1, 2, \ldots, M \), we have

\[
\begin{align*}
l = 1 & \implies a_0^* r[1] + a_1^* r[0] + a_2^* r[-1] + \cdots + a_M^* r[1-M] = 0 \\
l = 2 & \implies a_0^* r[2] + a_1^* r[1] + a_2^* r[0] + \cdots + a_M^* r[2-M] = 0 \\
& \vdots \\
l = M & \implies a_0^* r[M] + a_1^* r[M-1] + a_2^* r[M-2] + \cdots + a_M^* r[0] = 0
\end{align*}
\]
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,

\[
Rw = r \quad \text{or} \quad w = R^{-1}r.
\]

- Note that \( E[v[n]u^*[n]] = E[v[n]v^*[n]] = \sigma_v^2 \) 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
  \[ u_N[n] = \begin{cases} u[n], & n = 0,1,\ldots,N-1 \\ 0, & \text{otherwise} \end{cases} \]
  Then,
  \[ U_N(\omega) = \sum_{n=0}^{N-1} u_N[n] e^{-j\omega n}, \quad U_N^*(\omega) = \sum_{k=0}^{N-1} u_N^*[k] e^{j\omega k}, \]
  and
  \[ \|U_N(\omega)\|^2 = \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} u_N[n] u_N^*[k] e^{-j\omega(n-k)}. \]

- We take expected values of both sides as
  \[ E\left[\|U_N(\omega)\|^2\right] = \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} E\left[u_N[n] u_N^*[k]\right] e^{-j\omega(n-k)} = \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} r_N[n-k] e^{-j\omega(n-k)}. \]

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

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

  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
  \[ P_N(\omega) = \frac{1}{N} \|U_N(\omega)\|^2. \]

- We define the **power spectral density** of a WSS discrete-time stochastic process as
  \[ S(\omega) = \lim_{N \to \infty} E\left[P_N(\omega)\right] = \sum_{l=-\infty}^{\infty} r[l] e^{-j\omega l} \text{ with } r[l] = E\left[u[n] u^*[n-l]\right]. \]

(2) **Properties of PSD**
- $S(\omega) = \sum_{l=-\infty}^{\infty} r[l]e^{-j\omega l}$, $-\pi < \omega \leq \pi \leftrightarrow r[l] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega)e^{j\omega l}d\omega$, $l = 0, \pm 1, \pm 2, \cdots$
- $S(\omega + 2\pi k) = S(\omega)$ for integer $k$. $-\pi < \omega \leq \pi$ 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 $\Omega$ resistor
- $S(\omega) \geq 0$ for all $\omega$

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

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

WSS process with $S(\omega)$ → Discrete-time Linear Filter → WSS process with $S_o(\omega)$
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 $Rq = \lambda q$, $\lambda$ 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, $\det(R - \lambda I) = 0$ provide all $\lambda_i$ and $q_i$.

(2) Properties of eigenvalues and eigenvectors of $R$ from WSS process
- $R^k q = \lambda^k q$
- $\{\lambda_i\}_{i=1}^M$ are all real and nonnegative.
- If $\{\lambda_i\}_{i=1}^M$ are distinct, $\{q_i\}_{i=1}^M$ are linearly independent.
- If $\{\lambda_i\}_{i=1}^M$ are distinct, $\{q_i\}_{i=1}^M$ are orthogonal. $\{q_i\}_{i=1}^M$ is an orthogonal basis of $\text{span}(R)$.
- If $\{\lambda_i\}_{i=1}^M$ are distinct and $\{q_i\}_{i=1}^M$ are normalized so that $q_j^H q_j = \delta_{jj}$,

  $$Q^H R Q = \Lambda$$

  (unitary similarity transform)

  with $Q = [q_1, q_2, \ldots, q_M]$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$. Since $RQ = QA$, $Q^H Q = I$ or $Q^{-1} = Q^H$ (unitary matrix),

  $$R = QAQ^H = \sum_{i=1}^M \lambda_i q_i q_i^H$$

  (spectral theorem or Mercer's theorem).

- $\text{tr}(Q^H R Q) = \text{tr}(RQQ^H) = \text{tr}(R) = \text{tr}(\Lambda) = \sum_{i=1}^M \lambda_i$.

- The condition number of $R$ is $\chi(R) = \|R\| \|R^{-1}\| = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}}$. $R$ is ill-conditioned if $\chi(R)$ is large.
- $S_{\text{min}} \leq \lambda_i \leq S_{\text{max}}$ where $S_{\text{min}}$ and $S_{\text{max}}$ are the minimum and maximum of the power spectral density of the process. Therefore, $\chi(R) = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \leq \frac{S_{\text{max}}}{S_{\text{min}}}$. 

• If $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$, $\lambda_k = \min_{\text{dim}(T) = k} \max_{x \neq 0, x^T x = 1} \frac{x^T R x}{x^T x}$, $k = 1, 2, \ldots, M$.

• *(Karhunen-Loeve expansion)* $u[n] = \sum_{i=1}^{M} c_i[n] q_i$ with $c_i[n] = q_i^T u[n]$ for $i = 1, 2, \ldots, M$.

  Note

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

  Also,

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

(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 $u[n]$ by

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

  - $\text{span}\{q_i\}_{i=1}^{p}$ is a *feature space* whereas $\text{span}\{q_i\}_{i=1}^{M}$ is a *data space*.

  - The *reconstruction error vector* is

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

  - The *mean-square error* is

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

(4) Eigenfilter

- We want to maximize the output signal-to-noise ratio (SNR).

- Eigendecomposition of the correlation matrix $R$ is performed.

- The eigenvector $q_{\text{max}}$ 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 $w = [w_0, w_1, \ldots, w_{M-1}]^T \in \mathbb{C}^M$ and $u[n] = [u[n]u[n-1] \cdots u[n-M+1]]^T$. Let $y[n] = w^H u[n]$ and $e[n] = d[n] - y[n]$. Define

$$J(w) = E\{e[n]^2\} = E\{|d[n] - w^H u[n]|^2\}$$

and $w_o = \arg \min \limits_{w \in \mathbb{C}^M} J(w)$.

From the orthogonality principle (OP),

$$E\{u[n]e_o^*[n]\} = 0 \quad \text{where} \quad e_o[n] = d[n] - w_o^H u[n].$$

Therefore,

$$E\{u[n](d[n] - w_o^H u[n])^H\} = E\{u[n]d^*[n]\} - E\{u[n]u^H[n]\} w_o = 0 \quad \text{and}$$

$$Rw_o = p$$

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

The solution of the normal equation is

$$w_o = R^{-1}p$$

and

$$\hat{d}[n | U_o] = y_o[n] = w_o^H u[n].$$

Let $\sigma_d^2 = E\{|d[n]|^2\}$ and
\[ \sigma_d^2 = E \left\{ |d[n|U_n]|^2 \right\} = E \left\{ |y_o[n]|^2 \right\} = E \left\{ w_o^H u[n] u^H[n] w_o \right\} = w_o^H R w_o = w_o^H p = p^H w_o \]

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,

\[ J_{\text{min}} = J(w_o) = E \left\{ |e_o[n]|^2 \right\} = \sigma_d^2 - \sigma_d^2 \quad \text{and} \quad \varepsilon = \frac{J_{\text{min}}}{\sigma_d^2} = 1 - \frac{\sigma_d^2}{\sigma_d^2}. \]

(2) Error-performance surface in canonical coordinate

Let

\[ J(w) = E \left\{ (d[n] - w^H u[n])(d[n] - w^H u[n])^H \right\} \]

\[ = \sigma_d^2 - p^H w - w^H p + w^H R w \]

\[ = \sigma_d^2 + (w - w_o)^H R (w - w_o) - w_o^H R w_o \]

\[ = \sigma_d^2 + (w - w_o)^H R (w - w_o) - p^H R^{-1} p \]

then \( J_{\text{min}} = J(w_o) = \sigma_d^2 - p^H R^{-1} p \). Since \( R = QA\Lambda Q^H \), \( \Lambda = \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_M] \), and \( Q = [q_1, q_2, \ldots, q_M] \) with \( R q_i = \lambda_i q_i \),

\[ J(w) = J_{\text{min}} + (w - w_o)^H R (w - w_o) \]

\[ = J_{\text{min}} + \nabla^H \Lambda \nabla \]

\[ = J_{\text{min}} + \sum_{k=1}^{M} \lambda_k |v_k|^2 \]

with the principal axis \( v = Q^H (w - w_o) \).
Linear Prediction

- We assume a discrete-time WSS process.
- Consider a discrete-time time series of infinite duration;
  \[ \ldots u[n-M]\ldots u[0]\ldots u[n]u[n+1]\ldots. \]

(1) Forward prediction

- Let \( U_{n-1} = \text{span}\{u[n-k]\}_{k=1}^{M} \). Also let the \text{predicted value} of \( u[n] \) and the \text{forward prediction error} be
  \[ \hat{u}[n|U_{n-1}] = \sum_{k=1}^{M} w_{f,k}^* u[n-k] \quad \text{and} \quad f_M[n] = u[n] - \hat{u}[n|U_{n-1}] \]  
  with the \text{tap-weight vector} \( w_f = [w_{f,1}, w_{f,2}, \ldots, w_{f,M}]^T = \text{arg min}_{w_f} E\left[|f_M[n]|^2\right] \).

- Let \( u[n-1] = [u[n-1], u[n-2], \ldots, u[n-M]]^T \) and \( d[n] = u[n] \), then
  \[ \hat{u}[n|U_{n-1}] = w_f^H u[n-1]. \] Compare this with Wiener filter.

- From the \text{normal equation or Wiener-Hopf equation}, \( R w_f = p \) with
  \[ R = E\{u[n-1]u^H[n-1]\} = \begin{bmatrix} r[0] & r[1] & \cdots & r[M-1] \\ r[-1] & r[0] & \cdots & r[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ r[-M+1] & r[-M+2] & \cdots & r[0] \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} r[-1] \\ r[-2] \\ \vdots \\ r[-M] \end{bmatrix} = r, \]
  \[ p = E\{u[n-1]d^*[n]\} = E\{u[n-1]u^*[n]\} = \begin{bmatrix} r[0] \\ r[1] \\ \vdots \\ r[M] \end{bmatrix} \]
  \[ w_f = R^{-1} r. \]

- The \text{forward prediction error power} is
  \[ P_M = \text{min} E\left[|f_M[n]|^2\right] = \sigma_d^2 - p^H w_0 = r[0] - p^H w_f. \]

- Connection with AR process;
  \[ f_M[n] = u[n] - w_f^H u[n-1] = u[n] - \sum_{k=1}^{M} w_{f,k}^* u[n-k] \]

13
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;

\[
\begin{bmatrix}
  r[0] & r'[n] \\
  r' & R
\end{bmatrix}
\begin{bmatrix}
  1 \\
  -w_f
\end{bmatrix}
= \begin{bmatrix}
P_M \\
0
\end{bmatrix}.
\]

(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 \( \hat{u}[n-1|U_n] = \sum_{k=1}^{M} w_{b,k} u[n-k+1] \) and \( b_m[n] = u[n-M] - \hat{u}[n-M|U_n] \).

with the tap-weight vector \( w_b = \left[w_{b,1}, w_{b,2}, \ldots, w_{b,M}\right]^T = \text{arg min}_{w_b \in \mathbb{C}^b} E\left[|b_m[n]|^2\right] \).

- Let \( u[n] = [u[n], u[n-1], \ldots, u[n-M+1]]^T \) and \( d[n] = u[n-M] \), then \( \hat{u}[n-M|U_n] = w_b^H u[n] \). Compare this with Wiener filter.

- From the normal equation or Wiener-Hopf equation, \( R w_b = p \) with

\[
R = E\left[u[n]u'[n]\right] = \begin{bmatrix}
  r[0] & r[1] & \cdots & r[M-1] \\
  r[-1] & r[0] & \cdots & r[M-2] \\
  \vdots & \vdots & \ddots & \vdots \\
  r[-M+1] & r[-M+2] & \cdots & r[0]
\end{bmatrix}
\]

and

\[
p = E\left[u[n]d'[n]\right] = E\left[u[n]u'[n-M]\right] = \begin{bmatrix}
  r[M] \\
  r[M-1] \\
  \vdots \\
  r[1]
\end{bmatrix} = r^{BR}.
\]

\[
w_b = R^{-1} r^{BR}.
\]

- The backward prediction error power is

\[
P_M = \min E\left[|b_m[n]|^2\right] = \sigma_d^2 - p^{H} w_0 = r[0] - r^{BR} w_b.
\]