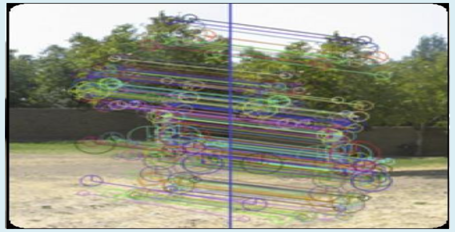
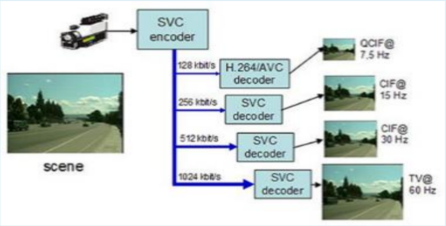
Adaptive System and Signal Processing Theory
(#7: Stationary Process and Models/Wiener Filters)

2023 Spring

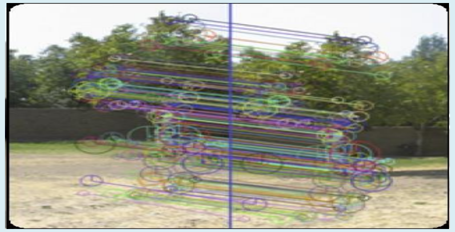
Prof. Byung-Gyu Kim
Intelligent Vision Processing Lab. (IVPL)
http://ivpl.sookmyung.ac.kr
Dept. of IT Engineering, Sookmyung Women's University
E-mail: bg.kim@sookmyung.ac.kr

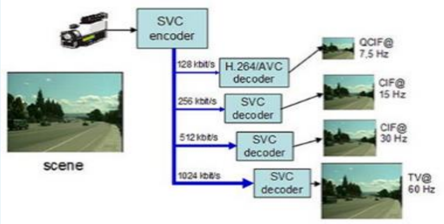




Contents

- **Summary of the Previous Lecture**
- **Stochastic Models**
  - **Yule-Walker Equation**
- **Wiener Filter**





Contents

- **Summary of the Previous Lecture**
- Stochastic Models
  - Yule-Walker Equation
- Wiener Filter

Remind (1): Stationary Process and Models

- Stochastic (Random) Process
  - A function of time, defined on some observation interval (sampling period (T) in the discrete-time).
- Strictly Stationary Process
- Partial Characterization of a Discrete-time Stochastic Process

For a time series $u[n],\ u[n-1],\ u[n-2],\cdots$

- Mean-value of function of the process

$$\mu(n) = E[u(n)] \text{ for } k = 0, \pm 1, \pm 2, \ldots.$$

where E() is the statistical expectation operator.

- Autocorrelation function

$$r(n, n-k) = E[u(n)u^*(n-k)] \text{ for } k = 0, \pm 1, \pm 2, \ldots.$$

- Auto-covariance function

$$c(n, n-k) = E[(u(n)-\mu(n))(u(n)-\mu(n))^*] \text{ for } k = 0, \pm 1, \pm 2, \ldots.$$

Remind (2): Stationary Process and Models

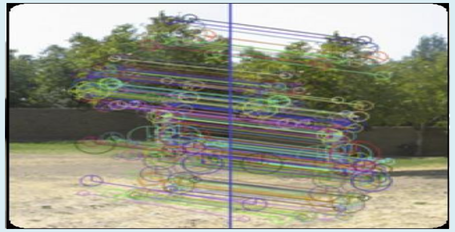
- Stationary process in Wide Sense
  - Condition: For $u[n]$,

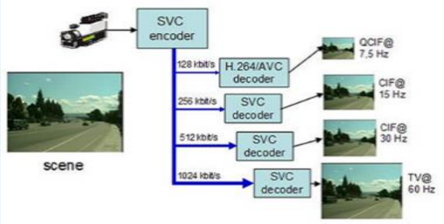
$$E[|u(n)|^2] < \infty \text{ for all } n.$$

- Mean Ergodic (in the sense of MSE)
  - If the mean square value of error between μ and $\widehat{\mu}$ as $N \to \infty$.

$$\lim_{N\to\infty} [(\mu - \widehat{\mu})^2] = 0.$$

- Correlation Ergodic
- Correlation Matrix of $\mathbf{u}(n) = [u[n],\ u[n-1],\ u[n-2], \cdots, u[n-(M-1)]]^T.$
- Properties of Correlation Matrix $\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^H(n)]$
- How to estimate the parameters of sinusoid signal in the presence of additive noise.
- Yule-Walker Equation
  - A system (set) of linear equations in AR model





Contents

- Summary of the Previous Lecture
- Stochastic Models
    - Yule-Walker Equation
- Wiener Filter

Stationary processes and Models (1): Stochastic Models (1)

- What is Model?
  - Any hypothesis that may be applied to explain or describe the hidden laws that are supposed to govern or constrain the generation of physical data of interest. ➔ Yule (1927).

Purely random Process $\nu[n]$ → **Discrete-time Linear filter** → **Discrete-time stochastic process** $u[n]$

White Gaussian noise | *Highly correlated observation*

  - In general, input-output relation for the stochastic model is linearly as

$$\boxed{\text{Present value of model output}} + \boxed{\text{Linear combination of past value of model output}} = \boxed{\text{Linear combination of present/past value of model input}}$$

IVPL Intelligent Vision Processing Lab

- Three Popular Types of Stochastic Models
  - AR (autoregressive) : Not use past input of the model.
  - MA (moving average) : Not use past output of the model.
  - ARMA (autoregressive-moving average): both are available.

- AR Models

For a time series $u[n],\ u[n-1],\ u[n-2], \cdots, u[n-M]$,

Order of AR process

$$u[n] + \underbrace{a_1^* u[n-1] + a_2^* u[n-2] + \cdots + a_M^*}_{\text{AR parameters}} u[n-M] = \nu[n].$$

AR parameters

In terms of $u[n]$,

$$u[n] = \underbrace{w_1^* u[n-1] + w_2^* u[n-2] + \cdots + w_M^* u[n-M]}_{\text{Linear combination of the past values}} + \nu[n].$$

Linear combination of the past values

Small error term

- For Usual Linear Regression Model, we can express as

$$y = \sum_{k=1}^{M} w_k^* x[k] + \nu.$$

Then

$$u[n] = \sum_{k=1}^{M} w_k^* u[n-k] + \nu[n].$$

  - $u[n]$ is regressed on the previous value of itself : "Autoregressive"

- Convolution Form of input sequence $u[n]$ and seq. of a_n^* like as,

$$\sum_{k=0}^{M} a_k^* u[n-k] = \nu[n].$$

By taking z-transform on both sides,

$$H_A(z)U(z) = V(z).$$

- AR process u[n] is
  - Input: noise $\nu[n]$ is output. Transfer function $H_A(z) = \frac{V(z)}{U(z)}.$ (finite duration)
  - Output: noise $\nu[n]$ is input. Transfer function $H_A(z) = \frac{U(z)}{V(z)}.$ (infinite duration)

AR analyzer ***AR generator***

- MA Models
  - For the given input $\nu[n]$, the output $u[n]$ is described by the difference equation:

Order of MA process

$$u[n] = b_1^*\nu[n-1]+b_2^*\nu[n-2]+\cdots+b_K^*\nu[n-K]+\nu[n].$$

MA parameters

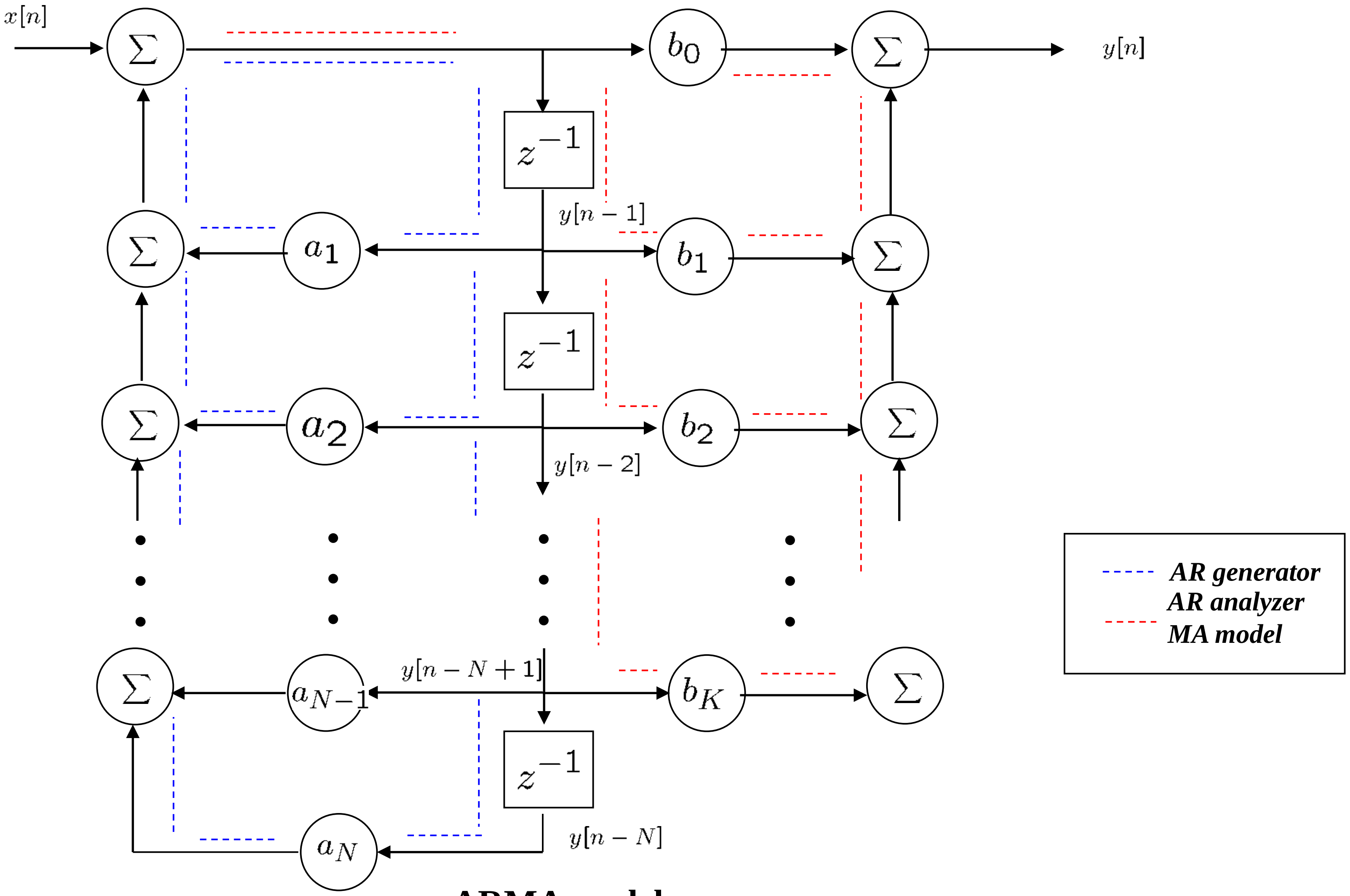
    - Output $u[n]$ is a form of weighted average with $\nu[n]$.

- ARMA Models
  - Has transfer function that contains both poles and zeros and model parameters of order (M, K).

$$u[n]+a_1^*u[n-1]+a_2^*u[n-2]+\cdots+a_M^*u[n-M] =$$

$$b_1^*\nu[n-1]+b_2^*\nu[n-2]+\cdots+b_K^*\nu[n-K]+\nu[n].$$

IVPL Intelligent Vision Processing Lab



$x[n]$
Σ
b_0
$y[n]$
z^{-1}
$y[n-1]$
a_1
b_1
z^{-1}
a_2
b_2
$y[n-2]$
$y[n-N+1]$
a_{N-1}
b_K
z^{-1}
$y[n-N]$
a_N
AR generator
AR analyzer
MA model

ARMA model

- Considerations
  - In terms of computation, AR model is usually better easy.
    - A system (set) of linear equations → Yule-Walker equations
    - ARMA/MA model: very complex to solve and so many nonlinear cases.

- Yule-Walker
  - How to define AR Process Model?
    - AR coefficients $\{a_k^*\}$
    - The variance of the white noise $\nu[n]$

For any AR process $u[n]$,

$$\sum_{k=0}^{M} a_k^* u[n-k] = \nu[n].$$

Stationary processes and Models (8): Stochastic Models (8) : Yule-Walker Eq.

By multiplying $u^*[n-l]$ and taking expectation operator (E(-)),

$$E[\sum_{k=0}^{M} a_k^* u[n-k]u*[n-l]] = \cancelto{0}{E[\nu[n]u^*[n-l]]} \text{ for } l > 0.$$

Then,

$$\boxed{E[}\sum_{k=0}^{M} a_k^* \boxed{u[n-k]u^*[n-l]]} = 0.$$

$$\sum_{k=0}^{M} a_k^* \boxed{r(l-k)} = 0 \quad \text{where} \quad a_0 = 1.$$

And

$$a_0^* r(l) + a_1^* r(l-1) + a_2^* r(l-2) + \cdots + a_M^* r(l-M) = 0,$$

$$\therefore \; r(l) = -a_1^* r(l-1) - a_2^* r(l-2) - \cdots - a_M^* r(l-M).$$

Here, let's evaluate it for each *l*, then we have the following matrix form that has been as so famous **Yule-Walker Equation**:

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

$$\mathbf{R}\mathbf{w} = \mathbf{r}$$

So we can solve this as if $\mathbf{R}$ is nonsingular,

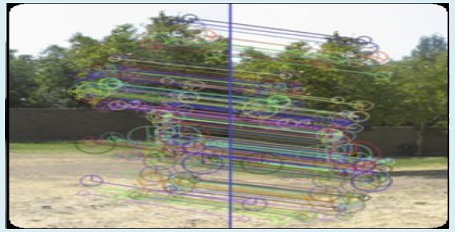
$$\underline{\mathbf{w}} = \mathbf{R}^{-1}\mathbf{r}$$

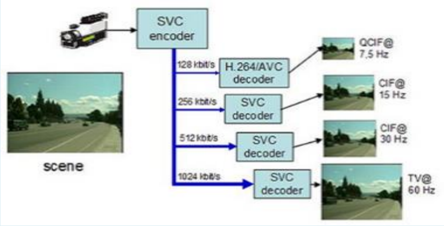
AR coefficients

- The variance of the white noise:

Let l = 0, then $$E[\nu[n]u^*[n]] = E[\sum_{k=0}^{M} a_k^* u[n-k]u*[n]]$$

$$\therefore \quad \sum_{k=0}^{M} a_k^* r(k) = \sigma_\nu^2.$$

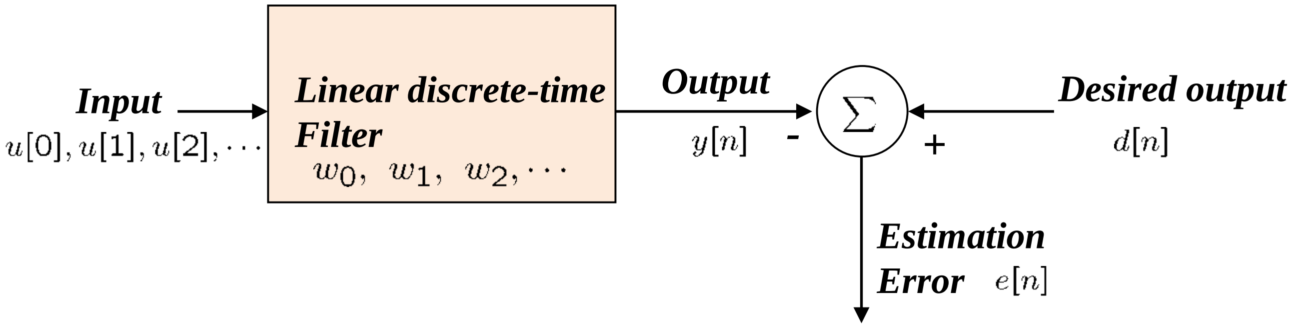




Contents

- Summary of the Previous Lecture
- Stochastic Models
    - Yule-Walker Equation
- **Wiener Filter**

- Linear Optimum Filtering: Problem Statement
  - Filtering problem can be formulated as the general case of the complex-valued time series in terms of filter's impulse response.
    - Input: $u[0], u[1], u[2], \cdots$
    - Filter: weight (coeff.) as impulse response
    - Desired response $d[n]$: signal of interest.
    - Output: **estimate** of the desired output
    - Error $e[n]$: the difference bet. the desired response and the filter output.



  - Goal : **How to set or control weights of the filter** to give the desired response?

Wiener Filters: Linear Optimum Filters (2)

- Main requirement:
  - Make $e[n]$ as smaller as possible in some statistical sense.
  - Two restrictions:
    - The filter is linear
    - The filter output is discrete-time for implementing on the digital HW.
- Factors for Filter Design
  - Whether the impulse response of the filter has finite or infinite duration.
  - What kind of statistical criterion used for the optimization.

Cost function or index of performance

- What kinds of Cost Functions:
  - Mean-square value of the estimation error.
  - Expectation of the absolute value of the estimation error.
  - Expectation of the third or higher order of ~ .

- How to Solve Mathematically the statistical Optimization Problem:
  - Principle of orthogonality
  - Error-performance surface on the filter's coefficients

- Principle of Orthogonality

  For any time series $u[n],\ u[n-1],\ u[n-2],\cdots$

  - Assumption:
    - Filter input and the desired response are jointly wide-sense stationary stochastic process.
  - Filter output : $y = \sum_{k=1}^{\infty} w_k^* u[n-k],\ k = 0, 1, 2, \cdots .$
  - Estimation error $e[n] = d[n] - y[n].$

    To optimize the filter design, we choose to minimize the mean-square value of the estimation error as cost function. Then

Wiener Filters: Linear Optimum Filters (4)-Orthogonality

- Cost function:

$$J = E[e[n]e^*[n]] = E[|e^2[n]|]$$

Finally, the problem is to determine the operating conditions for which ***J*** attains its minimum value.

$\Rightarrow$ minimize $J\,|_{\textit{Filter operating condition } w_k}$

- How to obtain the optimum operating condition of the filter?
  - Linear Filter is composed of tap-weights ($w_0, w_1, w_2, \cdots$).
  - Gradient operation:
    - Partial derivative on the specified variable.
    - Used in the context of finding the stationary points of a function of interest.

- Simple examples of gradient operation

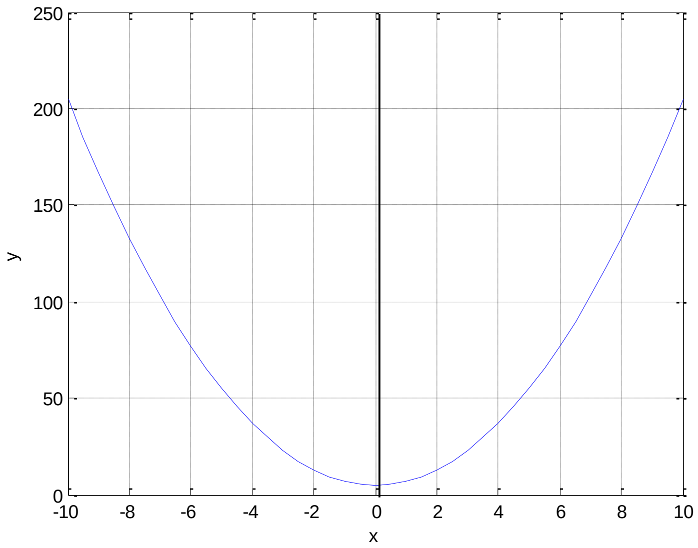
$$y = f(x) = 2x^2 + 5$$

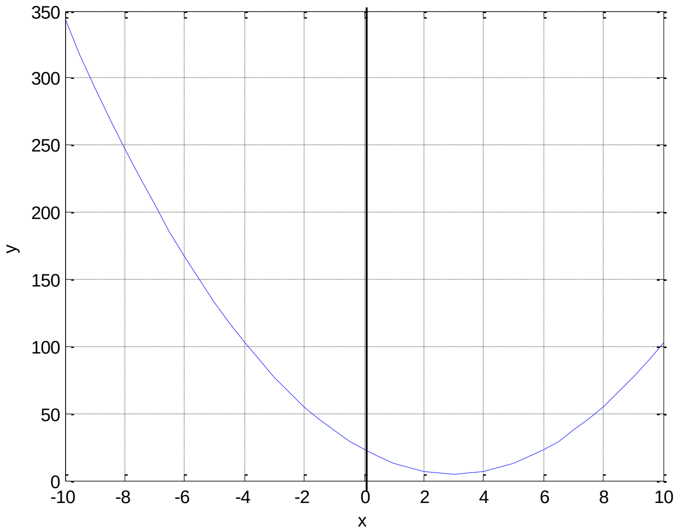
To find minimum point, we take partial derivative of $f(x)$ on the variable x and set to zero.

$$\frac{\partial f(x)}{\partial x} = 0$$

If we solve the above Eq., then we can find the optimum condition of x.

$$y = f(x) = 2(x-3)^2 + 5$$





Wiener Filters: Linear Optimum Filters (6)-Orthogonality

- General definition of gradient operation

Since w_k is complex value like $w_k = a_k + jb_k, \;\; k = 0, 1, 2, \cdots$

We can define the gradient operator as the following:

$$\nabla_k = \frac{\partial}{\partial a_k} + j\frac{\partial}{\partial b_k}, \;\; k = 0, 1, 2, \cdots$$

Let's apply this to the cost function,

$$\nabla_k J = \frac{\partial J}{\partial a_k} + j\frac{\partial J}{\partial b_k}, \;\; k = 0, 1, 2, \cdots .$$

Similarly, we can find the optimum condition of J on the filter w_k as the result of the gradient set to zero.

$$\nabla_k J = \frac{\partial J}{\partial a_k} + j\frac{\partial J}{\partial b_k} = 0, \;\; k = 0, 1, 2, \cdots .$$

Optimum in the MSE sense

IVPL Intelligent Vision Processing Lab

Therefore,

$$\nabla_k J = E\left[\frac{\partial e[n]}{\partial a_k}e^*[n]+\frac{\partial e^*[n]}{\partial a_k}e[n]+j\frac{\partial e[n]}{\partial b_k}e^*[n]+\frac{\partial e^*[n]}{\partial b_k}e[n]\right] = 0,\ k = 0,1,2,\cdots.$$

because $J = E[e[n]e^*[n]]$

Then

$$\frac{\partial e[n]}{\partial a_k} = -u[n-k]$$

$$\frac{\partial e^*[n]}{\partial a_k} = -u*[n-k]$$

$$\frac{\partial e[n]}{\partial b_k} = ju[n-k]$$

$$\frac{\partial e^*[n]}{\partial b_k} = -ju*[n-k]$$

$$\therefore\ \nabla_k J = -2E\left[u[n-k]e^*[n]\right] = 0.$$

$\therefore\ E\left[u[n-k]e^*[n]\right] = 0,\ k = 0,1,\cdots$, in the optimum condition.

Let $e_o[n]$ denote the special value of the estimation error that results when the filter operates in its optimum condition. Then

$$\therefore \; E\Big[u[n-k]e_o^*[n]\Big] = 0, \;\; k = 0, 1, \cdots .$$

- Principle of orthogonality
  - The estimation error $e_o[n]$ is orthogonal to each input sample that enters into the estimation of the desired response at time n.

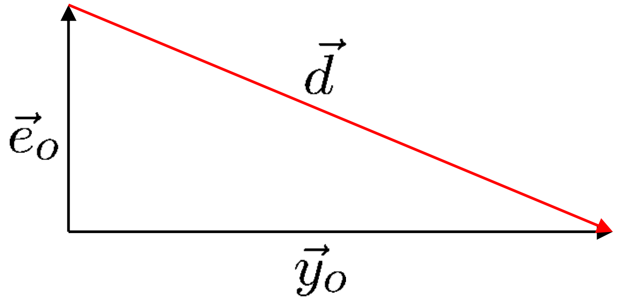
This is basic procedure for testing whether linear filter is operating in its optimum condition.

- Corollary to the Principle of Orthogonality
  - If we check on $E\Big[y[n]e^*[n]\Big]$,

$$E\Big[y[n]e^*[n]\Big] = E\Big[\sum_{k=0}^{\infty} w_k^* u[n-k]e^*[n]\Big]$$
$$= \sum_{k=0}^{\infty} w_k^* E\Big[u[n-k]e^*[n]\Big].$$

If the filter is in the optimum condition,

$$\therefore \; E\Big[y_o[n]e_o^*[n]\Big] = 0.$$



Geometric representation of the relationship

- Minimum Mean-Square Error (MSE)

  If the filter is in the optimum condition,

$$e_o[n] = d[n] - y_o[n],$$
$$= d[n] - \hat{d}[n|\mathcal{U}_n].$$

Then $$d[n] = e_o[n] + \hat{d}[n|\mathcal{U}_n].$$ ← ***Since*** $J = E\left[|e_o[n]|^2\right]$**,**

$$\sigma_d^2 = \sigma_{\hat{d}}^2 + J_{min}.$$ ← ***Both of these parameters are assumed to be of zero mean***

$$\therefore\ J_{min} = \sigma_d^2 - \sigma_{\hat{d}}^2 \text{ in MSE sense.}$$

- Normalized MSE: $\frac{J_{min}}{\sigma_d^2} = 1 - \frac{\sigma_{\hat{d}}^2}{\sigma_d^2} = \varepsilon.$

- Wiener-Hopf Equations

If the filter is in the optimum condition,

$$E\left[u[n-k]e_o^*[n]\right] = 0, \; k = 0, 1, \cdots .$$

$$e_o[n] = d[n] - y_o[n], \quad = d[n] - \sum_{i=0}^{\infty} w_{oi} u[n-i].$$

Then

$$E\left[\boxed{u[n-k]}\left(\boxed{d[n]} - \boxed{\sum_{i=0}^{\infty} w_{oi} u[n-i]}\right)\right] = 0.$$

Cross correlation function *Autocorrelation function*

- Auto correlation function: $E\left[u[n-k]u^*[n-i]\right] = r(i-k), \; k = 0, 1, \cdots .$
- Cross correlation function: $E\left[u[n-k]d^*[n]\right] = p(-k), \; k = 0, 1, \cdots .$

$$\therefore \; \underline{\sum_{i=0}^{\infty} w_{oi} r(i-k) = p(-k), \; \text{for } k = 0, 1, \cdots .}$$

Wiener-Hopf Equation

In Next Class?

- We will talk about the remained part of Wiener Filters and Linear Prediction.
  - Solution of Wiener-Hopf filter.
  - Error-performance surface.
  - Introduction of the linear prediction.

HW#4

- Solve the following Problems:
  - (Chap. 5) P. 2, P. 3 , P. 4,

- Due date: ~ to the next week.

Thank you for your attention.!!!
QnA

http://ivpl.sookmyung.ac.kr