EE6401 Advanced Digital Signal Processing

Part 2

Assoc Prof. Lin Zhiping

Office: S2 B2a-14
Tel: (65)6790-6857
Email: ezplin@ntu.edu.sg
http://www.ntu.edu.sg/home/ezplin

Sem 1 AY 2019-2020

Outline

Outline

Outline

Outline

References

Textbook:

- J. G. Proakis and D. G. Manolakis,"Digital Signal Processing: Principles, Algorithms, and Applications", Fourth Edition, Pearson Prentice Hall,2007.

Introduction

- Definition and taxonomy of systems and signals
- Random variables and statistics
- Stochastic processes and ensemble statistics
- Power density spectrum
- Innovations representation of a stationary random process

Definition of systems

Definition

A **system** *is an entity that manipulates one or more signals to accomplish a function and in turn yields new signals.*

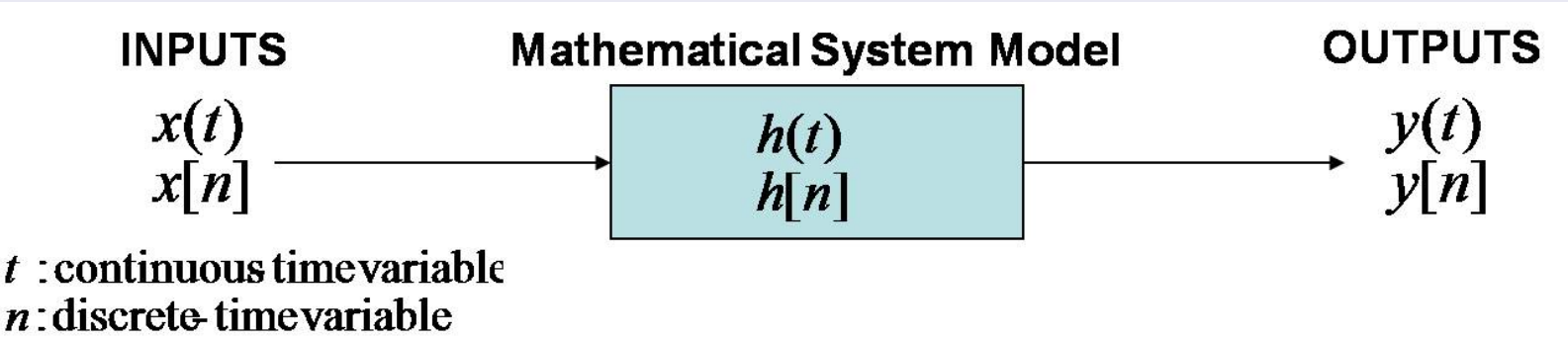


Figure 1: Every physical system is characterized by its ability to accept an input signal and to produce an output signal.

Example

Communication, control, biomedical systems.

Taxonomy of systems

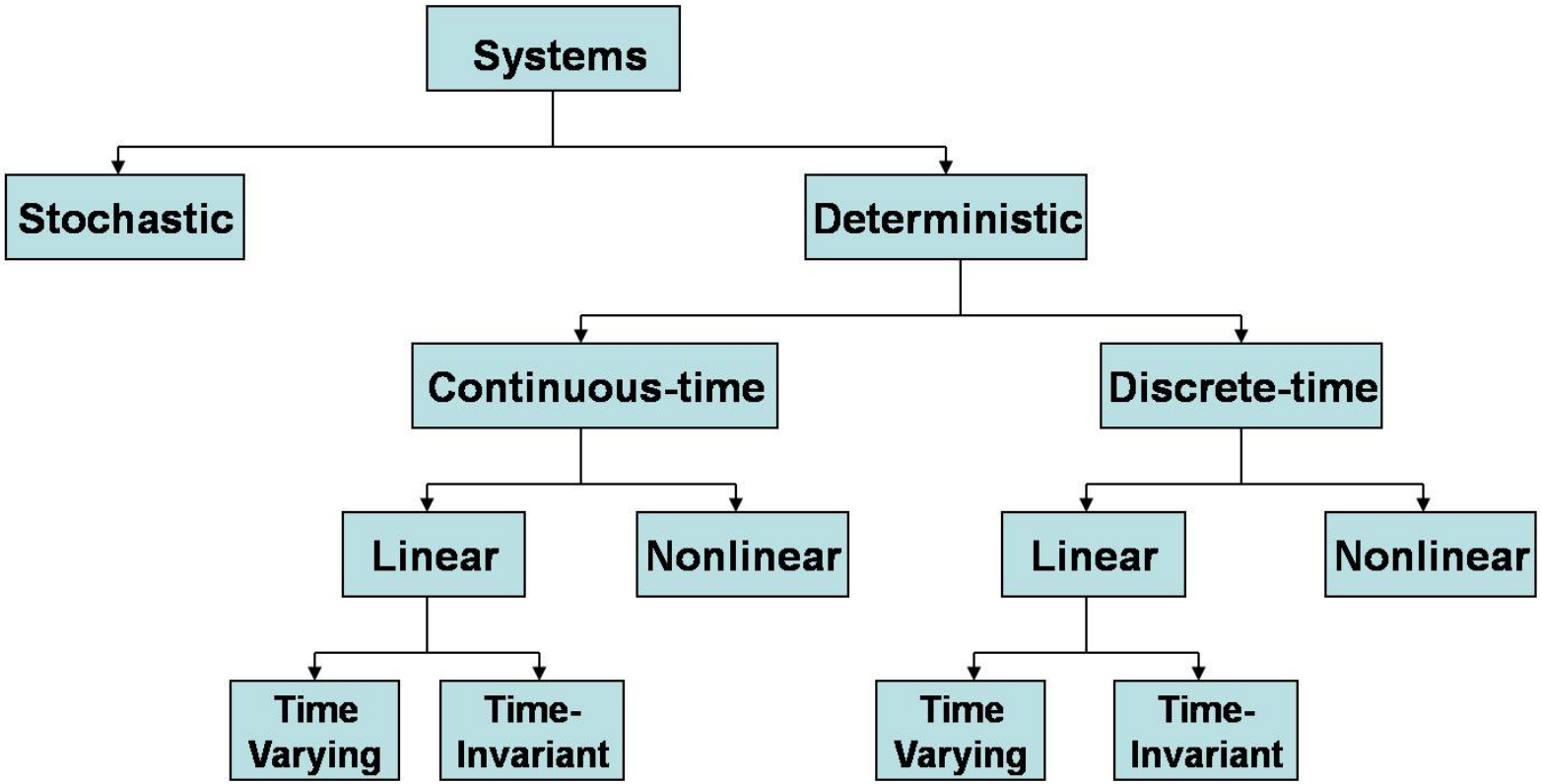


Figure 2: Different types of systems

System analysis

The goal of system analysis is to determine the system response to input or excitation signals so as to

- establish a performance specification
- aid in component selection
- uncover and study system deficiencies
- explain unusual or unexpected system operation
- produce quantitative data related to system operation for various needs

Definition of a signal

Definition

A **signal** *conveys information on the nature of a physical phenomenon and is a function of time, space or any other independent variable(s).*

Example

Communication, speech, biomedical signals, images.

Taxonomy of signals

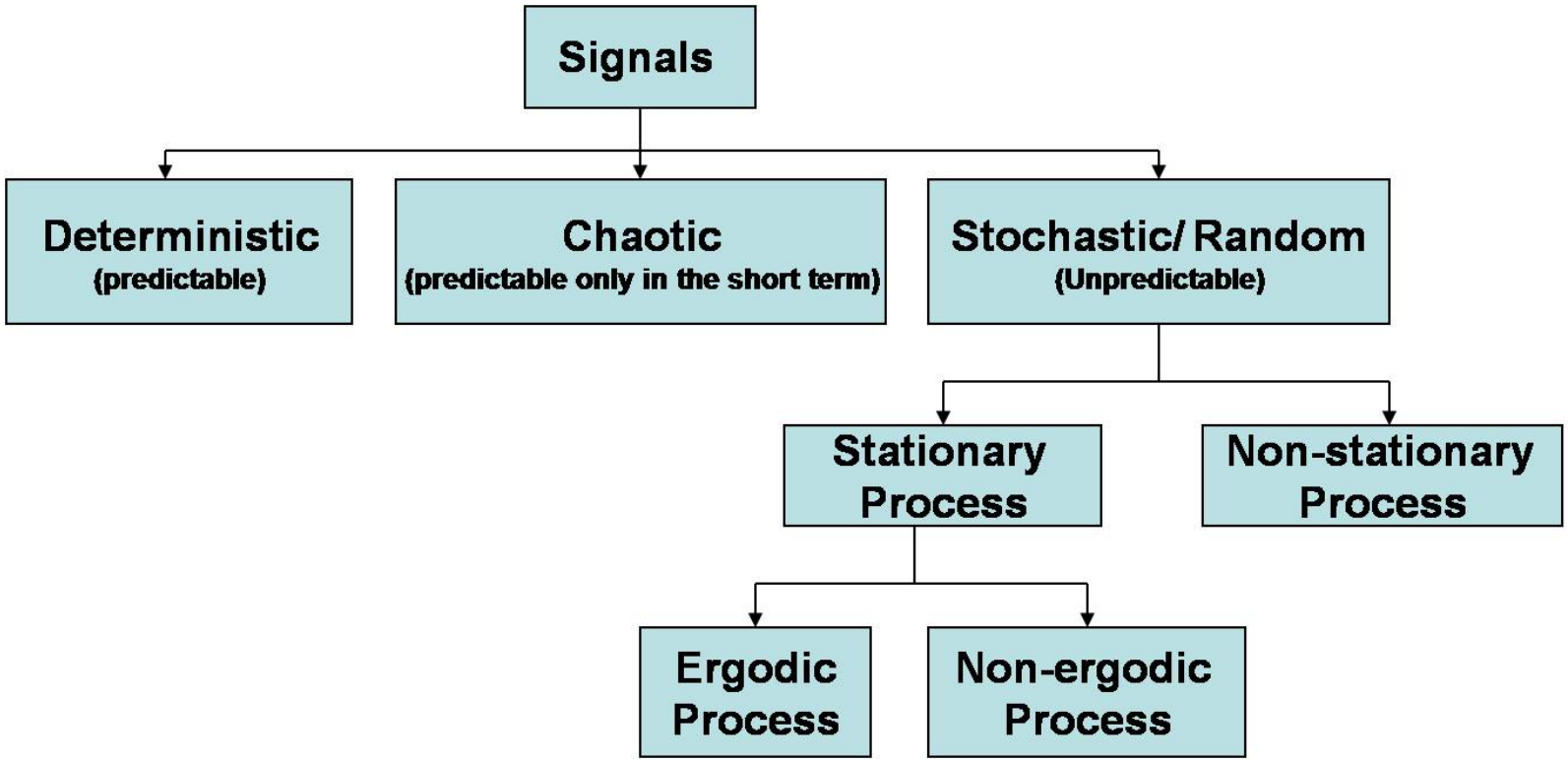


Figure 3: Three types of signals: deterministic, chaotic, stochastic.

Deterministic signals

Definition

A **deterministic** *signal is one that can be reproduced exactly by repeating the process.*

Example

Speech if it is seen as a set of specific waveform signals or

$$x[n] = 10\cos(2\pi n - 0.5), n \in \mathbb{Z}. \quad (1)$$

Chaotic signals

Definition

A **chaotic** *signal comes from a chaotic system which is defined by a coupled system of nonlinear differential equations whose parameters are fixed. Chaotic systems are sensitive to initial conditions and therefore are only predictable in the short term.*

Example

Fluid turbulence and radar sea clutter.

Random signals

Definition

A **random** *signal is one that is not repeatable in a predictable manner, that is, it contains a random variable.*

Example

- *Speech if it is seen as all possible signals emanating from a general speech process*
- *Thermal noise voltages generated in resistors of electronic devices such as radio or television receive*
- *Meteorological phenomena such as air temperature and air pressure fluctuates randomly as a function of time or* $x[n] = 10\cos(2\pi n - \theta)$*, where* θ *is a random variable, statistically uniformly distributed in* $(-\pi, \pi)$

Review: Random variables and statistics

- Discrete random variables and probability mass functions (p.m.f)
- Continuous random variables and probability density functions (p.d.f.)
- Expectation, moment generating function, variance
- Joint p.m.f./p.d.f.
- Independence, covariance, correlation
- Exercises

Discrete random variables

Consider the following variables:

S: Sample space is the set of all possible outcomes of an experiment where the outcome is unpredictable or random.

E: An event is a subset of the sample space S

$P(E)$: Probability of the event E such that

(i) $0 \leq P(E) \leq 1$

(ii) $P(\mathrm{S}) = 1$

(iii) $P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$ where the events E_n are mutually exclusive, i.e. $E_n \cap E_m = \emptyset$ when $n \neq m$.

X: Real valued function defined on the sample space S.

Discrete random variables

Definition

A **discrete random variable (r.v.)** X *can take on at most a countable number of possible values.*

Example

Let X *denote the r.v. that is defined as the sum of the two fair dice. The sample space is given by* $S=\{(1,1),(1,2),(1,3),\ldots,(6,4),(6,5),(6,6)\}$ *and is of size* 36. *Consider the event such that* $X=4$, *then* $E=\{(1,3),(2,2),(3,1)\}\subset S$ *and* $P(E)=P(\{X=4\})=\frac{3}{36}$.

Probability mass function

Definition

The **probability mass function (p.m.f)** *of a r.v. X is defined by*

$$p(a) = P(\{X = a\}) > 0 \tag{2}$$

for at most a countable number of values of a

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P(\{X = x_i\}) = 1 \tag{3}$$

Example

Cumulative distribution function

Definition

The **cumulative distribution function (c.d.f)** *F of a r.v. X is defined by*

$$\begin{aligned} F(a) &= P(\{X \leq a\}) = \sum_{x_i < a} P(\{X = x_i\}) \\ &= \sum_{x_i < a} p(x_i) \end{aligned} \tag{4}$$

Example

Discrete random variable distributions:

- *Bernouilli distribution*
- *Binomial distribution*
- *Geometric distribution*
- *Poisson distribution*

For a summary of corresponding definitions and statistics, see Slide 27.

Continuous random variables

Definition

A **continuous random variable (r.v.)** *X can take on an uncountable number of possible values.*

Example

Probability density function

Definition

The **probability density function (p.d.f)** $f(x)$ *of a continuous r.v.* X *is such that for any set* B *of real numbers:*

$$P(\{X \in B\}) = \int_{x \in B} f(x)\, dx \text{ where } f(x) > 0 \text{ for all } x \in \mathbb{R}$$

$$P(\{X \in \mathbb{R}\}) = \int_{x \in \mathbb{R}} f(x)dx = 1 \tag{5}$$

Example

Cumulative distribution function

Definition

The **cumulative distribution function (c.d.f)** *F of a r.v. X is defined by*

$$F(a) = P(\{X \in (-\infty, a)\}) = \int_{-\infty}^{a} f(x)\, dx \tag{6}$$

$\Rightarrow \frac{d}{da} F(a) = f(a)$

Example

Continuous random variable distributions:

- *Uniform distribution*
- *Exponential distribution*
- *Gamma distribution*
- *Normal (Gaussian) distribution*

For a summary of corresponding definitions and statistics, see Slide 28.

Expectation of a random variable

Definition

The **expected value** *of the r.v. X is the weighted average of the possible values that X can take on, each value being the probability that X assumes that value*

$$\mu_X = E[X] = \begin{cases} \sum\limits_{x:p(x)>0} xp(x) & \text{if } X \text{ is a discrete r.v.} \\ \int\limits_{x\in\mathbb{R}} xf(x)\,dx & \text{if } X \text{ is a continuous r.v.} \end{cases} \tag{7}$$

The expected value of the r.v. X is referred to as the mean or the first moment of X.

Example

Expectation of a real-valued function of a r.v.

Definition

The **expected value of a real-valued function** $g(X)$ *of a r.v.* X *is defined by*

$$E[g(X)] = \begin{cases} \sum\limits_{x:p(x)>0} g(x)p(x) & \text{if } X \text{ is a discrete r.v.} \\ \int\limits_{x\in\mathbb{R}} g(x)f(x)\,dx & \text{if } X \text{ is a continuous r.v.} \end{cases} \tag{8}$$

If $g(X) = X^m, m \geq 1$ *then* $E[X^m], m \geq 1$ *is called the* m^{th} *moment of* X.

Example

Moment generating function

Definition

The **moment generating function** *of the r.v. X is defined by*

$$\phi(t) = E[e^{tX}] = \begin{cases} \sum\limits_{x:p(x)>0} e^{tx} p(x) & \text{if } X \text{ is a discrete r.v.} \\ \int\limits_{x\in\mathbb{R}} e^{tx} f(x)\, dx & \text{if } X \text{ is a continuous r.v.} \end{cases} \tag{9}$$

for all values $t \in \mathbb{R}$.

Example

The moments of X are obtained by successively differentiating $\phi(t)$ and evaluating it at $t = 0$:

$$\phi^{(1)}(t) = \frac{d}{dt}E[e^{tX}] = E[\frac{d}{dt}e^{tX}] = \quad E[Xe^{tX}] \quad \Rightarrow E[X] = \phi^{(1)}(0)$$

$$\phi^{(2)}(t) = \frac{d}{dt}\phi^{(1)}(t) = E[\frac{d}{dt}Xe^{tX}] = \quad E[X^2e^{tX}] \quad \Rightarrow E[X^2] = \phi^{(2)}(0)$$

$$\vdots$$

$$\phi^{[m]}(t) = \frac{d}{dt}\phi^{(m-1)}(t) = E[\frac{d}{dt}X^{m-1}e^{tX}] = \quad E[X^me^{tX}] \quad \Rightarrow E[X^m] = \phi^{[m]}(0)$$

Variance of a random variable

Definition

The **variance** *of a random variable X measures the expected square of the deviation of X from its expected value and is defined by*

$$\sigma_X = Var(X) = E[(X - E[X])^2] \tag{10}$$

Example

Suppose X is a continuous r.v. with p.d.f $f(x)$ and let $E[X] = \mu_X$ then

$$\begin{aligned}
Var(X) &= E[(X - E[X])^2] = E[(X - \mu_X)^2] = E[X^2 - 2\mu_X X + \mu_X^2] \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu_X x + \mu_X^2) f(x)\ dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)\ dx - 2\mu_X \int_{-\infty}^{\infty} x f(x)\ dx + \mu_X^2 \int_{-\infty}^{\infty} f(x)\ dx \\
&= E[X^2] - 2\mu_X E[X] + \mu_X^2 \cdot 1 \\
&= E[X^2] - \mu_X^2 \\
&= E[X^2] - (E[X])^2 \qquad \left[= \phi^{(2)}(0) - (\phi^{(1)}(0))^2 \right]
\end{aligned}$$

Statistics of a discrete random variable

Discrete probability distribution	Probability mass function, $p(x)$	Moment generating function, $\phi(t)$	Mean	Variance
Binomial with parameters n, p $0 \le p \le 1$	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	$(pe^t + (1-p))^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \le p \le 1$	$p(1-p)^{x-1}$, $x = 1, 2, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Statistics of continuous random variable

Continuous probability distribution	Probability density function, $f(x)$	Moment generating function, $\phi(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters (n, λ) $\lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^n$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$, $-\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

Jointly distributed discrete random variables

Consider two discrete r.v. X and Y.

Definition

The **joint probability mass function** *of X and Y is defined by*

$$p(x,y) = P(\{X = x, Y = y\}). \tag{11}$$

Example

The joint cumulative probability distribution function of X and Y is defined by

$$F(a,b) = P(\{X \leq a, Y \leq b\}) \text{ with } a,b \in \mathbb{R}. \tag{12}$$

Note that the p.m.f. of X can be obtained from the joint distribution function of X and Y

$$p_X(x) = \sum_{y:p(x,y)>0} p(x,y). \tag{13}$$

Similarly for the p.m.f. of Y.

Jointly distributed continuous random variables

Consider two continuous r.v. X and Y.

Definition

The **joint probability density function** *of X and Y is defined by $f(x,y)$ if the following is satisfied*

$$P(\{X \in A, Y \in B\}) = \int_B \int_A f(x,y)dx\ dy. \tag{14}$$

Example

The p.d.f. of X can be obtained from the joint p.d.f. of X and Y

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\ dy. \tag{15}$$

Similarly for the p.d.f. of Y.

Independent random variables

Definition

The random variables X and Y are **independent** *if for all a, b*

$$P(\{X \leq a, Y \leq b\}) = P(\{X \leq a\})P(\{Y \leq b\}) \tag{16}$$

Example

The p.d.f. of X can be obtained from the joint p.d.f. of X and Y

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\ dy. \tag{17}$$

Corollary

If X and Y are independent then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for any functions g and h.

Correlation and Covariance of two random variables

Definition

The correlation of two random variables X and Y is defined by

$$\rho_{XY} = corr(X, Y) = E[XY] \tag{18}$$

and measures how the two variables affect one another.

Definition

The covariance of two random variables X and Y is defined by

$$\begin{aligned} \sigma_{XY} = cov(X, Y) &= E[(X - E[X])(Y - E[Y])] && (19)\\ &= E[XY] - E[X]E[Y] && (20)\\ &= \rho_{XY} - \mu_X \mu_Y && (21) \end{aligned}$$

and measures the deviation of one r.v to another r.v.

Note that if X and Y are independent r.v. then $cov(X, Y) = 0$.

Exercises

1. Calculate the mean and the variance of the random variable uniformly distributed over the interval $(-\pi, \pi)$
2. Show that $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$

Stochastic/random process

Definition

A **stochastic/random process** *is a family or an ensemble of all possible time functions which describes the evolution through time of some (physical) process :*

$$\{X(t,S), t \in T\} \tag{22}$$

where t is the time index and S is the set (sample space) of all possible sample functions.

If T is a countable set then $X[n,S]$ is a discrete-time process.
If T is an interval of the real line then $X(t,S)$ is a continuous-time process.

Example

$x(t,\Phi) = A\sin(2\pi f_0 t + \Phi)$ *where* Φ *is a random variable uniformly distributed on the interval* $[-\pi,\pi]$*and* A, f_0 *are constants.*

Realization of a Stochastic Process

Definition

A sample function $x(t,s)$ of the ensemble, is called a realization of the process.

For simplicity, we denote a random process by $X(t)$ and a single realization of the random process by $x(t)$.

Example

The Figure on the next page illustrates a set of waveforms corresponding to the air temperature in different cities around the world. This is an ensemble of time functions or a random process. One waveform of the temperature in any particular city is a single realization or a sample function of the random process.

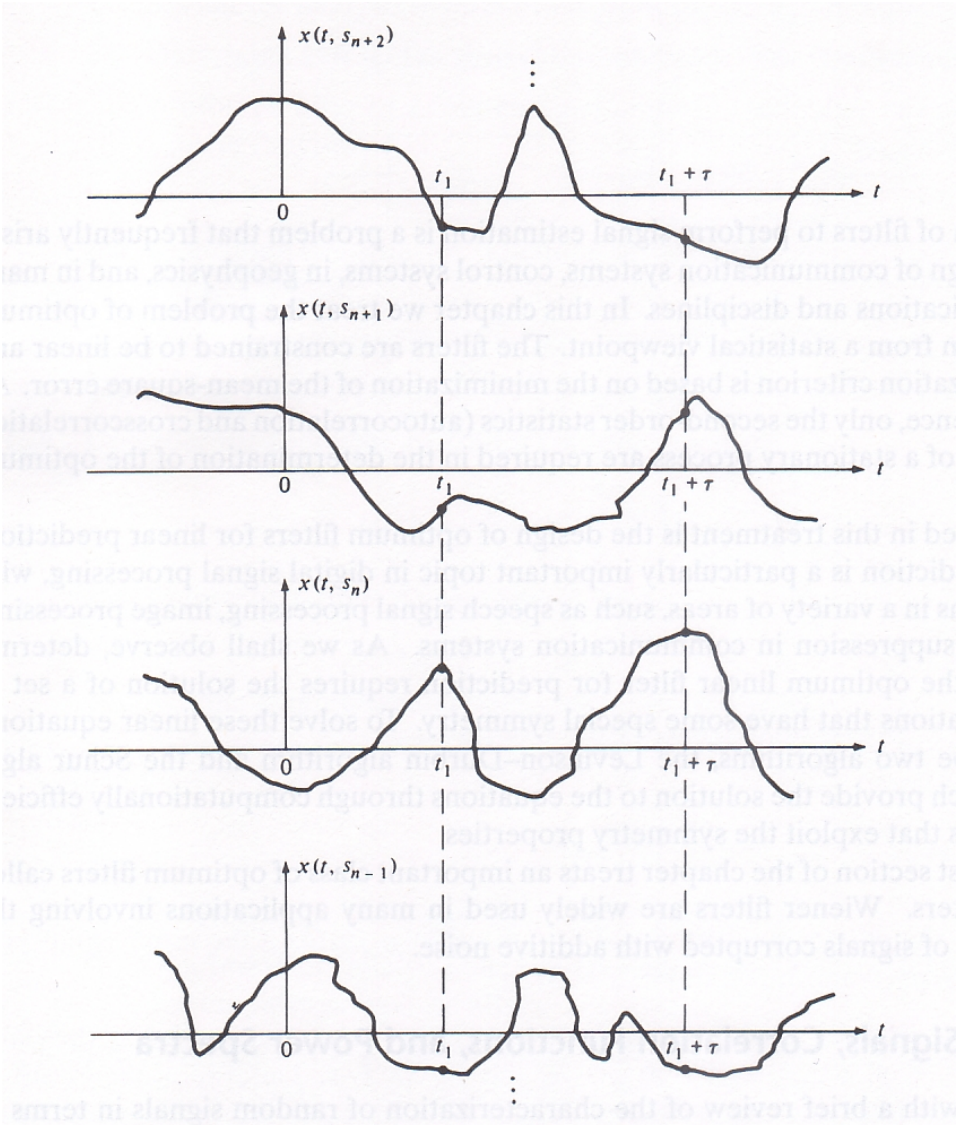


Figure 4: Sample functions of a random process.

Statistical (ensemble) averages

Definition

Consider a continuous random process $X(t)$ sampled at time instant $t = t_i$ then $X_{t_i} = X(t_i)$ is a random variable with corresponding p.d.f. $f(x_{t_i})$and the m^{th} **moment** *of the r.v. is given by*

$$E[X_{t_i}^m] = \int_{-\infty}^{\infty} x_{t_i}^m f(x_{t_i})\, dx_{t_i} \tag{23}$$

The statistical (Ensemble) **mean** is $\mu_X(t_i) = E[X_{t_i}]$.
The statistical (Ensemble) **mean square** is $E[X_{t_i}^2]$.
The statistical (Ensemble) **variance** is $\sigma_X(t_i) = E\left[\left(X_{t_i} - \mu_X(t_i)\right)^2\right]$.

Correlation

Definition

Consider two random variables X_{t_1}, X_{t_2} where $X_{t_i} = X(t_i), i = 1, 2$. The **statistical (ensemble) correlation** *between X_{t_1} and X_{t_2} is given by the joint moment*

$$\gamma_{xx}(t_1, t_2) = E[X_{t_1} X_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{t_1} x_{t_2} f(x_{t_1}, x_{t_2}) \, dx_{t_1} \, dx_{t_2} \tag{24}$$

which depends on time instants t_1, t_2 and is called the **autocorrelation** *of a random process.*

Autocovariance

Definition

The **autocovariance** *function is defined by*

$$c_{xx}(t_1,t_2) = E\left[\left(X_{t_1}-E[X_{t_1}]\right)\left(X_{t_2}-E[X_{t_2}]\right)\right] \quad (25)$$

$$= \gamma_{xx}(t_1,t_2)-\mu_X(t_1)\mu_X(t_2) \quad (26)$$

where $\mu_X(t_i)=E[X_{t_i}], i=1,2.$

Statistical (ensemble) averages for joint random processes

Definition

The **cross-correlation** *function of two processes $X(t)$ and $Y(t)$ is defined by the joint moment*

$$\gamma_{XY}(t_1, t_2) = E[X_{t_1} Y_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{t_1} y_{t_2} f(x_{t_1}, y_{t_2}) \, dx_{t_1} \, dy_{t_2} \tag{27}$$

Note that the cross-correlation is used to measure time delays, determine transmission paths, detection and recovery of signals in noise.

Cross-covariance

Definition

The **cross-covariance** *function of two processes* $X(t)$ *and* $Y(t)$ *is defined by*

$$c_{XY}(t_1, t_2) = \gamma_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2) \tag{28}$$

where $\mu_X(t_1) = E[X_{t_1}], \mu_Y(t_2) = E[Y_{t_2}]$.

If $c_{XY}(t_1, t_2) = 0$, that is,

$$\gamma_{XY}(t_1, t_2) = \mu_X(t_1)\mu_Y(t_2) \Leftrightarrow E[X_{t_1} Y_{t_2}] = E[X_{t_1}]E[Y_{t_2}] \tag{29}$$

then $X(t)$ and $Y(t)$ are uncorrelated.

If $E[X_{t_1} Y_{t_2}] = 0$ then $X(t)$ and $Y(t)$ are statistically orthogonal.

Stationary stochastic process

Suppose we have two sets of n samples of the random process $X(t)$:

Set 1 $X_{t_i} \equiv X(t_i)$ with joint p.d.f. $f(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$

Set 2 $X_{t_i+\tau} \equiv X(t_i+\tau)$ with joint p.d.f. $f(x_{t_1+\tau}, x_{t_2+\tau}, \ldots, x_{t_n+\tau})$

that is, the second set is displaced in time from the first set by an amount τ for all $i = 1, 2, \ldots, n$.

Definition

A stochastic process is **stationary** *if the joint probability density functions of the two sets of random variables are equal,*

$$f(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = f(x_{t_1+\tau}, x_{t_2+\tau}, \ldots, x_{t_n+\tau}) \tag{30}$$

for all τ and all n.

That is, the statistical properties of a stationary stochastic process are invariant with respect to translation of the time axis.

Auto-correlation, auto-covariance, cross-correlation of stationary processes

When the process $X(t)$ is stationary then the joint p.d.f. 's of the two pairs are equal, i.e., $f(x_{t_1}, x_{t_2}) = f(x_{t_1+\tau}, x_{t_2+\tau})$. Then the auto-correlation and auto-covariance functions depend on the time difference $t_1 - t_2 = \tau$, i.e.,

$$\gamma_{xx}(t_1, t_2) = \gamma_{xx}(t_1, t_1+\tau) = E[X_{t_1} X_{t_1+\tau}] = \gamma_{xx}(\tau) \quad (31)$$

$$c_{xx}(t_1, t_2) = c_{xx}(t_1 - t_2) = c_{xx}(\tau) = \gamma_{xx}(\tau) - \mu_X^2 \quad (32)$$

respectively and the cross-correlation of two jointly and individually stationary processes is

$$\gamma_{XY}(t_1, t_2) = \gamma_{XY}(t_1 - t_2) = \gamma_{XY}(\tau) \text{ and } \gamma_{XY}(-\tau) = \gamma_{YX}(\tau) \quad (33)$$

For more details see http://cnx.org/content/m10676/latest/

Wide-sense or weakly stationary random process

Definition

A process is **wide-sense** *or* **weakly stationary** *if both its mean and variance are finite and constant and its auto-correlation function depends only on the difference between the times of occurrence of the samples or the lag time, i.e.*

1. *Mean:* $\mu_X(t_1) = E[X_{t_1}] = \mu_X$
2. *Variance:* $\sigma_X(t_1) = E\left[\left(X_{t_1} - \mu_X\right)^2\right] = \sigma_X$
3. *Auto-correlation:* $\gamma_{xx}(t_1, t_2) = \gamma_{xx}(t_1 - t_2) = \gamma_{xx}(\tau)$

Exercise

Consider a random process defined by

$$X(t) = A\cos(\Omega_0 t - \Phi) \tag{34}$$

where Ω_0 and Φ are two constants, while A is a random variable uniformly distributed within $(0.5, 1)$,i.e. $f(a) = \begin{cases} \frac{1}{1-0.5} & \text{if } 0.5 < a < 1 \\ 0 & \text{otherwise} \end{cases}$

Is $X(t)$ a weakly stationary random process?

Solution:

$$\begin{aligned} \mu_X(t) &= E[X(t)] = \int_{-\infty}^{+\infty} x(t) f(a) da = \int_{0.5}^{1} a\cos(\Omega_0 t - \Phi)\frac{1}{1-0.5} da \\ &= (1^2 - 0.5^2)\cos(\Omega_0 t - \Phi) = 0.75\cos(\Omega_0 t - \Phi) \end{aligned}$$

Since $\mu_X(t)$ is not a constant $X(t)$ is not weakly stationary.

Exercise

Consider a random process defined by $X(t) = A\cos(\Omega_0 t - \Phi)$ where Ω_0 and A are two constants, while Φ is a random variable uniformly distributed within $(-\pi, \pi)$,i.e.$f(\phi) = \begin{cases} \frac{1}{\pi-(-\pi)} & \text{if } -\pi < \phi < \pi \\ 0 & \text{otherwise} \end{cases}$

Is $X(t)$ a weakly stationary random process?

Solution:

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{+\infty} x(t) f(\phi) d\phi \quad (35)$$

$$= \int_{\pi}^{-\pi} A\cos(\Omega_0 t - \phi) \frac{1}{\pi - (-\pi)} d\phi \quad (36)$$

$$= 0 \quad (37)$$

Solution:

$$\begin{aligned}
\gamma_{xx}(t_1, t_2) &= E[X(t_1)X(t_2)] = \int_{-\infty}^{+\infty} x(t_1)x(t_2)f(\phi)d\phi \qquad (t_2 = t_1 - \tau) \\
&= \int_{\pi}^{-\pi} A^2 \cos(\Omega_0 t_1 - \phi)cos(\Omega_0 t_2 - \phi) \times \frac{1}{2\pi} d\phi \\
&= \ldots = \frac{A^2}{2} \cos(\Omega_0 \tau) = \gamma_{xx}(\tau).
\end{aligned}$$

Since $\mu_X(t)$ is constant and $\gamma_{xx}(t_1, t_2) = \gamma_{xx}(\tau)$, then $X(t)$ is weakly stationary.

Ergodicity

Definition

A stochastic process is **ergodic** *if, with probability 1, the time averages obtained from a single sample function or a single realization are equal to the statistical (ensemble) averages.*

Note: This is useful in practice as usually only one realization of the stochastic process is available.

Discrete-time random signals

Discrete-time random signals $X[n]$ can be obtained by uniform sampling continuous-time random signals $X(t)$, thus similar statistical properties can be derived.

White noise/sequences

Definition

A **white sequence** $w[n]$ *is defined as uncorrelated random variables such that mean* $\mu_W = 0$ *and variance is* σ_W^2.

Definition

A wide-sense stationary process $W[n]$ *is said to be white if the autocovariance is given by*

$$c_{ww}[m] = \sigma_W^2 \delta[m] \tag{38}$$

where $\delta[m] = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0 \end{cases}$ *is the unit impulse function/signal.*

Thus the autocorrelation of a white random process is

$$\gamma_{ww}[m] = \sigma_W^2 \delta[m]. \tag{39}$$

Discrete ergodic random signal statistics

The statistical properties for discrete-time ergodic random processes is summarized in the following table:

Mean value of $\{x[k]\}$,	$\bar{x} = E\{x[k]\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x[k]$
Mean square value of $\{x[k]\}$,	$\Psi_x^2 = E\{(x[k])^2\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x^2[k]$
Variance of $\{x[k]\}$,	$\sigma_x^2 = E\{(x[k]-\bar{x})^2\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (x^2[k]-\bar{x}]$
	$\sigma_x^2 = E\{(x[k])^2\} - \bar{x}^2 = \Psi_x{}^2 - \bar{x}^2$
Autocorrelation of $\{x[k]\}$,	$r_x(m) = E\{x[k]x[k+m]\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x[k]x[k+m]$
Autocovariance of $\{x[k]\}$,	$c_x(m) = E\{(x[k]-\bar{x})(x[k+m]-\bar{x})\}$
Cross - correlation of $\{x[k]\}$ & $\{y[k]\}$,	$r_{xy}(m) = E\{x[k]y[k+m]\}$
Cross - covariance of $\{x[k]\}$ & $\{y[k]\}$,	$c_{xy}(m) = E\{(x[k]-\bar{x})(y[k+m]-\bar{x})\}$
	where : m is the lag factor

Exercise

Consider $x[n]$ and $w[n]$, two discrete real-valued weakly stationary and uncorrelated (i.e. $E[x[n]w[\tilde{n}]] = E[x[n]]E[w[\tilde{n}]]$) random processes and assume that $\mu_X = E[x[n]] = 0$.
Show that $y[n] = x[n] + w[n]$ is weakly stationary and

$$\gamma_{YY}[m] = \gamma_{xx}[m] + \gamma_{ww}[m]. \tag{40}$$

Solution: We need to show that:

- mean is constant

$$\begin{aligned} \mu_Y[n] &= E[y[n]] = E[x[n] + w[n]] \\ &= E[x[n]] + E[w[n]] = \mu_W = \text{constant} \end{aligned}$$

  since by hypothesis $E[x[n]] = 0$ and $w[n]$ is weakly stationary.

- variance is constant

$$var(y[n]) = E[(y[n] - \mu_Y)^2] = \ldots \tag{41}$$

cont.d Exercise

Solution:

- autocorrelation depends on the lag $m = n - \tilde{n}$:

$$\begin{aligned}
\gamma_{YY}(n,\tilde{n}) &= E[y[n]y[\tilde{n}]] \\
&= E\left[(x[n]+w[n])(x[\tilde{n}]+w[\tilde{n}])\right] \\
&= E[x[n]x[\tilde{n}]]+E[x[n]w[\tilde{n}]]+E[w[n]x[\tilde{n}]]+E[w[n]w[\tilde{n}]] \\
&= \gamma_{xx}[m]+E[x[n]]E[w[\tilde{n}]]+E[x[\tilde{n}]]E[w[n]]+\gamma_{ww}[m] \\
&\qquad \text{since } x[n] \text{ and } w[n] \text{ are uncorrelated} \\
&= \gamma_{xx}[m]+\gamma_{ww}[m] \\
&\qquad \text{since } E[x[n]] = 0
\end{aligned}$$

Therefore, $y[n]$ is weakly stationary and $\gamma_{YY}[m] = \gamma_{xx}[m] + \gamma_{ww}[m]$.

Power density spectrum

A stationary random process is an infinite energy signal and therefore its Fourier transform does not exist. Instead,

Definition

The **power density spectrum** *of a random process gives the spectral characteristics of a random process and is defined as the Fourier transform of the auto-correlation function* $\gamma_{xx}(\tau)$

$$\Gamma_{xx}(F) = \int_{-\infty}^{\infty} \gamma_{xx}(\tau)\, e^{-j2\pi F\tau}\, d\tau. \tag{42}$$

The inverse Fourier transform is given by

$$\gamma_{xx}(\tau) = \int_{-\infty}^{\infty} \Gamma_{xx}(F)\, e^{j2\pi F\tau}\, dF. \tag{43}$$

Note that

$$\gamma_{xx}(0) = \int_{-\infty}^{\infty} \Gamma_{xx}(F)\, dF = E[X_t^2] \geq 0.$$

Since $E[X_t^2] = \gamma_{xx}(0)$ represents the average power of the random process, which is the area under $\Gamma_{xx}(F)$, then $\Gamma_{xx}(F)$ is the distribution of power as a function of frequency.

Cross-power density spectrum

Definition

Consider two jointly stationary random processes $X(t)$ and $Y(t)$, with cross-correlation $\gamma_{XY}(\tau)$, then the **cross-power density spectrum** *is defined as the Fourier transform of $\gamma_{XY}(\tau)$*

$$\Gamma_{XY}(F) = \int_{-\infty}^{\infty} \gamma_{XY}(\tau)\, e^{-j2\pi F\tau}\, d\tau \tag{44}$$

Power density spectrum of discrete random processes

For wide-sense stationary signals,

Definition

The **Power density spectrum of the discrete random process** $X[n]$ *is*

$$\Gamma_{xx}(f) = \sum_{m=-\infty}^{\infty} \gamma_{xx}[m]\, e^{-j2\pi fm} \tag{45}$$

The inverse Fourier transform gives

$$\gamma_{xx}[m] = \int_{-1/2}^{1/2} \Gamma_{xx}(f)\, e^{-j2\pi fm}\, df \tag{46}$$

and the average power is $\gamma_{xx}[0] = \int_{-1/2}^{1/2} \Gamma_{xx}(f)\, df$.

Exercise

Consider $x[n] = A\cos(2\pi f_0 n - \Phi)$ where A, f_0 are constants and Φ is uniformly distributed in $(-\pi, \pi)$ and let $w[n]$ be defined as a white noise random process. Suppose that $x[n]$ and $w[n]$ are uncorrelated.

Determine the autocorrelation $\gamma_{YY}[m]$ and power spectral density $\Gamma_{YY}(f)$ of $y[n] = x[n] + w[n]$.

Solution

In the previous exercise, we showed that $x[n]$ is wide sense stationary. Since $w[n]$ is also wide sense stationary, following Ex. 3 we can conclude

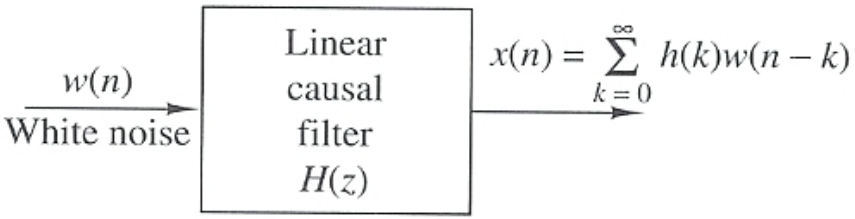
$$\gamma_{YY}[m] = \gamma_{xx}[m] + \gamma_{ww}[m] = \frac{A^2}{2}\cos(2\pi f_0 m) + \sigma_W^2 \delta[m] \qquad (47)$$

and by taking the DTFT of $\gamma_{YY}[m]$, we obtain the power spectral density of $y[n]$

$$\Gamma_{YY}(f) = \Gamma_{xx}(f) + \Gamma_{ww}(f) = \frac{A^2}{4}[\delta(f - f_0) + \delta(f + f_0)] + \sigma_W^2. \qquad (48)$$

Innovations representation of a stationary random process

Here we show that a wide-sense stationary random process $x[n]$ can be respresented as the output of a causal and causally invertible linear system $H(z)$ excited by a white noise process $w[n]$.



Conversely, if the input to the inverse causal filter is a wide-sense stationary random process $x[n]$ then the output is a white noise process $w[n]$.

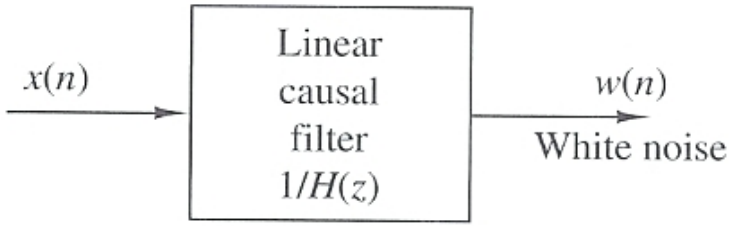


Figure 5: $1/H(z)$ is the noise whitening filter and $w[n]$ is the innovations process associated with the stationary random process $x[n]$.

Let $x[n]$ be a weakly stationary process with auto-correlation $\gamma_{xx}[m]$. The power spectral density $\Gamma_{xx}(f)$ can be obtained by substituting $z = e^{j2\pi f}$ into the $Z-$transform of $\gamma_{xx}[m]$,

$$\Gamma_{xx}(z) = \sum_{m=-\infty}^{\infty} \gamma_{xx}[m] z^{-m} \tag{49}$$

Assume that $\Gamma_{xx}(f)$ is real and continuous for all $|f| \leq 1/2$ and that $\ln \Gamma_{xx}(z)$ is analytic (possesses derivatives of all order) in an annular region that includes the unit circle, i.e.

$$r_1 < |z| < r_2 \text{ with } r_1 < 1 \text{ and } r_2 > 1. \tag{50}$$

Then one has the following Laurent series

$$\ln \Gamma_{xx}(z) = \sum_{m=-\infty}^{+\infty} \nu[m] z^{-m} \tag{51}$$

where $\{\nu[m]\}$ are the coefficients in the series expansion. Also $\nu[m]$ can be seen as the sequence with z-transform $V(z) = \ln \Gamma_{xx}(z)$.

We can evaluate Eq. (51) on the unit circle $z = e^{j2\pi f}$

$$\ln \Gamma_{xx}(f) = \sum_{m=-\infty}^{\infty} \nu[m] e^{-j2\pi fm} \tag{52}$$

so that

$$\nu[m] = \int_{-1/2}^{1/2} \ln \Gamma_{xx}(e^{j2\pi f}) e^{-j2\pi fm} df, \ \forall m \tag{53}$$

are the Fourier coefficients in the Fourier series expansion of the periodic function $\ln \Gamma_{xx}(f)$. Since $\Gamma_{xx}(f)$ is a real and even function of f, we have that $\nu[m]$ is real with $\nu[m] = \nu[-m]$, $\forall m$ and from Eq. (51) we have that

$$\begin{aligned}
\Gamma_{xx}(z) &= e^{\sum_{m=-\infty}^{+\infty} \nu[m] z^{-m}} && (54)\\
&= \underbrace{e^{\nu[0]}}_{\sigma_w^2} \underbrace{e^{\sum_{m=1}^{\infty} \nu[m] z^{-m}}}_{H(z)} \underbrace{e^{\sum_{m=1}^{+\infty} \nu[m] z^{m}}}_{H(z^{-1})} && (55)\\
&= \sigma_w^2 H(z) H(z^{-1}) && (56)
\end{aligned}$$

where $\sigma_w^2 = e^{\nu[0]}$ and

$$H(z) = e^{\sum_{m=1}^{\infty} \nu[m] z^{-m}} \tag{57}$$

is causal and stable with all poles within $|z| < r_1 < 1$.

It can be further assumed that $H(z)$ has its zeros all inside $|z| = 1$ (i.e., $H(z)$ is of minimum-phase) when $\Gamma_{xx}(z)$ has no zeros on $|z| = 1$.

Finally, if we evaluate Eq. (56) on the unit circle, we obtain

$$\Gamma_{xx}(f) = \sigma_w^2 H(f) H^*(f) \tag{58}$$

$$= \sigma_w^2 |H(f)|^2 \tag{59}$$

and therefore the weakly stationary random process $x[n]$ can be represented as the output of a causal filter $H(z)$ when excited with a white noise random process $w[n]$ of variance σ_W^2.

Rational power spectra

Suppose that the power spectral density of a weakly stationary random process $x[n]$ is a rational function

$$\Gamma_{xx}(z) = \sigma_w^2 \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}, \quad r_1 < |z| < r_2 \tag{60}$$

where the roots of $B(z)$ and $A(z)$ are inside the unit circle of the $z-$plane. Then the linear filter for generating $x[n]$ from $w[n]$ is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}}, \quad |z| > r_1 \tag{61}$$

where $\{b_k\}$ and $\{a_k\}$ are the filter coefficients that determine the location of the zeros and poles of $H(z)$, respectively.

Thus $H(z)$ is causal, stable and minimum phase and $1/H(z)$ is also causal, stable, minimum phase linear system.
Therefore the random process $x[n]$ uniquely represents the statistical properties of the innovations process $w[n]$, and vice versa.
For a linear system defined by $H(z)$ in Eq.(61), the output $x[n]$ is related to the input $w[n]$ by

$$x[n] + \sum_{k=1}^{p} a_k x[n-k] = \sum_{k=0}^{q} b_k w[n-k]. \tag{62}$$

Next,

- Autoregressive (AR) process
- Moving average (MA) process
- Autroregressive and moving average (ARMA) process

Autoregressive ($AR(p)$) process

Let $b_0 = 1, b_k = 0, k > 0$ then the linear filter

$$H(z) = \frac{1}{A(z)} \tag{63}$$

is an all-pole filter and the input-output relationship is

$$x[n] + \sum_{k=1}^{p} a_k x[n-k] = w[n]. \tag{64}$$

Consequently, the noise-whitening filter $(1/H(z))$ for generating the innovations is an all-zero filter.

Moving average ($MA(q)$) process

Let $a_k = 0, k \geq 1$ then the linear filter

$$H(z) = B(z) \tag{65}$$

is an all-zero filter and the input-output relationship is

$$x[n] = \sum_{k=0}^{q} b_k w[n-k]. \tag{66}$$

Consequently, the noise-whitening filter $(1/H(z))$ for the MA process is an all-pole filter.

Autroregressive and moving average ($ARMA(p,q)$) process

Consider the linear filter with a finite number of zeros and poles in the $z-$plane

$$H(z) = \frac{B(z)}{A(z)} \tag{67}$$

then the corresponding difference equation is given by

$$x[n] + \sum_{k=1}^{p} a_k x[n-k] = \sum_{k=0}^{q} b_k w[n-k]. \tag{68}$$

Consequently, the noise-whitening filter $1/H(z) = A(z)/B(z)$ for generating the innovations process $w[n]$ from $x[n]$ is a pole-zero filter.

Relationships between filter parameters and autocorrelation sequence

Multiplying both sides of Eq. (62) with $x[n-m]$ and applying $E[\cdot]$, we get

$$\underbrace{E[x[n]x[n-m]]}_{\gamma_{xx}[m]} + \sum_{k=1}^{p} a_k \underbrace{E[x[n-k]x[n-m]]}_{\gamma_{xx}[m-k]} = \sum_{k=0}^{q} b_k \underbrace{E[w[n-k]x[n-m]]}_{\gamma_{WX}[m-k]}. \tag{69}$$

Since $x[n] = \sum_{l=0}^{\infty} h[l]w[n-l]$, we can show that

$$E[w[n-k]x[n-m]] = E[w[n-k]\sum_{l=0}^{\infty} h[l]w[n-m-l]] \tag{70}$$

$$= \sum_{l=0}^{\infty} h[l] \underbrace{E[w(n-k)w[n-m-l]]}_{\sigma_W^2 \delta[l-(k-m)]} \tag{71}$$

since $w[n]$ is a white noise process.

Therefore Eq. (69) becomes

$$\gamma_{xx}[m] = -\sum_{k=1}^{p} a_k \gamma_{xx}[m-k] + \sigma_W^2 \sum_{k=0}^{q} h[k-m] b_k, \ \forall m \tag{72}$$

since $h[n]$ is causal, i.e. $h[n] = 0, \forall n < 0$ we have the following nonlinear relationship between $\gamma_{xx}[m]$ and the parameters $\{a_k\}, \{b_k\}$

$$\gamma_{xx}[m] = \begin{cases} -\sum_{k=1}^{p} a_k \gamma_{xx}[m-k] & m > q \\ -\sum_{k=1}^{p} a_k \gamma_{xx}[m-k] + \sigma_W^2 \sum_{k=0}^{q-m} h[k] b_{k+m} & 0 \le m \le q \\ \gamma_{xx}[-m] & m < 0 \end{cases} \tag{73}$$

Exercise

Consider $x[n]$ to be a weakly stationary AR process with power density spectrum given by

$$\Gamma_{xx}(z) = \frac{\sigma_w^2}{|A(z)|^2} \tag{74}$$

where $\sigma_w^2 = 25$ is the variance of the input sequence and $A(z) = 1 - z + 0.5z^2$.

a. Determine the difference equation for generating the AR process when the excitation is white noise.

b. Determine the system function for the whitening filter.

Linear prediction and optimum linear filters

- Introduction
- Forward and backward linear prediction
- Solution of the normal equations
- Wiener filters for filtering and prediction

Introduction

Design of filters to perform signal estimation comes about in design of communication systems, control systems, geophysics.
Problem Optimum filter design from a statistical viewpoint. Here we consider

- Linear filters
- "Best" $\leftrightarrow$ minimization mean squared error(MSE)
- Second order statistics (autocorrelation and crosscorrelation) of a stationary process.

Forward linear prediction

Let $x[n]$ be a stationary random process.
Question: How can we predict a future value of a stationary random process from observation of past values of the process?

One-step forward linear predictor of order p, $\hat{x}[n]$, gives the prediction of the value of $x[n]$ by a weighted linear combination of the past values $x[n-1], x[n-2], \cdots, x[n-p]$. It is defined by

$$\hat{x}[n] = -\sum_{k=1}^{p} a_p[k]x[n-k] \tag{75}$$

where $\{-a_p[k]\}_{k=1}^{p}$ are the weights in the linear combination and are called the **prediction coefficients** of the one-step forward linear predictor of order p.

Example

Take $p = 3$, *and let* $n = 0$.
Given $x[-3], x[-2], x[-1]$, *we would like to find* $a_p[k], k = 1, 2, 3$ *such that*

$$\hat{x}[0] = -a_p[1]x[-1] - a_p[2]x[-2] - a_p[3]x[-3] \tag{76}$$

is a forward linear predictor of $x[0]$.

Forward prediction error

The difference between the value $x[n]$ and $\hat{x}[n]$,the predicted value of $x[n]$, is called the **forward prediction error:**

$$\begin{aligned} f_p[n] &= x[n] - \hat{x}[n] \\ &= x[n] + \sum_{k=1}^{p} a_p[k]x[n-k] \qquad (77) \\ &= \sum_{k=0}^{p} a_p[k]x[n-k] \qquad (78) \end{aligned}$$

where $a_p[0] = 1$.

Prediction-error filter

Linear prediction is equivalent to linear filtering where the predictor is embedded in the linear filter (see Figure 6)

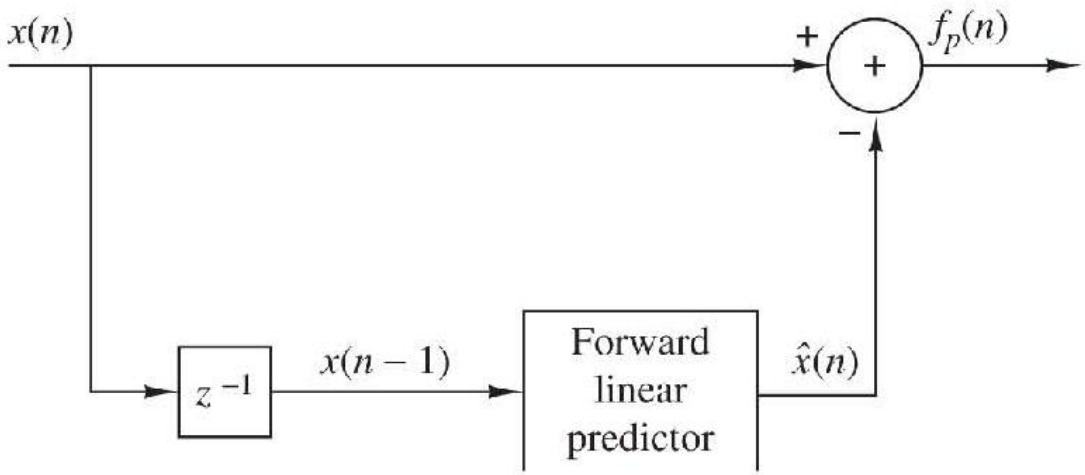


Figure 6: Forward linear prediction

where $\{x[n]\}$ is the input sequence, $\{f_p[n]\}$ is the output sequence.

Prediction-error filter

An equivalent realisation for the prediction filter is given by the direct-form FIR filter illustrated in Fig. 7,

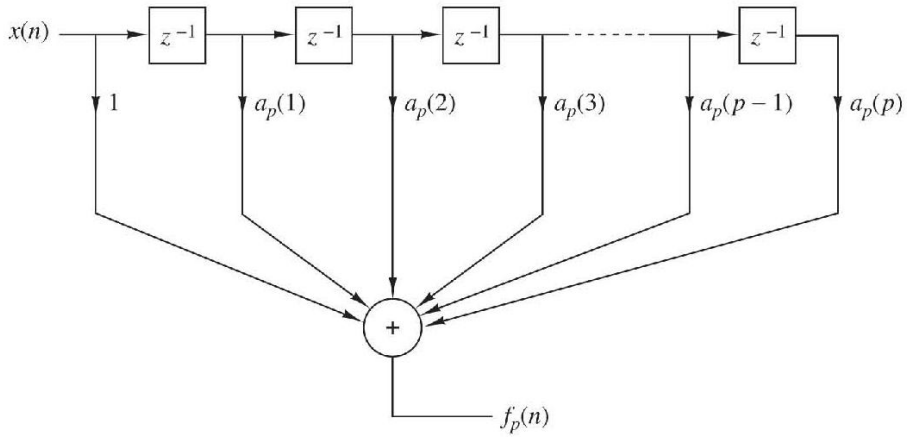


Figure 7: Prediction-error filter is $\{a_p[k]\}_{k=1}^{p}, a[0] = 1$

The $z-$transform of $f_p[n]$ in Eq. (78) is given by

$$F_p(z) = A_p(z) \cdot X(z) \Rightarrow A_p(z) = \frac{F_p(z)}{X(z)} = \frac{F_p(z)}{F_0(z)} \tag{79}$$

where the system function is

$$A_p(z) = 1 + \sum_{k=1}^{p} a_p[k] z^{-k}. \tag{80}$$

Mean-square forward linear prediction error

The mean-square value of the forward linear prediction error $f_p[n]$ is

$$\mathcal{E}_p^f = E[|f_p[n]|^2] \tag{81}$$

$$= \gamma_{xx}[0] + 2Re\left\{\sum_{k=1}^{p} a_p^*[k]\gamma_{xx}[k]\right\} + \sum_{k=1}^{p}\sum_{l=1}^{p} a_p^*[l]a_p^*[k]\gamma_{xx}[l-k] \tag{82}$$

Note that $\mathcal{E}_p^f$ is a quadratic function of the predictor coefficients $a_p[k]$.

Minimum mean-square prediction error

The minimization of $\mathcal{E}_p^f$, leads to the following set of linear equations

$$\gamma_{xx}[l] = -\sum_{k=1}^{p} a_p[k]\gamma_{xx}[l-k], \quad l = 1, 2, \ldots, p \tag{83}$$

which are called the **normal equations** for the coefficients of the linear predictor. The minimum mean squared prediction error is simply

$$\min_{a_p[k]} \mathcal{E}_p^f = \varepsilon_p^f = \gamma_{xx}[0] + \sum_{k=1}^{p} a_p[k]\gamma_{xx}[-k] \tag{84}$$

Proof: ■

Backward linear prediction

Suppose the data sequence $x[n], x[n-1], \ldots, x[n-(p-1)]$ is obtained from a stationary random process.
Question: How can we predict the value $x[n-p]$ from observation of past values of the process?

One-step backward linear predictor of order p, $\hat{x}[n-p]$, gives the prediction of the value of $x[n-p]$ by a weighted linear combination of the past values $x[n-1], x[n-2], \cdots, x[n-(p-1)]$. It is defined by

$$\hat{x}[n-p] = -\sum_{k=1}^{p} b_p[k]x[n-k] \tag{85}$$

where $\{-b_p[k]\}_{k=1}^{p}$ are called the **prediction coefficients** of the one-step backward linear predictor of order p

Example

Take $p = 3$, and let $n = 0$.
Given $x[-2], x[-1], x[0]$, we would like to find $b_p[k], k = 0, 1, 2$ such that

$$\hat{x}[-3] = -b_p[0]x[0] - b_p[1]x[-1] - b_p[2]x[-2] \tag{86}$$

is a backward linear predictor of $x[-3]$.

Backward prediction error

The difference between the value $x[n-p]$ and $\hat{x}[n-p]$,the predicted value of $x[n-p]$, is called the **backward prediction error:**

$$g_p[n] = x[n-p] - \hat{x}[n-p] \quad (87)$$

$$= x[n-p] + \sum_{k=0}^{p-1} b_p[k]x[n-k] \quad (88)$$

$$= \sum_{k=0}^{p} b_p[k]x[n-k] \quad (89)$$

where $b_p[p] = 1$.

In the $z-$domain Eq. (89), becomes

$$G_p(z) = B_p(z)X(z) \Rightarrow B_p(z) = \frac{G_p(z)}{X(z)} = \frac{G_p(z)}{G_0(z)} = \sum_{k=0}^{p} b_p[k]z^{-k} \quad (90)$$

The weighting coefficients in the backward linear predictor are complex conjugates of the coefficients for the forward linear predictor but in reverse order, i.e.

$$b_p[k] = a_p^*[p-k], k = 0, 1, \ldots, p \quad (91)$$

therefore

$$B_p(z) = \sum_{k=0}^{p} a_p^*[p-k]z^{-k} = z^{-p}\sum_{k=0}^{p} a_p^*[k]z^{k} = z^{-p}A_p^*(z^{-1}) \quad (92)$$

implies that the zeros of FIR filter with system function $B_p(z)$ are conjugate reciprocals of the zeros of $A_p(z)$.
$B_p(z)$ is called the reciprocal or reverse polynomial of $A_p(z)$.

Minimum mean-square backward linear prediction error

The mean-square value of the backward linear prediction error $g_p[n]$ is

$$\mathcal{E}_p^g = E[|g_p[n]|^2] \tag{93}$$

where

$$g_p[n] = x[n-p] + \sum_{k=1}^{p} a_p^*[k]x[n-p+k] \tag{94}$$

Note that $\mathcal{E}_p^g$ is a quadratic function of the predictor coefficients $a_p[k]$ and the minimization of $\mathcal{E}_p^g$ leads to the same set of linear equations given in Eq. (83). The minimum mean square backward linear prediction error is

$$\varepsilon_p^g = \min_{a_p[k]} \mathcal{E}_p^g = \varepsilon_p^f \tag{95}$$

Relationship of an AR process to linear prediction

Recall that in an $AR(p)$ process, the autocorrelation sequence is related to the parameters by the following:

$$\gamma_{xx}[m] = \begin{cases} -\sum_{k=1}^{p} a_k \gamma_{xx}[m-k] & m > 0 \\ -\sum_{k=1}^{p} a_k \gamma_{xx}[m-k] + \sigma_w^2 & m = 0 \\ \gamma_{xx}[-m] & m < 0 \end{cases} \tag{96}$$

which is called the Yule-Walker equations or in matrix form

$$\begin{bmatrix} \gamma_{xx}[0] & \gamma_{xx}[-1] & \gamma_{xx}[-2] & \cdots & \gamma_{xx}[-p] \\ \gamma_{xx}[1] & \gamma_{xx}[0] & \gamma_{xx}[-1] & \cdots & \gamma_{xx}[-(p-1)] \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{xx}[p] & \gamma_{xx}[p-1] & \gamma_{xx}[p-2] & \cdots & \gamma_{xx}[0] \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{97}$$

where σ_w^2 is variance of the white noise process

Note that if the underlying process $x[n]$ is $AR(p)$ then the parameters $\{a_k\}$ of the $AR(p)$ process are exactly the predictor coefficients $a_p[k]$ of the pth-order predictor (ie. compare Eq. (96) and the Normal Equations in Eq. (83).
The minimium mean-square error (MMSE) in the pth order predictor is equal to the variance of the white noise process, i.e.,

$$\varepsilon_p^f = \sigma_w^2. \tag{98}$$

Therefore the prediction-error filter with system function $A_p(z)$ is a noise whitening filter which produces the innovations process $w[n]$.

Solution of the normal equations

Consider the following compact form of the normal equations

$$\sum_{k=0}^{p} a_p(k)\gamma_{xx}(l-k) = 0, \quad l = 1, 2, \ldots, p \tag{99}$$

with $a_p(0) = 1$.
The resulting MMSE is given by Eq. (84) and Eq. (99) can be expressed by the following set of augmented normal equations

$$\sum_{k=0}^{p} a_p(k)\gamma_{xx}(l-k) = \begin{cases} \varepsilon_p^f & l = 0 \\ 0 & l = 1, 2, \ldots, p \end{cases} \tag{100}$$

The minimization of the mean-square value of the forward prediction error comes down to solving the normal equations for the prediction coefficients.

Matrix form of augmented normal equations

Let $p = 3$, then Eq. (100) becomes

$$\gamma_{xx}(l)+a_3(1)\gamma_{xx}(l-1)+a_3(2)\gamma_{xx}(l-2)+a_3(3)\gamma_{xx}(l-3)=\begin{cases}\varepsilon_p^f & l=0\\ 0 & l=1,2,3\end{cases} \quad (101)$$

or in matrix form

$$\begin{bmatrix}\gamma_{xx}[0] & \gamma_{xx}[-1] & \gamma_{xx}[-2] & \gamma_{xx}[-3]\\ \gamma_{xx}[1] & \gamma_{xx}[0] & \gamma_{xx}[-1] & \gamma_{xx}[-2]\\ \gamma_{xx}[2] & \gamma_{xx}[1] & \gamma_{xx}[0] & \gamma_{xx}[-1]\\ \gamma_{xx}[3] & \gamma_{xx}[2] & \gamma_{xx}[1] & \gamma_{xx}[0]\end{bmatrix}\begin{bmatrix}1\\ a_3[1]\\ a_3[2]\\ a_3[3]\end{bmatrix}=\begin{bmatrix}\varepsilon_p^f\\ 0\\ 0\\ 0\end{bmatrix} \quad (102)$$

where $\gamma_{xx}[-m] = \gamma_{xx}[m]$.

Note the Toeplitz[1] symmetry property of the autocorrelation matrix.

[1] A matrix is Toeplitz if all the diagonals are constant.

Two computationally efficient iterative solving methods are:

- Levinson-Durbin algorithm : Due to Levinson in 1947 and modified by Durbin in 1959, this algorithm is suitable for serial processing and has a computation complexity of $O(p^2)$.
- Schur Algorithm : Due to Schur in 1917, this algorithm computes the reflection coefficients[2] in $O(p^2)$ but with parallel processors, the computations can be performed in $O(p)$ time.

Both algorithms exploit the Toeplitz symmetry property of the autocorrelation matrix.

[2] Reflection coefficients are found in lattice structures of FIR filters.

Levinson-Durbin algorithm

The method begins with a predictor of order $p = 1$ and then increases the order recursively, using the lower-order solutions to obtain the solution of the next higher order.
For $m = 1$,

$$\gamma_{xx}[0] + a_1[1]\gamma_{xx}[-1] = \varepsilon_1^f \quad (103)$$

$$\gamma_{xx}[1] + a_1[1]\gamma_{xx}[0] = 0 \Rightarrow a_1[1] = -\frac{\gamma_{xx}[1]}{\gamma_{xx}[0]} \quad (104)$$

and substituting in Eq.(103) to obtain

$$\varepsilon_1^f = \gamma_{xx}[0]\left(1 + a_1[1]\frac{\gamma_{xx}[-1]}{\gamma_{xx}[0]}\right) = \gamma_{xx}[0]\left(1 - |a_1[1]|^2\right) \quad (105)$$

where $\gamma_{xx}[-1] = \gamma_x x[1]$.

The next step is to solve for the second-order ($p = 2$)predictor coefficients $\{a_2[1], a_2[2]\}$ in terms of $a_1[1]$:

$$\begin{aligned} \gamma_{xx}[0] + a_2[1]\gamma_{xx}[-1] + a_2[2]\gamma_{xx}[-2] &= \varepsilon_2^f \\ \gamma_{xx}[1] + a_2[1]\gamma_{xx}[0] + a_2[2]\gamma_{xx}[-1] &= 0 \\ \gamma_{xx}[2] + a_2[1]\gamma_{xx}[1] + a_2[2]\gamma_{xx}[0] &= 0 \end{aligned} \tag{106}$$

i.e., first solve

$$\begin{aligned} a_2[1]\gamma_{xx}[0] + a_2[2]\gamma_{xx}[-1] &= -\gamma_{xx}[1] \\ a_2[1]\gamma_{xx}[1] + a_2[2]\gamma_{xx}[0] &= -\gamma_{xx}[2] \end{aligned} \tag{107}$$

By using Eq. (104)

$$\begin{aligned} \Rightarrow a_2[2] &= -\frac{\gamma_{xx}[2] + a_1[1]\gamma_{xx}[1]}{\gamma_{xx}[0](1 - |a_1[1]|^2)} = -\frac{\gamma_{xx}[2] + a_1[1]\gamma_{xx}[1]}{\varepsilon_1^f} \\ a_2[1] &= a_1[1] + a_2[2]a_1[1] \end{aligned}$$

then substitute in Eq. (106) to obtain ε_2^f.

Properties of the Linear Prediction-Error Filters

- Minimum-phase property of the forward prediction-error filter
- Maximum-phase property of the backward prediction-error filter
- Whitening property
- Orthogonality of the backward prediction errors

Wiener filter for Filtering and Prediction

Consider an input signal $x[n] = s[n] + w[n]$ where $s[n]$ is a desired signal and $w[n]$ is undesired noise or interference.

Problem: Design a linear filter $H(z)$ that eliminates or filters out the undesired noise or interference while keeping the characteristics of the desired signal $s[n]$ and so that the output of the filter, $y(n)$, approximates some specified (reference) desired signal $d(n)$, see Fig. 8.

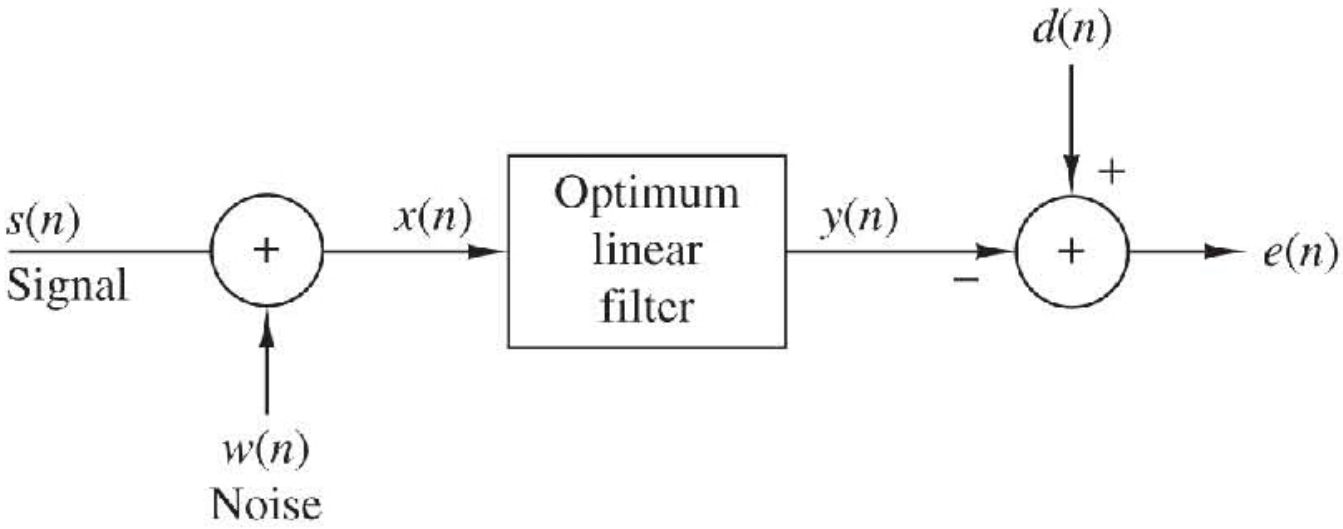


Figure 8: Model for linear estimation problem.

Three types of linear estimation problems

The linear estimation problem is referred to as

- filtering if reference signal $d[n] = s[n]$, i.e. design $H(z)$ to suppress $w[n]$ and $y[n] \approx s[n]$
- signal prediction if reference signal $d[n] = s[n+D], D > 0$. Note that this is different than the prediction considered before where $d[n] = x[n+D], D \geq 0$.
- signal smoothing if reference signal $d[n] = s[n-D], D > 0$

In what follows, we assume that $s[n], w[n], d[n]$

- have a zero mean
- are wide-sense stationary.

Wiener filter

The optimum linear filter $h[n]$ is called the **Wiener filter** in the sense of minimum mean-square error

$$\varepsilon_M^h = \min_{h[n]} \mathcal{E}_M^h \tag{108}$$

where $\mathcal{E}_M^h = E\left[|e[n]|^2\right]$ and $e[n] = d[n] - y[n]$.

Wiener filters may be designed as FIR or IIR. In the case of IIR, it is assumed that the input data $x[n]$ is available over the infinite past.

Designing FIR Wiener filters

Suppose that the filter must be of length M with coefficients $\{h[k]\}_{k=0}^{M-1}$, then the output $y[n]$ depends on the finite data set $x[n], x[n-1], \ldots, x[n-(M-1)]$,

$$y[n] = \sum_{k=0}^{M-1} h[k]x[n-k] \tag{109}$$

The optimal Wiener FIR filter is found by minimizing the mean-square value of the error between the desired output $d[n]$ and $y[n]$ is

$$\mathcal{E}_M^h = E\left[|e[n]|^2\right] = E\left[\left|d[n] - \sum_{k=0}^{M-1} h[k]x[n-k]\right|^2\right] \tag{110}$$

Wiener-Hopf Equation

Since Eq. (110) is a quadratic function of the filter coefficients $\mathbf{h} = \left[h[0], h[1], \ldots, h[M-1]\right]^t$, then the minimization of $\mathcal{E}_M^h$ comes down to solving the linear system of equations

$$\sum_{k=0}^{M-1} h[k]\gamma_{xx}[l-k] = \gamma_{dx}[l], \quad l = 0, \ldots, M-1 \tag{111}$$

where

- $\gamma_{xx}[l]$ is the autocorrelation of the input sequence $x[n]$
- $\gamma_{dx}[l]$ is the crosscorrelation between the desired sequence $d[n]$ and the input sequence $x[n], n = 0, \ldots, M-1$.

Eq. (111) is known as **Wiener-Hopf** equation and corresponds to the normal equation in the context of linear prediction.

Matrix form of the Wiener-Hopf equation

Eq. (111) can be written in matrix form as

$$\mathbf{\Gamma}_M \mathbf{h}_M = \boldsymbol{\gamma}_d \tag{112}$$

where

- $\mathbf{\Gamma}_M$ is an $M \times M$ (Hermitian) Toeplitz matrix with elements $\Gamma_{lk} = \gamma_{xx}[l-k]$ and
- $\boldsymbol{\gamma}_d$ is an $M \times 1$ crosscorrelation vector with elements $\gamma_{dx}[l]$, $l = 0, 1, \ldots, M-1$.

Therefore the optimum filter coefficients are given by

$$\mathbf{h}_{opt} = \mathbf{\Gamma}_M^{-1} \boldsymbol{\gamma}_d \tag{113}$$

and the resulting MSE using the Wiener filter is

$$\varepsilon_M^h = \min_{\mathbf{h}_M} \mathcal{E}_M^h = \sigma_d^2 - \sum_{k=0}^{M-1} h_{opt}[k] \gamma_{dx}^*[k] \tag{114}$$

$$= \sigma_d^2 - \boldsymbol{\gamma}_d^{*t} \mathbf{\Gamma}_M^{-1} \boldsymbol{\gamma}_d \tag{115}$$

where $\sigma_d^2 = E[|d[n]|^2]$

Wiener-Hopf equations for filtering

If we are considering the problem of filtering $d[n] = s[n]$ and we further assume $s[n]$ and $w[n]$ are uncorrelated random sequences, i.e.

$$\gamma_{xx}[k] = \gamma_{ss}[k] + \gamma_{ww}[k] \quad (116)$$

$$\gamma_{dx}[k] = \gamma_{ss}[k] \quad (117)$$

with known $\gamma_{ss}[k], \gamma_{ww}[k]$, then the Wiener-Hopf or normal equations are given by

$$\sum_{k=0}^{M-1} h[k]\Big(\gamma_{ss}[l-k] + \gamma_{ww}[l-k]\Big) = \gamma_{ss}[l], \quad l = 0, \ldots, M-1 \quad (118)$$

Wiener-Hopf equations for prediction

If we are considering the problem of prediction $d[n] = s[n+D], D > 0$ and we further assume $s[n]$ and $w[n]$ are uncorrelated random sequences, i.e.

$$\gamma_{xx}[k] = \gamma_{ss}[k] + \gamma_{ww}[k] \quad (119)$$

$$\gamma_{dx}[k] = \gamma_{ss}[l+D] \quad (120)$$

with known $\gamma_{ss}[k], \gamma_{ww}[k]$, then the Wiener-Hopf equations for prediction filter are

$$\sum_{k=0}^{M-1} h[k]\Big(\gamma_{ss}[l-k] + \gamma_{ww}[l-k]\Big) = \gamma_{ss}[l+D], \quad l = 0, \ldots, M-1 \quad (121)$$

Note that the correlation matrix to be inverted is Toeplitz and the Levinson-Durbin algorithm may be used to solve for the optimum filter coeffiicents.

Example: FIR Wiener filter design

Consider an input signal $x[n] = s[n] + w[n]$ whereby $s[n]$ is an $AR(1)$ process satisfying the following difference equation

$$s[n] - 0.6s[n-1] = v[n] \tag{122}$$

where $v[n]$ is a white noise random process with variance $\sigma_v^2 = 0.64$, statistically independent of the white noise random process $w[n]$ with variance $\sigma_w^2 = 1$.

Problem: Design an FIR Wiener filter of length $M = 2$ which estimates $s[n]$.

Solution: Estimatimg $s[n]$ is equivalent to a filtering problem $\Rightarrow d[n] = s[n]$ and therefore $\gamma_{dx}[k] = \gamma_{ss}(k)$.

Note that $s[n]$ is an $AR(1)$ process and can be obtained by exciting a single-pole filter with system function $H(z) = \frac{1}{1-0.6z^{-1}}$ by white noise $v[n]$. From Eq. (56), the power spectral density of $s[n]$

$$\begin{aligned} \Gamma_{ss}(z) &= \sigma_v^2 H(z)H(z^{-1}) && (123)\\ &= \sigma_v^2 \frac{1}{1-0.6z^{-1}} \frac{1}{1-0.6z} && (124)\\ &= \frac{\sigma_v^2}{1-0.6^2} \frac{1-0.6^2}{(1+0.6^2)-2\times 0.6(z+z^{-1})}. && (125) \end{aligned}$$

By taking the inverse Z-transform, we get the corresponding autocorrelation function

$$\gamma_{ss}[m] = 0.6^{|m|} \text{ and } \gamma_{xx}[m] = \gamma_{ss}[m] + \gamma_{ww}[m] = 0.6^{|m|} + \delta[m]. \quad (126)$$

which leads to solving the following normal equations:

$$\begin{bmatrix} \gamma_{xx}[0] & \gamma_{xx}[-1] \\ \gamma_{xx}[1] & \gamma_{xx}[0] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix} = \begin{bmatrix} \gamma_{ss}[0] \\ \gamma_{ss}[1] \end{bmatrix} \tag{127}$$

$$\Leftrightarrow \begin{bmatrix} 2 & 0.6 \\ 0.6 & 2 \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix} = \begin{bmatrix} 1 \\ 0.6 \end{bmatrix} \tag{128}$$

Therefore $h_{opt}[0] = 0.451,\ h_{opt}[1] = 0.165$ and hence

$$\varepsilon_2^h = \min \mathcal{E}_2^h = \gamma_{ss}[0] - \begin{bmatrix} \gamma_{ss}[0]\gamma_{ss}[1] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix} = 0.45 \tag{129}$$

Orthogonality principle in linear mean-square estimation

Let $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2$ be three given vectors and c_1, c_2 be two scalars such that

$$\widehat{\mathbf{v}}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2. \tag{130}$$

For which values of c_1, c_2 is the distance $\mathbf{v}_e = \mathbf{v}_0 - \widehat{\mathbf{v}}_0$ minimized?

It is well known that the smallest distance is achieved when c_1 and c_2 are chosen such that $\mathbf{v}_e$ is orthogonal (i.e.perpendicular) to both $\mathbf{v}_1$ and $\mathbf{v}_2$.

Consider the output of the filter, which is the estimate

$$\hat{d}[n] = \sum_{k=0}^{M-1} h[k]x[n-k] \tag{131}$$

and is a vector in the subspace spanned by the data

$$x[n-k], k = 0, 1, \cdots, M-1.$$

The error of estimation $e[n] = d[n] - \hat{d}[n]$ is a vector from $d[n]$ to $\hat{d}[n]$ (see Fig. 9) and the orthogonality principle states that the mean-square error $\mathcal{E}_M^h$ is minimized when h[k] are selected such that the error is orthogonal to each of the data ponts in the estimate

$$E[e[n]x[n-k]] = 0, k = 0, 1, 2, \cdots, M-1 \tag{132}$$

where $e[n] = d[n] - \sum_{k=0}^{M-1} h[k]x[n-k]$. Therefore, the minimum MSE is

$$\varepsilon_M = \min \mathcal{E}_M^h = E[e[n]d[n]]. \tag{133}$$

The optimum FIR Wiener filter or the solution to the Wiener-Hopf equations is unique if and only if $\mathbf{\Gamma}_{xx}$ in Eq.(112) is non-singular.

Geometric interpretation of the linear MSE problem

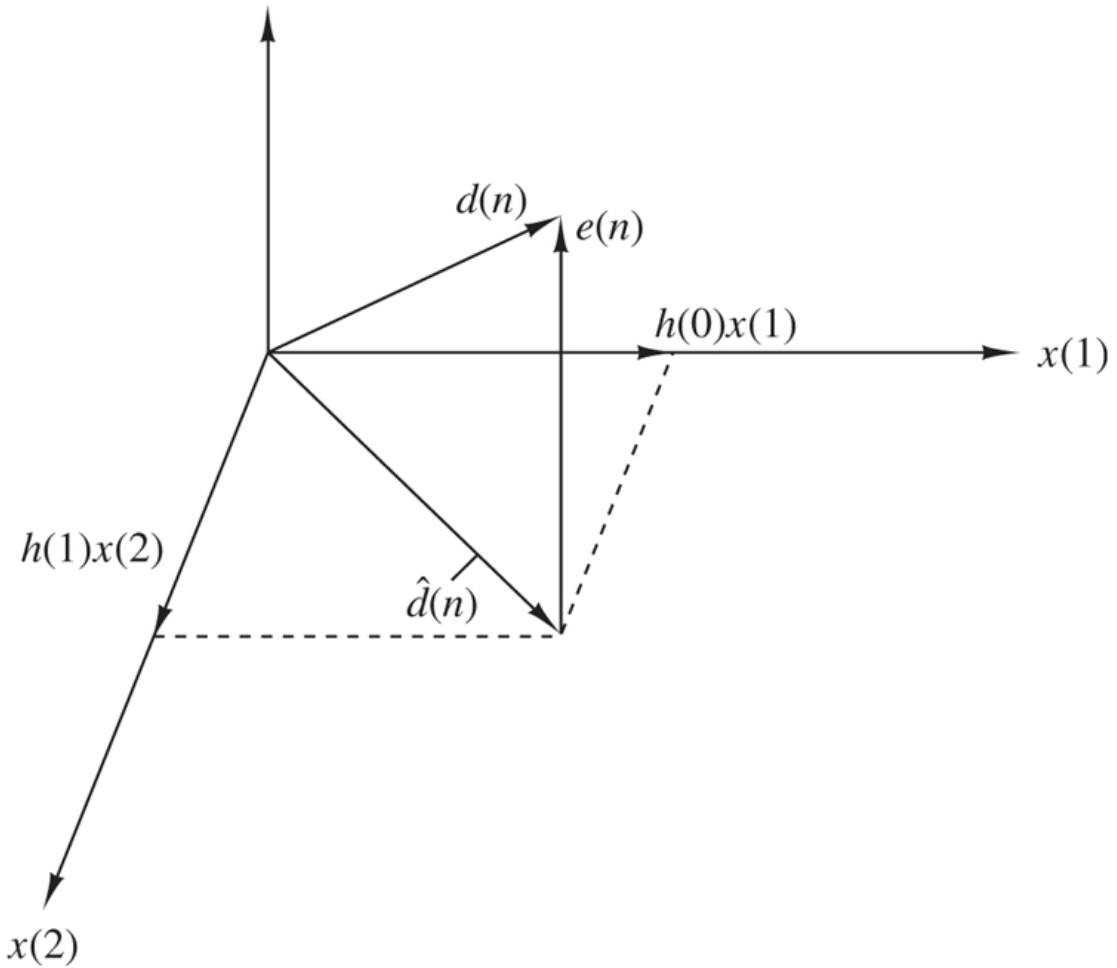


Figure 9: Geometric interpretation of the linear MSE problem.

IIR Wiener filter

When M goes to infinity, the FIR filter $H(z)$ becomes an IIR *causal* filter

$$H(z) = \sum_{k=0}^{\infty} h(k) z^{-k} \quad \Rightarrow \quad \hat{d}[n] = \sum_{k=0}^{\infty} h[k] x[n-k].$$

Then the mean-square value of the error $e(n) = d(n) - \hat{d}(n)$ is

$$\mathcal{E}_{\infty} = E[|e[n]|^2] = E\left[\left|d[n] - \sum_{k=0}^{\infty} h[k] x[n-k]\right|^2\right],$$

which is a quadratic function in the unit impulse response $\{h[k]\}$ of $H(z)$.

Applying the orthogonality principle, i.e., $E[e[n]x[n-m]] = 0,\ \forall m \geq 0$, leads to the following *Wiener-Hopf* equations:

$$\sum_{k=0}^{\infty} h[k]\gamma_{xx}[m-k] = \gamma_{dx}[m],\ \forall m \geq 0. \tag{134}$$

With $h_{opt}[k]$ as the solution, the corresponding minimum MSE is

$$\varepsilon_{\infty} = \min \mathcal{E}_{\infty} = \sigma_d^2 - \sum_{k=0}^{\infty} h_{opt}[k]\gamma_{dx}[k]. \tag{135}$$

Noting that (134) holds only for $m > 0$, the Wiener-Hopf equation can not be obtained directly with z-transform.

How do we solve the problem?

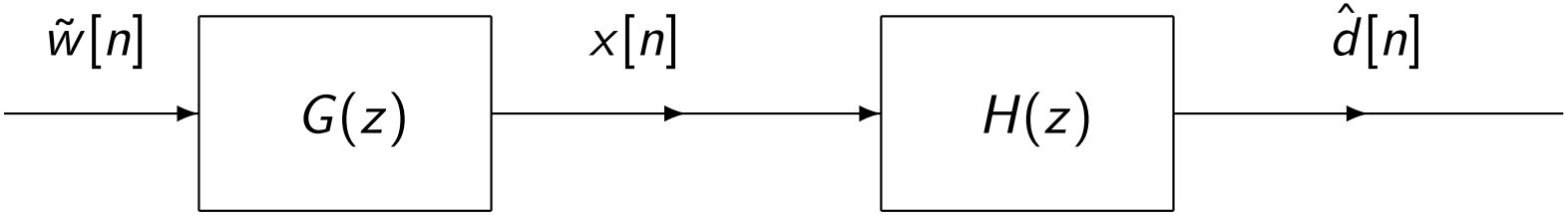


Figure 10: $Q(z) = H(z)G(z)$ is overall system, $\tilde{w}[n]$ is an innovation process, $x[n]$ stationary process, $\hat{d}[n] = \sum_{k=0}^{\infty} q[k]\tilde{w}[n-k]$.

Recall: a stationary random process $x[n]$ can be represented by the output of a minimum-phase system $G(z)$ excited with an innovations process (white sequence) $\tilde{w}[n]$ of variance $\sigma_{\tilde{w}}^2 = 1$, where $G(z)$ can be obtained from the spectral factorization of $\Gamma_{xx}(z)$ (the z-transform of $\gamma_{xx}[m]$),

$$\Gamma_{xx}(z) = \sigma_{\tilde{w}}^2 G(z)G(z^{-1}). \quad (136)$$

Therefore, by injecting $\tilde{w}[n]$ in the overall system

$$Q(z) = H(z)G(z) = \sum_{k=0}^{\infty} q[k]z^{-k} \tag{137}$$

we obtain the output

$$\hat{d}[n] = \sum_{k=0}^{\infty} q[k]\tilde{w}[n-k]. \tag{138}$$

To find the optimum $q[k]$, we apply the orthogonality principle, which leads to the following *Wiener-Hopf* equations:

$$\sum_{k=0}^{\infty} q[k]\gamma_{\tilde{w}}[m-k] = \gamma_{d\tilde{w}}[m], \ \forall m \geq 0 \tag{139}$$

where $\gamma_{\tilde{w}\tilde{w}}[m-k] = \delta[m-k]$ since $\tilde{w}[n]$ is white with unit variance and $\gamma_{d\tilde{w}}[m] = E[d[n]\tilde{w}[n-m]]$. So, $q[m] = \gamma_{d\tilde{w}}[m]$, $\forall m \geq 0$.

Hence

$$Q(z) = \sum_{k=0}^{+\infty} q[k]z^{-k} = \sum_{k=0}^{+\infty} \gamma_{d\tilde{w}}[k]z^{-k} = [\Gamma_{d\tilde{w}}(z)]_+$$

$\gamma_{dx}[m]$, as assumed, is available but not $\gamma_{d\tilde{w}}[m]$. Since $\tilde{w}[n]$ is the output of $1/G(z)$ excited with $x[n]$, it then can be shown that

$$\Gamma_{d\tilde{w}}(z) = \frac{\Gamma_{dx}(z)}{G(z^{-1})}. \tag{140}$$

Finally, the optimum causal IIR Wiener filter is given by

$$H_{opt}(z) = \frac{Q(z)}{G(z)} = \frac{\left[\frac{\Gamma_{dx}(z)}{G(z^{-1})}\right]_+}{G(z)}. \tag{141}$$

Example: IIR Wiener filter design

Consider a signal $x[n] = s[n] + w[n]$, where $s[n]$ is an AR(1) process given by

$$s[n] - 0.6s[n-1] = v[n] \tag{142}$$

where $v[n]$ is a white signal with $\sigma_v^2 = 0.64$, statistically independent of the white noise $w[n]$ of $\sigma_w^2 = 1$.

Problem: Determine the optimum causal IIR Wiener filter to estimate $s[n]$.

Solution: Once again, estimating $s[n]$ is a filtering problem $\Rightarrow d[n] = s[n]$ and therefore, $\gamma_{dx}[m] = \gamma_{ss}[m]$.
Noting

$$\Gamma_{dx}(z) = \Gamma_{ss}(z) = \frac{\sigma_v^2}{(1-0.6z^{-1})(1-0.6z)} \tag{143}$$

one has

$$\begin{aligned} \Gamma_{xx}(z) &= \Gamma_{ss}(z) + \Gamma_{ww}(z) = \Gamma_{ss}(z) + 1 \\ &= \frac{1.8(1-\frac{1}{3}z^{-1})(1-\frac{1}{3}z)}{(1-0.6z^{-1})(1-0.6z)} = G(z)G(z^{-1}) \end{aligned} \tag{144}$$

where the minimum phase $G(z)$ is

$$G(z) = \frac{\sqrt{1.8}(1-\frac{1}{3}z^{-1})}{(1-0.6z^{-1})}. \tag{145}$$

Hence,

$$\begin{aligned}\frac{\Gamma_{dx}(z)}{G(z^{-1})} &= \frac{1}{\sqrt{1.8}}\frac{0.64}{(1-0.6z^{-1})(1-\frac{1}{3}z)} \\ &= \frac{1}{\sqrt{1.8}}\frac{0.8}{1-0.6z^{-1}} + \frac{1}{\sqrt{1.8}}\frac{0.266z}{1-\frac{1}{3}z} \end{aligned} \tag{146}$$

and

$$\left[\frac{\Gamma_{dx}(z)}{G(z^{-1})}\right]_+ = \frac{1}{\sqrt{1.8}}\frac{0.8}{1-0.6z^{-1}}. \tag{147}$$

Therefore,

$$H_{opt}(z) = \frac{\left[\frac{\Gamma_{dx}(z)}{G(z^{-1})}\right]_+}{G(z)} = \frac{4/9}{1-\frac{1}{3}z^{-1}} \tag{148}$$

and $h_{opt}[k] = \frac{4}{9}(\frac{1}{3})^k u[k]$, where $u[k] = \begin{cases} 1 & k \geq 0 \\ 0 & k < 0 \end{cases}$ is the unit step function.

Once we have $\gamma_{dx}[m] = \gamma_{ss}[m] = 0.6^{|m|}$, we can compute the minimum MSE using (135)

$$\varepsilon_\infty = 0.444 \tag{149}$$

The optimum causal IIR Wiener filter is only slightly better than the optimal two-tap FIR Wiener filter obtained previously where $\varepsilon_2 = 0.45$.

Adaptive Filters

- Introduction
- Applications of Adaptive Filters
- Adaptive Direct-form FIR filters: LMS algorithm

Introduction

So far, we have considered designing filters based on second order statistics of the signals. Recall that a Wiener filter requires the autocorrelation sequence.
In many DSP applications, second order statistics cannot be specified *a priori*. So filters with adjustable coefficients, or *adaptive filters*, will be used.
These filters include algorithms that enable the filter coefficients to adapt to the signal characteristics. Two basic algorithms for adaptive filtering are:

- Least-Mean-Square (LMS) algorithm which is based on a gradient optimization for determining the coefficients
- Recursive least-squares (RLS) algorithm which includes both direct-form FIR and lattice structrues realizations.

Direct form adaptive filter

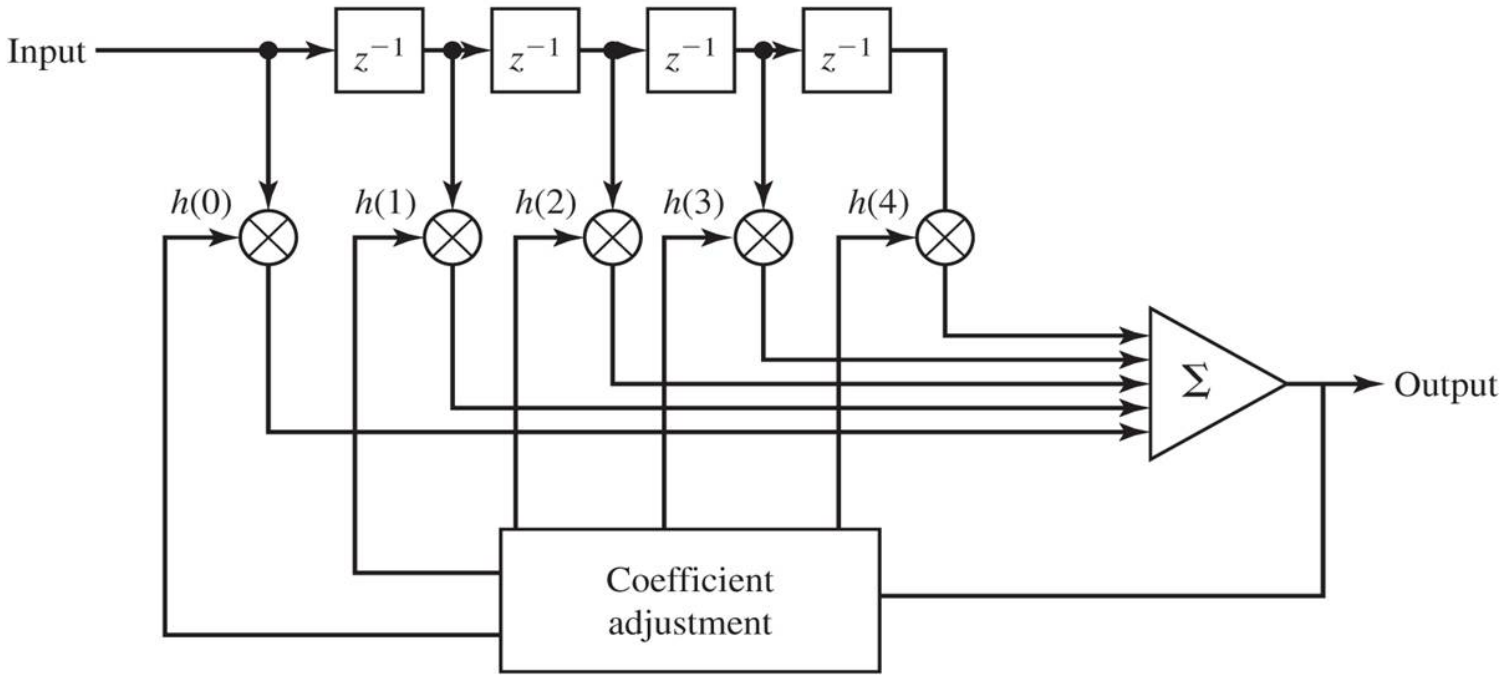


Figure 11: Direct-form adaptive FIR filter

Performance measures

In adaptive filtering the criterion for optimizing the adjustable filter parameters

- must provide a meaningful measure of filter performance
- must result in a practically realizable algorithm

Two criteria that provide good measures of performance in adaptive filtering applications are

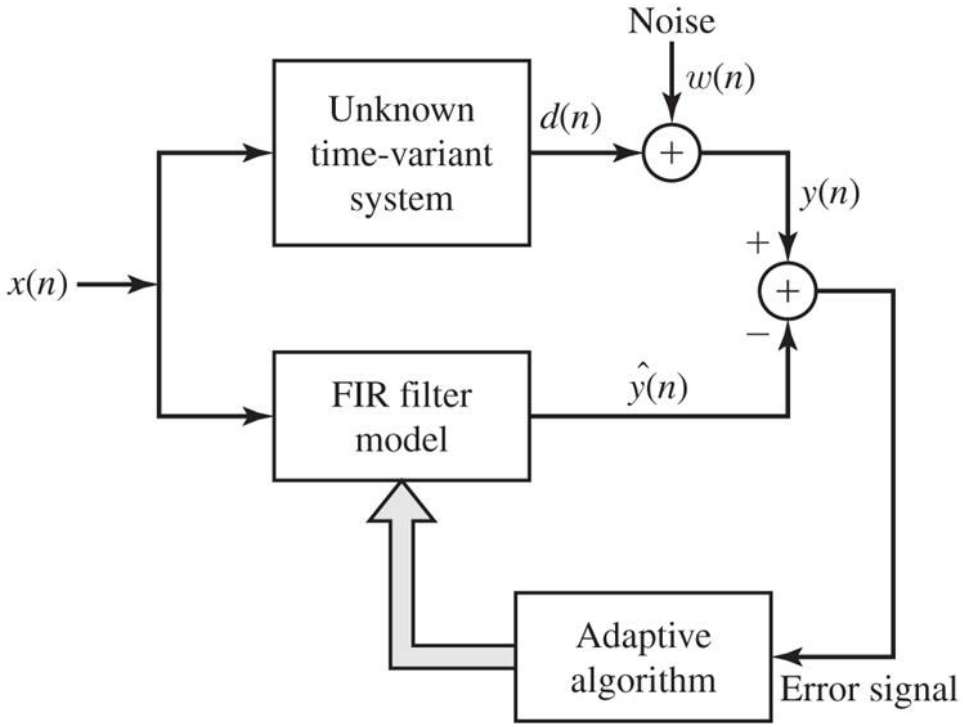
- least squares criterion
- mean square error criterion

Both criteria result in a quadratic performance index as a function of the filter coefficients and therefore has a single minimum.

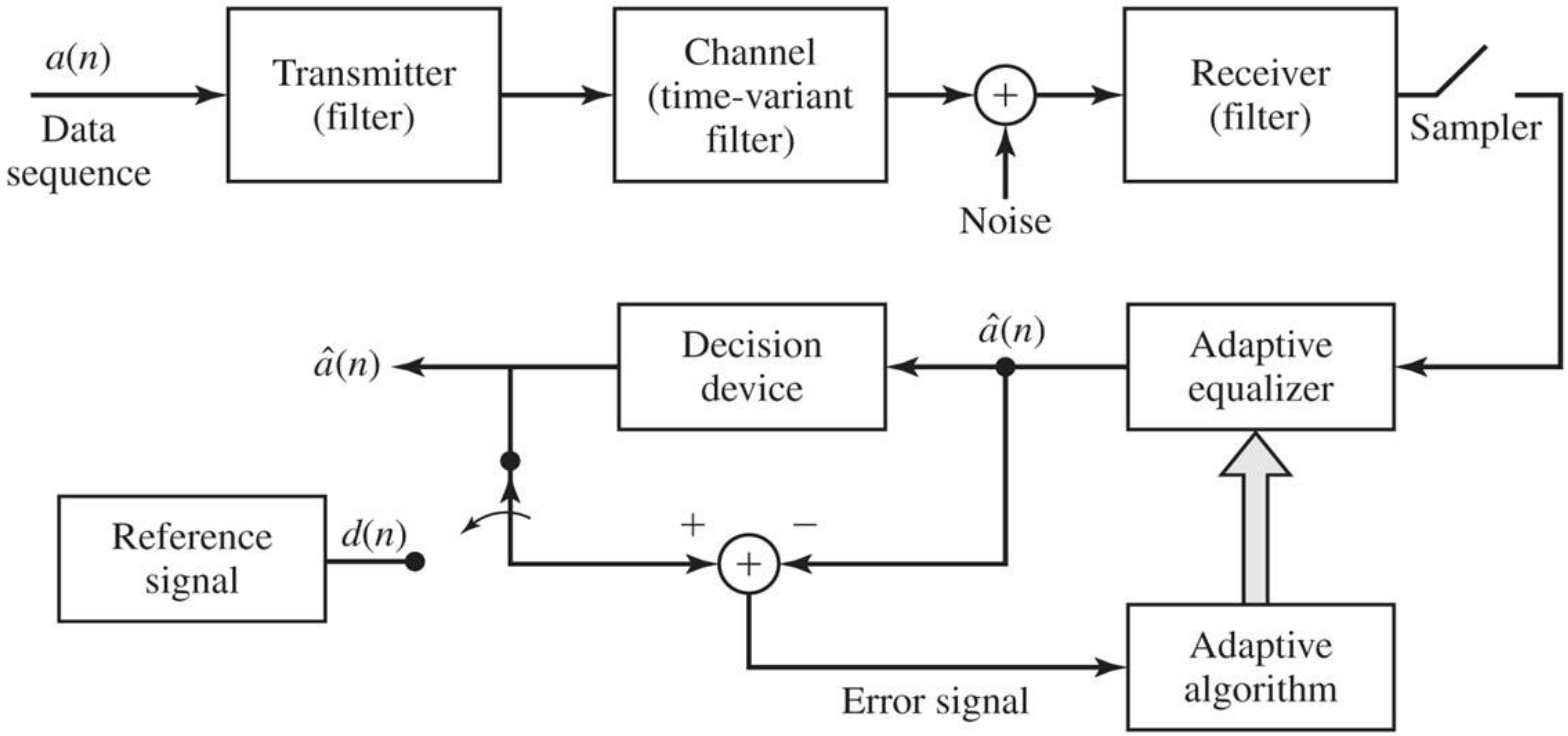
Applications of adaptive filters

- System identification or system modeling
- Adaptive channel equalization
- Echo cancellation in data transmission over telephone channels
- Suppression of narrowband interference in a wideband signal
- Adaptive line enhancer
- Adaptive noise cancelling
- Linear predicitive coding of speech signals

System identification or system modeling



Adaptive channel equalization



Suppression of narrowband interference in a wideband signal

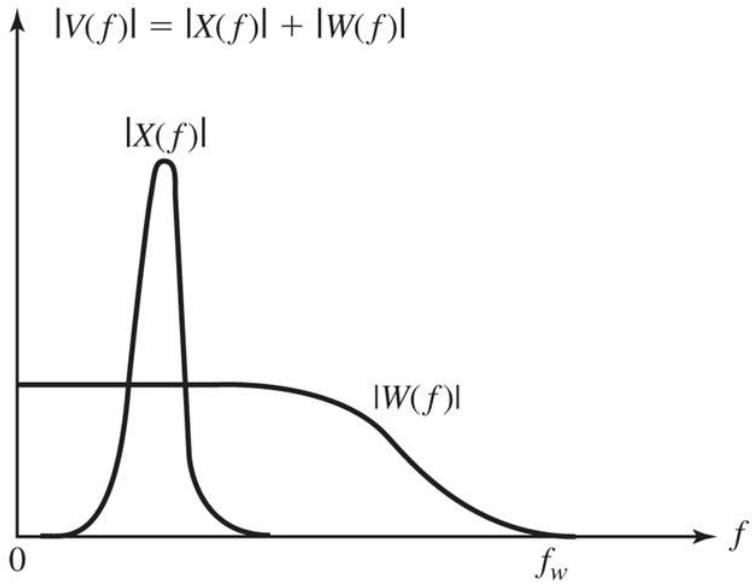
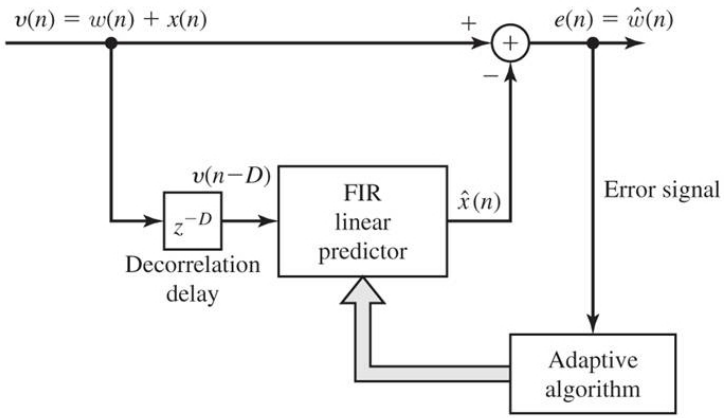
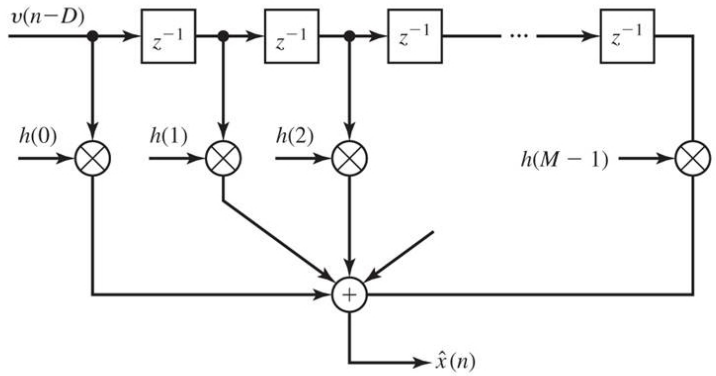


Figure 12: Strong narrowband interference $X(f)$ in a wideband signal $W(f)$

Adaptive filter for estimating and suppressing a narrowband interference in a wideband signal



(a)



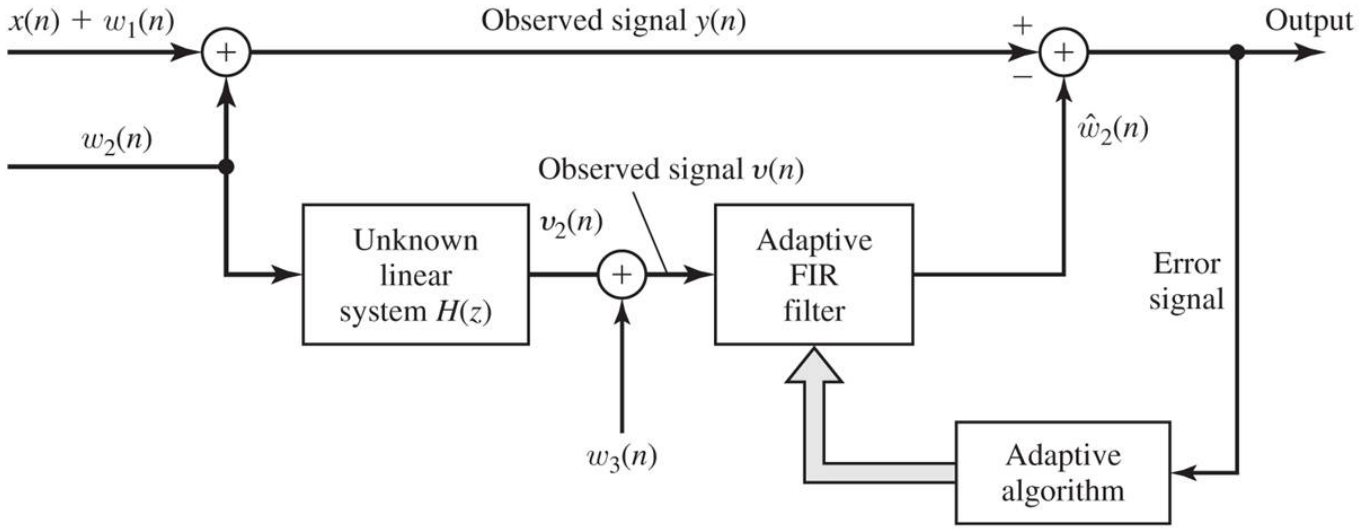
(b)

Adaptive line enhancer

Same configuration as the interference suppression filter but objective is different.

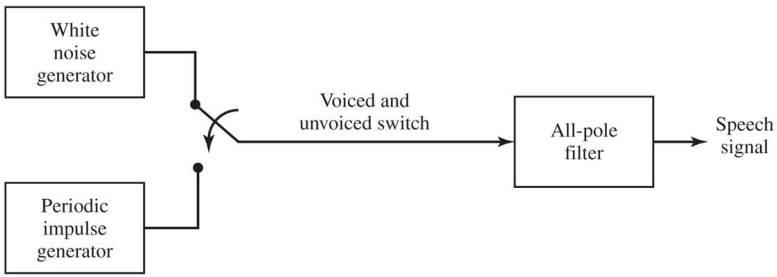
- $x[n]$ is the desired signal and is either a spectral line or a relatively narrowband signal
- $w[n]$ is a wideband noise component that masks $x[n]$.

Adaptive noise cancelling

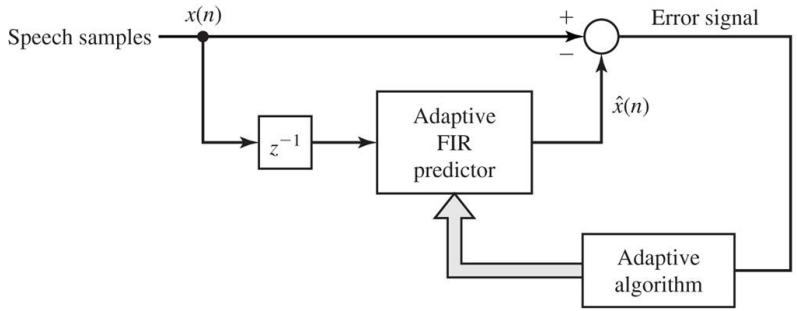


Linear predictive coding of speech signals

The primary objective in designing speech encoders[3] is to minimize the number of bits required to represent the speech signal, while maintaining speech intelligibility.



Generation of speech signal



Estimation of pole parameters in LPC

[3]Pulse Code Modulation, Differential PCM, Delta Modulation and Adaptive DPCM are examples of waveform coding.

Adaptive direct-form FIR filters–LMS algorithm

In all the adaptive filtering applications there is a least-squares criterion that is adopted and which leads to

$$\sum_{k=0}^{M-1} h[k] r_{xx}[l-k] = r_{dx}[l+D], \quad l = 0, 1, \ldots, M-1 \qquad (150)$$

where

- $r_{xx}[l]$ is the autocorrelation of the sequence $x[n]$
- $r_{dx}[l]$ is the cross-correlation of the sequences $d[n]$ and $x[n]$
- D is the delay parameter

Two problems in implementing an adaptive filter

1. Note that $r_{xx}[l]$ and $r_{dx}[l]$ are obtained from the data and, hence, represent estimates of the true (statistical) autocorrelation $\gamma_{xx}[l]$, cross-correlation $\gamma_{dx}[l]$ sequences, respectively.
   Therefore the coefficients $h[k]$ obtained from Eq. (150) are estimates of the true coefficients. The quality of the estimates depends on the length of the recorded data that is available for estimating $r_{xx}[l]$ and $r_{dx}[l]$.
2. The underlying random process $x[n]$ is usually non-stationary $\Rightarrow$ statistical autocorrelation and crosscorrelation sequences and hence their estimates vary with time.
   This implies that the the coefficients of the adaptive filter must change with time to incorporate the time-variant statistical characteristics of the signal into the filter.

Minimum Mean-Square-Error Criterion

The LMS algorithm is obtained by formulating the optimization of the FIR coefficients as an estimation problem based on the minimization of the mean square error. We substitute $r_{xx}[l]$ and $r_{dx}[l]$ by $\gamma_{xx}[l]$ and $\gamma_{dx}[l]$, respectively, in Eq. (150) and we obtain in matrix form

$$\mathbf{\Gamma}_M \mathbf{h}_M = \boldsymbol{\gamma}_d. \tag{151}$$

Recall that when $\mathbf{\Gamma}_M, \boldsymbol{\gamma}_d$ are known, the optimum filter coefficients are given by

$$\mathbf{h}_{opt} = \mathbf{\Gamma}_M^{-1} \boldsymbol{\gamma}_d \tag{152}$$

and the resulting MSE using the Wiener filter is

$$\varepsilon_M^h = \min_{\mathbf{h}_M} \mathcal{E}_M^h = \sigma_d^2 - \sum_{k=0}^{M-1} h_{opt}[k] \gamma_{dx}^*[k] \tag{153}$$

$$= \sigma_d^2 - \boldsymbol{\gamma}_d^{*t} \mathbf{\Gamma}_M^{-1} \boldsymbol{\gamma}_d \tag{154}$$

where $\sigma_d^2 = E\left[|d[n]|^2\right]$

LMS algorithm

The LMS algorithm uses an iterative gradient method for solving for $\mathbf{h}_{opt}$. In general, algorithms for recursively computing the filter coefficients and therefore searching for the minimum of $\mathcal{E}_M^h = E\left[|e[n]|^2\right]$ have the form

$$\mathbf{h}_M(n+1) = \mathbf{h}_M(n) + \frac{1}{2}\Delta(n)\mathbf{S}(n), \qquad n = 0,1,2,\ldots \tag{155}$$

where

- $\mathbf{h}_M(n)$ is the vector of filter coefficients at the n^{th} iteration
- $\Delta(n)$ is a step size at the n^{th} iteration
- $\mathbf{S}(n)$ is a direction vector for the n^{th} iteration

The initial vector $\mathbf{h}_M(0)$ is chosen arbitrarily.
The steepest descent search is a simple method for minimizing $\mathcal{E}_M^h$.

Steepest descent search

In the method of steepest descent, the direction vector $\mathbf{S}(n) = -\mathbf{g}(n)$ where the gradient vector at the n^{th} iteration is defined by

$$\mathbf{g}(n) = \frac{d\mathcal{E}_M^h(n)}{d\mathbf{h}_M(n)} \tag{156}$$

$$= 2[\Gamma_M \mathbf{h}_M(n) - \gamma(n)], \qquad n = 0, 1, 2, \ldots \tag{157}$$

I.e., we compute the gradient vector at each iteration and change the values of $\mathbf{h}_M[n]$ in a direction opposite the gradient.This leads to

$$\mathbf{h}_M(n+1) = \mathbf{h}_M(n) - \frac{1}{2}\Delta(n)\mathbf{g}(n), \qquad n = 0, 1, 2, \ldots \tag{158}$$

or equivalently,

$$\mathbf{h}_M(n+1) = [\mathbf{I} - \Delta(n)\mathbf{\Gamma}_M]\mathbf{h}_M(n) + \Delta(n)\gamma_d, \qquad n = 0, 1, 2, \ldots \tag{159}$$

If $\lim_{n\to\infty}\Delta(n)=0$ and $\sum_{n=0}^{\infty}|\Delta(n)|<\infty$ or $\lim_{n\to\infty}\mathbf{g}(n)=0$ then $\lim_{n\to\infty}\mathbf{h}_M(n)=\mathbf{h}_{opt}$.
Two other algorithms that provide faster convergence and which differ in their direction vectors are

- Conjugate Gradient Algorithm

$$\mathbf{S}(n)=\beta(n-1)\mathbf{S}(n-1)-\mathbf{g}(n) \tag{160}$$

- Fletcher-Powell Algorithm

$$\mathbf{S}(n)=-\mathbf{H}(n)\mathbf{g}(n) \tag{161}$$

with $\lim_{n\to\infty}\mathbf{H}(n)=\mathbf{\Gamma}_M$

In all three algorithms $\mathbf{\Gamma}_M$ and γ_d are assumed to be known which is not the case in adaptive filtering applications.

Unknown $\mathbf{\Gamma}_M$ and γ_d

When $\mathbf{\Gamma}_M$ and γ_d are unknown, we can substitute the actual direction vectors $\mathbf{S}(n)$ by some estimate $\hat{\mathbf{S}}(n)$.
Also note that the gradient vector $\mathbf{g}[n]$ may be exprssed in terms of the orthogonality conditions $E[e[n]x^*[n-l]] = 0, l = 0, 1, \ldots, M-1$ ie.,

$$E[e[n]\mathbf{X}_M^*(n)] = \gamma_d - \mathbf{\Gamma}_M \mathbf{h}_M(n) \tag{162}$$

where $\mathbf{X}_M^*(n) = [x[n], x[n-1], \cdots, x[n-(M-1)]^T$. The gradient vector is

$$\mathbf{g}(n) = -2E[e[n]\mathbf{X}_M^*(n)] \tag{163}$$

Ntoe that $\mathbf{g}(n) = 0$ when the error $e[n]$ is orthogonal to the data in the estimate $\hat{d}[n] = \sum_{k=0}^{M-1} h[k]x[n-k]$ of $d[n]$.

Stochastic-gradient-descent algorithm

The stochastic gradient descent algorithm is given by

$$\mathbf{h}_M(n+1) = \mathbf{h}_M(n) + \Delta(n)e(n)\mathbf{X}_M^*(n), \qquad n = 0, 1, 2, \ldots \tag{164}$$

It is obtained by substituting the gradient vector at the n^{th} iteration, $\mathbf{g}(n)$, with an unbiased estimate

$$\hat{\mathbf{g}}(n) = -2e[n]\mathbf{X}_M^*(n) \tag{165}$$

where $e[n] = d[n] - \hat{d}[n]$ and $\mathbf{X}_M(n)$ is the set of M signal samples in the filter at the n^{th} iteration.
Note that this algorithm has a variable step size $\Delta(n)$.

Stochastic Gradient-Descent Algorithm: Least-Mean-Squares (LMS) algorithm

The LMS algorithm is a stochastic gradient algorithm and is given by:

$$\mathbf{h}_M(n+1) = \mathbf{h}_M(n) + \mu e(n)\mathbf{X}_M^*(n), \qquad n = 0, 1, 2, \ldots \tag{166}$$

where μ is a fixed step size.
In practice, we use a fixed step size for two reasons:

- it can be easily implemented in hardware and software
- it is more appropriate for tracking time-variant signal statistics
  If $\Delta(n) \to 0$ as $n \to \infty$ algorithm cannot adapt to signal variations.

Normalized LMS (NLMS)

NLMS is frequently used in practice and is given by

$$\mathbf{h}_M(n+1) = \mathbf{h}_M(n) + \frac{\mu}{\|\mathbf{X}_M(n)\|^2} e(n)\mathbf{X}_M^*(n), \qquad n = 0, 1, 2, \ldots. \quad (167)$$

Here the step size is divided by the norm of $\mathbf{X}_M^*(n)$ which in turn is equivalent to using a variable step size

$$\Delta[n] = \frac{\mu}{\|\mathbf{X}_M(n)\|^2} \quad (168)$$

Note that the step size at each iteration is inversely proportional to the energy in the received data vector $\mathbf{X}_M(n)$,which is useful in adaptive filtering applications where the dynamic range of the input to the adaptive filter is large (eg. adaptive equalizers for slowly fading communication channels).

Convergence and Stability of LMS

Let $\bar{\mathbf{h}}_M(n) = E[\mathbf{h}_M(n)]$ be the mean value of $\mathbf{h}_M(n)$ then

$$\begin{aligned} E[\mathbf{h}_M(n+1)] &= E[\mathbf{h}_M(n) + \mu e(n)\mathbf{X}_M^*(n)] && (169)\\ \Leftrightarrow \bar{\mathbf{h}}_M(n+1) &= \bar{\mathbf{h}}_M(n) + \mu E[e(n)\mathbf{X}_M^*(n)] && (170)\\ &= \bar{\mathbf{h}}_M(n) + \mu[\boldsymbol{\gamma}_d - \boldsymbol{\Gamma}_M\bar{\mathbf{h}}_M(n)] && (171)\\ &= (\mathbf{I} - \mu\boldsymbol{\Gamma}_M)\bar{\mathbf{h}}_M(n) + \mu\boldsymbol{\gamma}_d && (172) \end{aligned}$$

where $\mathbf{I}$ is the identity matrix and the autocorrelation matrix $\boldsymbol{\Gamma}_M$ is Hermitian.

Convergence of LMS

Note that

$$\mathbf{\Gamma}_M = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{173}$$

where $\mathbf{U}$ is the normalized modal matrix of $\mathbf{\Gamma}_M$ and $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements $\lambda_k, 0 \leq k \leq M-1$ corresponding to the eigenvalues of $\mathbf{\Gamma}_M$. By substituting $\mathbf{\Gamma}_M = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, the convergence depends on the following

$$\bar{h}^0(k,n) = C(1-\mu\lambda_k)^n u[n], \qquad k = 0,1,2,\ldots,M-1 \tag{174}$$

where C is a constant and $u[n]$ is a unit step sequence. Note that

$$\bar{h}^0(k,n) \to 0 \iff |1-\mu\lambda_k| < 1 \tag{175}$$

$$\iff 0 < \mu < \frac{2}{\lambda_k} \tag{176}$$

$\forall k = 0,1,\ldots,M-1.$

The range of values of μ which ensures convergence of the mean of the coefficient vector in the LMS algorithm is

$$0 < \mu < \frac{2}{\lambda_{max}} \tag{177}$$

where λ_{max} is the largest eigenvalue of $\mathbf{\Gamma}_M$. Note that the eigenvalues of an autocorrelation matrix $\mathbf{\Gamma}_M$ are nonnegative, hence

$$\lambda_{max} < \sum_{k=0}^{M-1} \lambda_k = \text{trace}\mathbf{\Gamma}_M = M\gamma_{xx}[0] \tag{178}$$

where $\gamma_{xx}[0]$ is the input signal power, which can be estimated from the received signal.
Therefore an upper bound on the step size μ is given by

$$\mu < \frac{2}{M\gamma_{xx}[0]} \tag{179}$$

Speed of Convergence

The LMS algorithm will converge rapidly when $|1-\mu\lambda_k|$ is small, i.e. when the poles of the closed-loop system are far from the unit circle. But this conditions cannot be achieved simultaneously with the condition on the upper bound when there is a large difference between λ_{min} and λ_{max}. Take $\mu = \frac{1}{\lambda_{max}}$ and substitute in Eq. (174), we have

$$\bar{h}^0(k,n) = C\left(1-\frac{\lambda_{min}}{\lambda_{max}}\right)^n u[n], \qquad k = 0,1,2,\ldots,M-1 \qquad (180)$$

and consequently $\frac{\lambda_{min}}{\lambda_{max}}$ determines the convergence rate.
If $\frac{\lambda_{min}}{\lambda_{max}}$ is small ($<<1$) then convergence rate will be slow.
If $\frac{\lambda_{min}}{\lambda_{max}}$ is close to one, then convergence rate of the algorithm is fast.

Exercise 1

Consider the autocorrelation matrix of $x[n]$, $\mathbf{\Gamma}_M = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ and the cross-correlation vector between $x[n]$ and $d[n]$, $\boldsymbol{\gamma}_d = \begin{bmatrix} 0.25 \\ -0.25 \end{bmatrix}$.

1. Find $\mathbf{h}_{opt} = \begin{bmatrix} h_{opt}[0] \\ h_{opt}[1] \end{bmatrix}$ directly.
2. Let $\mu = 0.1$ and $\mathbf{h}_M(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ use the LMS algorithm to find $\mathbf{h}_{opt}$.
3. Let $\mu = 0.35$ and $\mathbf{h}_M(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ use the LMS algorithm to find $\mathbf{h}_{opt}$.

For each case plot the evolution graphs of the filter parameters , i.e., $h_M[0](n)$ versus n and $h_M[1](n)$ versus n.

Exercise 2

Consider the primary signal $d[n] = cos(2\pi\omega_0 n) + y_0[n]$ where the noise $y_0[n]$ is the output of the $F_0(z) = 0.5 - 0.5z^{-1}$ when excited with a white noise sequence $w[n]$.
Consider an input signal $x[n] = w[n]$.

1. Let $\mu = 0.00125$ and $\mathbf{h}_M(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ use the LMS algorithm to find $\mathbf{h}_{opt}$.
2. Plot the primary signal $d[n]$.
3. Plot $e[n] = d[n] - y[n]$ and the desired sine wave signal.
4. Plot the evolution of the two filter parameters.

Power Spectrum Estimation (6hrs)

- Introduction
- From finite duration observations of signals
- Nonparametric methods
- Parametric methods
- Filterbank methods
- Eigenanalysis methods