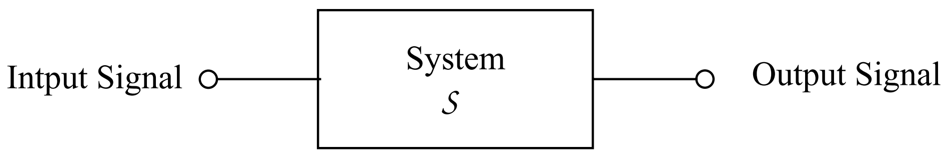
1. Response of Linear Time-Invariant Systems to Random Inputs

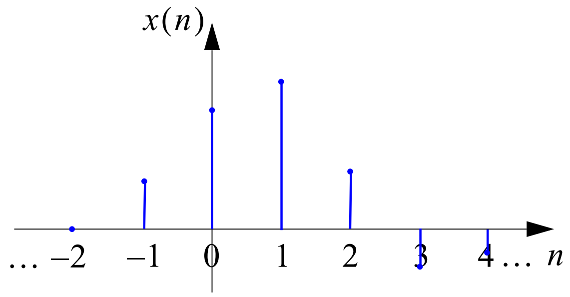
System:



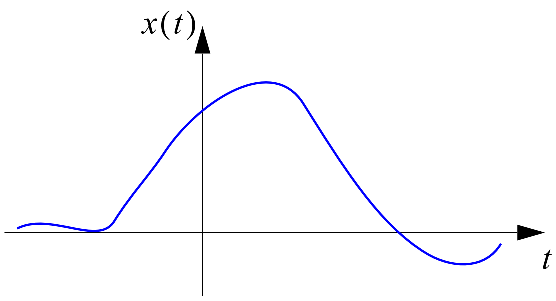
We look at a system as a black box which generates an output signal depending on the input signal and possibly some initial conditions.

We consider two types of signals:

- ***Discrete-time signals or sequences***



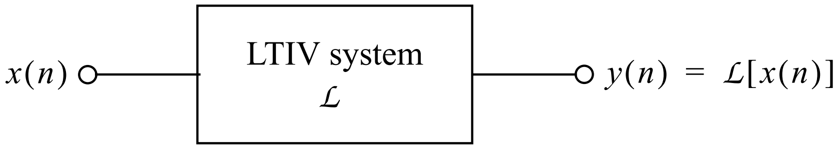
- ***Continuous-time signals***



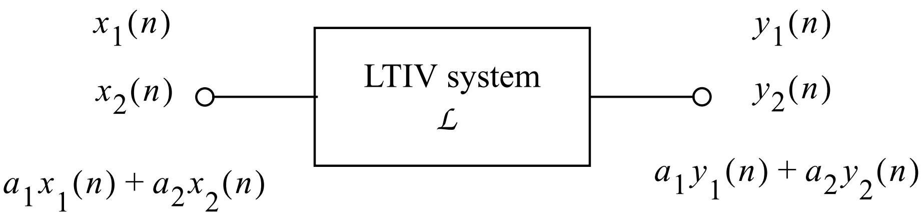
Discrete-time signals are obtained by sampling continuous-time signals.

1.1. Discrete-time linear time-invariant (LTIV) systems

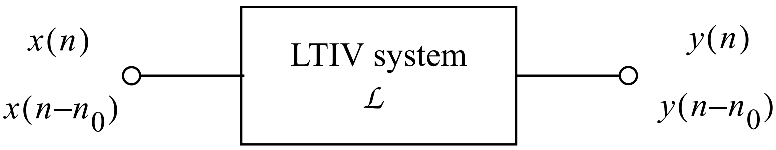
1.1.1. Discrete-time LTIV system



Linear:



Time-invariant:



1.1.2. Steady-state description of a LTIV system

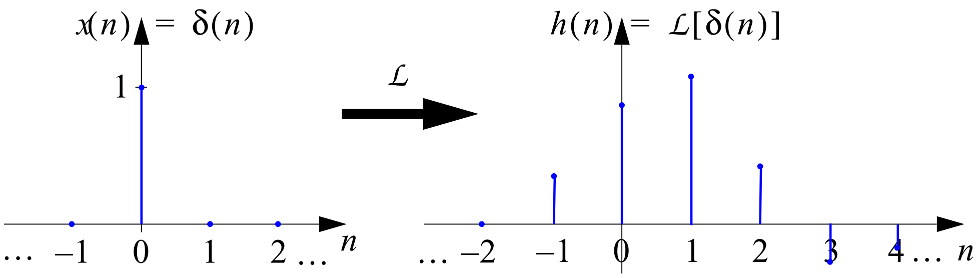
- ***Impulse response:***

  The impulse response (IR) $h(n)$ of $\mathcal{L}$ is the response of $\mathcal{L}$ to the unit pulse

$$\boxed{\delta(n) \equiv \begin{cases} 1 \; ; & n = 0 \\ 0 \; ; & n \neq 0 \end{cases} ,}$$

  namely

$$\boxed{h(n) = \mathcal{L}[\delta(n)]}$$



- ***Stable LTIV system:***

  A LTIV system is stable if its response to a bounded signal is bounded.

  It can be shown [to this end we need (1.1)] that a LTIV system is stable if, and only if,

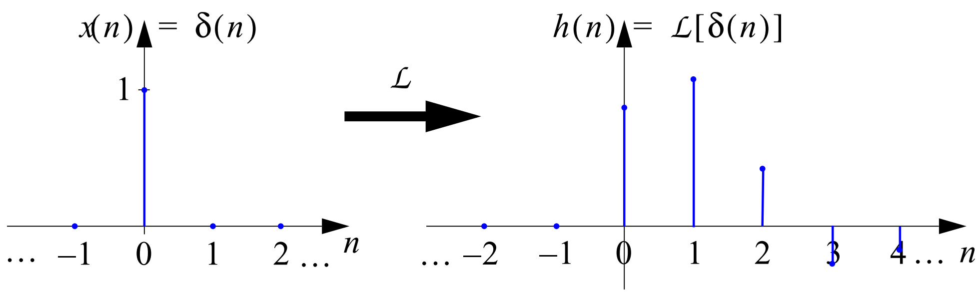
$$\sum_{m=-\infty}^{\infty} |h(m)| < \infty$$

Proof:

$\square$

- ***Causal LTIV system:***

$$h(n) = 0 \qquad \text{for} \qquad n < 0$$



- ***Input-output (I-O) relationship of a LTIV system (time domain):***

$$\boxed{\begin{aligned} y(n) &= \sum_{m=-\infty}^{\infty} h(m)x(n-m) \\ &= h(n) * x(n) \end{aligned}} \qquad (1.1)$$

The symbol * denotes the discrete convolution operation.

Proof:

$\square$

- ***(Discrete) Fourier transform:***
  Here, $z(n)$ denotes an arbitrary sequence.

$$Z(f) = \mathcal{F}\{z(n)\} \equiv \sum_{n=-\infty}^{\infty} z(n)\exp(-j2\pi nf) \qquad (|f| < 1/2)$$

$$z(n) = \mathcal{F}^{-1}\{Z(f)\} = \int_{-1/2}^{1/2} Z(f)\exp(j2\pi nf)df$$

Useful property extensively used in the sequel:

$$z(n-n_0) \;\circ\!\!-\!\!\!-\!\!\!-\!\!\bullet\; \exp(-j2\pi n_0 f)Z(f)$$

Proof:

$\square$

- ***(Frequency) transfer function of a LTIV system:***

$$\boxed{H(f) \equiv \mathcal{F}\{h(n)\} \equiv \sum_{n=-\infty}^{\infty} h(n)\exp(-j2\pi nf)}$$

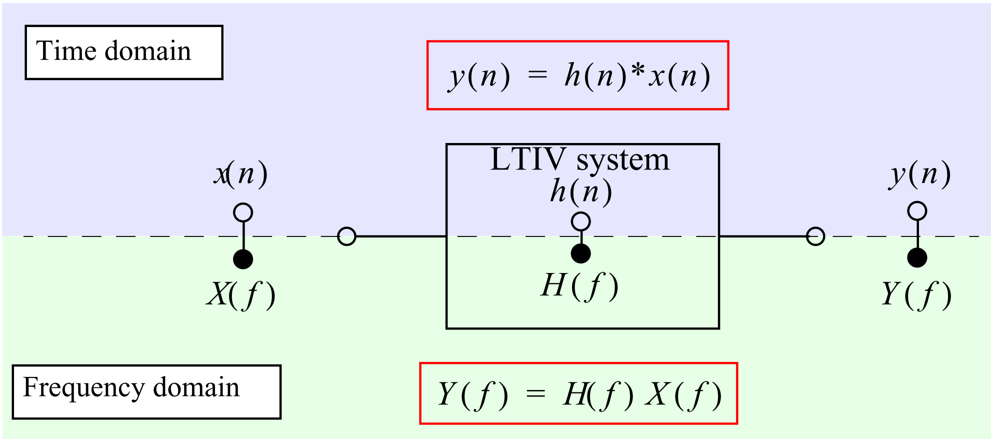
- ***I-O relationship of a LTIV system (frequency domain):***

$$\boxed{Y(f) = H(f)X(f)}$$

Proof:

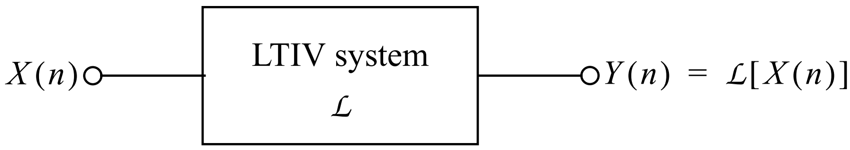
$\square$

- ***Summary: I-O relationship of a LTIV system:***



1.1.3. First- and second-order characterization of a LTIV system

- ***Random input and output sequences:***
  If $X(n)$ is a random sequence (or process), so is $Y(n)$.



- ***Second-order characterization of random sequences:***
  Here, $Z(n)$ denotes an arbitrary random sequence.
  - Expectation:

$$\mu_Z(n) \equiv \mathbf{E}[Z(n)]$$

  - Autocorrelation function:

$$R_{ZZ}(n_1, n_2) \equiv \mathbf{E}[Z(n_1)Z(n_2)]$$

- ***Second-order properties of the output sequence*** $Y(n)$***:***
  - Expectation:

$$\mu_Y(n) = h(n)*\mu_X(n)$$

- Autocorrelation function:

$$R_{YY}(n_1, n_2) = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} h(m_1)h(m_2)R_{XX}(n_1 - m_1, n_2 - m_2)$$

Proof:

□

1.1.4. Wide-sense-stationary (WSS) processes

- ***Definition:***

  A random sequence $Z(n)$ is WSS if the following conditions are satisfied:

  - Expectation:

$$\mu_Z(n) = \mathbf{E}[Z(n)] = \mu_Z$$

  - Autocorrelation function:

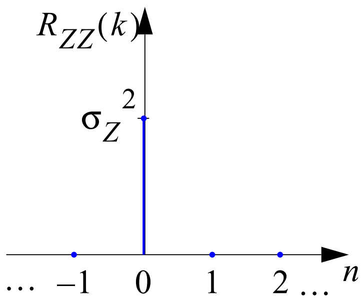
$$R_{ZZ}(n_1, n_1 + k) = \mathbf{E}[Z(n_1)Z(n_1 + k)] = R_{ZZ}(k)$$

- ***White process:***

  $Z(n)$ is a white process if it satisfies the following conditions:

  - $Z(n)$ is a random process
  - $\mu_Z(n) = \mathbf{E}[Z(n)] = 0$

- $R_{ZZ}(n, n+k) = \mathbf{E}[Z(n)Z(n+k)] = R_{ZZ}(k) = \sigma_Z^2\delta(k)$



- ***Autocorrelation function of the impulse response:***

$$\boxed{\begin{aligned} R_{hh}(k) &= \sum_{m=-\infty}^{\infty} h(m)h(m+k) \\ &= h(k)*h(-k) \end{aligned}}$$

Proof:

□

- ***Second-order I-O relationship of a LTIV system (time domain):***

  - Expectation:

$$\boxed{\mu_Y = \left[\sum_{m=-\infty}^{\infty} h(m)\right]\mu_X = H(0)\mu_X}$$

  - Autocorrelation function:

$$\boxed{R_{YY}(k) = R_{hh}(k)*R_{XX}(k)}$$

Proof:

□

- ***Power spectrum of a WSS process:***
  The power spectrum of the WSS process $Z(n)$ with the autocorrelation function $R_{ZZ}(k)$ is defined to be

$$\boxed{S_{ZZ}(f) = \mathcal{F}\{R_{ZZ}(k)\} = \sum_{n=-\infty}^{\infty} R_{ZZ}(k)\exp(-j2\pi kf)}$$

  Notice that from the inverse Fourier transformation

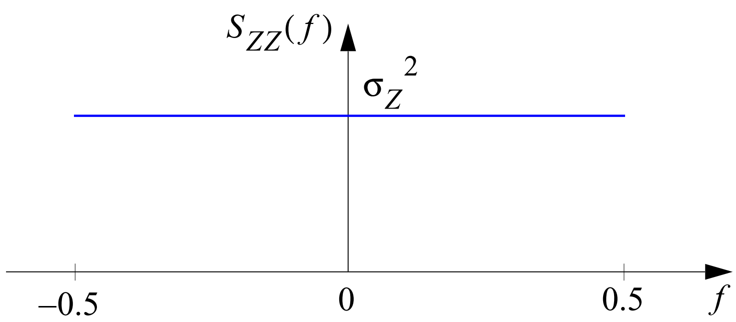
$$R_{ZZ}(k) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{ZZ}(f)\exp(j2\pi kf)df$$

  we conclude that

$$\mathbf{E}[Z(n)^2] = R_{ZZ}(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{ZZ}(f)df$$

- ***Spectrum of a white process:***
  If $Z(n)$ is a white process:

$$\boxed{S_{ZZ}(f) = \sigma_Z^2}$$



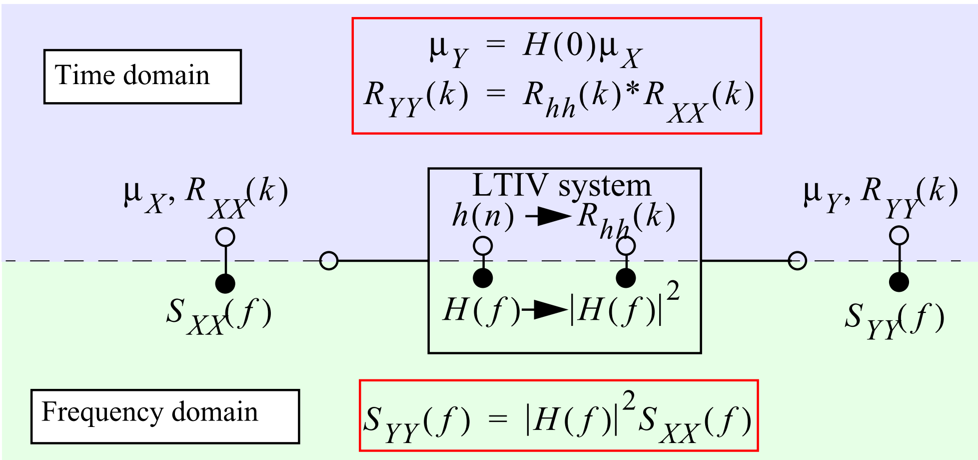
- ***Second-order I-O relationship of a LTIV system (frequency domain):***

$$\boxed{S_{YY}(f) = |H(f)|^2 S_{XX}(f)}$$

Proof:

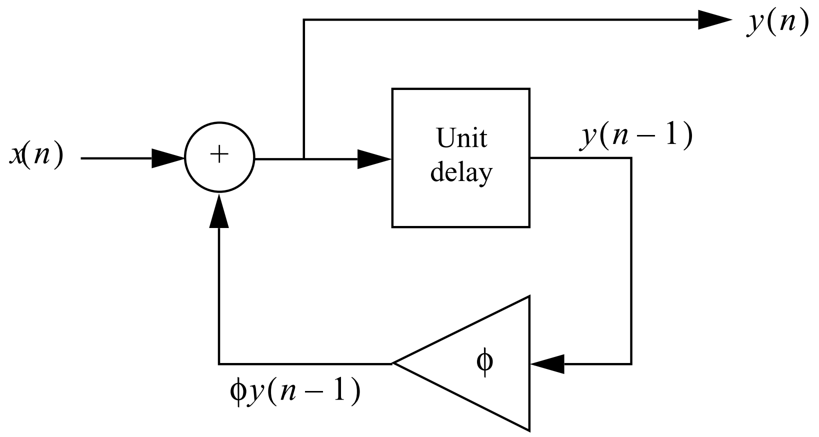
$\square$

- ***Summary: Second-order I-O relationship of a LTIV system:***



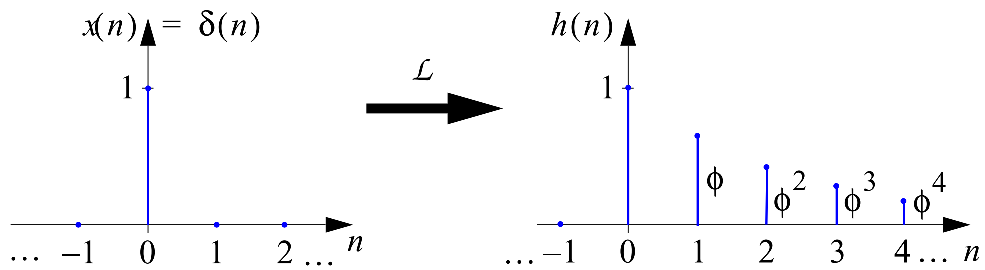
1.1.5. Example: First order recursive filter

- ***Block diagram and recursive equation:***



$$y(n) = x(n) + \phi y(n-1) \qquad (y(n) = 0 \qquad n < 0)$$

- ***Impulse response:***



$$h(n) = \begin{cases} 0 & ; \quad n < 0 \\ \phi^n & ; \quad n \geq 0 \end{cases}$$

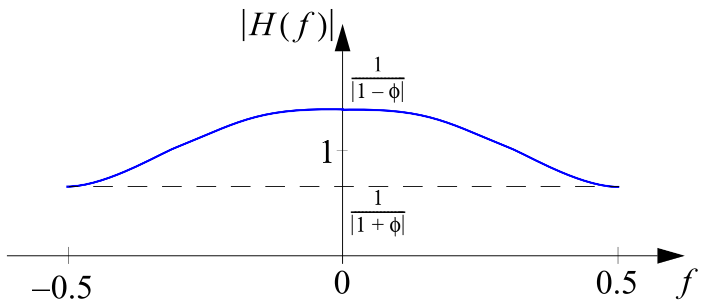
- ***Stability condition:***

$$\sum_{n=-\infty}^{\infty} |h(n)| = \sum_{n=0}^{\infty} |\phi|^n = \lim_{N \to \infty} \frac{1 - |\phi|^N}{1 - |\phi|} < \infty \qquad \Leftrightarrow \qquad |\phi| < 1$$

$$\left[\sum_{n=0}^{N} a^n = \frac{1 - a^{N+1}}{1 - a} \right]$$

- ***Transfer function:***

$$
\begin{aligned}
H(f) &= \mathcal{F}\{h(n)\} = \sum_{n=0}^{\infty} \phi^n \exp(-j2\pi n f) \\
&= \sum_{n=0}^{\infty} [\phi \exp(-j2\pi f)]^n \\
&= \frac{1}{1-\phi\exp(-j2\pi f)}
\end{aligned}
$$

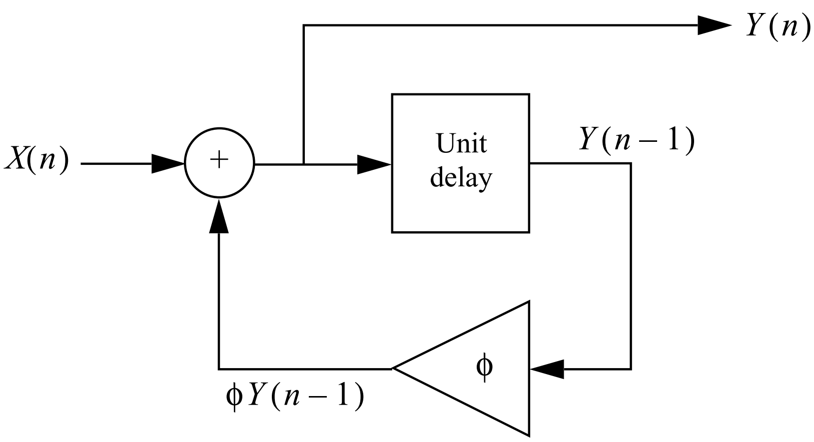


Another more direct way to compute the transfer function:

$$y(n) = x(n) + \phi y(n-1)$$

$$Y(f) = X(f) + \phi\exp(-j2\pi f)Y(f)$$

- ***Random input and output:***



$$Y(n) = X(n) + \phi Y(n-1)$$

- ***Second-order I-O relationship:***
  - Time domain:

$$\mu_Y = H(0)\mu_X = \frac{1}{1-\phi}\mu_X$$

$$R_{YY}(k) = R_{hh}(k) * R_{XX}(k) = \frac{\phi^{|k|}}{1-\phi^2} * R_{XX}(k)$$

$$\left[R_{hh}(k) = \sum_{m=0}^{\infty} \phi^m \phi^{m+|k|} = \phi^{|k|} \sum_{m=0}^{\infty} \phi^{2m} = \phi^{|k|} \frac{1}{1-\phi^2} \right]$$

  - Frequency domain:

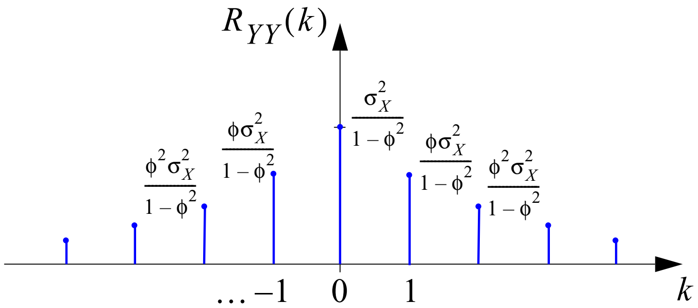
$$S_{YY}(f) = |H(f)|^2 S_{XX}(f) = \frac{1}{|1-\phi\exp(-j2\pi f)|^2} S_{XX}(f)$$

- ***Special case: AR(1) process (see Section 2.2):***

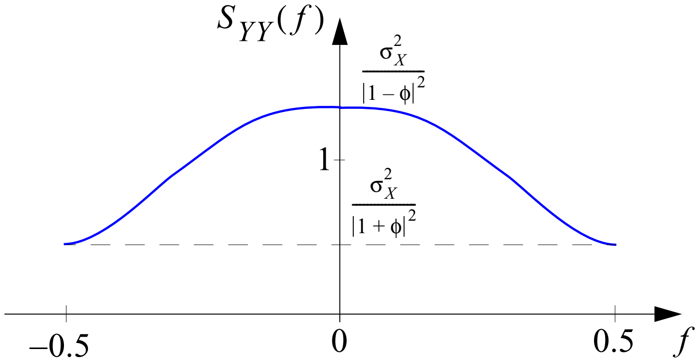
  If $X(n)$ is a white Gaussian process,

$$\mu_Y = 0$$

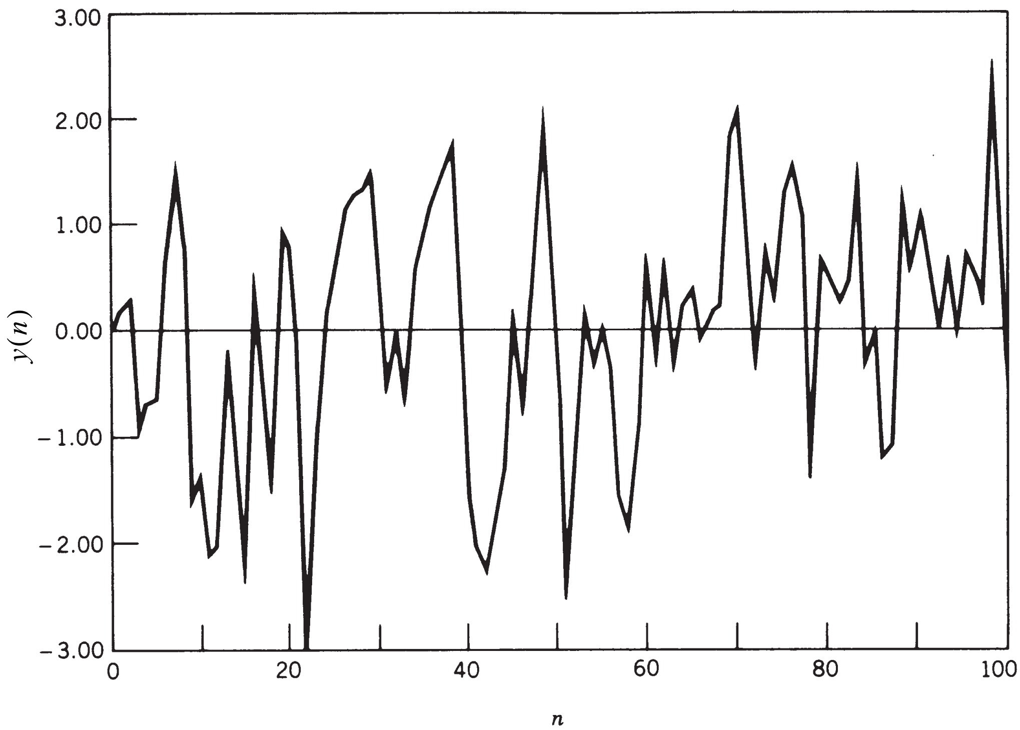
$$R_{YY}(k) = \frac{\phi^{|k|}}{1-\phi^2}\sigma_X^2$$



$$S_{YY}(f) = \frac{\sigma_X^2}{|1-\phi\exp(-j2\pi f)|^2}$$

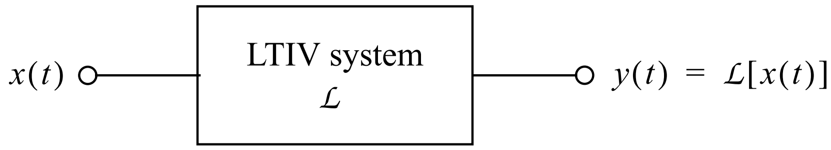


One realization of $Y(n)$:

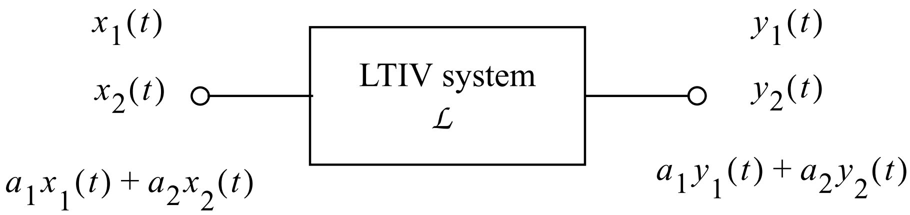


1.2. Continuous-time LTIV systems

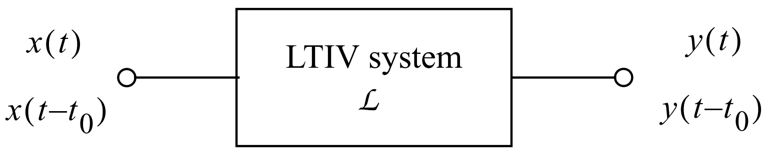
1.2.1. Continuous-time LTIV system



Linear:



Time-invariant:



1.2.2. Steady-state description of a LTIV system

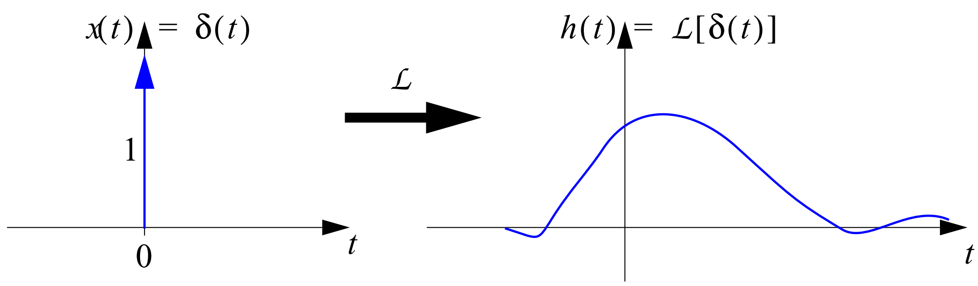
- ***Impulse response:***

  The impulse response $h(t)$ of $\mathcal{L}$ is the response of $\mathcal{L}$ to the Dirac impulse

$$\boxed{\delta(t) \equiv \begin{cases} \infty \ ; & t = 0 \\ 0 \ ; & t \neq 0 \end{cases} \qquad \int_{-\infty}^{\infty} \delta(t)dt = 1,}$$

namely

$$\boxed{h(t) = \mathcal{L}[\delta(t)]}$$



- ***Stable LTIV system:***

  A LTIV system is stable if its response to a bounded signal is bounded.

  It can be shown [to this end we need (1.2)] that a LTIV system is stable if, and only if,

$$\int_{-\infty}^{\infty} |h(t)| dt < \infty$$

- ***Causal LTIV system:***

$$h(t) = 0 \qquad t < 0$$

- ***Input-output relationship of a LTIV system (time domain):***

$$\boxed{\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \\ &= h(t)*x(t) \end{aligned}} \tag{1.2}$$

  Here, the symbol * denotes the continuous convolution operation.

- ***(Continuous) Fourier transform:***

$$Z(f) = \mathcal{F}\{z(t)\} \equiv \int_{-\infty}^{\infty} z(t)\exp(-j2\pi ft)d\tau$$

$$z(t) = \mathcal{F}^{-1}\{Z(f)\} = \int_{-\infty}^{\infty} Z(f)\exp(j2\pi ft)df$$

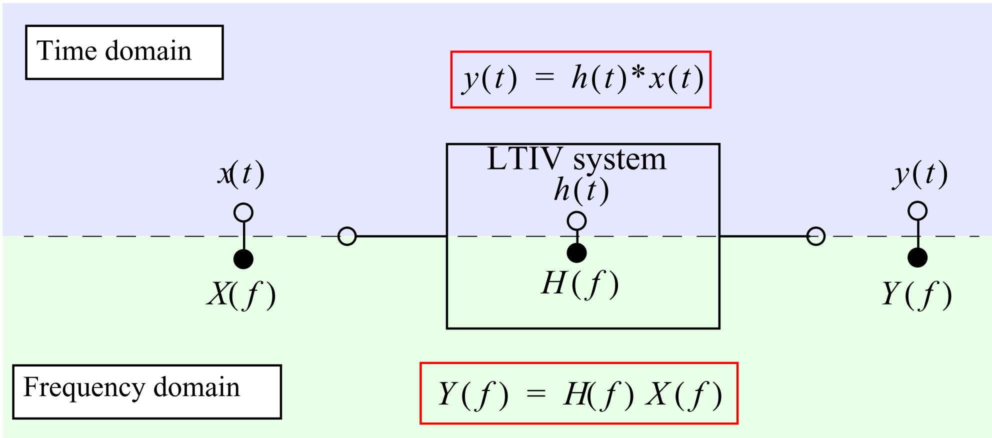
- ***(Frequency) transfer function of a LTIV system***

$$\boxed{H(f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t)\exp(-j2\pi ft)dt}$$

- ***I-O relation ship of a LTIV system (frequency domain):***

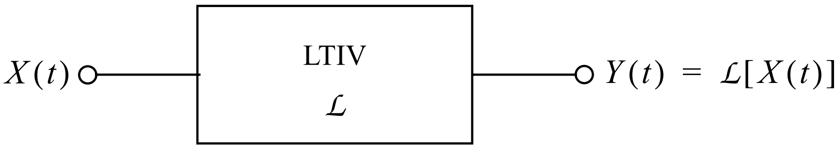
$$\boxed{Y(f) = H(f)X(f)}$$

- ***Summary: I-O relation ship of a LTIV system:***



1.2.3. Second-order characterization of LTIV

- ***Random input and output processes:***



- ***Second-order characterization of random processes:***
  Here, $Z(t)$ denotes an arbitrary random process.
  - Expectation:

$$\mu_Z(t) \equiv \mathbf{E}[Z(t)]$$

  - Autocorrelation function:

$$R_{ZZ}(t_1, t_2) \equiv \mathbf{E}[Z(t_1)Z(t_2)]$$

- ***Second-order properties of the output process*** $Y(t)$***:***
  - Expectation:

$$\mu_Y(t) = h(t)*\mu_X(t)$$

  - Autocorrelation function:

$$R_{YY}(t_1, t_2) \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u_1)h(u_2)R_{XX}(t_1 - u_1, t_2 - u_2)du_1 du_2$$

1.2.4. Wide-sense-stationary (WSS) processes

- ***Wide-sense stationarity:***

  A random process $Z(t)$ is WSS if the following conditions are satisfied:

  - Expectation:

$$\mu_Z(t) \equiv \mu_Z$$

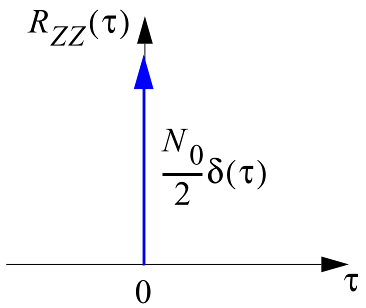
  - Autocorrelation function:

$$R_{ZZ}(t_1, t_1 + \tau) \equiv R_{ZZ}(\tau)$$

- ***White process:***

  $Z(t)$ is a white process if it satisfies the following conditions:

  - $Z(t)$ is a random process
  - $\mu_Z(t) = \mathbf{E}[Z(t)] = 0$
  - $R_{ZZ}(t, t+\tau) = \mathbf{E}[Z(t)Z(t+\tau)] = R_{ZZ}(\tau) = \frac{N_0}{2}\delta(\tau)$



- ***Autocorrelation function of the impulse response:***

$$\begin{aligned} R_{hh}(\tau) &= \int_{-\infty}^{\infty} h(u)h(u+\tau)du \\ &= h(\tau)*h(-\tau) \end{aligned}$$

- ***Second-moment I-O relationship of a LTIV system (time domain):***

  - Expectation:

$$\boxed{\mu_Y = \left[\int_{-\infty}^{\infty} h(u)du\right]\mu_X = H(0)\mu_X}$$

- Autocorrelation function:

$$\boxed{R_{YY}(\tau) = R_{hh}(\tau) * R_{XX}(\tau)}$$

- ***Power spectrum of a WSS process:***
  $Z(t)$ is WSS with the autocorrelation function $R_{ZZ}(\tau)$.

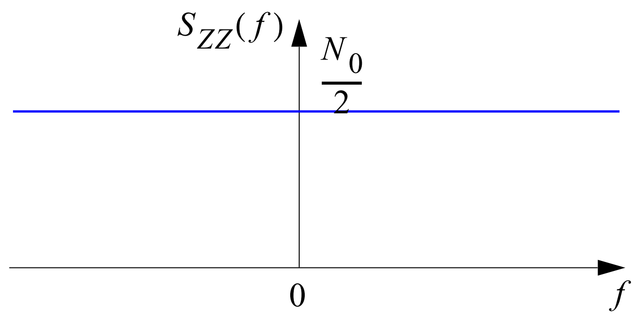
$$\boxed{S_{ZZ}(f) = \mathcal{F}\{R_{ZZ}(\tau)\}}$$

  Notice the identities

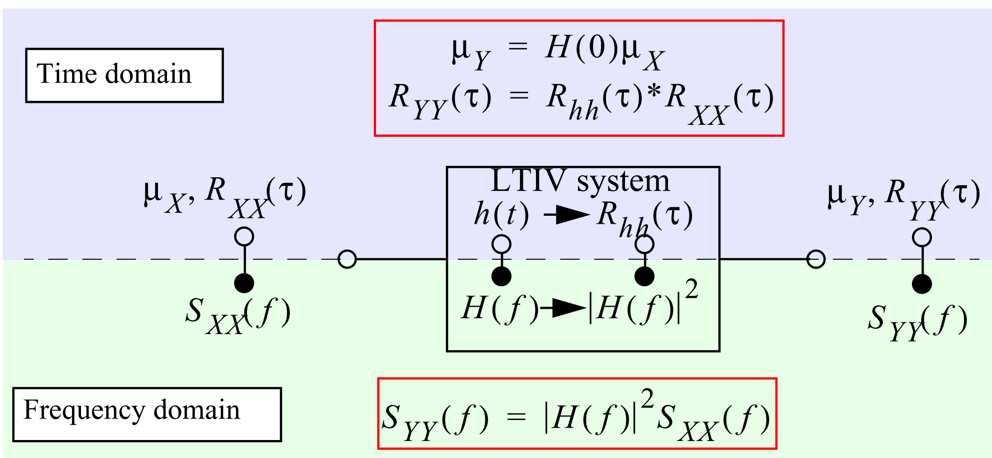
$$\mathbf{E}[Z(t)^2] = R_{ZZ}(0) = \int_{-\infty}^{\infty} S_{ZZ}(f)df$$

- ***Spectrum of a white process:***
  If $Z(t)$ is a white process:

$$\boxed{S_{ZZ}(f) = \frac{N_0}{2}}$$

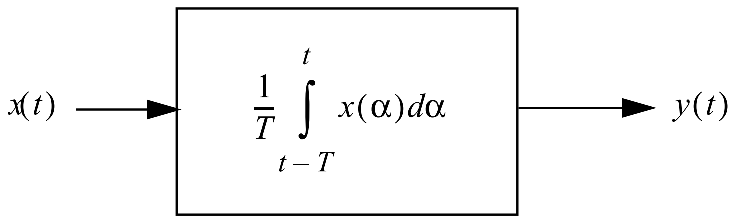


- ***Summary: Second order I-O relationship of a LTIV system:***



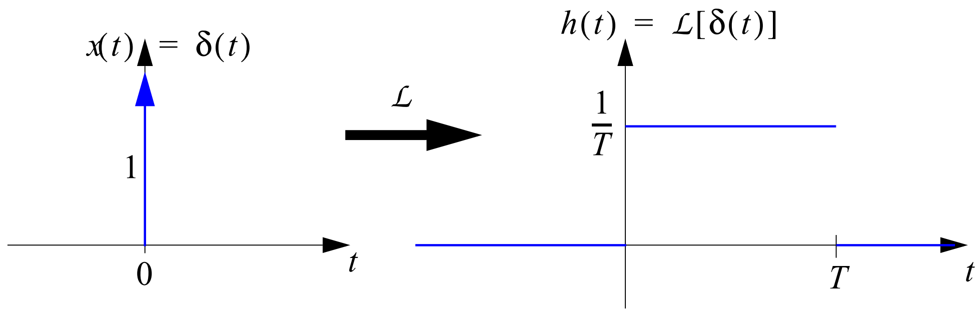
1.2.5. Example: Ideal integrator

- ***Block diagram and input-output relationship:***



$$y(t) = \frac{1}{T}\int_{t-T}^{t} x(\alpha)d\alpha$$

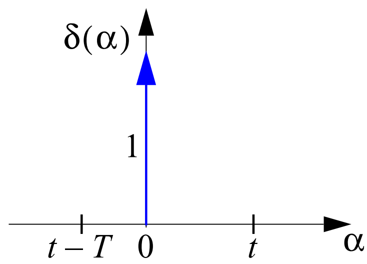
- ***Impulse response:***



$$h(t) = \begin{cases} \frac{1}{T} \ ; & 0 < t < T \\ 0 \ ; & \text{elsewhere} \end{cases}$$

Proof:

$$\frac{1}{T}\int_{t-T}^{t} \delta(\alpha)d\alpha = \begin{cases} 1 \ ; & 0 \in (t-T, t) \\ 0 \ ; & \text{elsewhere} \end{cases}$$



□

- ***Stability condition:***

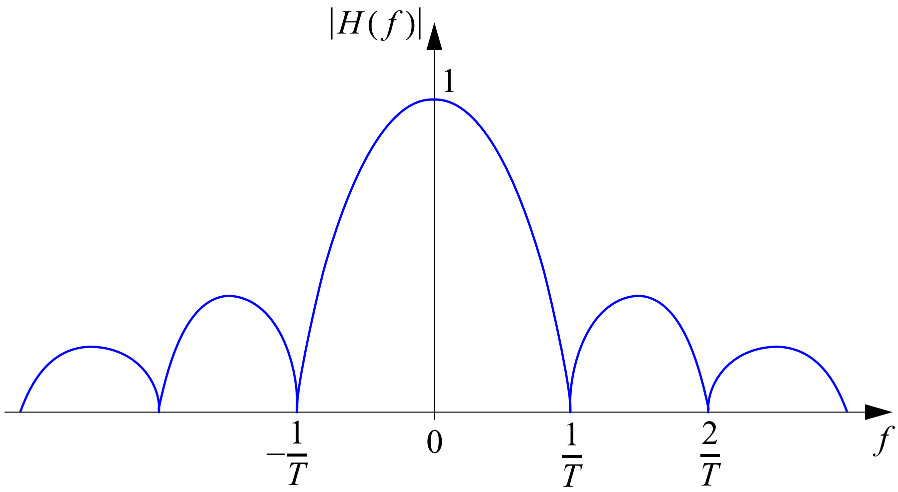
$$\int_{-\infty}^{\infty} |h(t)|dt = 1$$

- ***Transfer function:***

$$H(f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t)\exp(-j2\pi ft)dt$$

$$= \frac{1}{T}\int_{0}^{T} \exp(-j2\pi ft)dt = \exp(-j\pi fT)\frac{\sin(\pi fT)}{\pi fT}$$

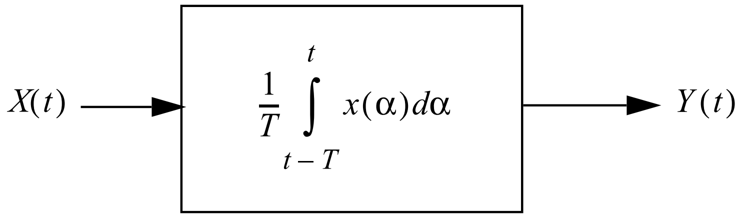
$$= \exp(-j\pi fT)\,\mathrm{sinc}(fT)$$



Proof:

□

- ***Random input and output:***



$$Y(t) = \frac{1}{T}\int_{t-T}^{t} X(\alpha)d\alpha$$

- ***Second-order I-O relationship:***

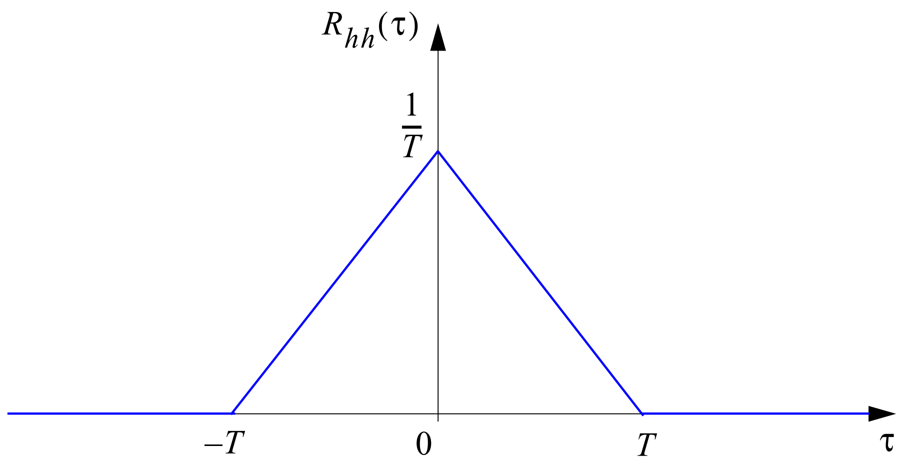
  - Time domain:

$$\mu_Y = H(0)\mu_X$$
$$= \mu_X$$

$$R_{YY}(\tau) = R_{hh}(\tau) * R_{XX}(\tau)$$

with

$$R_{hh}(\tau) = \begin{cases} \frac{1}{T}\left(1 - \frac{|\tau|}{T}\right) & ; \quad -T < \tau < T \\ 0 & ; \quad \text{elsewhere} \end{cases}$$



Proof:

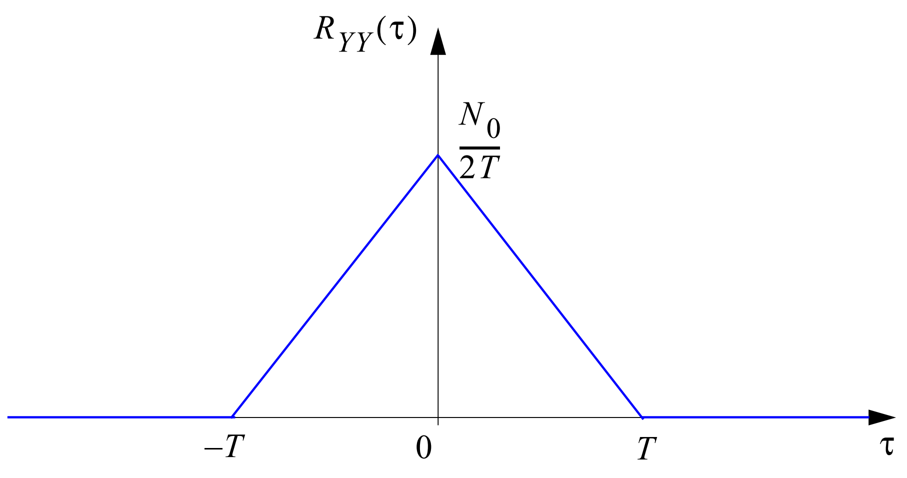
$\square$

- Frequency domain:

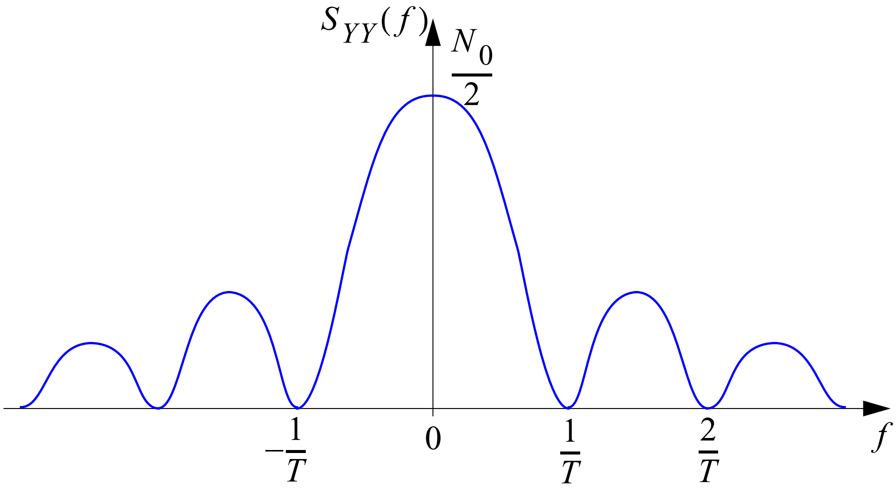
$$S_{YY}(f) = |H(f)|^2 S_{XX}(f)$$
$$= \operatorname{sinc}(fT)^2 S_{XX}(f)$$

- ***Special case:*** $X(t)$ ***is a white process,***

$$\mu_Y = 0 \qquad\qquad R_{YY}(\tau) = \begin{cases} \dfrac{N_0}{2T}\left(1 - \dfrac{|\tau|}{T}\right) & ; \quad -T < \tau < T \\ 0 & ; \quad \text{elsewhere} \end{cases}$$



$$S_{YY}(f) = \frac{N_0}{2}\mathrm{sinc}(fT)^2$$



2. Discrete Linear Stochastic Processes/ Models

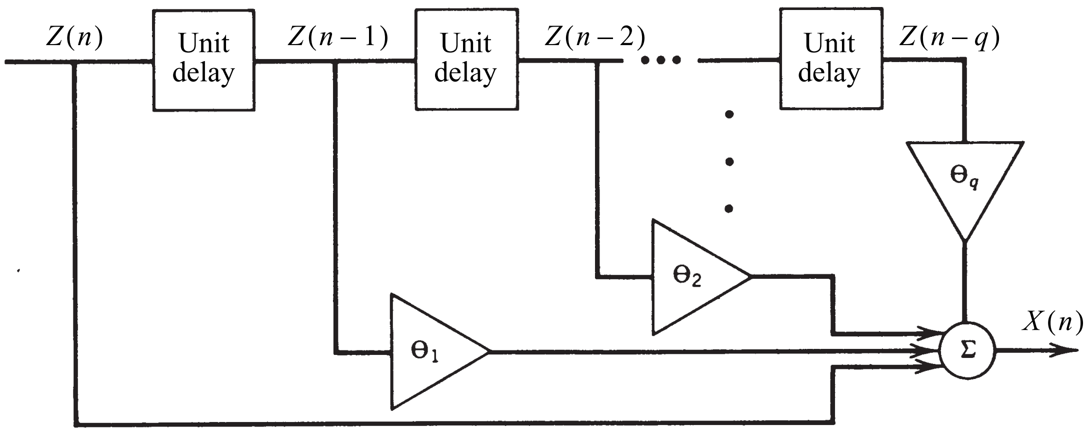
2.1. Moving average processes

- ***Definition:***
  A random sequence $X(n)$ is a moving average process of order q (MA(q)) if for any n:

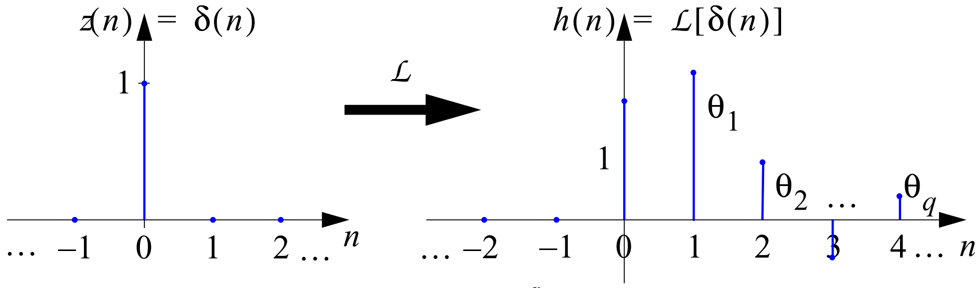
$$\boxed{X(n) = Z(n) + \sum_{i=1}^{q} \theta_i Z(n-i)}$$

  where $Z(n)$ is a white Gaussian process.

- ***Transversal filter implementation of a MA(q) process:***



- ***Impulse response of the transversal filter:***



$$h(n) = \delta(n) + \sum_{i=1}^{q} \theta_i \delta(n-i)$$

- ***Stability and causality:***
  Transversal filters are stable and causal.

- ***Transfer function of the transversal filter:***

$$\boxed{H(f) = 1 + \sum_{i=1}^{q} \theta_i \exp(-j2\pi i f)}$$

Proof:

$$x(n) = \sum_{i=1}^{q} \theta_i z(n-i) + z(n)$$

$$X(f) = \sum_{i=1}^{q} \theta_i \exp(-j2\pi i f) Z(f) + Z(f)$$

$$= \left[1 + \sum_{i=1}^{q} \theta_i \exp(-j2\pi i f)\right] Z(f)$$

$\square$

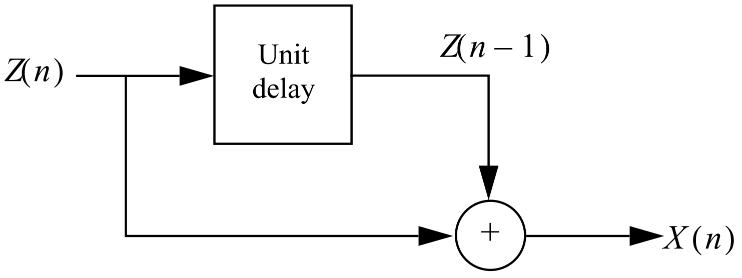
- ***Power spectrum of a MA(q) process:***

$$\boxed{S_{XX}(f) = \left|1 + \sum_{i=1}^{q} \theta_i \exp(-j2\pi i f)\right|^2 \sigma_Z^2}$$

- ***Mean value and autocorrelation function of a MA(q) process:***

$$\boxed{\begin{aligned} \mu_X &= 0 \\ R_{XX}(k) &= \sigma_Z^2 R_{hh}(k) \end{aligned}}$$

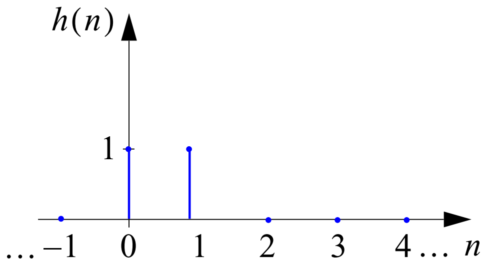
- ***Example: MA(1)***



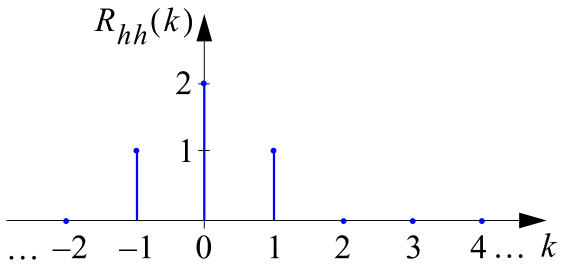
$$X(n) = Z(n) + Z(n-1) \qquad (\theta_1 = 1)$$

- Impulse response and autocorrelation function of the transversal filter

$$h(n) = \begin{cases} 1\ ; & n \in \{0, 1\} \\ 0\ ; & \text{elsewhere} \end{cases}$$



$$R_{hh}(k) = \begin{cases} 2\ ; & k \in \{0\} \\ 1\ ; & k \in \{-1, 1\} \\ 0\ ; & \text{elsewhere} \end{cases}$$

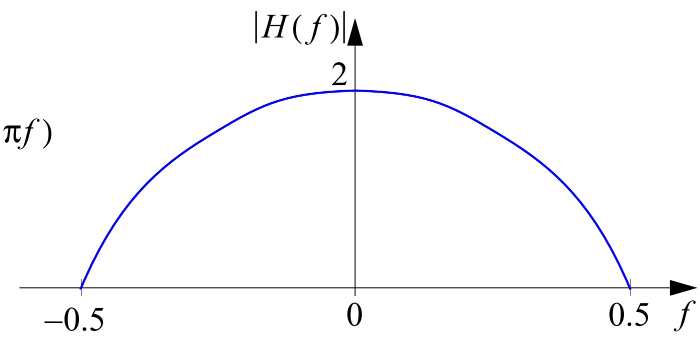


- Transfer function:

$$\begin{aligned} H(f) &= 1 + \exp(-j2\pi f) \qquad |f| \le 0.5 \\ &= \exp(-j\pi f)[\exp(j\pi f) + \exp(-j\pi f)] \\ &= 2\exp(-j\pi f)\cos(\pi f) \end{aligned}$$

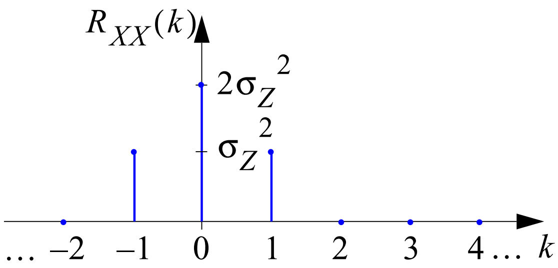
$$|H(f)| = 2\cos(\pi f)$$

$$(|f| \le 0.5)$$



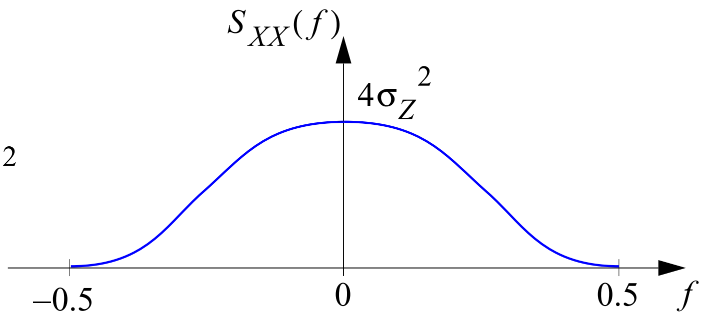
- Autocorrelation function of $X(n)$:

$$R_{XX}(k) = \sigma_Z^2 R_{hh}(k)$$
$$= \begin{cases} 2\sigma_Z^2 \; ; & k \in \{0\} \\ \sigma_Z^2 \; ; & k \in \{-1, 1\} \\ 0 \; ; & \text{elsewhere} \end{cases}$$



- Power spectrum of $X(n)$:

$$S_{XX}(f) = \sigma_Z^2 |H(f)|^2$$
$$= 4\sigma_Z^2 \cos(\pi f)^2$$



2.2. Autoregressive processes

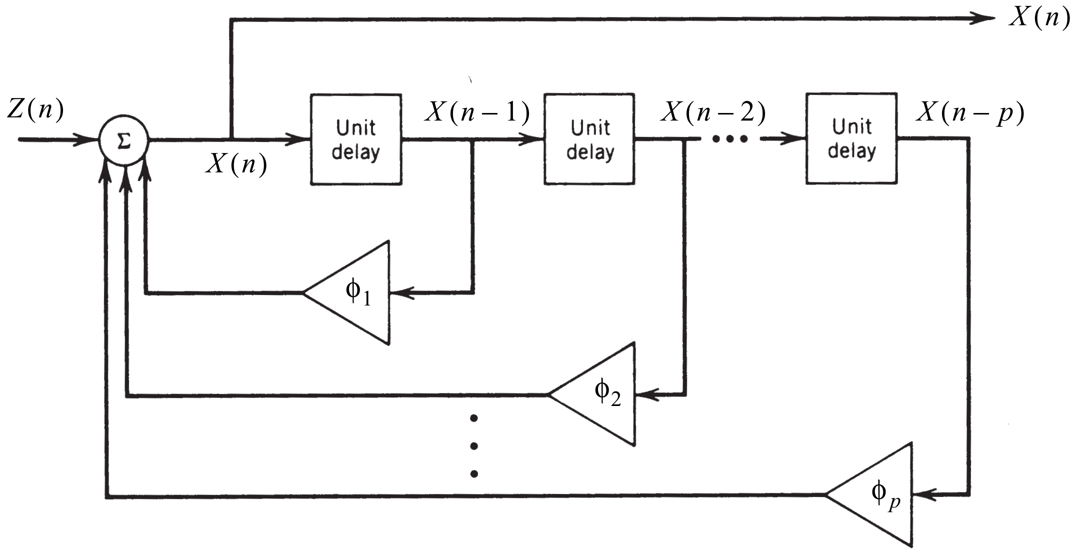
- ***Definition:***

A random sequence $X(n)$ is an autoregressive process of order p (AR(p)) if it is WSS and for any n:

$$\boxed{X(n) = \sum_{i=1}^{p} \phi_i X(n-i) + Z(n)}$$

where $Z(n)$ is a white Gaussian process.

- ***Recursive filter implementation:***



- ***Causal and stable AR processes:***

An AR(p) process $X(n)$ is called causal and stable if there exists an infinite causal and stable transversal filter with impulse response $h(n)$ such that

$$X(n) = \sum_{i=0}^{\infty} h(i)Z(n-i)$$

$$= h(n)*Z(n)$$

Let us define the polynomial

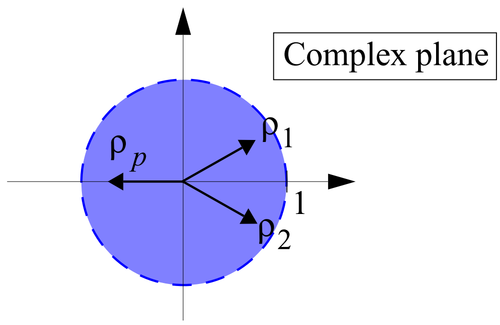
$$\phi(z) \equiv 1 - \sum_{i=1}^{p} \phi_i z^{-i} \qquad z : \text{complex variable.}$$

Then, the AR process $X(n)$ is causal and stable, if, and only if, the roots of $\phi(z)$ are located inside the unit circle, i.e. if $\phi(z)$ factorizes according to

$$\phi(z) = \prod_{i=1}^{p} (1 - \rho_i z^{-i})$$

with $|\rho_i| < 1$, $i = 1, \ldots, p$.

Location of the roots of $\phi(z)$ in the complex plane when $X(n)$ is causal and stable:



The impulse response of a causal and stable AR(p) process is determined by the identity

$$\boxed{\sum_{i=0}^{\infty} h(i) z^{-i} = \frac{1}{\phi(z)} \qquad |z| \geq 1}$$

- ***Transfer function of the recursive filter:***

$$\boxed{H(f) = \frac{1}{1 - \sum_{i=1}^{p} \phi_i \exp(-j2\pi i f)}}$$

Proof:

$$x(n) = \sum_{i=1}^{p} \phi_i x(n-i) + z(n)$$

$$\circ\!\!-\!\!\bullet$$

$$\begin{aligned} X(f) &= \sum_{i=1}^{p} \phi_i \exp(-j2\pi i f) X(f) + Z(f) \\ &= \left[\sum_{i=1}^{p} \phi_i \exp(-j2\pi i f)\right] X(f) + Z(f) \end{aligned}$$

□

- ***Power spectrum of an AR(p) process:***

$$\boxed{S_{XX}(f) = \frac{\sigma_Z^2}{\left|1 - \sum_{i=1}^{p} \phi_i \exp(-j2\pi i f)\right|^2}}$$

- ***Mean value and autocorrelation function of a causal AR(p) process:***
  If the AR process $X(n)$ is causal,

$$\boxed{\begin{aligned} \mu_X &= 0 \\ R_{XX}(k) &= \sigma_Z^2 R_{hh}(k) \end{aligned}}$$

- ***Example: AR(1):***
  The first-order recursive filter discussed in the previous chapter with a white Gaussian process as the input signal generates an AR(1) process.
- ***Yule-Walker equations:***
  Let be $k \geq 0$:

$$X(n) = \sum_{i=1}^{p} \phi_i X(n-i) + Z(n)$$

$$X(n)\, X(n-k) = \sum_{i=1}^{p} \phi_i X(n-i) X(n-k) + Z(n) X(n-k)$$

$$\mathbf{E}[X(n)\, X(n-k)] = \sum_{i=1}^{p} \phi_i \mathbf{E}[X(n-i) X(n-k)] + \mathbf{E}[Z(n) X(n-k)]$$

$$R_{XX}(n, n-k) = \sum_{i=1}^{p} \phi_i R_{XX}(n-i, n-k) + \sigma_Z^2 \delta(k)$$

$$R_{XX}(k) = R_{XX}(-k) = \sum_{i=1}^{p} \phi_i R_{XX}(i-k) + \sigma_Z^2 \delta(k)$$

Using a vector notation, for $0 \leq k \leq p$

$$R_{XX}(k) = [R_{XX}(1-k), \ldots, R_{XX}(p-k)] \begin{bmatrix} \phi_1 \\ \ldots \\ \phi_p \end{bmatrix} + \sigma_Z^2 \delta(k) \qquad (2.1)$$

For $k > p$:

$$R_{XX}(k) = [R_{XX}(k-1), \ldots, R_{XX}(k-p)] \begin{bmatrix} \phi_1 \\ \ldots \\ \phi_p \end{bmatrix} \qquad (2.2)$$

Let us define

$$\Phi \equiv \begin{bmatrix} \phi_1 \\ \ldots \\ \phi_p \end{bmatrix} \qquad \gamma \equiv \begin{bmatrix} R_{XX}(1) \\ R_{XX}(2) \\ \ldots \\ R_{XX}(p) \end{bmatrix}$$

$$\Gamma \equiv \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & \ldots & R_{XX}(p-1) \\ R_{XX}(-1) & R_{XX}(0) & \ldots & R_{XX}(p-2) \\ \ldots & \ldots & \ldots & \ldots \\ R_{XX}(-(p-1)) & R_{XX}(-(p-2)) & \ldots & R_{XX}(0) \end{bmatrix}$$

Note that Γ is symmetric.
Then, for $k = 0$ Identity (2.1) becomes

$$\boxed{R_{XX}(0) = \gamma^T \Phi + \sigma_Z^2}$$

Inserting $k = 1, \ldots, p$ in (2.1) yields p identities that can be concatenated in a matrix form according to

$$\begin{bmatrix} R_{XX}(1) \\ R_{XX}(2) \\ \ldots \\ R_{XX}(p) \end{bmatrix} = \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & \ldots & R_{XX}(p-1) \\ R_{XX}(-1) & R_{XX}(0) & \ldots & R_{XX}(p-2) \\ \ldots & \ldots & \ldots & \ldots \\ R_{XX}(-(p-1)) & R_{XX}(-(p-2)) & \ldots & R_{XX}(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \ldots \\ \phi_p \end{bmatrix}$$

$$\boxed{\gamma = \Gamma \Phi}$$

Comments:

- The feed-back coefficients $\phi_1, \ldots, \phi_p$ of the recursiv filter and the variance σ_Z^2 of the white Gaussian input process $Z(n)$ can be computed from $R_{XX}(0), \ldots, R_{XX}(p)$ via the Yule-Walker equations and vice-versa.
- The samples $R_{XX}(k)$, $k > p$ can be recursively computed from $\phi_1, \ldots, \phi_p$ and $R_{XX}(k-1), \ldots, R_{XX}(k-p)$ by using Identity (2.2).

2.3. Autoregressive moving average processes

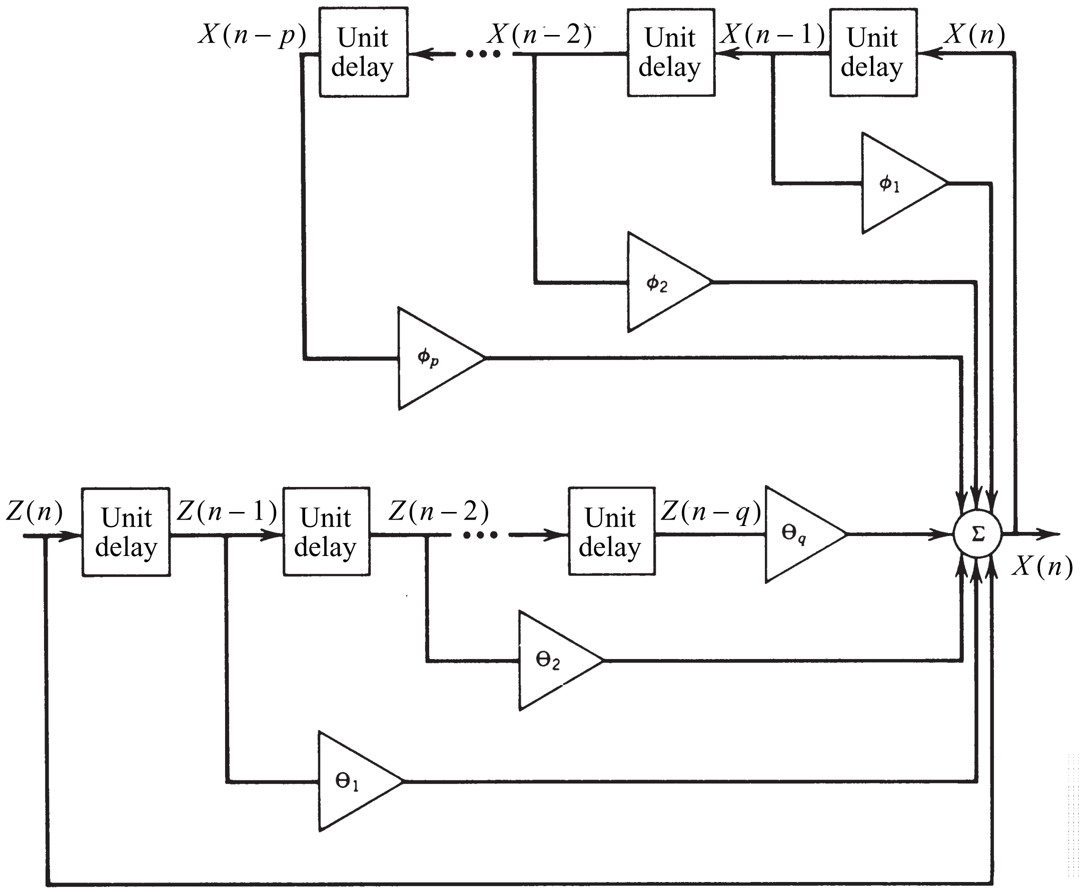
- ***Definition:***

  A random sequence $X(n)$ is an autoregressive moving average process (p, q) th order (ARMA((p, q))) if it is WSS and for any n:

$$\boxed{X(n) = \sum_{i=1}^{p} \phi_i X(n-i) + \sum_{i=1}^{q} \theta_i Z(n-i) + Z(n)}$$

  where $Z(n)$ is a white Gaussian process.

- ***Filter implementation:***



- ***Causal and stable ARMA processes:***

  An ARMA(p, q) process $X(n)$ is called causal and stable if there exists an infinite causal and stable transversal filter with impulse response $h(n)$ such that

$$X(n) = \sum_{i=0}^{\infty} h(i)Z(n-i) = h(n)*Z(n)$$

  Let us define

$$\theta(z) \equiv 1 + \sum_{i=1}^{q} \theta_i z^{-i} \qquad \text{and} \qquad \phi(z) \equiv 1 - \sum_{i=1}^{p} \phi_i z^{-i}$$

  A necessary and sufficient condition for an ARMA(p, q) process to be causal and stable is that the polynomial $\phi(z)$ has its roots inside the unit circle.

  The impulse response of a causal and stable ARMA(p, q) process is then determined by the identity

$$\boxed{\sum_{i=0}^{\infty} h_i z^{-i} = \frac{\theta(z)}{\phi(z)} \qquad |z| \geq 1}$$

  In the above considerations we assume that $\theta(z)$ and $\phi(z)$ have no common root.

- ***Transfer function of the filter:***

$$\boxed{H(f) = \frac{1 + \sum_{i=1}^{q} \theta_i \exp(-j2\pi i f)}{1 - \sum_{i=1}^{p} \phi_i \exp(-j2\pi i f)}}$$

  Proof: Similar as before.

- ***Power spectrum of an ARMA(p,q) process:***

$$\boxed{S_{XX}(f) = \frac{\left|1 + \sum_{i=1}^{q} \theta_i \exp(-j2\pi i f)\right|^2}{\left|1 - \sum_{i=1}^{p} \phi_i \exp(-j2\pi i f)\right|^2} \sigma_Z^2}$$

- ***Mean value and autocorrelation function of a causal ARMA(p,q) process:***

  If the ARMA process $X(n)$ is causal,

$$\boxed{\begin{aligned} \mu_X &= 0 \\ R_{XX}(k) &= \sigma_Z^2 R_{hh}(k) \end{aligned}}$$

- ***Importance of ARMA(p,q) processes:***

  - Because of the linearity property of ARMA(p,q) processes, analytical expressions can be derived which describe their statistical behavior, i.e. their autocorrelation and power spectrum.
  - For any given zero-mean WSS process $Y(n)$ with autocorrelation function $R_{YY}(k)$ there exists an ARMA(p,q) process $X(n)$ such that

$$R_{YY}(k) = R_{XX}(k) \qquad |k| \leq K.$$

  In this sense, any WSS process can be approximated by an ARMA(p,q) process.

3. Signal Detection

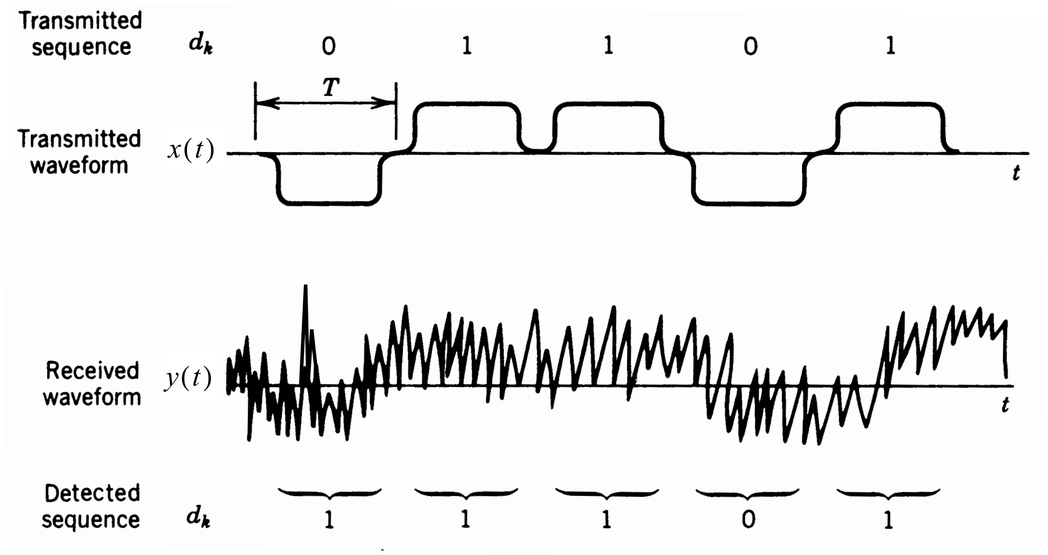
3.1. Hypothesis testing

- ***Method:***
  In hypothesis testing, a decision is made based on the observation of a random variable as to which of several hypotheses to accept.
  In binary hypothesis testing the choice is made among two hypotheses.

  Example 1: *Detection of a BPAM signal:*

  H_0: "0" is transmitted Null hypothesis

  H_1: "1" is transmitted Alternate hypothesis



  Example 2: *Target detection in radar technique:*

  H_0: The target is not present

  H_1: The target is present

- ***Mathematical framework for binary hypothesis testing:***

  - **Two hypotheses:** H_0 and H_1

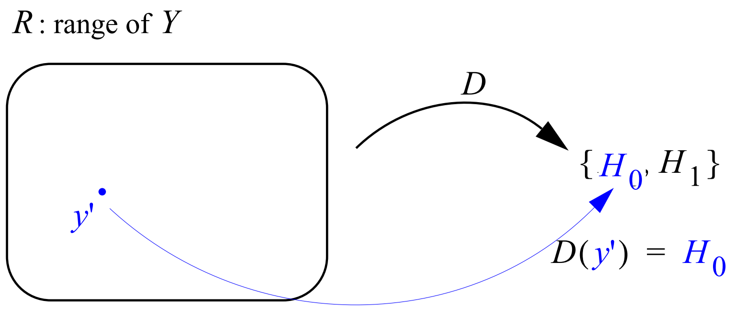
  - **One observation** y **of a random variable** Y **whose probability density function (pdf) under each hypothesis is known.**
    We denote these pdfs as:

$$f(y|H_0) \text{ and } f(y|H_1)$$

- **A decision rule** D, i.e. a mapping

$$D : R \to \{H_0, H_1\},\ y \to D(y),$$

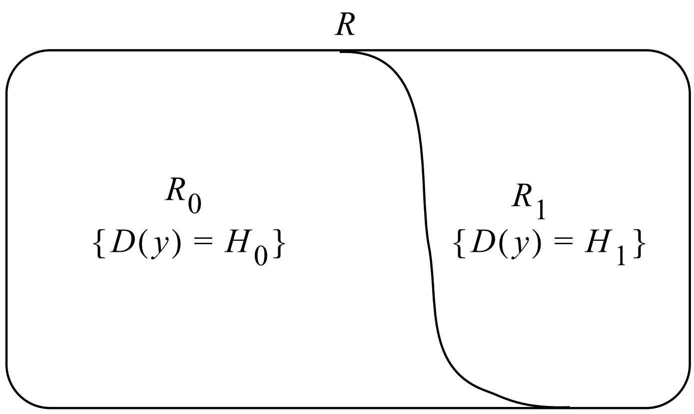
where R denotes the range of Y.



In the above figure, when $Y = y'$ is observed, the decision $D(y') = H_0$ is made.

The decision rule D determines two **decision regions** in R:

$$R_0 : y \in R_0 \qquad \Rightarrow \qquad D(y) = H_0$$
$$R_1 : y \in R_1 \qquad \Rightarrow \qquad D(y) = H_1$$



Properties of R_0 and R_1:

- $R_0 \cup R_1 = R$
- $R_0 \cap R_1 = \varnothing$

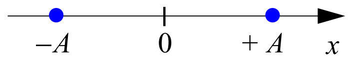
Example: Binary pulse amplitude modulation (BPAM)

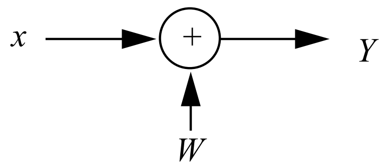
- Signal model:

$$Y = x + W$$

where

$$x = \begin{cases} -A & \text{under} \quad H_0 \\ +A & \text{under} \quad H_1 \end{cases}$$





W is a Gaussian noise, i.e.:

- W is a Gaussian random variable,
- with expectation $\mathbf{E}[W] = 0$,
- and variance $\mathbf{E}[W^2] = \sigma^2$.

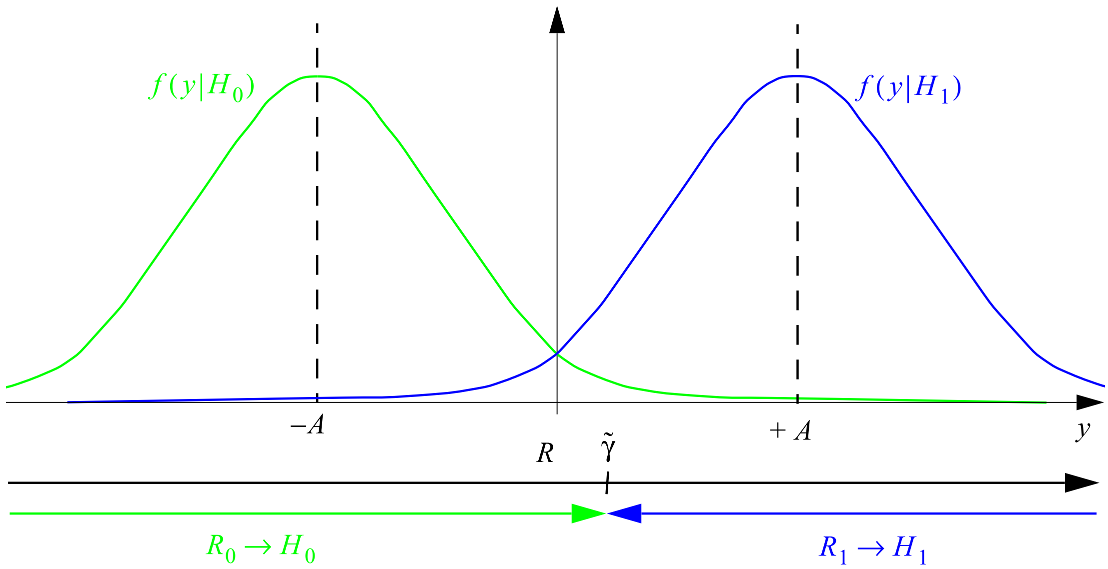
- Probability density function (pdf) of W:

$$f(w) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2} w^2\right\}$$

- Pdf of Y under H_0 and H_1 and decision regions:

$$f(y|H_0) = f(w)|_{w = y+A} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y+A)^2\right\}$$

$$f(y|H_1) = f(w)|_{w = y-A} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y-A)^2\right\}$$



- ***Decision table:***

<table>
<tr><td rowspan="2">Decision
D</td><td colspan="2">True hypothesis H</td></tr>
<tr><td>H_0</td><td>H_1</td></tr>
<tr><td>H_0</td><td>(H_0, H_0)</td><td>(H_0, H_1)</td></tr>
<tr><td>H_1</td><td>(H_1, H_0)</td><td>(H_1, H_1)</td></tr>
</table>

- (green) : Correct decision
- (red) : Incorrect decision:

The pair (H_i, H_j) $(i, j = 0, 1)$ in the above table means $D = H_i$ and $H = H_j$.

- ***Probabilities of correct decision and of making an error:***

- Probability of correct decision:

$$\boxed{\begin{aligned} P_c &= P[D = H_0, H = H_0] + P[D = H_1, H = H_1] \\ &= P[D = H_0 | H_0]P[H_0] + P[D = H_1 | H_1]P[H_1] \end{aligned}}$$

- Probability of incorrect decision:

$$\boxed{\begin{aligned} P_e &= P[D = H_1, H = H_0] + P[D = H_0, H = H_1] \\ &= P[D = H_1 | H_0]P[H_0] + P[D = H_0 | H_1]P[H_1] \end{aligned}}$$

Obviously,

$$P_c = 1 - P_e.$$

- ***Types of error and their probability:***

- **False alarm** (Type I error): $D = H_1$ when H_0 is true.

False alarm probability:

$$\boxed{P_f \equiv P[D = H_1 | H_0]}.$$

- **Miss** (Type II error): $D = H_0$ when H_1 is true.

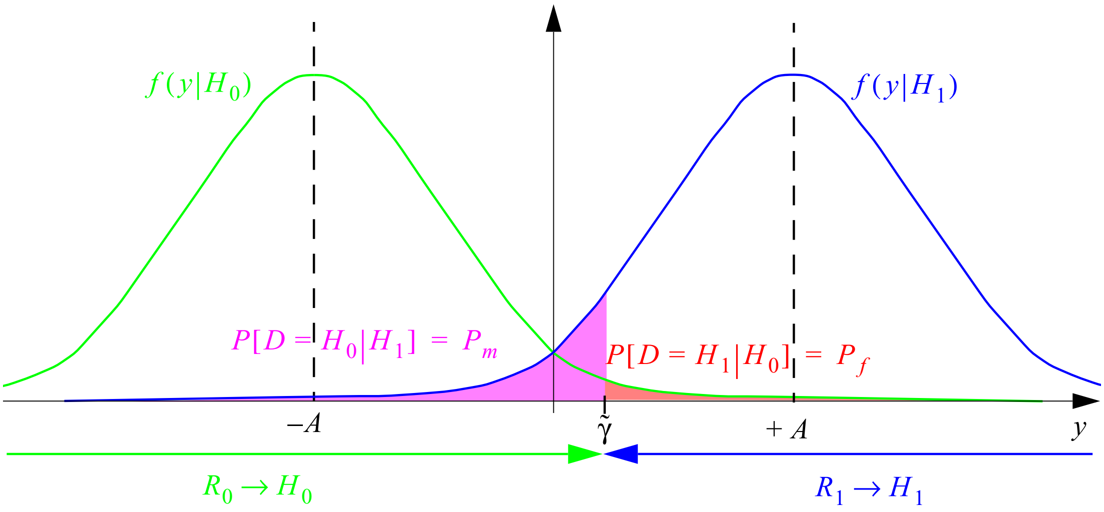
Probability of a miss:

$$\boxed{P_m \equiv P[D = H_0 | H_1]}.$$

- Probability of incorrect decision:

$$P_e = P_f P[H_0] + P_m P[H_1]$$

***Example**:* BPAM (cont'd):



3.2. Decision rules

3.2.1. Maximum "a posteriori" (MAP) decision rule

We seek a decision rule which minimizes the probability of error P_e.

- ***MAP decision rule:***
  Such a rule exists. It is of the form:

$$\boxed{P[H_1|y] \underset{H_0}{\overset{H_1}{\gtrless}} P[H_0|y]}$$

where $P[H_i|y]$, $i = 0, 1$, is the "a posteriori" probability of H_i when $Y = y$ is observed.

- ***Bayes rule:***
  The "a posteriori" probability $P[H_i|y]$ can be obtained by invoking Bayes' rule:

$$P[H_i|y] = \frac{f(y|H_i)P[H_i]}{f(y)}.$$

- ***MAP decision rule (cont'd):***
  Using the last identity, the MAP decision rule can be recast as:

$$f(y|H_1)P[H_1] \underset{H_0}{\overset{H_1}{\gtrless}} f(y|H_0)P[H_0]$$

or equivalently

$$\boxed{L(y) \equiv \frac{f(y|H_1)}{f(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P[H_0]}{P[H_1]}}$$

The function $L(y)$ is called the likelihood ratio.
It is more common to use the log-likelihood ratio

$$l(y) \equiv \ln(L(y))$$

instead:

$$\boxed{l(y) = \ln\left(\frac{f(y|H_1)}{f(y|H_0)}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\frac{P[H_0]}{P[H_1]}\right)}$$

- ***Derivation of the MAP decision rule:***

$\square$

Example: BPAM (cont'd)

- Likelihood ratio:

$$L(y) = \frac{f(y|H_1)}{f(y|H_0)}$$

$$= \frac{\exp\left\{-\frac{1}{2\sigma^2}(y-A)^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2}(y+A)^2\right\}} = \exp\left\{\frac{1}{2\sigma^2}[(y+A)^2-(y-A)^2]\right\}$$

- Loglikelihood ratio:

$$l(y) = \ln\left(\frac{f(y|H_1)}{f(y|H_0)}\right)$$

$$= \frac{1}{2\sigma^2}[(y+A)^2-(y-A)^2]$$

$$= \frac{2Ay}{\sigma^2}$$

- MAP decision rule:

$$\frac{2Ay}{\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\frac{P[H_0]}{P[H_1]}\right)$$

$$y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma^2}{2A}\ln\left(\frac{P[H_0]}{P[H_1]}\right) \equiv \tilde{\gamma}_{MAP}$$

- ***Maximum likelihood (ML) decision rule:***
  Selecting a uniform "a priori" pdf for the hypotheses, i.e.

$$P[H_0] = P[H_1] = \frac{1}{2},$$

the MAP decision rule reduces to the ML decision rule:

$$f(y|H_1) \quad \underset{H_0}{\overset{H_1}{\gtrless}} \quad f(y|H_0)$$

The ML decision rule selects the hypothesis which maximizes the likelihood function $H \to f(y|H)$.

3.2.2. Bayes decision rule

- ***Cost function:***

  In many engineering branches costs have to be taken into account depending on the decision and the true hypothesis.

Decision D	True hypothesis	
	H_0	H_1
H_0	C_{00}	C_{01}
H_1	C_{10}	C_{11}

  Usually, the cost of making a wrong decision is higher than that of making a correct decision:

$$C_{10} \geq C_{00} \qquad \text{and} \qquad C_{01} \geq C_{11}.$$

- ***Average cost:***

$$\boxed{\begin{aligned}\bar{C} &= C_{00}P[D = H_0|H_0]P[H_0] + C_{10}P[D = H_1|H_0]P[H_0] + \\ &\quad + C_{01}P[D = H_0|H_1]P[H_1] + C_{11}P[D = H_1|H_1]P[H_1]\end{aligned}}$$

- ***Bayes decision rule:***

  A Bayes decision rule is a decision rule which minimizes the average cost $\bar{C}$. It is of the form:

$$\boxed{L(y) = \frac{f(y|H_1)}{f(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P[H_0](C_{10} - C_{00})}{P[H_1](C_{01} - C_{11})}}$$

  Proof:

  The average cost $\bar{C}$ can be written as:

$$\bar{C} = C_{10}P[H_0] + C_{11}P[H_1] +$$
$$+ \int_{R_0} \{P[H_1](C_{01} - C_{11})f(y|H_1) - P[H_0](C_{10} - C_{00})f(y|H_0)\}dy$$

$\square$

  Note that the Bayes rule reduces to the MAP rule when the cost is selected to be:

$$C_{00} = C_{11} = 0$$
$$C_{10} = C_{01} = 1$$

3.2.3. Minimax and Neyman-Pearson decision rule

Bayes decision rules necessitate the specification of an “a priori” pdf:

$$P[H_0], P[H_1] = 1 - P[H_0].$$

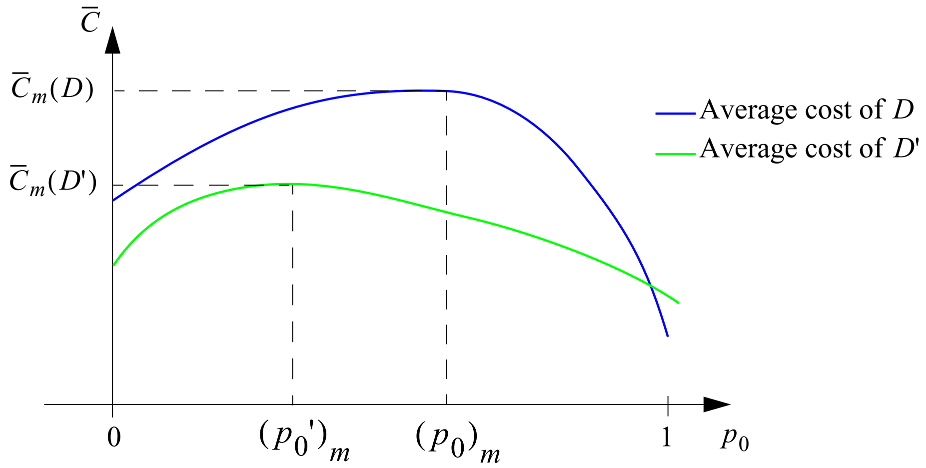
In some situations such a pdf is unknown and difficult to assess or even the definition of such a pdf does not make sense. In this case we have to resort to alternative decision rules.

- ***Minimax decision rule:***

  The minimax decision rule is employed when the “a priori” pdf is unknown.

  - Maximum average cost of a decision rule:

  Let us consider the behavior of $\bar{C}$ for a fixed decision rule D as $p_0 \equiv P[H_0]$ varies:



  $\bar{C}_m(D)$ is the maximum average cost and

  $P[H_0] = (p_0)_m, P[H_1] = 1 - (p_0)_m$ is the worst case “a priori” pdf

  when employing decision rule D.

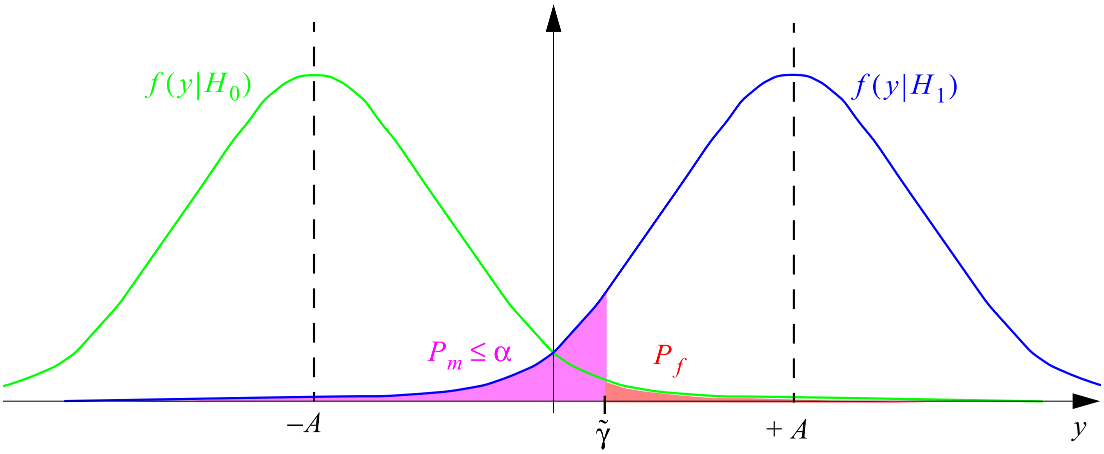
  - Minimax decision rule:

  A minimax decision rule, say D_{MM} minimizes $\bar{C}_m$:

$$\boxed{\bar{C}_m(D_{MM}) \le \bar{C}_m(D) \qquad \text{for any decision rule } D}$$

  Minimax $\equiv$ **mini**mize the **max**imum of $\bar{C}_m$.

- ***Neyman-Pearson decision rule:***
  The Neyman-Pearson (NP) decision rule is used when neither an “a priori” pdf nor cost assignments are given.
  A NP decision rule minimizes the probability of false alarm
  $P_f = P[D = H_1|H_0]$ while keeping the probability of a miss
  $P_m = P[D = H_0|H_1]$ below a certain specified level, say α.
  Example: BPAM (cont’d)



Thus, a NP decision rule D_{NP_α} satisfies the inequality

$$\boxed{P_f(D_{NP_\alpha}) \leq P_f(D) \text{ for any decision rule } D \text{ such that } P_m(D) \leq \alpha}$$

3.2.4. General form of a binary decision rule:

Both minimax and NP decision rules can be shown to be of the form:

$$\boxed{L(y) = \frac{f(y|H_1)}{f(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma}$$

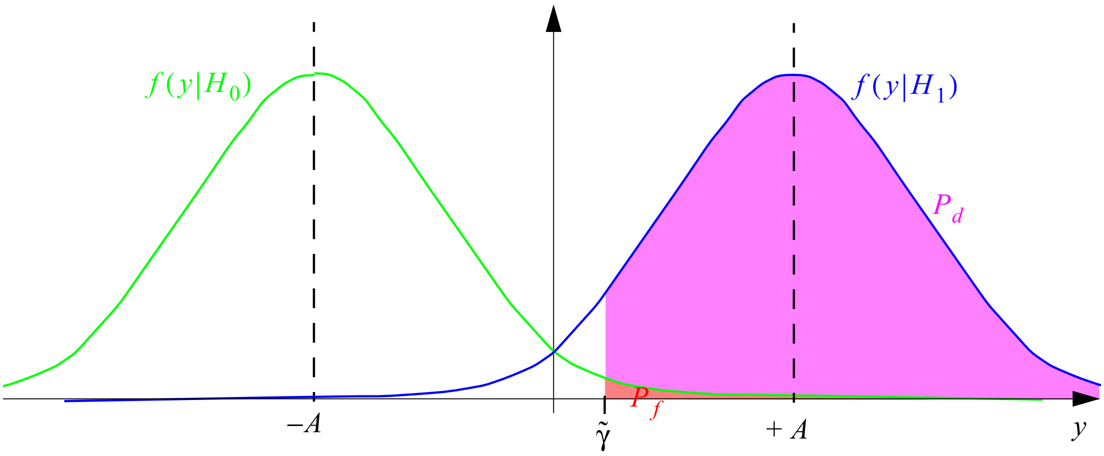
or equivalently

$$\boxed{l(y) = \ln\left(\frac{f(y|H_1)}{f(y|H_0)}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\gamma)}$$

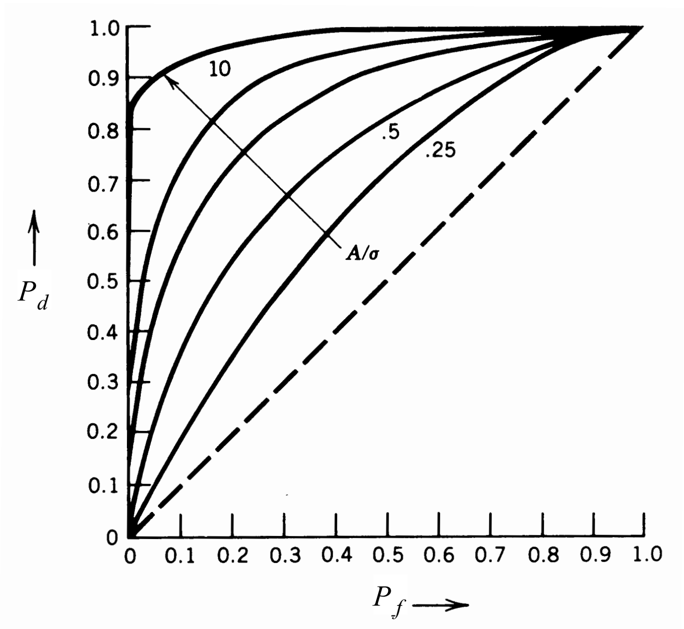
for some decision threshold γ.
Notice that the Bayes decision rule, and therefore the MAP decision rule as well, are also of the same form with

$$\gamma = \frac{P[H_0](C_{10} - C_{00})}{P[H_1](C_{01} - C_{11})}.$$

- ***Receiver operating characteristics (ROC):***
  In radar technique the performance of detectors are described in terms of a graph representing the probability of correct detection $P_d = P[D = H_1|H_1]$ versus the false alarm probability $P_f = P[D = H_1|H_0]$:
  Example: BPAM (cont'd)



ROC for $\gamma = \gamma_{MAP} = 1 \Rightarrow \tilde{\gamma} = \tilde{\gamma}_{MAP} = 0$:



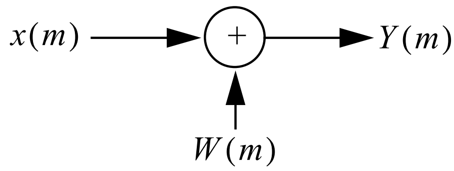
3.3. Binary detection of discrete-time signals

3.3.1. Time-limited discrete-time signals

- ***Signal model:***

$$Y(m) = x(m) + W(m)$$

$$m = 0, \ldots, M-1$$

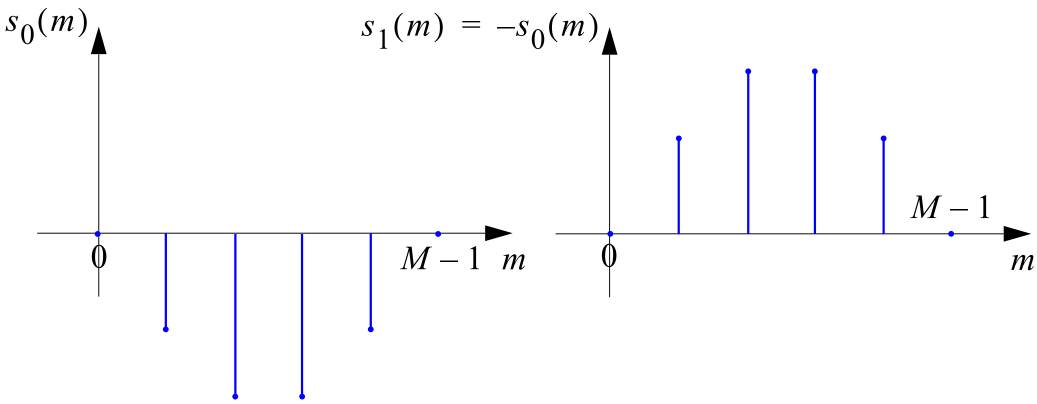


where

$$\text{- } x(m) = \begin{cases} s_0(m) & \text{under} \quad H_0 \\ s_1(m) & \text{under} \quad H_1 \end{cases}$$

- $W(m)$ is a white Gaussian noise:
    - $W(m)$ is a Gaussian process,
    - $\mathbf{E}[W(m)] = 0$,
    - $\mathbf{E}[W(m)W(m+k)] = \sigma^2\delta(k)$.

Example: *Detection of BPAM signals*



- ***Vector representation of finite sequences:***
    - Deterministic signals:

$$\boldsymbol{u} \equiv [u(0), \ldots, u(M-1)]^T$$

- Random sequences:

$$\boldsymbol{U} \equiv [U(0), \ldots, U(M-1)]^T$$

- ***Pdf of Y under H_0 and H_1:***

  - Vector representation of the received signal:

$$\boldsymbol{Y} = \boldsymbol{x} + \boldsymbol{W}$$

where

$$\boldsymbol{x} = \begin{cases} \boldsymbol{s}_0 & \text{under} \quad H_0 \\ \boldsymbol{s}_1 & \text{under} \quad H_1 \end{cases}$$

  - Pdf of $\boldsymbol{W}$:

$$\begin{aligned} f(\boldsymbol{w}) &= \prod_{m=0}^{M-1} f(w(m)) \\ &= \prod_{m=0}^{M-1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2} w(m)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{w}\|^2\right\} \end{aligned}$$

where

$$\|\boldsymbol{w}\| \equiv \sqrt{\sum_{m=0}^{M-1} w(m)^2}$$

is the norm of $\boldsymbol{w}$.

  - Pdf of $\boldsymbol{Y}$ under H_0 and H_1:

$$\begin{aligned} f(\boldsymbol{y}|H_i) &= f(\boldsymbol{w})\big|_{\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{s}_i} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{s}_i\|^2\right\} \end{aligned}$$

- ***Likelihood and loglikelihood ratios:***
  - Likelihood ratio:

$$L(\boldsymbol{y}) = \frac{f(\boldsymbol{y}|H_1)}{f(\boldsymbol{y}|H_0)}$$

$$= \frac{\exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{s}_1\|^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{s}_0\|^2\right\}} = \exp\left\{\frac{1}{2\sigma^2}[\|\boldsymbol{y}-\boldsymbol{s}_0\|^2 - \|\boldsymbol{y}-\boldsymbol{s}_1\|^2]\right\}$$

  - Loglikelihood ratio:

$$l(\boldsymbol{y}) = \ln\left(\frac{f(\boldsymbol{y}|H_1)}{f(\boldsymbol{y}|H_0)}\right)$$

$$= \frac{1}{2\sigma^2}[\|\boldsymbol{y}-\boldsymbol{s}_0\|^2 - \|\boldsymbol{y}-\boldsymbol{s}_1\|^2]$$

$$= \frac{1}{\sigma^2}\left[\boldsymbol{y}^T(\boldsymbol{s}_1-\boldsymbol{s}_0) + \frac{1}{2}(\|\boldsymbol{s}_0\|^2 - \|\boldsymbol{s}_1\|^2)\right]$$

- ***Decision rules:***

$$l(\boldsymbol{y}) = \frac{1}{\sigma^2}\left[\boldsymbol{y}^T(\boldsymbol{s}_1-\boldsymbol{s}_0) + \frac{1}{2}(\|\boldsymbol{s}_0\|^2 - \|\boldsymbol{s}_1\|^2)\right] \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\gamma)$$

or equivalently

$$\boxed{\boldsymbol{y}^T(\boldsymbol{s}_1-\boldsymbol{s}_0) \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2\ln(\gamma) + \frac{1}{2}(E_{\boldsymbol{s}_1} - E_{\boldsymbol{s}_0})}$$

where

$$E_{\boldsymbol{s}_i} \equiv \|\boldsymbol{s}_i\|^2 = \sum_{m=0}^{M-1} s_i(m)^2$$

is the energy of the signal $\boldsymbol{s}_i$, $i = 0, 1$.

Comment:

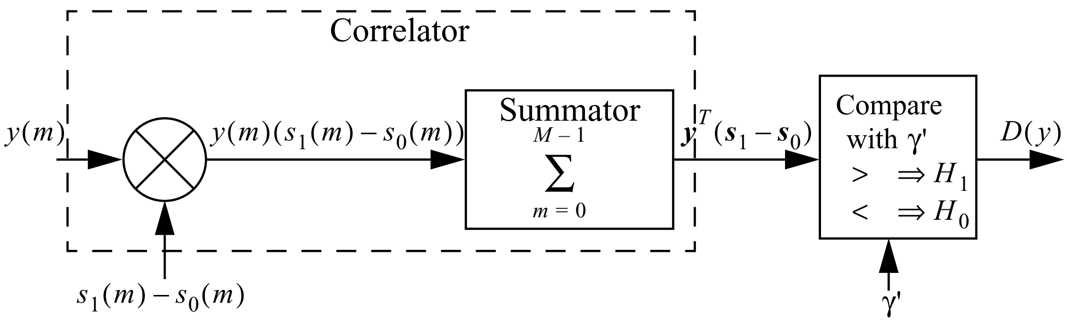
Notice that $\boldsymbol{y}^T(\boldsymbol{s}_1 - \boldsymbol{s}_0)$ is the scalar product of $\boldsymbol{y}$ and $\boldsymbol{s}_1 - \boldsymbol{s}_0$, i.e.

$$\boldsymbol{y}^T(\boldsymbol{s}_1 - \boldsymbol{s}_0) = \sum_{m=0}^{M-1} y(m)(s_1(m) - s_0(m))$$

Special case: MAP decision rule:

$$\boldsymbol{y}^T(\boldsymbol{s}_1 - \boldsymbol{s}_0) \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2 \ln\left(\frac{P[H_0]}{P[H_1]}\right) + \frac{1}{2}(E_{\boldsymbol{s}_1} - E_{\boldsymbol{s}_0})$$

- ***Block diagram of a binary detector for time-limited discrete-time signals:***



$$\gamma' = \sigma^2 \ln(\gamma) + \frac{1}{2}(E_{\boldsymbol{s}_1} - E_{\boldsymbol{s}_0})$$

3.3.2. Discrete-time signals with finite energy

- ***Signal model:***

$$Y(m) = x(m) + W(m) \qquad m = \ldots, -2, -1, 0, 1, 2, \ldots$$

where

- $x(m) = \begin{cases} s_0(m) & \text{under} \quad H_0 \\ s_1(m) & \text{under} \quad H_1 \end{cases}$

  where $s_i(m)$, $i = 0, 1$ has finite energy, i.e.

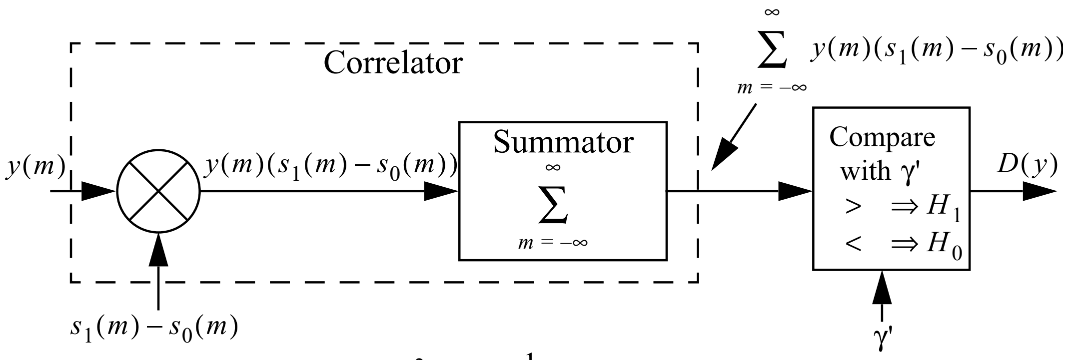
$$E_{s_i} \equiv \sum_{m=-\infty}^{\infty} s_i(m)^2 < \infty$$

- $W(m)$ is a white Gaussian noise sequence with variance σ^2.

- ***Decision rules:***
  Decision rules can be obtained in this case as well which prove to be of the form:

$$\boxed{\sum_{m=-\infty}^{\infty} y(m)(s_1(m)-s_0(m)) \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2 \ln(\gamma) + \frac{1}{2}(E_{s_1} - E_{s_0})}$$

  Hence, these decision rules are essentially the same as those derived for time-limited sequences.



$$\gamma' = \sigma^2 \ln(\gamma) + \frac{1}{2}(E_{s_1} - E_{s_0})$$

3.4. Binary detection of continuous-time signals

- ***Signal model:***

with
$$Y(t) = x(t) + W(t)$$

- $x(t) = \begin{cases} s_0(t) & \text{under} \quad H_0 \\ s_1(t) & \text{under} \quad H_1 \end{cases}$.

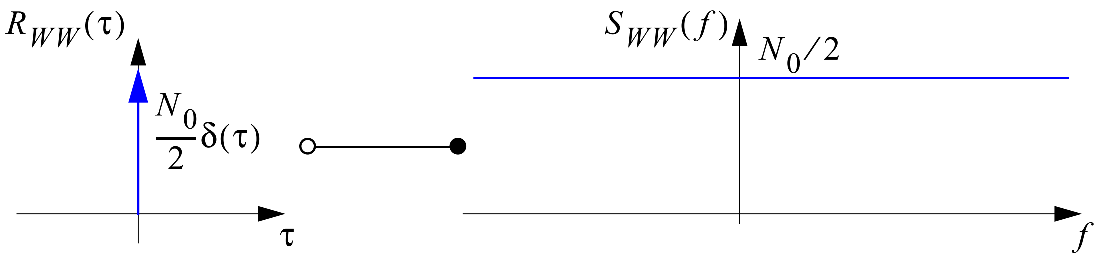
  The signals $s_0(t)$ and $s_1(t)$ have finite energy, i.e.

$$E_{s_i} \equiv \int_{-\infty}^{\infty} s_i(t)^2 dt < \infty \qquad i = 0, 1$$

- $W(t)$ is a white Gaussian noise:
  - $W(t)$ is a Gaussian process,
  - $\mathbf{E}[W(t)] = 0$,
  - $R_{WW}(\tau) = \mathbf{E}[W(t)W(t+\tau)] = \frac{N_0}{2}\delta(\tau)$.

The power spectral density function of $W(t)$ reads:

$$S_{WW}(f) = \mathcal{F}\{R_{WW}(\tau)\} = \frac{N_0}{2}$$

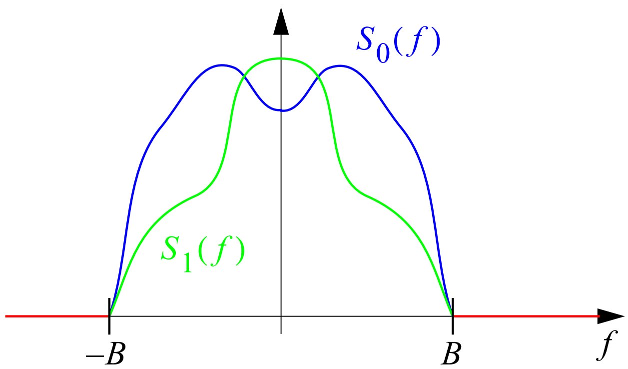


- ***Key issue:***
  Subsequently we consider two situations which prove to be equivalent to that previously considered in Sect. 3.3.2.

3.4.1. Bandwidth-limited continuous-time signals

- ***Bandwidth-limited signals:***
  The signals $s_0(t)$ and $s_1(t)$ are bandwidth-limited with bandwidth B:

Spectrum of $s_0(t)$ and $s_1(t)$:



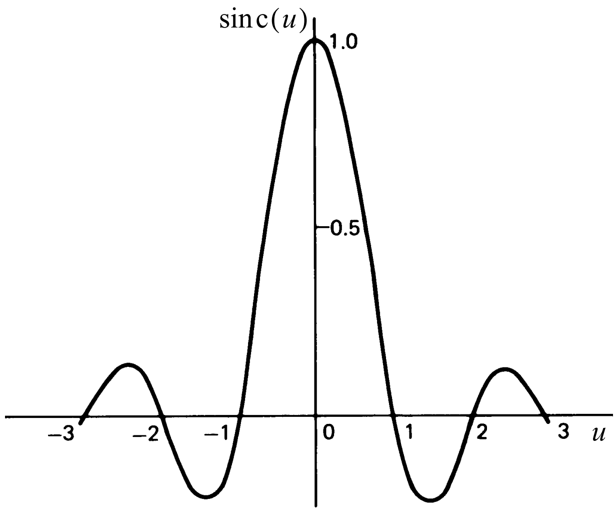
- ***Sampling theorem for bandwidth-limited signals:***
  The signal $s_i(t)$, $i = 0, 1$, can be represented as

$$\boxed{s_i(t) = \sum_{m=-\infty}^{\infty} s_i(mT_s)\,\mathrm{sinc}\left(\frac{1}{T_s}(t - mT_s)\right)},$$

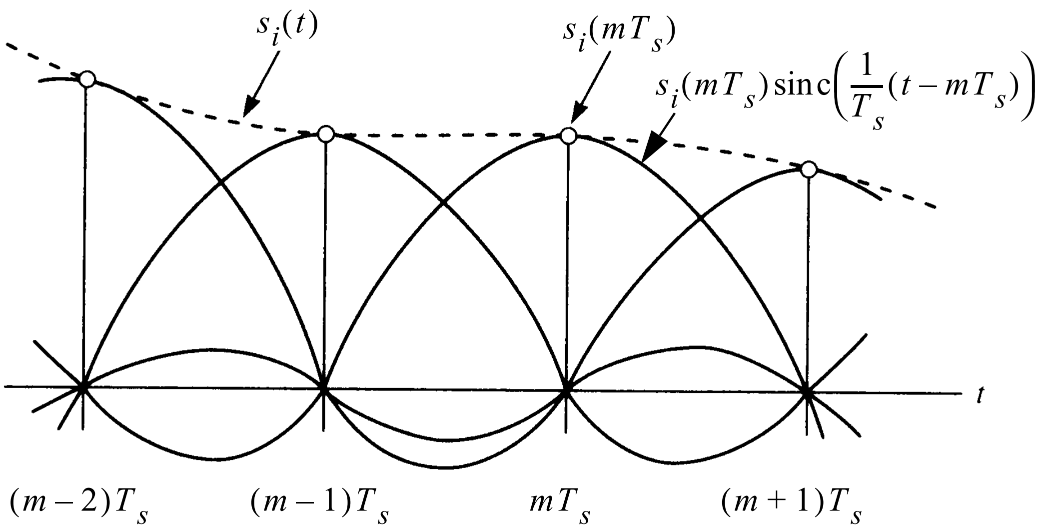
where

- $\frac{1}{T_s} = 2B$ is the Nyquist rate

- $\mathrm{sinc}(u) \equiv \frac{\sin(\pi u)}{\pi u}$:



According to the above sum, $s_i(t)$ is entirely determined by its samples $s_i(mT_s)$, $m = \ldots, -1, 0, 1, \ldots$:



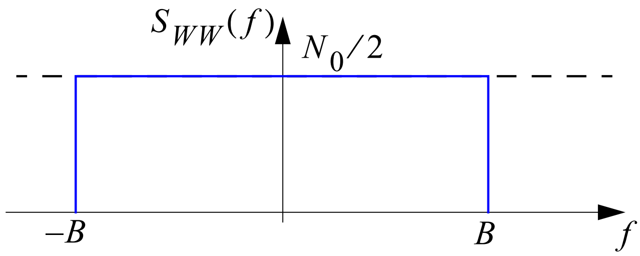
- ***Parseval relationship for bandwidth-limited finite-energy signals:***
  Let $u(t)$ and $v(t)$ denote two finite-energy signals with bandwidth B. Then

$$\boxed{\int_{-\infty}^{\infty} u(t)v(t)dt = T_s \sum_{m=-\infty}^{\infty} u(mT_s)v(mT_s)}\ .$$

In particular,

$$\boxed{E_u = \int_{-\infty}^{\infty} u(t)^2 dt = T_s \sum_{m=-\infty}^{\infty} u(mT_s)^2} \, .$$

- ***Sampling theorem for bandwidth-limited processes:***
  Without loss of generality we can assume that $W(t)$ is bandwidth-limited with bandwidth B.



Then,

$$\boxed{W(t) = \sum_{m=-\infty}^{\infty} W(mT_s) \operatorname{sinc}\left(\frac{1}{T_s}(t - mT_s)\right)},$$

where $W(mT_s)$ is the sample of $W(t)$ at $t = mT_s$.

Moreover, it can be shown that the sequence $W(mT_s)$, $m = \ldots, -1, 0, 1, \ldots$, is a white Gaussian sequence with variance

$$\mathbf{E}[W(mT_s)^2] = \int S_{WW}(f) df = N_0 B \, .$$

- ***Decision rules:***
  The identification

$$s_i(t) \quad \leftrightarrow \quad s_i(m) = s_i(mT_s), \, m = \ldots, -1, 0, 1, \ldots$$

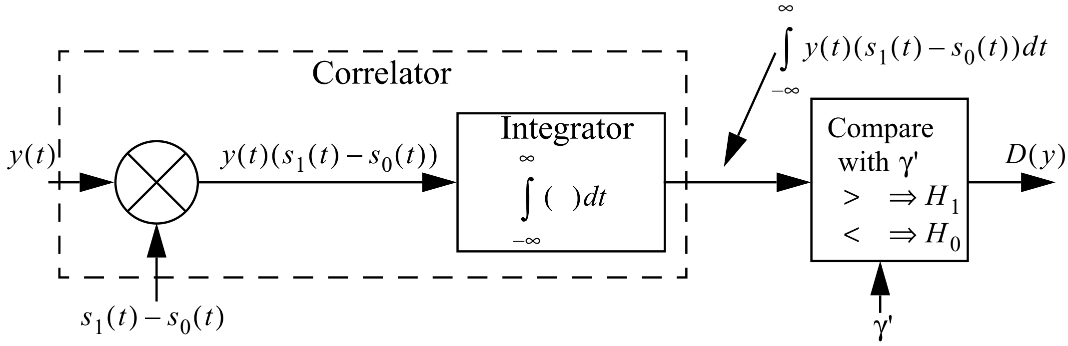
$$W(t) \quad \leftrightarrow \quad W_i(m) = W_i(mT_s), \, m = \ldots, -1, 0, 1, \ldots$$

in combination with the Parseval relation allows to reduce the situation of continuous-time bandlimited signals with finite energy to the case of infinite sequences with finite energy considered in Sect. 3.3.2.

Invoking the Parseval relation again and the result obtained in Sect. 3.3.2, the decision rules are found to be of the form:

$$\boxed{\int_{-\infty}^{\infty} y(t)(s_1(t) - s_0(t))dt \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N_0}{2} \ln(\gamma) + \frac{1}{2}(E_{s_1} - E_{s_0})}$$

- ***Block diagram of a binary detector for bandwidth-limited continuous-time signals:***



$$\gamma' = \sigma^2 \ln(\gamma) + \frac{1}{2}(E_{s_1} - E_{s_0})$$

3.4.2. Time-limited continuous-time signals

A similar rational as used above can be applied to time-limited (but possibly bandwidth unlimited) finite-energy signals.

- ***Time-limited signals:***

  Let $[T_u, T_o]$ be an interval outside which $s_0(t)$ and $s_1(t)$ vanish, i.e.

$$s_i(t) = 0 \qquad t \notin [T_u, T_o] \qquad i = 0, 1\,.$$

- ***Decision rules:***

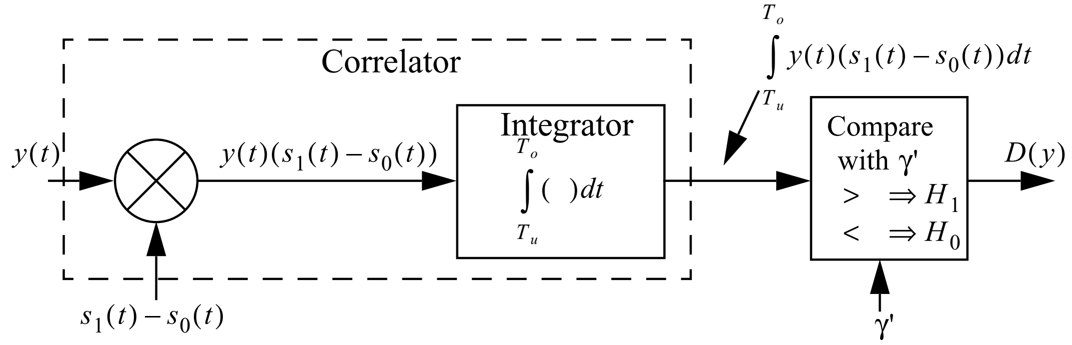
  The decision rules are found to be of the form:

$$\boxed{\int_{T_u}^{T_o} y(t)(s_1(t) - s_0(t))dt \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N_0}{2}\ln(\gamma) + \frac{1}{2}(E_{s_1} - E_{s_0})}$$

where

$$E_{s_i} = \int_{T_u}^{T_o} s_i(t)^2 dt$$

- ***Block diagram of a binary detector for time-limited continuous-time signals:***

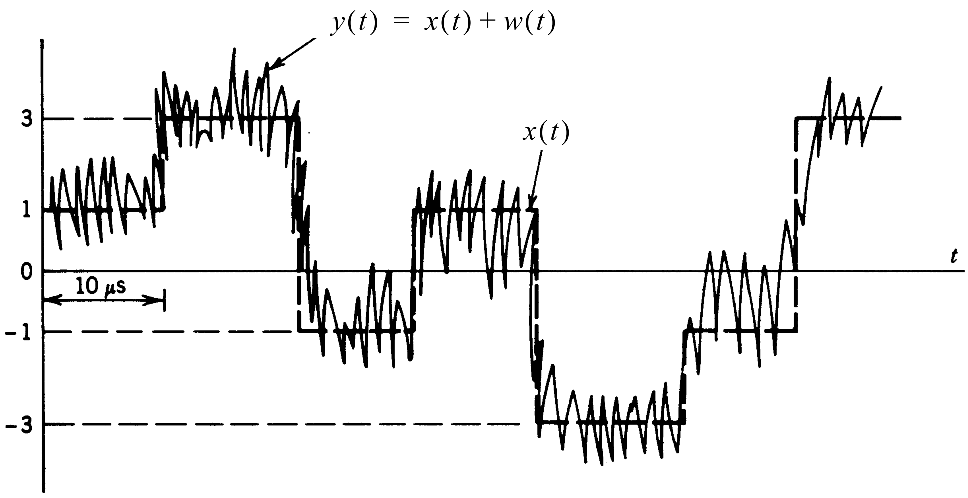


$$\gamma' = \sigma^2 \ln(\gamma) + \frac{1}{2}(E_{s_1} - E_{s_0})$$

3.5. M-ary detection

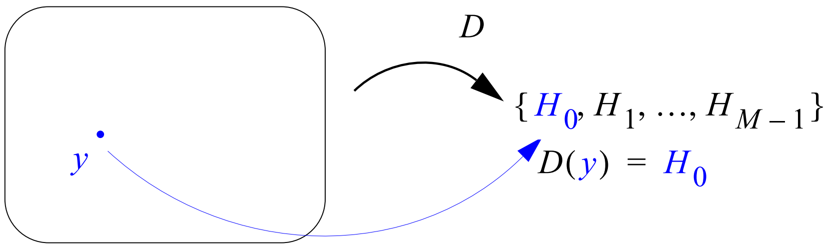
- ***Multiple hypothesis testing:***
  So far, we have considered the problem of deciding between one among two hypotheses. In many engineering problems a decision must be taken between more that two possibilities, say $H_0, \ldots, H_{M-1}$.

  Example: 4-PAM

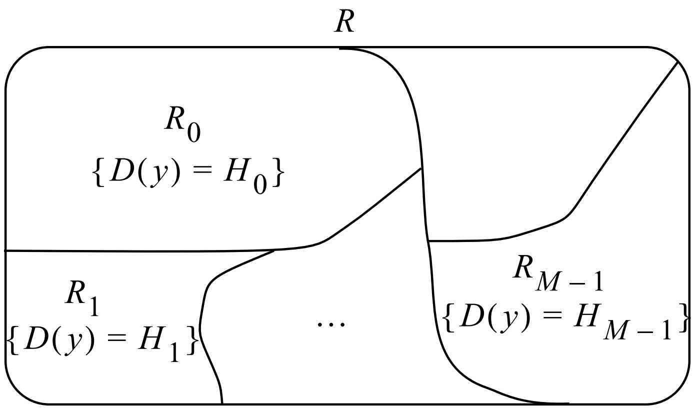


- ***Decision rule and decision regions:***
    - Decision rule:

R : range of Y



    - Decision regions:



$$R_0 \cup R_1 \cup \ldots \cup R_{M-1} = R$$
$$R_0 \cap R_1 \cap \ldots \cap R_{M-1} = \varnothing$$

- ***Probabilities of correct decision and of making an error:***
    - Probability of correct decision:

$$\boxed{\begin{aligned} P_c &= \sum_{i=0}^{M-1} P[D = H_i, H = H_i] \\ &= \sum_{i=0}^{M-1} P[D = H_i | H_i] P[H_i] \end{aligned}}$$

    - Probability of incorrect decision:

$$\boxed{P_e = 1 - P_c}$$

- ***MAP for M-ary detection:***

  In order to minimize P_e a decision rule must select an hypothesis whose "a posteriori" probability is maximum:

$$\boxed{\text{Select } H_i \text{ if } P[H_i|y] \geq P[H_j|y] \text{ for any } j = 0, \ldots, M-1.}$$

or equivalently, by invoking Bayes' rule:

$$\boxed{\text{Select } H_i \text{ if } \frac{f(y|H_i)}{f(y|H_j)} \geq \frac{P[H_j]}{P[H_i]} \text{ for any } j = 0, \ldots, M-1.}$$

Using the log-likelihood function:

$$\boxed{\text{Select } H_i \text{ if } \ln\left(\frac{f(y|H_i)}{f(y|H_j)}\right) \geq \ln\left(\frac{P[H_j]}{P[H_i]}\right) \text{ for any } j = 0, \ldots, M-1.}$$

Example: M-ary MAP decision rule for time-limited discrete-time signals

$$\text{Select } H_i \text{ if}$$

$$\boldsymbol{y}^T\boldsymbol{s}_i + \sigma^2\ln(P[H_i]) - \frac{1}{2}E_{\boldsymbol{s}_i} \geq \boldsymbol{y}^T\boldsymbol{s}_j + \sigma^2\ln(P[H_j]) - \frac{1}{2}E_{\boldsymbol{s}_j}$$

$$\text{for any } j = 0, \ldots, M-1.$$

Compact formulation of the MAP decision rule:

$$\hat{H} = H_{\hat{i}} \text{ where}$$

$$\hat{i} = \operatorname{argmax}_i\left\{\boldsymbol{y}^T\boldsymbol{s}_i + \sigma^2\ln(P[H_i]) - \frac{1}{2}E_{\boldsymbol{s}_i}\right\}$$

Special case: ML decision rule:

Selecting the uniform "a priori" pdf

$$P[H_0] = P[H_1] = \ldots = P[H_{M-1}] = \frac{1}{M}$$

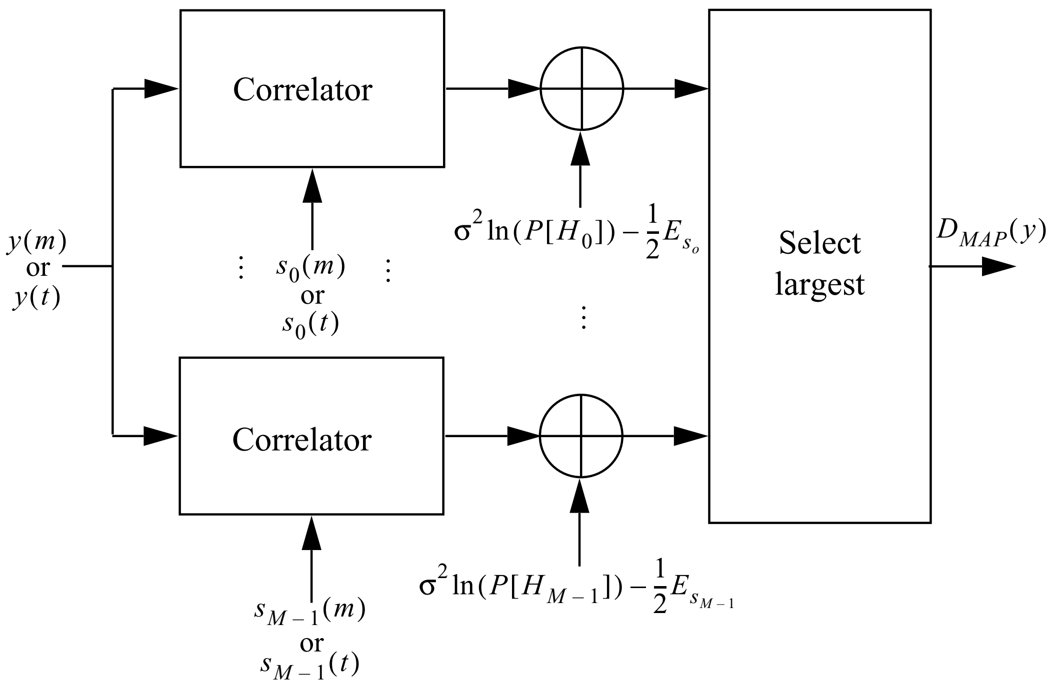
yields the ML decision rule:

$$\hat{H} = H_{\hat{i}} \text{ where}$$

$$\hat{i} = \operatorname{argmax}_i\left\{\boldsymbol{y}^T\boldsymbol{s}_i - \frac{1}{2}E_{\boldsymbol{s}_i}\right\}$$

Comment: The M-ary MAP decision rules for all other situations previously considered are obtained by appropriately replacing the scalar product in the above decision rules.

- ***Block diagram of a M-ary MAP detector:***



4. Linear Minimum Mean Squared Error Estimation

4.1. Linear minimum mean squared error estimators

- ***Situation considered:***
  - A random sequence $X(1), \ldots, X(M)$ whose realizations can be observed.
  - A random variable Y which has to be estimated.
  - We seek an estimate of Y with a **linear estimator** of the form:

$$\boxed{\hat{Y} = h_0 + \sum_{m=1}^{M} h_m X(m)}\ .$$

  - A measure of the goodness of $\hat{Y}$ is the mean squared error (MSE):

$$\boxed{\mathbf{E}[(\hat{Y} - Y)^2]}\ .$$

- ***Covariance and variance of random variables:***
  Let U and V denote two random variables with expectation $\mu_U \equiv \mathbf{E}[U]$ and $\mu_V \equiv \mathbf{E}[V]$.
  - The **covariance** of U and V is defined to be:

$$\begin{aligned} \Sigma_{UV} &\equiv \mathbf{E}[(U - \mu_U)(V - \mu_V)] \\ &= \mathbf{E}[UV] - \mu_U \mu_V \end{aligned}$$

  - The **variance** of U is defined to be:

$$\begin{aligned} \sigma_U^2 &\equiv \mathbf{E}[(U - \mu_U)^2] = \Sigma_{UU} \\ &= \mathbf{E}[U^2] - (\mu_U)^2 \end{aligned}$$

Let $\boldsymbol{U} \equiv [U(1), \dots, U(M)]^T$ and $\boldsymbol{V} \equiv [V(1), \dots, V(M')]^T$ denote two random vectors.

The covariance matrix of $\boldsymbol{U}$ and $\boldsymbol{V}$ is defined as

$$\Sigma_{\boldsymbol{UV}} \equiv \begin{bmatrix} \Sigma_{U(1)V(1)} & \cdots & \Sigma_{U(1)V(M')} \\ \cdots & \cdots & \cdots \\ \Sigma_{U(M)V(1)} & \cdots & \Sigma_{U(M)V(M')} \end{bmatrix}$$

A direct way to obtain $\Sigma_{\boldsymbol{UV}}$:

$$\begin{aligned} \Sigma_{\boldsymbol{UV}} &= \mathbf{E}[(\boldsymbol{U} - \mu_{\boldsymbol{U}})(\boldsymbol{V} - \mu_{\boldsymbol{V}})^T] \\ &= \mathbf{E}[\boldsymbol{U}\boldsymbol{V}^T] - \mu_{\boldsymbol{U}}(\mu_{\boldsymbol{V}})^T \end{aligned}$$

where

$$\begin{aligned} \mu_{\boldsymbol{U}} &\equiv \mathbf{E}[\boldsymbol{U}] = [\mathbf{E}[U(1)], \dots, \mathbf{E}[U(M)]]^T \\ \mu_{\boldsymbol{V}} &\equiv \mathbf{E}[\boldsymbol{V}] \end{aligned}$$

Examples: $\boldsymbol{U} = \boldsymbol{X} \equiv [X(1), \dots, X(M)]^T$ and $\boldsymbol{V} = Y$.

In the sequel we shall frequently make use of the following covariance matrix and vector:

(i) $$\begin{aligned} \Sigma_{\boldsymbol{XX}} &= \mathbf{E}[(\boldsymbol{X} - \mu_{\boldsymbol{X}})(\boldsymbol{X} - \mu_{\boldsymbol{X}})^T] \\ &= \begin{bmatrix} \sigma^2_{X(1)} & \cdots & \Sigma_{X(1)X(M)} \\ \cdots & \cdots & \cdots \\ \Sigma_{X(M)X(1)} & \cdots & \sigma^2_{X(M)} \end{bmatrix} \end{aligned}$$

(ii) $$\begin{aligned} \Sigma_{\boldsymbol{X}Y} &= \mathbf{E}[(\boldsymbol{X} - \mu_{\boldsymbol{X}})(Y - \mu_Y)] \\ &= \begin{bmatrix} \Sigma_{X(1)Y} & \cdots & \Sigma_{X(M)Y} \end{bmatrix}^T \end{aligned}$$

- ***Linear minimum mean squared error estimator (LMMSEE)***

  A LMMSEE of Y is a linear estimator, i.e. an estimator of the form

$$\hat{Y} = h_0 + \sum_{m=1}^{M} h_m X(m) \ ,$$

which minimizes the MSE $\mathbf{E}[(\hat{Y} - Y)^2]$.

A linear estimator is entirely determined by the $(M+1)$-dimensional vector $\boldsymbol{h} \equiv [h_0, \ldots, h_M]^T$.

- ***Orthogonality principle:***

> **Orthogonality principle:**
> A necessary condition for $\boldsymbol{h} \equiv [h_0, \ldots, h_M]^T$ to be the coefficient vector of the LMMSEE is that its entries fulfils the $(M+1)$ identities:
>
> $$\mathbf{E}[Y - \hat{Y}] = \mathbf{E}\left[Y - \left(h_0 + \sum_{m=1}^{M} h_m X(m)\right)\right] = 0 \qquad (4.1a)$$
>
> $$\mathbf{E}[(Y - \hat{Y})X(j)] = \mathbf{E}\left[\left\{Y - \left(h_0 + \sum_{m=1}^{M} h_m X(m)\right)\right\}X(j)\right] = 0, \qquad (4.1b)$$
>
> $$j = 1, \ldots, M$$

Proof:

Because the coefficient vector of the LMMSEE minimizes $\mathbf{E}[(\hat{Y} - Y)^2]$, its components must satisfy the set of equations:

$$\frac{\partial}{\partial h_j}\mathbf{E}[(\hat{Y} - Y)^2] = 0 \qquad j = 0, \ldots, M.$$

$\square$

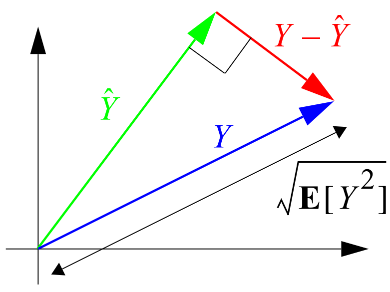
Notice that the two expressions in (4.1) can be rewritten as:

$$\mathbf{E}[Y - \hat{Y}] = 0 \qquad (4.2a)$$

$$\mathbf{E}[(Y - \hat{Y})X(j)] = 0 \qquad j = 1, \ldots, M \qquad (4.2b)$$

Important consequences of the orthogonality principle:

> $$\mathbf{E}[(Y - \hat{Y})\hat{Y}] = 0 \qquad (4.3a)$$
>
> $$\mathbf{E}[(Y - \hat{Y})^2] = \mathbf{E}[Y^2] - \mathbf{E}[\hat{Y}^2] = \mathbf{E}[(Y - \hat{Y})Y] \qquad (4.3b)$$



Geometrical interpretation:
Let U and V denote two random variables with finite second moment, i.e.
$\mathbf{E}[U^2] < \infty$ and $\mathbf{E}[V^2] < \infty$.
Then, the expectation $\mathbf{E}[UV]$ can be viewed as the **scalar or inner product** of U and V.

Within this interpretation:

- U and V are uncorrelated, i.e. $\mathbf{E}[UV] = 0$ if and only if, they are **orthogonal**,
- $\sqrt{\mathbf{E}[U^2]}$ is the norm (length) of U.

Interpretation of both equations in (4.3):

- (4.3a): the estimation error $Y - \hat{Y}$ and the estimate $\hat{Y}$ are orthogonal.
- (4.3b): results from Pythagoras' Theorem.

- ***Computation of the coefficient vector of the LMMSEE:***
The coefficients of the LMMSEE satisfy the relationships:

$$\boxed{\begin{aligned} \mu_Y &= h_0 + \sum_{m=1}^{M} h_m \mu_{X(m)} = h_0 + (\bar{\boldsymbol{h}})^T \mu_{\boldsymbol{X}} \\ \Sigma_{\boldsymbol{X}Y} &= \Sigma_{\boldsymbol{XX}} \bar{\boldsymbol{h}} \end{aligned}}$$

where $\bar{\boldsymbol{h}} \equiv [h_1, \ldots, h_M]^T$ and $\boldsymbol{X} \equiv [X_1, \ldots, X_M]^T$.
Proof:
Both identities follow by appropriately reformulating relations (4.1a) and (4.1b) and using a matrix notation for the latter one.

$\square$

Thus, provided $(\Sigma_{\boldsymbol{XX}})^{-1}$ exists the coefficients of the LMMSEE are given by:

$$\boxed{\begin{aligned} \bar{\boldsymbol{h}} &= (\Sigma_{\boldsymbol{XX}})^{-1}\Sigma_{\boldsymbol{X}Y} && (4.4a) \\ h_0 &= \mu_Y - (\bar{\boldsymbol{h}})^T\mu_{\boldsymbol{X}} = \mu_Y - \Sigma_{\boldsymbol{X}Y}{}^T(\Sigma_{\boldsymbol{XX}})^{-1}\mu_{\boldsymbol{X}} && (4.4b) \end{aligned}}$$

- ***Example: Linear prediction of a WSS process***

  Let $Y(n)$ denote a WSS process with

  - zero mean, i.e $\mathbf{E}[Y(n)] = 0$,
  - autocorrelation function $\mathbf{E}[Y(n)Y(n+k)] = R_{YY}(k)$

  We seek the LMMSEE for the present value of $Y(n)$ based on the M past observations $Y(n-1), \ldots, Y(n-M)$ of the process. Hence,

  - $Y = Y(n)$
  - $X(m) = Y(n-m)$, $m = 1, \ldots, M$, i.e.

    $\boldsymbol{X} = [Y(n-1), \ldots, Y(n-M)]^T$

  Because $\mu_Y = 0$ and $\mu_{\boldsymbol{X}} = 0$, it follows from $(4.4b)$ that

  $$h_0 = 0$$

  Computation of $\Sigma_{\boldsymbol{X}Y}$ and $\Sigma_{\boldsymbol{XX}}$:

  - $\Sigma_{\boldsymbol{X}Y} = [\mathbf{E}[Y(n-1)Y(n)], \ldots, \mathbf{E}[Y(n-M)Y(n)]]^T$

    $= [R_{YY}(1), \ldots, R_{YY}(M)]^T$

  - $\Sigma_{\boldsymbol{XX}} =$

$$= \begin{bmatrix} \mathbf{E}[Y(n-1)^2] & \mathbf{E}[Y(n-1)Y(n-2)] & \ldots & \mathbf{E}[Y(n-1)Y(n-M)] \\ \mathbf{E}[Y(n-2)Y(n-1)] & \mathbf{E}[Y(n-2)^2] & \ldots & \mathbf{E}[Y(n-2)Y(n-M)] \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{E}[Y(n-M)Y(n-1)] & \mathbf{E}[Y(n-M)Y(n-2)] & \ldots & \mathbf{E}[Y(n-M)^2] \end{bmatrix}$$

$$= \begin{bmatrix} R_{YY}(0) & R_{YY}(1) & R_{YY}(2) & \ldots & R_{YY}(M-1) \\ R_{YY}(1) & R_{YY}(0) & R_{YY}(1) & \ldots & R_{YY}(M-2) \\ R_{YY}(2) & R_{YY}(1) & R_{YY}(0) & \ldots & R_{YY}(M-3) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ R_{YY}(M-1) & R_{YY}(M-2) & R_{YY}(M-3) & \ldots & R_{YY}(0) \end{bmatrix}$$

- ***Residual error using a LMMSEE:***
  The MSE resulting when using a LMMSEE is

$$\boxed{\mathbf{E}[(\hat{Y}-Y)^2] = \sigma_Y{}^2 - (\boldsymbol{h}^{-})^T \Sigma_{\boldsymbol{X}Y} \qquad (4.5)}$$

Proof:

$$\begin{aligned}\mathbf{E}[(Y-\hat{Y})^2] &= \mathbf{E}[(Y-\hat{Y})Y] \\ &= \mathbf{E}[Y^2] - \mathbf{E}[\hat{Y}Y]\end{aligned}$$

$\square$

Notice the similarity between Identities (4.4a) and (4.5) and the **Yule-Walker equations** (pp. 2.7-2.8).

4.2. Minimum mean squared error estimators

- ***Conditional expectation:***
  Let U and V denote two random variables.
  The conditional expectation of V given $U = u$ is observed is defined to be

$$\mathbf{E}[V|u] \equiv \int vp(v|u)dv .$$

Notice that $\mathbf{E}[V|U]$ is a random variable. In the sequel we shall make use of the following important property of conditional expectations:

$$\mathbf{E}[\mathbf{E}[V|U]] = \mathbf{E}[V]$$

Proof:

$\square$

- ***Minimum mean squared error estimator (MMSEE):***

  The MMSEE of Y based on the observation of $X(1), \ldots, X(M)$ is of the form:

$$\boxed{\widehat{Y}(X(1), \ldots, X(M)) = \mathbf{E}[Y|X(1), \ldots, X(M)]}$$

Hence if $X(1) = x(1), \ldots, X(M) = x(M)$ is observed, then

$$\begin{aligned} \widehat{Y}(x(1), \ldots, x(M)) &= \mathbf{E}[Y|x(1), \ldots, x(M)] \\ &= \int yp(y|x(1), \ldots, x(M))dy \end{aligned}$$

Proof:

Let $\hat{Y}$ denote an arbitrary estimator. Then,

$$\begin{aligned} \mathbf{E}[(\hat{Y}-Y)^2] &= \mathbf{E}[((\hat{Y}-\widehat{Y})-(Y-\widehat{Y}))^2] \\ &= \mathbf{E}[(\hat{Y}-\widehat{Y})^2] - \underbrace{2\mathbf{E}[(\hat{Y}-\widehat{Y})(Y-\widehat{Y})]}_{=\,0} + \mathbf{E}[(Y-\widehat{Y})^2] \\ &= \mathbf{E}[(\hat{Y}-\widehat{Y})^2] + \mathbf{E}[(Y-\widehat{Y})^2] \end{aligned}$$

Thus,

$$\mathbf{E}[(\hat{Y}-Y)^2] \geq \mathbf{E}[(Y-\widehat{Y})^2]$$

with equality if, and only if, $\hat{Y} = \widehat{Y}$. We still have to prove that $\mathbf{E}[(\hat{Y}-\widehat{Y})(Y-\widehat{Y})] = 0$.

$\square$

Example: Multivariate Gaussian variables:

$[Y, X(1), \ldots, X(M)]^T \sim \mathcal{N}(\mu, \Sigma)$ with

- $\mu \equiv [\mu_Y, \mu_{X(1)}, \ldots, \mu_{X(M)}]^T$

- $\Sigma \equiv \begin{bmatrix} \sigma_Y^2 & (\Sigma_{XY})^T \\ \Sigma_{XY} & \Sigma_{\boldsymbol{XX}} \end{bmatrix}$

From Equation (2.66a) in [Shanmugan] it follows that

$$\boxed{\widehat{Y} = \mathbf{E}[Y|\boldsymbol{X}] = \mu_Y + (\Sigma_{\boldsymbol{X}Y})^T(\Sigma_{\boldsymbol{XX}})^{-1}(\boldsymbol{X} - \mu_{\boldsymbol{X}})}$$

Bivariate case: $M = 1, X(1) = X$

- $\Sigma_{XX} = \sigma_X^2$,

- $\Sigma_{XY} = \rho\sigma_X\sigma_Y$, where $\rho \equiv \dfrac{\Sigma_{XY}}{\sigma_X\sigma_Y}$ is the correlation coefficient of Y and X.

In this case,

$$\begin{aligned} \widehat{Y} &= \mathbf{E}[Y|X] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - \mu_X) \\ &= \underbrace{\left(\mu_Y - \frac{\rho\sigma_Y}{\sigma_X}\mu_X\right)}_{h_0} + \underbrace{\left(\frac{\rho\sigma_Y}{\sigma_X}\right)}_{h_1} X \end{aligned}$$

We can observe that $\widehat{Y}$ is linear, i.e. is the LMMSEE $\hat{Y} = \widehat{Y}$ in the bivariate case. This is also true in the general multivariate Gaussian case. In fact,

$$\boxed{\hat{Y} = \widehat{Y} \text{ if, and only if, } [Y, X(1), \ldots, X(M)]^T \text{ is a Gaussian random vector.}}$$

4.3. Time-discrete Wiener filters

- ***Problem:***

  Estimation of a WSS random sequence $Y(n)$ based on the observation of another sequence $X(n)$. Without loss of generality we assume that $\mathbf{E}[Y(n)] = \mathbf{E}[X(n)] = 0$.

  The goodness of the estimator $\hat{Y}(n)$ is described by the MSE

$$\mathbf{E}[(\hat{Y}(n) - Y(n))^2].$$

  We distinguish between two cases:

  - **Prediction:**

    $\hat{Y}(n)$ depends on one or several past observations of $X(n)$ only, i.e.

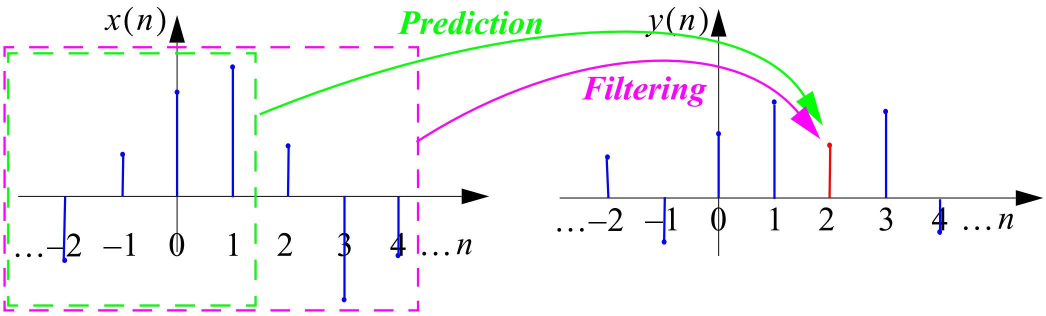
$$\hat{Y}(n) = \hat{Y}(X(n_1), X(n_2), \dots) \text{ with } n_1, n_2, \dots < n$$

  - **Filtering:**

    $\hat{Y}(n)$ depends on the present observation and/or one or many future observation(s) of $X(n)$, i.e.

$$\hat{Y}(n) = \hat{Y}(X(n_1), X(n_2), \dots) \text{ where at least one } n_i \geq n$$

    If all $n_i \leq n$, the filter is **causal** otherwise it is **noncausal**.



  Typical application: WSS signal embedded in additive white noise

$$X(n) = Y(n) + W(n) \ .$$

  where,

  - $W(n)$ is a white noise sequence,
  - $Y(n)$ is a WSS process
  - $Y(n)$ and $W(n)$ are uncorrelated.

  However, the theoretical treatment is more general as shown below.

4.3.1. Noncausal Wiener filters

- ***Linear Minimum Mean Squared Error Filter***

  We seek a linear filter

$$\hat{Y}(n) = \sum_{m=-\infty}^{\infty} h(m)X(n-m) = h(n)*X(n)$$

  which minimizes the MSE $\mathbf{E}[(\hat{Y}(n) - Y(n))^2]$.

  Such a filter exists. It is called a **Wiener filter** in honour of his inventor.

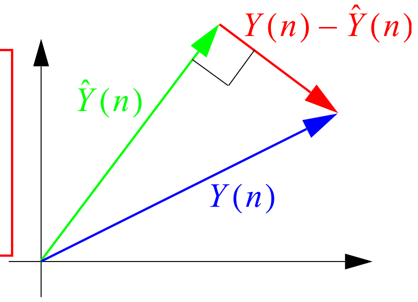
- ***Orthogonality principle (time domain):***

  The coefficients of a Wiener filter satisfy the conditions:

$$\boxed{\begin{aligned}&\mathbf{E}[(Y(n) - \hat{Y}(n))X(n-k)] = \\ &= \mathbf{E}\left[\left(Y(n) - \sum_{m=-\infty}^{\infty} h(m)X(n-m)\right)X(n-k)\right] = 0, \quad k = \ldots, -1, 0, 1, \ldots\end{aligned}}$$

  It follows from these identities (see also (4.3)) that

$$\boxed{\begin{aligned}\mathbf{E}[(Y(n) - \hat{Y}(n))\hat{Y}(n)] &= 0 \\ \mathbf{E}[(Y(n) - \hat{Y}(n))^2] &= \mathbf{E}[Y(n)^2] - \mathbf{E}[\hat{Y}(n)^2] \\ &= \mathbf{E}[(Y(n) - \hat{Y}(n))Y(n)]\end{aligned}}$$



  With the definitions

$$R_{XX}(k) \equiv \mathbf{E}[X(n)X(n+k)], \quad R_{XY}(k) \equiv \mathbf{E}[X(n)Y(n+k)]$$

  we can recast the orthogonality conditions as follows:

$$\boxed{\begin{aligned}R_{XY}(k) &= \sum_{m=-\infty}^{\infty} h(m)R_{XX}(k-m) \qquad k = \ldots, -1, 0, 1, \ldots \\ R_{XY}(k) &= h(k)*R_{XX}(k) \qquad \text{Wiener-Hopf equation}\end{aligned}}$$

- ***Orthogonality principle (frequency domain):***

$$\boxed{S_{XY}(f) = H(f)S_{XX}(f)}$$

where

$$S_{XY}(f) \equiv \mathcal{F}\{R_{XY}(k)\}$$

- ***Transfer function of the Wiener filter:***

$$\boxed{H(f) = \frac{S_{XY}(f)}{S_{XX}(f)}}$$

- ***MSE of the Wiener filter (time-domain formulation):***

$$\boxed{\mathbf{E}[(Y(n) - \hat{Y}(n))^2] = \sigma_Y^2 - \sum_{m=-\infty}^{\infty} h(m)R_{XY}(m)}$$

Proof:

$$\mathbf{E}[(Y(n) - \hat{Y}(n))^2] = \mathbf{E}[Y(n)^2] - \mathbf{E}[\hat{Y}(n)Y(n)]$$

$\square$

- ***MSE of the Wiener filter (frequency-domain formulation):***
  We can rewrite the above identity as:

$$\mathbf{E}[(Y(n) - \hat{Y}(n))^2] = R_{YY}(0) - \sum_{m=-\infty}^{\infty} h(m)R_{YX}(-m)$$

$\mathbf{E}[(Y(n)-\hat{Y}(n))^2]$ is the value $p(0)$ of the sequence

$$p(k) = R_{YY}(k) - \sum_{m=-\infty}^{\infty} h(m)R_{YX}(k-m)$$

$$= R_{YY}(k) - h(k) * R_{YX}(k)$$

$$P(f) = S_{YY}(f) - H(f)S_{YX}(f) = S_{YY}(f) - \frac{|S_{XY}(f)|^2}{S_{XX}(f)}$$

Hence,

$$\mathbf{E}[(Y(n)-\hat{Y}(n))^2] = p(0) = \int_{-1/2}^{1/2} P(f)df$$

$$\boxed{\mathbf{E}[(Y(n)-\hat{Y}(n))^2] = \int_{-1/2}^{1/2} \left[S_{YY}(f) - \frac{|S_{XY}(f)|^2}{S_{XX}(f)}\right]df}$$

4.3.2. Causal Wiener filters

A. $X(n)$ is a white noise.

We first assume that $X(n)$ is a white noise with unit variance, i.e. $\mathbf{E}[X(n)X(n+k)] = \delta(k)$.

- ***Derivation of the causal Wiener filter from the noncausal Wiener filter:***

  Let us consider the noncausal Wiener filter

$$\hat{Y}(n) = \sum_{m=-\infty}^{\infty} h(m)X(n-m)$$

whose transfer function is given by

$$H(f) = \frac{S_{XY}(f)}{S_{XX}(f)} = S_{XY}(f).$$

Then, the causal Wiener filter $\hat{Y}_c(n)$ results by cancelling the noncausal part of the non-causal Wiener filter:

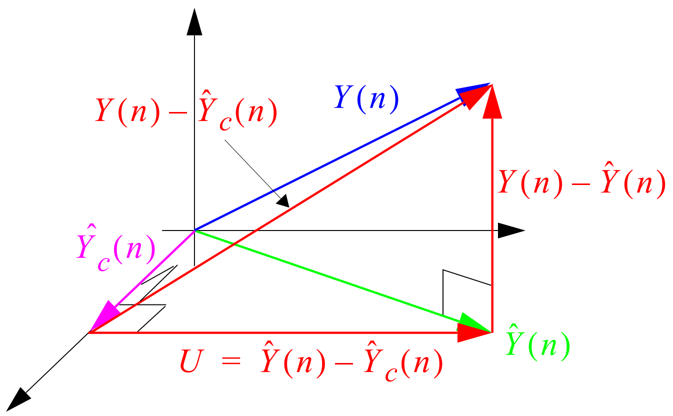
$$\boxed{\hat{Y}_c(n) = \sum_{m=0}^{\infty} h(m)X(n-m)}$$

Sketch of the proof:

$\hat{Y}(n)$ can be written as

$$\hat{Y}(n) = \underbrace{\sum_{m=-\infty}^{-1} h(m)X(n-m)}_{\equiv U} + \underbrace{\sum_{m=0}^{\infty} h(m)X(n-m)}_{V \equiv \hat{Y}_c(n)}$$

Because $X(n)$ is a white noise, the causal part $V = \hat{Y}_c(n)$ and the noncausal part $U = \hat{Y}(n) - \hat{Y}_c(n)$ of $\hat{Y}(n)$ are orthogonal. It follows from this property



that $\hat{Y}_c(n)$ and $Y(n)$ are orthogonal, i.e. that $\hat{Y}_c(n)$ minimizes the MSE within the class of linear causal estimators.

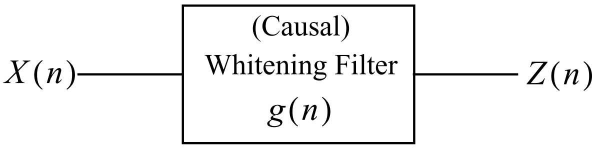
$\square$

B. $X(n)$ **is an arbitrary WSS process whose spectrum satisfies the Paley-Wiener condition.**

Usually, the above truncation procedure to obtain $\hat{Y}_c(n)$ does not apply because U and V are correlated and therefore not orthogonal in the general case.

- ***Causal whitening filter:***

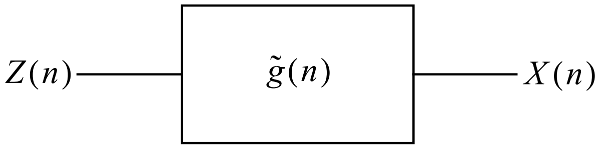
  However, we can show (see the Spectral Decomposition Theorem below) that provided $S_{XX}(f)$ satisfies the Paley-Wiener condition (see below) then $X(n)$ can be **converted into an equivalent white noise sequence** $Z(n)$ **with unit variance** by filtering it with an appropriate causal filter $g(n)$,



$$\mathbf{E}[Z(n)Z(n+k)] = \delta(k)$$

This operation is called **whitening** and the filter $g(n)$ is called a **whitening filter**.

equivalent $\equiv$ there exists another causal filter $\tilde{g}(n)$ so that $X(n) = \tilde{g}(n)*Z(n)$:



Notice that if

$$G(f) \equiv \mathcal{F}\{g(n)\}$$

$$\tilde{G}(f) \equiv \mathcal{F}\{\tilde{g}(n)\}$$

then

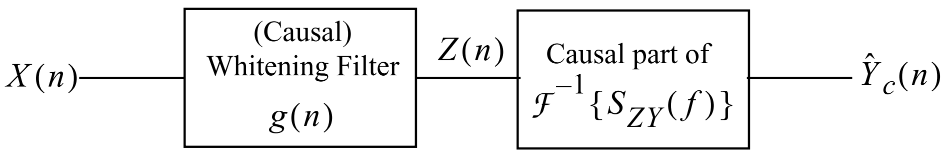
$$|G(f)|^2 = S_{XX}(f)^{-1}$$

$$|\tilde{G}(f)|^2 = S_{XX}(f) \tag{4.6}$$

We shall see that a whitening filter exists such that

$$\tilde{G}(f) = G(f)^{-1}$$

- ***Causal Wiener filter***
  Making use of the result in Part A, the block diagram of the causal Wiener filter is



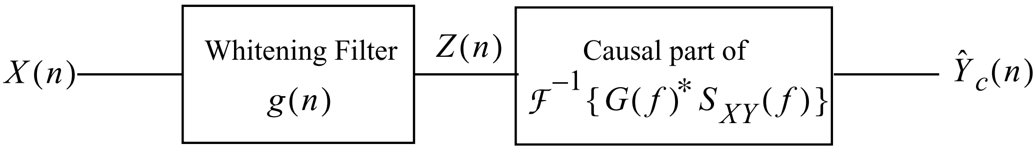
$S_{ZY}(f)$ is obtained from $S_{XY}(f)$ according to

$$\boxed{S_{ZY}(f) = G(f)^* S_{XY}(f)}$$

Proof:

$\square$

Hence, the block diagram of the causal Wiener filter is:



- ***Spectral Decomposition Theorem:***

Let $S_{XX}(f)$ satisfies the so-called **Paley-Wiener condition**:

$$\boxed{\int_{-1/2}^{1/2} \log(S_{XX}(f))df > -\infty}$$

Then $S_{XX}(f)$ can be written as

$$S_{XX}(f) = G(f)^{+} G(f)^{-}$$

with $G(f)^{+}$ and $G(f)^{-}$ satisfying

$$\boxed{\left|G(f)^{+}\right|^2 = \left|G(f)^{-}\right|^2 = S_{XX}(f)}.$$

Moreover, the sequences

$$\boxed{\begin{aligned} g(n)^{+} &\equiv \mathcal{F}^{-1}\{G(f)^{+}\} \\ g(n)^{-} &\equiv \mathcal{F}^{-1}\{G(f)^{-}\} \\ g^{-1}(n)^{+} &\equiv \mathcal{F}^{-1}\{1/G(f)^{+}\} \\ g^{-1}(n)^{-} &\equiv \mathcal{F}^{-1}\{1/G(f)^{-}\} \end{aligned}}$$

satisfy

$$\boxed{\begin{array}{llll} g(n)^{+} = g^{-1}(n)^{+} = 0 & n < 0 & \text{Causal sequences} \\ g(n)^{-} = g^{-1}(n)^{-} = 0 & n > 0 & \text{Anticausal sequences} \end{array}}$$

- ***Whitening filter (cont'd):***
  The sought whitening filter used to obtained $Z(n)$ is
$$\boxed{g(n) = g^{-1}(n)^{+}}$$
  and
$$\tilde{g}(n) = g(n)^{+}.$$
  It can be easily verified that both sequences satisfy the identities in (4.6).

4.3.3. Finite Wiener filters

- ***Finite linear filter:***
$$\hat{Y}(n) = \sum_{m=-M_1}^{M_2} h(m)X(n-m)$$
- ***Wiener-Hopf equation:***
  By applying the orthogonality principle we obtain the Wiener-Hopf system of equations:
$$\boxed{\Sigma_{XY} = \Sigma_{XX}\boldsymbol{h}}$$
  where
$$\boldsymbol{h} \equiv [h(-M_1), \ldots, h(M_2)]^T$$
  and
$$\Sigma_{XY} \equiv [R_{XY}(-M_1), \ldots, R_{XY}(M_2)]^T$$
$$\Sigma_{XX} \equiv$$
$$\equiv \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & R_{XX}(2) & \ldots & R_{XX}(M_1+M_2) \\ R_{XX}(1) & R_{XX}(0) & R_{XX}(1) & \ldots & R_{XX}(M_1+M_2-1) \\ R_{XX}(2) & R_{XX}(1) & R_{XX}(0) & \ldots & R_{XX}(M_1+M_2-2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ R_{XX}(M_1+M_2) & R_{XX}(M_1+M_2-1) & R_{XX}(M_1+M_2-2) & \ldots & R_{XX}(0) \end{bmatrix}$$
  Coefficient vector of the finite Wiener filter:
$$\boxed{\boldsymbol{h} = (\Sigma_{XX})^{-1}\Sigma_{XY}}$$

provided Σ_{XX} is invertible.

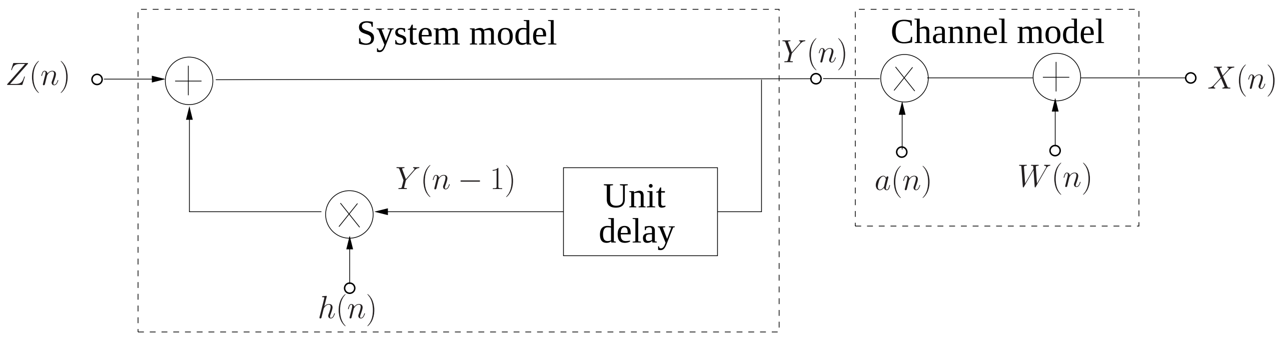
5 Kalman filters

5.1 Scalar Kalman filter

Signal model and Kalman filter to be derived in the sequel:

- **Complete signal model:**

$$\boxed{\begin{aligned} Y(n) &= h(n)Y(n-1) + Z(n), n = 1, 2, \ldots \\ X(n) &= a(n)Y(n) + W(n) \quad , n = 1, 2, \ldots \end{aligned}}$$



- **Scalar Kalman filter**

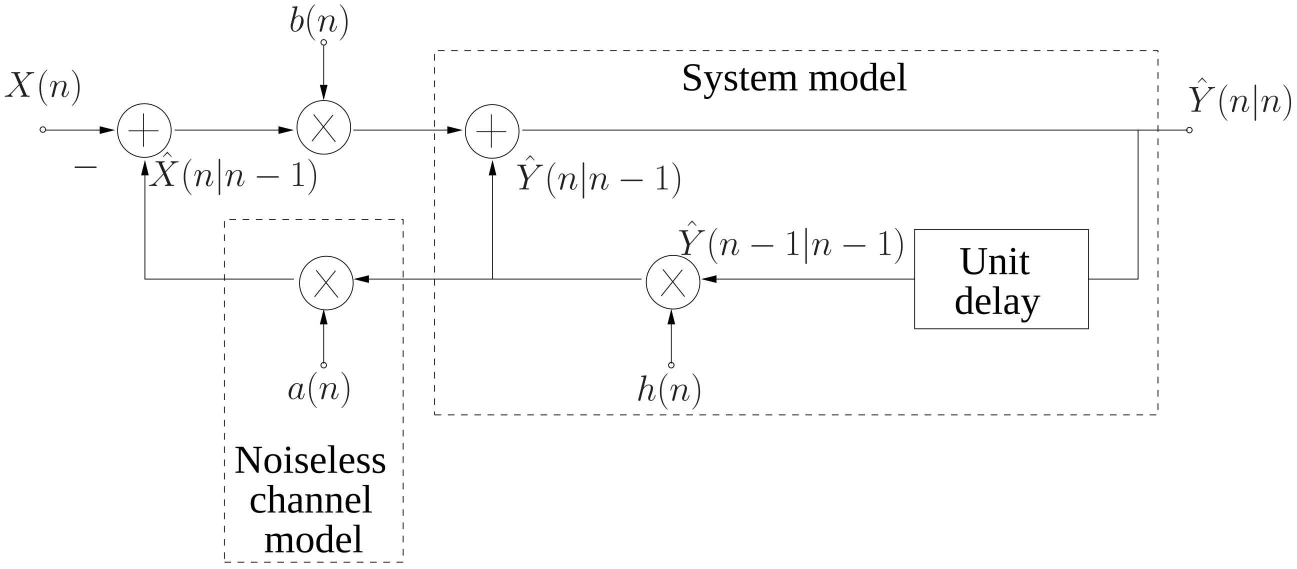
Prediction step:

$$\boxed{\begin{aligned} \hat{Y}(n+1 \mid n) &= h(n+1)\hat{Y}(n \mid n) \\ R(n+1 \mid n) &= h^2(n+1)R(n \mid n) + \sigma_{ZZ}^2(n+1) \\ \hat{X}(n+1 \mid n) &= a(n+1)\hat{Y}(n+1 \mid n) \end{aligned}}$$

Updating step:

$$\boxed{\begin{aligned} \hat{Y}(n+1 \mid n+1) &= \hat{Y}(n+1 \mid n) + b(n+1)[X(n+1) - \hat{X}(n+1 \mid n)] \\ R(n+1 \mid n+1) &= [1 - b(n+1)a(n+1)]R(n+1 \mid n) \\ &\text{with} \\ b(n+1) &\equiv \frac{a(n+1)R(n+1 \mid n)}{a(n+1)^2 R(n+1 \mid n) + \sigma_{WW}^2(n+1)} \end{aligned}}$$

Block diagram of the scalar Kalman filter:



Initialization:

$$\boxed{\begin{aligned} \hat{Y}(0 \mid 0) &= \mu_{Y(0)} \\ R(0 \mid 0) &= \sigma_{Y(0)}^2 \end{aligned}}$$

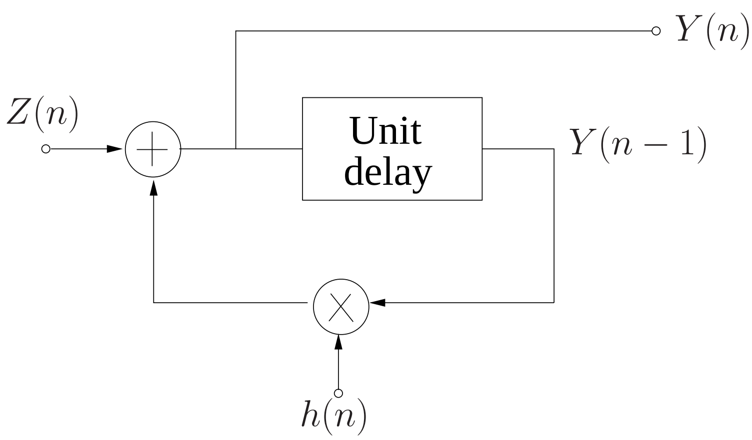
5.1.1 Signal model

- **System model**

  $\{Y(n)\}$ is an unobservable sequence which is described by the following ***state or system equation***:

$$\boxed{Y(n) = h(n)Y(n-1) + Z(n), n = 1, 2, \ldots \qquad (5.1)}$$

Block Diagram Representation of (5.1)



Initialization:

$Y(0)$ is a random variable whose expectation $\mu_{Y(0)} \equiv E[Y(0)]$ and variance $\sigma^2_{Y(0)} \equiv E[(Y(0) - \mu_{Y(0)})^2]$ are known.

Property of the driving process/noise $\{Z(n)\}$***:***

$\{Z(n)\}$ is an uncorrelated noise with a possibly time-varying variance (non-stationary correlated noise) :

- $E[Z(n)] = 0$
- $E[Z(n)Z(n+k)] = \sigma^2_{ZZ}(n)\delta(k)$

Property of the feedback coefficients $\{h(n)\}$:

$\{h(n)\}$ is a known deterministic sequence.

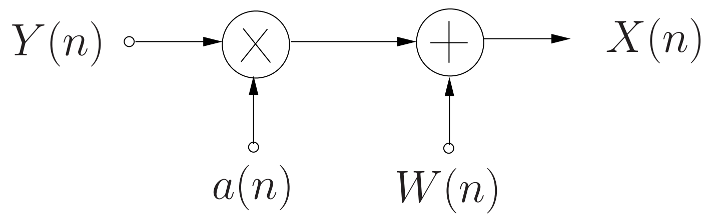
Remark: Provided $\{h(n)\}$ and $\{\sigma^2_{ZZ}(n)\}$ are constant and $\{Z(n)\}$ is a Gaussian random process, then $\{Y(n)\}$ is an AR(1) process.

- **Observation (or channel) model**

  The observable sequence $X(n)$ is given by

$$\boxed{X(n) = a(n)Y(n) + W(n) \quad , n = 1, 2, \ldots \qquad (5.2)}$$

Block diagram representation of (5.2)



Property of the weighting sequence $\{a(n)\}$:

$\{a(n)\}$ is a known deterministic sequence.

Property of the noise $\{W(n)\}$:

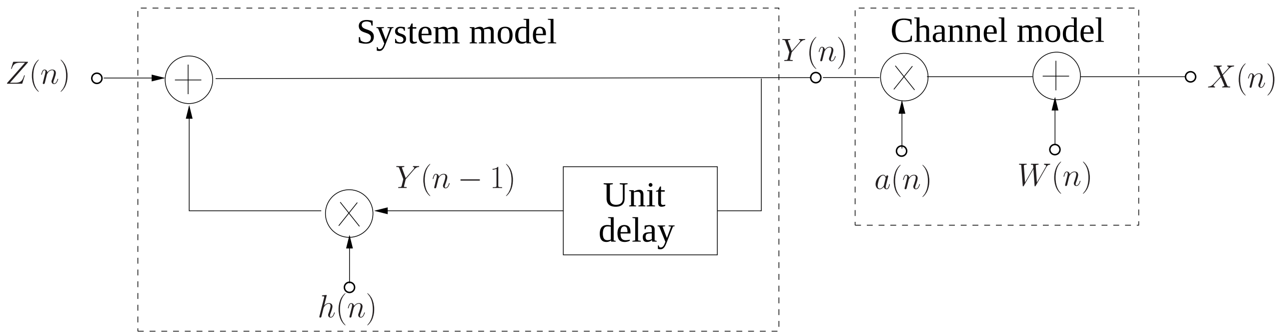
$\{W(n)\}$ is a non-stationary uncorrelated noise:

- $E[W(n)] = 0$
- $E[W(n)W(n+k)] = \sigma^2_{WW}(n)\delta(k)$

- **Additional "weak independence" assumptions**

  $Y(0)$, $\{Z(n)\}$, and $\{W(n)\}$ are uncorrelated.

- **Block diagram of the complete signal model**



5.1.2 Recursive implementation of the LMMSEE

- **Objective:**

  To find a ***recursive implementation***[1] of the LMMSEE of $Y(n)$ based on the observation of $X(1), \ldots, X(n)$.

- **Recursive implementation:**

  We need the following definitions:

  – $\hat{Y}(n \mid n) \equiv$ LMMSEE of $Y(n)$ based on the observation of $X(1), \ldots, X(n)$
  Estimation of $Y(n)$.

  – $\hat{Y}(n+1 \mid n) \equiv$ LMMSEE of $Y(n+1)$ based on the observation of $X(1), \ldots, X(n)$
  One-step prediction of $Y(n+1)$ ***at time*** n.

  – $\hat{X}(n+1 \mid n) \equiv$ LMMSEE of $X(n+1)$ based on the observation of $X(1), \ldots, X(n)$
  One-step prediction of $X(n+1)$.

  Recursive implementation of the LMMSEE of $Y(n)$:

$$\underbrace{\hat{Y}(n+1 \mid n+1)}_{\substack{\text{Estimation}\\ \text{at time } n+1}} \equiv \mathcal{LF}(\underbrace{\hat{Y}(n \mid n)}_{\substack{\text{Estimation}\\ \text{at time } n}}, \underbrace{X(n+1)}_{\substack{\text{Observation}\\ \text{at time } n+1}})$$

  where $\mathcal{LF}$ denotes a linear function to be found.

[1] See Section 5.3 for an example of a recursive estimator.

- **We shall show:**

  1. Such a recursive implementation of the LMMSEE exists. It is called the ***Kalman Filter***.

  2. The recursion is split into two steps:

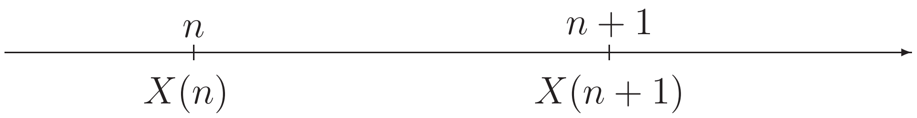
     – Step 1: ***One-step prediction***:

$$P : \hat{Y}(n \mid n) \underset{P}{\longrightarrow} \hat{Y}(n+1 \mid n)$$

     – Step 2: ***Updating***:

$$U : \hat{Y}(n+1 \mid n) \xrightarrow[U]{X(n+1)} \hat{Y}(n+1 \mid n+1)$$

Temporal evolution of the recursive estimation procedure in the Kalman filter:



$$\hat{Y}(n \mid n-1) \xrightarrow[U]{X(n)} \hat{Y}(n \mid n) \underset{P}{\longrightarrow} \hat{Y}(n+1 \mid n) \xrightarrow[U]{X(n+1)} \hat{Y}(n+1 \mid n+1) \underset{P}{\longrightarrow} \ldots$$

  3. The mean-squared estimation error $E[(Y(n) - \hat{Y}(n \mid n))^2]$ can also be computed recursively.

     We shall need the following definitions:

     - $R(n|n) \equiv E[(Y(n) - \hat{Y}(n|n))^2]$
       $\equiv$ mean-squared estimation error at time n

     - $R(n+1|n) \equiv E[(Y(n+1) - \hat{Y}(n+1|n))^2]$
       $\equiv$ mean-squared one-step prediction error at time n

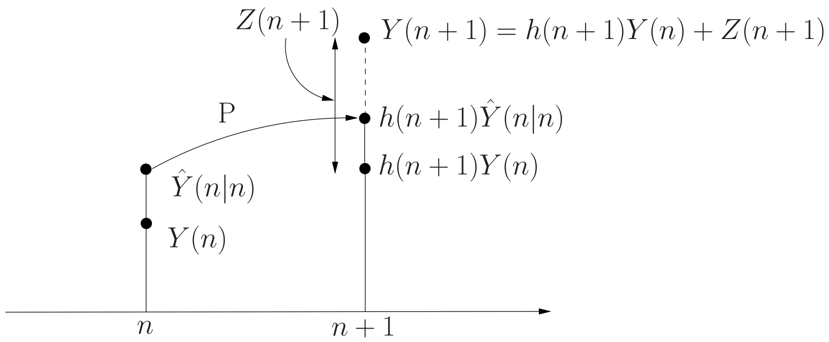
5.1.3 Derivation of the equations of the Kalman filter

- **Prediction step:**

$$\boxed{\begin{aligned} \hat{Y}(n+1 \mid n) &= h(n+1)\hat{Y}(n \mid n) && (5.3) \\ R(n+1 \mid n) &= h^2(n+1)R(n \mid n) + \sigma_{ZZ}^2(n+1) && (5.4) \end{aligned}}$$

Proof of (5.3):

(5.3) follows from the linearity property of the expectation.



Let us show that (5.3) satisfies the orthogonality principle (OP) and therefore is the LMMSEE:

Let $m = 1, \ldots, n$:

$$\begin{aligned} & E[(\qquad Y(n+1) \qquad - \qquad \hat{Y}(n+1 \mid n))X(m)] \\ &= E[(\overbrace{h(n+1)Y(n) + Z(n+1)} - \overbrace{h(n+1)\hat{Y}(n \mid n))}\, X(m)] \\ &= h(n+1)\; \underbrace{E[(Y(n) - \hat{Y}(n \mid n))X(m)]}_{\substack{=0 \\ \text{OP for } \hat{Y}(n|n)}} + \underbrace{E[Z(n+1)X(m)]}_{\substack{=0 \\ Z(n+1) \text{ and } X(n) \\ \text{are uncorrelated}}} \\ &= 0 \quad \surd \end{aligned}$$

Proof of (5.4):

$$
\begin{aligned}
R(n+1 \mid n) &= E[(Y(n+1) - \hat{Y}(n+1 \mid n))^2] \\
&= E[(\overbrace{h(n+1)Y(n) + Z(n+1)} - \overbrace{h(n+1)\hat{Y}(n \mid n)})^2] \\
&= E[(\underbrace{h(n+1)[Y(n) - \hat{Y}(n \mid n)]} + \underbrace{Z(n+1)})^2]
\end{aligned}
$$

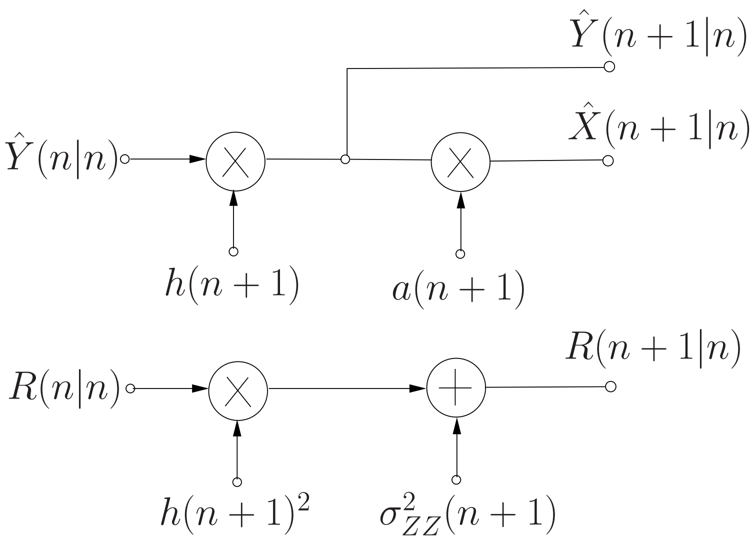
These two random variables are uncorrelated

$$
\begin{aligned}
&= h(n+1)^2 E[(Y(n) - \hat{Y}(n \mid n))^2] + E[Z(n+1)^2] \\
&= h(n+1)^2 R(n \mid n) + \sigma_{ZZ}^2(n+1)
\end{aligned}
$$

With the same argument as that used for the proof of (5.3) we show that

$$\boxed{\hat{X}(n+1 \mid n) = a(n+1)\hat{Y}(n+1 \mid n)}$$

Block diagram of the prediction step:



- **Updating step:**

$$\hat{Y}(n+1 \mid n+1) = \hat{Y}(n+1 \mid n) + b(n+1)[X(n+1) - \hat{X}(n+1 \mid n)] \quad (5.5)$$

$$R(n+1 \mid n+1) = [1 - b(n+1)a(n+1)]R(n+1 \mid n) \quad (5.6)$$

with

$$b(n+1) \equiv \frac{a(n+1)R(n+1 \mid n)}{a(n+1)^2 R(n+1 \mid n) + \sigma_{WW}^2(n+1)}$$

Interpretation of (5.5):

$$\hat{Y}(n+1 \mid n+1) = \underbrace{\hat{Y}(n+1 \mid n)}_{\substack{\text{One-step}\\ \text{prediction}\\ \text{of } Y(n+1)}} \underbrace{+\, b(n+1)\,\underbrace{[\,\overbrace{X(n+1)}^{\substack{\text{New}\\ \text{observation}}} - \overbrace{\hat{X}(n+1 \mid n)}^{\substack{\text{One-step}\\ \text{prediction}\\ \text{of } X(n+1)}}\,]}_{\substack{\text{Residual error}\\ \text{of } \hat{X}(n+1 \mid n)}}}_{\text{Correction factor}}$$

Kalman gain:

The coefficient $b(n)$ is called the ***Kalman gain*** of the filter.

Proof of (5.5) :

We seek an updating equation given by (5.5) and determine $b(n+1)$ so that (5.5) satisfies the orthogonality principle.

1st case: $m = 1, \ldots, n$

$$E[(Y(n+1) - \hat{Y}(n+1 \mid n+1))X(m)]$$

$$= \underbrace{E[(Y(n+1) - \hat{Y}(n+1 \mid n))X(m)]}_{\substack{=0 \\ \text{OP for } \hat{Y}(n+1 \mid n)}} - b(n+1)\underbrace{E[(X(n+1) - \hat{X}(n+1 \mid n))X(m)]}_{\substack{=0 \\ \text{OP for } \hat{X}(n+1 \mid n)}}$$

$$= 0 \quad \surd$$

2nd case: $m = n+1$

$$\begin{aligned} &E[(Y(n+1) - \hat{Y}(n+1 \mid n+1))X(n+1)] \\ &\qquad = E[(Y(n+1) - \hat{Y}(n+1 \mid n))X(n+1)] \\ &\qquad\qquad - b(n+1)E[(X(n+1) - \hat{X}(n+1 \mid n))X(n+1)] \end{aligned}$$

We determine $b(n+1)$ such that the above expression vanishes:

$$b(n+1) = \frac{E[(Y(n+1) - \hat{Y}(n+1 \mid n))X(n+1)]}{E[(X(n+1) - \hat{X}(n+1 \mid n))X(n+1)]} = \frac{I}{II}$$

→ ***Computation of*** I**:**

$$
\begin{aligned}
I &= E[(Y(n+1) - \hat{Y}(n+1 \mid n))\overbrace{(a(n+1)Y(n+1) + W(n+1))}^{X(n+1)}] \\
&= a(n+1)\,\underbrace{E[(Y(n+1) - \hat{Y}(n+1 \mid n))Y(n+1)]}_{\substack{= E[(Y(n+1) - \hat{Y}(n+1 \mid n))^2] \\ \uparrow OP \Rightarrow E[(Y - \hat{Y})Y] = E[(Y - \hat{Y})^2]}} \\
&\qquad + \underbrace{E[(Y(n+1) - \hat{Y}(n+1 \mid n))W(n+1)]}_{=0}
\end{aligned}
$$

$Y(n+1) - \hat{Y}(n+1 \mid n)$ and $W(n+1)$ are uncorrelated because $Y(n+1) - \hat{Y}(n+1 \mid n)$ depends on $Y(0), Y(1), \ldots, Y(n), Y(n+1), X(1), \ldots, X(n)$

$$
\begin{aligned}
&= a(n+1)E[(Y(n+1) - \hat{Y}(n+1 \mid n))^2] \\
I &= a(n+1)R(n+1 \mid n)
\end{aligned}
$$

→ ***Computation of*** II**:**

$$
\begin{aligned}
II &= E[(X(n+1) - \hat{X}(n+1 \mid n))X(n+1)] \\
&= E[(X(n+1) - \hat{X}(n+1 \mid n))^2] \quad \leftarrow OP \\
&= E[(\overbrace{a(n+1)Y(n+1) + W(n+1)}^{X(n+1)} - \overbrace{a(n+1)\hat{Y}(n+1 \mid n)}^{\hat{X}(n+1 \mid n)})^2] \\
&= E[(a(n+1)\,\underbrace{(Y(n+1) - \hat{Y}(n+1 \mid n))} + \underbrace{W(n+1)})^2]
\end{aligned}
$$

these random variables are uncorrelated

$$
\begin{aligned}
&= a(n+1)^2 E[(Y(n+1) - \hat{Y}(n+1 \mid n))^2] + E[W(n+1)^2] \\
II &= a(n+1)^2 R(n+1 \mid n) + \sigma^2_{WW}(n+1)
\end{aligned}
$$

Proof of (5.6):

$$
\begin{aligned}
R(n+1 \mid n+1) &= E[(Y(n+1) - \hat{Y}(n+1 \mid n+1))^2] \\
&\overset{OP}{=} E[(Y(n+1) - \hat{Y}(n+1 \mid n+1))Y(n+1)] \\
&= \underbrace{E[(Y(n+1) - \hat{Y}(n+1 \mid n))Y(n+1)]}_{\bigcirc} \\
&\quad - b(n+1)\,\underbrace{E[(X(n+1) - \hat{X}(n+1 \mid n))Y(n+1)]}_{\bigstar}
\end{aligned}
$$

Computation of $\bigcirc$:

$$
\begin{aligned}
\bigcirc &\overset{OP}{=} E[(Y(n+1) - \hat{Y}(n+1 \mid n))^2] \\
&= R(n+1 \mid n)
\end{aligned}
$$

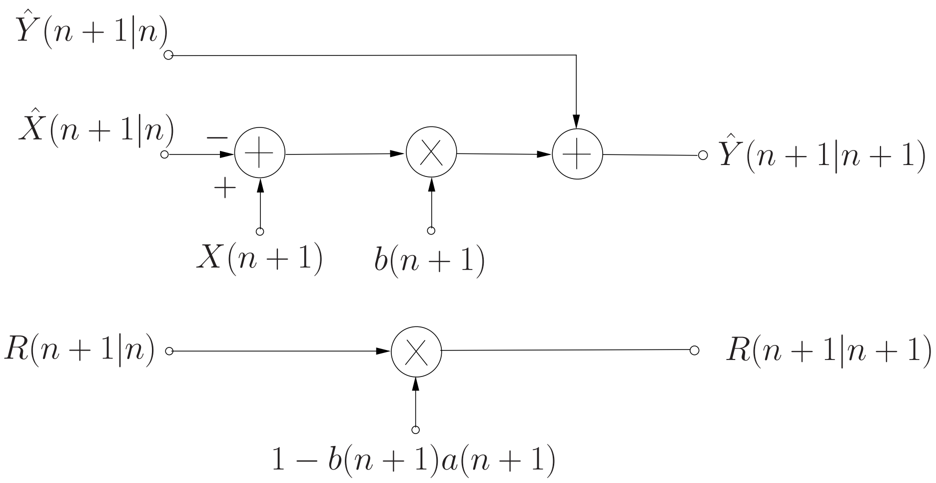
Computation of $\bigstar$:

$$
\begin{aligned}
\bigstar &= E\Big[\Big(a(n+1)Y(n+1) + W(n+1) - a(n+1)\hat{Y}(n+1 \mid n)\Big)Y(n+1)\Big] \\
&= a(n+1)E[(Y(n+1) - \hat{Y}(n+1 \mid n))Y(n+1)] + \underbrace{E[W(n+1)Y(n+1)]}_{=0} \\
&\overset{OP}{=} a(n+1)E[(Y(n+1) - \hat{Y}(n+1 \mid n))^2] \\
&= a(n+1)R(n+1 \mid n)
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
R(n+1 \mid n+1) &= R(n+1 \mid n) - b(n+1)a(n+1)R(n+1 \mid n) \\
&= [1 - b(n+1)a(n+1)]R(n+1 \mid n)
\end{aligned}
$$

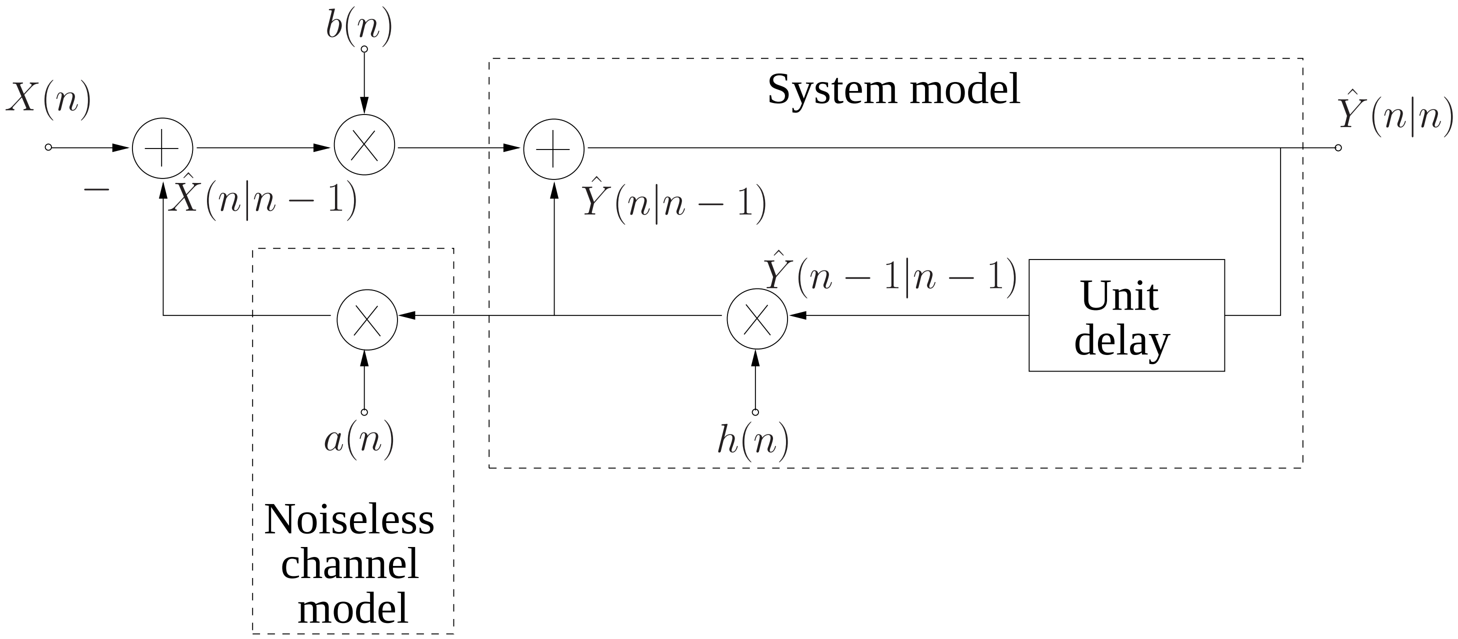
Block diagram of the updating step:



- **Initialization:**

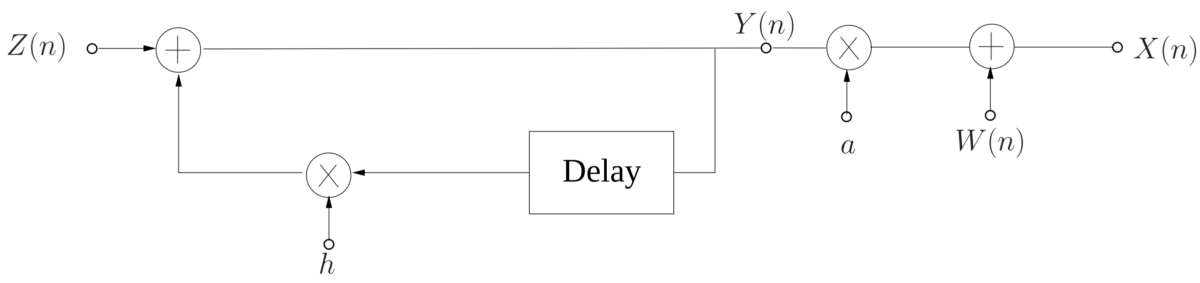
$$\boxed{\begin{aligned} \hat{Y}(0 \mid 0) &= \mu_{Y(0)} \\ R(0 \mid 0) &= \sigma^2_{Y(0)} \end{aligned}}$$

Block diagram of the Scalar Kalman filter:



5.1.4 Steady-state Kalman filter when the system and channel models are time-invariant

We consider the time-invariant system and channel models as depicted below:



The system-driving process $Z(n)$ and the channel noise $W(n)$ are uncorrelated wide-sense stationary process:

- $E[Z(n)Z(n+k)] = \sigma_{ZZ}^2(n)\delta(k)$
- $E[W(n)W(n+k)] = \sigma_{WW}^2\delta(k)$

Equations of the Kalman filter estimating $Y(n)$:

$$R(n+1 \mid n) = h^2 R(n \mid n) + \sigma_{ZZ}^2$$

$$b(n+1) = \frac{aR(n+1 \mid n)}{a^2 R(n+1 \mid n) + \sigma_{WW}^2}$$

$$R(n+1 \mid n+1) = [1 - ab(n+1)]R(n+1 \mid n)$$

For $n \to \infty$ the three sequences $\{R(n+1 \mid n)\}$, $\{b(n)\}$, and $\{R(n+1 \mid n+1)\}$ converge, i.e.

$$\left.\begin{aligned} R(n+1 \mid n) &\to R_{p\infty} \\ b(n) &\to b_\infty \\ R(n+1 \mid n+1) &\to R_\infty \end{aligned}\right. \qquad\qquad n \to \infty$$

The Kalman filter converges to its ***steady state***.

The above limits can be calculated by inserting them into the equations of the Kalman filter:

$$R_{p\infty} = h^2 R_\infty + \sigma_{ZZ}^2 \tag{5.7}$$

$$b_\infty = \frac{aR_{p\infty}}{a^2 R_{p\infty} + \sigma_{WW}^2} \tag{5.8}$$

$$R_\infty = [1 - ab_\infty] R_{p\infty} \tag{5.9}$$

Inserting (5.8) into (5.9), we obtain

$$R_\infty = \left[1 - \frac{a^2 R_{p\,\infty}}{a^2 R_{p\infty} + \sigma_{WW}^2}\right] R_{p\infty}$$

$$= \frac{\sigma_{WW}^2 R_{p\,\infty}}{a^2 R_{p\,\infty} + \sigma_{WW}^2}$$

Substituting (5.7) into the last expression yields the so-called steady-state Ricatti equation

$$R_\infty = \frac{\sigma_{WW}^2[h^2 R_\infty + \sigma_{ZZ}^2]}{a^2[h^2 R_\infty + \sigma_{ZZ}^2] + \sigma_{WW}^2}$$

The Ricatti equation is a quadratic equation that can be solved numerically. e.g. by using Newton's method.

Then, $R_{p\infty}$ and b_∞ follow by inserting the numerical solution for R_∞ into (5.7) and (5.8), respectively.

Example:

The steady-state solutions for the model with parameter setting

- $h = 0.9$
- $a = 1$
- $\sigma_{ZZ} = \sigma_{WW} = 1$

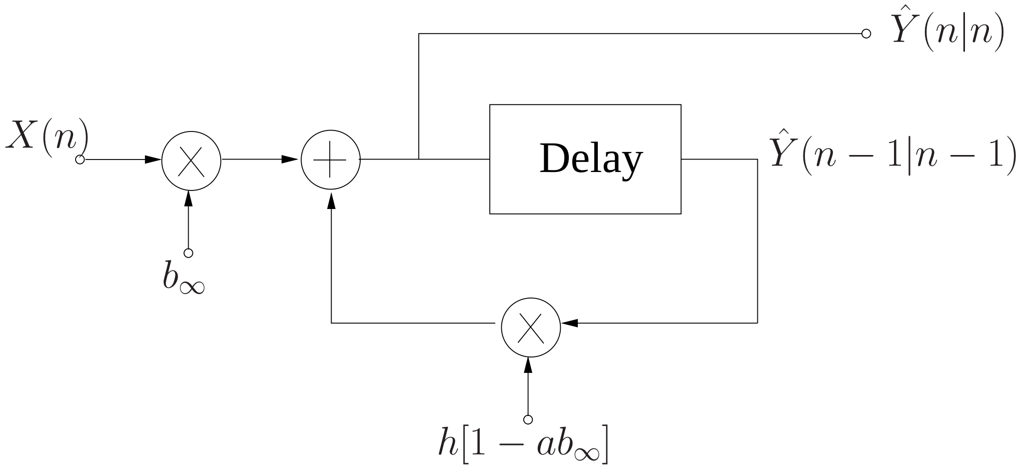
read

- $R_\infty = 0.5974$
- $R_{p\infty} = 1.4839$
- $b_\infty = 0.5974$

Input-output relationship of the steady-state Kalman filter:

$$\begin{aligned}\hat{Y}(n \mid n) &= h\hat{Y}(n-1 \mid n-1) + b_\infty[X(n) - ah\hat{Y}(n-1 \mid n-1)] \\ &= b_\infty X(n) + h[1 - ab_\infty]\hat{Y}(n-1 \mid n-1)\end{aligned}$$

Block-diagram of the steady-state Kalman filter



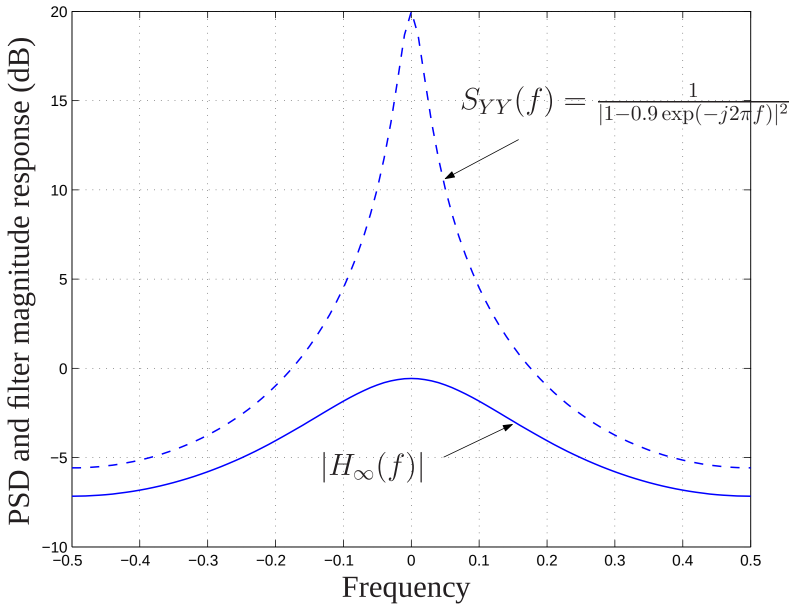
The steady-state Kalman filter is an infinite impulse response (IIR) filter with transfer function

$$H_\infty(f) = \frac{b_\infty}{1 - h[1 - ab_\infty]\exp(-j2\pi f)}$$

$$H_\infty(z) = \frac{b_\infty}{1 - h[1 - ab_\infty]z^{-1}}$$

Example (cont'd):

$$H_\infty(f) = \frac{0.5974}{1 - 0.3623 \cdot \exp(-j2\pi f)}$$



Comment:

The steady-state Kalman filter calculates the LMMSEE of $Y(n)$ based on the observation of the sequence $\{X(n)\}$ in the time window $[n, n-1, n-2, \ldots]$

Hence, the steady-state Kalman filter implements the ***Causal Wiener filter***.

5.2 Vector Kalman Filter

5.2.1 Signal Model

- **System model**

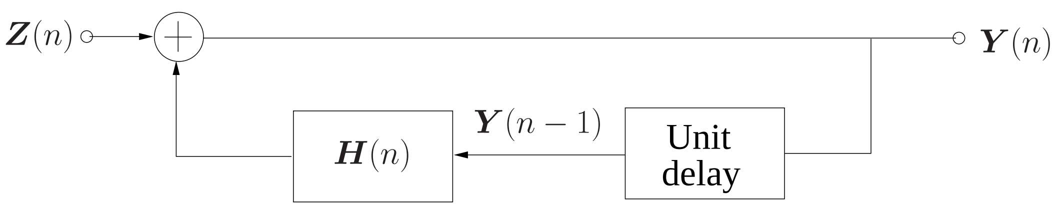
$$\boxed{\mathbf{Y}(n) = \mathbf{H}(n)\mathbf{Y}(n-1) + \mathbf{Z}(n), \qquad n = 1, 2, \ldots \quad (5.10)}$$

where:

- $\mathbf{Y}(n) = [Y_1(n), \ldots, Y_r(n)]^{\mathrm{T}}$: r-dimensional (-D) random vector.
- $\{\mathbf{Z}(n)\}$: r-D non-stationary uncorrelated noise vector:
  - $E[\mathbf{Z}(n)] = \mathbf{0}$
  - $\sum_{\mathbf{Z}(n)\mathbf{Z}(n+k)} = \mathbf{Q}_{\mathbf{Z}}(n)\delta(k)$
- $\{\mathbf{H}(n)\}$: sequence of known $r \times r$ matrices.

See the example discussed in Section 5.4.

Block diagram:



Initialization

$\mathbf{Y}(0)$ is a random vector specified by its expectation $\mu_{\mathbf{Y}(0)}$ and covariance matrix $\Sigma_{\mathbf{Y}(0)\mathbf{Y}(0)}$.

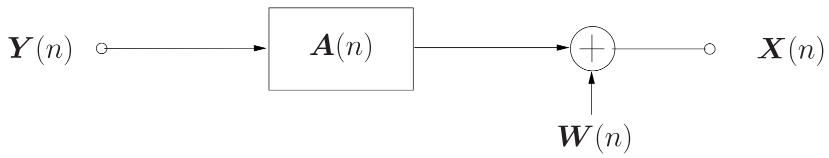
- **Observation Model**

$$\boxed{\mathbf{X}(n) = \mathbf{A}(n)\mathbf{Y}(n) + \mathbf{W}(n), n = 1, 2, \ldots \qquad (5.11)}$$

where:

- $\mathbf{X}(n) = [X_1(n), \ldots, X_s(n)]^{\mathrm{T}}$: s-D random vector.
- $\{\mathbf{W}(n)\}$: s-D non-stationary uncorrelated noise vector with autocovariance
  $\Sigma_{\mathbf{W}(n)\mathbf{W}(n+k)} = \mathbf{Q}_{\mathbf{W}(n)}\delta(k)$
- $\{\mathbf{A}(n)\}$: sequence of known $s \times r$ matrices.

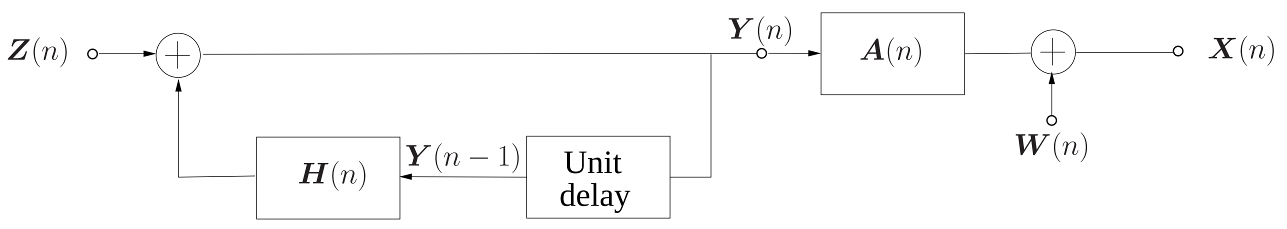
Block Diagram:



- **Additional independence assumption**

$\mathbf{Y}(0)$, $\{\mathbf{Z}(n)\}$, and $\{\mathbf{W}(n)\}$ are uncorrelated.

- **Complete Signal Model**



5.2.2 Equation of the vector Kalman filter

Let us define

- $\hat{\mathbf{Y}}(n \mid n) \quad \equiv \quad$ LMMSEE of $\quad \mathbf{Y}(n) \quad$ based on $\quad \mathbf{X}(1), \ldots, \mathbf{X}(n)$
- $\hat{\mathbf{Y}}(n+1 \mid n) \quad \equiv \quad$ LMMSEE of $\quad \mathbf{Y}(n+1) \quad$ based on $\quad \mathbf{X}(1), \ldots, \mathbf{X}(n)$
- $\hat{\mathbf{X}}(n+1 \mid n) \quad \equiv \quad$ LMMSEE of $\quad \mathbf{X}(n+1) \quad$ based on $\quad \mathbf{X}(1), \ldots, \mathbf{X}(n)$
- $\mathbf{R}(n \mid n) \quad \equiv \quad E[(\mathbf{Y}(n) - \hat{\mathbf{Y}}(n \mid n))(\mathbf{Y}(n) - \hat{\mathbf{Y}}(n|n))^{\mathrm{T}}]$
- $\mathbf{R}(n+1 \mid n) \quad \equiv \quad E[(\mathbf{Y}(n+1) - \hat{\mathbf{Y}}(n+1 \mid n))(\mathbf{Y}(n+1) - \hat{\mathbf{Y}}(n+1|n))^{\mathrm{T}}]$

We can apply the same reasoning as used for the scalar Kalman filter to show that the recursive equations of the vector Kalman filter are given as follows.

- **Recursive equations of the Kalman filter**

  Prediction Step :

$$\boxed{\begin{aligned}\hat{\mathbf{Y}}(n+1 \mid n) &= \mathbf{H}(n+1)\hat{\mathbf{Y}}(n \mid n) \\ \hat{\mathbf{X}}(n+1 \mid n) &= \mathbf{A}(n+1)\hat{\mathbf{Y}}(n+1 \mid n) \\ \mathbf{R}(n+1 \mid n) &= \mathbf{H}(n+1)\mathbf{R}(n \mid n)\mathbf{H}(n+1)^{\mathrm{T}} + \mathbf{Q}_{\mathbf{Z}}(n+1)\end{aligned}}$$

  Updating Step :

$$\hat{\mathbf{Y}}(n+1 \mid n+1) = \hat{\mathbf{Y}}(n+1 \mid n) + \mathbf{B}(n+1)[\mathbf{X}(n+1) - \hat{\mathbf{X}}(n+1 \mid n)]$$

$$\mathbf{R}(n+1 \mid n+1) = [\mathbf{I} - \mathbf{B}(n+1)\mathbf{A}(n+1)]\mathbf{R}(n+1 \mid n)$$

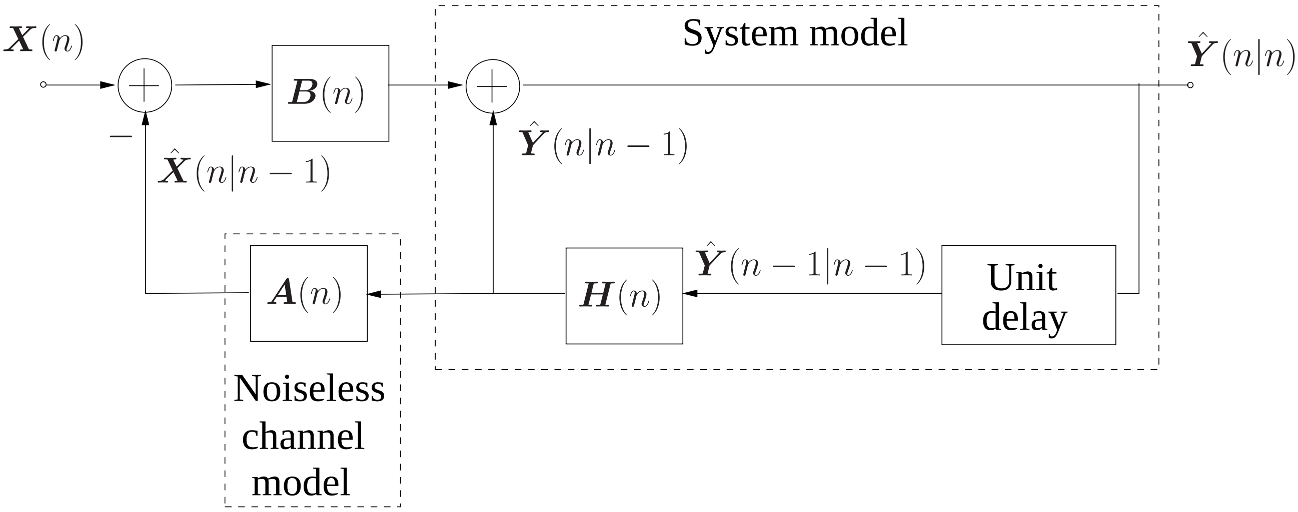
with the Kalman matrix

$$\mathbf{B}(n+1) \equiv \mathbf{R}(n+1 \mid n)\mathbf{A}(n+1)^{\mathrm{T}}[\mathbf{A}(n+1)\mathbf{R}(n+1 \mid n)\mathbf{A}(n+1)^{\mathrm{T}} + \mathbf{Q}_{\mathbf{W}}(n+1)]^{-1}$$

Initialization :

$$\boxed{\begin{aligned}&\hat{\mathbf{Y}}(0 \mid 0) = \mu_{\mathbf{Y}}(0) \\ &\mathbf{R}(0 \mid 0) = \sum\nolimits_{\mathbf{Y}(0)\mathbf{Y}(0)}\end{aligned}}$$

- **Block diagram of the vector Kalman filter**



5.3 Example of a recursive estimator

- **Signal model**

$$X(n) = Y + W(n) \qquad n = 1, 2, 3, \ldots$$

Where:

  - Y is and unknown constant to be estimated based on the observation of $\{X(n)\}$.
  - $\{W(n)\}$ is a white noise sequence.

- **Arithmetic mean**

An appealing linear estimator for Y is the ***arithmetic mean***

$$\hat{Y}(n) = \frac{1}{n}\sum_{m=1}^{n} X(m)$$

Drawback: To compute $\hat{Y}(n)$ based on the above formula, $X(1), \ldots, X(n)$ need to be stored. The required memory grows linearly with n.

- **Recursive implementation**

$$\hat{Y}(n+1) = \frac{1}{n+1}\sum_{m=1}^{n} X(m) + \frac{1}{n+1}X(n+1)$$

$$\boxed{\hat{Y}(n+1) = \frac{n}{n+1}\hat{Y}(n) + \frac{1}{n+1}X(n+1)}$$

This estimator requires storage of one value, i.e. $\hat{Y}(n)$, only.

5.4 Example of a signal model: Target tracking

- **Equations of the movement of a target:**

Position: $U(t) = \int_0^t V(t')dt' + U(0)$, $U(0)$: initial position

Velocity: $V(t) = \int_0^t G(t')dt' + V(0)$, $V(0)$: initial velocity

Acceleration: $G(t)$ is assumed to be white noise.

- **Discrete-time model:**

$\frac{du}{dt}(t) = V(t)$ $\qquad \frac{du}{dt}(nT_s) \approx [U((n+1)T_s) - U(nT_s)]/T_s$

$\frac{dv}{dt}(t) = G(t)$ $\qquad \frac{dv}{dt}(nT_s) \approx [V((n+1)T_s) - V(nT_s)]/T_s$

$U((n+1)T_s) - U(nT_s) = V(nT_s) \cdot T_s$ $\qquad T_s$: Sampling interval

$V((n+1)T_s) - V(nT_s) = \tilde{G}(nT_s) \cdot T_s$ $\qquad \tilde{G} = \tilde{G}(t)$ low-pass filtered with bandwidth $\frac{1}{2T_s}$.

State model

$$\underbrace{\begin{bmatrix} U(nT_s) \\ V(nT_s) \end{bmatrix}}_{\boldsymbol{Y}(n)} = \underbrace{\begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{H}(n)} \underbrace{\begin{bmatrix} U((n-1)T_s) \\ V((n-1)T_s) \end{bmatrix}}_{\boldsymbol{Y}(n-1)} + \underbrace{\begin{bmatrix} 0 \\ T_s\tilde{G}((n-1)T_s) \end{bmatrix}}_{\boldsymbol{Z}(n)}$$

$$\boldsymbol{Y}(n) = \boldsymbol{H}(n) \quad \boldsymbol{Y}(n-1) + \boldsymbol{Z}(n)$$

with

$$\boldsymbol{Y}(0) = [U(0), V(0)]^{\mathrm{T}},$$

$$\boldsymbol{Q}_{\boldsymbol{Z}}(n) = \begin{bmatrix} 0 & 0 \\ 0 & T_s^2 E[\tilde{G}(nT_s)^2] \end{bmatrix}.$$

Observation model

$$X(n) = U(nT_s) + \underbrace{W(n)}_{\text{Measurement error}}$$

$$X(n) = \underbrace{[1 \quad 0]}_{\boldsymbol{A}(n)} \boldsymbol{Y}(n) + W(n)$$

where $W(n)$ is white noise with variance σ^2_{WW}.

6. Model-Free and Model-Based Estimation of Random Processes

6.1. Model-free estimation of random processes

In this section $\{X(n)\}$ is a WSS process with

- mean value: $\mu_X \equiv \mathbf{E}[X(n)]$

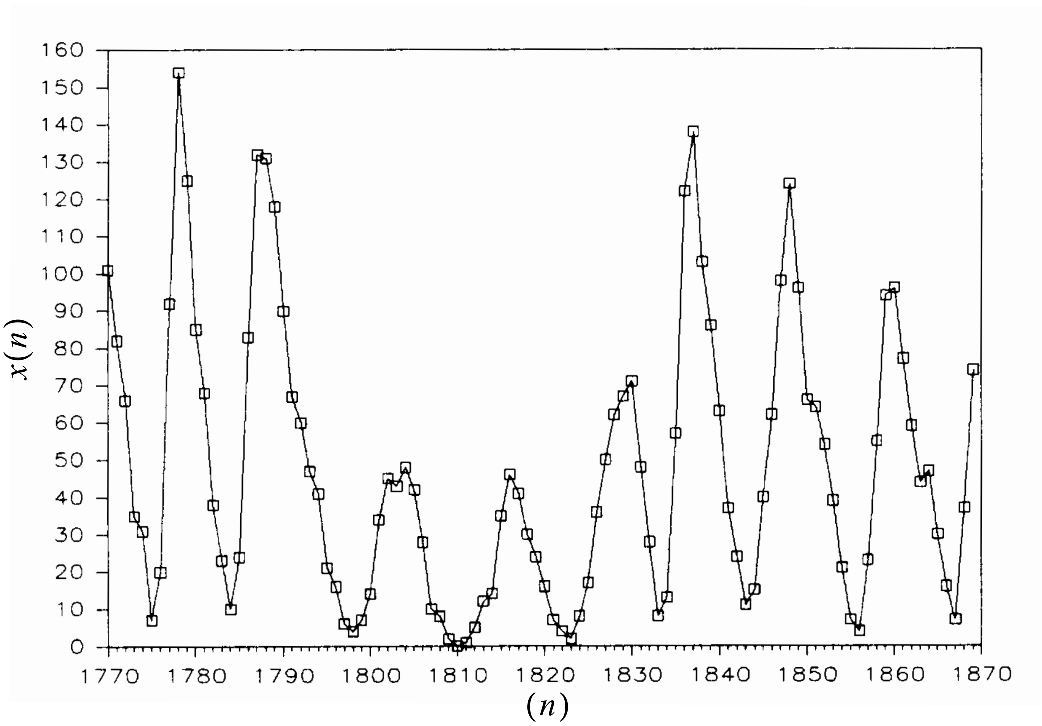
- autocorrelation function: $R_{XX}(k) \equiv \mathbf{E}[X(n)X(n+k)]$

The autocovariance function of $\{X(n)\}$ is

$$C_{XX}(k) \equiv \mathbf{E}[(X(n)-\mu_X)(X(n+k)-\mu_X)] = R_{XX}(k)-\mu_X^2$$

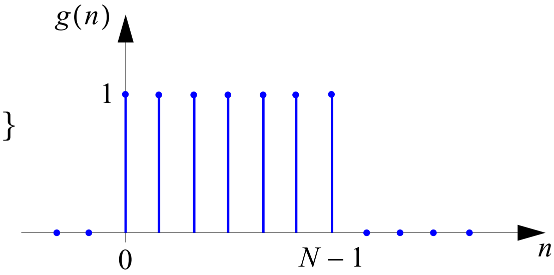
- ***Observed sequence:***
  We assume that $\{X(0), \ldots, X(N-1)\}$ can be observed.

Example 1: Wölfer sunspot numbers



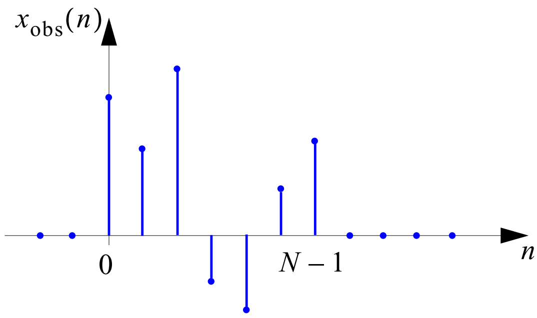
Defining the window function

$$g(n) \equiv \begin{cases} 1\ ; & n \in \{0, \ldots, N-1\} \\ 0\ ; & \text{otherwise} \end{cases}$$



the observed sequence reads:

$$\boxed{X_{\text{obs}}(n) = g(n)X(n)}$$



6.1.1. Estimation of the mean-value

- ***Arithmetic mean:***

$$\boxed{\hat{\mu}_X \equiv \overline{X} = \frac{1}{N}\sum_{n=0}^{N-1} X(n)}$$

- ***Mean and variance of $\overline{X}$:***

  - Mean: $\overline{X}$ is an unbiased estimator of μ_X:

$$\mu_{\overline{X}} = \mu_X$$

  - Variance:

$$\sigma^2_{\overline{X}} = \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left[1 - \frac{|k|}{N}\right] C_{XX}(k)$$

Special case: When $\{X(n)\}$ is an uncorrelated process:

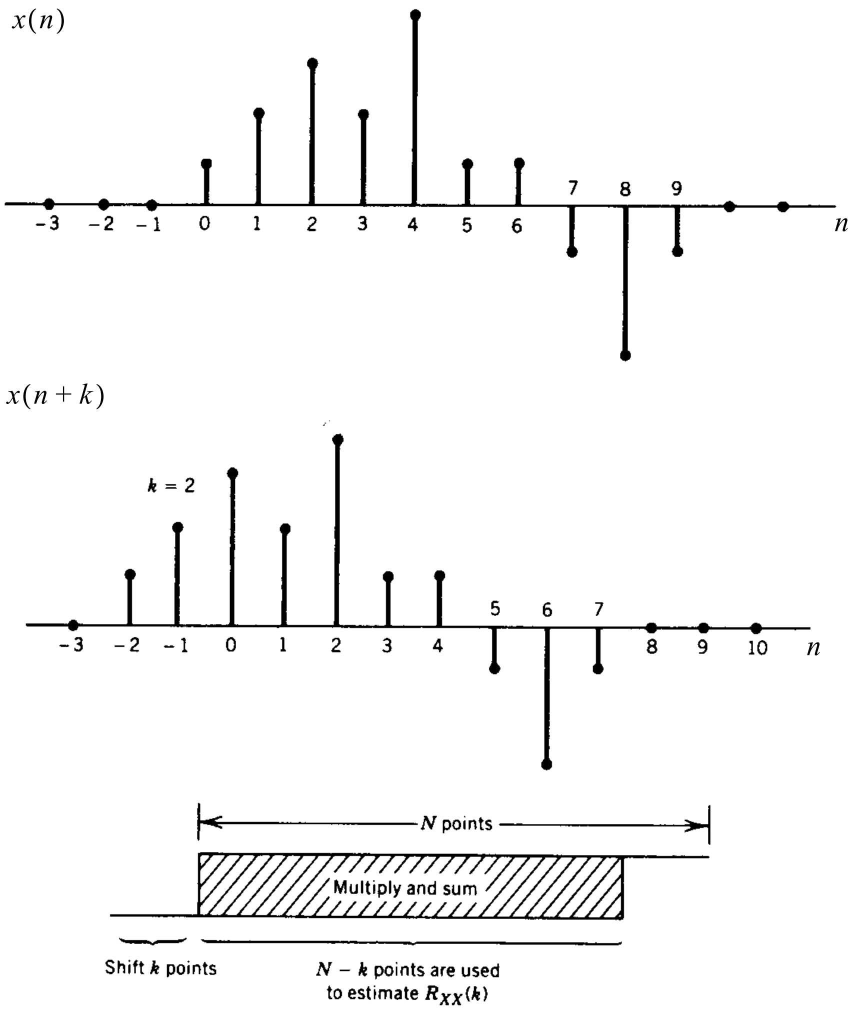
$$\sigma^2_{\overline{X}} = \frac{1}{N}C_{XX}(0) = \frac{1}{N}\sigma^2_X$$

Proof: See Exercise 9.1.

6.1.2. Estimation of the autocorrelation function:

- ***Biased sample autocorrelation function:***

$$\hat{R}_{XX}(k) \equiv \begin{cases} \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-k-1} X(n)X(n+k) \; ; & k = 0, \ldots, N-1 \\ \hat{R}_{XX}(-k) \; ; & k = -(N-1), \ldots, -1 \\ 0 \; ; & |k| \geq N \end{cases} \tag{6.1}$$



To show that the sample autocorrelation function $\hat{R}_{XX}(k)$ is biased we recast it as:

$$\begin{aligned}\hat{R}_{XX}(k) &= \frac{1}{N}\sum_{n=-\infty}^{\infty} X_{\text{obs}}(n)X_{\text{obs}}(n+k) \\ &= \frac{1}{N}\sum_{n=-\infty}^{\infty} g(n)g(n+k)X(n)X(n+k)\end{aligned}$$

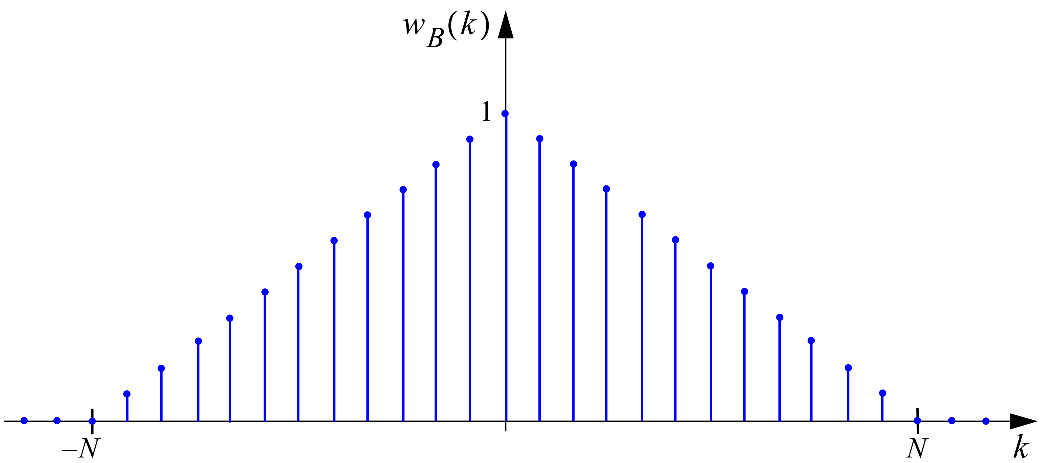
Taking the expectation on both side yields

$$\mathbf{E}[\hat{R}_{XX}(k)] = \frac{1}{N}R_{gg}(k)R_{XX}(k)$$

The function

$$\boxed{w_B(k) \equiv \frac{1}{N}R_{gg}(k) = \begin{cases} 1-\dfrac{|k|}{N} & ; \quad |k| < N \\ 0 & ; \quad \text{otherwise} \end{cases}}$$

is called the **Bartlett window**.



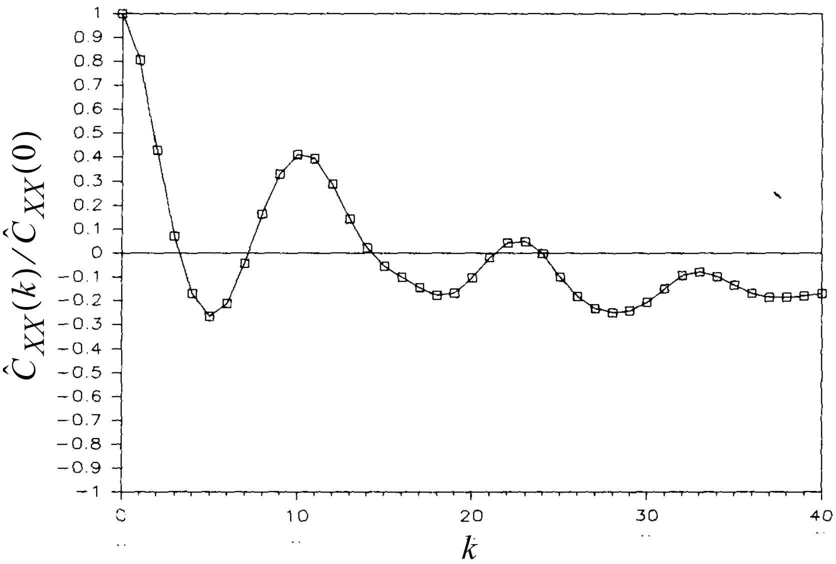
With this definition, the bias of $\hat{R}_{XX}(k)$ can be recast as

$$\boxed{\mathbf{E}[\hat{R}_{XX}(k)] = w_B(k)R_{XX}(k)} \qquad (6.2)$$

- ***Biased sample autocovariance:***

$$\boxed{\hat{C}_{XX}(k) = \hat{R}_{XX}(k) - \hat{\mu}_X^2}$$

Example 1: Wölfer sunspot numbers



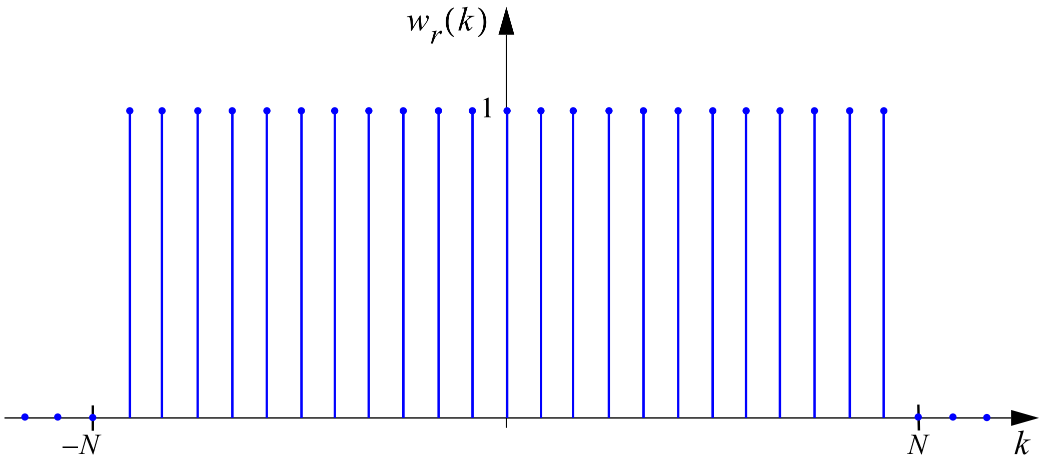
- ***Unbiased sample autocorrelation function:***

$$\boxed{\widehat{R}_{XX}(k) \equiv \begin{cases} \dfrac{1}{N-k}\displaystyle\sum_{n=0}^{N-k-1} X(n)X(n+k) \ ; & k = 0, \ldots, N-1 \\ \widehat{R}_{XX}(-k) \ ; & k = -(N-1), \ldots, -1 \\ 0 \ ; & |k| \geq N \end{cases}}$$

$\widehat{R}_{XX}(k)$ is unbiased for $|k| < N$:

$$\boxed{\mathbf{E}[\widehat{R}_{XX}(k)] = w_r(k)R_{XX}(k)}$$

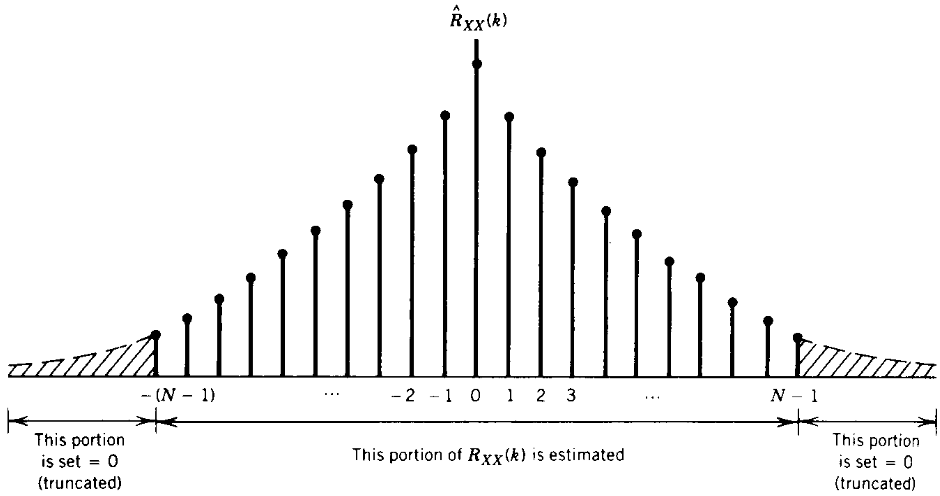
where $w_r(k)$ is the centered rectangular function:



- ***Properties of the sample autocorrelation functions:***

  - $\hat{R}_{XX}(k) = w_B(k)\widehat{R}_{XX}(k)$

  - With N observations, we can only estimate $R_{XX}(k)$ for $|k| < N$.



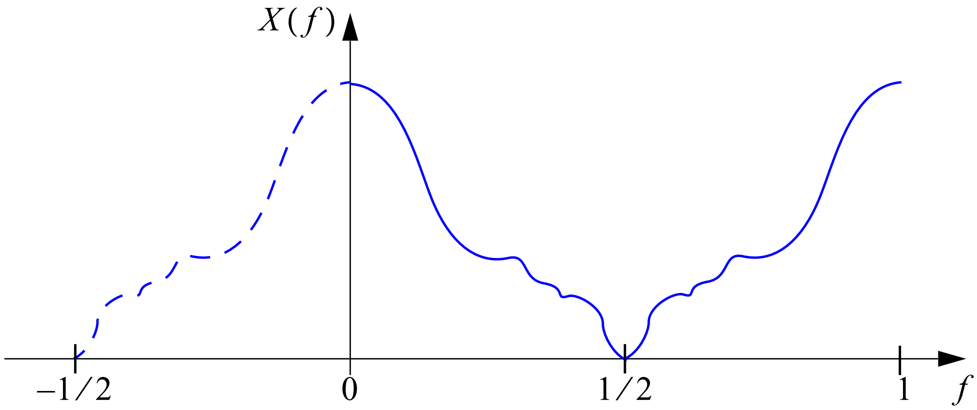
  - In general, it is difficult to calculate the variance of the sample autocorrelation functions since the computation involves fourth moments of the form $\mathbf{E}[X(n)X(n+m)X(k)X(k+m)]$.
    In the Gaussian case these moments can be evaluated and the variance of the sample autocorrelation functions can be calculated (See Exercise 9.8 of [Shanmugan]).

- A general conclusion is that the variance of $\hat{R}_{XX}(k)$ and $\widehat{R}_{XX}(k)$ increases with $|k|$ since the number of observations considered in the computation of these values is $N-|k|$.

6.1.3. Estimation of the power spectral density:

- ***Continuous-frequency periodogram:***
  Let us start from the slightly differently reformulated Fourier transform:

$$X(f) = \sum_{n-0}^{N-1} x(n)\exp(-j2\pi nf) \qquad f \in [0, 1)$$



The periodogram of $X_{\text{obs}}(n)$ is defined to be

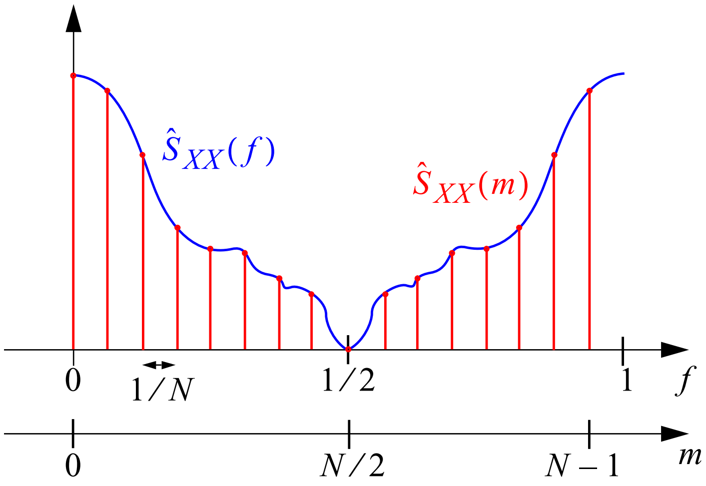
$$\hat{S}_{XX}(f) = \mathcal{F}\{\hat{R}_{XX}(k)\}$$
$$= \frac{1}{N}\left|\sum_{n-0}^{N-1} X(n)\exp(-j2\pi nf)\right|^2 = \frac{1}{N}\left|\mathcal{F}\{X_{\text{obs}}(n)\}(f)\right|^2 \qquad f \in [0, 1)$$

Proof:

□

- ***Discrete-frequency periodogram:***

$$\boxed{\hat{S}_{XX}(m) = \hat{S}_{XX}(f)\Big|_{f = m/N} \qquad m = 0, \ldots, N-1}$$



- ***Discrete Fourier transform:***
  The discrete Fourier transform and the inverse DFT are defined according to

$$\boxed{\begin{aligned} X_d(m) &= \mathcal{F}_d\{x(n)\} \equiv \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \exp\left(-j2\pi\frac{nm}{N}\right) \\ x(n) &= \mathcal{F}_d^{-1}\{X_d(m)\} \equiv \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_d(m) \exp\left(j2\pi\frac{nm}{N}\right) \end{aligned}}$$

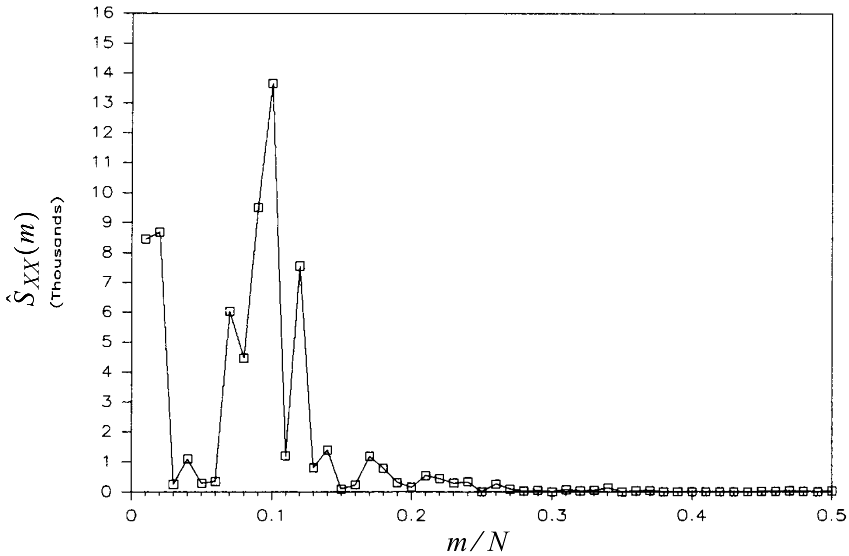
  Relation between the discrete Fourier transform and the (continuous-frequency) Fourier transform:

$$X_d(m) = \frac{1}{\sqrt{N}} X(f)\Big|_{f = m/N} \qquad m = 0, \ldots, N-1$$

  In particular, the discrete-frequency periodogram can be computed as

$$\boxed{\hat{S}_{XX}(m) \equiv \left|\mathcal{F}_d\{X_{obs}(n)\}(m)\right|^2}$$

Example 1: Wölfer sunspot numbers



- ***Bias of the periodogram:***

  Because the Fourier transform is a linear operation, we have

$$\mathbf{E}[\hat{S}_{XX}(f)] = \mathcal{F}\{\mathbf{E}[\hat{R}_{XX}(k)]\}$$

It follows from (6.2) that:

$$\mathbf{E}[\hat{S}_{XX}(f)] = \mathcal{F}\{w_B(k)R_{XX}(k)\}$$

$$= W_F(f) * S_{XX}(f)$$

The Fourier transform

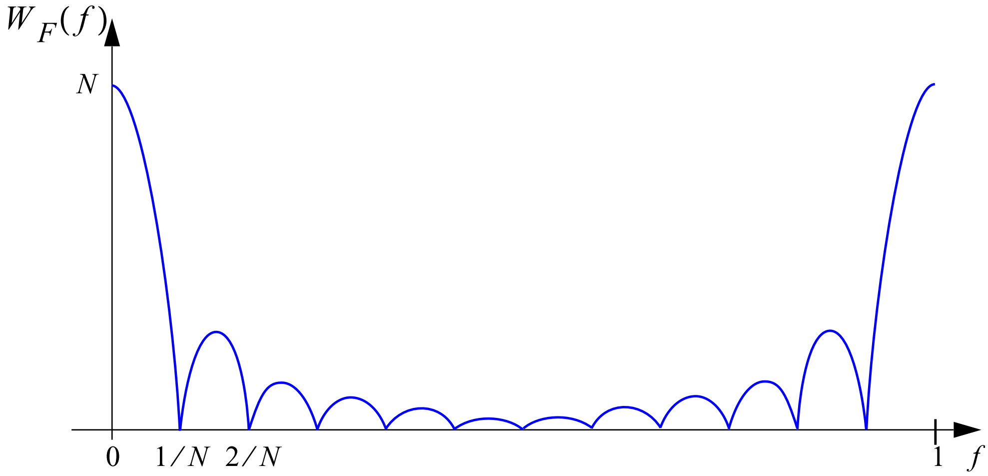
$$\boxed{W_F(f) \equiv \mathcal{F}\{w_B(k)\} = \frac{1}{N}\left(\frac{\sin(\pi f N)}{\sin(\pi f)}\right)^2}$$

of the Bartlett window is called the Féjer kernel.

Proof: It can be easily shown that the Fourier spectrum of $R_{gg}(k)$ is

$$|G(f)|^2 = \left(\frac{\sin(\pi f N)}{\sin(\pi f)}\right)^2$$

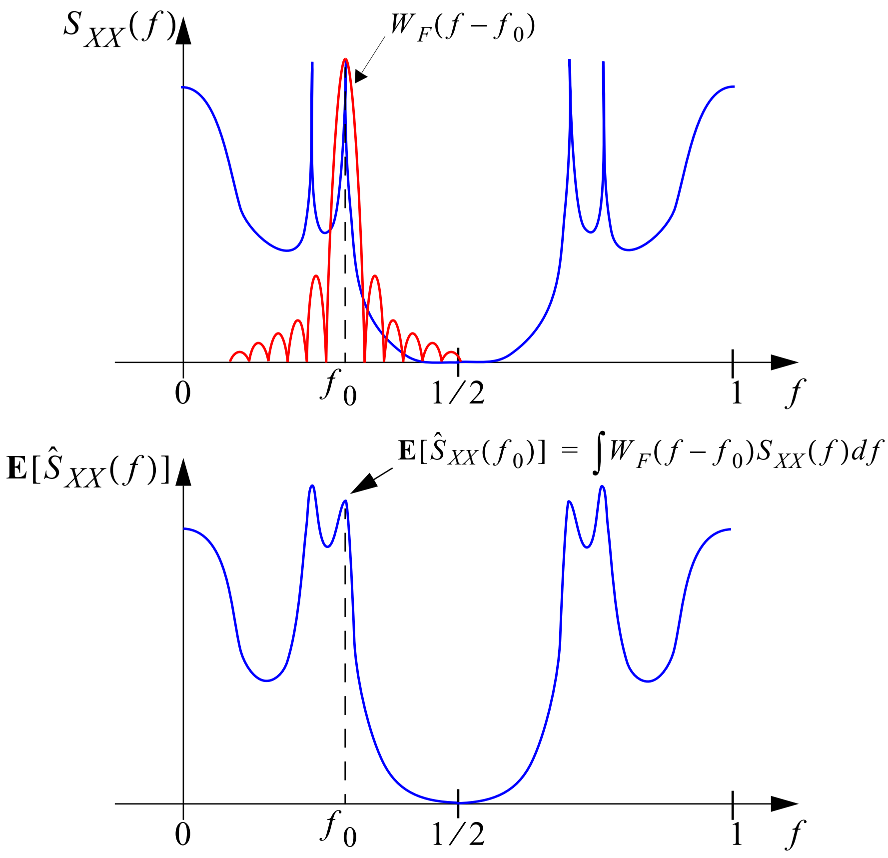
where $G(f) \equiv \mathcal{F}\{g(n)\}$.



In summary, the bias of $\hat{S}_{XX}(f)$ and $\hat{S}_{XX}(m)$ are given by

$$\boxed{\begin{aligned} \mathbf{E}[\hat{S}_{XX}(f)] &= W_F(f) * S_{XX}(f) \\ \mathbf{E}[\hat{S}_{XX}(m)] &= [W_F(f) * S_{XX}(f)]\big|_{f = m/N} \end{aligned}}$$

- ***Spectral leakage:***



As N increases to infinity, $W_F(f) \to \delta(f)$, so that

$$\mathbf{E}[\hat{S}_{XX}(f)] \to S_{XX}(f),$$

i.e. $\hat{S}_{XX}(f)$ and $\hat{S}_{XX}(m)$ are asymptotically unbiased.

- ***Variance of the periodogram:***
  The following asymptotic results are valid for a large classes of stochastic processes, and in particular for ARMA processes.

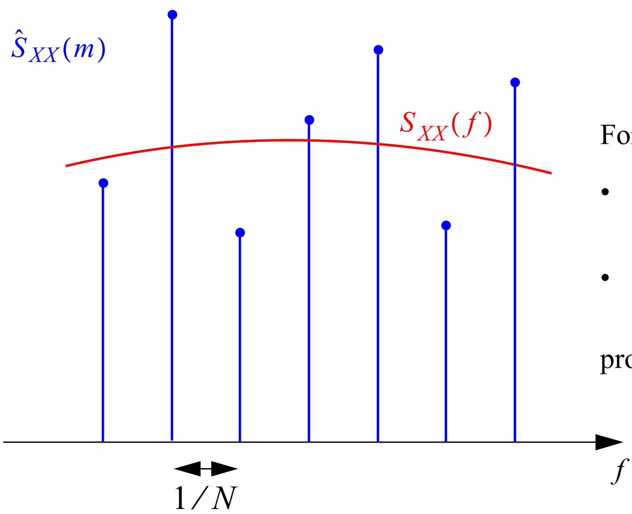
As the number N of observations tends to infinity,

$$\sigma^2_{\hat{S}_{XX}(f)} \to \begin{cases} 2S_{XX}(f)^2 & ; \quad f = 0, 1/2 \\ S_{XX}(f)^2 & ; \quad \text{otherwise} \end{cases}$$

$$\Sigma_{\hat{S}_{XX}(f_1)\hat{S}_{XX}(f_2)} \to 0 \qquad \text{for any} \qquad f_1, f_2 \in \left[0, \frac{1}{2}\right], f_1 \neq f_2$$

Hence,

- Any two “different” samples of the periodogram are asymptotically uncorrelated.
  Remember that $\hat{S}_{XX}(f)$ and consequently $\hat{S}_{XX}(m)$ are even functions.
- As N increases the variance of the periodogram does not vanish but stabilizes to a value. This value coincides with the asymptotic mean of the periodogram when $f \neq 0, 1/2$.

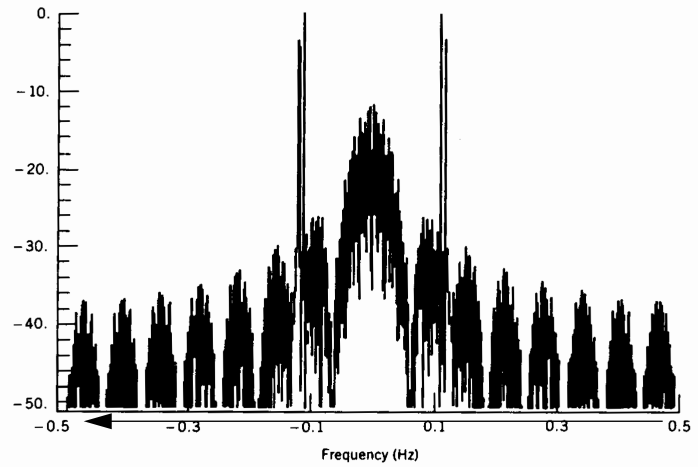


For large values of N:

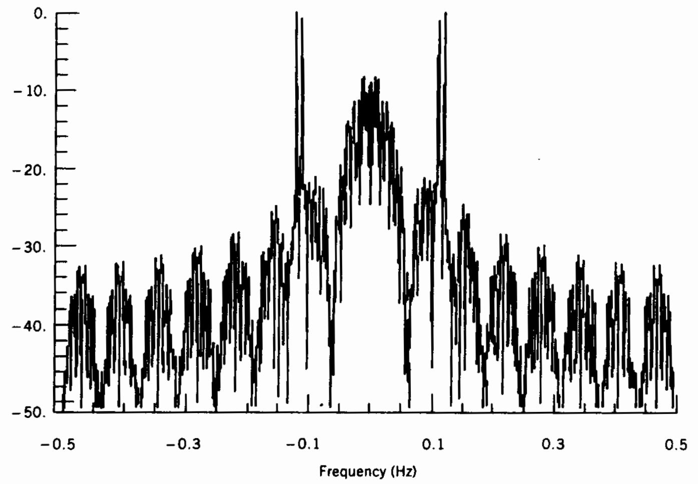
- $\mathbf{E}[\hat{S}_{XX}(m)] \approx S_{XX}(f)\big|_{f = \frac{m}{N}}$
- $\sigma_{\hat{S}_{XX}(m)} \approx S_{XX}(f)\big|_{f = \frac{m}{N}}$

provided $m \neq 0, N/2$

These two properties are responsible of the erratic nature of the periodogram (see the periodogram of the sunspot numbers).
Increasing the number of samples increases the spectral resolution only.



(a) $N = 4096$



(b) $N = 1024$

- ***Smoothing through windowing:***

Windowing aims at reducing the variability of the estimated spectrum.

A **lag window** $w(k)$ is a sequence satisfying the following properties:

- $w(k)$ is even, i.e $w(k) = w(-k)$.
- $w(k) = 0$ for $|k| > N$
- $w(0) = 1$

The **Blackman-Tukey estimator** of the spectrum is of the form

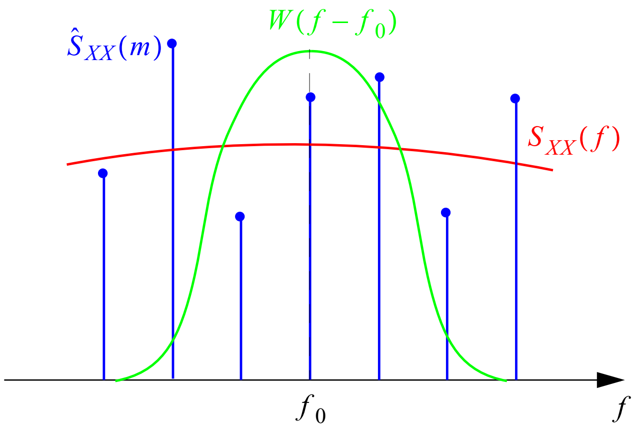
$$\boxed{\hat{S}_{XX}^{(W)}(f) = \mathcal{F}\{w(k)\hat{R}_{XX}(k)\}}$$

where $w(k)$ is a given lag window with Fourier transform $W(f)$.

Making use of the property of the Fourier transform, we obtain

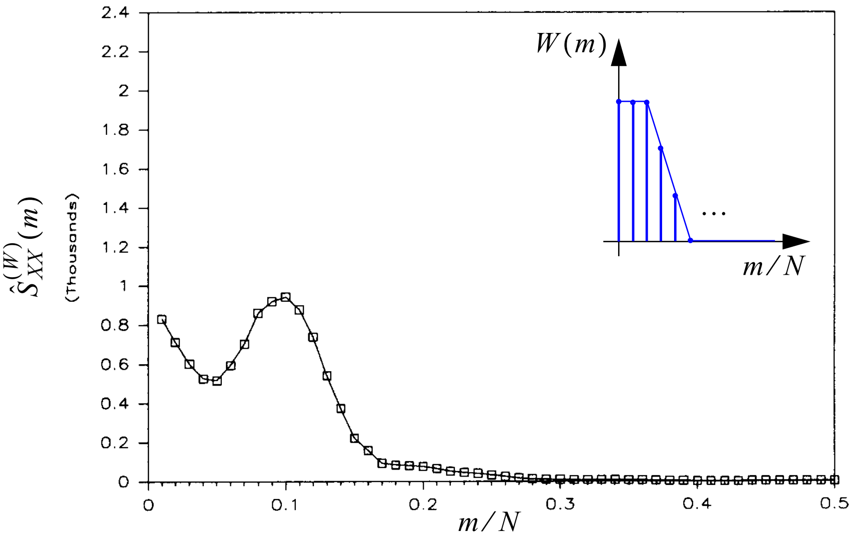
$$\boxed{\hat{S}_{XX}^{(W)}(f) = W(f) * \hat{S}_{XX}(f)}$$

Usually, the **spectral window** $W(f)$ is selected to have a narrow main lobe and low sidelobes. The above convolution corresponds to a local weighted averaging of $\hat{S}_{XX}(f)$.

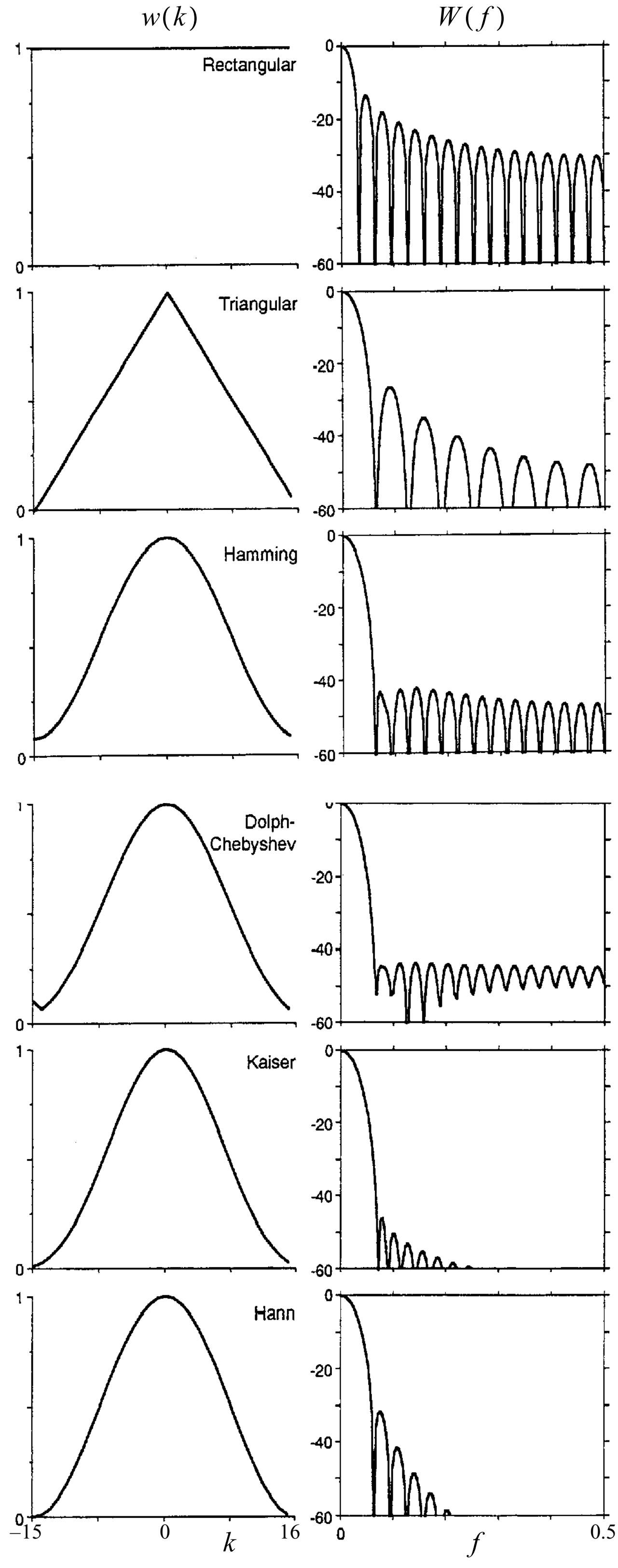


This averaging operation reduces the variability of $\hat{S}_{XX}^{(W)}(f)$ but also leads to a reduction of the spectral resolution.

Example 1: Wölfer sunspot numbers



Some well-known lag windows:



6.2. Parametric (model-based) estimation of random processes

6.2.1. Box-Jenkins method:

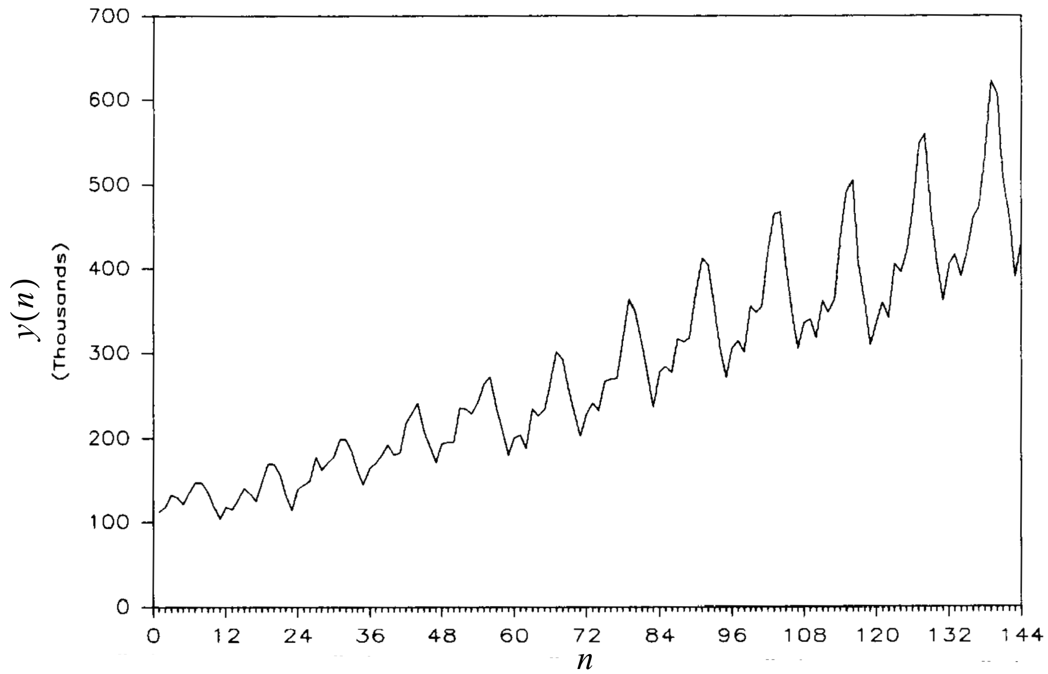
- ***Key idea of the method:***

  - The observed sequence $\{y(0), \ldots, y(N'-1)\}$ is transformed in such a way that the transformed sequence $\{x(0), \ldots, x(N-1)\}$ can be reasonably assumed to be the realization of a WSS process $\{X(n)\}$.

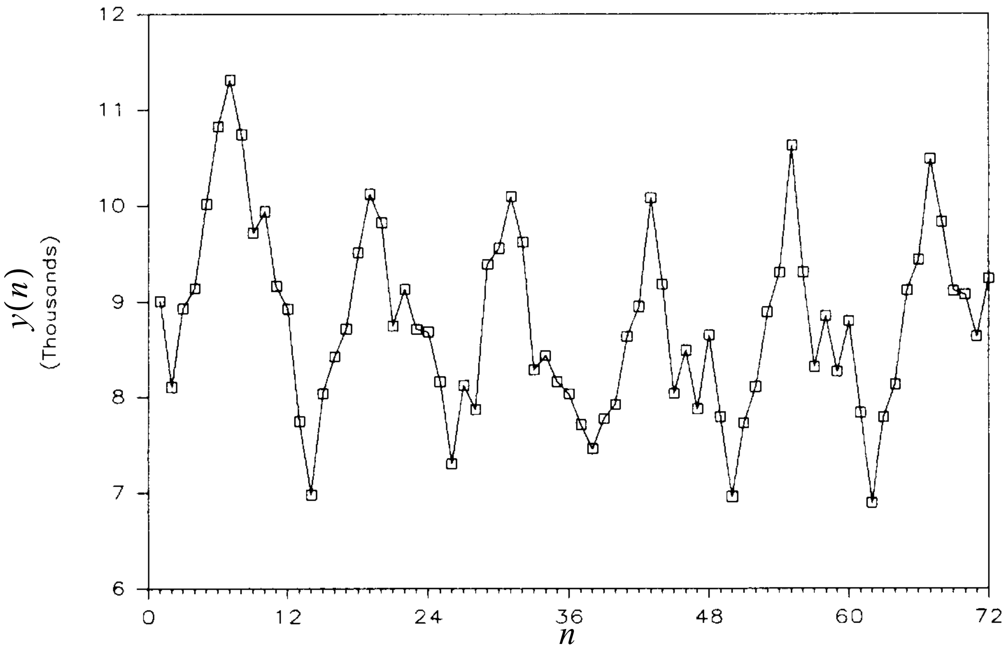
  - An ARMA*(p,q)* process is fitted to $\{x(0), \ldots, x(N-1)\}$.

  - The estimated autocorrelation function and power spectrum are identified to the autocorrelation function and the power spectrum of the estimated ARMA*(p,q)* process.

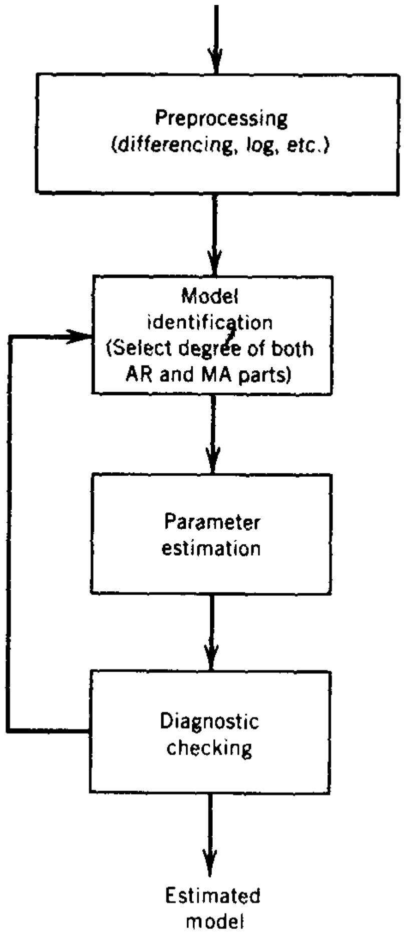
Example 2: International airline passengers.



Example 3: Monthly accidental deaths in the U.S.A.



- **The different steps of the Box-Jenkins method:**



6.2.2. Preprocessing:

- ***Objective:***

  The observed sequence $\{y(0), \ldots, y(N'-1)\}$ is transformed in such a way that the transformed sequence

  $$\boxed{\{x(0), \ldots, x(N-1)\} = T[\{y(0), \ldots, y(N'-1)\}]}$$

  can be reasonably assumed to be the realization of a WSS process $\{X(n)\}$.

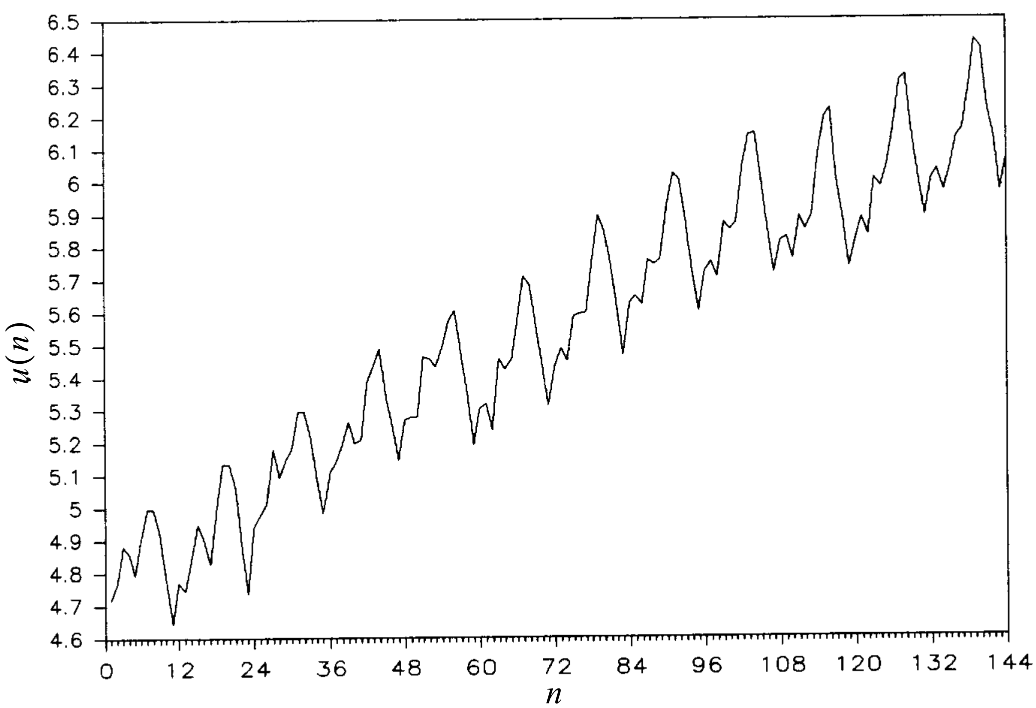
- **Non-linear transformation to create stationarity:**

  Let $\{Y(n)\}$ be a sequence which exhibits some non-stationary features. We can apply a non-linear transformation T to $\{Y(n)\}$ to obtain a new sequence $\{X(n)\} = T[\{Y(n)\}]$ where these features are eliminated or at least reduced.

  Example 2: International airline passengers.

  The variability of the serie increases linearly as a function of the level of the serie. This variability is stabilized by applying the following transformation:

  $$\boxed{U(n) = \ln(Y(n))}$$



To understand how the transformation $Y(n) \to \ln(Y(n))$ stabilizes the variability, let us assume that the standard deviation of $\{Y(n)\}$ increases proportionally to its expectation:

$$\sigma_{Y(n)} = c\mu_{Y(n)}$$

Equivalently,

$$E\left[\left(\frac{Y(n)}{\mu_{Y(n)}} - 1\right)^2\right] = c^2.$$

We can rewrite $U(n) = \ln(Y(n))$ as

$$U(n) = \ln(\mu_{Y(n)}) + \ln\left(\frac{Y(n)}{\mu_{Y(n)}}\right) = \ln(\mu_{Y(n)}) + \ln\left(\frac{Y(n) - \mu_{Y(n)}}{\mu_{Y(n)}} + 1\right)$$

Considering the first order Taylor approximation $\ln(v+1) \approx v$ around 1, $U(n)$ can be approximated according to

$$U(n) \approx \ln(\mu_{Y(n)}) + \left(\frac{Y(n)}{\mu_{Y(n)}} - 1\right)$$

Approximation of the expectation and standard deviation of $U(n)$:

$$\mu_{U(n)} \approx \ln(\mu_{Y(n)})$$

$$\sigma_{U(n)} \approx c$$

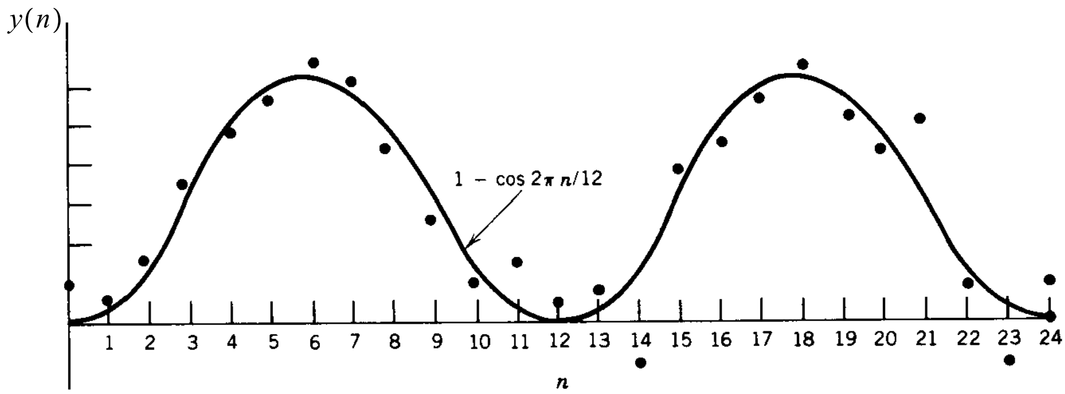
- ***Differentiating to remove periodicity (seasonality):***

Theoretical example 1:

Let consider the sequence $\{Y(n)\}$ where

$$Y(n) = \underbrace{\left[1 - \cos\left(2\pi\frac{n}{12}\right)\right]}_{\text{Periodic components of period 12}} + V(n)$$

where $\{V(n)\}$ is a WSS process.



For example, $\{Y(n)\}$ might represent a monthly average (see Examples 2 to 3). Let

$$\{X(n)\} = \Delta_{12}\{Y(n)\}$$

be the sequence obtained by transforming $\{Y(n)\}$ according to

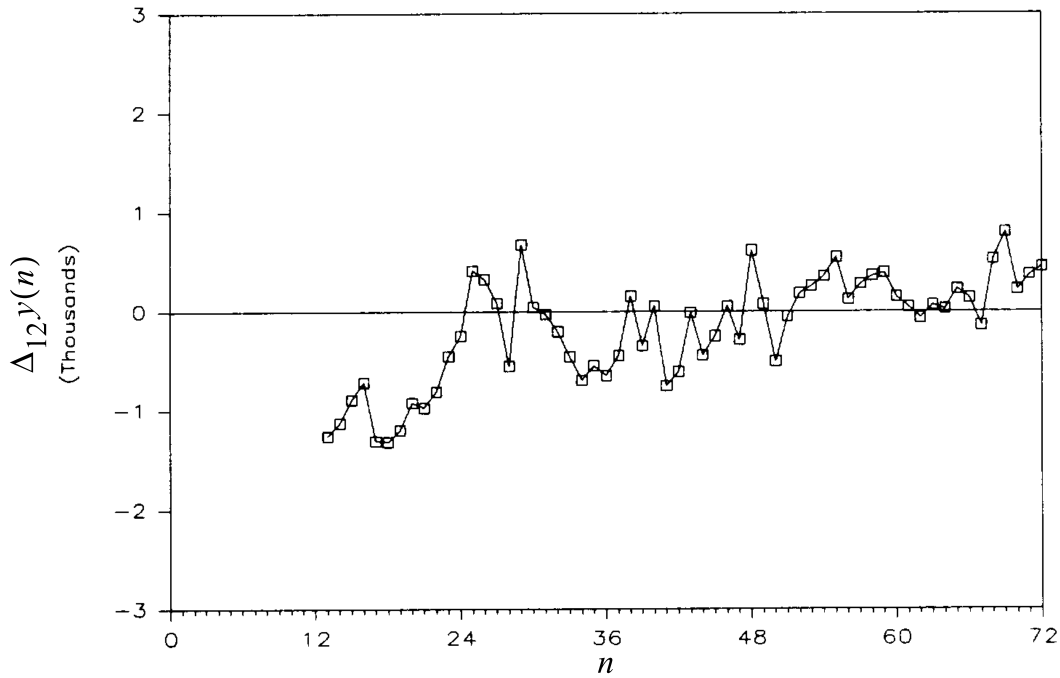
$$X(n) = Y(n) - Y(n-12)$$

Then

$$X(n) = V(n) - V(n-12)$$

Hence, the sequence $\{X(n)\}$ is stationary.

Example 3: Monthly accidental deaths in the U.S.A.



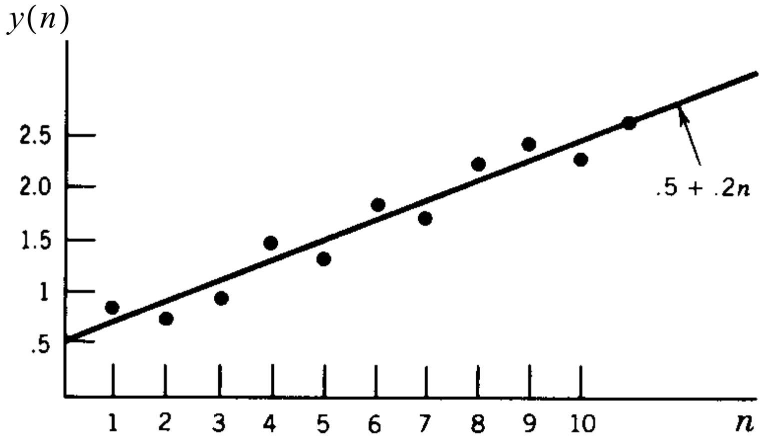
- ***Differentiating to remove trends:***

Theoretical example 2:

Let consider the sequence $\{Y(n)\}$ where

$$Y(n) = \underbrace{\left[\frac{1}{2} + \frac{1}{5}n\right]}_{\text{Trend}} + V(n)$$

where $\{V(n)\}$ is a WSS process.



Let us consider the transformation

$$X(n) = Y(n) - Y(n-1).$$

Then,

$$X(n) = V(n) - V(n-1) + \frac{1}{5} .$$

Hence, $\{X(n)\}$ is a WSS process, which can be modelled as an ARMA process.

- ***ARIMA(p,d,q) processes:***

  Notice that the above process $\{X(n)\}$ is the "discrete derivative" of $\{V(n)\}$.

  Let us introduce the following notation for discrete derivative:

$$\{X(n)\} = \Delta\{Y(n)\} \text{ if } X(n) = Y(n) - Y(n-1) \text{ for all } n.$$

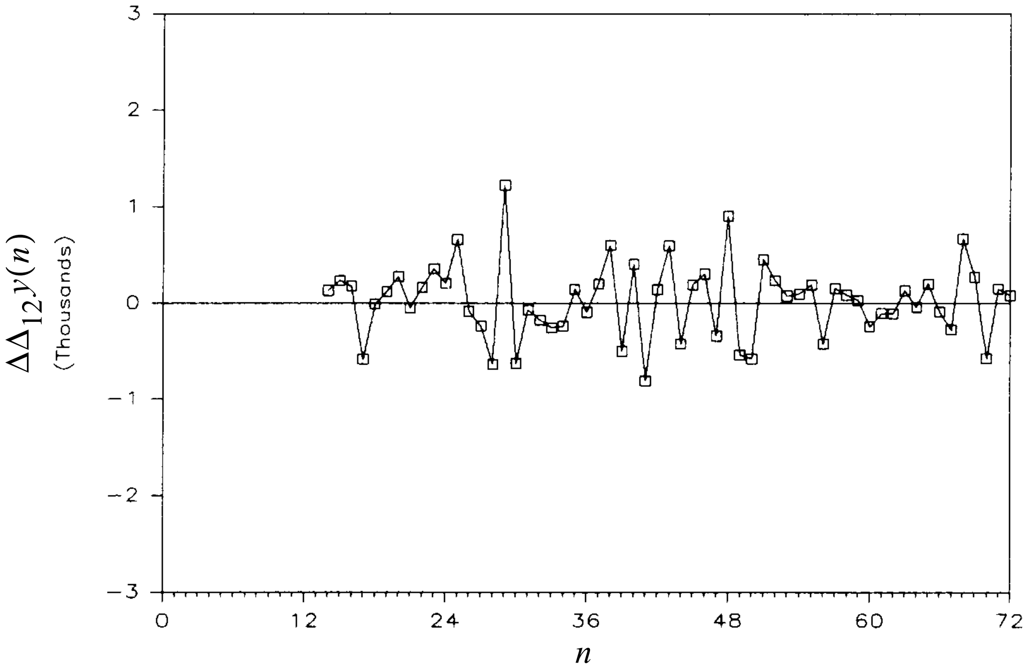
Notice that according to the previously introduced notation

$$\Delta\{Y(n)\} = \Delta_1\{Y(n)\}.$$

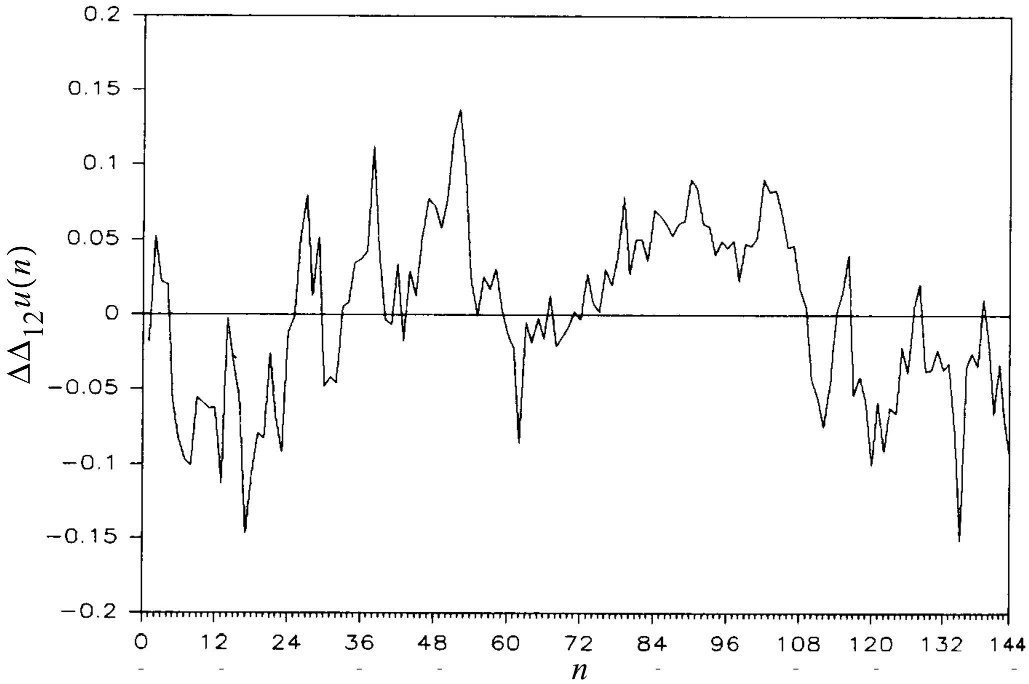
A process $\{Y(n)\}$ is an **ARIMA(*p,d,q*) process** if its *d*th discrete derivative $\{X(n)\} = \Delta^{(d)}\{Y(n)\}$ is an ARMA(*p,q*) process.

An ARIMA process reduces after differentiating finitely many times to an ARMA process. The letter **I** in ARIMA stands for **"integrated"**.
Notice that if $\{X(n)\} = \Delta\{Y(n)\}$ then $\{Y(n)\}$ can be obtained by carrying out a discrete integration of $\{X(n)\}$.

Example 3: Monthly accidental deaths in the U.S.A.



Example 2: International airline passengers.



6.2.3. Fitting ARMA(p,q) processes:

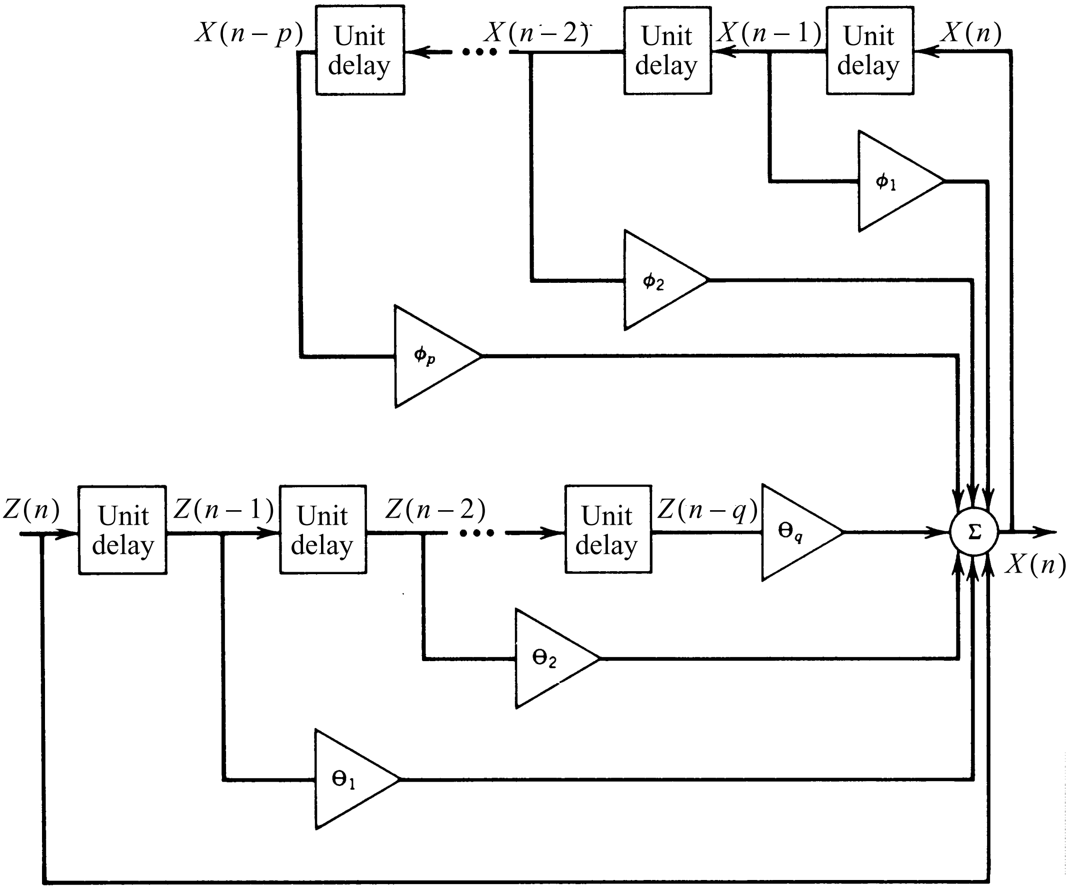
- ***Definition (review):***

  A random sequence $\{X(n)\}$ is an autoregressive moving average process (p, q) th order (ARMA((p, q))) if it is WSS and for any n:

$$\boxed{X(n) = \sum_{i=1}^{p} \phi_i X(n-i) + \sum_{i=1}^{q} \theta_i Z(n-i) + Z(n)}$$

  where $Z(n)$ is a white Gaussian process with variance σ_Z^2.

- ***Filter implementation:***



- ***Parameter estimation:***

  - Model order p, q**:**

  p and q are estimated by applying the Akaike information criterion (AIC) or the minimum description length (MDL) criterion.

- Coefficients $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$:

1. The parameters of an AR process can be estimated by solving the Yule-Walker equations:

$$\boxed{\begin{aligned} \hat{\gamma} &= \hat{\Gamma}\hat{\Phi} \\ \hat{R}_{XX}(0) &= \hat{\gamma}^T\hat{\Phi} + \hat{\sigma}_Z^{\ 2} \end{aligned}}$$

where

$$\hat{\Phi} \equiv \begin{bmatrix} \hat{\phi}_1 \\ \ldots \\ \hat{\phi}_p \end{bmatrix} \qquad \hat{\gamma} \equiv \begin{bmatrix} \hat{R}_{XX}(1) \\ \hat{R}_{XX}(2) \\ \ldots \\ \hat{R}_{XX}(p) \end{bmatrix}$$

$$\hat{\Gamma} \equiv \begin{bmatrix} \hat{R}_{XX}(0) & \hat{R}_{XX}(1) & \ldots & \hat{R}_{XX}(p-1) \\ \hat{R}_{XX}(-1) & \hat{R}_{XX}(0) & \ldots & \hat{R}_{XX}(p-2) \\ \ldots & \ldots & \ldots & \ldots \\ \hat{R}_{XX}(-(p-1)) & \hat{R}_{XX}(-(p-2)) & \ldots & \hat{R}_{XX}(0) \end{bmatrix}$$

Example 1: Wölfer sunspot numbers

The estimated AR model for the mean-corrected data is found to be

a) $\hat{p} = 3$,

b) $X(n) - \hat{\phi}_1 X(n-1) + \hat{\phi}_2 X(n-2) - \hat{\phi}_3 X(n-3) = Z(n)$

2. In the general case of an ARMA process, $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ can be estimated by using the **maximum likelihood method**.

Example 1: Wölfer sunspot numbers

The estimated ARMA model for the mean-corrected data is found to be

a) $\hat{p} = 9$, $\hat{q} = 1$,

b) $X(n) - 1.475X(n-1) + 0.937X(n-2)$
$\qquad -0.218X(n-3) + 0.134X(n-9) = Z(n)$

- ***Estimate of the power spectrum:***

  - Estimate of the transfer function:

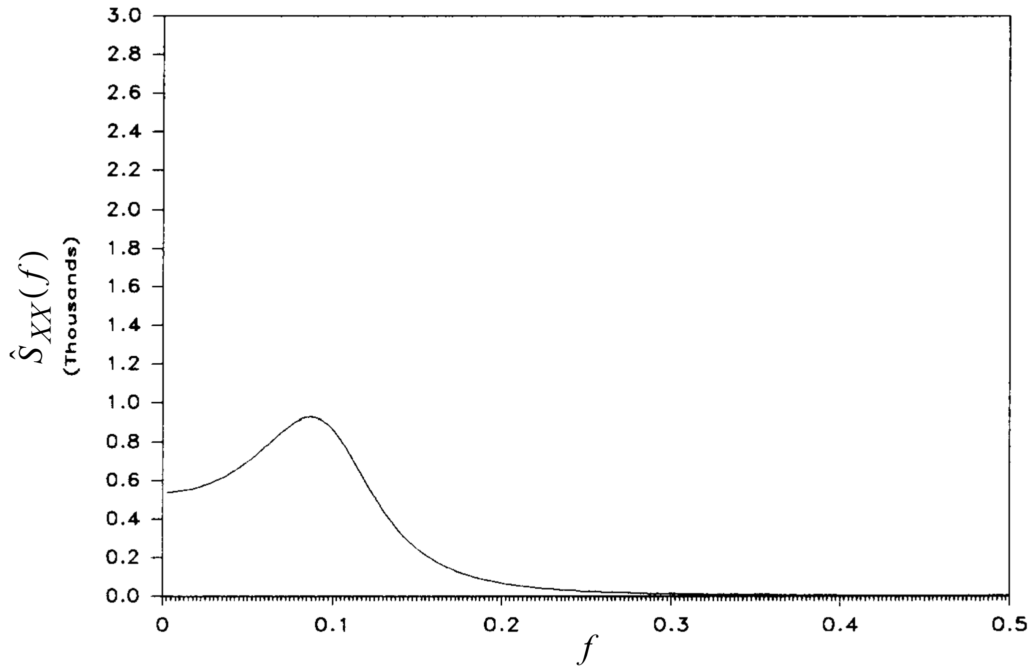
$$\hat{H}(f) = \frac{1 + \sum_{i=1}^{\hat{q}} \hat{\theta}_i \exp(-j2\pi i f)}{1 - \sum_{i=1}^{\hat{p}} \hat{\phi}_i \exp(-j2\pi i f)}$$

  - Estimate of the power spectrum:

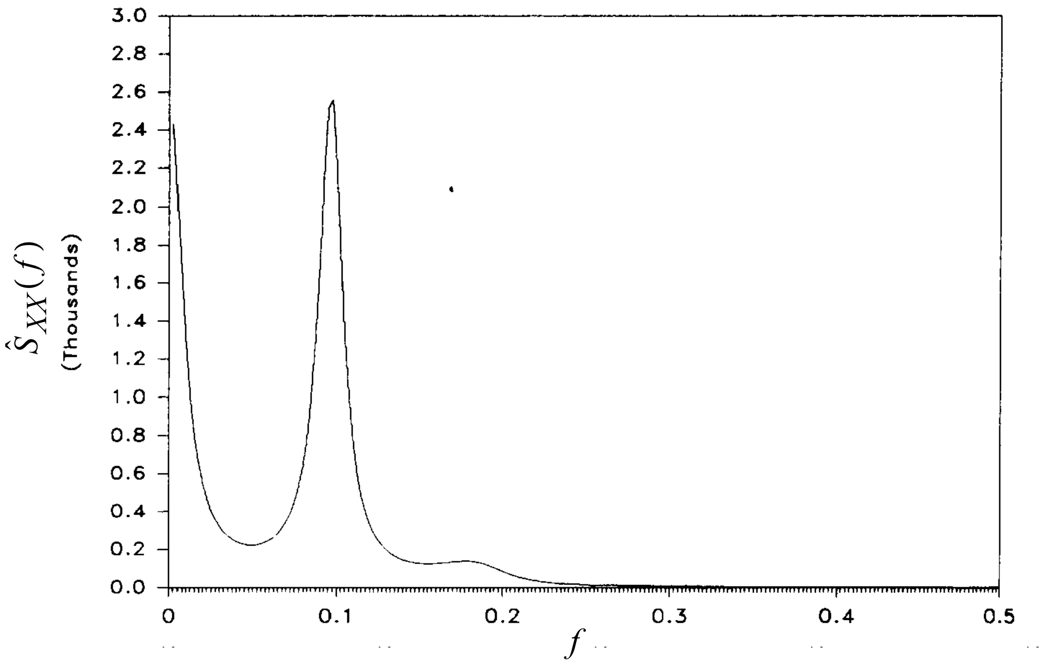
$$\boxed{\hat{S}_{XX}(f) = \frac{\left|1 + \sum_{i=1}^{\hat{q}} \hat{\theta}_i \exp(-j2\pi i f)\right|^2}{\left|1 - \sum_{i=1}^{\hat{p}} \hat{\phi}_i \exp(-j2\pi i f)\right|^2} \hat{\sigma}_Z^{\,2}}$$

Example 1: Wölfer sunspot numbers:

- Estimate with the AR(3) model:



- Estimate with the ARMA(9,1) model:



- ***Estimate of the autocorrelation function:***

$$\boxed{\hat{R}_{XX}(k) = \mathcal{F}^{-1}\{\hat{S}_{XX}(f)\}}$$