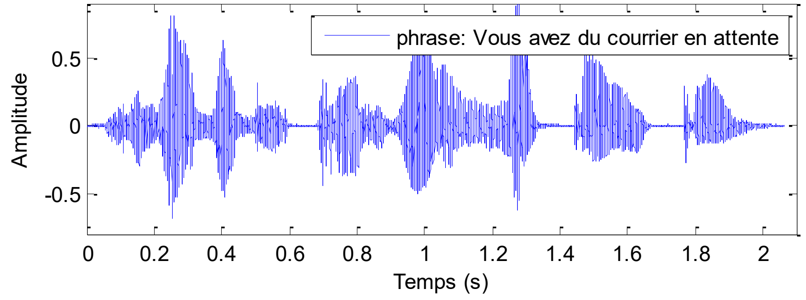
II. Random Processes

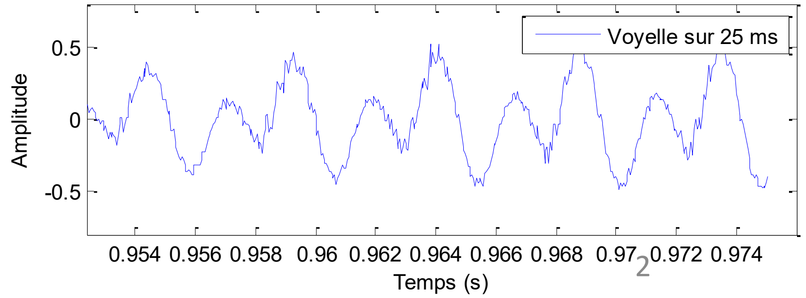
Introduction

- Random (stochastic) processes describe the evolution of a **random quantity** as a **function of time** (or space).
- A stochastic process is a **family of random variables indexed by time** s defined on the same probability space Ω .
- A random process cannot be represented analytically in time.
- **Each observed random signal** represents **a particular realization** of this process.

Examples

- ❑The speech signal,
- ❑the radar signal,
- ❑electrocardiogram, electroencephalogram,
- ❑seismic signals

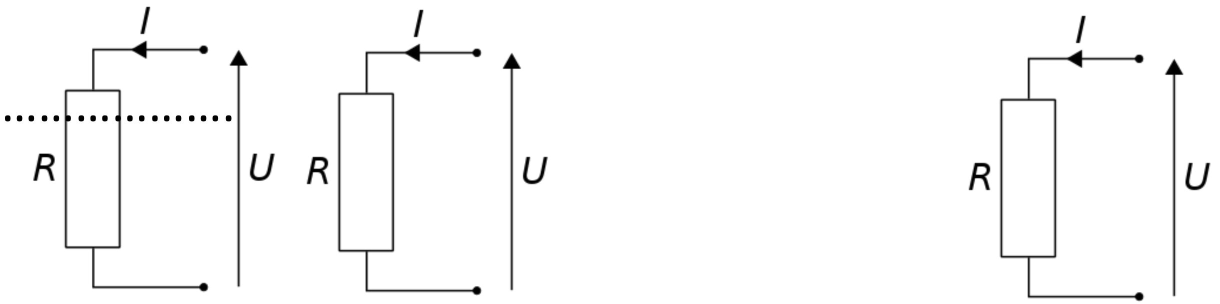




Introduction

Example 1

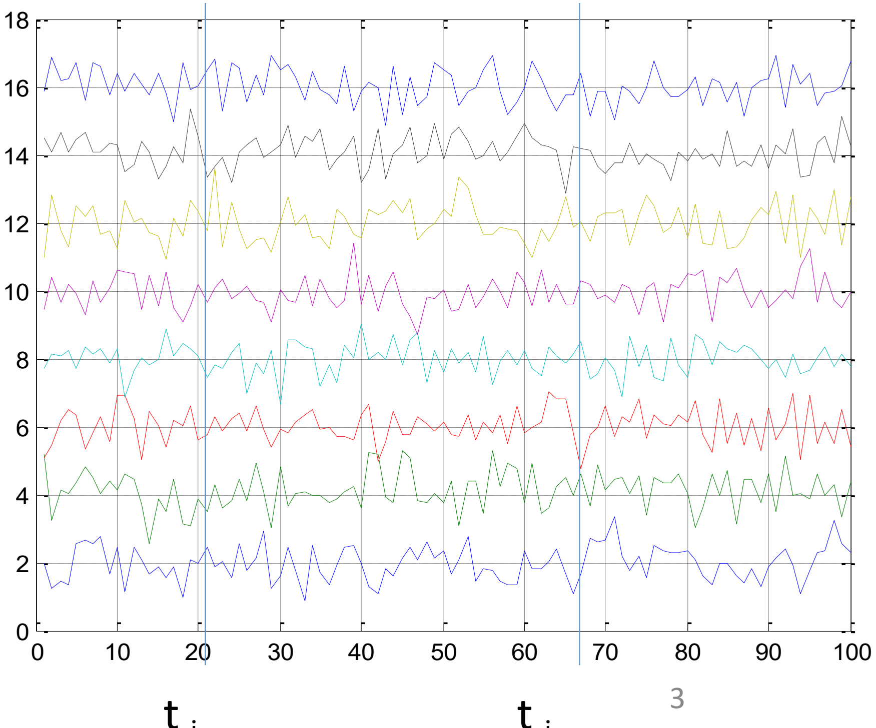
- Several identical resistors (of the same value) → Non-zero voltage measurement due to thermal agitation of free electrons in the resistor.
- The voltages measured over time on all the resistors → random signals all different which constitute the random process.



- Each trace provides a random signal
- at t_i, the process is reduced to a RV x_i
- Two moments t_i and t_j : 2 RV . x_i and x_j →

$$p(x;t_i) = p(x_i)$$

$$p(x;t_i,t_j) = p(x_i,x_j)$$

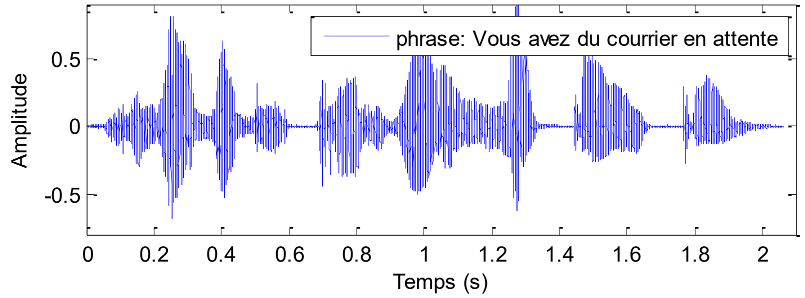


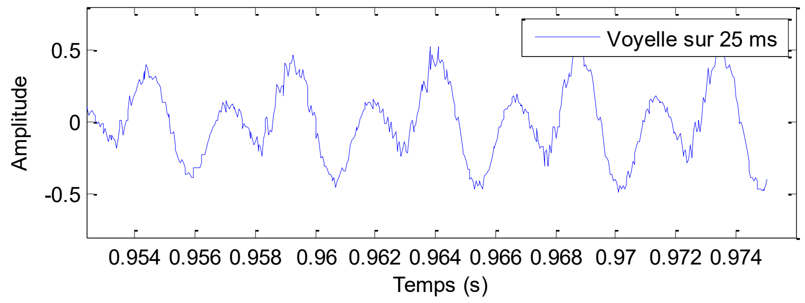
Random Variable Reminders

- Every physical signal has a random component, whether it is an external disturbance such as a random voltage across a resistor or an atmospheric disturbance.

The useful signal itself is random if we consider

- ❑the dice roll,
- ❑the lottery,
- ❑match results
- ❑weather phenomena,
- ❑etc., ...





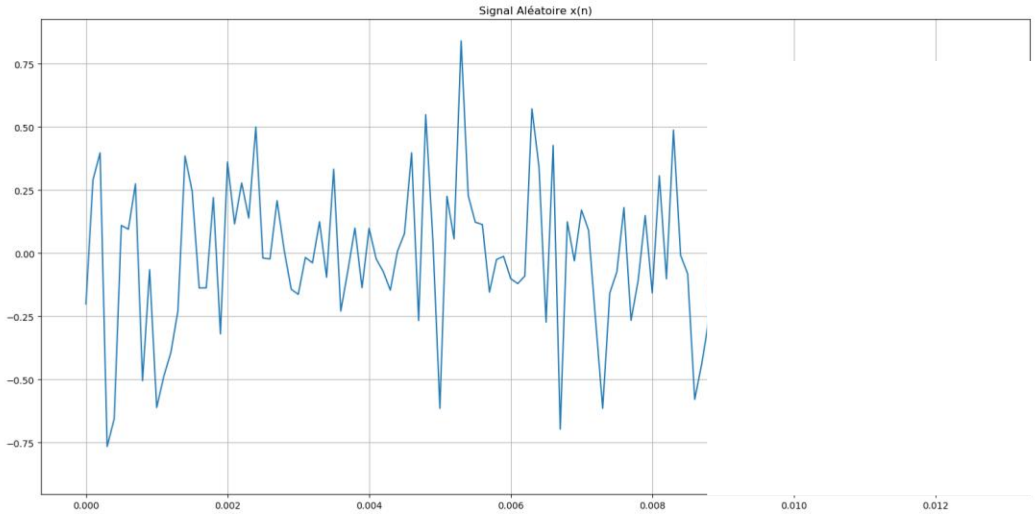
Random Variable Reminders

- A deterministic signal → a mathematical formulation that allows us to know its value at any time.
- Knowing the value of a random signal at time t does not allow us to know its value at $t+\tau$.

- Making predictions by refering

to ***statistical parameters*** defining

the *probabilities* of signal evolution.

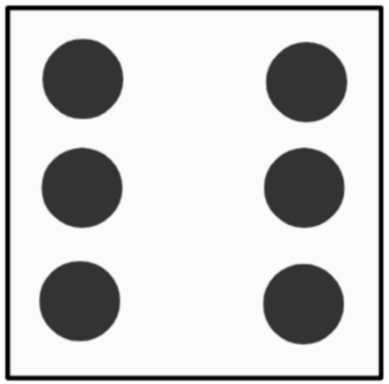


- To study the evolution of **random phenomena** , we use **probabilistic models** .

1.1. Random variables and probabilities

- A probability only makes sense if it is calculated on a large number. Let's imagine that we throw a die 6 times, we will rarely get 6 different faces, or even as many faces for 12 or 60 throws.
- (N >>) that the 6 faces will appear with almost the same probabilities (1/6)

$$p(A) \overset{déf.}{=} \lim_{N \to +\infty} \frac{n_A}{N}$$



- Joint probability

$$p(A, B) \overset{déf.}{=} \lim_{N \to +\infty} \frac{n_{AB}}{N}$$

1.1. Random variables and probabilities

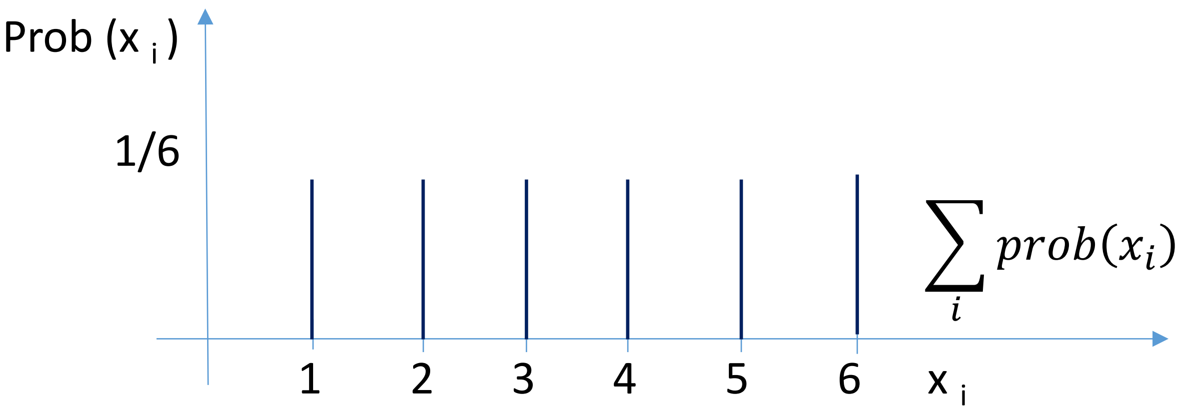
A RV is discrete or continuous: the set of possible values of the random experiment is respectively discrete (x_i) or continuous (x). Nothing to do with time !!

- Number of students present today (discrete)
- Temperature now (continuous)

A random variable is characterized by:

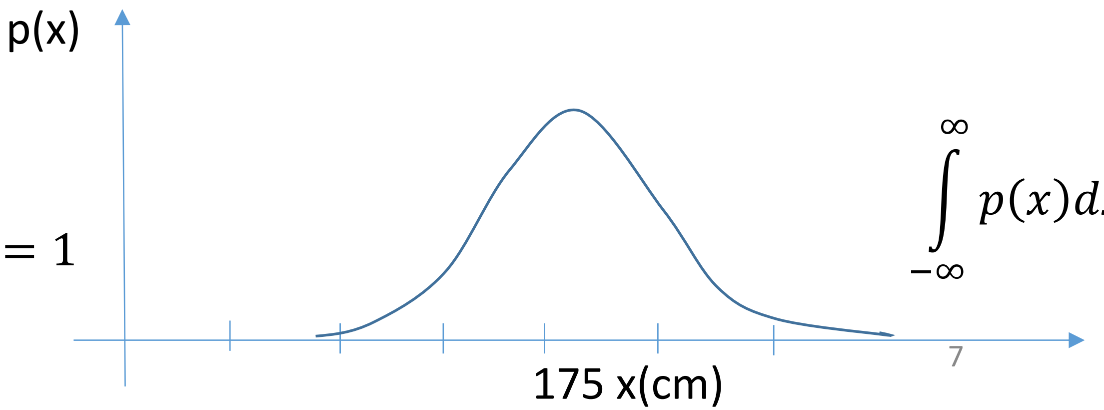
- ✓ the set of values it can take
- ✓ the mathematical expression of the probability density function (PDF) of these values prob ($X=x_i$) or p(x)

Dice Roll



$$\sum_i prob(x_i) = 1$$

Adult Size



$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Faculty of Electrical Engineering, USTHB [akourgli@usthb.dz
http://perso.usthb.dz/~akourgli/

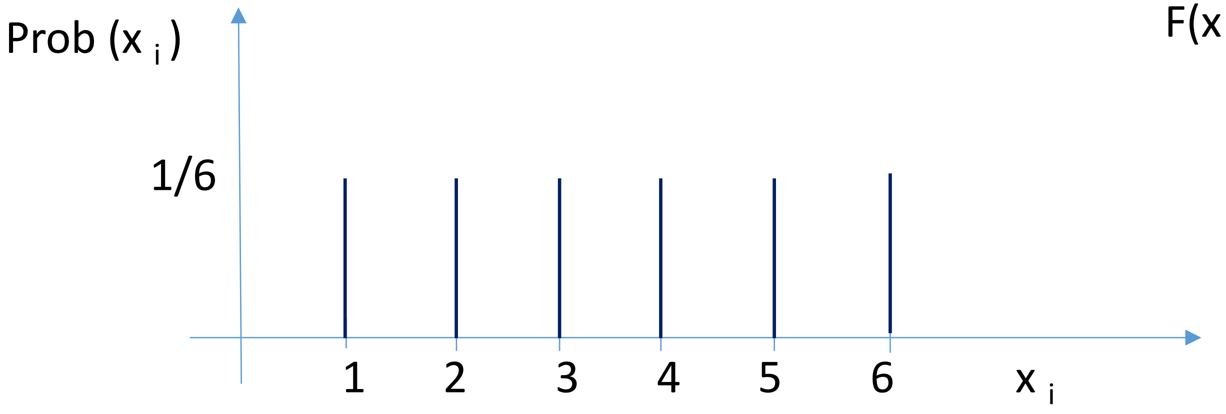
1.1. Random variables and probabilities

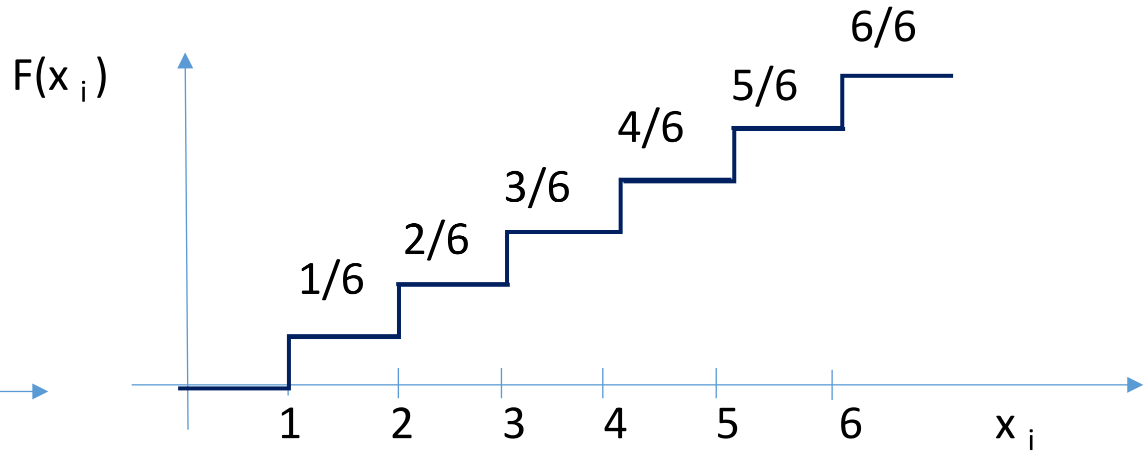
- Cumulative Distribution Function (CFD)

$$F(x_i) = \sum_{j=-\infty}^{i} prob(x_j)U(x - x_j)$$

$$F(x) = \int_{-\infty}^{x} p_x(u)du$$

Example: Rolling the dice





F(- ∞) = 0 F(+ ∞) = 1

1.1. Random variables and probabilities

- Joint probability

$$p(A,B) \overset{déf.}{=} \lim_{N\to+\infty} \frac{n_{AB}}{N}$$

- conditional probability

Conditional probability

$$p(A,B) = \lim_{N\to+\infty}\left(\frac{n_{AB}}{n_A}\frac{n_A}{N}\right) = \lim_{N\to+\infty}\left(\frac{n_{AB}}{n_A}\right)\cdot \lim_{N\to+\infty}\left(\frac{n_A}{N}\right) \overset{déf}{=} p(B/A)\cdot p(A)$$

$$p(A,B) = \lim_{N\to+\infty}\left(\frac{n_{AB}}{n_B}\frac{n_B}{N}\right) = \lim_{N\to+\infty}\left(\frac{n_{AB}}{n_B}\right)\cdot \lim_{N\to+\infty}\left(\frac{n_B}{N}\right) \overset{déf}{=} p(A/B)\cdot p(B) = p(B/A)\cdot p(A)$$

Bayes' Theorem

- Independence

Example: Probability that it rains in Algiers (Event A) knowing that it is sunny in Rio (Event B) → *p(A/B)=p(A)* likewise *p(B/A)=p(B)* hence ***p(A,B)=p(A).p(B)***

A and B are independent

1.2. Statistical moments

✓ **Insufficient** probability density or distribution function

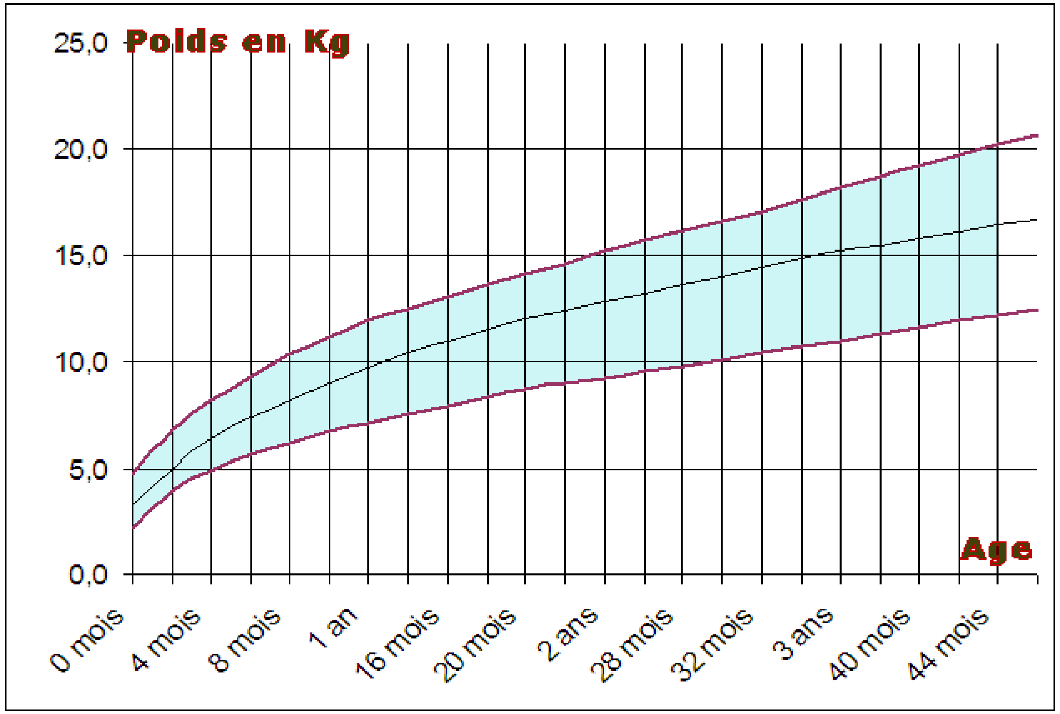
✓ Quantify **dispersion** of the random variable across:

- **Mean**
- **Standard deviation,**
- **etc.**

Example 1 : Monitoring an infant's weight

Weight of 3 children of 1 year:

- 11 kg
- 15 kg
- 6 kg



Example 2 : Choosing a resistor (tolerance on its value ±10%, ±5%, , ±2%,)

1.2. Statistical moments

✓ For a discrete random variable

$$m = \mu_x = \sum_{i}^{N} x_i p(x_i)$$

$$Var = \sigma_x{}^2 = \sum_{i}^{N} (x_i - \mu_x)^2 p(x_i) = \sum_{i}^{N} x_i{}^2 p(x_i) - \mu_x{}^2$$

✓ **x** a continuous RV defined on [a,b]

$$m = \mu_x == E\{x\} = \int_a^b x.p(x)dx$$

$$Var = \sigma_x{}^2 = E\{(x-\mu_x)^2\} = \int_a^b (x-\mu_x)^2 p(x)dx = \int_a^b x^2\, p(x) - \mu_x{}^2$$

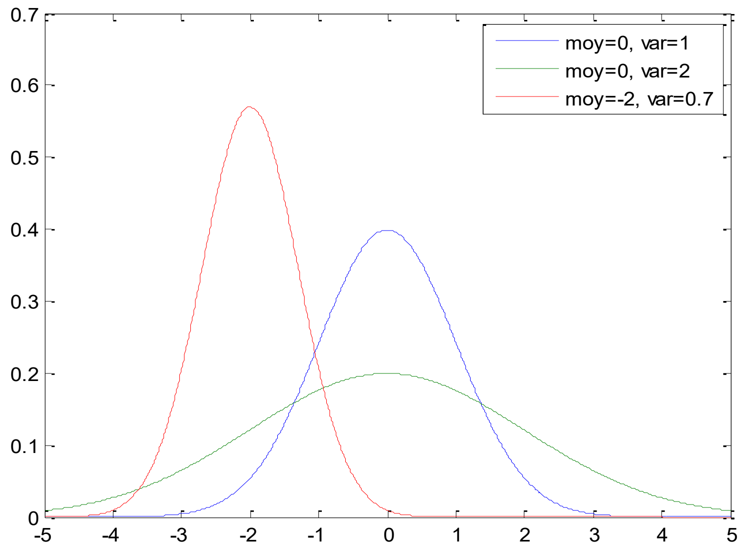
✓ Notation

$$\mu_x = E\{x\} \;\; \sigma_x^2 = E\{(x-\mu_x)^2\}$$

1.3. Common distributions

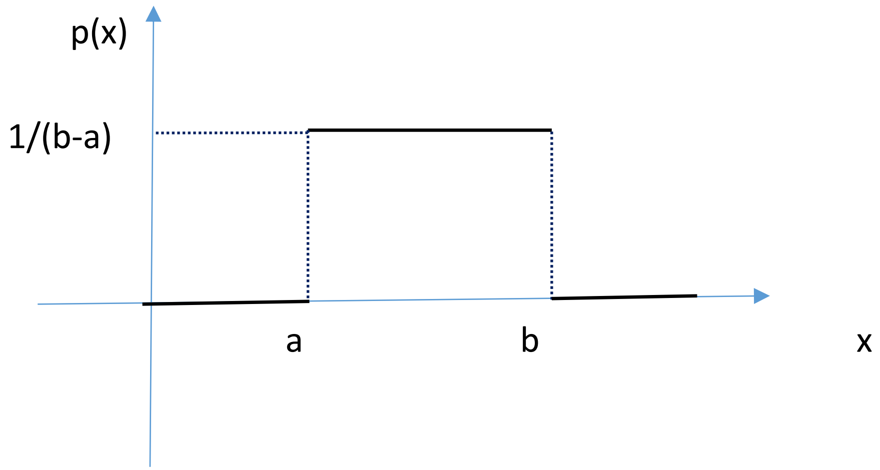
✓ Normal (Gaussian) distribution

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-m)^2}{2\sigma^2})$$



✓ Uniform distribution

$$p_X(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \mu_x = (b+a)/2, \\ 0 ailleurs \sigma_x{}^2 = (b-a)^2/12 \end{cases},$$



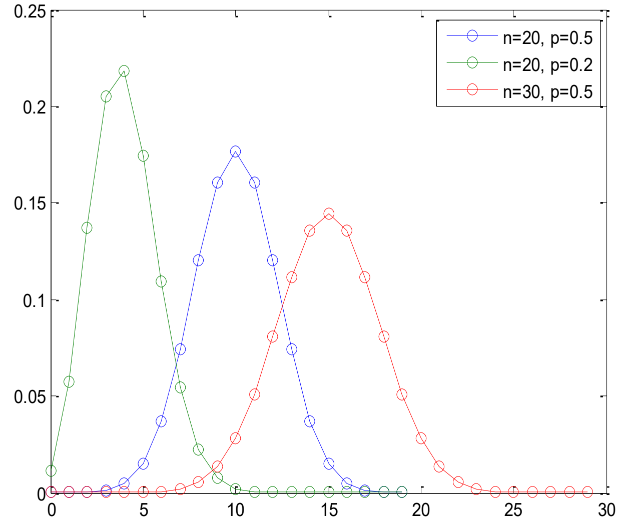
✓ Bernoulli's distribution

pX (x)=p . δ(xa) + (1-p) . δ(xb) *F(x) = p. U(xa) + (1-p) . U(xb)*

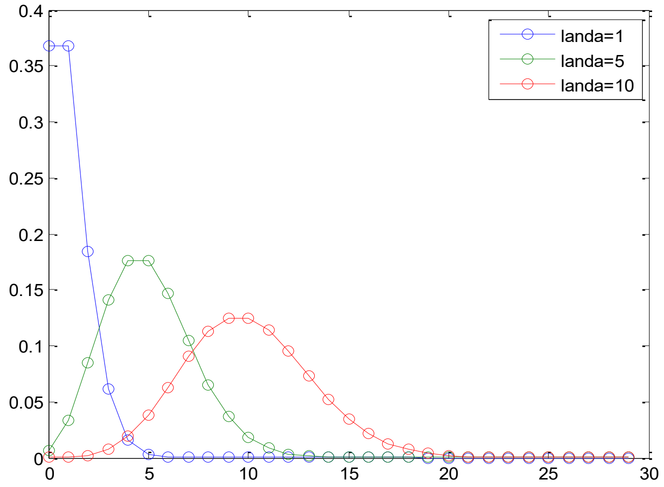
1.3. Common distributions

✓ Binomial distribution

$$p_X(x) = C_n^x p^x (1-p)^{(n-x)} \qquad C_n^k = \frac{n!}{k \rightleftarrows !\,(n-k)!}$$



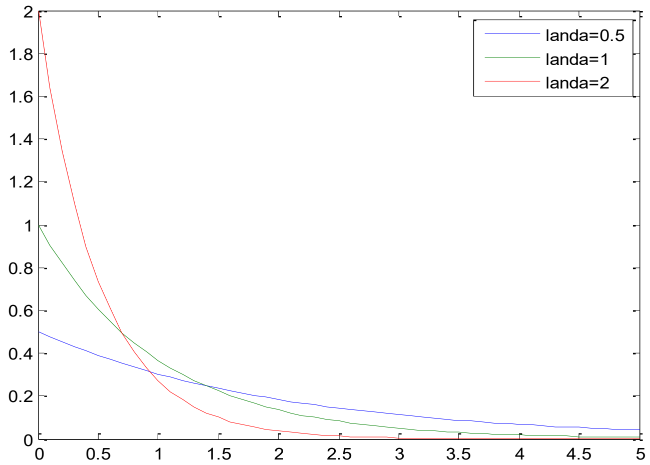
✓ Poisson's distribution



$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \lambda > 0, x \geq 0$$

$$\mu_x = \lambda,\ {\sigma_x}^2 = \lambda$$

✓ Exponential distribution

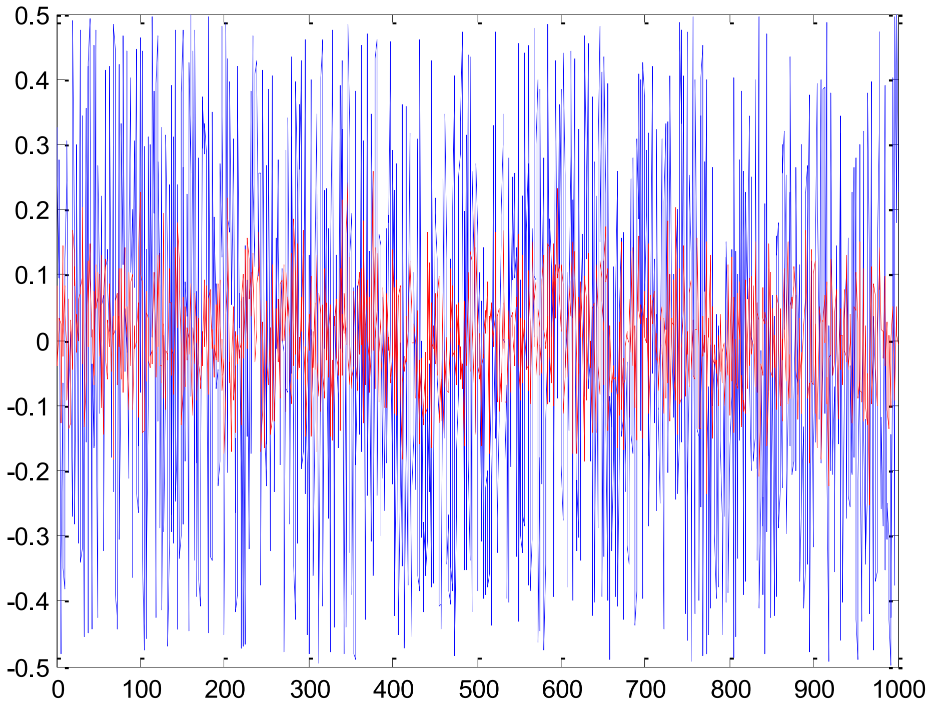


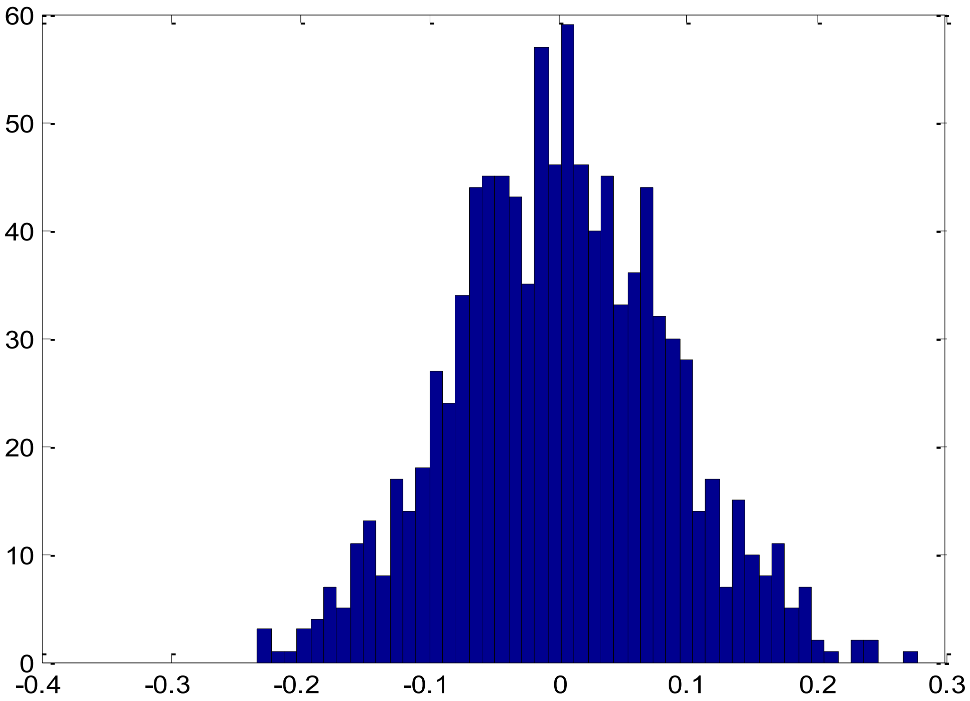
$$p(x) = \begin{cases} 0x < 0 \\ \lambda e^{-\lambda x}\ x \geq 0 \end{cases}$$

$$\mu_x = 1/\lambda,\ {\sigma_x}^2 = 1/\lambda^2$$

1.4. Central limit theorem

The statistical distribution of the sum of n independent RAs with the same distribution tends towards the Normal distribution when n tends towards infinity.





Faculty of Electrical Engineering, USTHB [akourgli@usthb.dz
http://perso.usthb.dz/~akourgli/

1.5. Random Vectors

A multidimensional random variable (also called a random vector) is the dependent result of several random characteristics.

Examples

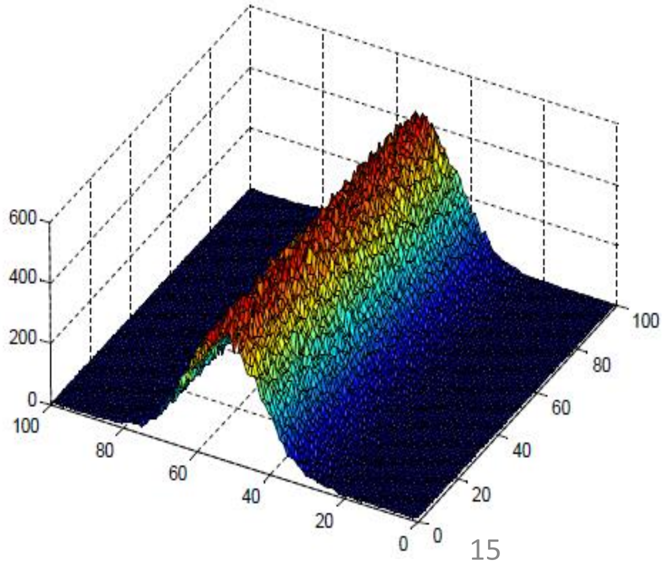
- Roll 2 dice: 2-dimensional discrete random vector
- Weight and height of a person: 2-dimensional continuous random vector
- X is a Gaussian and Y is uniform

Properties

$$\int_{-\infty}^{x1} \cdots \int_{-\infty}^{xn} p_{X1,X2,\ldots,Xn}(x_1, x_2, \ldots\ldots, x_n)\, dx_1 \ldots\ldots dx_n = 1$$

$$F_{X1,X2,\ldots,Xn}(x1, x2, \ldots\ldots, xn) = \int_{-\infty}^{x1} \cdots \int_{-\infty}^{xn} p_{X1,X2,\ldots,Xn}(u_1, u_2, \ldots\ldots, u_n)\, du_1 \ldots\ldots du_n$$

$$p_{X1,X2,\ldots,Xm}(x1, x2, \ldots\ldots, xm) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_{X1,X2,\ldots,Xn}(u_1, \ldots u_m, u_{m+1}, \ldots\ldots, u_n)\, du_{m+1} \ldots\ldots du_n$$



1.5. Random Vectors

A multidimensional random variable (also called a random vector) is the dependent result of several random characteristics.

Examples

- Roll 2 dice: 2-dimensional discrete random vector (x and y)

$$prob(x, y)$$

$$
\begin{array}{c}
\quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{bmatrix}
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36 \\
1/36 & 1/36 & 1/36 & 1/36 & 1/36 & 1/36
\end{bmatrix}
\end{array}
$$

- Generalization: Continuous case

1.5. Random Vectors

Example 1: Discrete RV

$$prob(x,y) = \overbrace{\begin{bmatrix} 1/9 & 1/3 & 0 \\ 2/9 & 1/9 & 2/9 \end{bmatrix}}^{y_1 \quad y_2 \quad y_3}$$

$\Longrightarrow$

prob (x) = [1/9+1/3=4/9; 2/9+1/9+2/9=5/9]

prob (y)= [1/9+2/9=1/3; 1/3+1/9=4/9; 0+2/9=2/9]

Example 2: Let X and Y be two RV uniformly distributed on [0, T]× [0, T],

$$p_{X,Y}(x,y) = \begin{cases} 1/T^2 pour x \in [0,T] et y \in [0,T] \\ 0 ailleurs \end{cases}$$

$\Longrightarrow$

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)dy = \int_0^T 1/T^2 dy = 1/T^2 \int_0^T dy$$
$$= 1/T^2 . y]_0^T = 1/T = p_Y(y)$$

Example 3:

$$p_{X,Y}(x,y) = \begin{cases} 2 pour 0 \leq x \leq 1 0 \leq y \leq x \\ 0 ailleurs \end{cases}$$

$\Longrightarrow$

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)dy = \int_0^x 2dy = 2.x$$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)dx = \int_y^1 2dx = 2(1-y)$$

1.6. Covariance and correlation coefficient

$$C_{x1x2} = \sigma_{x1x2} = E\{(x_1 - \mu_{x1})(x_2 - \mu_{x2})^*\} = E\{x_1 . x_2{}^*\} - \mu_{x1}\mu_{x2}{}^* = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1 x_2{}^* p(x_1, x_2) dx_1 dx_2 - \mu_{x1} . \mu_{x2}{}^*$$

$$\rho_{x1x2} = C_{x1x2}/(\sigma_{x1}\sigma_{x2})$$

- ✓ It measures the degree of linear dependence between X1 and X2. Between -1 and 1.
- ✓ Positive if X1 and X2 vary in the same direction and negative if the two variables vary in opposite directions.
- ✓ If it is zero, X1 and X2 are uncorrelated .

- ✓ If X_1 and X_2 are independent C_{x1x2} =0.
- ✓ The converse is false.

1.6. Covariance and correlation coefficient

- ✓ Matrix notation: $\underline{X}$ the vector of rv . , $\underline{m}_X$ the vector of means $\underline{\mu}_{Xi}$ and $\underline{Xc}$ the vector of centered values :

$$\underline{X} = \begin{bmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{bmatrix}, \underline{\mu}_X = \begin{bmatrix} \mu_{X1} \\ . \\ . \\ . \\ \mu_{Xn} \end{bmatrix}, \underline{X}_C = \begin{bmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{bmatrix} - \begin{bmatrix} \mu_{X1} \\ . \\ . \\ . \\ \mu_{Xn} \end{bmatrix} \Rightarrow C_X = E\{\underline{X}_C . \underline{X}_C{}^T\} = \begin{bmatrix} \sigma_1{}^2 & C_{12} & . & . & C_{1n} \\ C_{21} & \sigma_2{}^2 & . & . & C_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ C_{n1} & C_{n2} & . & . & C_{nn} \end{bmatrix}$$

- ✓ Special case: Gaussian Random Vector

A vector X: (X1,….., Xn) is said to be Gaussian if any linear combination of X1,….., Xn is a Gaussian vector .

$$Y = \sum_{k=1}^{n} \alpha_k X_k = \alpha^T \underline{X}$$

1.7. Gaussian Random Vector

✓ Special case: Gaussian Random Vector

A vector X: (X1,….., Xn) is said to be Gaussian if any linear combination of X1,….., Xn is a Gaussian vector .

$$Y = \sum_{k=1}^{n} \alpha_k X_k = \alpha^T \underline{X}$$

Consequences

- .
$$p(\underline{X}) = \frac{1}{\sqrt{(2\pi)^n \det(C_X)}} e^{-\frac{1}{2}(\underline{X}-\underline{\mu}x)^T C_X{}^{-1}(\underline{X}-\underline{\mu}x)}$$

- A combination of independent Gaussian . follows a Gaussian distribution.

Example :

If Z1 and Z2 are independent random variables with normal distribution with respective parameters (μ1, σ 1) and (μ2, σ 2), then, whatever the real numbers a and b, aZ1 + bZ2 is with normal distribution with parameters (a μ1+b μ2,)

$$\sqrt{a^2\ \sigma_1^2 + b^2\ \sigma_2^2}$$

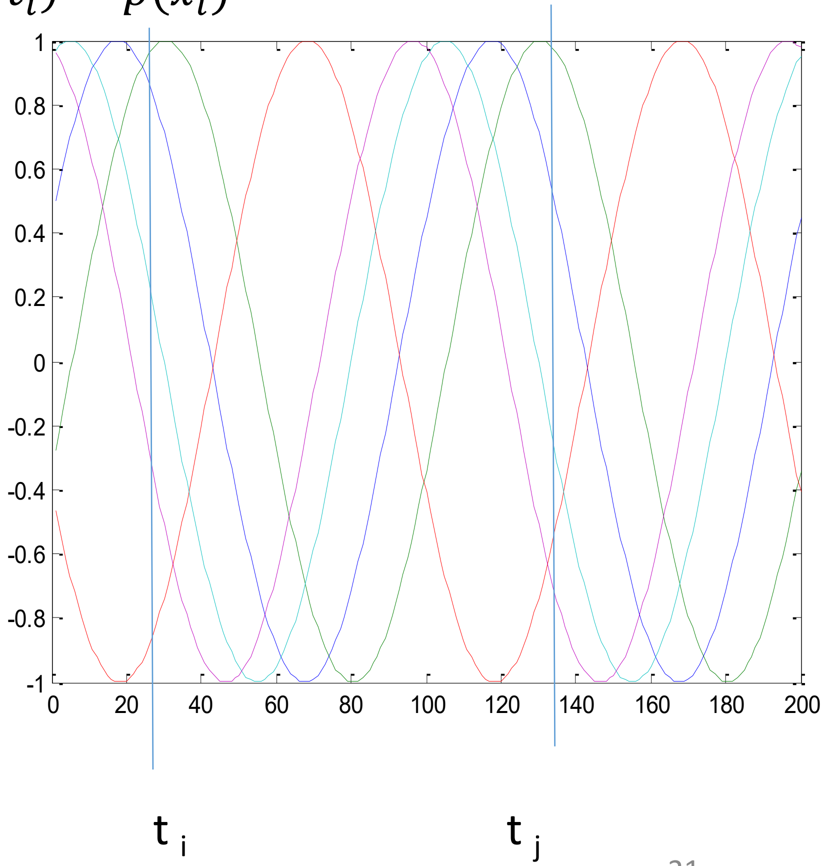
- Decorrelated Gaussian vas are independent

2. Statistical Moments of a Random Process

Example 2

- Random-phase sinusoidal signal $X(t,\phi) = a\ \cos(\omega\ t + \phi)$ with phase φ uniformly distributed between 0 and 2 π
- In practice, it is not easy to obtain the density of Probability
- We are satisfied with moments of order 1 and 2

$$p(x;t_i) = p(x_i)$$



First Order Statistics

- Average
- Variance $\mu_x(t_i) = E\{x(t_i)\} = E\{x_i\} = \int_{-\infty}^{+\infty} x(t_i)\cdot p(x,t_i)dx$

$$\sigma_x^2(t_i) = E\{x(t_i) - \mu_x(t_i)^2\} = E\{x_i^2\} - \mu_x(t_i)^2$$

2. Statistical Moments of a Random Process

Example 1: Let the stochastic process x(t) be defined by: *x(t) = a + bt* where *a* and *b* are RVs whose probabilities are known.

$$\mu_X(t) = E[a + b\ t] = \mu_a + \mu_b\ t$$

$$\sigma^2{}_X(t) = E[(a + b\ t)^2] = \mu_a^2 + \sigma_a^2 + \ t^2(\mu_b^2 + \sigma_b^2) + 2tE[ab]$$

Example 2: Let the stochastic process x(t) be defined by $X(t) = a\ \cos(\omega\ t + \phi)$ with φ uniformly distributed between 0 and 2 π

$$\mu_X(t) = E[X(t)] = \int_0^{2\pi} f_\phi(\phi)\, a \cos(\omega\ t + \phi) d\phi = \frac{1}{2\pi} a \int_0^{2\pi} \cos(\omega\ t + \phi) d\phi = 0$$

$$\sigma_X^2(t) = E[X^2(t)] - \mu_X^2(t) = E[X^2(t)] = \int_0^{2\pi} f_\phi(\phi) a^2 \cos^2(\omega\ t + \phi) d\phi = \frac{a^2}{2}$$

Example 3: Let the stochastic process x(t) be defined by with φ cst and a . with mean μ_a and variance σ_a^2

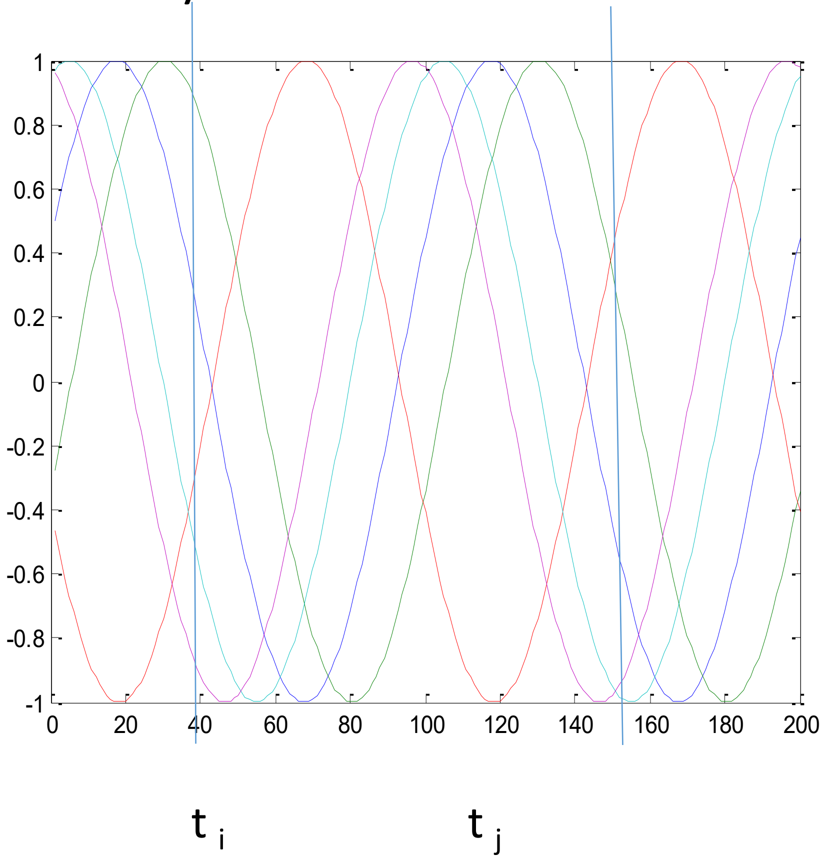
$$\mu_X(t) = E[a \cos(\omega\ t + \phi)] = \mu_a . \cos(\omega\ t + \phi)$$

$$\sigma_X^2(t) = \sigma_a^2 . \cos^2(\omega\ t + \phi)$$

2. Statistical Moments of a Random Process

2nd order statistics

- The process is known at 2 times t_1 and t_2, if $\forall\, t_1, t_2$, the joint probability is known
- **2nd order** statistics → **2 RV** .→ 2 instants → Correlation measure → Memory effect?



- *Statistical autocorrelation* function

$$R_x(t_1,t_2) = E\{x(t_1)\cdot x^*(t_2)\} = E\{x_1\cdot {x_2}^*\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1\cdot {x_2}^*\cdot p(x_1,x_2)\cdot dx_1\cdot dx_2$$

- *Statistical autocovariance* function

$$C_x(t_1,t_2) = E\{[x(t_1)-\mu_x(t_1)]\cdot[x(t_2)-\mu_x(t_2)]^*\} = R_x(t_1,t_2) - \mu_x(t_1)\cdot\mu_x(t_2)^*$$

Purpose : To illustrate the relationship between the statistics taken at **2 different** times t1 and t2 of the process.

2. Statistical Moments of a Random Process

2nd order statistics

Measuring the link between 2 different processes:

- *Statistical cross-correlation* function

$$R_{xy}(t_1,t_2) = E\{x(t_1)\cdot y^*(t_2)\} = E\{x_1 \cdot {y^*}_2\}$$

- *Statistical cross-covariancee* function

$$C_{xy}(t_1,t_2) = R_{xy}(t_1,t_2) - \mu_x(t_1)\cdot {\mu_y}^*(t_2)$$

- *Correlation coefficient*

$$\rho_{xy}(t_1,t_2) = \frac{C_{xy}(t_1,t_2)}{\sigma_x(t_1)\sigma_y(t_2)} \quad -1 \le \rho_{xy}(t_1,t_2) \le 1$$

2. Statistical Moments of a Random Process

Example 1: Let the stochastic process x(t) be defined by: $x(t) = a + bt$ where *a* and *b* are rv whose probabilities are known.

Example 2: Let the stochastic process x(t) be defined by $X(t) = a \cos(\omega t + \phi)$ with φ uniformly distributed between 0 and 2 π

Example 3: Let the stochastic process x(t) be defined by $X(t) = a \cos(\omega t + \phi)$ with φ cst and are random variables with mean μ_a and variance σ_a^2

3. Stationarity

- Many random processes observed in practice have statistical properties that do not depend on the time at which the observation is made → They are **stationary**
- *Stationary* processes are those whose ***statistical characteristics*** are *independent of* ***the origin of time.***

- Stationarity ≠ the process is independent of time
- *Its statistical properties do not depend on the time at which they are estimated.*

✓ temperature is a **non-stationary process**

✓ dice rolling is a **stationary process.**

Stationarity in the strict sense

$$\Leftrightarrow p(x, t_i) = p(x);$$
$$\mu_x(t_i) = \mu_x$$
$$R_{xy}(t_1, t_2) = R_{xy}(\tau) \text{ avec } \tau = t_2 - t_1$$
$$\ldots$$

3. Stationarity

- Many random processes observed in practice have statistical properties that do not depend on the time the observation is made → They are **stationary**
- *Stationary* processes are those whose ***statistical characteristics*** are *independent of* ***the origin of time.***

First Order Stationarity

…. Moments of order 1 $\mu_x(t) = \mu_x = cste$ $\sigma_x^2(t_i) = \sigma_x^2$

Second Order Stationarity

…. Moments of order 2 $R_x(t_1, t_2) = R_x(\tau)$ avec $\tau = t_2 - t_1$

Stationarity in the broad sense : wide-sense stationarity (WSS)

✓ Not always access *p(x,t* $_i$*)* → Be satisfied with

$$\begin{cases} \mu_x(t) = \mu_x = cste \\ R_x(t_1, t_2) = R_x(\tau) \text{ avec } \tau = t_2 - t_1 \end{cases}$$

3. Stationarity

Example 1: Let the stochastic process x(t) be defined by: *x(t) = a + bt* where *a* and *b* are rvs whose probabilities are known.

Stationary step 1

$$\mu_X(t) = E[a + b\ t] = \mu_a + \mu_b\ t$$

Example 2: Let the stochastic process x(t) be defined by $X(t) = a\ \cos(\omega\ t + \phi)$ with φ cst and a rv . of mean μ_a and variance σ_a^2

Stationary of first order 1

$$\mu_X(t) = E[a \cos(\omega\ t + \phi)] = \mu_a . \cos(\omega\ t + \phi)$$

$$\sigma_X^2(t) = \sigma_a^2 . \cos^2(\omega\ t + \phi)$$

3. Stationarity

Example 3: Let the stochastic process x(t) be defined by $X(t) = a\ \cos(\omega\ t + \phi)$ with φ uniformly distributed between 0 and 2 π

$$\mu_X(t) = E[X(t)] = \int_0^{2\pi} f_\phi(\phi)\, a\cos(\omega\ t + \phi)d\phi = \frac{1}{2\pi} a \int_0^{2\pi} \cos(\omega\ t + \phi)d\phi = 0$$

$$\sigma_X^2(t) = E[X^2(t)] - \mu_X^2(t) = E[X^2(t)] = \int_0^{2\pi} f_\phi(\phi) a^2 \cos^2(\omega\ t + \phi)d\phi = \frac{a^2}{2}$$

$$R_{xx}(t_1, t_2) = E[a\cos(\omega t_1 + \phi) a\cos(\omega t_2 + \phi)] = E[\frac{a^2}{2}(\cos(\omega(t_1 + t_2) + 2\phi) + \cos(\omega(t_1 - t_2)))] = \frac{a^2}{2}\cos(\omega(t_1 - t_2))$$

Stationary of order 1 and 2 WSS

Example 4: Show that the process x(t) = a.cos (2πf $_0$ t) + b.sin (2πf $_0$ t) where a, b are variables uncorrelated with zero mean and unit variance is stationary in the broad sense.

3. Stationarity

Properties of the autocorrelation function for real random x WSS

✓ The correlation function is even

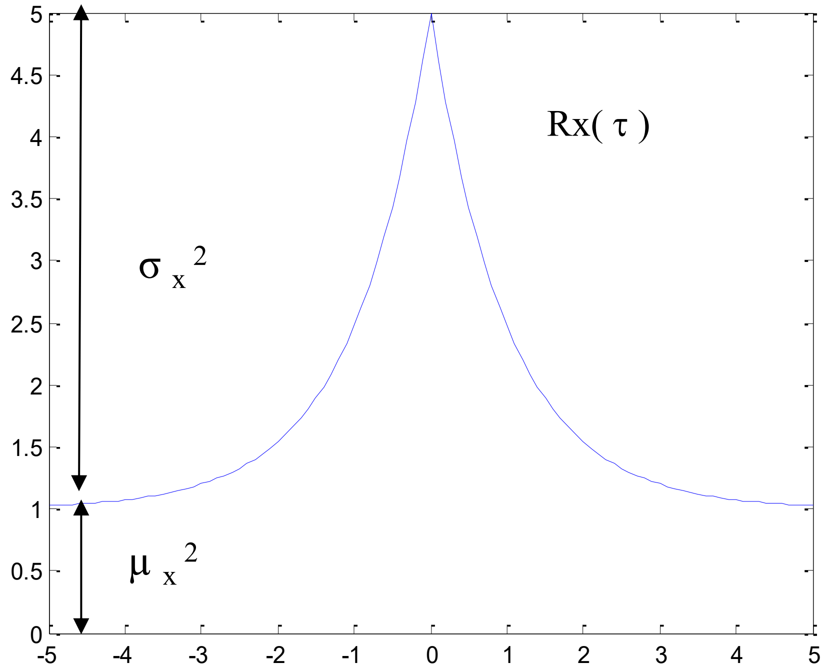
$$R_x(\tau) = R_x(-\tau) \text{ avec } \tau = t_2 - t_1$$

✓ The function is maximum at the origin

$$R_x(0) \geq |R_x(\tau)|$$

✓ Values at origin and infinity:

$$R_x(0) = E(x(t)^2) = \mu_x{}^2 + \sigma_x{}^2$$



if *x(t)* contains neither a periodic component nor a time-independent component, when t→> ∞ , the variables *x(t)* and *x(t + τ)* become statistically uncorrelated .

$$\lim_{\tau\to\infty} R_x(\tau) = \lim_{\tau\to\infty} E\{x(t)x(t-\tau)\} = \mu_x{}^2$$

3. Stationarity

WSS: Discrete Case

We will replace *t* pat *n* ,

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

$$R_x(n_1,n_2) = E\{x(n_1)x^*(n_2)\}$$

$$R_x(n_1,n_2) = \begin{bmatrix} R_x(1,1) & R_x(1,2) & R_x(1,3) & \dots\dots\dots & R_x(1,N) \\ R_x(2,1) & R_x(2,2) & R_x(2,3) & \dots\dots\dots & R_x(2,N) \\ R_x(3,1) & R_x(3,2) & R_x(3,3) & \dots\dots\dots & R_x(3,N) \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \\ R_x(N,1) & R_x(N,2) & R_x(N,3) & \dots\dots\dots & R_x(N,N) \end{bmatrix}$$

A discrete WSS random process X will have a constant mean and a statistical autocorrelation depending only on $k = n_1 - n_2$

WSS → $\mu_x(n) = \mu_x = cste et R_x(n_1,n_2) = R_x(k)$ avec $k = n_1 - n_2$

Toeplitz

$$R_x(n_1,n_2) = R_x(n_1-n_2) = \begin{bmatrix} R_x(0) & R_x(1) & R_x(2) & \dots\dots\dots & R_x(N-1) \\ R_x(1) & R_x(0) & R_x(1) & \dots\dots\dots & R_x(N-2) \\ R_x(2) & R_x(1) & R_x(0) & \dots\dots\dots & R_x(N-3) \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \\ R_x(N-1) & R_x(N-2) & R_x(N-3) & \dots\dots\dots & R_x(0) \end{bmatrix}$$

3. Stationarity

WSS: Discrete Case

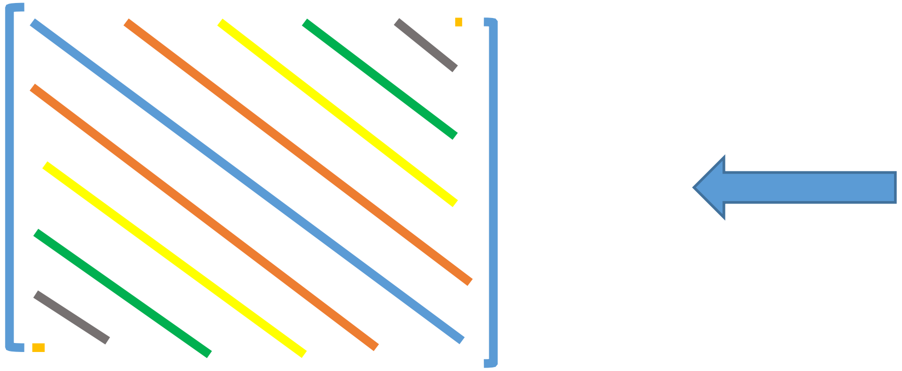
A discrete WSS random process X will have a constant mean and a statistical autocorrelation depending only on $k = n_1 - n_2$

WSS $\Rightarrow$

$$\mu_x(n) = \mu_x = cste \; et \; R_x(n_1, n_2) = R_x(k) \text{ avec } k = n_1 - n_2$$

Toeplitz

$$R_x(n_1, n_2) = R_x(n_1 - n_2) = \begin{bmatrix} R_x(0) & R_x(1) & R_x(2) & \ldots\ldots\ldots\ldots & R_x(N-1) \\ R_x(1) & R_x(0) & R_x(1) & \ldots\ldots\ldots\ldots & R_x(N-2) \\ R_x(2) & R_x(1) & R_x(0) & \ldots\ldots\ldots\ldots & R_x(N-3) \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots & \\ R_x(N-1) & R_x(N-2) & R_x(N-3) & \ldots\ldots\ldots\ldots & R_x(0) \end{bmatrix}$$



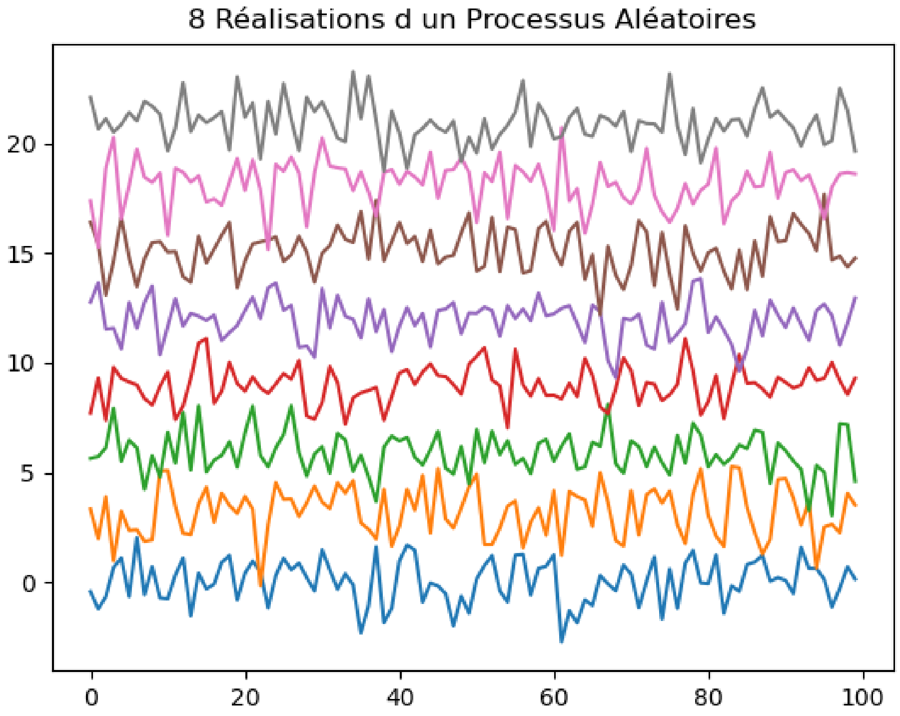
3. Stationarity

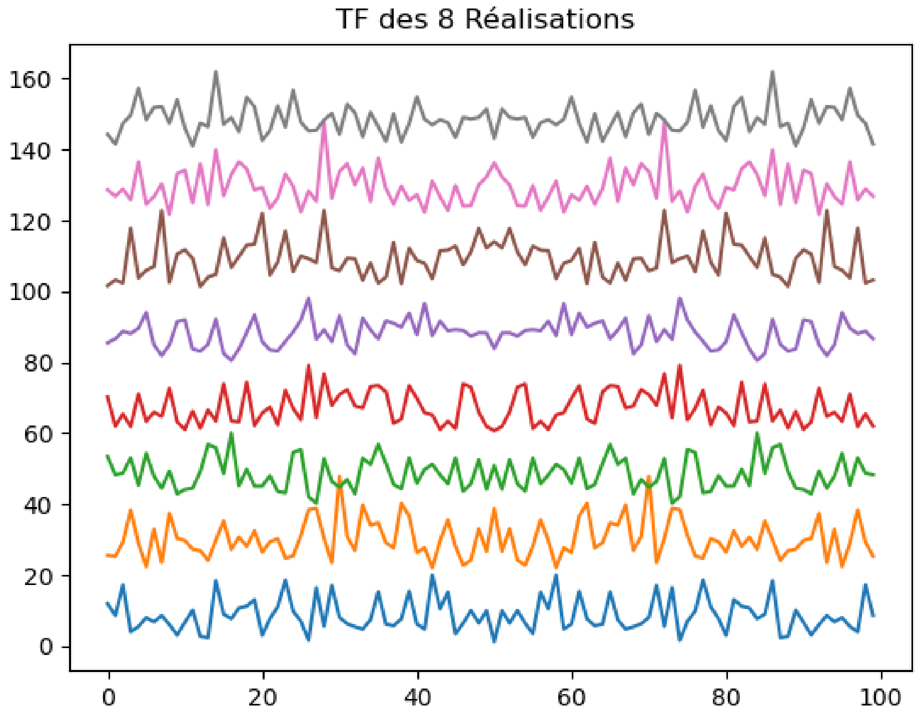
Power and PSD

$$P_X = E(x(t)^2) = R_x(0) = \mu_x{}^2 + \sigma_x{}^2$$

$$E_X = \int P_X dt = \int E(x(t)^2) dt$$

- Each realization will provide a different spectrum





3. Stationarity

Power and PSD

- **Solution:** Consider the power spectral density of a random signal X(t) WSS

Let the Fourier Transform of Rx (τ)

$$S_x(f) = \int_{-\infty}^{+\infty} R_x(\tau)\ e^{-2\pi j f \tau} d\tau \qquad P_x = \int_{-\infty}^{+\infty} S_x(f)\ df = R_x(0)$$

- Provide information on the **frequency distribution of the average signal power** .

White Noise

It is a second-order stationary random signal whose power spectral density is constant along the entire frequency axis.

$$S_x(f) = \sigma_X^2 \Rightarrow$$

$$R_x(\tau) = \sigma_x^2 \delta(\tau)$$ Continuous

$$R_x(k) = \sigma_x^2 \delta(k)$$ discrete

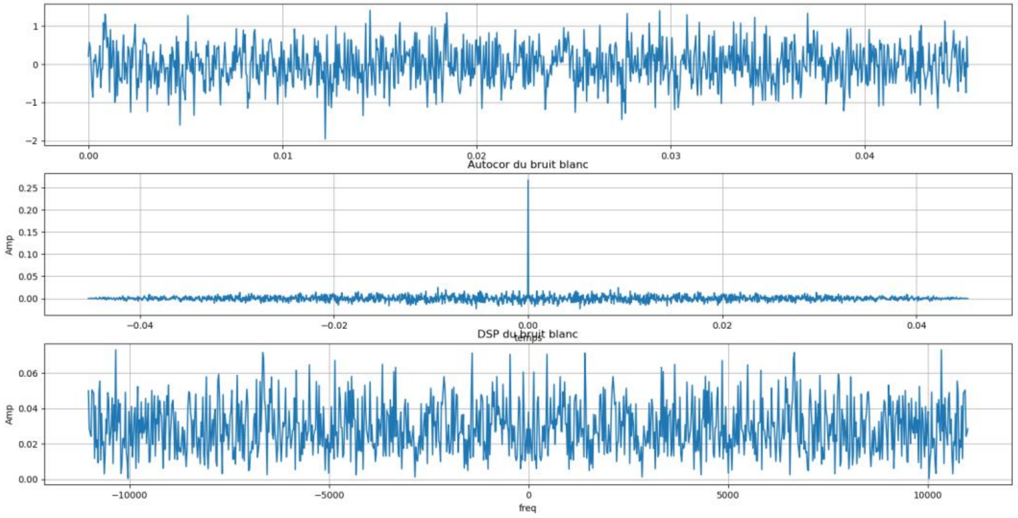
3. Stationarity

White Noise

- Analogy with White Light

$$S_x(f) = \sigma_X^2 \quad \left\{ \begin{array}{l} R_x(\tau) = \sigma_x^2 \delta(\tau) \\ * \\ R_x(k) = \sigma_x^2 \delta(k) \end{array} \right.$$

$$\Rightarrow \quad R_x(k) = 0 pour k \neq 0$$



4. Ergodicity

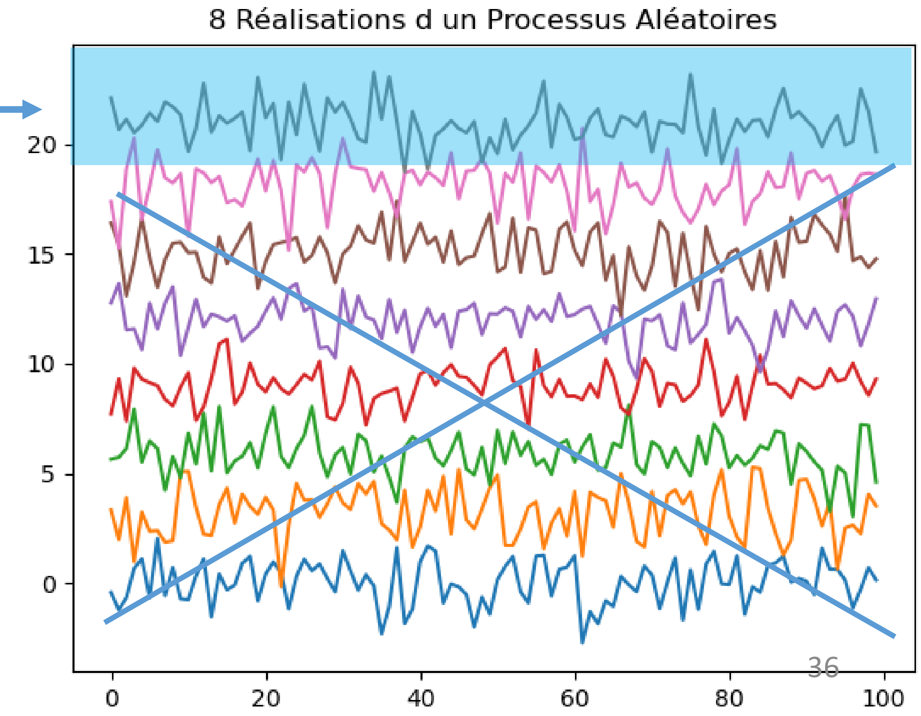
- It is not always possible to carry out a sufficient number of measurements to establish the statistical properties of a random process.
- Easier to roll a die 1000 times than to requisition 1000 people to roll 1000 dice.
- Assimilate the results obtained on a realization to those obtained for a given instant t_i for different realizations → Use **temporal properties** to **replace non-calculable statistical properties** due to the non-availability of a random process.

$$\overline{\mu_x} = \lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \cdot dt \Rightarrow \mu_x = \overline{\mu_x}$$

$$\overline{R_x(\tau)} = \lim_{T \to +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)x(t-\tau) \cdot dt \Rightarrow R_x(\tau) = \overline{R_x(\tau)}$$

$$S_x(f) = TF\{R_x(\tau)\} = \lim_{T \to \infty} \frac{1}{T} |X_T(f)|^2$$

Only 1 realization available



5. Estimators of statistical moments

- The quantities $E\{x\}$, $E\{x^2\}$, , $Sxx(f)$,... are impossible to calculate on a computer because they would require an infinite number of points.

- On the computer, we only have a discrete and finite sequence of *N* points.

- Estimates of these quantities are calculated by generally assuming the signal to be stationary and erVAdic and then replacing the calculation of statistical averages with time averages.

5. Estimators of statistical moments

A. Estimated mean

Let N samples (x_0, x_1, ...,x_{N-1}) be independent and identically distributed (same distribution with same parameter) of a stationary random signal:

$$\hat{m} = \frac{1}{N}\sum_{i=0}^{N-1} x_i \qquad E\{\hat{m}\} = \frac{1}{N}\sum_{i=0}^{N-1} E\{x_i\} = \frac{1}{N}.N.m = m \quad \Rightarrow \quad b\{\hat{m}\} = 0$$

B. Variance estimation

Let N samples (x_0, x_1, ...,x_{N-1}) be independent and identically distributed (same distribution with same parameter) of a stationary random signal:

$$\hat{\sigma}_1{}^2 = \frac{1}{N}\sum_{k=0}^{N-1}(x_k - m)^2 \qquad \hat{\sigma}_2{}^2 = \frac{1}{N-1}\sum_{k=0}^{N-1}(x_k - \hat{m})^2 \qquad \hat{m} = \frac{1}{N}\sum_{i=0}^{N-1} x_i$$

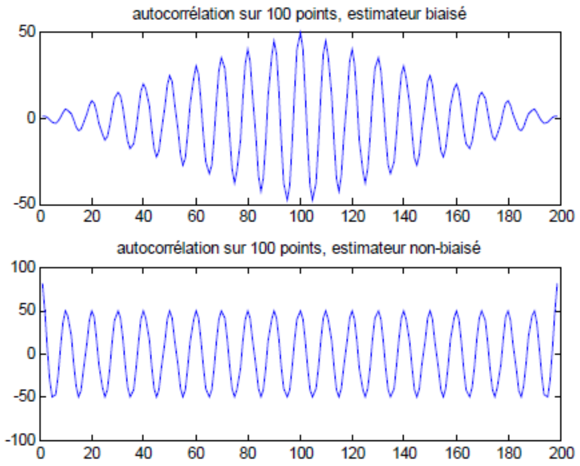
5. Estimators of statistical moments

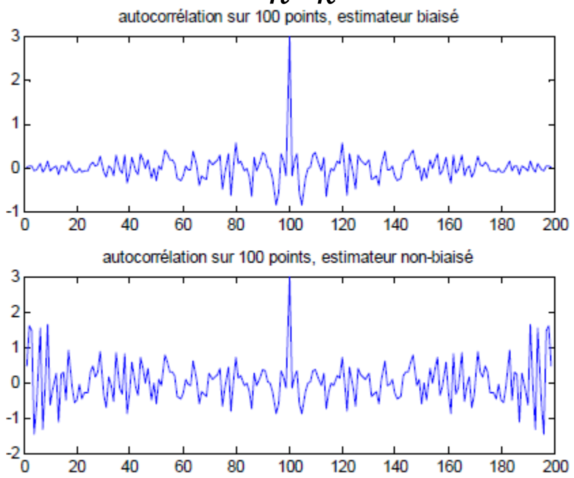
C. Estimation of Correlation

Let N samples ($x_0, x_1, \ldots, x_{N-1}$) be independent and identically distributed (same distribution with same parameter) of a stationary random signal:

$$\hat{R}_{xx}(k) = \frac{1}{N}\sum_{n=k}^{N-1} x(n)x*(n-k) \qquad\qquad \hat{R}_{xx}(k) = \frac{1}{N-k}\sum_{n=k}^{N-1} x(n)x*(n-k)$$

$$E\{\hat{R}_{xx}(k)\} = \frac{1}{N-k}\sum_{n=k}^{N-1} E\{x(n)x*(n-k)\} = \frac{1}{N-k}\sum_{n=k}^{N-1} R_{xx}(k) = R_{xx}(k)$$





5. Estimators of statistical moments

D. Estimation of Power Spectral Density

- Let N samples (x_0, x_1, …..,x_{N-1}) be independent and identically distributed (same distribution with same parameter) of a stationary random signal.
- So let the sequence y_k be obtained from the sequence x_k weighted by the window f_k :

$$y_k = x_k . f_k$$

- The Fourier transform calculated on the finite sequence will therefore be convolved with the spectrum of the rectangular window, i.e. a cardinal sine.
- **But** the spectral properties of the cardinal sinus are not well suited to spectral analysis of the signal (large side lobes).
- To remedy this, we use weighting windows (Hamming , Hanning , Kaiser, etc.)

5. Estimators of statistical moments

D. Estimation of Power Spectral Density

1. Periodogram method

- Let N samples (x_0, x_1, …..,x_{N-1}) be independent and identically distributed (same distribution with same parameter) of a stationary random signal.
- So let the sequence y_k be obtained from the sequence x_k weighted by the window f_k : $y_k = x_k . f_k$

$$\hat{S}_{xx}(f) = \frac{1}{N}|Y(f)|^2 \; avec Y(f) = \sum_{k=0}^{N-1} y_k e^{-2\pi jfk}$$

$$\hat{S}_{xx}(f) = \frac{1}{N}\sum_{l=0}^{N-1} y_l e^{-2\pi jfl} \sum_{m=0}^{N-1} y_m e^{2\pi jfm} \frac{1}{N}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1} y_l y_m e^{-2\pi jf(l-m)}$$

$$= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{l=k}^{N-1} y_l y_{l-k} e^{-2\pi jfk} = \sum_{k=0}^{N-1} \hat{R}_{yy}(k) e^{-2\pi jfk} = TF\left(\hat{R}_{yy}(k)\right)$$

5. Estimators of statistical moments

D. Estimation of Power Spectral Density

1. Periodogram method

- Let N samples ($x_0, x_1, \ldots, x_{N-1}$) be independent and identically distributed (same distribution with same parameter) of a stationary random signal.
- So let the sequence y_k be obtained from the sequence x_k weighted by the window f_k : $y_k = x_k . f_k$

$$E\{\hat{S}_{xx}(f)\} = E\left\{\sum_{k=0}^{N-1} \hat{R}_{yy}(k) e^{-2\pi jfk}\right\} = E\left\{\frac{1}{N}\sum_{k=0}^{N-1}\sum_{l=k}^{N-1} y_l y_{l-k} e^{-2\pi jfk}\right\}$$

$$E\{\hat{S}_{xx}(f)\} = \frac{1}{N}\sum_{k=0}^{N-1}\sum_{l=k}^{N-1} E\{y_l y_{l-k}\} e^{-2\pi jfk} = \sum_{k=0}^{N-1} \frac{N-k}{N} R_{yy}(k)\, e^{-2\pi jfk} = S_{yy}(f) * N \sin c\,(fN)^2$$

5. Estimators of statistical moments

D. Estimation of Power Spectral Density

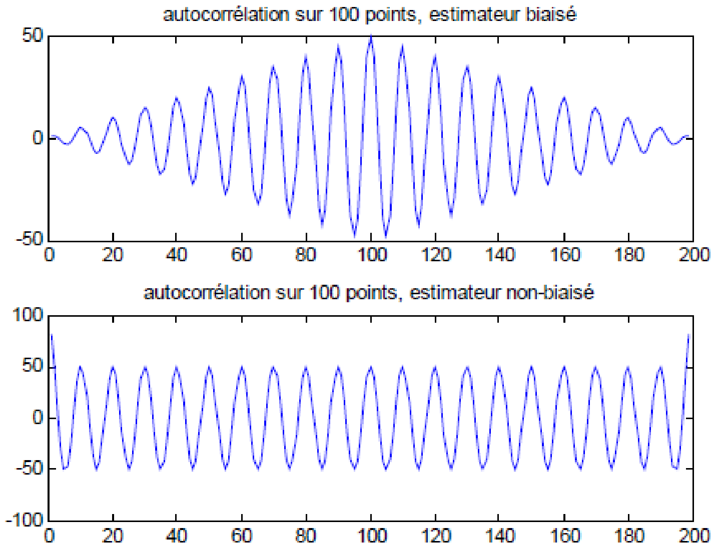
- The periodogram is a biased estimator. When $N \to \infty$, the bias becomes zero. The variance is practically independent of N and proportional to the spectrum: ⇨ $\simeq Sx(f)^2$
- To reduce the variance of this estimator, use an averaged periodogram (Separate the signal into K slices (of length *N/K*), calculate the periodogram on each slice and take the average)
- Due to the K averages , the variance is almost divided by K: however, the slices being shorter, the resolution decreases ($\Delta f = fe / N \to fe / K$).

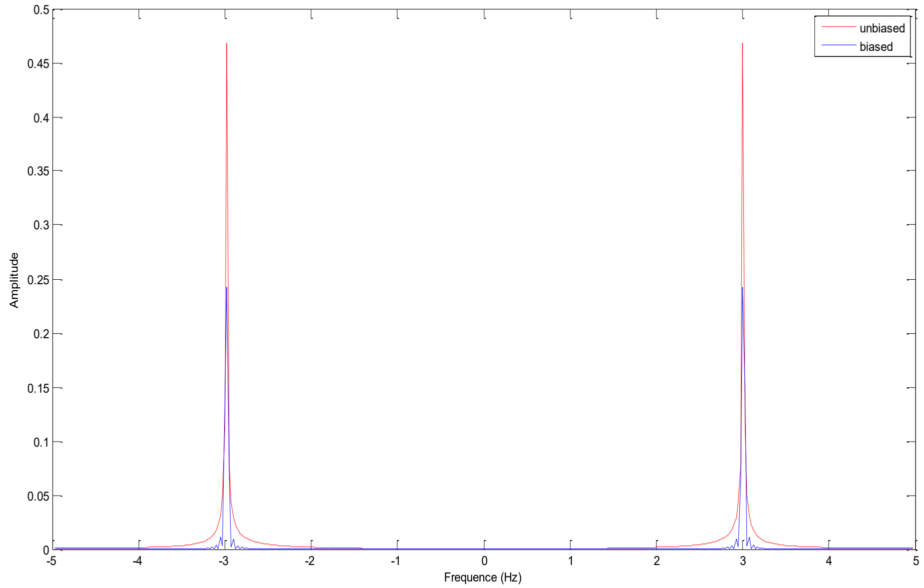
5. Estimators of statistical moments

D. Estimation of Power Spectral Density

Correlogram

First calculate the estimate of *Rxx* (*k*) of the autocorrelation function and then take the TF of this estimate as the estimate of the spectral density.





➢ The unbiased correlogram allows to better highlight the frequency of the sinusoid but at the expense of a greater variance on the edges

5. Estimators of statistical moments

D. Estimation of Power Spectral Density

2. Modeling method

We are looking for an AR, MA or ARMA model for the sequence x_k. In the case of an autoregressive model $x_k = a_1 x_{k-1} + ... + a_r x_{kr} + u_k$

Linear Filtering of Random Signals

6. Random processes and LIT system

A **random signal** is to be **transmitted, analyzed, transformed,** etc.

1. Does it retain its randomness?
2. Its stationarity?
3. What happens to its statistical mean and autocorrelation when linearly filtered?

Aim :

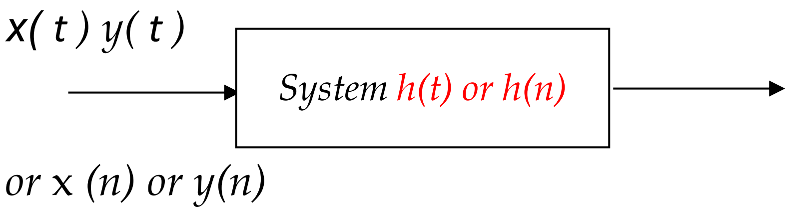
- Examine the transformation of signal characteristics in the frequency domain which will allow us to approach **the notion of shaping filter** .
- 2 direct applications:

1. Optimal and matching filtering.
2. AR, MA and ARMA modeling

6. Random processes and LIT system

A **random signal** is to be **transmitted, analyzed, transformed,** etc.

1. Does it retain its randomness?



$$y(t) = x(t) * h(t) = \int x(\tau)h(t-\tau)d\tau$$

2. What happens to its statistical mean and statistical autocorrelation when linearly filtered?

If x(t) is WSS

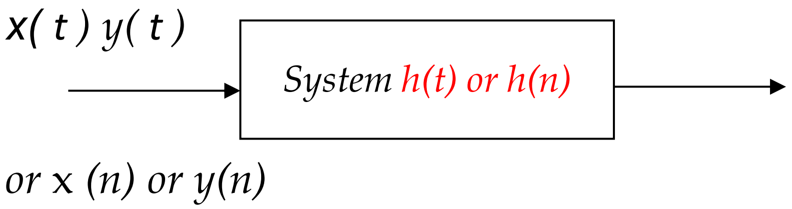
$$\mu_y(t) = E\{y(t)\} = E\{x(t) * h(t)\} = E\left\{\int h(\tau)x(t-\tau)d\tau\right\}$$

$$\mu_y(t) = \int h(\tau)E\{x(t-\tau)\}\, d\tau = \int h(\tau)\mu_x(t-\tau) = h(t) * \mu_x(t)$$

$$\mu_y(t) = \int h(\tau)\mu_x(t-\tau)d\tau = \int h(\tau)\mu_x d\tau = \mu_x \int h(\tau)d\tau = \mu_x H(0) \;\; (f = 0)$$

$$\boxed{\mu_y(t) = \mu_x H(0) \;\; (f = 0)}$$

6. Random processes and LIT system



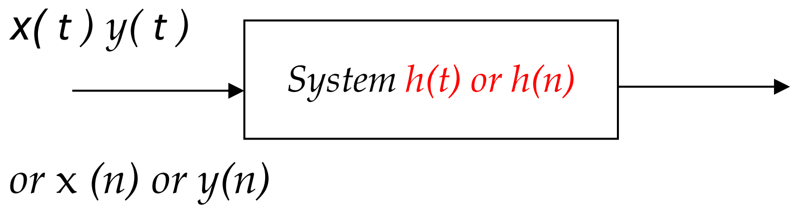
2. What happens to the statistical mean and Statistical autocorrelation during linear filtering?

$$R_y(t_1, t_2) = E\{y(t_1)y^*(t_2)\} = E\{x(t_1) * h(t_1).x^*(t_2) * h^*(t_2)\}$$
$$= E\left\{\int h(\tau_1)x(t_1 - \tau_1)d\tau_1 \int h^*(\tau_2)x^*(t_2 - \tau_2)d\tau_2\right\}$$
$$= \int\int h(\tau_1)h^*(\tau_2)E\{x(t_1 - \tau_1)x^*(t_2 - \tau_2)\}d\tau_1 d\tau_2$$

If x(t) is WSS

$$R_y(t_1, t_2) = \int\int h(\tau_1)h^*(\tau_2)R_x(\tau - \tau_1 + \tau 2)d\tau_1 d\tau_2 = fct(\tau) = R_y(\tau)$$

6. Random processes and LIT system



2. What happens to the statistical mean and Statistical autocorrelation during linear filtering?

If x(t) is WSS

$$R_y(t_1, t_2) = \int\int h(\tau_1)h^*(\tau_2)R_x(\tau - \tau_1 + \tau_2)d\tau_1 d\tau_2 = fct(\tau)$$

$$= R_y(\tau)$$

Reminders

$$\begin{cases} f(\tau) * h(\tau) = \int f(\tau_2)h(\tau - \tau_2)d\tau_2 = \int f(\tau - \tau_2)h(\tau_2)d\tau_2 \\ f(\tau) * h(-\tau) = \int f(\tau_2)h(\tau_2 + \tau)d\tau_2 = \int f(\tau + \tau_2)h(\tau_2)d\tau_2 \end{cases}$$

$\Longrightarrow$ $R_y(\tau) = R_x(\tau) * h(\tau) * h^*(-\tau)$

6. Random processes and LIT system

$x(t)$ $y(t)$

System $h(t)$ or $h(n)$

or $x(n)$ or $y(n)$

2. What happens to the statistical mean and Statistical autocorrelation during linear filtering?

If x(t) is WSS

$$R_y(\tau) = R_x(\tau) * h(\tau) * h^*(-\tau)$$

$$\longrightarrow \quad S_y(f) = TF(R_y(\tau)) = TF(R_x(\tau) * h(\tau) * h^*(-\tau)) = S_x(f).H(f).H^*(f) = |H(f)|^2 S_x(f)$$

$$\boxed{S_y(f) = TF(R_y(\tau)) = |H(f)|^2 S_x(f)} \longrightarrow R_y(\tau) = TF^{-1}(|H(f)|^2 S_x(f)) = \int |H(f)|^2 S_x(f) e^{2\pi j f \tau} df$$

$$E\{y(t)^2\} = R_y(0) = \int |H(f)|^2 S_x(f) df$$

Interference formula

$$\begin{cases} R_{xy}(\tau) = R_x(\tau) * h^*(-\tau) \\ R_{yx}(\tau) = R_x(\tau) * h(\tau) \end{cases}$$

$$\begin{cases} S_{y1y2}(f) = H_1(f) S_{x1x2}(f) H_2^*(f) \\ S_{xy}(f) = S_x(f) H^*(f) \\ S_{yx}(f) = H(f) S_x(f) \end{cases}$$

6. Random processes and LIT system

Example 1

Let the stochastic process x(t) be WSS of statistical correlation $R_x(\tau) = \sigma^2 e^{-\frac{|\tau|}{\theta}}$ and the LIT system of impulse response $h(t)=5e^{-2t}$

- is y(t) WSS?

- Determine its average and its DSP.

6. Random processes and LIT system

Example 1

Let the stochastic process x(t) be WSS of statistical correlation $R_x(\tau) = \sigma^2 e^{-\frac{|\tau|}{\theta}}$ and the LIT system of impulse response $h(t)=5e^{-2t}$

- is y(t) WSS?

- Determine its average and its DSP.

$$\mu_y(t) = \mu_x H(0) = 0 \times \left.\frac{5}{2+2\pi jf}\right|_{f=0} = 0 \times 5/2 = 0$$

$$S_y(f) = |H(f)|^2 S_x(f) = \left|\frac{5}{2+2\pi jf}\right|^2 . \quad TF\left(\sigma^2 e^{\frac{-|\tau|}{\theta}}\right) = \frac{25}{4+(4\pi^2 f^2)} \cdot \frac{2\sigma^2\theta}{1+4\pi^2 f^2 \theta^2}$$

6. Random processes and LIT system

Example 2

If the input signal is white noise → $R_x(\tau) = \sigma^2\delta(\tau)$

- Determine its average and its DSP.

$$\mu_y(t) = \mu_x H(0) \ \ (f = 0)$$

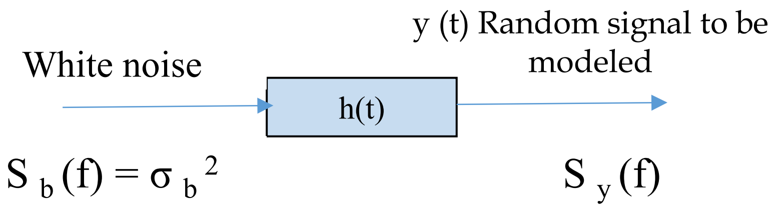
$$S_y(f) = TF(R_y(\tau)) = |H(f)|^2 S_x(f)$$

6. Random processes and LIT system

Notion of shaping filter

If the input signal is white noise → $S_y(f) = \sigma_x^2 |H(f)|^2$ → $|H(f)|^2 = S_y(f)/{\sigma_b}^2$

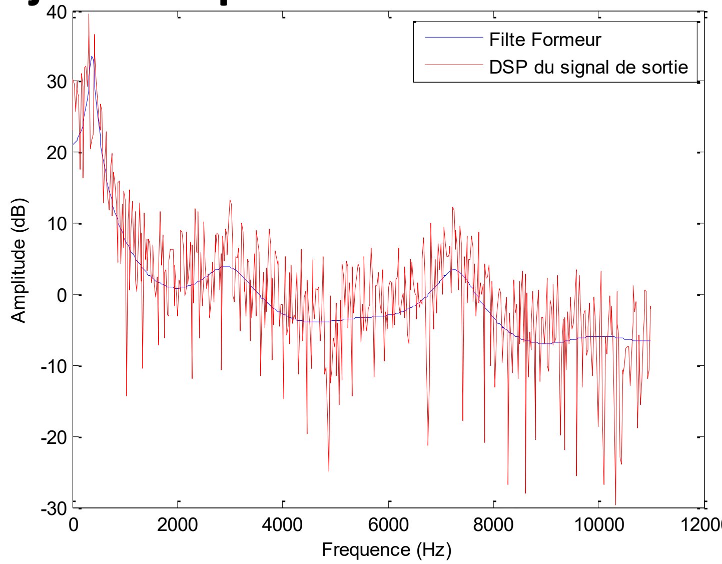
We call the h(t) shaping filter; the filter with transfer function H(f); such that y(t) is generated by passing a white noise b(t) through H(f).



So if we pass the white noise through **a linear and stationary filter with adjustable parameters** and we obtain the desired output signal (to be modeled)

⇒all spectral information is contained

in the **filter represented by its coefficients**



Faculty of Electrical Engineering, USTHB [akourgli@usthb.dz
http://perso.usthb.dz/~akourgli/

7. Optimal and matching filtering

- Transmission of a signal → distortions (due to transmission media)

- Eliminate or at least attenuate before any further treatment.

➢ If we know the original (deterministic) signal, we will speak of detection.

➢ If the signal is random we will use the term estimation.

- Many detection approaches:

Filtering that will enhance the signal drowned in noise.

Etc.

7. Optimal and matching filtering

Hypotheses:

- Consider **a deterministic signal x(t) assumed to be known** , whose possible presence in an observation s(t) we wish to test.
- **The observation noise is assumed to be stationary WSS** with spectral density $S_b(f)$.

→ We are looking for **a filter H(f) which maximizes the SNR** at a precise moment T_0.

- We therefore assume that the useful signal x(t) is drowned in an **additive WSS stationary noise b(t)** , hence: s(t) = x(t)+b(t)

- The signal is filtered by a linear filter whose impulse response is h(t).

At the filter output: $y(t) = s(t) * h(t) = x(t)*h(t)+b(t)*h(t) = x_2(t)+ b_2(t)$

7. Optimal and matching filtering

Hypotheses:

- Consider **a deterministic signal x(t) assumed to be known** , whose possible presence in an observation s(t) we wish to test.
- **The observation noise is assumed to be stationary WSS** with spectral density $S_b(f)$.

→ We are looking for **a filter H(f) which maximizes the SNR** at a precise moment T_0.

7. Optimal and matching filtering

The filter output:

$y(t) = s(t) * h(t) = x(t)*h(t)+b(t)*h(t) = x_2(t)+ b_2(t)$

At time T_0, the SNR is written as $SNR(T_0) = \frac{Puis(x_2(T_0))}{Puis(b_2(T_0))}$

VAAL : Find *h(t)* ,

How? Express the SNR as a function of *h(t)* (or *H(f)*)

- The signal in the numerator $x_2(t)$ is deterministic so its power:

$$Puis(x_2(T_0)) = |x_2(T_0)|^2 = |TF^{-1}(X_2(f))|^2 = \left|\int X(f)H(f)e^{2\pi jfT_0}df\right|^2$$

7. Optimal and matching filtering

The filter output:

y(t) = s(t) * h(t) =x(t)*h(t)+b(t)*h(t) =x_2(t)+ b_2(t)

At time T_0, the SNR is written as $SNR(T_0) = \frac{Puis(x_2(T_0))}{Puis(b_2(T_0))}$

VAAL : Find *h(t)* ,

How? Express the SNR as a function of *h(t)* (or *H(f)*)

- The signal *b_2(t)* comes from the filtering of a random signal *b(t)* WSS so it is also WSS

$$Puis(b_2(T_0)) = E\{|b_2(T_0)^2|\} = R_{b2}(\tau = 0) = TF^{-1}(S_{b2}(f))\Big|_{\tau=0} = \int S_b(f)|H(f)^2|df$$

7. Optimal and matching filtering

$$SNR(T_0) = \frac{Puis(x_2(T_0))}{Puis(b_2(T_0))}$$

$$\begin{cases} Puis(x_2(T_0)) = |x_2(T_0)|^2 = |TF^{-1}(X_2(f))|^2 = \left|\int X(f)H(f)e^{2\pi jfT_0}df\right|^2 \\ Puis(b_2(T_0)) = E\{|b_2(T_0)^2|\} = R_{b2}(\tau = 0) = TF^{-1}(S_{b2}(f))\Big|_{\tau=0} = \int S_b(f)|H(f)^2|df \end{cases}$$

$$\Rightarrow \quad SNR(T_0) = \frac{Puis(x_2(T_0))}{Puis(b_2(T_0))} = \frac{\left|\int X(f)H(f)e^{2\pi jfT_0}df\right|^2}{\int S_b(f)|H(f)^2|df}$$

$$\begin{cases} SNR(T_0) = \frac{\left|\int a(f)b^*(f)df\right|^2}{\int a(f)a^*(f)df} \\ a(f) = \sqrt{S_b(f)}H(f) \\ b(f) = X^*(f)e^{-2\pi jfT0}/\sqrt{S_b(f)} \end{cases}$$

$$SNR(T_0) = \frac{\left|\int a(f)b^*(f)df\right|^2}{\int a(f)a^*(f)df} \leq \int b(f)b^*(f)df$$

Equality if *a(f)=kb(f)* $\Rightarrow$

$$\begin{cases} SNR(T_0) \int \frac{|X(f)|^2}{S_b(f)}df \quad \max \\ H(f)_{optimal} = k.X^*(f)e^{-2\pi jfT0}/S_b(f) \end{cases}$$

7. Optimal and matching filtering

Optimal filtering

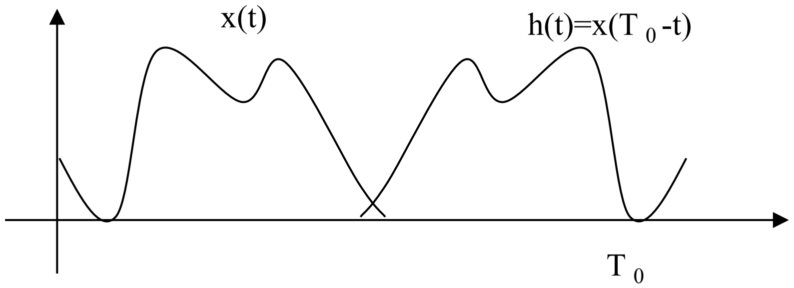
$$\Rightarrow \begin{cases} SNR(T_0) \int \frac{|X(f)|^2}{S_b(f)} df \ _{\max} \\ H(f)_{optimal} = k.X^*(f)e^{-2\pi jfT0}/S_b(f) \end{cases}$$

Special case: The noise is white

$$\begin{cases} H(f)_{adaptél} = k/\sigma_b{}^2 X^*(f)e^{-2\pi jfT0} \\ SNR(T_0)\frac{E_x}{\sigma_b^2}{}_{\max} \end{cases} \Rightarrow h(t)_{adaptél} = k/\sigma_b{}^2 x^*(T_0 - t)$$

✓ The impulse response of the filter represents the useful signal x(t) reversed and shifted by T_0.

✓ matching filtering amounts to performing cross-correlation between the observation and the signal to be detected.



7. Optimal and matching filtering

Example 1: Detecting a pulse

Consider a system emitting a rectangular pulse x(t) of duration T_0 and amplitude A. The additive noise is white.

$$h(t)_{adapt\acute{e}l} = k/{\sigma_b}^2 x^*(T_0 - t)$$

The filter will have the same expression, i.e. h(t)=x(t).

7. Optimal and matching filtering

Example 2: In sonar or radar, we seek to locate a target (boat, plane, etc.)

- A signal x(t) is emitted, which travels the distance *d* to the target, where it will be reflected towards a receiver.
- The receiver then receives the noisy signal *y(t)* , attenuated (by *a*) and delayed by T_{AR}

$y(t) = ax(t - T_{AR}) + b(t) \rightarrow h(t) = x^*(T_0 - t)$

The output of the filter: the cross-correlation of the signals *y(t)* and *x(t)* :

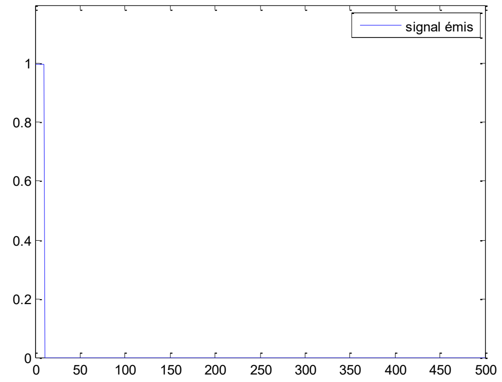
$R_{yx}(t-T_0) = aR_{xx}(t - T_0 - T_{AR}) + R_{bx}(t-T_0)$ the maximum cross-correlation for ⟹ $t = T_0 + T_{AR}$

7. Optimal and matching filtering

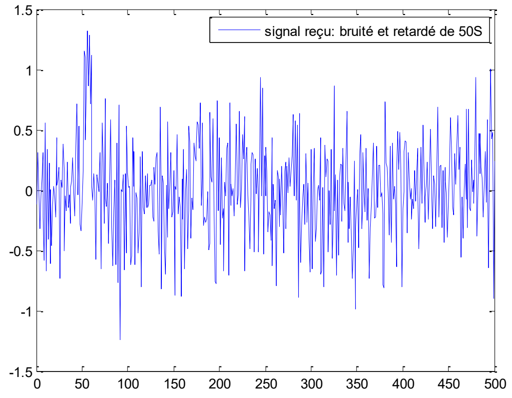
Example 2: In sonar or radar, we seek to locate a target (boat, plane, etc.)

- A signal x(t) is emitted, which travels the distance *d* to the target, where it will be reflected towards a receiver.
- The receiver then receives the noisy signal *y(t)* , attenuated (by *a*) and delayed by T_{AR}

$$y(t) = ax(t - T_{AR}) + b(t)$$



$$\rightarrow h(t) = x^*(T_0 - t)$$

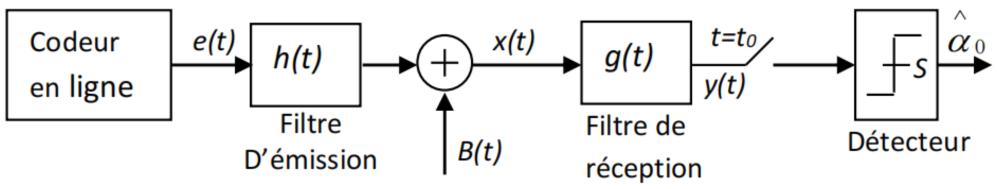


The output of the filter: the cross-correlation of the signals *y(t)* and *x(t)* :

$R_{yx}(t-T_0) = aR_{xx}(t - T_0 - T_{AR}) + R_{bx}(t-T_0)$ the maximum cross-correlation for $\Rightarrow$ $t = T_0 + T_{AR}$

7. Optimal and matching filtering

Transmission d'un symbole sur un canal idéal



$e(t) = \sum_k s_m(t - kT); \quad m = 0,1$

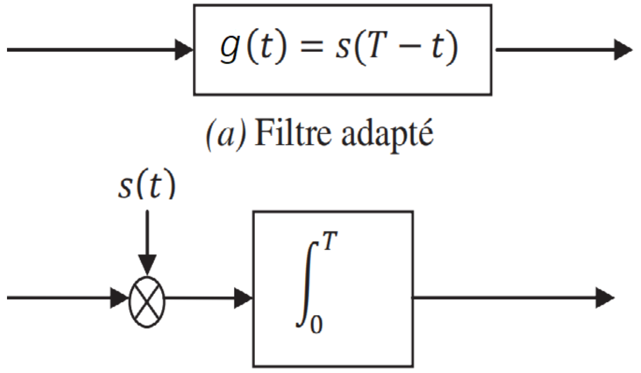
Récepteur optimal

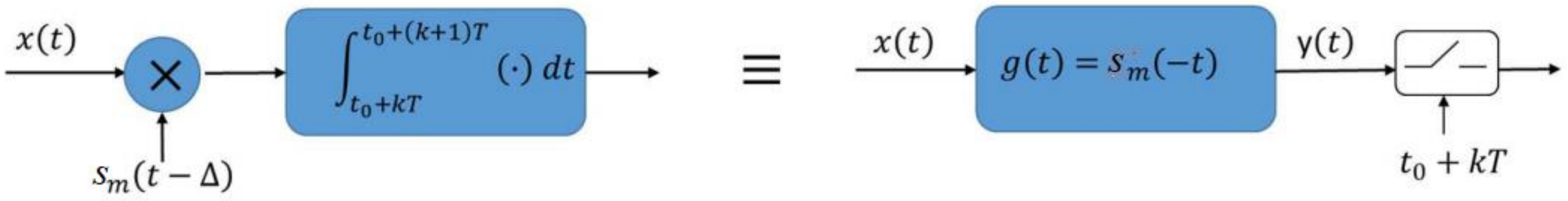
Le filtre g(t) devrait maximiser le rapport signal sur bruit à un instant T

$g(t) = Ks_m(T - t)$ → Le filtre récepteur est le signal émis (Filtrage adapté) et $SNR_T = \frac{2}{N_0}\int_{-\infty}^{\infty} |s_m(t)|^2 dt$

Remarque :

Convolution (On inverse le signal) → Remplacer le filtrage g(t) par une corrélation avec $g(-t) = Ks_m(t - T)$





8. Wiener filtering

- The aim of estimation techniques is to use (random) observations *X* to extract information about the quantity of interest *Y* . Depending on whether *Y* is deterministic or random, different techniques will be used. Thus:

- To estimate **a certain variable** , in the most general case, we want the estimator to be **unbiased** , so we seek the **minimum variance estimator** , the best known of which is the **maximum likelihood (ML**) based on the use of the probability of the observed values *p(X/Y).* If we allow a **bias** , we then seek to **minimize the mean square error** .

Reminders:

✓ **Bias** : Mean of the deviation $E\{\hat{x} - x\}$

✓ **Variance** : Power of the deviation $\sigma_{\hat{x}}^2 = E\{(\hat{x} - E\{\hat{x}\})^2\}$

8. Wiener filtering

- When *Y* the random variable to be estimated is **random , we will use** Bayesian estimators . Several possible scenarios:

➢ When we assume that ***p(Y)* is known , we will use different estimators, such as the maximum a posteriori** (MAP), depending on the cost function to be minimized .

➢ If only the **first and second order moments of *p(Y)* and *p(X*)** *are known, then the* **minimum variance unbiased linear** estimator is often used . **The Wiener filter is an example.**

➢ In the case where we have no **statistical information** on *X* and *Y* and the only one we have is that *X* is a noisy measurement of *fct (Y),* we will adopt **the least squares estimator** to estimate *Y* which is free from any probabilistic framework.

8. Wiener filtering

- It consists of looking for an estimate $\hat{y}$ of y which is a linear function of the observations, that is to say

$$\hat{y} = \sum_{i=1}^{n} h_i \, x_i = h^T x$$

$$\mathrm{E}\{(y-\hat{y})^2\} = \mathrm{E}\{(y-h^T x)^T (y-h^T x)\} = E\{yy^T\} - E\{y^T h^T x\} - E\{hx^T y\} + E\{hx^T h^T x\}$$

We start by deriving with respect to *h* then we cancel the derivative,

$$-E\{y^T x\} - E\{x^T y\} + 2h \quad E\{x^T x\} = 0 \Rightarrow h = E\{y^T x\}/E\{x^T x\} = R_{yx}/R_{xx}$$

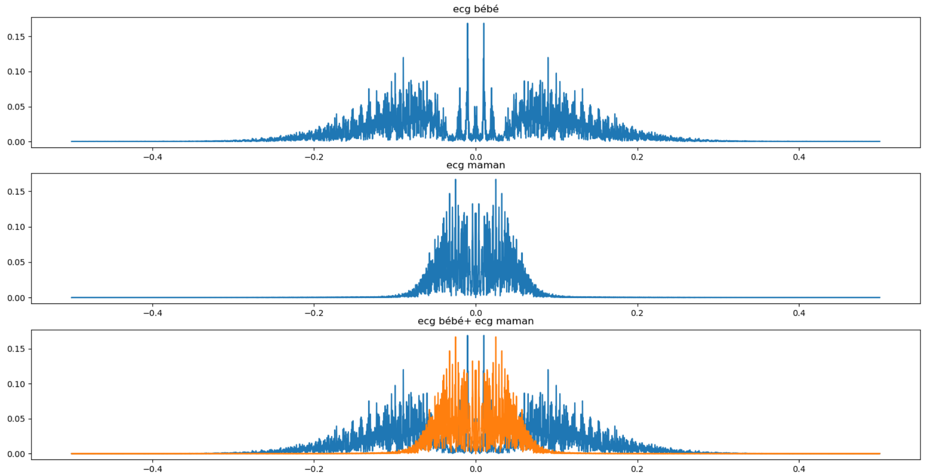
<u>Application: </u>***Wiener filter***

8. Wiener filtering

<u>Application:</u> ***Wiener filter***

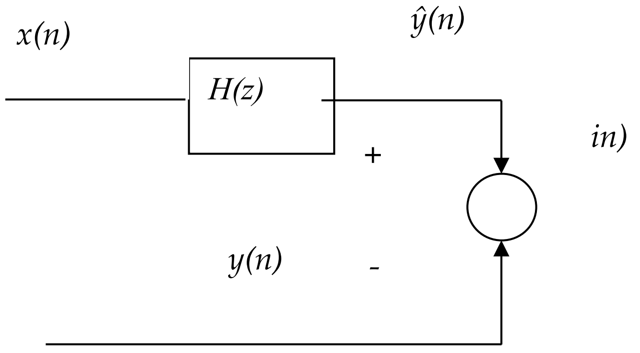
In many applications, time signals are tainted by noise that one wishes to remove or at least reduce.

Since the **random useful signal occupies the same frequency bands as the parasitic signal** , classical filtering cannot be used.



The Wiener filter provides a solution to this problem when the process is stationary.

Wiener filters are said to **be optimum** in the sense of the criterion of **the mean square error** between their output and a desired output.



$$\xi(n) = E\{e^2(n)\}$$

8. Wiener filtering

Application: ***Wiener filter***

y(n) corresponds to the stationary random signal that interests us but is not directly accessible. Only *x(n)* is. $\mathrm{h} = [b_0 \quad b_1 \quad \dots \quad b_{N-1}]^T$

If we assume that the desired filter h is a FIR filter of length N, we can calculate its coefficients by solving a linear system of equations.

The estimated signal $\hat{y}(n) = \sum_{i=0}^{N-1} b_i x(n-i)$

We seek to minimize

$$\xi(n) = E\{e^2(n)\} = E\left\{\left(y(n) - \sum_{i=0}^{N-1} b_i x(n-i)\right)^2\right\}$$

→ derive and cancel the cost/variable function b_i of the impulse response of the filter.

$$\frac{\partial \xi}{\partial b_j} = E\left\{\frac{\partial}{\partial b_j}\{e^2(n)\}\right\} = E\left\{2e(n)\frac{\partial e(n)}{\partial b_j}\right\} = E\left\{2e(n)\frac{\partial}{\partial b_j}\{-b_j x(n-j)\}\right\}$$

$$\frac{\partial \xi}{\partial b_j} = -E\{2e(n)x(n-j)\} = -E\left\{2\left(y(n) - \sum_{i=0}^{N-1} b_i x(n-i)\right)x(n-j)\right\}$$

$$-2R_{yx}(j) + 2\sum_{i=0}^{N-1} b_i R_{xx}(j-i) = 0$$

$$R_{yx}(j) = \sum_{i=0}^{N-1} b_i R_{xx}(j-i)$$

8. Wiener filtering

Application: ***Wiener filter***

y(n) corresponds to the stationary random signal that interests us but is not directly accessible. Only *x(n)* is. $\mathrm{h} = [b_0 \quad b_1 \quad \dots \quad b_{N-1}]^T$

If we assume that the desired filter h is a FIR filter of length N, we can calculate its coefficients by solving a linear system of equations.

The estimated signal
$$\hat{y}(n) = \sum_{i=0}^{N-1} b_i x(n-i)$$

We seek to minimize
$$\xi(n) = E\{e^2(n)\} = E\left\{\left(y(n) - \sum_{i=0}^{N-1} b_i x(n-i)\right)^2\right\}$$

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \dots\dots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \dots\dots & R_{xx}(N-2) \\ & & \dots\dots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \dots\dots & R_{xx}(0) \end{bmatrix} . \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}$$

$$R_{yx}(j) = \sum_{i=0}^{N-1} b_i R_{xx}(j-i)$$

8. Wiener filtering

Example 1

Assume that the observation *x(n)=y(n)+ bb (n)* and that the additive noise *bb (n)* is centered and uncorrelated to the signal. Let us simplify the Wiener- Hopf equations accordingly:

- $R_{yx}(k)=E\{ y(n) [y(n-k)+ bb(n-k)] \}= R_{yy}(k)$
- $R_{xx}(k)=E\{(y(n)+ bb(n))(y(n-k)+ bb(n-k))\}= R_{yy}(k)+ R_{bb}(k)$

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \cdots\cdots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \cdots\cdots & R_{xx}(N-2) \\ & & \cdots\cdots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \cdots\cdots & R_{xx}(0) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}$$

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \cdots\cdots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \cdots\cdots & R_{xx}(N-2) \\ & & \cdots\cdots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \cdots\cdots & R_{xx}(0) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} R_{xx}(0) - R_{bb}(0) \\ R_{xx}(1) - R_{bb}(1) \\ . \\ R_{xx}(N-1) - R_{bb}(N-1) \end{bmatrix}$$

8. Wiener filtering

Example 2

We assume that the observation *x(n) = y(n) + bb (n)*

The signal to be estimated *y(n)* has for the autocorrelation function $R_y(k) = \alpha^{|k|}$ $0 < \alpha < 1$

It is uncorrelated from white noise *bb (n*) of variance σ_b^2.

Let us find h(n) such that $H(z) = b_0 + b_1 z^{-1}$.

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}$$

$$\begin{bmatrix} 1+\sigma_b^2 & \alpha \\ \alpha & 1+\sigma_b^2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix}$$

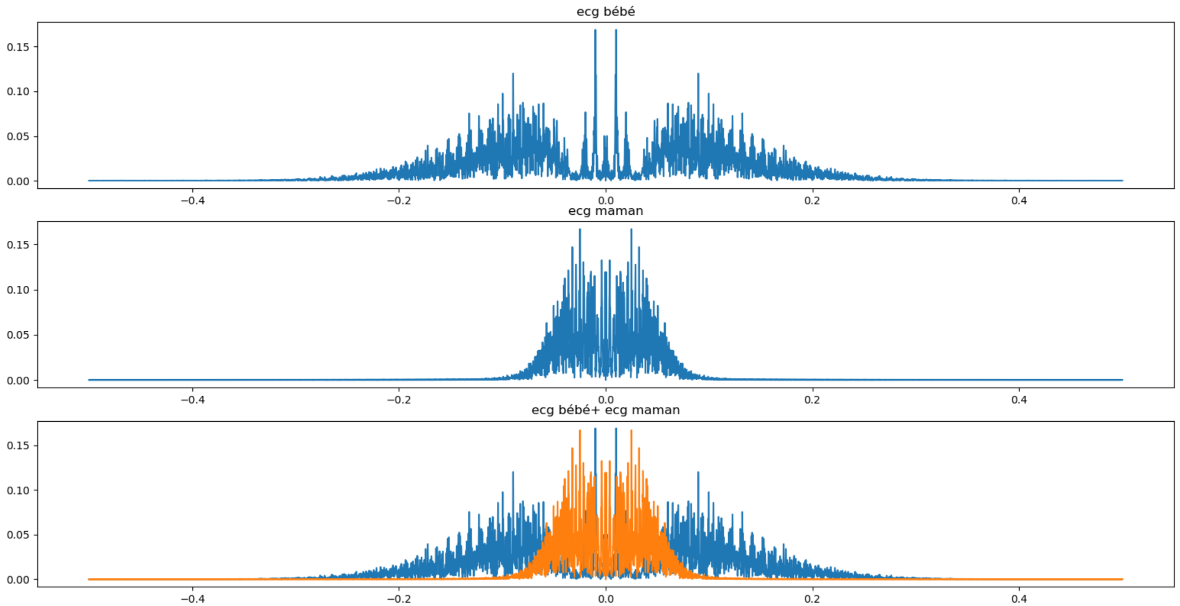
$$H(z) = \frac{1}{(1+\sigma_b^2)^2 - \alpha^2}\left[(1+\sigma_b^2-\alpha^2) + \alpha\sigma_b^2 z^{-1}\right]$$

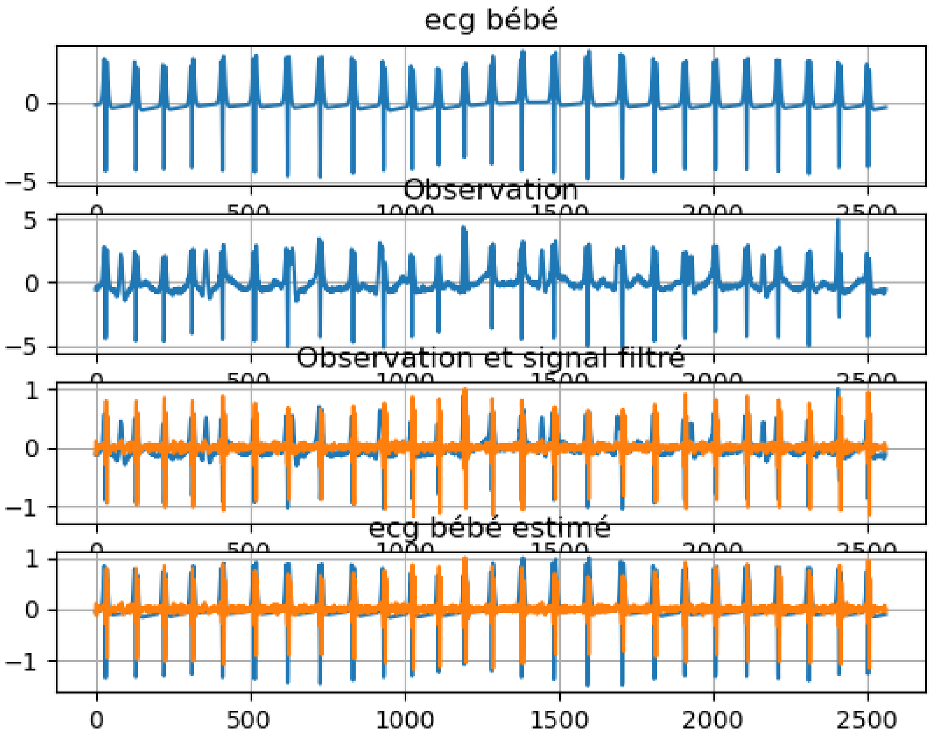
8. Wiener filtering

Example 3

Measurement of the cardiac activity of a fetus using an electrocardiogram (ECG) taken at the level of the mother's abdomen (the signal *x(n))* and which will naturally be disturbed by the latter's ECG to which is added the thermal noise of the electrodes and electronic equipment.

To find the fetal ECG , a second measurement is made providing the mother's ECG (the disturbing signal). The Wiener filter can then be used to estimate the signal *y(n) representing the* fetal ECG.





9. Adaptive digital filtering

*Reminder: **Wiener Filter***

- Versions: continuous (≈ 30s) discrete (≈1947)

Assumptions: SSL signal → Coefficients of the optimal filter calculated only once.

- ≈60s, processing power >> ability to implement complex algorithms in real time

→ Processing **non-stationary signals**

→ Filter coefficients **change** when the signal changes

→ **Adaptive** filtering

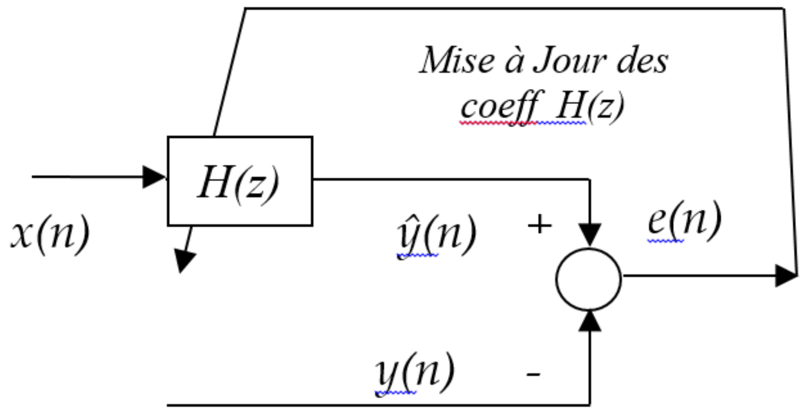
Adaptive filter: It is a digital filter with coefficients that are determined and updated by an adaptive algorithm.

9. Adaptive digital filtering

Adaptive filter: It is a digital filter with coefficients that are determined and updated by an adaptive algorithm.

- x(n): input,
- y(n) : desired response,
- e (n) : error which is the difference between y(n) and the output of th

It is used to control (adapt) the values of the filter coefficients.



Applications:

- Identification of systems: Known input and output → identify H(z) (eg: automatic)

- Prediction: Output knows up to n, predict next output values (eg: target position)

- Deconvolution (equalization): Find signal distorted by unknown process.

- Interference cancellation: Echo cancellation, noise

- Neural networks (weight update with the back-propagation algorithm)

9. Adaptive digital filtering

Reminder: ***Wiener Filter***

x(n) : available observation linked to the process *y(n)* of interest (not directly accessible).

We are looking for the filter $h = [b_0 \quad b_1 \quad \cdots \quad b_{N-1}]^T$ → The estimated signal $\hat{y}(n) = \sum_{i=0}^{N-1} b_i x(n-i)$

We seek to minimize the mean squared error.

$$\xi(n) = E\{e^2(n)\} = E\left\{\left(y(n) - \sum_{i=0}^{N-1} b_i x(n-i)\right)^2\right\} \rightarrow \frac{\partial \xi}{\partial b_j} = -E\{2e(n)x(n-j)\} = -2R_{yx}(j) + 2\sum_{i=0}^{N-1} b_i R_{xx}(j-i) = 0 \rightarrow R_{yx}(j) = \sum_{i=0}^{N-1} b_i R_{xx}(j-i)$$

→ $R_{xx}(k)$ and $R_{yx}(k)$ assumed **to be known**

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix} . \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix} = \begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}$$

2 Problems

- Inadequate Wiener filtering if signal or noise are non-stationary

- In practice: $\boldsymbol{R_{xx}(k)}$ **and** $\boldsymbol{R_{yx}(k)}$ **unknowns**

→Optimal filter must be time-varying: **Adaptive filtering.**

Recursive update of parameters (coefficients) to adapt to the process by a feedback loop (depending on the error).

9. Adaptive digital filtering

Reminder: ***Wiener Filter***

$$\frac{\partial \xi}{\partial b_j} = -E\{2e(n)x(n-j)\} = -2R_{yx}(j) + 2\sum_{i=0}^{N-1} b_i R_{xx}(j-i) = 0$$

$$\overbrace{\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix}}^{R_{xx}} \overbrace{\begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix}}^{h} . = \overbrace{\begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}}^{R_{yx}} \rightarrow -2R_{yx} + 2\,R_{xx}h = -2E\{y(n)X(n)\} + 2E\{X(n)X(n)^T\}\,h = 0$$

Matrix Notation

$$X(n) = [x(n), x(n-1), \ldots\ldots\ldots x(n-(N-1)]$$

$$R_{xx} = E\{X(n)X(n)^T\} = \mathrm{E}\left\{ \overbrace{\begin{bmatrix} x(n) \\ x(n-1) \\ . \\ x(n-(N-1)) \end{bmatrix}}^{X(n)} \overbrace{\begin{bmatrix} x(n) & x(n-1) & \ldots & x(n-(N-1) \end{bmatrix}}^{X(n)^T} \right\} = \begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix}$$

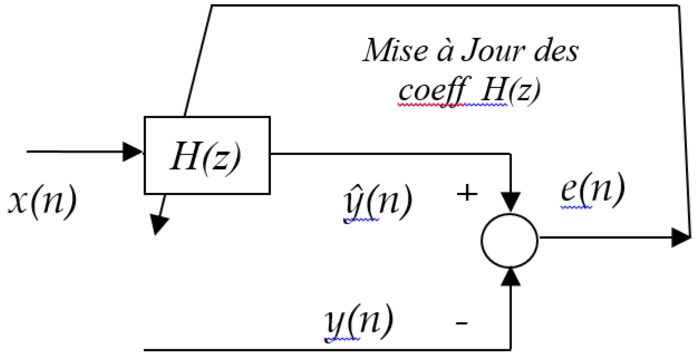
$$R_{yx} = E\{y(n)X(n)\} = E\left\{ y(n) \overbrace{\begin{bmatrix} x(n) \\ x(n-1) \\ . \\ x(n-(N-1)) \end{bmatrix}}^{X(n)} \right\} = \begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}$$

9. Adaptive digital filtering

Reminder: ***Wiener Filter***

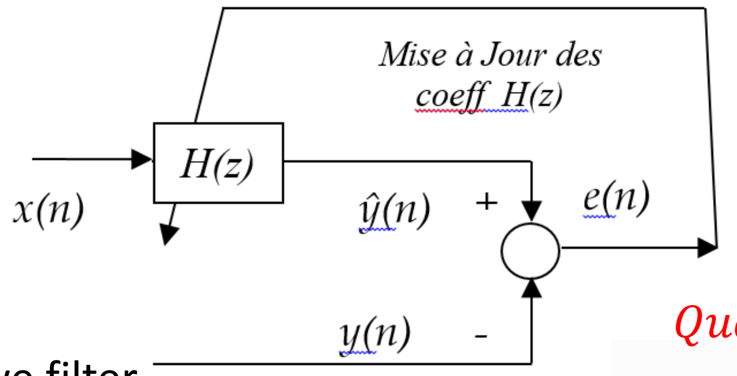
$$\overbrace{\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \dots\dots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \dots\dots & R_{xx}(N-2) \\ & & \dots\dots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \dots\dots & R_{xx}(0) \end{bmatrix}}^{R_{xx}} \overbrace{\begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix}}^{h} . = \overbrace{\begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}}^{R_{yx}} \rightarrow -2R_{yx} + 2\,R_{xx}h = -2E\{y(n)X(n)\} + 2E\{X(n)X(n)^T\}\,h = 0$$

Adaptive filtering does not require a priori knowledge of the statistical properties of the processed random signals. It is built as samples arrive. It adapts over time to changes in the statistical properties of the signals. In the stationary case, the filter converges towards the "optimal Wiener filter"



$$Quantity\ to\ minimise = -2R_{yx} + 2hR_{xx}$$

9. Adaptive digital filtering



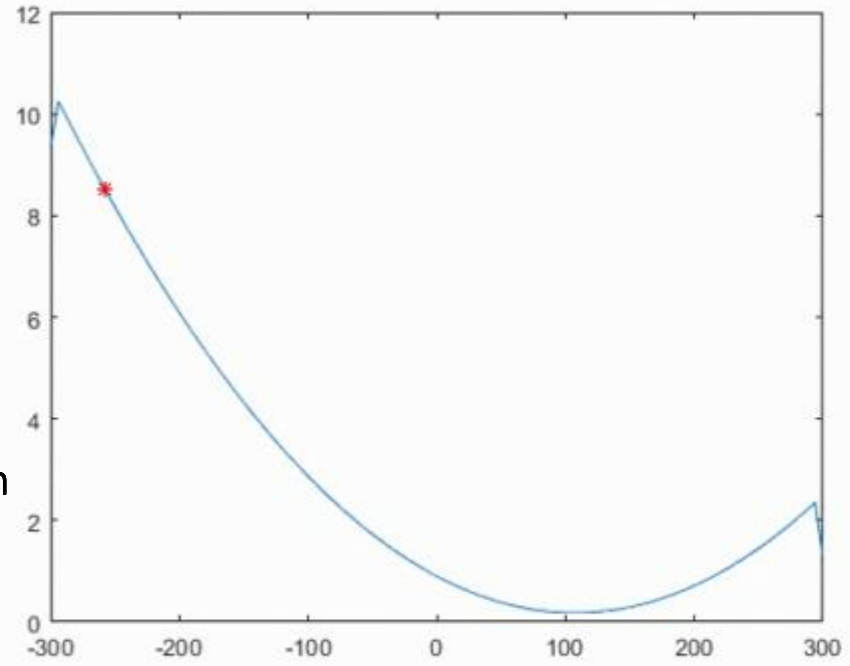
Iterative algorithms: No matrix inversion

$Quantity\ to\ minimise = -2R_{yx} + 2hR_{xx}$

1. Initialization (n = 0), h(0)=[0, 0, ….. 0] then we filter

2. At each instant n, we calculate the gradient of the error (the derivative) $\nabla_{h(n)}\xi(n)$with respect to the vector h(n)

3. Adjust the parameters: $h(n+1) = h(n) - \frac{\mu}{2}\nabla_{h(n)}\xi(n)$



https://tenor.com/view/gradientdescent-graph-lines-wave-gif-17776126

Adjustments in the opposite direction to the gradient (gradient descent) → h(n) towards the hopt value (ξ min). If the derivative is positive then we will decrease h(n) so as to move towards the minimum of the cost function. The corrective term will therefore be chosen negative and vice versa.

Not $\frac{\mu}{2}$:small enough for the algorithm to converge and large enough for the algorithm to reach its optimal value as quickly as possible. This is called the learning phase.

9. Adaptive digital filtering

$$\overbrace{\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix}}^{R_{xx}} \overbrace{\begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix}}^{h} .= \overbrace{\begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}}^{R_{yx}} \rightarrow -2R_{yx} + 2\,R_{xx}h = -2E\{y(n)X(n)\} + 2E\{X(n)X(n)^T\}\,h$$

Steepest (gradient) descent

Recursive algorithm for calculating Wiener filter coefficients (no inversion of R_{xx})

1. Initialization (n = 0), h(0)=[0, 0, ….. 0]
2. At each instant n, we calculate the gradient of the error (the derivative) $\nabla_{h(n)}\xi(n)$with respect to the vector h(n):

$$\nabla_{h(n)}\xi(n) = -2R_{yx} + 2\,R_{xx}\,h(n)$$

3. Adjust the parameters:$h(n+1) = h(n) - \frac{\mu}{2}\nabla_{h(n)}\xi(n) = h(n) + \mu\,(R_{yx} - R_{xx}h(n))$

9. Adaptive digital filtering

$$\overbrace{\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix}}^{R_{xx}} \overbrace{\begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix}}^{h} .= \overbrace{\begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}}^{R_{yx}} \rightarrow -2R_{yx} + 2\,R_{xx}h = -2E\{y(n)X(n)\} + 2E\{X(n)X(n)^T\}\,h$$

Least Mean square (LMS)

Replace $R_{xx}$and with R_{yx}instant estimates. $\hat{R}_{xx} = \mathrm{X}(n)\mathrm{X}^T(n)\;\; et\; \hat{R}_{yx} = y(n)\mathrm{X}(n)$

1. Initialization (n = 0), h(0)=[0, 0, ….. 0]

2. At each instant n, we calculate the gradient of the error (the derivative) $\nabla_{h(n)}\xi(n)$with respect to the vector h(n):

$$\hat{\nabla}_{h(n)}\xi(n) = -2y(n)\mathrm{X}(n) + 2\mathrm{X}(n)\mathrm{X}^T(n)h(n) \qquad \mathrm{X}(n) = [x(n), x(n-1), \ldots\ldots x(n-N+1)]^T.$$

3. Adjust the filter parameters according to$\hat{h}(n+1) = \hat{h}(n) + \mu\,\mathrm{X}(n)[\overbrace{y(n) - \mathrm{X}^T(n)\hat{h}(n)}^{e(n)}$

Note: LMS is based on minimizing the instantaneous error → Oscillations around the minimum error ξ_{min}.

9. Adaptive digital filtering

$$\overbrace{\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix}}^{R_{xx}} \overbrace{\begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix}}^{h} . = \overbrace{\begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}}^{R_{yx}} \rightarrow -2R_{yx} + 2\, R_{xx}h = -2E\{y(n)X(n)\} + 2E\{X(n)X(n)^T\}\, h$$

Least Squares (RLS) Algorithm :

Minimization of the sum of squared errors: $\xi(n) = \sum_{k=1}^{n} \lambda^{n-k} e^2(k)$

1. Initialization (n = 0), $R_{xx\lambda}^{-1}(0) = \left(\sum_{k=-n_0}^{0} \mathrm{X}(k)\mathrm{X}^T(k)\right)^{-1}$ and $R_{yx\lambda}(0) = \lambda \sum_{k=-n_0}^{0} y(k)\mathrm{X}(k)$

2. At each time n, we calculate the correlation matrices:

- $R_{xx\lambda}(n) = \sum_{k=1}^{n} \lambda^{n-k}\mathrm{X}(k)\mathrm{X}^T(k) = \lambda R_{xx\lambda}(n-1) + \mathrm{X}(k)\mathrm{X}^T(k)$
- $R_{yx\lambda}(n) = \sum_{k=1}^{n} \lambda^{n-k} y(k)\mathrm{X}(k) = \lambda R_{yx\lambda}(n) + y(k)\mathrm{X}(k)$

3. Adjust the filter parameters $\hat{h}(n) = R_{xx\lambda}^{-1}(n) R_{yx\lambda}(n)$

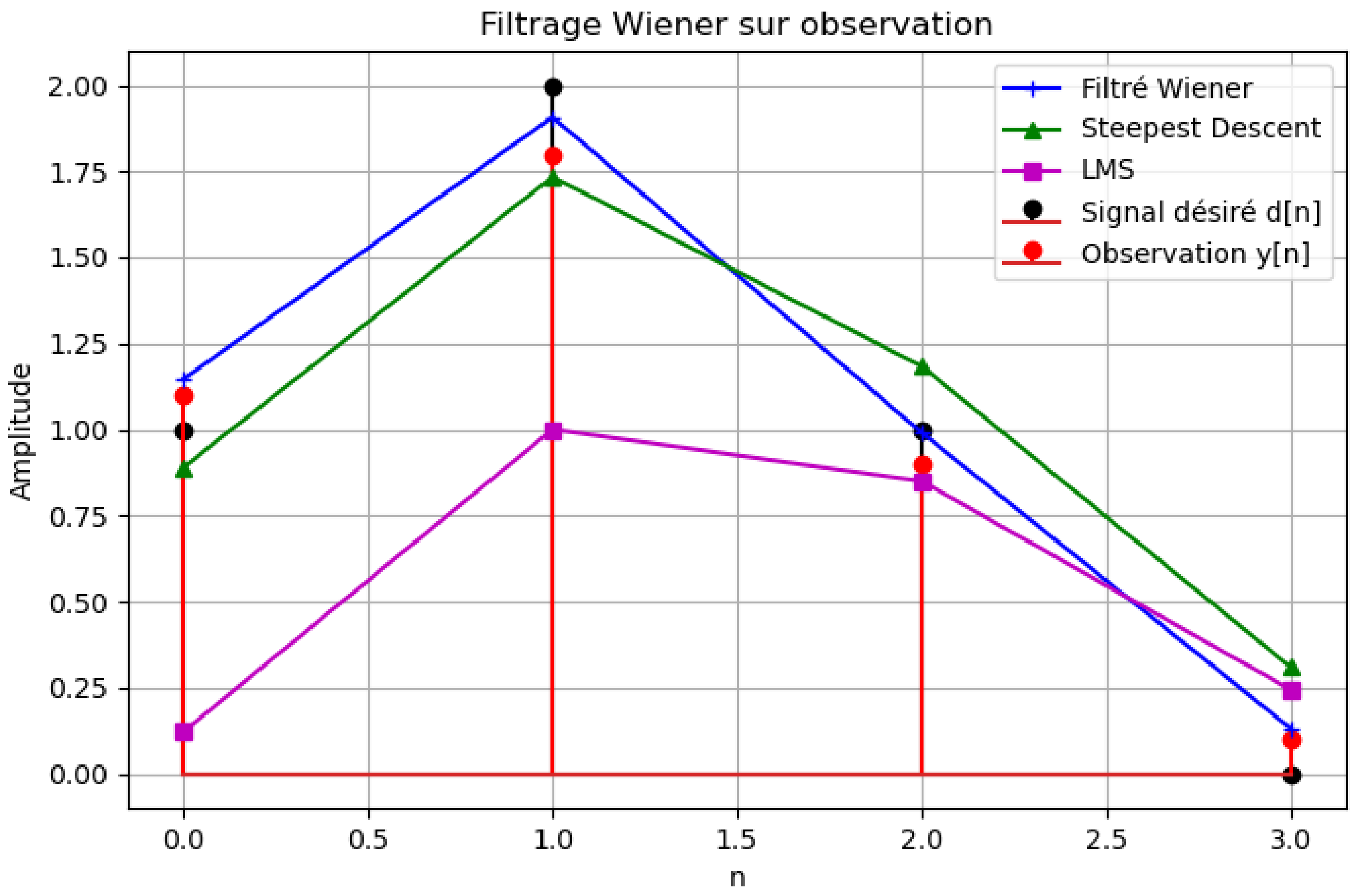
Note: For $\lambda = 1$, the matrix $R_{xx\lambda}(n)$ is an approximation of the autocorrelation matrix .

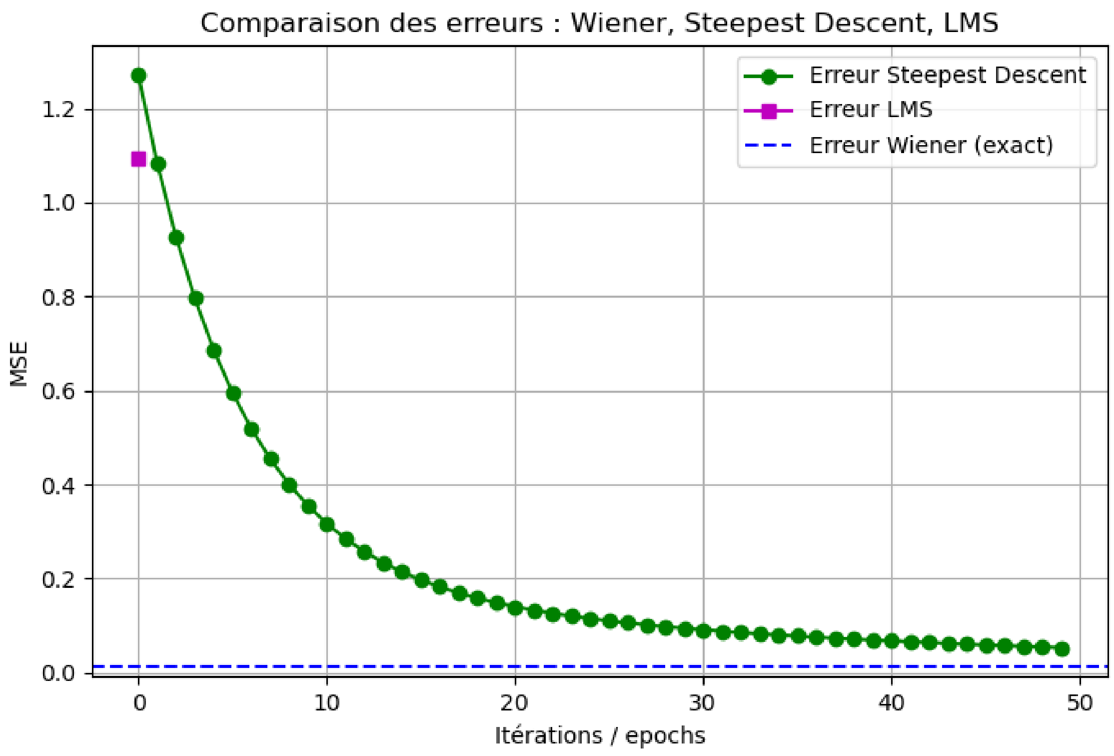
9. Adaptive digital filtering

Wiener /SD/LMS :

Observed signal $x(n) = [1.1, 1.8, 0.9, 0.1]$

Desired signal $y(n) = [1, 2, 1, 0]$,



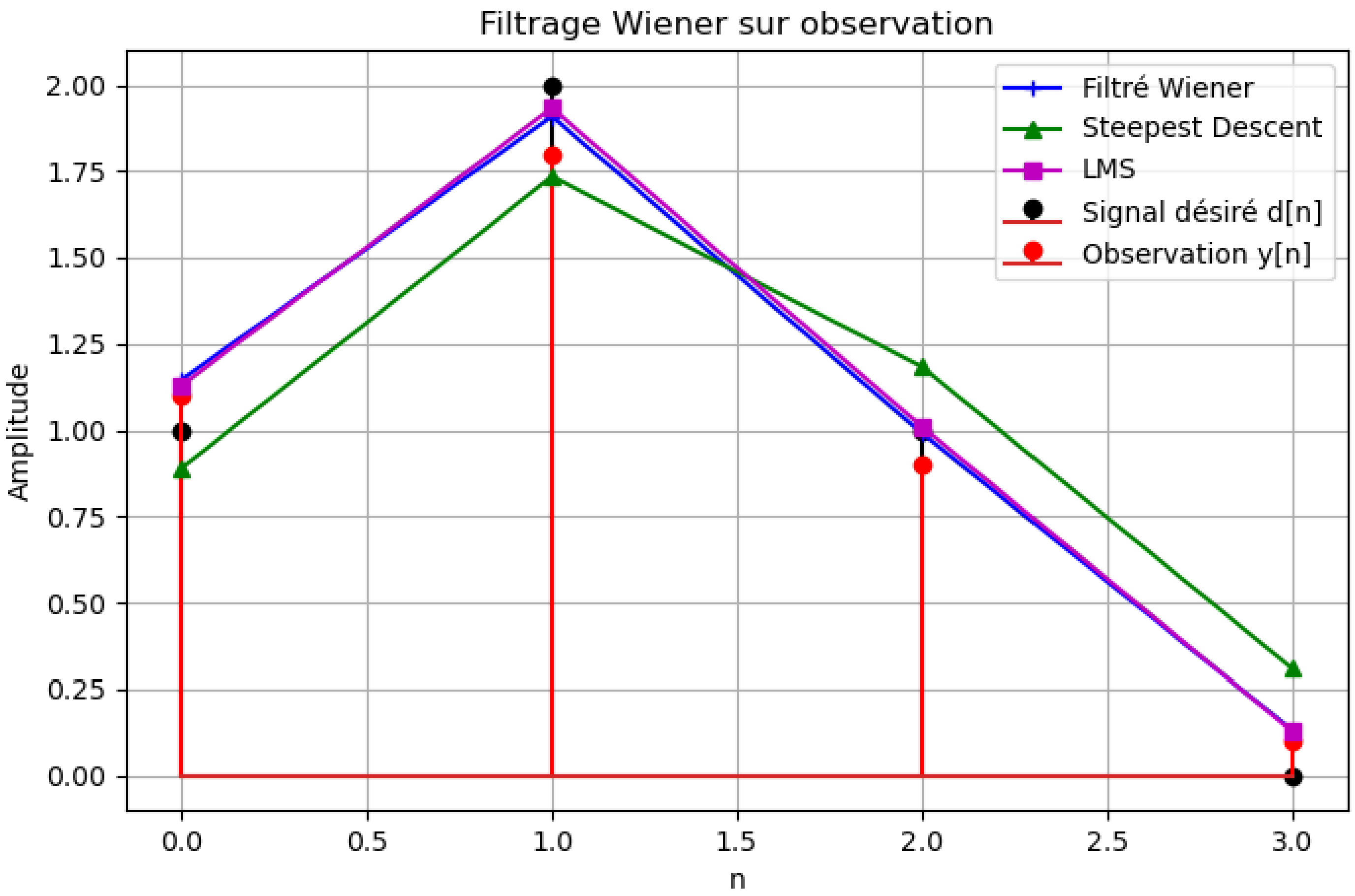


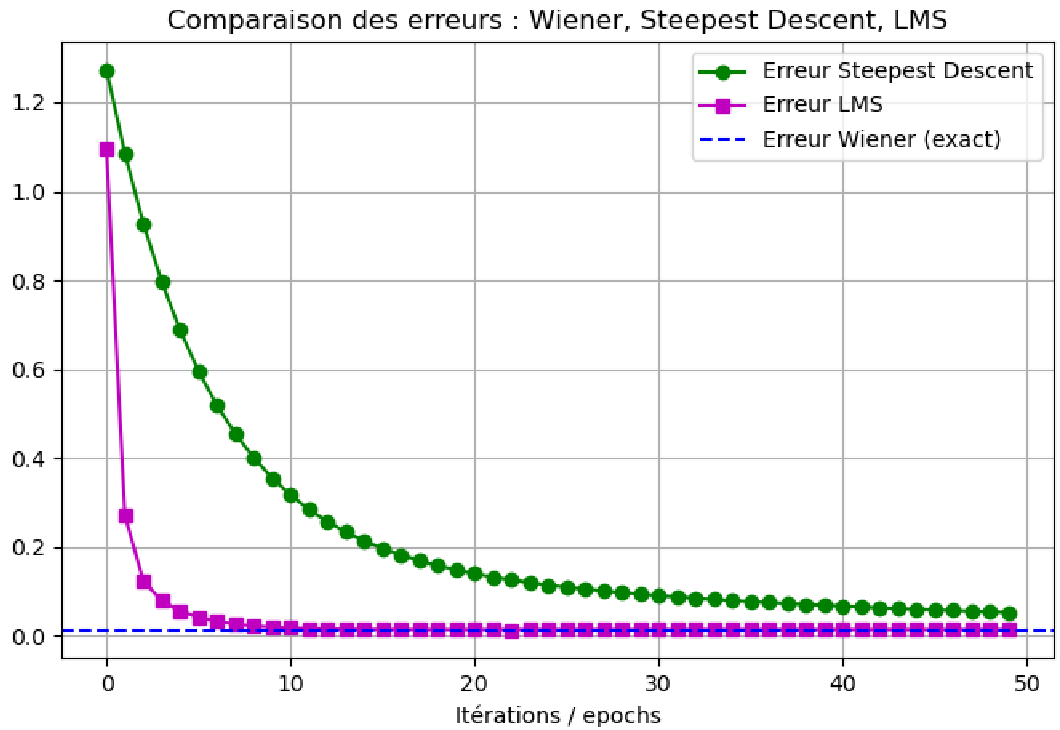
9. Adaptive digital filtering

Wiener /SD/LMS :

Observed signal $x(n) = [1.1, 1.8, 0.9, 0.1]$

Desired signal $y(n) = [1, 2, 1, 0]$,



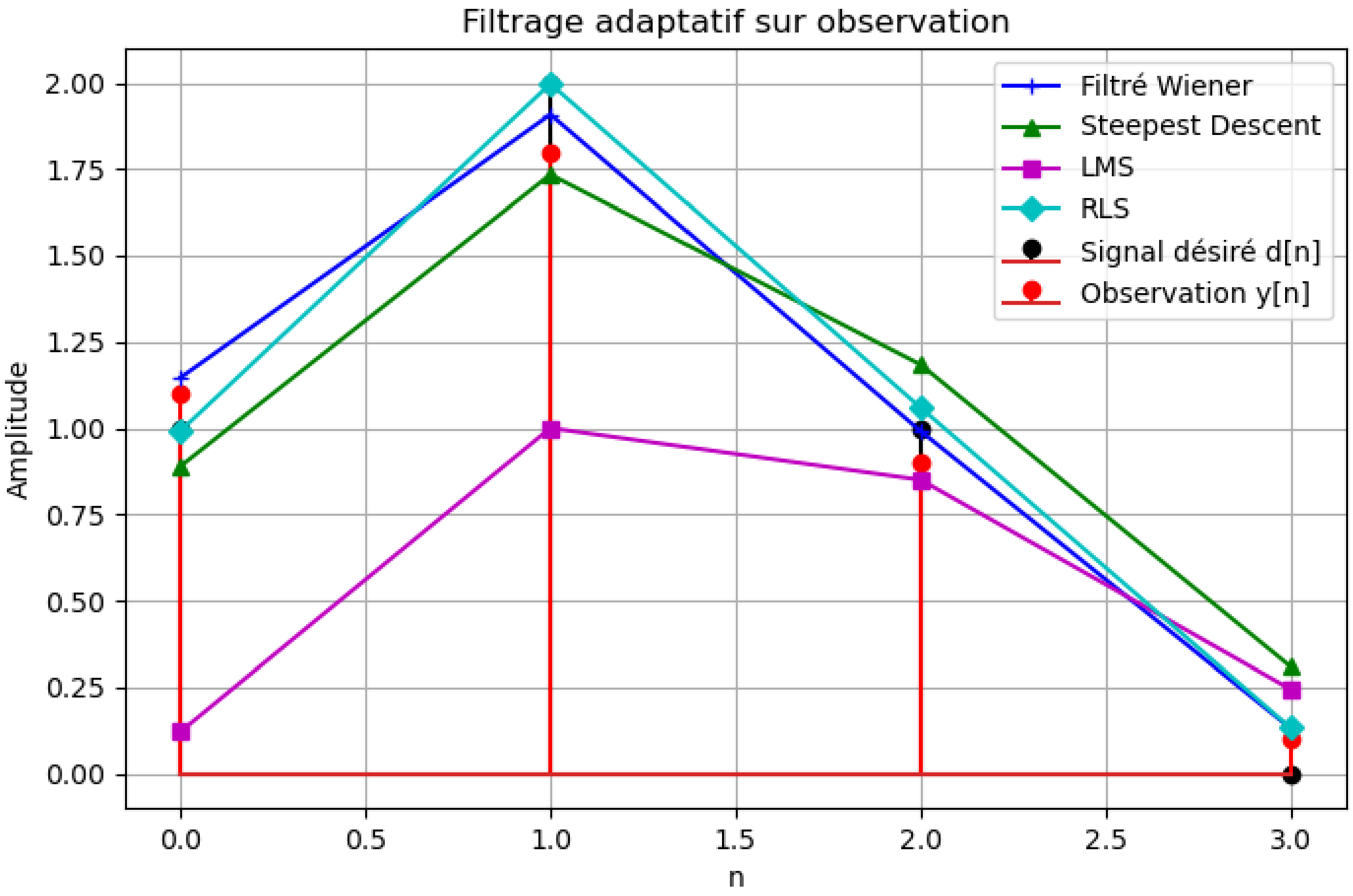


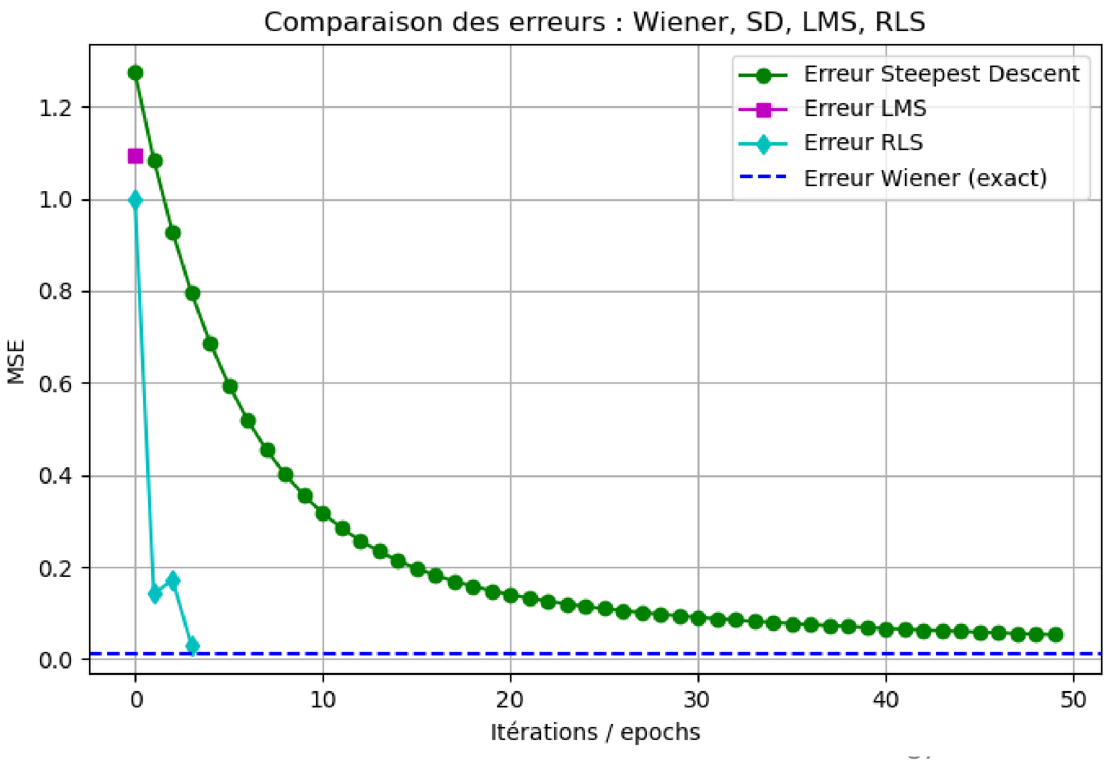
9. Adaptive digital filtering

Wiener /SD/LMS/RLS :

Observed signal $x(n) = [1.1, 1.8, 0.9, 0.1]$

Desired signal $y(n) = [1, 2, 1, 0]$,



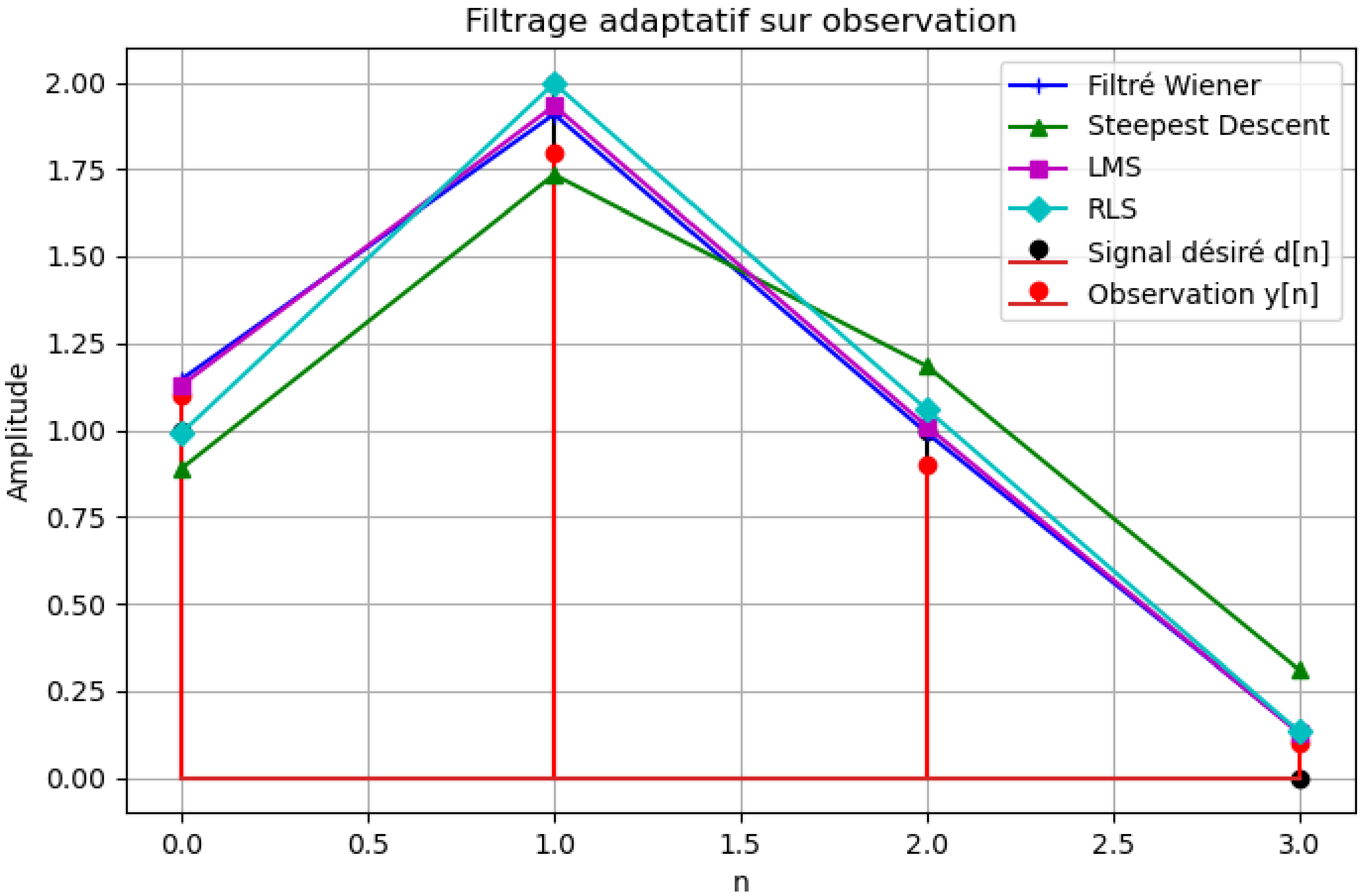


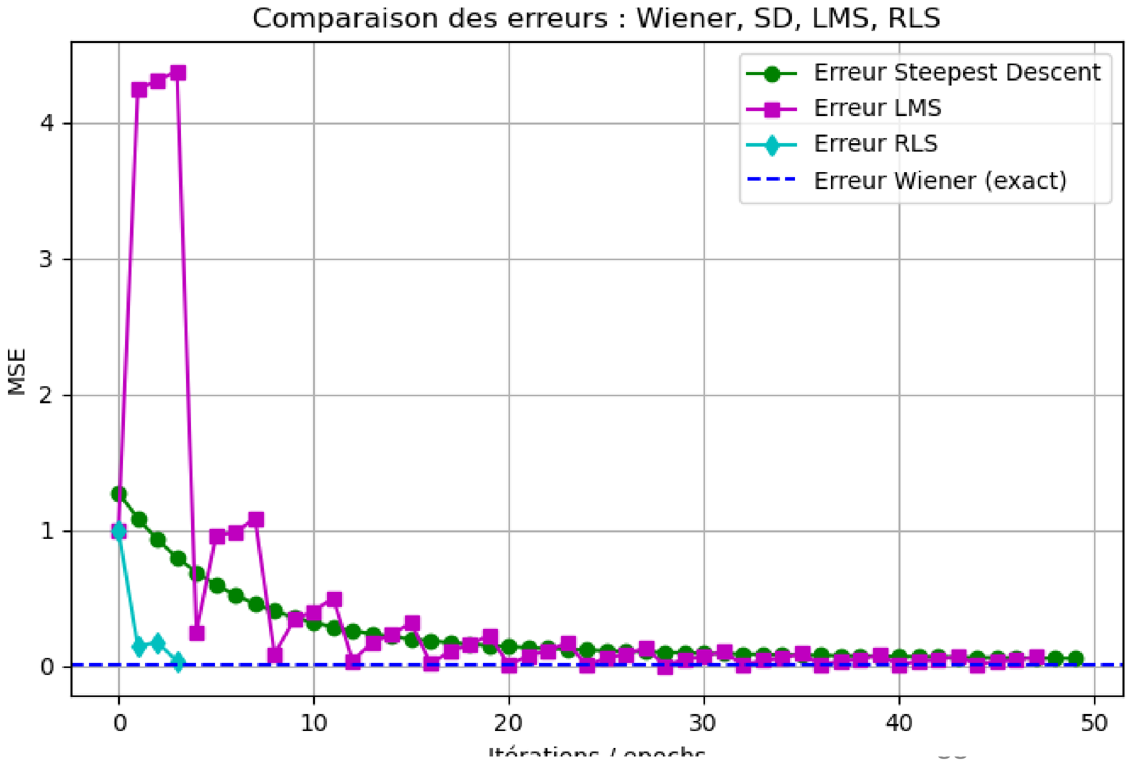
9. Adaptive digital filtering

Wiener /SD/LMS/RLS :

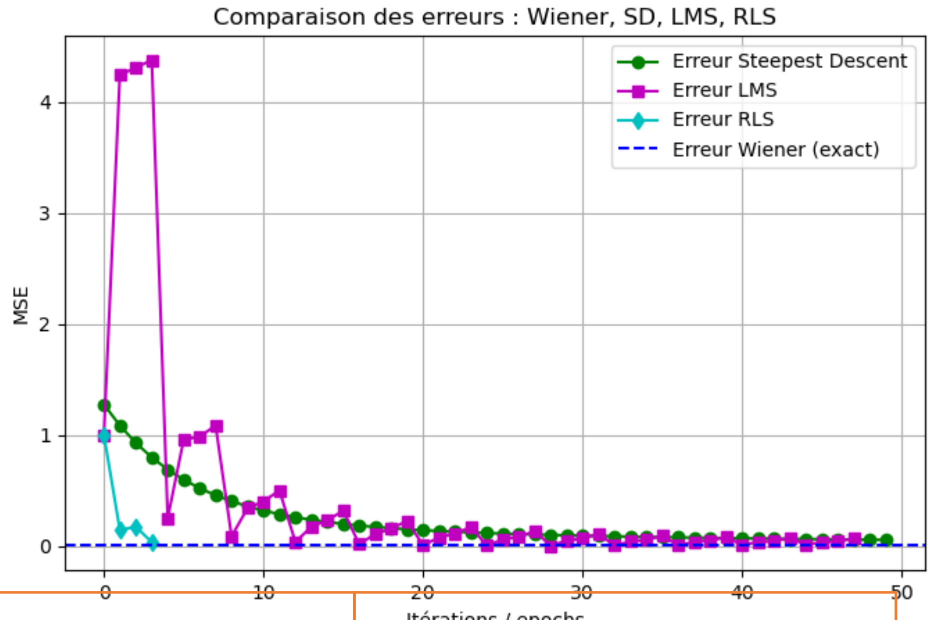
Observed signal $x(n) = [1.1, 1.8, 0.9, 0.1]$

Desired signal $y(n) = [1, 2, 1, 0]$,





9. Adaptive digital filtering



Wiener /SD/LMS/RLS :

Observed signal $x(n) = [1.1, 1.8, 0.9, 0.1]$

Desired signal $y(n) = [1, 2, 1, 0]$,

Method	Convergence Speed	Simplicity / Cost	Robustness	Practical Example	Main Limitation
Wiener	Immediate	Moderate	Good	Noise removal in pre-recorded audio or ECG, image restoration	Offline only, requires exact correlation knowledge
Steepest Descent (SD)	Slow	Simple	Moderate	Channel equalization in telecom, adaptive filtering of static audio noise	Slow convergence, sensitive to step size
LMS	Moderate	Very simple, low cost	Good	ANC (microphone), real-time channel equalization, adaptive tracking of portable biological signals	Irregular convergence, step size must be tuned
RLS	Very fast	More complex, higher cost	Excellent	High-speed channel equalization, radar/sonar, precise adaptive tracking of industrial or biological signals	Computationally heavy, requires recursive matrix inversion

9. Adaptive digital filtering

$$\overbrace{\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & \ldots\ldots & R_{xx}(N-1) \\ R_{xx}(1) & R_{xx}(0) & \ldots\ldots & R_{xx}(N-2) \\ & & \ldots\ldots & \\ R_{xx}(N-1) & R_{xx}(N-2) & \ldots\ldots & R_{xx}(0) \end{bmatrix}}^{R_{xx}} \overbrace{\begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{N-1} \end{bmatrix}}^{h} . = \overbrace{\begin{bmatrix} R_{yx}(0) \\ R_{yx}(1) \\ . \\ R_{yx}(N-1) \end{bmatrix}}^{R_{yx}} \rightarrow -2R_{yx} + 2\,R_{xx}h = -2E\{y(n)X(n)\} + 2E\{X(n)X(n)^T\}\,h$$

Other Algorithms :

- RIF: Generalized Normalized Gradient Descent (GNGD), Normalized Nonlinear Gradient Descent (NNGD), Fully Adaptive NNGD (FANNGD), Normalized Gradient / Normalized LMS (NLMS) filter , Leaky LMS filter , Gradient Adaptive Lattice (GAL) filter , Lattice LMS filter with Joint Process Estimation

- RII: Recursive Least Squares (RLS) filter / Forgetting Factor, Kernel RLS (KRLS) filter , Sliding Window RLS filter , IIR Wiener filter

9. Adaptive digital filtering

Kalman :

- Used to estimate parameters of a system evolving over time from noisy measurements.
- Suitable for tracking because it not only allows to predict parameters but also to correct measurement and model errors.
- Prediction is made by using the previous estimate. The error calculation then allows this prediction to be updated using the new measurements.

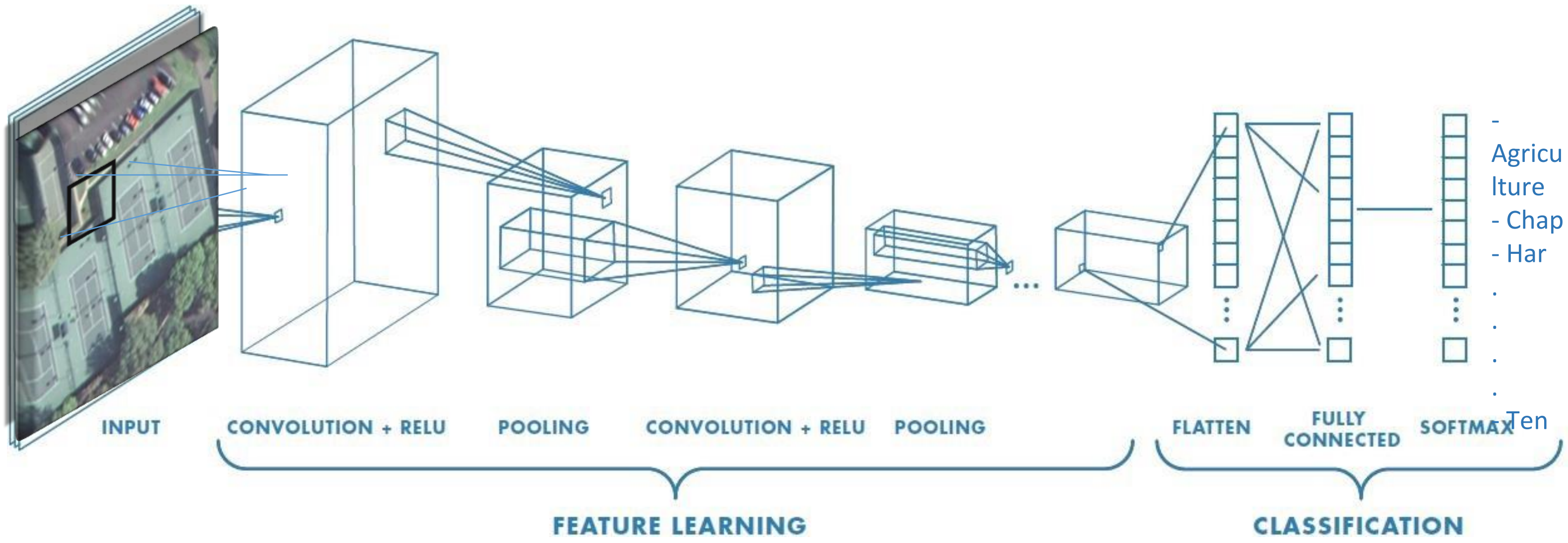
Kalman Filter allows to take into account non-linear modeling.

Particulate (sequential Monte Carlo method):

- Alternative to Extended Kalman Filters (more accurate if Samples >>)
- Estimate the posterior density of the variables that we seek to estimate (called state variables) as a function of the observations (observation variables) assuming that the state variables constitute a first-order Markov chain.
- Used in navigation applications: target tracking, missile guidance, robotics, etc.

9. Architecture of a CNN

An artificial neural network consists of interconnected neurons organized into different layers.



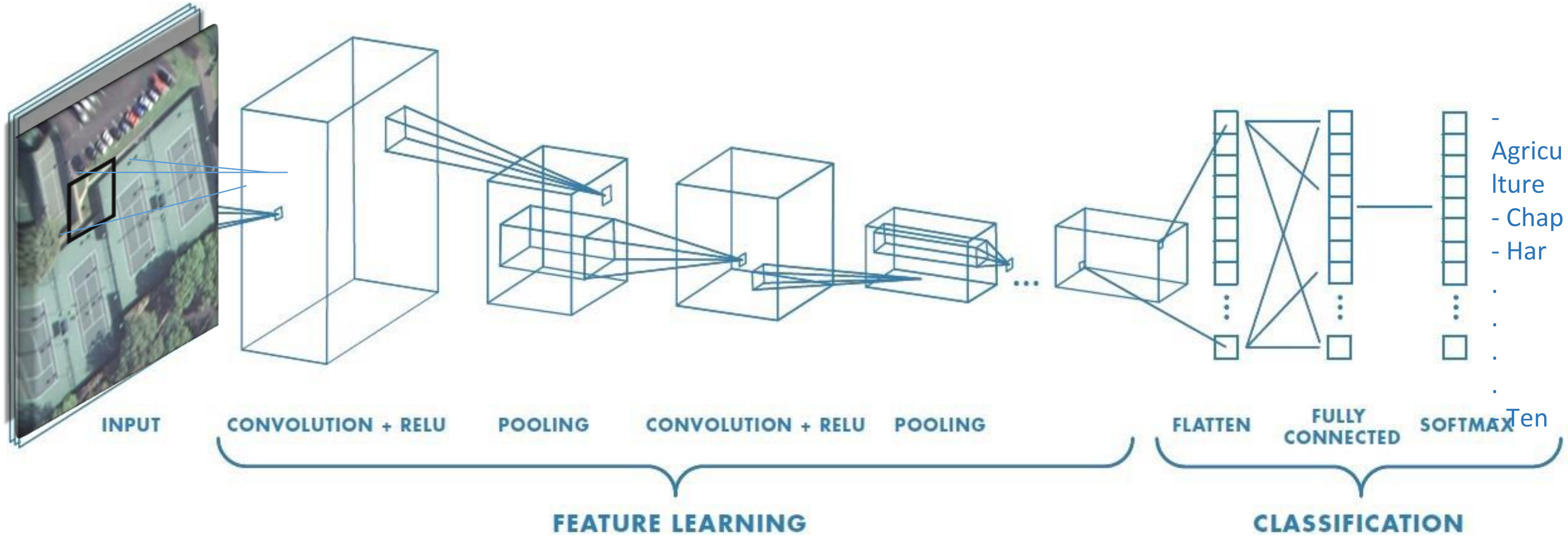
Two very distinct parts:

Convolutional part : extract specific features from the input images using a succession of filters –> Convolution maps which are concatenated into a feature vector called CNN code.

Classification part: Combine and match the features of the CNN code with the features of each class to classify the input images

9. Training a CNN

This is the process of adjusting the value of the weights. Training a neural network usually consists of two phases: forward and backward .



Training steps:

- The weights (filter coefficients, biases, neurons) are assigned random values.

- A large dataset of labeled images is provided by CNN.

The CNN network processes each image with the random weights, and then performs comparisons between the predicted value and the class label of the input image (loss function). If there is no match between the output and the class label, a small adjustment of the weights is performed by backpropagation (gradients are backpropagated) → Tuning. CNN goes through several rounds of iterations during the training process, gradually adjusting its weights.