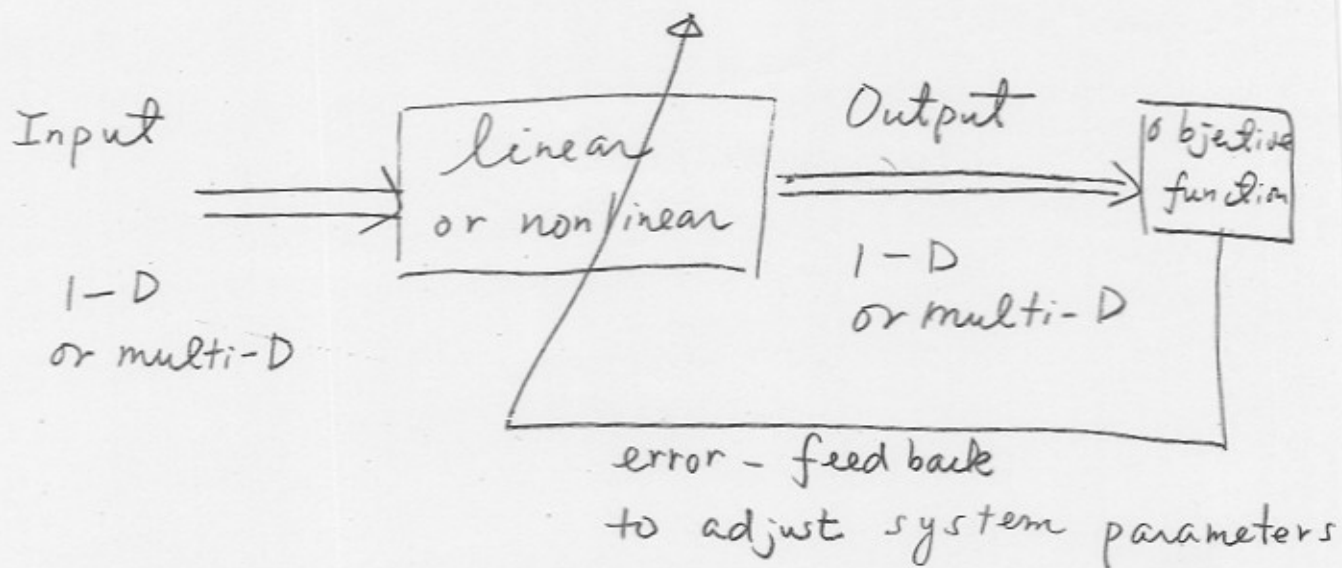


Adaptive Filter Theory

General Adaptive Filtering System

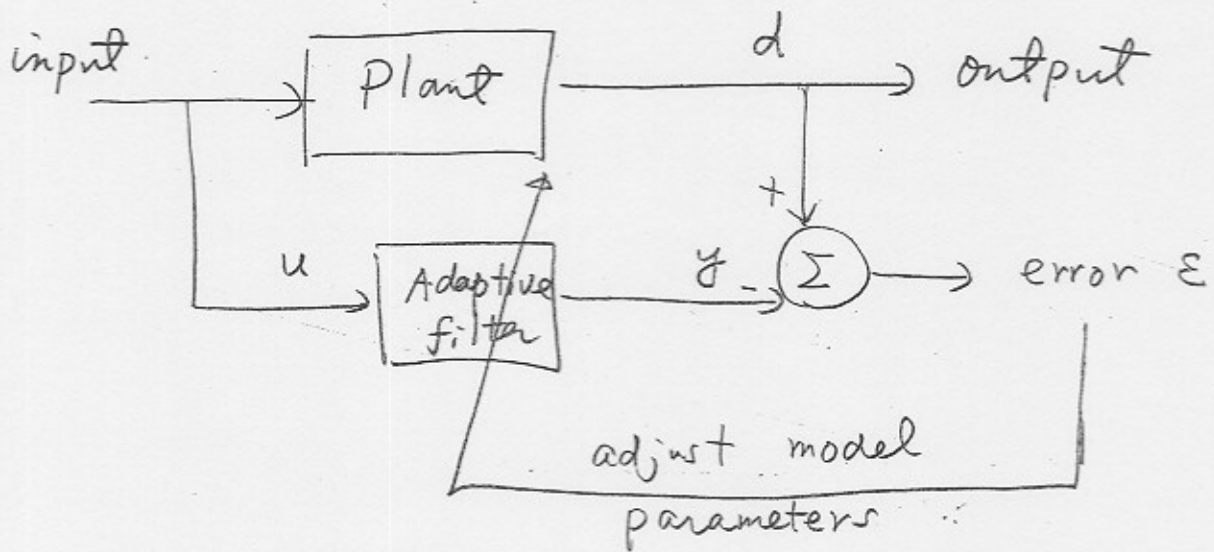


In this course, we focus on the
Linear Time-invariant System (LTI)
model.

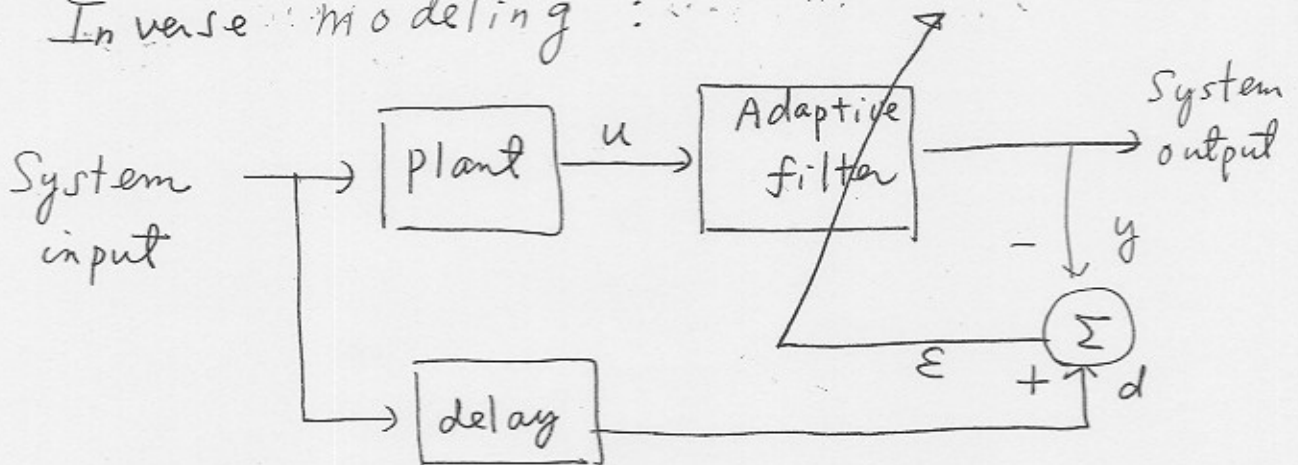
Reference: Probability, Random Variables
and Stochastic Processes
A. Papoulis and S. U. Pillai
4th ed. McGraw-Hill, 2002

Application of Adaptive Filters

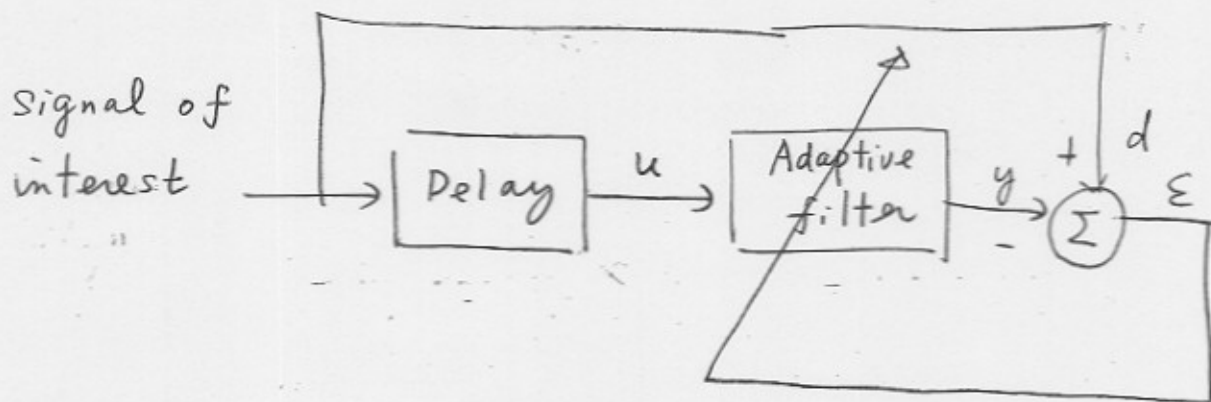
1. Identification :



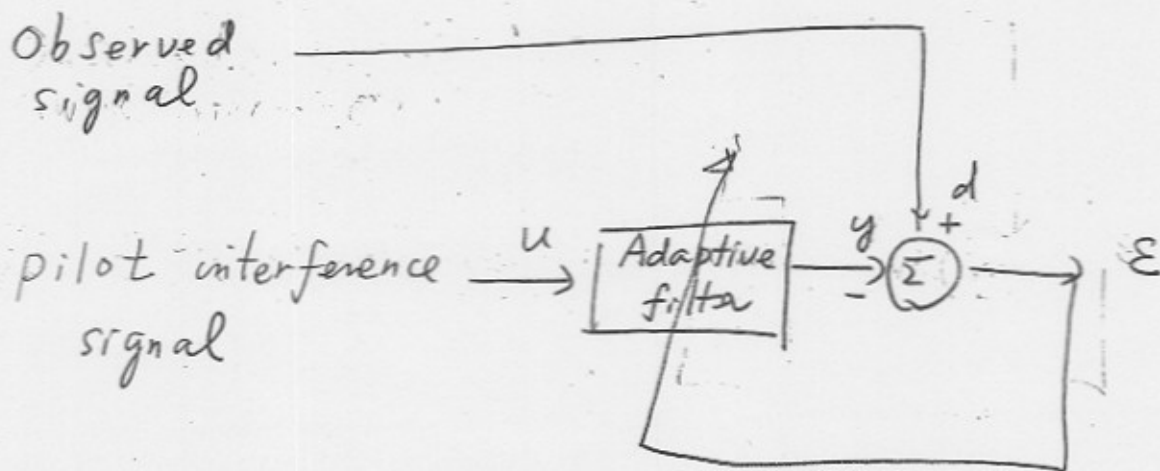
2. Inverse modeling :



3. Prediction :



4. Interference cancellation :



Reference : Signals and Systems, A.V.
Oppenheim et al, 2nd ed.
Prentice-Hall

Chapter 1. Stochastic Processes and Models

1-1

* Statistical expectation: For a random process X

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

where $f_X(x)$ is the probability density function of random variable X .

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

where $g(x)$ is an arbitrary function of X .

* i.i.d. process: an random process $X(n)$ is called statistically independent, identical distributed (i.i.d.) in time

if and only if $f_{X_n}(x(n)) = f_{X_m}(x(m))$

$\forall n, m$ and m, n are arbitrary

discrete-time indices, $m, n \in \mathbb{Z}$

The i.i.d. assumption is generally made in the Adaptive Filter Theory.

Hence, we can discuss the statistical expectation on any arbitrary sample $x(n)$, such as

$$\mu(n) = E\{x(n)\} = \mu$$

$$r(n, n-k) = E\{x(n)x^*(n-k)\}, \quad k=0, \pm 1, \pm 2, \dots$$
$$= r(k)$$

$$c(n, n-k) = r(n, n-k) - \mu(n)\mu^*(n-k)$$
$$= r(k) - |\mu|^2$$

Since $x(n)$ is also a stationary process.

1.2

The statistical expectation operator has to be taken through many many trials, which are impossible to be obtained in reality.

Therefore, the random process $x(n)$ can be assumed to be ERGODIC, and the statistical expectation $E\{\cdot\}$ can be

estimated through a time average, i.e.,

$$E\{g(x(n))\} = \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} g(x(n))}_{\text{estimator}}$$

Example :

mean estimate: $\hat{\mu}(N) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)$

where $g(x(n)) = x(n)$

energy estimate: $\hat{e}(N) = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n)$

where $g(x(n)) = x^2(n)$

How good is the time-average estimator (the rule of the estimate)?

The robustness of an estimator is defined as the mean-square error between the statistical expectation of the true value and the estimate.

If the statistical expectation of the estimator is equivalent to the true value, such an estimator is called "unbiased", otherwise "biased".

Example: For an i.i.d. $x^{(n)}$, what is the mean square error of a time-average mean estimator?

Answer:

True mean value: $\mu = E\{x^{(n)}\}$

Time-average mean estimator:

$$\hat{\mu}(N) = \frac{1}{N} \sum_{n=0}^{N-1} x^{(n)}$$

statistical expectation of mean estimate:

$$\begin{aligned} E[\hat{\mu}(N)] &= \frac{1}{N} \sum_{n=0}^{N-1} E\{x^{(n)}\} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \mu = \mu \end{aligned}$$

Since $E[\hat{\mu}(N)] = \mu$, it is unbiased!

mean square error: $E[|\hat{\mu}(N) - \mu|^2]$

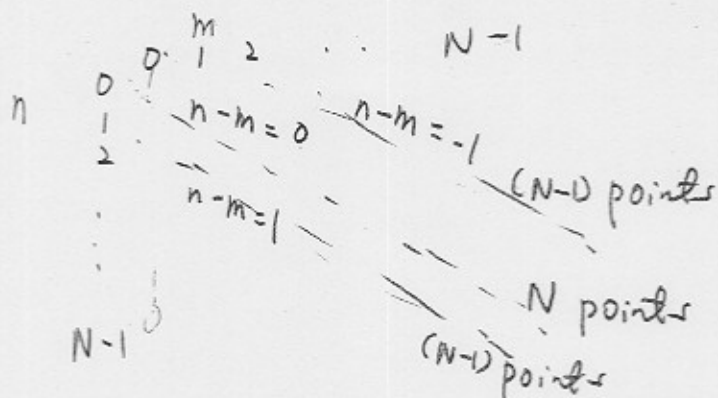
$$= E[|\hat{\mu}(N)|^2 - \mu \hat{\mu}(N) - \mu^* \hat{\mu}(N)] - |\mu|^2$$

$$= E[|\hat{\mu}(N)|^2] - |\mu|^2$$

$$\begin{aligned}
& E \left[\left| \hat{\mu}(N) \right|^2 \right] - |\mu|^2 \\
&= E \left[\frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} x(n) x(m)^* \right] - |\mu|^2 \\
&= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} E \left[x(n) x(m)^* \right] - |\mu|^2 \\
&= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \left\{ E \left[x(n) x(m)^* \right] - |\mu|^2 \right\} \\
&= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \left\{ r(n-m) - |\mu|^2 \right\} = \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} c(n-m)
\end{aligned}$$

Let $l = n - m$

$$= \frac{1}{N} \sum_{l=-N+1}^{N+1} c(l) \left(1 - \frac{|l|}{N} \right)$$



The asymptotical variance of the estimator

is $\lim_{N \rightarrow \infty}$ mean square error of estimator

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=-N+1}^{N+1} c(l) \left[1 - \frac{|l|}{N} \right]$$

$$= 0 \quad \text{if } |c(l)| < \infty, \forall l$$

1.3

Usually the observation has to form a vector $\vec{u}(n)$ such that

$$\vec{u}(n) = [u(n), u(n-1), \dots, u(n-M+1)]^T_{M \times 1}$$

The correlation matrix is defined as

$$\tilde{R} = E [\vec{u}(n) \vec{u}^H(n)]_{M \times M}$$

$$= \begin{bmatrix} r(0) & r(1) & \dots & r(M-1) \\ r(-1) & r(0) & \dots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-M+1) & r(-M+2) & \dots & r(0) \end{bmatrix}$$

Properties of Correlation Matrix

Property 1. The correlation matrix of a stationary discrete-time random process is always Hermitian.

Proof: $r(-k) = r(n, n+k) = E[u(n)u^*(n+k)]$

in Eq. (1.2)

$$= E[u^*(n+k)u(n+k-k)]$$

Let $m = n+k$,

$$r(-k) = E[u^*(m)u(m-k)]$$

$$= \left\{ E[u(m)u^*(m-k)] \right\}^*$$

$$= r^*(k)$$

$$\tilde{R} = \begin{bmatrix} r(0) & r(1) & \dots & r(M-1) \\ r(-1) & r(0) & \dots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-M+1) & r(-M+2) & \dots & r(0) \end{bmatrix}$$

$$= \begin{bmatrix} r(0) & r(1) & \dots & r(M-1) \\ r^*(1) & r(0) & \dots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M-1) & r^*(M-2) & \dots & r(0) \end{bmatrix}$$

$$\Rightarrow \tilde{R} = (\tilde{R}^T)^* = \tilde{R}^H$$

$\therefore \tilde{R}$ is a Hermitian matrix

\therefore "H" is a Hermitian operator,

which is equivalent to a transpose operation and then a complex-conjugate operation on a matrix, or in the reverse order.

Property 2. The correlation matrix of a stationary discrete-time random process is Toeplitz.

Proof:
$$\tilde{R} = \left[E [u(n-p+1) u^*(n-q+1)] \right]_{p,q}$$

$$= [r(q-p)]_{p,q}$$

row, column indices

according to Eq. (1.2) & (1.6)

Therefore \tilde{R} is a Toeplitz matrix.

Property 3 The correlation matrix of a discrete-time random process is always positive semi-definite and positive definite most of the time.

Remark: A square matrix \tilde{R}_c is called

(i) positive definite iff

$$\vec{a}^H \tilde{R} \vec{a} > 0, \text{ for any arbitrary nontrivial vector } \vec{a}, \vec{a}^H \vec{a} \neq 0 \text{ (nontrivial)}$$

Or equivalently, the eigenvalues of \tilde{R} are all positive!!
(if \tilde{R} is Hermitian)

(ii) positive semi-definite iff

$$\vec{a}^H \tilde{R} \vec{a} \geq 0, \text{ for any arbitrary nontrivial vector } \vec{a}, \vec{a}^H \vec{a} \neq 0.$$

Or equivalently, the eigen

values of \tilde{R} are all non-negative (\tilde{R} is Hermitian)

(iii) negative definite iff

$$\vec{a} \tilde{R} \vec{a}^H < 0, \text{ for any arbitrary nontrivial vector } \vec{a}, \vec{a}^H \vec{a} \neq 0.$$

Or equivalently, the eigen values of \tilde{R} are all negative (\tilde{R} is Hermitian)

(iv) negative semi-definite iff

$$\vec{a} \tilde{R} \vec{a}^H \leq 0, \text{ for any arbitrary nontrivial vector } \vec{a}, \vec{a}^H \vec{a} \neq 0$$

Or equivalently, the eigen values of \tilde{R} are all non-positive (\tilde{R} is Hermitian)

Therefore we always check the eigen values to categorize a Hermitian matrix ::

- Example :
- (a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ is positive-definite
- (b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is positive-semi-definite.
- (c) $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix}$ is negative definite.

Proof: Given any arbitrary vector \vec{a} ,
we can construct a new random
variable $y \equiv \vec{a}^H \vec{x}$ (a transformation
from random vector \vec{x} to scalar y).

$$\begin{aligned} \text{Then } \vec{a}^H \tilde{R} \vec{a} &= \vec{a}^H E[\vec{x} \vec{x}^H] \vec{a} \\ &= E[\underbrace{\vec{a}^H \vec{x}}_y \underbrace{\vec{x}^H \vec{a}}_{y^*}] = E[|y|^2] \geq 0 \end{aligned}$$

Hence \tilde{R} is positive semi-definite.

Property 4. The correlation matrix of a
wide-sense stationary process is
nonsingular due to the unavoidable
presence of additive noise. If the
signal is uncorrelated with noise, the
correlation matrix is positive definite.

Proof: A non-singular matrix is invertible,
that means, its inverse matrix exists.

Since $\tilde{R}^{-1} = \frac{\text{adj}(\tilde{R})}{\det(\tilde{R})}$, \tilde{R}^{-1} exists

iff $\det(\tilde{R}) \neq 0$

If we model $\tilde{R} = E[(\vec{x} + \vec{w})(\vec{x} + \vec{w})^H]$

where \vec{x} is the signal vector and \vec{w} is the noise vector of the same dimensionality, and $E[\vec{x}^H \vec{w}] = 0$ (noise is uncorrelated with signal), then

$$\tilde{R} = E[\vec{x} \vec{x}^H] + E[\vec{w} \vec{w}^H]$$

Usually, the noise correlation matrix $E[\vec{w} \vec{w}^H]$ is positive definite.

Therefore, for any arbitrary nontrivial

vector \vec{a} ,

$$\vec{a}^H \tilde{R} \vec{a} = \underbrace{\vec{a}^H E[\vec{x} \vec{x}^H] \vec{a}}_{\text{positive semi-definite}} + \underbrace{\vec{a}^H E[\vec{w} \vec{w}^H] \vec{a}}_{\substack{\text{positive definite} \\ > 0}} > 0 \geq 0$$

$\therefore \tilde{R}$ has all positive eigen values.

$\det(\tilde{R})$ is equivalent to the product of all eigen values $\neq 0 \Rightarrow \tilde{R}$ is invertible and non-singular.

Property 5. The correlation matrix of a reverse-order random vector is the transpose matrix of the correlation matrix of the random vector in the original order.

$$\vec{x}_1 = [x(n) \quad x(n-1), \dots, x(n-M+1)]$$

$$\vec{x}_2 = [x(n-M+1), x(n-M), \dots, x(n)]$$

$$E[\vec{x}_1 \vec{x}_1^H] = \tilde{R}$$

$$E[\vec{x}_2 \vec{x}_2^H] = \tilde{R}^T$$

It is easy to prove.

Property 6. The correlation matrices \tilde{R}_M and \tilde{R}_{M+1} of a stationary discrete-time random process pertaining to M and $M+1$ observations of process, respectively, are related by

$$\tilde{R}_{M+1} = \begin{bmatrix} r(0) & \vec{r}^H \\ \vec{r} & \tilde{R}_M \end{bmatrix} \quad \text{or}$$

equivalently,

$$\tilde{R}_{M+1} = \begin{bmatrix} \tilde{R}_M & \vec{r}^{B*} \\ \vec{r}^{BT} & r(0) \end{bmatrix}$$

where $r(0) = E[x(n)x^*(n)]$

$$\vec{r}^H = [r(1), r(2), \dots, r(M)]$$

$$\vec{r}^{BT} = [r(-M), r(-M+1), \dots, r(-1)]$$

Proof:

$$\tilde{R}_{M+1} = \begin{bmatrix} r(0) & r(1) & r(2) & \dots & r(M) \\ r^*(1) & r(0) & r(1) & \dots & r(M-1) \\ r^*(2) & r^*(1) & \dots & \dots & r(M-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^*(M) & r^*(M-1) & \dots & \dots & r(0) \end{bmatrix}$$

Example: Correlation Matrix of

a sine wave plus noise

$$x(n) = \alpha e^{j\omega n} + v(n), \quad n=0, 1, \dots, N-1$$

Assume the noise $v(n)$ is white process, then $E[v(n) v^*(n-k)]$

$$= \begin{cases} \sigma_v^2, & k=0 \\ 0, & k \neq 0 \end{cases}$$

The autocorrelation function of $x(n)$ for a lag k is given by

$$r(k) = E[x(n) x^*(n-k)] = \begin{cases} |\alpha|^2 + \sigma_v^2, & k=0 \\ |\alpha|^2 \exp(j\omega k), & k \neq 0 \end{cases}$$

Hence $\tilde{R} = E[\tilde{x} \tilde{x}^H]$

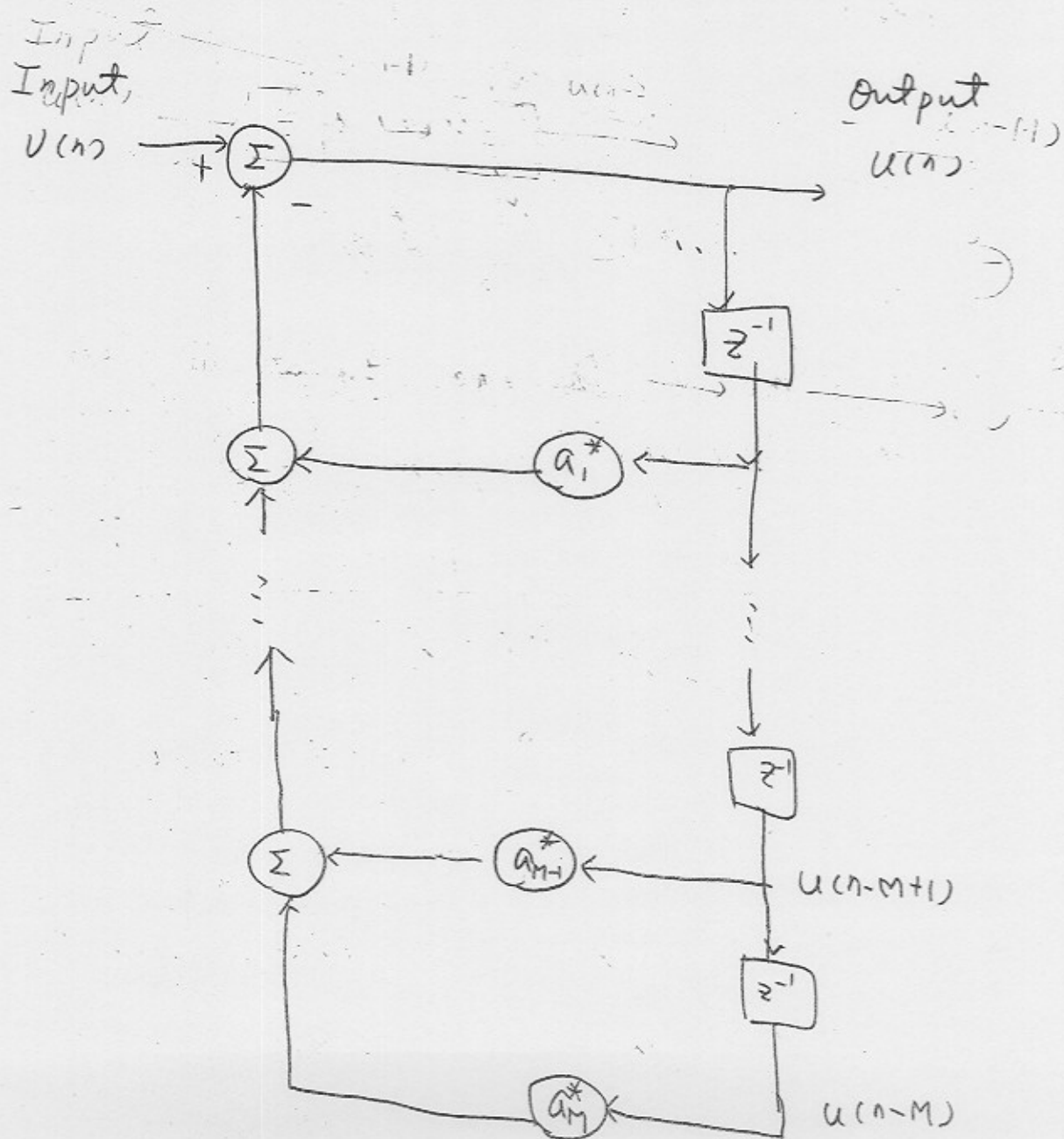
$$= |\alpha|^2 \begin{bmatrix} 1 + \frac{1}{\rho} \exp(j\omega) & \dots & \exp[j\omega(M-1)] \\ \exp(-j\omega) & 1 + \frac{1}{\rho} & \dots & \exp[j\omega(M-2)] \\ \vdots & & & \\ \exp[-j\omega(M-1)] & \exp[-j\omega(M-2)] & \dots & 1 + \frac{1}{\rho} \end{bmatrix}$$

where $\rho = \frac{|\alpha|^2}{\sigma_v^2}$

1.5 Linear filter Models

* Auto regressive Models

A linear filter can be applied to build an adaptive system, such as an all-pole filter as depicted as follows:



The difference equation associated with such an AR model is

$$V(n) = u(n) + a_1^* u(n-1) + \dots + a_M^* u(n-M)$$

Take z -transform,

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

where $V(z) = \sum_{n=0}^{\infty} V(n) z^{-n}$

$$U(z) = \sum_{n=0}^{\infty} u(n) z^{-n}$$

$$H_A(z) = \sum_{n=0}^M a_n^* z^{-n}$$

The model transfer function can be

described as $H_A(z) = \frac{1}{H_A(z)} = \frac{1}{\sum_{n=0}^M a_n^* z^{-n}}$

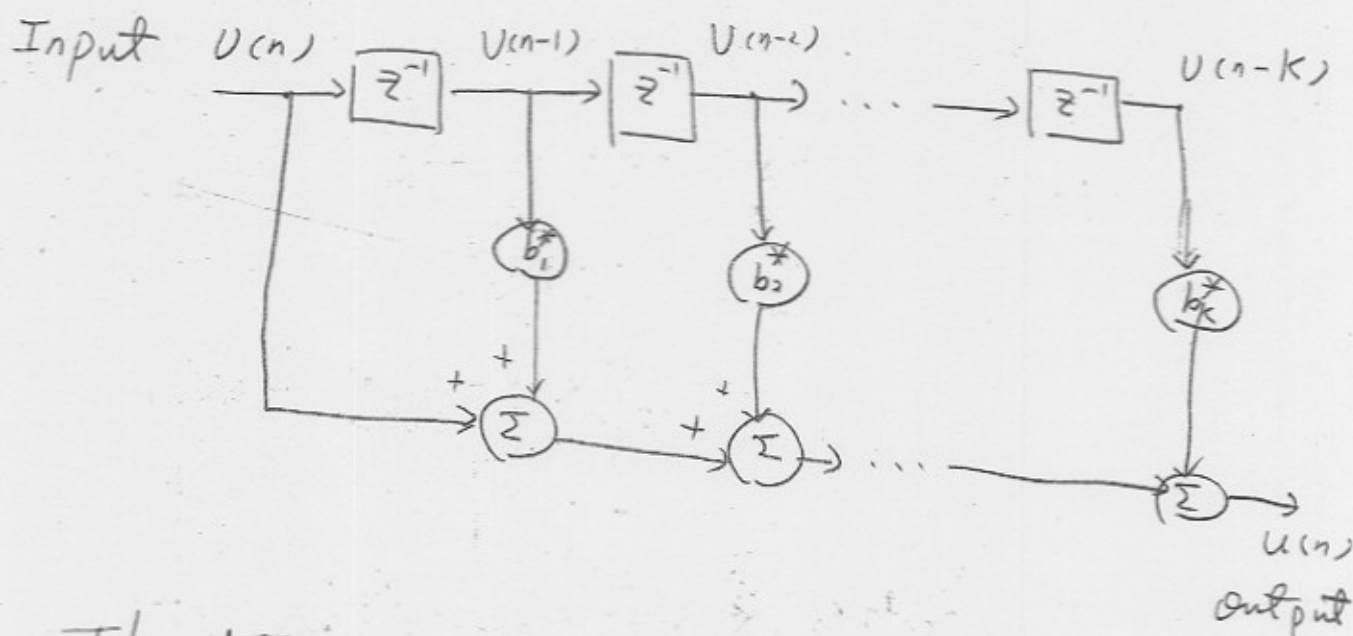
$$= \frac{1}{(1 - p_1 z^{-1})(1 - p_2 z^{-1}) \dots (1 - p_M z^{-1})}$$

where p_1, p_2, \dots, p_M are the poles or the roots of $\sum_{n=0}^M a_n^* z^{-n} = 0$.

The AR model can characterize the spectral peaks in the frequency domain.

* Moving-Average Models

A linear filter can be applied to build an adaptive system, such as an all-zero filter as depicted as follows:



The difference equation associated with such an MA model is

$$U(n) = V(n) + b_1^* U(n-1) + \dots + b_k^* U(n-k)$$

Take z -transform,

$$U(z) = H_M(z) V(z)$$

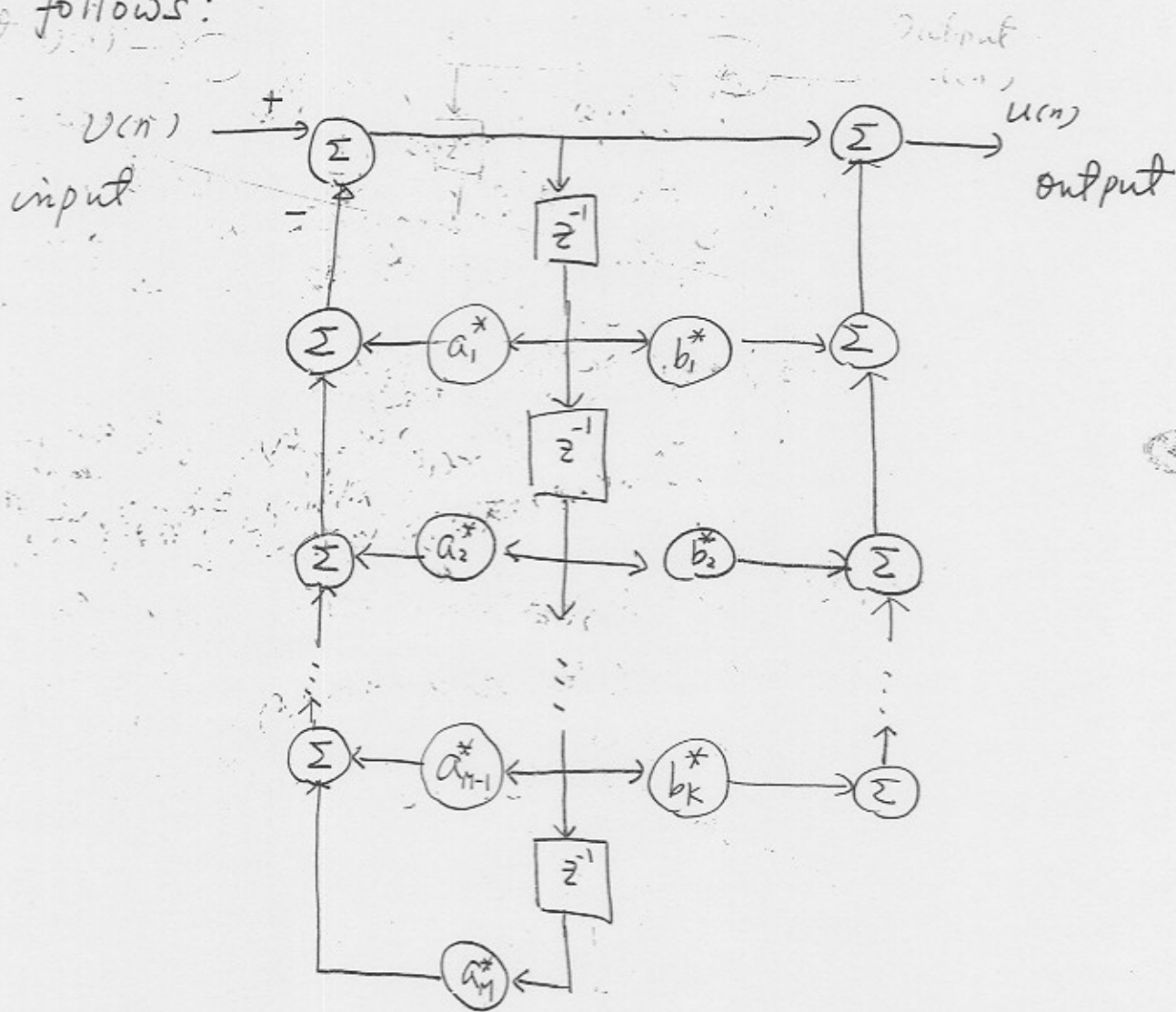
$$\text{where } U(z) = \sum_{n=0}^{\infty} U(n) z^{-n}$$

$$V(z) = \sum_{n=0}^{\infty} U(n) z^{-n}$$

$$H_M(z) = \sum_{n=0}^k b_n^* z^{-n}, \quad b_0 = 1$$

The MA model can characterize the spectral nulls in the frequency domain. Auto regressive-moving-average Models: The third alternative to build a linear filter is pole-zero filter or autoregressive-moving-average model (ARMA), which is depicted

as follows:



The difference equation associated with such an ARMA model is

$$\begin{aligned}
 u(n) + a_1^* u(n-1) + \dots + a_M^* u(n-M) \\
 = u(n) + b_1^* u(n-1) + \dots + b_k^* u(n-k)
 \end{aligned}$$

The ARMA model can characterize the spectral peaks and nulls in the frequency domain.

1.10 Selecting the Model Order

The representation of a stochastic process by a stochastic process by a linear model may be used for synthesis or analysis. Both applications involve with

the demand of some a priori information, the model order (the degrees of freedom in the model, or the number of parameters to specify a linear filter). For example,

the previous AR model order equals M ;

the previous MA model order equals k ;

the previous ARMA model order equals (M, k) .

How can we determine the model order?

An information-theoretic criterion is applied to determine the model order.

Let $u_i = u(i)$, $i = 1, 2, \dots, N$, denote the data obtained by N independent observations of a stationary discrete-time random process.

Let $f_{\sigma}(u_i | \hat{\theta}_m)$ denote the conditional probability density function of u_i , given $\hat{\theta}_m$, the estimated vector of parameters that model the process. Let m be the model order so that we can write the set of parameters estimates as a vector:

$$\hat{\theta} = [\hat{\theta}_{1m}, \hat{\theta}_{2m}, \dots, \hat{\theta}_{mm}]^T$$

An information-theoretic criterion (AIC) proposed by Akaike selects the model for which the quantity

$$AIC(m) = -2L(\hat{\theta}_m) + 2m$$

is minimum, where $L(\hat{\theta}_m) = \sum_{i=1}^N \ln [f_{\sigma}(u_i | \hat{\theta}_m)]$

The AIC criterion is derived by minimizing the Kullback-Leibler divergence which is used to provide a measure of the divergence between the unknown true probability density function $g(\vec{u})$ and the conditional probability density function $f_{\vec{u}}(u_i | \hat{\theta}_m)$ given by the model in light of the observed data. If we plot $AIC(m)$ versus m , the graph will, in general, show a definite minimum value, and the optimum order of the model is determined by that value of m at which $AIC(m)$ attains its minimum value.