Random Variables and Stochastic Processes

Randomness

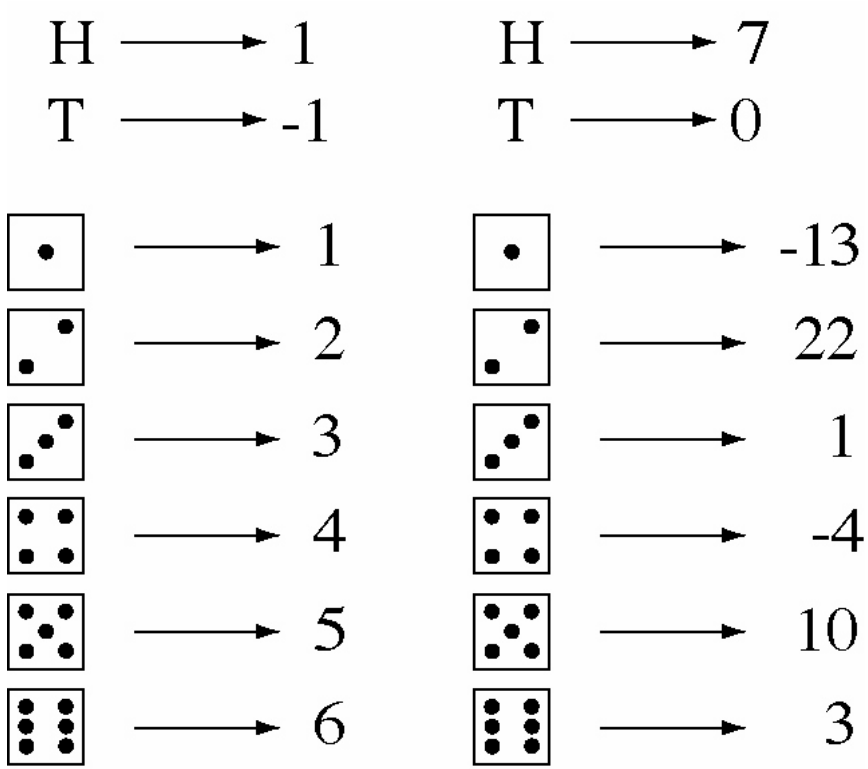
- The word *random* effectively means unpredictable
- In engineering practice we may treat some signals as random to simplify the analysis even though they may not actually be random

Random Variable Defined

A random variable $x(\zeta)$ is the assignment of numerical values to the outcomes ζ of experiments

Random Variables

Examples of assignments of numbers to the outcomes of experiments.



Discrete-Value vs Continuous-Value Random Variables

- A *discrete-value (DV)* random variable has a set of distinct values separated by values that cannot occur
- A random variable associated with the outcomes of coin flips, card draws, dice tosses, etc... would be DV random variable
- A *continuous-value (CV)* random variable may take on any value in a continuum of values which may be finite or infinite in size

Distribution Functions

The distribution function of a random variable X is the probability that it is less than or equal to some value, as a function of that value.

$$\mathrm{F}_X\left(x\right) = \mathrm{P}\left(X \le x\right)$$

Since the distribution function is a probability it must satisfy the requirements for a probability.

$$0 \le \mathrm{F}_X\left(x\right) \le 1 \;\; , \;\; -\infty < x < \infty$$

$$\mathrm{F}_X\left(-\infty\right) = 0 \;\; \text{and} \;\; \mathrm{F}_X\left(+\infty\right) = 1$$

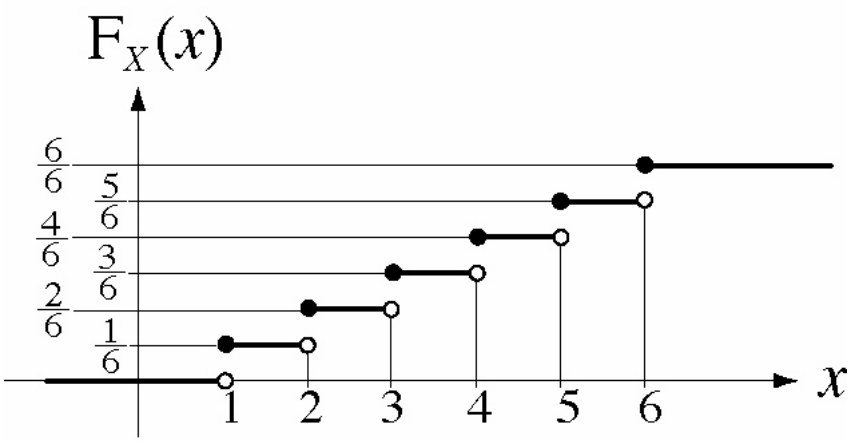
$$\mathrm{P}\left(x_1 < X \le x_2\right) = \mathrm{F}_X\left(x_2\right) - \mathrm{F}_X\left(x_1\right)$$

$\mathrm{F}_X\left(x\right)$ is a monotonic function and its derivative is never negative.

Distribution Functions

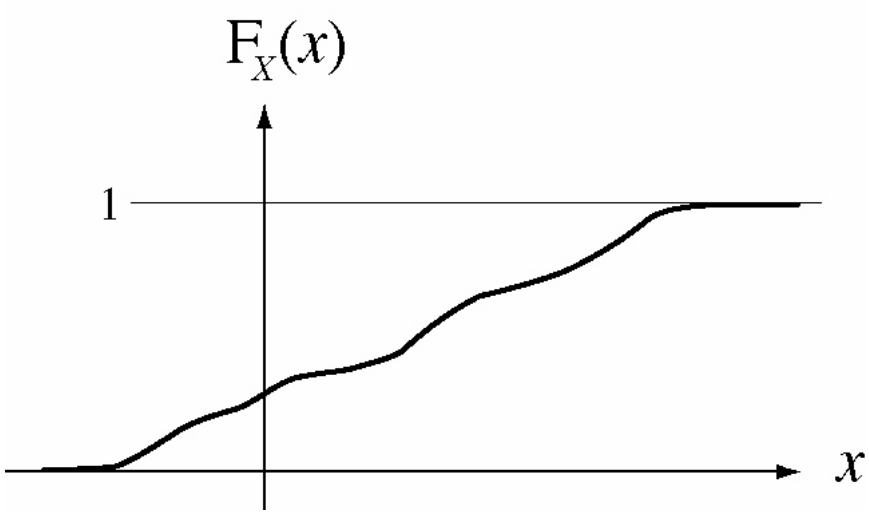
The distribution function for tossing a single die

$$F_X(x) = (1/6)\left[\begin{array}{l} u(x-1)+u(x-2)+u(x-3) \\ +u(x-4)+u(x-5)+u(x-6) \end{array}\right]$$



Distribution Functions

A possible distribution function for a continuous random variable



Probability Density

The derivative of the distribution function is the *probability density function (pdf)*.

$$f_X(x) \equiv \frac{d}{dx}\left(F_X(x)\right)$$

Probability density can also be defined by

$$f_X(x)\,dx = P(x < X \le x + dx)$$

Properties

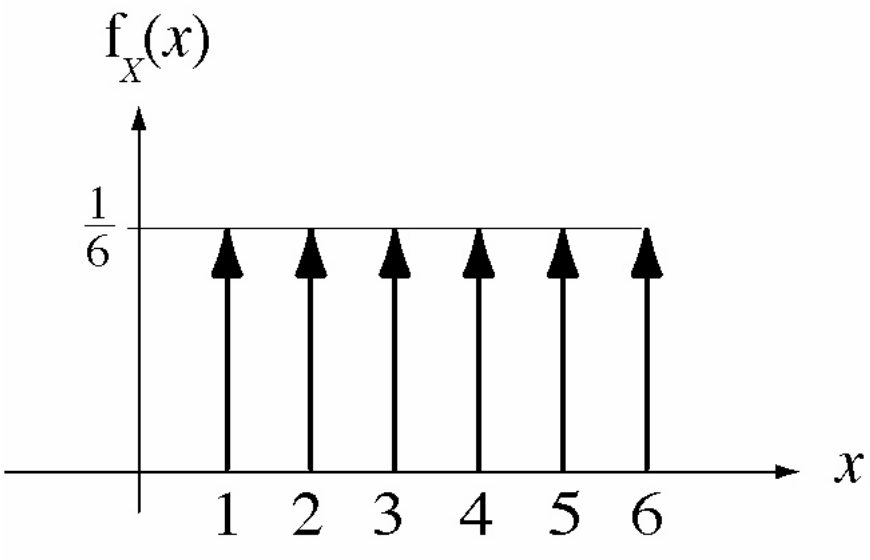
$$f_X(x) \ge 0 \quad , \quad -\infty < x < +\infty \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

$$F_X(x) = \int_{-\infty}^{x} f_X(\lambda)\,d\lambda \qquad P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)\,dx$$

(Proakis uses the notation $p(x)$ instead of $f_X(x)$ for probability density.)

Probability Mass and Density

The pdf for tossing a die



Expectation and Moments

Imagine an experiment with M possible distinct outcomes performed N times. The average of those N outcomes is

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{M} n_i x_i$$

where x_i is the ith distinct value of X and n_i is the number of times that value occurred. Then

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{M} n_i x_i = \sum_{i=1}^{M} \frac{n_i}{N} x_i = \sum_{i=1}^{M} r_i x_i$$

The *expected value* of X is

$$\mathrm{E}\left(X\right) = \lim_{N\to\infty} \sum_{i=1}^{M} \frac{n_i}{N} x_i = \lim_{N\to\infty} \sum_{i=1}^{M} r_i x_i = \sum_{i=1}^{M} \mathrm{P}\left(X = x_i\right) x_i$$

Expectation and Moments

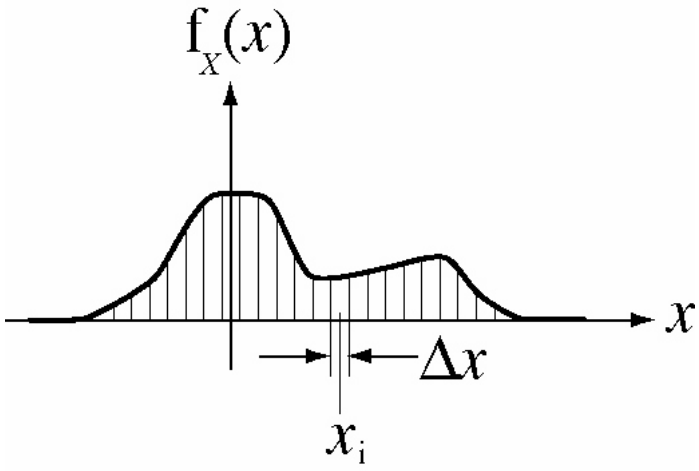
The probability that X lies within some small range can be approximated by

$$\mathrm{P}\left(x_i - \frac{\Delta x}{2} < X \le x_i + \frac{\Delta x}{2} \right) \cong \mathrm{f}_X\left(x_i\right)\Delta x$$

and the expected value is then approximated by

$$\mathrm{E}\left(X\right) = \sum_{i=1}^{M} \mathrm{P}\left(x_i - \frac{\Delta x}{2} < X \le x_i + \frac{\Delta x}{2} \right) x_i \cong \sum_{i=1}^{M} x_i \, \mathrm{f}_X\left(x_i\right)\Delta x$$

where M is now the number of subdivisions of width Δx of the range of the random variable.



Expectation and Moments

In the limit as Δx approaches zero,

$$\mathrm{E}\left(X\right) = \int_{-\infty}^{\infty} x\, \mathrm{f}_X\left(x\right) dx$$

Similarly

$$\mathrm{E}\left(\mathrm{g}\left(X\right)\right) = \int_{-\infty}^{\infty} \mathrm{g}\left(x\right) \mathrm{f}_X\left(x\right) dx$$

The *n*th *moment* of a random variable is

$$\mathrm{E}\left(X^n\right) = \int_{-\infty}^{\infty} x^n\, \mathrm{f}_X\left(x\right) dx$$

Expectation and Moments

The first moment of a random variable is its expected value

$$\mathrm{E}\left(X\right) = \int_{-\infty}^{\infty} x\, \mathrm{f}_X\left(x\right) dx$$

The second moment of a random variable is its *mean-squared value* (which is the mean of its square, not the square of its mean).

$$\mathrm{E}\left(X^2\right) = \int_{-\infty}^{\infty} x^2\, \mathrm{f}_X\left(x\right) dx$$

Expectation and Moments

A *central moment* of a random variable is the moment of that random variable after its expected value is subtracted.

$$\mathrm{E}\left(\left[X-\mathrm{E}\left(X\right)\right]^{n}\right)=\int_{-\infty}^{\infty}\left[x-\mathrm{E}\left(X\right)\right]^{n}\mathrm{f}_{X}\left(x\right)dx$$

The first central moment is always zero. The second central moment (for real-valued random variables) is the *variance*,

$$\sigma_{X}^{2}=\mathrm{E}\left(\left[X-\mathrm{E}\left(X\right)\right]^{2}\right)=\int_{-\infty}^{\infty}\left[x-\mathrm{E}\left(X\right)\right]^{2}\mathrm{f}_{X}\left(x\right)dx$$

The positive square root of the variance is the *standard deviation*.

Expectation and Moments

Properties of expectation

$$\mathrm{E}\left(a\right)=a \;\;,\;\; \mathrm{E}\left(aX\right)=a\,\mathrm{E}\left(X\right) \;\;,\;\; \mathrm{E}\left(\sum_n X_n\right)=\sum_n \mathrm{E}\left(X_n\right)$$

where *a* is a constant. These properties can be use to prove the handy relationship,

$$\sigma_X^2 = \mathrm{E}\left(X^2\right)-\mathrm{E}^2\left(X\right)$$

The variance of a random variable is the mean of its square minus the square of its mean.

Expectation and Moments

For complex-valued random variables *absolute moments* are useful. The *n*th absolute moment of a random variable is defined by

$$\mathrm{E}\left(\left|X\right|^n\right) = \int_{-\infty}^{\infty} \left|x\right|^n \mathrm{f}_X\left(x\right) dx$$

and the *n*th absolute central moment is defined by

$$\mathrm{E}\left(\left|X - \mathrm{E}\left(X\right)\right|^n\right) = \int_{-\infty}^{\infty} \left|x - \mathrm{E}\left(X\right)\right|^n \mathrm{f}_X\left(x\right) dx$$

Joint Probability Density

Let X and Y be two random variables. Their *joint distribution function* is

$$\mathrm{F}_{XY}\left(x,y\right)\equiv \mathrm{P}\left(X\le x\cap Y\le y\right)$$

$$0\le \mathrm{F}_{XY}\left(x,y\right)\le 1\ ,\ -\infty < x < \infty\ ,\ -\infty < y < \infty$$

$$\mathrm{F}_{XY}\left(-\infty,-\infty\right)=\mathrm{F}_{XY}\left(x,-\infty\right)=\mathrm{F}_{XY}\left(-\infty,y\right)=0$$

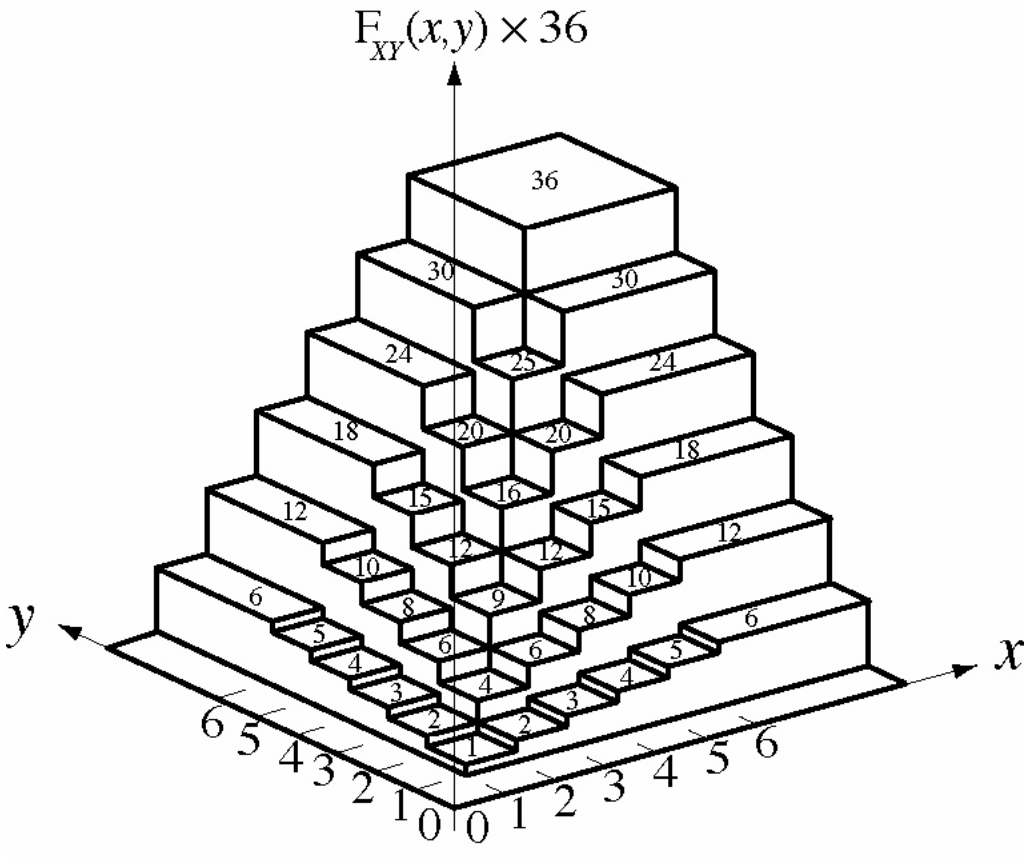
$$\mathrm{F}_{XY}\left(\infty,\infty\right)=1$$

$\mathrm{F}_{XY}\left(x,y\right)$ does not decrease if either x or y increases or both increase

$$\mathrm{F}_{XY}\left(\infty,y\right)=\mathrm{F}_{Y}\left(y\right)\ \text{and}\ \mathrm{F}_{XY}\left(x,\infty\right)=\mathrm{F}_{X}\left(x\right)$$

Joint Probability Density

Joint distribution function for tossing two dice



Joint Probability Density

$$\mathrm{f}_{XY}\left(x,y\right)=\frac{\partial^2}{\partial x\partial y}\left(\mathrm{F}_{XY}\left(x,y\right)\right)$$

$$\mathrm{f}_{XY}\left(x,y\right)\geq 0\ \ ,\ -\infty<x<\infty\ \ ,\ -\infty<y<\infty$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{f}_{XY}\left(x,y\right)dxdy=1 \qquad \mathrm{F}_{XY}\left(x,y\right)=\int_{-\infty}^{y}\int_{-\infty}^{x}\mathrm{f}_{XY}\left(\alpha,\beta\right)d\alpha d\beta$$

$$\mathrm{f}_{X}\left(x\right)=\int_{-\infty}^{\infty}\mathrm{f}_{XY}\left(x,y\right)dy \ \text{ and } \ \mathrm{f}_{Y}\left(y\right)=\int_{-\infty}^{\infty}\mathrm{f}_{XY}\left(x,y\right)dx$$

$$\mathrm{P}\left(\left(X,Y\right)\in R\right)=\iint_{R}\mathrm{f}_{XY}\left(x,y\right)dxdy$$

$$\mathrm{P}\left(x_1<X\leq x_2\ ,\ y_1<Y\leq y_2\right)=\int_{y_1}^{y_2}\int_{x_1}^{x_2}\mathrm{f}_{XY}\left(x,y\right)dxdy$$

$$\mathrm{E}\left(\mathrm{g}\left(X,Y\right)\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{g}\left(x,y\right)\mathrm{f}_{XY}\left(x,y\right)dxdy$$

Independent Random Variables

If two random variables X and Y are independent then

$$\mathrm{f}_{XY}\left(x,y\right)=\mathrm{f}_{X}\left(x\right)\mathrm{f}_{Y}\left(y\right)$$

and their correlation is the product of their expected values

$$\mathrm{E}\left(XY\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xy\,\mathrm{f}_{XY}\left(x,y\right)dxdy=\int_{-\infty}^{\infty}y\,\mathrm{f}_{Y}\left(y\right)dy\int_{-\infty}^{\infty}x\,\mathrm{f}_{X}\left(x\right)dx=\mathrm{E}\left(X\right)\mathrm{E}\left(Y\right)$$

Independent Random Variables

Covariance

$$\sigma_{XY} \equiv \mathrm{E}\left(\left[X - \mathrm{E}\left(X \right) \right] \left[Y - \mathrm{E}\left(Y \right) \right]^{*} \right)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x - \mathrm{E}\left(X \right) \right) \left(y^{*} - \mathrm{E}\left(Y^{*} \right) \right) \mathrm{f}_{XY}\left(x, y \right) dx dy$$

$$\sigma_{XY} = \mathrm{E}\left(XY^{*} \right) - \mathrm{E}\left(X \right) \mathrm{E}\left(Y^{*} \right)$$

If X and Y are independent,

$$\sigma_{XY} = \mathrm{E}\left(X \right) \mathrm{E}\left(Y^{*} \right) - \mathrm{E}\left(X \right) \mathrm{E}\left(Y^{*} \right) = 0$$

Independent Random Variables

If two random variables are independent, their covariance is zero.

However, if two random variables have a zero covariance that does not mean they are necessarily independent.

$$\text{Independence} \Rightarrow \text{Zero Covariance}$$

$$\cancel{\text{Zero Covariance} \Rightarrow \text{Independence}}$$

Independent Random Variables

In the traditional jargon of random variable analysis, two "uncorrelated" random variables have a covariance of zero.

Unfortunately, this does not also imply that their correlation is zero. If their correlation is zero they are said to be *orthogonal*.

$$X \text{ and } Y \text{ are "Uncorrelated"} \Rightarrow \sigma_{XY} = 0$$

~~X and Y are "Uncorrelated" $\Rightarrow \mathrm{E}(XY) = 0$~~

Independent Random Variables

The variance of a sum of random variables X and Y is

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y$$

If Z is a linear combination of random variables X_i

$$Z = a_0 + \sum_{i=1}^{N} a_i X_i$$

then

$$\mathrm{E}\left(Z\right) = a_0 + \sum_{i=1}^{N} a_i \,\mathrm{E}\left(X_i\right)$$

$$\sigma_Z^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j \sigma_{X_i X_j} = \sum_{i=1}^{N} a_i^2 \sigma_{X_i}^2 + \sum_{\substack{i=1 \\ i\neq j}}^{N}\sum_{j=1}^{N} a_i a_j \sigma_{X_i X_j}$$

Independent Random Variables

If the *X*'s are all independent of each other, the variance of the linear combination is a linear combination of the variances.

$$\sigma_Z^2 = \sum_{i=1}^{N} a_i^2 \sigma_{X_i}^2$$

If *Z* is simply the sum of the *X*'s, and the *X*'s are all independent of each other, then the variance of the sum is the sum of the variances.

$$\sigma_Z^2 = \sum_{i=1}^{N} \sigma_{X_i}^2$$

The Central Limit Theorem

If N independent random variables are added to form a resultant random variable Z

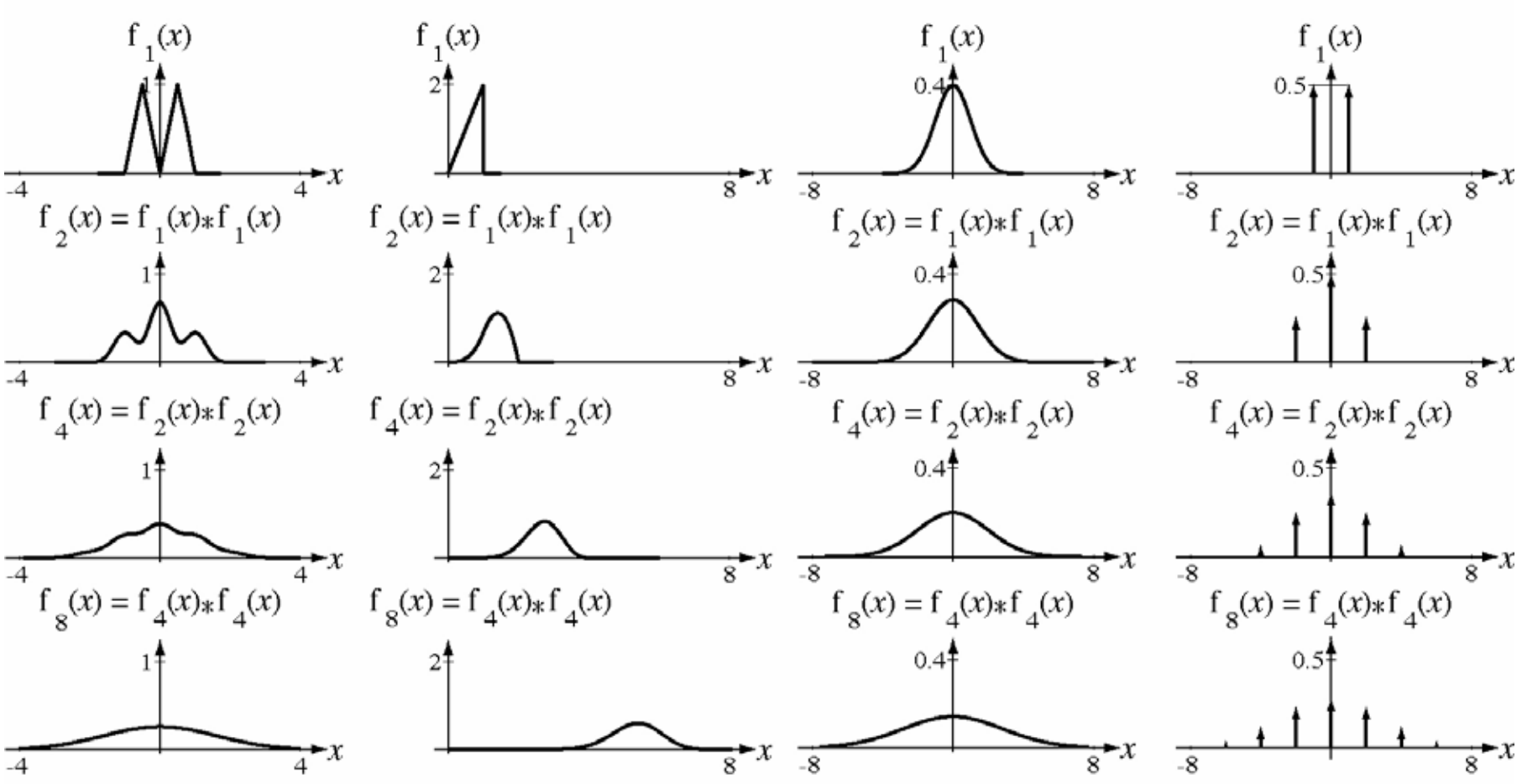
$$Z = \sum_{n=1}^{N} X_n$$

then

$$f_Z(z) = f_{X_1}(z) * f_{X_2}(z) * f_{X_2}(z) * \cdots * f_{X_N}(z)$$

and it can be shown that, under very general conditions, the pdf of a sum of a large number of independent random variables with continuous pdf's approaches a limiting shape called the "Gaussian" pdf regardless of the shapes of the individual pdf's.

The Central Limit Theorem

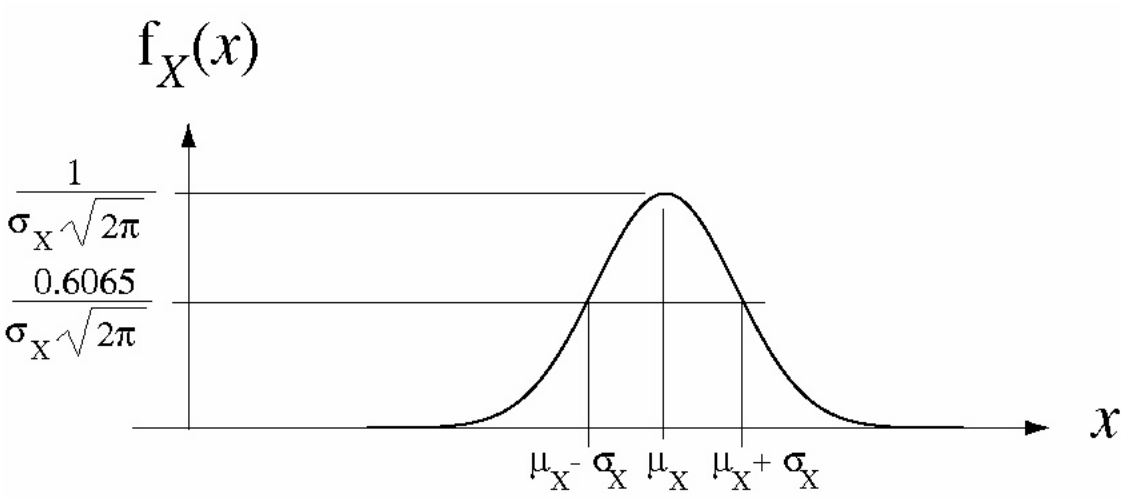


The Central Limit Theorem

The Gaussian pdf

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-(x-\mu_X)^2/2\sigma_X^2}$$

$$\mu_X = E(X) \text{ and } \sigma_X = \sqrt{E\left(\left[X - E(X)\right]^2\right)}$$



The Central Limit Theorem

The Gaussian pdf

Its maximum value occurs at the mean value of its argument

It is symmetrical about the mean value

The points of maximum absolute slope occur at one standard deviation above and below the mean

Its maximum value is inversely proportional to its standard deviation

The limit as the standard deviation approaches zero is a unit impulse

$$\delta\left(x-\mu_x\right)=\lim_{\sigma_X\to 0}\frac{1}{\sigma_X\sqrt{2\pi}}e^{-\left(x-\mu_X\right)^2/2\sigma_X^2}$$

The Central Limit Theorem

The *normal* pdf is a Gaussian pdf with a mean of zero and a variance of one.

$$\mathrm{f}_X\left(x\right) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

The central moments of the Gaussian pdf are

$$\mathrm{E}\left(\left[X - \mathrm{E}\left(X\right)\right]^n\right) = \begin{cases} 0 & , \ n \ \text{odd} \\ 1 \cdot 3 \cdot 5 \ldots \left(n-1\right)\sigma_X^n & , \ n \ \text{even} \end{cases}$$

Stochastic Processes

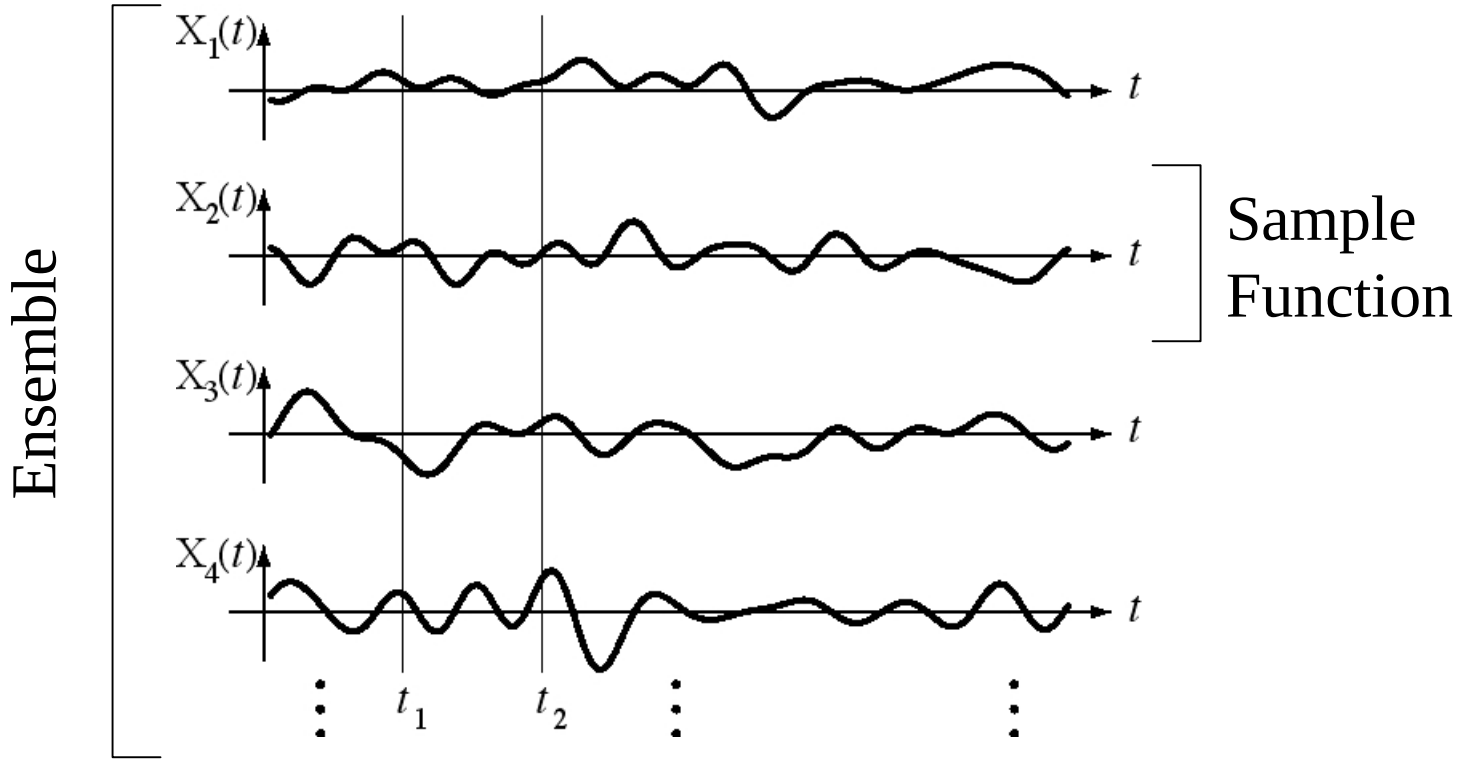
A random variable is a number $X(\zeta)$ assigned to every outcome of an experiment.

A *stochastic process* is the assignment of a function of t $X(t,\zeta)$ to each outcome of an experiment.

The set of functions $\{X(t,\zeta_1), X(t,\zeta_2), \cdots, X(t,\zeta_N)\}$ corresponding to the N outcomes of an experiment is called an *ensemble* and each member $X(t,\zeta_i)$ is called a *sample function* of the stochastic process.

A common convention in the notation describing stochastic processes is to write the sample functions as functions of t only and to indicate the stochastic process by $X(t)$ instead of $X(t,\zeta)$ and any particular sample function by $X_i(t)$ instead of $X(t,\zeta_i)$.

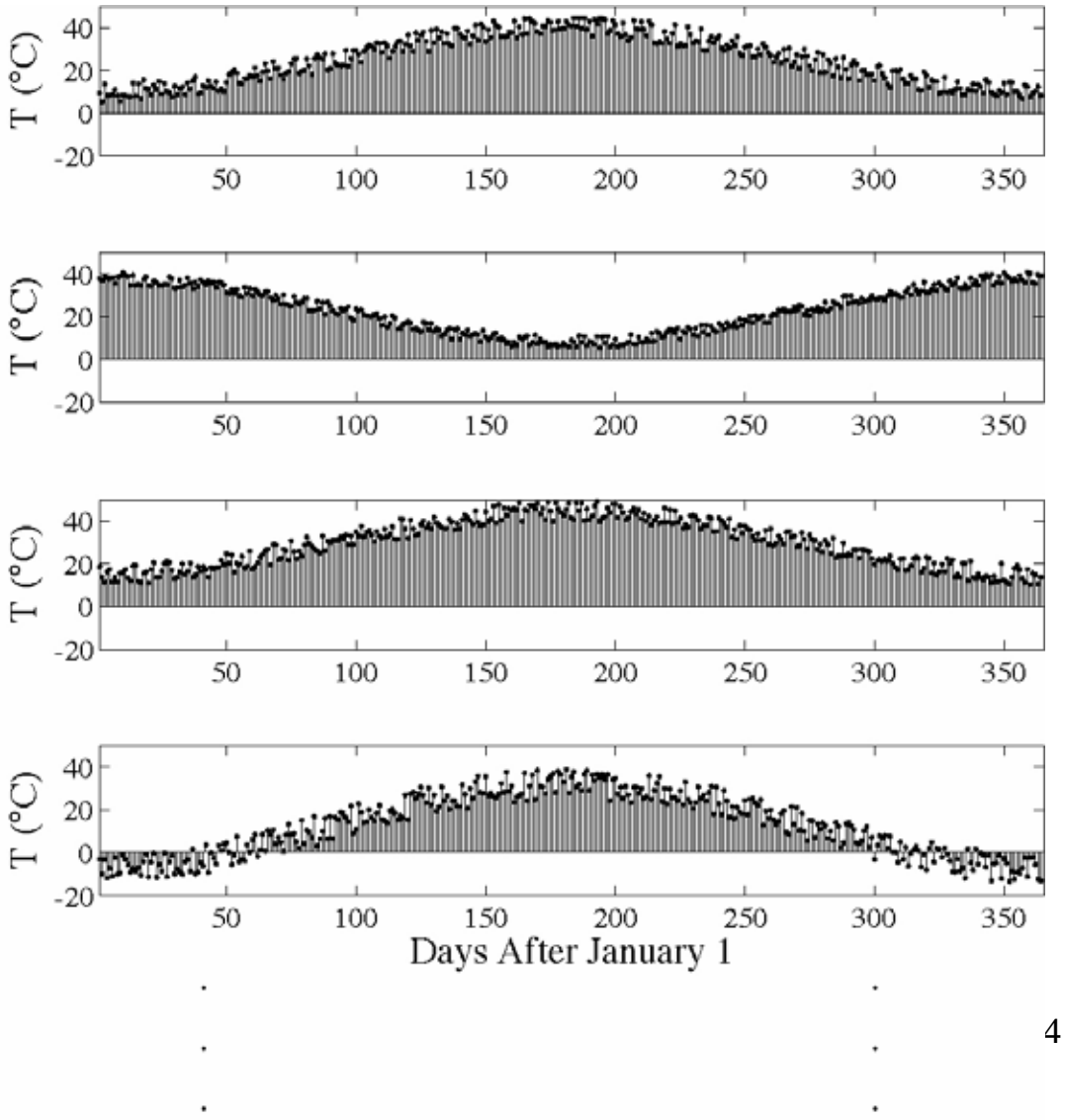
Stochastic Processes



The values of $X(t)$ at a particular time t_1 define a random variable $X(t_1)$ or just X_1.

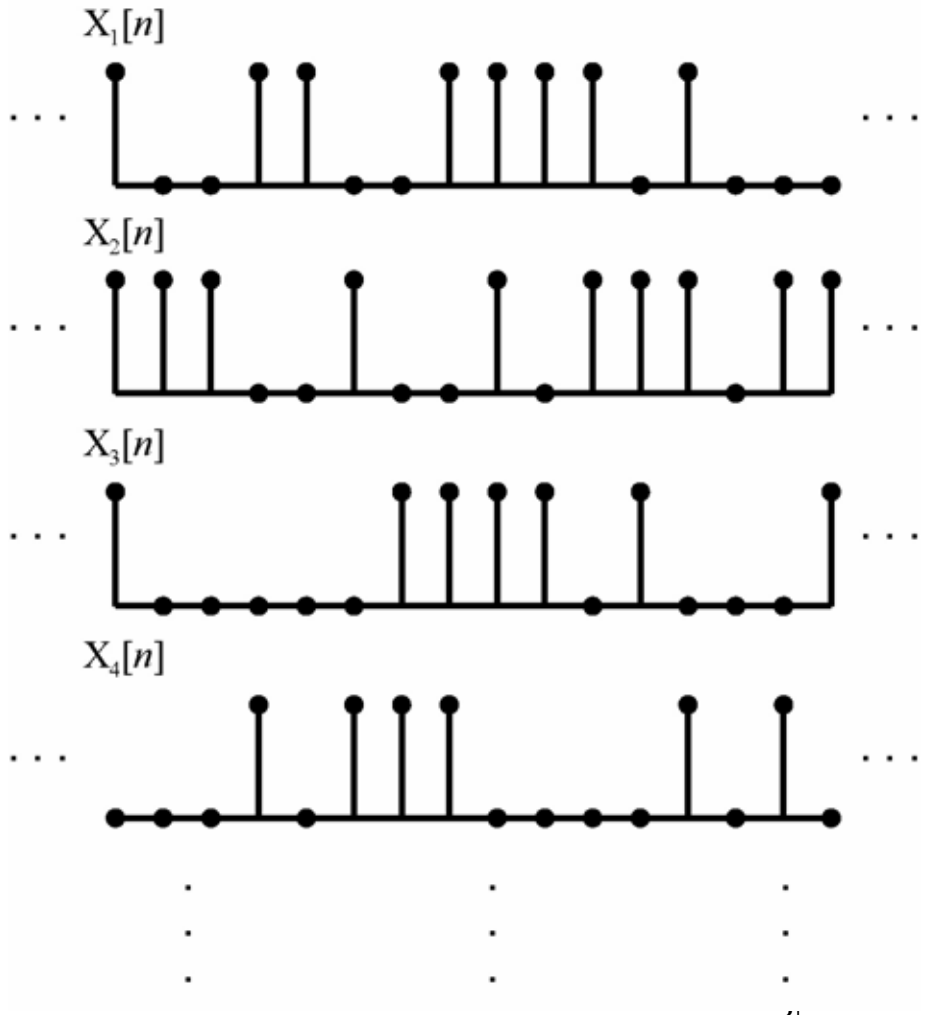
Example of a Stochastic Process

Suppose we place a temperature sensor at every airport control tower in the world and record the temperature at noon every day for a year. Then we have a discrete-time, continuous-value (DTCV) stochastic process.



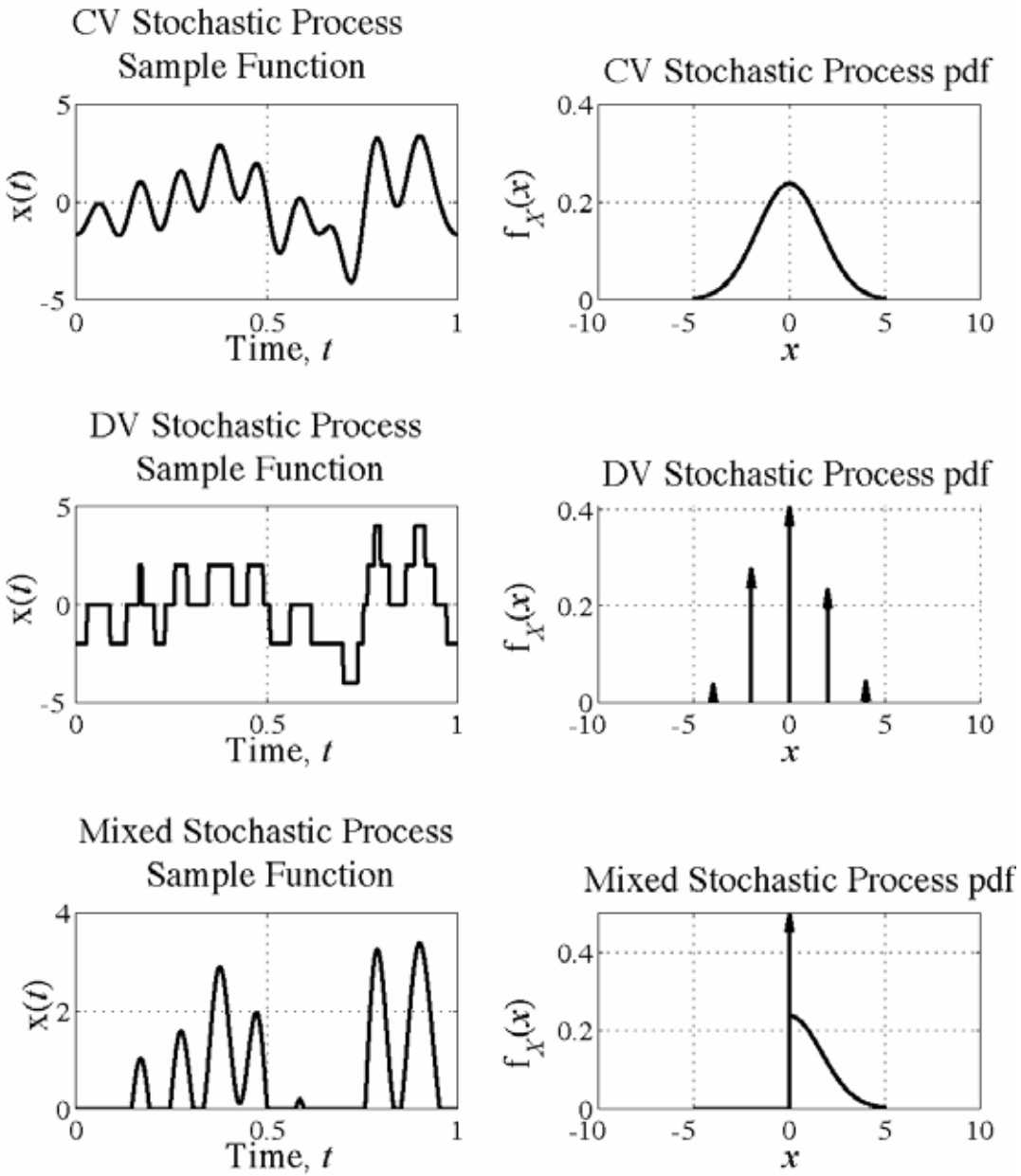
Example of a Stochastic Process

Suppose there is a large number of people, each flipping a fair coin every minute. If we assign the value 1 to a head and the value 0 to a tail we have a discrete-time, discrete-value (DTDV) stochastic process



Continuous-Value vs. Discrete-Value

A continuous-value (CV) random process has a pdf with no impulses. A discrete-value (DV) random process has a pdf consisting only of impulses. A mixed random process has a pdf with impulses, but not just impulses.



Deterministic vs. Non-Deterministic

A random process is deterministic if a sample function can be described by a mathematical function such that its future values can be computed. The randomness is in the ensemble, not in the time functions. For example, let the sample functions be of the form,

$$X(t) = A\cos(2\pi f_0 t + \theta)$$

and let the parameter θ be random over the ensemble but constant for any particular sample function.

All other random processes are non-deterministic.

Stationarity

If all the mltivariate statistical descriptors of a random process are not functions of time, the random process is said to be *strict-sense stationary (SSS)*.

A random process is *wide-sense stationary (WSS)* if

$E\left(X\left(t_1\right)\right)$ is independent of the choice of t_1

and

$E\left(X\left(t_1\right)X\left(t_2\right)\right)$ depends only on the difference between t_1 and t_2

Ergodicity

If all of the sample functions of a random process have the same statistical properties the random process is said to be *ergodic*. The most important consequence of ergodicity is that ensemble moments can be replaced by time moments.

$$\mathrm{E}\left(X^n\right) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} X^n\left(t\right) dt$$

Every ergodic random process is also stationary.

The Correlation Function

If X(t) is a sample function of one stochastic CT process and Y(t) is a sample function from another stochastic CT process and

$$X_1 = \mathrm{X}\left(t_1\right) \text{ and } Y_2 = \mathrm{Y}\left(t_2\right)$$

then

$$\mathrm{R}_{XY}\left(t_1,t_2\right) = \mathrm{E}\left(X_1 Y_2^*\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X_1 Y_2^* \, \mathrm{f}_{XY}\left(x_1, y_2; t_1, t_2\right) dx_1 dy_2$$

is the correlation function relating X and Y. For stationary stochastic CT processes this can be simplified to

$$\mathrm{R}_{XY}\left(\tau\right) = \mathrm{E}\left(\mathrm{X}\left(t\right)\mathrm{Y}^*\left(t+\tau\right)\right)$$

If the stochastic process is also ergodic then the *time* correlation function is

$$\mathrm{R}_{XY}\left(\tau\right) = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\mathrm{X}\left(t\right)\mathrm{Y}^*\left(t+\tau\right)dt = \left\langle \mathrm{X}\left(t\right)\mathrm{Y}^*\left(t+\tau\right)\right\rangle = \mathrm{R}_{XY}\left(\tau\right)$$

Autocorrelation

If *X* and *Y* represent the same stochastic CT process then the correlation function becomes the special case called *autocorrelation*.

$$\mathrm{R}_X(\tau) = \mathrm{E}\left[\mathrm{X}(t)\mathrm{X}^*(t+\tau)\right]$$

For an ergodic stochastic process,

$$\mathrm{R}_X(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\mathrm{X}(t)\mathrm{X}^*(t+\tau)\,dt = \left\langle \mathrm{X}(t)\mathrm{X}^*(t+\tau)\right\rangle = \mathrm{R}_X(\tau)$$

(Proakis uses the notation $\gamma_{xy}(\tau)$ for correlation.)

Autocorrelation

$$R_X(t,t) = \underbrace{E\left(X^2(t)\right)}_{\text{Mean-squared value of X}}$$

For WSS stochastic CT processes

$$R_X(0) = \underbrace{E\left(X^2(t)\right)}_{\text{Mean-squared value of X}} \quad \text{and} \quad R_X(0) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X^2(t)\,dt = \underbrace{\left\langle X^2(t)\right\rangle}_{\text{Average Signal Power of X}}$$

The Correlation Function

If $\mathrm{X}\left[n\right]$ is a sample function of one stochastic DT process and $\mathrm{Y}\left[n\right]$ is a sample function from another stochastic DT process and

$$X_1 = \mathrm{X}\left[n_1\right] \text{ and } Y_2 = \mathrm{Y}\left[n_2\right]$$

then

$$\mathrm{R}_{XY}\left[n_1, n_2\right] = \mathrm{E}\left(X_1 Y_2^*\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X_1 Y_2^* \, \mathrm{f}_{XY}\left(x_1, y_2; n_1, n_2\right) dx_1 dy_2$$

is the correlation function relating X and Y. For stationary stochastic DT processes this can be simplified to

$$\mathrm{R}_{XY}\left[m\right] = \mathrm{E}\left(\mathrm{X}\left[n\right]\mathrm{Y}^*\left[n+m\right]\right)$$

If the stochastic DT process is also ergodic then the *time* correlation function is

$$\mathrm{R}_{XY}\left[m\right] = \lim_{N\to\infty}\frac{1}{2N}\sum_{n=-N}^{N-1}\mathrm{X}\left[n\right]\mathrm{Y}^*\left[n+m\right] = \left\langle \mathrm{X}\left[n\right]\mathrm{Y}^*\left[n+m\right]\right\rangle = \mathrm{R}_{XY}\left[m\right]$$

Autocorrelation

If X and Y represent the same stochastic DT process then the correlation function becomes the special case called *autocorrelation*.

$$\mathrm{R}_X\left[m\right] = \mathrm{E}\left(\mathrm{X}\left[n\right]\mathrm{X}^*\left[n+m\right]\right)$$

For an ergodic stochastic DT process,

$$\mathrm{R}_X\left[m\right] = \lim_{N\to\infty}\frac{1}{2N}\sum_{n=-N}^{N-1}\mathrm{X}\left[n\right]\mathrm{X}^*\left[n+m\right] = \left\langle \mathrm{X}\left[n\right]\mathrm{X}^*\left[n+m\right]\right\rangle = \mathrm{R}_X\left[m\right]$$

Autocorrelation

$$R_X[n,n] = \underbrace{E\left(X^2[n]\right)}$$

Mean-squared value of X

For WSS stochastic DT processes

$$R_X[0] = \underbrace{E\left(X^2[n]\right)} \quad \text{and} \quad R_X[0] = \lim_{N\to\infty} \frac{1}{2N} \sum_{n=-N}^{N-1} X^2[n] = \underbrace{\left\langle X^2[n] \right\rangle}$$

Mean-squared value of X

Average Signal Power of X

Properties of Autocorrelation

Autocorrelation is an even function

$$\mathrm{R}_X(\tau) = \mathrm{R}_X(-\tau) \text{ or } \mathrm{R}_X[m] = \mathrm{R}_X[-m]$$

The magnitude of the autocorrelation value is never greater than at zero delay.

$$\left|\mathrm{R}_X(\tau)\right| \leq \mathrm{R}_X(0) \text{ or } \left|\mathrm{R}_X[m]\right| \leq \mathrm{R}_X[0]$$

If X has a non-zero expected value then $\mathrm{R}_X(\tau)$ or $\mathrm{R}_X[m]$ will also and it will be the square of the expected value of X.

If X has a periodic component then $\mathrm{R}_X(\tau)$ or $\mathrm{R}_X[m]$ will also, with the same period.

Properties of Autocorrelation

If $\{X(t)\}$ is ergodic with zero mean and no periodic components then

$$\lim_{|\tau|\to\infty} R_X(\tau) = 0 \text{ or } \lim_{|m|\to\infty} R_X[m] = 0$$

Only autocorrelation functions for which

$$\mathcal{F}\{R_X(\tau)\} \geq 0 \text{ for all } f \text{ or } \mathcal{F}\{R_X[m]\} \geq 0 \text{ for all } \Omega$$

are possible

A time shift of a function does not affect its autocorrelation

Autocovariance

Autocovariance is similar to autocorrelation. Autocovariance is the autocorrelation of the time-varying part of a signal.

$$C_X(\tau) = R_X(\tau) - E^2(X) \text{ or } C_X[m] = R_X[m] - E^2(X)$$

Crosscorrelation

Properties

$$\mathrm{R}_{XY}\left(\tau\right)=\mathrm{R}_{YX}\left(-\tau\right) \quad \text{or} \quad \mathrm{R}_{XY}\left[m\right]=\mathrm{R}_{YX}\left[-m\right]$$

$$\left|\mathrm{R}_{XY}\left(\tau\right)\right|\le\sqrt{\mathrm{R}_{X}\left(0\right)\mathrm{R}_{Y}\left(0\right)} \quad \text{or} \quad \left|\mathrm{R}_{XY}\left[m\right]\right|\le\sqrt{\mathrm{R}_{X}\left[0\right]\mathrm{R}_{Y}\left[0\right]}$$

If two stochastic processes X and Y are statistically independent

$$\mathrm{R}_{XY}\left(\tau\right)=\mathrm{E}\left(X\right)\mathrm{E}\left(Y^{*}\right)=\mathrm{R}_{YX}\left(\tau\right) \quad \text{or} \quad \mathrm{R}_{XY}\left[m\right]=\mathrm{E}\left(X\right)\mathrm{E}\left(Y^{*}\right)=\mathrm{R}_{YX}\left[m\right]$$

If X is stationary CT and X' is its time derivative

$$\mathrm{R}_{XX'}\left(\tau\right)=\frac{d}{d\tau}\left(\mathrm{R}_{X}\left(\tau\right)\right) \qquad \mathrm{R}_{X'}\left(\tau\right)=-\frac{d^{2}}{d\tau^{2}}\left(\mathrm{R}_{X}\left(\tau\right)\right)$$

If $\mathrm{Z}\left(t\right)=\mathrm{X}\left(t\right)\pm\mathrm{Y}\left(t\right)$ and X and Y are independent and at least one of them has a zero mean

$$\mathrm{R}_{Z}\left(\tau\right)=\mathrm{R}_{X}\left(\tau\right)+\mathrm{R}_{Y}\left(\tau\right)$$

Power Spectral Density

In applying frequency-domain techniques to the analysis of random signals the natural approach is to Fourier transform the signals.

Unfortunately the Fourier transform of a stochastic process does not, strictly speaking, exist because it has infinite signal energy.

But the Fourier transform of a truncated version of a stochastic process does exist.

Power Spectral Density

For a CT stochastic process let

$$\mathrm{X}_T\left(t\right)=\left\{\begin{matrix}\mathrm{X}\left(t\right) & , & \left|t\right|\le T/2\\ 0 & , & \left|t\right|>T/2\end{matrix}\right\}=\mathrm{X}\left(t\right)\mathrm{rect}\left(t/T\right)$$

The Fourier transform is

$$\mathsf{F}\left(\mathrm{X}_T\left(t\right)\right)=\int_{-\infty}^{\infty}\mathrm{X}_T\left(t\right)e^{-j2\pi ft}dt\quad,\quad T<\infty$$

Using Parseval's theorem,

$$\int_{-T/2}^{T/2}\left|\mathrm{X}_T\left(t\right)\right|^2dt=\int_{-\infty}^{\infty}\left|\mathsf{F}\left(\mathrm{X}_T\left(t\right)\right)\right|^2df$$

Dividing through by T,

$$\frac{1}{T}\int_{-T/2}^{T/2}\mathrm{X}_T^2\left(t\right)dt=\frac{1}{T}\int_{-\infty}^{\infty}\left|\mathsf{F}\left(\mathrm{X}_T\left(t\right)\right)\right|^2df$$

Power Spectral Density

$$\frac{1}{T}\int_{-T/2}^{T/2} \mathrm{X}_T^2(t)\,dt = \frac{1}{T}\int_{-\infty}^{\infty} \left|\mathsf{F}\left(\mathrm{X}_T(t)\right)\right|^2 df$$

← Average signal power over time, T

If we let T approach infinity, the left side becomes the average power over all time. On the right side, the Fourier transform is not defined in that limit. But it can be shown that even though the Fourier transform does not exist, *its expected value does*. Then

$$\mathrm{E}\left(\frac{1}{T}\int_{-T/2}^{T/2} \mathrm{X}_T^2(t)\,dt\right) = \mathrm{E}\left(\frac{1}{T}\int_{-\infty}^{\infty} \left|\mathsf{F}\left(\mathrm{X}_T(t)\right)\right|^2 df\right)$$

Power Spectral Density

Taking the limit as T approaches infinity,

$$\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\mathrm{E}\left(\mathrm{X}^2\right)dt=\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty}\mathrm{E}\left[\left|\mathsf{F}\left(\mathrm{X}_T\left(t\right)\right)\right|^2\right]df$$

$$\left\langle\mathrm{E}\left(X^2\right)\right\rangle=\int_{-\infty}^{\infty}\lim_{T\to\infty}\mathrm{E}\left[\frac{\left|\mathsf{F}\left(\mathrm{X}_T\left(t\right)\right)\right|^2}{T}\right]df$$

The integrand on the right side is identified as *power spectral density (PSD)*.

$$\mathrm{G}_X\left(f\right)=\lim_{T\to\infty}\mathrm{E}\left(\frac{\left|\mathsf{F}\left(\mathrm{X}_T\left(t\right)\right)\right|^2}{T}\right)$$

(Proakis uses the notation $\Gamma_{XX}\left(F\right)$ for power spectral density.)

Power Spectral Density

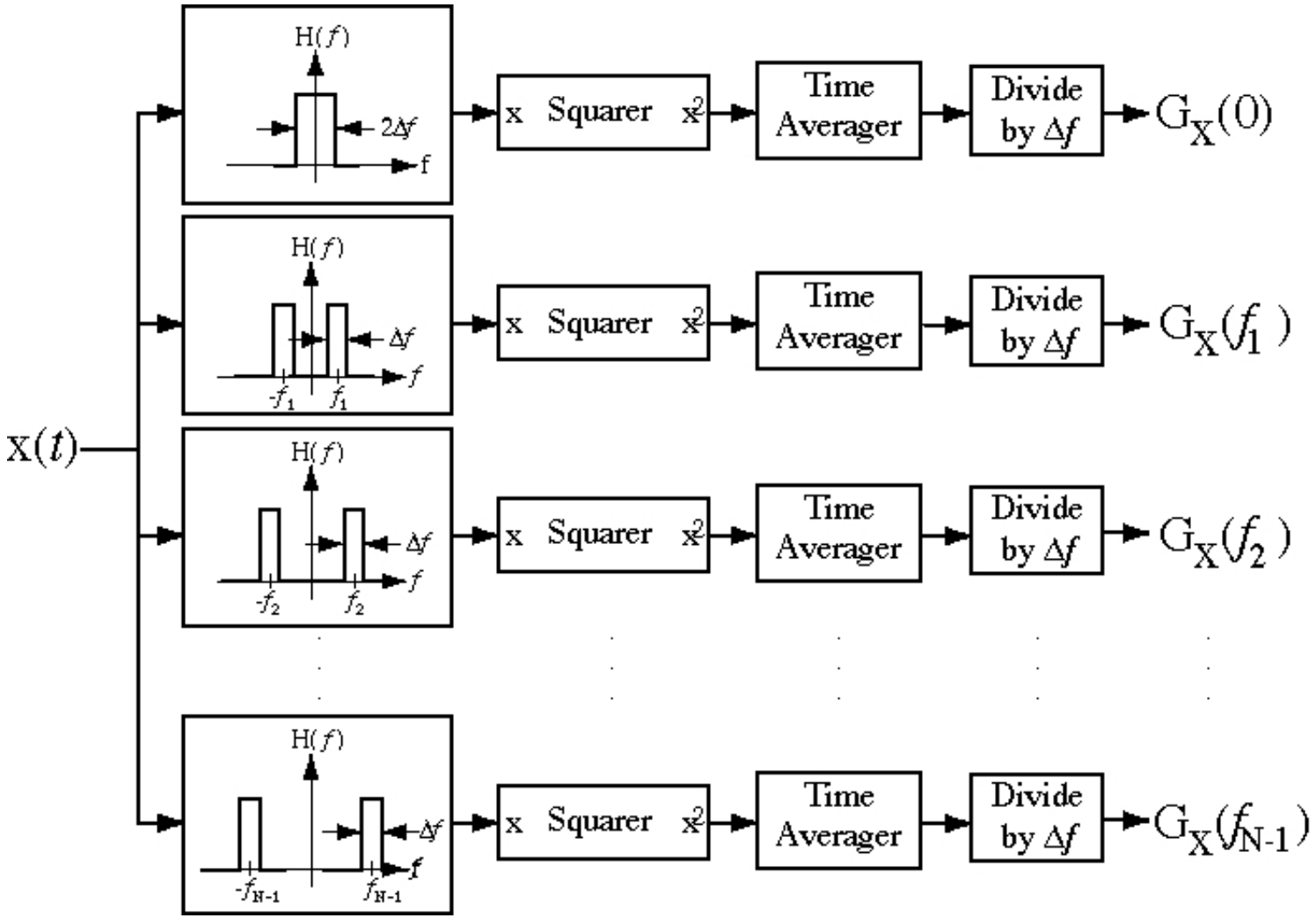
$$\int_{-\infty}^{\infty} G_X(f)\,df = \text{mean} - \text{squared value of } \{X(t)\}$$

$$\int_{-\infty}^{\infty} G_X(f)\,df = \text{average power of } \{X(t)\}$$

PSD is a description of the variation of a signal's power versus frequency.

PSD can be (and often is) conceived as single-sided, in which all the power is accounted for in positive frequency space.

PSD Concept


X(t)
H(f)
2Δf
f
x Squarer x²
Time Averager
Divide by Δf
G_X(0)
Δf
-f1
f1
G_X(f1)
-f2
f2
G_X(f2)
-fN-1
fN-1
G_X(fN-1)

PSD of DT Stochastic Processes

$$G_X(F) = \lim_{N\to\infty} E\left(\frac{\left| \mathcal{F}\left(X_T[n] \right) \right|^2}{N} \right)$$

$$\int_1 G_X(F)\, dF = \text{mean-squared value of } \{X[n]\}$$

or

$$\frac{1}{2\pi}\int_{2\pi} G_X(\Omega)\, d\Omega = \text{mean-squared value of } \{X[n]\}$$

where the Fourier transform is the discrete-time Fourier transform (DTFT) defined by

$$x[n] = \mathcal{F}^{-1}\left(X(F)\right) = \int_1 X(F) e^{j2\pi Fn} dF \xleftrightarrow{\mathcal{F}} X(F) = \mathcal{F}\left(x[n]\right) = \sum_{n=-\infty}^{\infty} x[n] e^{-j2\pi Fn}$$

$$x[n] = \mathcal{F}^{-1}\left(X(\Omega)\right) = \frac{1}{2\pi}\int_{2\pi} X(\Omega) e^{j\Omega n} d\Omega \xleftrightarrow{\mathcal{F}} X(\Omega) = \mathcal{F}\left(x[n]\right) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\Omega n}$$

PSD and Autocorrelation

It can be shown that PSD and autocorrelation form a Fourier transform pair.

$$G_X(f) = \mathcal{F}\left(R_X(\tau)\right) \text{ or } G_X(F) = \mathcal{F}\left(R_X[m]\right)$$

White Noise

White noise is a stochastic process whose PSD is constant.

$$G_X(f) = A \text{ or } G_X(F) = A$$

For CT signals, signal power is the integral of PSD over all frequency space. Therefore the power of white noise is infinite.

$$E(X^2) = \int_{-\infty}^{\infty} A\,df \to \infty$$

No real physical process may have infinite signal power. Therefore CT white noise cannot exist. However, many real and important CT stochastic processes have a PSD that is almost constant over a very wide frequency range.

White Noise

In many kinds of CT noise analysis, a type of random variable known as *bandlimited white noise* is used. Bandlimited white noise is simply the response of an ideal lowpass filter which is excited by white noise.

The PSD of bandlimited white noise is constant over a finite frequency range and zero outside that range.

Bandlimited white noise has finite signal power.

Cross Power Spectral Density

PSD is the Fourier transform of autocorrelation.

Cross power spectral density is the Fourier transform of cross correlation.

$$R_{XY}(t) \overset{F}{\longleftrightarrow} G_{XY}(f) \text{ or } R_{XY}[n] \overset{F}{\longleftrightarrow} G_{XY}(F)$$

Properties:

$$G_{XY}(f) = G^{*}_{YX}(f) \text{ or } G_{XY}(F) = G^{*}_{YX}(F)$$

$$\text{Re}\left(G_{XY}(f)\right) \text{ and } \text{Re}\left(G_{YX}(f)\right) \text{ are both even}$$

$$\text{Im}\left(G_{XY}(f)\right) \text{ and } \text{Im}\left(G_{YX}(f)\right) \text{ are both odd}$$

Time-Domain Linear System Analysis

For any linear, time-invariant (LTI) system, the response y is the convolution of the excitation x with the impulse response h.

$$y(t) = x(t) * h(t) \text{ or } y[n] = x[n] * h[n]$$

In the case of non-deterministic random processes this operation cannot be done because the signals are random and cannot, therefore, be described mathematically.

If $X(t)$ excites a system and $Y(t)$ is the response then the convolution integral is

$$Y(t) = \int_{-\infty}^{\infty} X(t-\tau)h(\tau)d\tau$$

Time-Domain Linear System Analysis

We cannot directly evaluate

$$\mathrm{Y}(t) = \int_{-\infty}^{\infty} \mathrm{X}(t-\tau)\mathrm{h}(\tau)\,d\tau$$

but we can find the expected value.

$$\mathrm{E}\left(\mathrm{Y}(t)\right) = \mathrm{E}\left(\int_{-\infty}^{\infty} \mathrm{X}(t-\tau)\mathrm{h}(\tau)\,d\tau\right)$$

If the random process is bounded and the system is stable

$$\mathrm{E}\left(\mathrm{Y}(t)\right) = \int_{-\infty}^{\infty} \mathrm{E}\left(\mathrm{X}(t-\tau)\right)\mathrm{h}(\tau)\,d\tau$$

Time-Domain Linear System Analysis

If the random process X is stationary

$$\mathrm{E}\left(\mathrm{Y}\left(t\right)\right) = \int_{-\infty}^{\infty} \mathrm{E}\left(\mathrm{X}\left(t-\tau\right)\right)\mathrm{h}\left(\tau\right)d\tau \Rightarrow \mathrm{E}\left(\mathrm{Y}\right) = \mathrm{E}\left(\mathrm{X}\right)\int_{-\infty}^{\infty} \mathrm{h}\left(\tau\right)d\tau$$

Using

$$\int_{-\infty}^{\infty} \mathrm{h}\left(t\right)dt = \mathrm{H}\left(0\right) \Rightarrow \mathrm{E}\left(Y\right) = \mathrm{E}\left(X\right)\mathrm{H}\left(0\right)$$

where H is the Fourier transform of h, we see that the expected value of the response is the expected value of the excitation multiplied by the zero-frequency response of the system. If the system is DT the corresponding result is

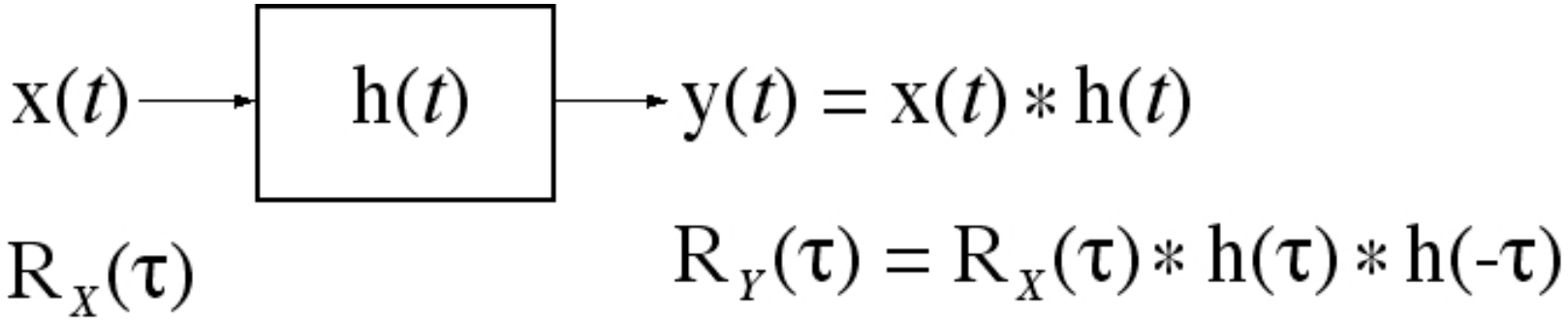
$$\mathrm{E}\left(Y\right) = \mathrm{E}\left(X\right)\sum_{n=-\infty}^{\infty} \mathrm{h}\left[n\right]$$

Time-Domain Linear System Analysis

It can be shown that the autocorrelation of the excitation and the autocorrelation of the response are related by

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau) \text{ or } R_Y[m] = R_X[m] * h[m] * h[-m]$$

This result leads to a way of thinking about the analysis of LTI systems with random excitation.



Time-Domain Linear System Analysis

It can be shown that the cross correlation between the excitation and the response is

$$\mathrm{R}_{XY}(\tau) = \mathrm{R}_{X}(\tau) * \mathrm{h}(\tau) \text{ or } \mathrm{R}_{XY}[m] = \mathrm{R}_{X}[m] * \mathrm{h}[m]$$

and

$$\mathrm{R}_{YX}(\tau) = \mathrm{R}_{X}(\tau) * \mathrm{h}(-\tau) \text{ or } \mathrm{R}_{YX}[m] = \mathrm{R}_{X}[m] * \mathrm{h}[-m]$$

Frequency-Domain Linear System Analysis

The frequency-domain relationship between excitation and response of an LTI system is the Fourier transform of the time-domain relationship.

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau) \overset{F}{\longleftrightarrow} G_Y(f) = G_X(f) H(f) H^*(f) = G_X(f) |H(f)|^2$$

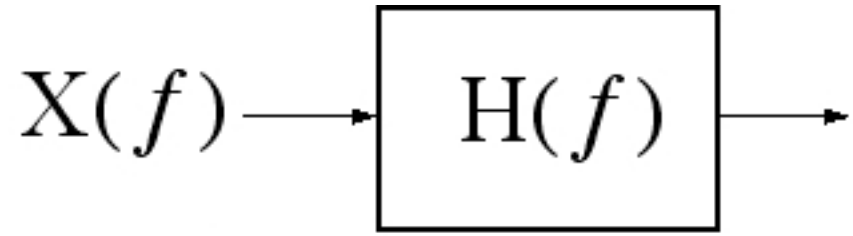
$$R_Y[m] = R_X[m] * h[m] * h[-m] \overset{F}{\longleftrightarrow} G_Y(F) = G_X(F) H(F) H^*(F) = G_X(F) |H(F)|^2$$

The mean-squared value of the response is

$$E(Y^2) = \int_{-\infty}^{\infty} G_Y(f)\,df = \int_{-\infty}^{\infty} G_X(f) |H(f)|^2\,df$$

$$E(Y^2) = \int_1 G_Y(F)\,dF = \int_1 G_X(F) |H(F)|^2\,dF$$

Frequency-Domain Linear System Analysis



$G_X(f)$

$$G_Y(f) = G_X(f)|H(f)|^2$$

White Noise Excitation of LTI Filters

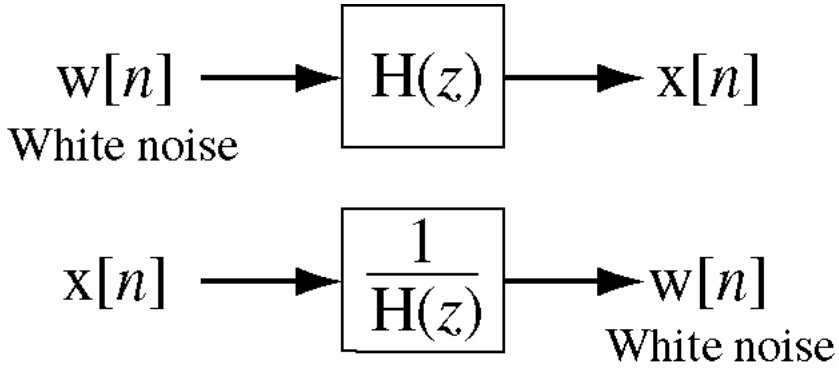
Consider an LTI filter excited by white noise $\mathrm{w}[n]$ with response $\mathrm{x}[n]$.

The autocorrelation of the response is

$$\mathrm{R}_{xx}[m] = \mathrm{R}_{ww}[m] * \mathrm{h}[m] * \mathrm{h}[-m].$$

If the excitation is white noise its autocorrelation is

$$\mathrm{R}_{xx}[m] = \sigma_w^2 \delta[m] * \mathrm{h}[m] * \mathrm{h}[-m] \ .$$



The power spectral density of the response is

$$\mathrm{G}_{xx}(F) = \mathrm{G}_{ww}(F)\left|\mathrm{H}(F)\right|^2 = \mathrm{G}_{ww}(F)\mathrm{H}(F)\mathrm{H}^*(F)$$

If the inverse $1/\mathrm{H}(z)$ exists and that system is excited by $\mathrm{x}[n]$ the response is $\mathrm{w}[n]$.

White Noise Excitation of LTI Filters

If the transfer function is the most common form, a ratio of polynomials in z, the power spectral density of the response is

$$\mathrm{G}_{xx}\left(F\right) = \sigma_w^2 \frac{\mathrm{B}\left(F\right)\mathrm{B}^*\left(F\right)}{\mathrm{A}\left(F\right)\mathrm{A}^*\left(F\right)}$$

where $\sum_{k=0}^{q} b_k \mathrm{w}\left[n-k\right] \overset{\mathrm{F}}{\longleftrightarrow} \mathrm{B}\left(F\right)$ and $1 + \sum_{k=1}^{p} a_k \mathrm{x}\left[n-k\right] \overset{\mathrm{F}}{\longleftrightarrow} \mathrm{A}\left(F\right)$

and where the excitation and response are related by the difference equation $\mathrm{x}\left[n\right] + \sum_{k=1}^{p} a_k \mathrm{x}\left[n-k\right] = \sum_{k=0}^{q} b_k \mathrm{w}\left[n-k\right].$

White Noise Excitation of LTI Filters

1. If $b_0 = 1$ and $b_k = 0$, $k > 0$ the frequency response of the system is $\mathrm{H}(F) = 1 / \mathrm{A}(F)$, it has no finite zeros and it is called an **autoregressive** (**AR**) system.
2. If $a_k = 0$, $k > 0$ the frequency response of the system is $\mathrm{H}(F) = \mathrm{B}(F)$ it has no finite poles and it is called a **moving average** (**MA**) system.
3. In the general case of both finite poles and finite zeros the system is called an **autoregressive moving average** (**ARMA**) system.

White Noise Excitation of LTI Filters

If we multiply both sides of $\mathrm{x}[n] + \sum_{k=1}^{p} a_k \mathrm{x}[n-k] = \sum_{k=0}^{q} b_k \mathrm{w}[n-k]$

by $\mathrm{x}^*[n+m]$ and then take the expectation of both sides we get

$$\mathrm{R}_{xx}[m] = -\sum_{k=1}^{p} a_k \mathrm{R}_{xx}[m+k] + \sum_{k=0}^{q} b_k \mathrm{R}_{wx}[m+k].$$

Using $\mathrm{x}[n] = \mathrm{h}[n] * \mathrm{w}[n] = \sum_{q=-\infty}^{\infty} \mathrm{h}[q]\mathrm{w}[n-q]$ we can show that

$\mathrm{R}_{wx}[m] = \mathrm{h}[m] * \mathrm{R}_{ww}[m]$ and, if $\mathrm{w}[n]$ is white noise

$\mathrm{R}_{wx}[m] = \mathrm{h}[m] * \sigma_w^2 \delta[m] = \sigma_w^2 \mathrm{h}[m].$

White Noise Excitation of LTI Filters

Combining $\mathrm{R}_{xx}[m] = -\sum_{k=1}^{p} a_k \mathrm{R}_{xx}[m+k] + \sum_{k=0}^{q} b_k \mathrm{R}_{wx}[m+k]$

with $\mathrm{R}_{wx}[m] = \sigma_w^2 \mathrm{h}[m]$ we get

$$\mathrm{R}_{xx}[m] = -\sum_{k=1}^{p} a_k \mathrm{R}_{xx}[m-k] + \sum_{k=0}^{q} b_k \sigma_w^2 \mathrm{h}[k-m]$$

For AR systems,

$$\mathrm{R}_{xx}[m] = -\sum_{k=1}^{p} a_k \mathrm{R}_{xx}[m-k] + \sigma_w^2 \mathrm{h}[-m]$$

For MA systems,

$$\mathrm{R}_{xx}[m] = \sigma_w^2 \sum_{k=0}^{q} b_k \mathrm{h}[k+m]$$

(These correspond to Eqs. 12.2.18, 12.2.19 and 12.2.21 in Proakis.)

White Noise Excitation of LTI Filters

Example

An AR system with frequency response $\mathrm{H}(F) = \dfrac{1}{1 - 0.2\,\mathrm{e}^{-j2\pi F}}$ is excited by white noise with $\mathrm{R}_{ww}[m] = \sigma_w^2 \delta[m]$. Find the impulse response and autocorrelation of the output signal.

$$\mathrm{H}(z) = \frac{1}{1 - 0.2\,\mathrm{e}^{-j2\pi F}} \Rightarrow \mathrm{h}[n] = 0.2^n\,\mathrm{u}[n]$$

$$\mathrm{G}_{xx}(F) = \sigma_w^2\,\mathrm{H}(F)\mathrm{H}^*(F) = \sigma_w^2 \frac{1}{1 - 0.2\,\mathrm{e}^{-j2\pi F}} \frac{1}{1 - 0.2\,\mathrm{e}^{j2\pi F}}$$

$$\mathrm{G}_{xx}(F) = \frac{\sigma_w^2}{1 - 0.4\cos(2\pi F) + 0.04} \Rightarrow \mathrm{R}_{xx}[m] = 1.042\sigma_w^2 (0.2)^{|m|}$$

White Noise Excitation of LTI Filters

Example

$$\frac{\mathrm{R}_{xx}\left[m\right]}{\mathrm{R}_{xx}\left[m-1\right]}=\frac{1.042\sigma_w^2\left(0.2\right)^{|m|}}{1.042\sigma_w^2\left(0.2\right)^{|m-1|}}=\frac{\left(0.2\right)^{m}}{\left(0.2\right)^{m-1}}=0.2\ \ ,\ \ m>0$$

Using the result for MA systems

$$\mathrm{R}_{xx}\left[m\right]=-\sum_{k=1}^{p}a_k\,\mathrm{R}_{xx}\left[m-k\right]+\sigma_w^2\,\mathrm{h}\left[-m\right]$$

$$\mathrm{R}_{xx}\left[m\right]=-\sum_{k=1}^{p}a_k\,\mathrm{R}_{xx}\left[m-k\right]\ \ ,\ \ m>0$$

$$\mathrm{R}_{xx}\left[m\right]=0.2\,\mathrm{R}_{xx}\left[m-1\right]\Rightarrow\frac{\mathrm{R}_{xx}\left[m\right]}{\mathrm{R}_{xx}\left[m-1\right]}=0.2\ \ ,\ \ m>0\ \ \text{Check.}$$

White Noise Excitation of LTI Filters

Example

$$\mathrm{R}_{xx}\left[m\right] = -\sum_{k=1}^{p} a_k \, \mathrm{R}_{xx}\left[m-k\right] + \sigma_w^2 \, \mathrm{h}\left[-m\right]$$

$$\mathrm{R}_{xx}\left[0\right] = -\sum_{k=1}^{p} a_k \, \mathrm{R}_{xx}\left[-k\right] + \sigma_w^2 \quad , \quad m = 0$$

$$\mathrm{R}_{xx}\left[0\right] = 0.2\,\mathrm{R}_{xx}\left[-1\right] + \sigma_w^2$$

$$1.042\sigma_w^2 = (0.2)1.042\sigma_w^2(0.2) + \sigma_w^2 = 1.042\sigma_w^2 \quad \text{Check}$$

White Noise Excitation of LTI Filters

Example

$$\mathrm{R}_{xx}\left[m\right] = -\sum_{k=1}^{p} a_k \,\mathrm{R}_{xx}\left[m-k\right] + \sigma_w^2 \,\mathrm{h}\left[-m\right] \quad , \quad m < 0$$

$$1.042\sigma_w^2 \left(0.2\right)^{|m|} = 0.2 \times 1.042\sigma_w^2 \left(0.2\right)^{|m-1|} + \sigma_w^2 \left(0.2\right)^{-m} \mathrm{u}\left[-m\right]$$

$$1.042\sigma_w^2 \left(0.2\right)^{-m} = 0.2 \times 1.042\sigma_w^2 \left(0.2\right)^{1-m} + \sigma_w^2 \left(0.2\right)^{-m}$$

$$1.042 = 0.2 \times 1.042 \times 0.2 + 1 = 1.042 \quad \text{Check}$$

White Noise Excitation of LTI Filters

Example

An MA system with frequency response $\mathrm{H}(F) = 1 + 0.8\,\mathrm{e}^{-j2\pi F}$ is excited by white noise with autocorrelation $\sigma_w^2 \delta[n]$. Find its impulse response and the autocorrelation of the output signal.

$$\mathrm{H}(F) = 1 + 0.8\,\mathrm{e}^{-j2\pi F} \Rightarrow \mathrm{h}[n] = \delta[n] + 0.8\delta[n-1]$$

$$\mathrm{R}_{xx}[m] = \sigma_w^2 \delta[m] * \mathrm{h}[m] * \mathrm{h}[-m]$$

$$= \sigma_w^2 \left(1.64\delta[m] + 0.8\delta[m-1] + 0.8\delta[m+1]\right)$$

Using the result $\mathrm{R}_{xx}[m] = \sigma_w^2 \sum_{k=0}^{q} b_k b_{k+m}$, $0 \le k+m \le q$

for MA systems,

$$\mathrm{R}_{xx}[m] = \sigma_w^2 \sum_{k=0}^{1} b_k b_{k+m} \quad , \quad 0 \le k+m \le 1$$

White Noise Excitation of LTI Filters

Example

$$R_{xx}[0] = \sigma_w^2 \left(b_0 b_0 + b_1 b_1 \right) = \sigma_w^2 \left(1 + 0.8^2 \right) = 1.64 \sigma_w^2$$

$$R_{xx}[1] = \sigma_w^2 \left(b_0 b_1 \right) = 0.8 \sigma_w^2$$

$$R_{xx}[-1] = \sigma_w^2 \left(b_1 b_0 \right) = 0.8 \sigma_w^2$$

Check.

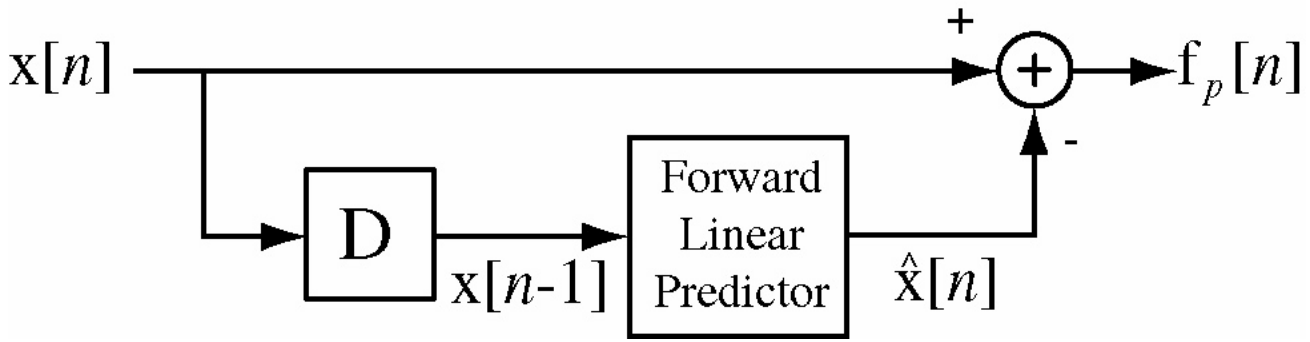
Forward and Backward Linear Prediction

Forward linear prediction is estimating a signal value at time n from a linear combination of signal values at previous times $n-1, n-2, \cdots, n-p$. The estimate is $\hat{x}[n] = -\sum_{k=1}^{p} a_p[k]\, x[n-k]$ where the "hat" on x means "an estimate of x". The difference between $x[n]$ and $\hat{x}[n]$ is called the **forward prediction error**

$$f_p[n] = x[n] - \hat{x}[n] = x[n] + \sum_{k=1}^{p} a_p[k]\, x[n-k]$$

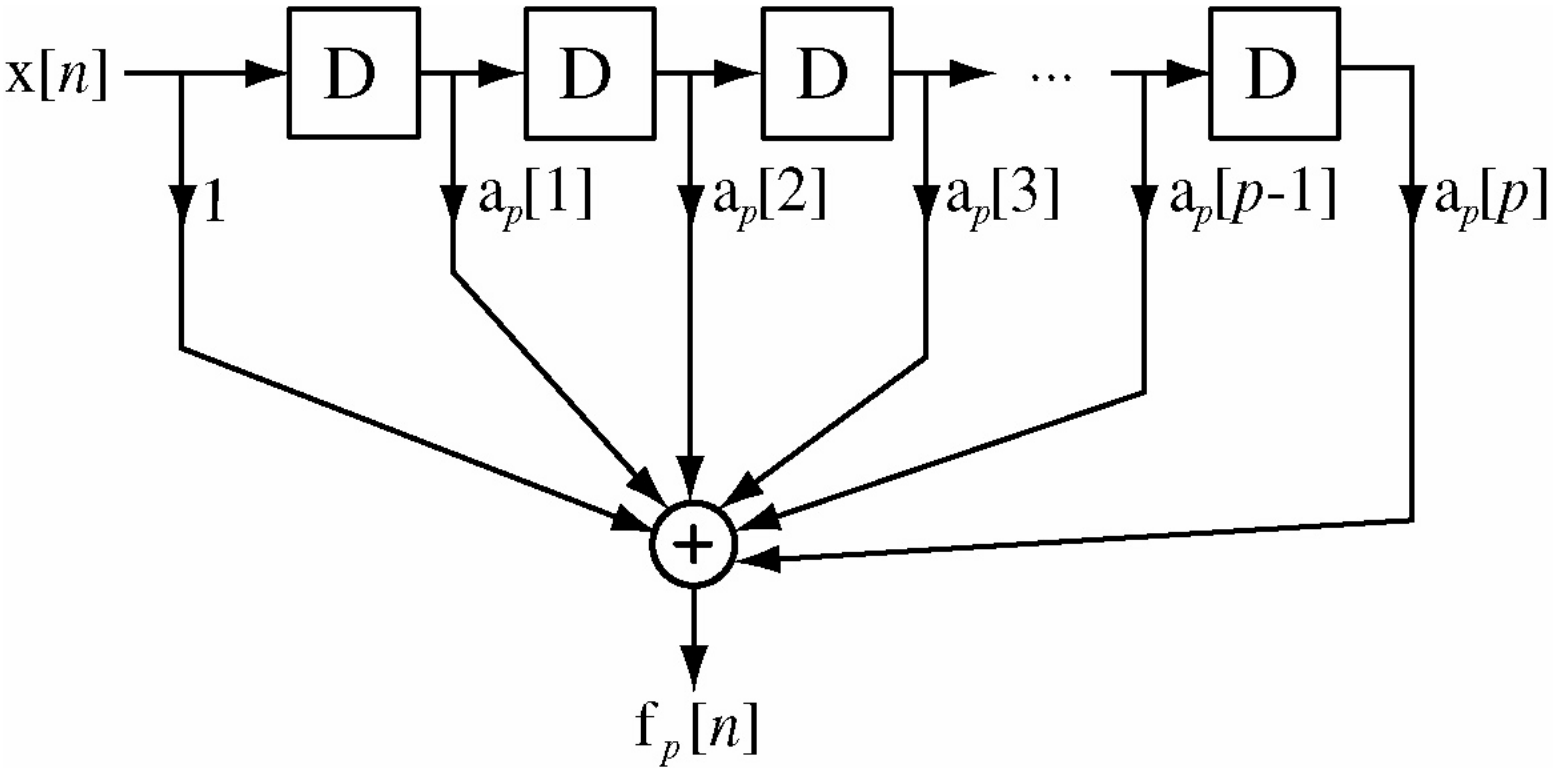
Forward and Backward Linear Prediction

A one-step forward linear predictor



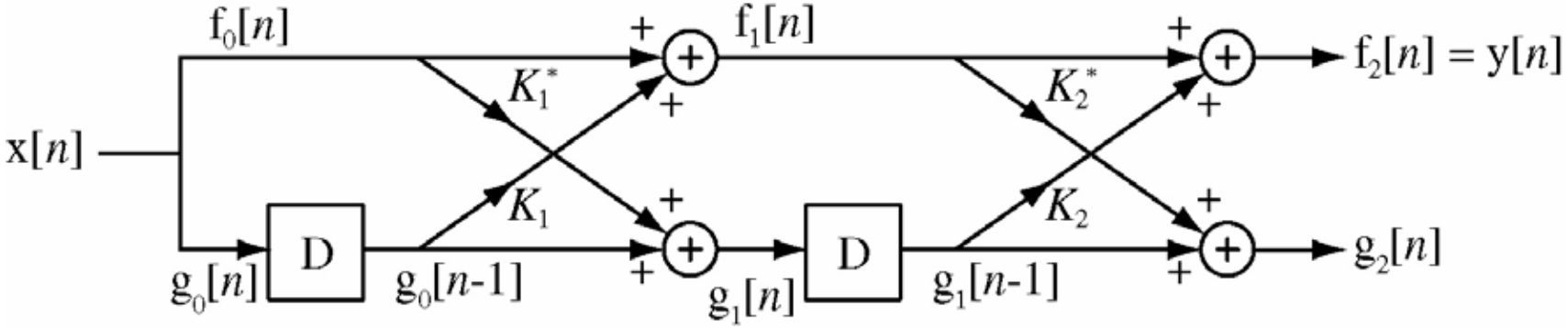
Forward and Backward Linear Prediction

A *p*-step prediction-error filter



Forward and Backward Linear Prediction

A forward linear predictor can be realized as a lattice.



$$f_0[n] = g_0[n] = x[n]$$

$$f_m[n] = f_{m-1}[n] + K_m\, g_{m-1}[n-1] \quad , \quad m = 0,1,2,\cdots,p$$

$$g_m[n] = K_m^*\, f_{m-1}[n] + g_{m-1}[n-1] \quad , \quad m = 0,1,2,\cdots,p$$

(Notice that the reflection coefficients are no longer identical in a single stage but are instead complex conjugates. This is done to handle the case of complex-valued signals.)

Forward and Backward Linear Prediction

From chapter 9

$$\mathrm{A}_0(z) = \mathrm{B}_0(z) = 1$$

$$\mathrm{A}_m(z) = \mathrm{A}_{m-1}(z) + K_m z^{-1} \mathrm{B}_{m-1}(z)$$

$$\mathrm{B}_m(z) = z^{-m} \mathrm{A}_m(1/z)$$

$$\mathrm{A}_{m-1}(z) = \frac{\mathrm{A}_m(z) - K_m \mathrm{B}_m(z)}{1 - K_m^2}$$

$$K_m = \alpha_m[m]$$

Forward and Backward Linear Prediction

The mean-squared forward prediction error is

$$\mathrm{E}\left(\left|\mathrm{f}_p[n]\right|^2\right) = \mathrm{E}\left(\sum_{k=0}^{p} \mathrm{a}_p[k]\,\mathrm{x}[n-k]\right)\mathrm{E}\left(\sum_{q=0}^{p} \mathrm{a}_p^*[q]\,\mathrm{x}^*[n-q]\right)$$

where it is understood that $\mathrm{a}_p[0] = 1$. This can be reduced to

$$\mathrm{E}\left(\left|\mathrm{f}_p[n]\right|^2\right) = \sum_{k=0}^{p}\sum_{q=0}^{p} \mathrm{a}_p[k]\,\mathrm{a}_p^*[q]\,\mathrm{R}_{xx}[k-q]$$

which can also be written (after considerable algebra) as

$$\mathrm{E}\left(\left|\mathrm{f}_p[n]\right|^2\right) = \mathrm{R}_{xx}[0] + 2\,\mathrm{Re}\left(\sum_{k=1}^{p} \mathrm{a}_p[k]\,\mathrm{R}_{xx}[k]\right)$$

$$+ \sum_{k=1}^{p}\left|\mathrm{a}_p[k]\right|^2 \mathrm{R}_{xx}[0] + 2\,\mathrm{Re}\left(\sum_{k=q+1}^{p}\sum_{q=1}^{p} \mathrm{a}_p[k]\,\mathrm{a}_p^*[q]\,\mathrm{R}_{xx}[k-q]\right)$$

Forward and Backward Linear Prediction

It can be shown that minimizing the mean-squared error leads a set of linear equations

$$\mathrm{R}_{xx}\left[l\right] = -\sum_{k=1}^{p} \mathrm{a}_{p}\left[k\right]\mathrm{R}_{xx}\left[l-k\right] \;\;,\;\; l = 1,2,\cdots,p$$

known as the **normal equations**. If the autocorrelation is known the "a" coefficients can be found from this set of equations.

The minimum mean-sqared error is

$$\min\left\{\mathrm{E}\left(\left|\mathrm{f}_{p}\left[n\right]\right|^{2}\right)\right\} = E_{p}^{\mathrm{f}} = \mathrm{R}_{xx}\left[0\right] + \sum_{k=1}^{p} \mathrm{a}_{p}\left[k\right]\mathrm{R}_{xx}\left[k\right]$$

Forward and Backward Linear Prediction

We can also "predict" in the backward direction. We can estimate $\mathrm{x}[n-p]$ from the values of $\mathrm{x}[n]$, $\mathrm{x}[n-1]$, $\cdots, \mathrm{x}[n-p+1]$.

The estimate is

$$\hat{\mathrm{x}}[n-p] = -\sum_{k=0}^{p-1} \mathrm{b}_p[k]\,\mathrm{x}[n-k]$$

and the backward "prediction" error is

$$\mathrm{g}_p[n] = \mathrm{x}[n-p] + \sum_{k=0}^{p-1} \mathrm{b}_p[k]\,\mathrm{x}[n-k] \; .$$

The minimum mean-squared error is the same as in the forward prediction case

$$\min\left(\mathrm{E}\left(\left|\mathrm{g}_p[n]\right|^2\right)\right) = E_p^g = E_p^{\mathrm{f}}$$

Optimal Reflection Coefficients

The forward prediction error in a lattice filter is

$$\mathrm{f}_m[n] = \mathrm{f}_{m-1}[n] + K_m\, \mathrm{g}_{m-1}[n-1]$$

Its mean-squared value is

$$\mathrm{E}\left(\left|\mathrm{f}_m[n]\right|^2\right) = \mathrm{E}\left\{\left(\mathrm{f}_{m-1}[n] + K_m\, \mathrm{g}_{m-1}[n-1]\right)\left(\mathrm{f}_{m-1}[n] + K_m\, \mathrm{g}_{m-1}[n-1]\right)\right\}$$

It can be shown that the optimal reflection coefficient to minimize the mean-squared error is

$$K_m = -\frac{\mathrm{E}\left(\mathrm{f}_{m-1}[n]\, \mathrm{g}_{m-1}[n-1]\right)}{\sqrt{E_{m-1}^f E_{m-1}^b}}$$

This is the negative of the correlation coefficient between the forward and backward errors. Then the prediction errors can be recursively computed by $E_m^f = \left(1 - K_m^2\right) E_{m-1}^f$.

AR Processes and Linear Prediction

There is a one-to-one correspondence between the parameters in an AR process and the predictor-coefficients of a *p*-th order predictor. If the process is actually AR and *p*th order, they are the same. If the output signal from an AR process excited by white noise is applied to the corresponding predictor, the output signal from the predictor will be white noise. So the prediction filter is often called a “whitening” filter.

Properties of Linear Prediction-Error Filters

It can be shown that if the reflection coefficients in a lattice-type linear prediction-error filter are all less than one in magnitude, that the zeros of A(z) all lie inside the unit circle. This makes A(z) **minimum-phase**. All the zeros of B(z) lie outside the unit circle making it **maximum-phase**.

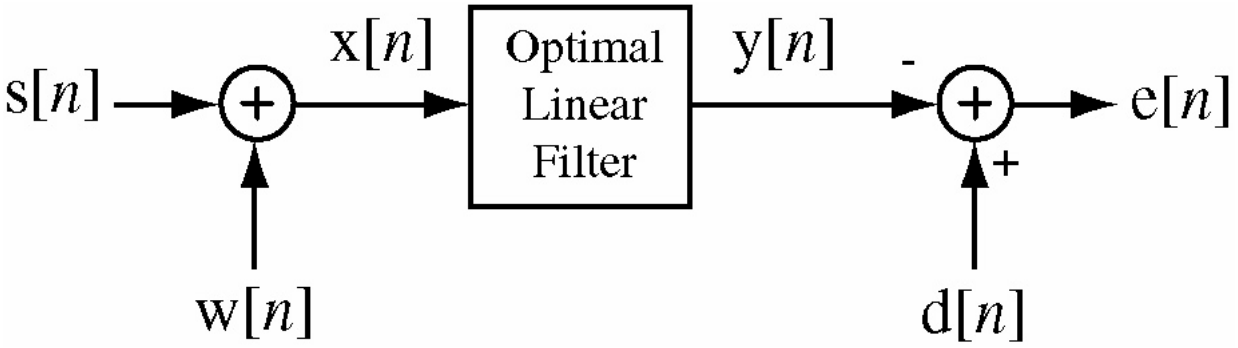
Wiener Filters

Below is a general system model for an optimal linear filter called a **Wiener** filter. It makes the best estimate of $d[n]$ based on $x[n]$, which contains a signal $s[n]$ plus noise $w[n]$, and the autocorrelation functions for $s[n]$ and $w[n]$. The estimation error is $e[n]$.

If $d[n] = s[n]$, the estimation problem is called **filtering**.

If $d[n] = s[n + n_0]$, $n_0 > 0$, it is called **prediction**.

If $d[n] = s[n - n_0]$, $n_0 > 0$, it is called **smoothing**.



Wiener Filters

The filter is optimized by minimizing the mean-squared estimation error

$$\mathrm{E}\left(\left|\mathrm{e}[n]\right|^2\right) = \mathrm{E}\left(\left|\mathrm{d}[n] - \sum_{m=0}^{M-1} \mathrm{h}[m]\,\mathrm{x}[n-m]\right|^2\right).$$

A set of equations called the **Wiener - Hopf** equations

$$\sum_{m=0}^{M-1} \mathrm{h}[m]\,\mathrm{R}_{xx}[l-m] = \mathrm{R}_{dx}[-l] \quad , \quad l = 0,1,\cdots,M-1$$

is used to find the impulse response of the optimal linear filter.

In Proakis' notation, $\sum_{k=0}^{M-1} \mathrm{h}(k)\gamma_{\mathrm{xx}}(l-k) = \gamma_{dx}(l)$. The sign difference on the right is caused by using a different definition of autocorrelation. But both sets of equations yield an optimal impulse response.

Wiener Filters

The Wiener-Hopf equations can be compactly written in matrix form as $\mathbf{R}_M \mathbf{h}_M = \mathbf{r}_d$, where $\mathbf{R}_M$ is an $M \times M$ matrix with elements $\mathrm{R}_{xx}\left[l-m\right]$ and $\mathbf{r}_d$ is an $M \times 1$ matrix with elements $\mathrm{R}_{dx}\left[-l\right]$. The solution is $\mathbf{h}_M = \mathbf{R}_M^{-1}\mathbf{r}_d$ and the minimum mean-squared error achieved is

$$\mathrm{E}\left(\left|\mathrm{e}\left[n\right]\right|^2\right)_{\min} = \sigma_d^2 - \mathbf{r}_{\mathrm{d}}^T \mathbf{R}_M^{-1}\mathbf{r}_d$$

If $\mathrm{d}\left[n\right] = \mathrm{s}\left[n\right]$ and if $\mathrm{s}\left[n\right]$ and $\mathrm{w}\left[n\right]$ are independent the Wiener-Hopf equations reduce to

$$\sum_{m=0}^{M-1} \mathrm{h}\left[m\right]\left(\mathrm{R}_{ss}\left[l-m\right] + \mathrm{R}_{ww}\left[l-m\right]\right) = \mathrm{R}_{ss}\left[-l\right] \quad , \quad l = 0,1,\cdots,M-1$$

Wiener Filters

The Wiener-Hopf equations for IIR filters are similar to those for FIR filters except that the impulse response

$$\sum_{m=0}^{\infty} \mathrm{h}[k]\,\mathrm{R}_{xx}[l-m] = \mathrm{R}_{dx}[-l] \quad , \quad l \geq 0$$

has an infinite duration and the minimum mean-squared error is

$$MMSE_{\infty} = \sigma_d^2 - \sum_{m=0}^{\infty} \mathrm{h}_{opt}[m]\,\mathrm{R}_{dx}^{*}[-m]$$

Wiener Filters

A stationary random process $\mathrm{x}[n]$ with autocorrelation $\mathrm{R}_{xx}[m]$ and power spectral density $\mathrm{G}_{\mathrm{xx}}(F)$ can be represented by an equivalent innovations process $\mathrm{i}[n]$ by passing $\mathrm{x}[n]$ through a whitening filter with transfer function $1/\mathrm{G}_{\min}(z)$ where $\mathrm{G}_{\min}(z)$ is the minimum-phase part from spectral factorization of $\mathrm{G}_{\mathrm{xx}}(F)$

$$\mathrm{G}_{xx}(F) = \sigma_i^2 \mathrm{G}_{\min}(F)\mathrm{G}_{\max}(F) = \sigma_i^2 \mathrm{G}_{\min}(F)\mathrm{G}_{\min}(-F)$$

Wiener Filters

It can be shown that the optimal IIR causal Wiener filter has the frequency response

$$\mathrm{H}_{opt}\left(F\right) = \frac{1}{\sigma_i^2 \, \mathrm{G}_{\min}\left(F\right)} \left[\frac{\mathrm{G}_{dx}^{*}\left(F\right)}{\mathrm{G}_{\min}\left(-F\right)} \right]_{+}$$

where $\mathrm{G}_{dx}\left(F\right)$ is the cross power spectral density between $\mathrm{d}\left[n\right]$ and $\mathrm{x}\left[n\right]$ and the subscript "+" on the square brackets means "the causal part".

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

Let $x[n] = s[n] + w[n]$ where $s[n]$ is an AR process that satisfies the equation $s[n] = 0.6\,s[n-1] + v[n]$ where $v[n]$ is a white noise sequence with variance $\sigma_v^2 = 0.64$ and $w[n]$ is a white noise sequency with variance $\sigma_w^2 = 1$. Design a Wiener filter to optimally estimate the signal $s[n]$ and delayed versions of the signal $s[n - n_0]$.

The system impulse response is $h[n] = (0.6)^n u[n]$ and the transfer function is $H(z) = \dfrac{1}{1 - 0.6z^{-1}} = \dfrac{z}{z - 0.6}$.

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

The power spectral density of $\mathrm{s}[n]$ is

$$\mathrm{G}_{ss}(z) = \mathrm{G}_{vv}(z)\left|\mathrm{H}(z)\right|^2 = \sigma_{\mathrm{v}}^2\left|\mathrm{H}(z)\right|^2$$

$$= 0.64\frac{1}{1-0.6z^{-1}}\frac{1}{1-0.6z} = \frac{0.64}{1.36-0.6\left(z^{-1}+z\right)}$$

The power spectral density of $\mathrm{x}[n]$ is

$$\mathrm{G}_{xx}(z) = \mathrm{G}_{ss}(z) + \mathrm{G}_{ww}(z) = \frac{0.64}{1.36-0.6\left(z^{-1}+z\right)} + 1 = \frac{2-0.6\left(z^{-1}+z\right)}{1.36-0.6\left(z^{-1}+z\right)}$$

This can be spectrally factored into the form

$$\mathrm{G}_{xx}(z) = \frac{\left(a-bz^{-1}\right)\left(a-bz\right)}{\left(1-0.6z^{-1}\right)\left(1-0.6z\right)}$$

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

After spectral factorization

$$G_{xx}(z) = 1.8\frac{\left(1-(1/3)z^{-1}\right)\left(1-(1/3)z\right)}{\left(1-0.6z^{-1}\right)\left(1-0.6z\right)}$$

Therefore, if $G_{xx}(z) = \sigma_i^2 G_{\min}(z) G_{\min}\left(z^{-1}\right)$, $\sigma_i^2 = 1.8$ and

$$G_{\min}(z) = \frac{1-(1/3)z^{-1}}{1-0.6z^{-1}} = \frac{z-1/3}{z-0.6}$$

The cross correlation between $d[n]$ and $x[n]$ is the same as the cross correlation between $s[n]$ and $x[n]$ because we are doing filtering and $d[n] = s[n]$.

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

$$R_{dx}[m] = R_{sx}[m] = E\left(s[n]\,x[n+m]\right) = E\left(s[n]\left(s[n+m] + w[n+m]\right)\right)$$

$$R_{dx}[m] = E\left(s[n]\,s[n+m]\right) + E\left(s[n]\,w[n+m]\right) = R_{ss}[m] + \underbrace{R_{sw}[m]}_{=0} = R_{ss}[m]$$

$$\text{Therefore } G_{dx}(z) = G_{ss}(z) = \frac{0.64}{1.36 - 0.6\left(z^{-1} + z\right)} = \frac{0.64}{\left(1 - 0.6z^{-1}\right)\left(1 - 0.6z\right)}$$

$$\text{and } \left[\frac{G_{dx}\left(z^{-1}\right)}{G_{\min}\left(z^{-1}\right)}\right]_+ = \left[\frac{\dfrac{0.64}{\left(1-0.6z\right)\left(1-0.6z^{-1}\right)}}{\dfrac{1-(1/3)z}{1-0.6z}}\right]_+ = \left[\frac{0.64}{\left(1-(1/3)z\right)\left(1-0.6z^{-1}\right)}\right]_+$$

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

We want to split this into the causal and anti-causal parts and retain only the causal part. The causal part has the poles inside the unit circle. So we want a partial-fraction expansion of the form

$$\frac{0.64}{\left(1-(1/3)z\right)\left(1-0.6z^{-1}\right)} = \frac{K_1 z}{1-(1/3)z} + \frac{K_2}{1-0.6z^{-1}}$$

Solving, $K_1 = 0.8/3,\ K_2 = 0.8$ and the causal part is

$$\left[\frac{\mathrm{G}_{dx}\left(z^{-1}\right)}{\mathrm{G}_{\min}\left(z^{-1}\right)}\right]_+ = \frac{0.8}{1-0.6z^{-1}}$$

$$\mathrm{H}_{opt}(z) = \frac{1}{\sigma_i^2 \dfrac{1-(1/3)z^{-1}}{1-0.6z^{-1}}} \frac{0.8}{1-0.6z^{-1}} = \frac{1}{1.8} \frac{0.8}{1-(1/3)z^{-1}} = \frac{4/9}{1-(1/3)z^{-1}}$$

$$\mathrm{h}_{opt}[n] = (4/9)(1/3)^n \,\mathrm{u}[n]$$

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

Next consider the case in which we are not filtering but instead smoothing. Now the cross correlation between $\mathrm{d}[n]$ and $\mathrm{x}[n]$ is not the same as the cross correlation between $\mathrm{d}[n]$ and $\mathrm{x}[n]$. Let $\mathrm{d}[n] = \mathrm{s}[n - n_0]$. Then

$$\mathrm{R}_{dx}[m] = \mathrm{E}\left(\mathrm{s}[n - n_0]\,\mathrm{x}[n + m]\right) = \mathrm{E}\left(\mathrm{s}[n - n_0]\left(\mathrm{s}[n + m] + \mathrm{w}[n + m]\right)\right)$$

$$\mathrm{R}_{dx}[m] = \mathrm{E}\left(\mathrm{s}[n - n_0]\,\mathrm{s}[n + m]\right) + \mathrm{E}\left(\mathrm{s}[n - n_0]\,\mathrm{w}[n + m]\right)$$

$$= \mathrm{R}_{\mathrm{ss}}[m + n_0] + \underbrace{\mathrm{R}_{\mathrm{sw}}[m + n_0]}_{=0} = \mathrm{R}_{\mathrm{ss}}[m + n_0]$$

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

$$\text{Therefore } G_{dx}(z) = G_{ss}(z)z^{n_0} = \frac{0.64z^{n_0}}{1.36 - 0.6\left(z^{-1} + z\right)} = \frac{0.64z^{n_0}}{\left(1 - 0.6z^{-1}\right)\left(1 - 0.6z\right)}$$

$$\text{and } \left[\frac{G_{dx}\left(z^{-1}\right)}{G_{\min}\left(z^{-1}\right)}\right]_+ = \left[\frac{\dfrac{0.64z^{-n_0}}{\left(1 - 0.6z\right)\left(1 - 0.6z^{-1}\right)}}{\dfrac{1 - (1/3)z}{1 - 0.6z}}\right]_+ = \left[\frac{0.64z^{-n_0}}{\left(1 - (1/3)z\right)\left(1 - 0.6z^{-1}\right)}\right]_+$$

Expanding in partial fractions as before

$$\frac{0.64z^{-n_0}}{\left(1 - (1/3)z\right)\left(1 - 0.6z^{-1}\right)} = z^{-n_0}\left[-\frac{0.8z}{z - 3} + \frac{0.8z}{z - 0.6}\right]$$

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

$$0.8(3)^n \, \mathrm{u}[-(n+1)] + 0.8(0.6)^n \, \mathrm{u}[n] \overset{z}{\longleftrightarrow} -\frac{0.8z}{z-3} + \frac{0.8z}{z-0.6}$$

$$0.8(3)^{n-n_0} \, \mathrm{u}[-(n-n_0+1)] + 0.8(0.6)^{n-n_0} \, \mathrm{u}[n-n_0] \overset{z}{\longleftrightarrow} z^{-n_0}\left[-\frac{0.8z}{z-3} + \frac{0.8z}{z-0.6}\right]$$

The causal part of the inverse transform is

$$\left\{0.8(3)^{n-n_0} \, \mathrm{u}[-(n-n_0+1)] + 0.8(0.6)^{n-n_0} \, \mathrm{u}[n-n_0]\right\} \mathrm{u}[n]$$

which can be written as

$$0.8(3)^{n-n_0}\left(\mathrm{u}[n] - \mathrm{u}[n-n_0]\right) + 0.8(0.6)^{n-n_0} \, \mathrm{u}[n-n_0]$$

Its z transform is

$$0.8\left\{\frac{1}{3}\frac{(3)^{-n_0} - z^{-n_0}}{(1/3) - z^{-1}} + \frac{z^{-n_0}}{1 - 0.6z^{-1}}\right\}$$

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

$$\left[\frac{\mathrm{G}_{dx}\left(z^{-1}\right)}{\mathrm{G}_{\min}\left(z^{-1}\right)}\right]_{+} = \left(0.8/3\right)\frac{\left(1/3\right)^{n_0} - z^{-n_0}}{1/3 - z^{-1}} + \frac{0.8z^{-n_0}}{1-0.6z^{-1}}$$

$$\mathrm{H}_{opt}\left(z\right) = \frac{1}{\sigma_i^2 \dfrac{1-\left(1/3\right)z^{-1}}{1-0.6z^{-1}}}\left[\left(0.8/3\right)\frac{\left(1/3\right)^{n_0} - z^{-n_0}}{\left(1/3\right) - z^{-1}} + \frac{0.8z^{-n_0}}{1-0.6z^{-1}}\right]$$

which can be written as

$$\mathrm{H}_{opt}\left(z\right) = \frac{5}{9}z^{-\left(n_0-1\right)}\left[\left(0.8/3\right)\left(1/3\right)^{n_0-1}\frac{z-0.6}{z-1/3}\frac{z^{n_0} - \left(3\right)^{n_0}}{z-3} + \frac{0.8}{z-1/3}\right]$$

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

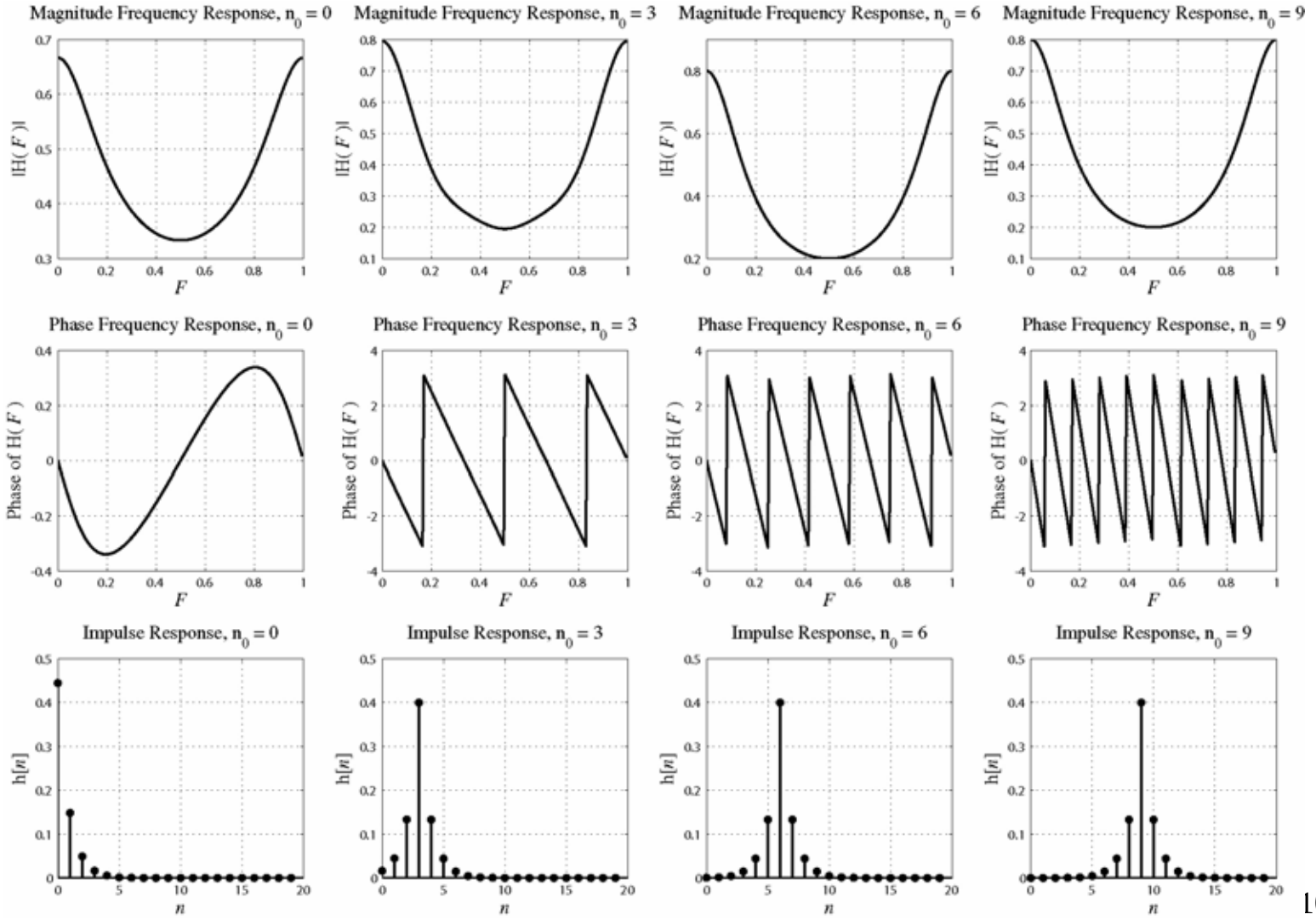
The impulse response of the filter is the inverse transform of this transfer function. Finding a general expression for it is very tedious. But we can see its form by finding the frequency response

$$\mathrm{H}_{opt}\left(e^{j\Omega}\right)=\frac{5}{9}e^{-j\Omega\left(n_0-1\right)}\left[\left(0.8/3\right)\left(1/3\right)^{n_0-1}\frac{e^{j\Omega}-0.6}{e^{j\Omega}-1/3}\frac{e^{j\Omega n_0}-\left(3\right)^{n_0}}{e^{j\Omega}-3}+\frac{0.8}{e^{j\Omega}-1/3}\right]$$

and then computing the impulse response numerically, using the fast Fourier transform.

Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)



Wiener Filters

IIR Wiener Filter Example (An extension of Example 12.7.2 in Proakis)

Now consider the case in which we use a non-causal filter.

Then

$$H_{opt}(z) = \frac{G_{dx}(z^{-1})}{G_{xx}(z^{-1})} = \frac{\dfrac{0.64}{(1-0.6z^{-1})(1-0.6z)}}{1.8\dfrac{(1-(1/3)z^{-1})(1-(1/3)z)}{(1-0.6z^{-1})(1-0.6z)}}$$

$$H_{opt}(z) = \frac{0.3555}{(1-0.333z^{-1})(1-0.333z)} = -1.067\left[\frac{-1/8}{z-(1/3)} + \frac{9/8}{z-3}\right]$$

$$h_{opt}[n] = 1.067(3/8)(1/3)^{|n|} = 0.4(1/3)^{|n|}$$

This impulse response is virtually identical to the $n_0 = 9$ impulse response for the smoothing case except that it is centered at $n = 0$ instead of $n = 9$.