Random signals and Wiener's theorem

slide set # 8

Alessandro Falaschi

Signal Processing and Information Theory
Bioinformatics Degree, Sapienza University of Roma

Didactic material on TeoriadeiSegnali.it

May 2023

Content index

Overwiev

Summary of the presentation

Since it is not possible to compute a Fourier transform for random signals, a statistical description of it must be adopted. The concept of a random process is therefore introduced, and how to calculate its ensemble and time averages is discussed, then the conditions for defining it as a stationary and ergodic process are explained. After some examples, a 2D random variable is extracted from the process, defining correlation and covariance between the two components, both from a probabilistic and a statistical point of view

The autocorrelation function defined for both random processes and deterministic signals is then introduced, as well as the intercorrelation between different signals, and links to convolution are clarified. The properties of autocorrelation are then enunciated, and a few words are spent on the link between Pearson's coefficient and Schwartz's inequality, just as the matched filter can also be thought of in this way

Wiener's theorem is a valuable tool for obtaining the spectral density for any type of signal, be it deterministic or random, as well as for opening an alternative path towards already known results. Examples of its application are given. Particular emphasis is given to the multivariate Gaussian and to its peculiar characteristics, as well as to its suitability to express the statistics of the samples of an ergodic Gaussian process. Spectral estimation by means of a periodogram is also explored.

Finally, the issues of random signals and filtering are jointly addressed: in fact, the expression of the spectral density at the filter output is considered, and then its statistical description is evaluated. The set of slides ends by analyzing the result of adding or multiplying random processes.

Now we talk about...

1. **Probability and statistics**
   - Building a ground of common terms
   - Stationary and ergodic processes
   - Correlation and covariance for signals
2. **A common point of view**
   - Autocorrelation and intercorrelation
   - Geometry and adaptation
3. **Power density spectrum**
   - Wiener's theorem
   - Multidimensional Gaussian and process
   - Spectral estimation
4. **Filtering and combining of signals and processes**
   - Spectral density at the output of a filter
   - Statistical characteristics at the output of a filter
   - Sum and product of random and deterministic signals

Random variables and probability density

Random variable

Let X to be a *random variable* (or r.v.), whose values x are taken from a *sample space* Ω, and are called *a realization* of the r.v. The Ω set can accommodate a finite or infinite number of values, making X a *discrete* or *continuous* r.v.

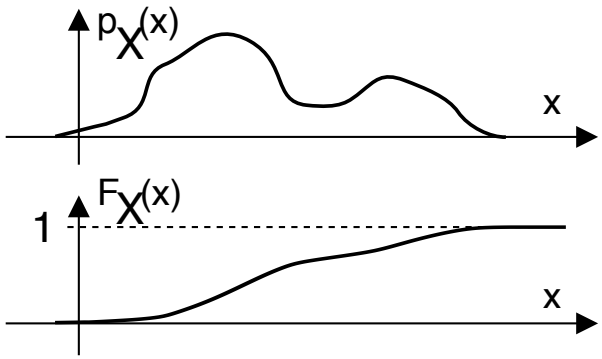
Probability density function (or p.d.f.)

Is a *non-negative* function $p_X(x)$ of the r.v. X with unit area, whose definite integral $\int_a^b p_X(x)\,dx$ gives the probability to find a realization x of X such that $a < x \leq b$

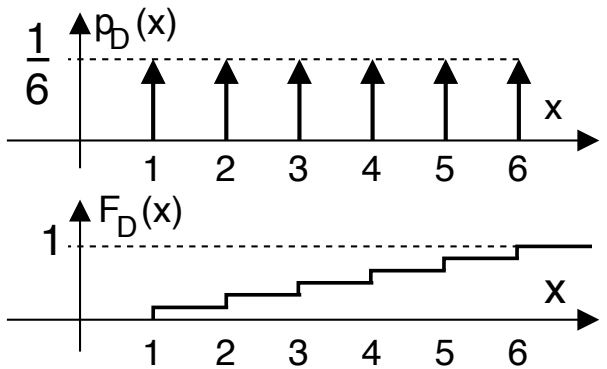
Distribution function

Is a *non-decreasing* function $0 \leq F_X(a) \leq 1$ of a variable $-\infty < a < \infty$, which gives the probability of finding a realization x of X less than or equal to a, resulting $F_X(a) = \int_{-\infty}^{a} p_X(x)\,dx = Pr\{X \leq a\}$.

Continuous random variable



Discrete r.v. (a dice)

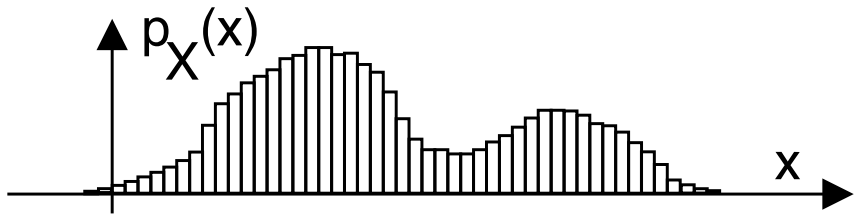


Histogram

Instead of postulating which p.d.f. your r.v. should obey, you can

- observe a (large) amount of results, repeating N times some sort of experiment
- note the range of observed values
- divide the range into a finite number of bins and count how many results (N_k) fall into each k^{th} bin
- draw a rectangle for each bin, with a heigth equal to the ratio N_k/N of results in the k^{th} bin compared to the total number of attempts

For each bin k, you get an *estimate* of the probability p_k that a future realization of X falls into the k^{th} bin



Expected value and moments

Expected value (or *ensemble average)*

It's a weighted mean, i.e. it weighs a function of a r.v. $g(x)$ with the probability of each possible outcome

$$E_X\{g(x)\} = \int_{-\infty}^{\infty} g(x)\, p_X(x)\, dx$$

It can also be defined for functions of a pair (or more generally of an $n-$ple) of r.v. using their *joint* p.d.f., that is

$$E_{X,Y}\{g(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, p_{XY}(x,y)\, dxdy$$

or, if a *conditional* p.d.f. $p_{X/Y}(x/y)$ is used, we will obtain a *conditional* expected value $E_{X/Y}\{g(x)\}$, which depends on the value of y

Moment

By setting $g(x) = x^n$ we obtain the *moment of order n*

$$m_X^{(n)} = E\{x^n\} = \int_{-\infty}^{\infty} x^n p_X(x)\, dx$$

which for discrete r.v. is written as $m_X^{(n)} = \sum_i x_i^n p_i$, thus weighting the possible realizations x_i with the respective probabilities $p_i = Pr\{x = x_i\}$

Now note that $m_X^{(0)} = \int_{-\infty}^{\infty} p_X(x)\, dx = 1$. With order n equal to 1 or 2 we get...

Mean value and quadratic mean

Mean value (*or centroid)* is the *first-order* moment

$$m_X = m_X^{(1)} = \int_{-\infty}^{\infty} x p_X(x)\, dx$$

Quadratic mean (or *square mean*) is the *second order* moment

$$m_X^{(2)} = \int_{-\infty}^{\infty} x^2 p_X(x)\, dx$$

Example

Think at the r.v. X as at the height of a population: its *mean value* m_X can be *estimated* as the arithmetic mean of the measurements

$$\hat{m}_X = \frac{\overbrace{x_1 + x_1 + \cdots}^{N_1 \text{ times}} + \overbrace{x_2 + x_2 + \cdots}^{N_2 \text{ times}} + \ldots + \overbrace{x_n + + x_n + \cdots}^{N_n \text{ times}}}{N} = \frac{x_1 N_1 + x_2 N_2 + \ldots + x_n N_n}{N}$$

When $N = \sum_{i=1}^{n} N_i$ approaches ∞, the formula for $\hat{m}_X$ coincides with that for the mean m_X **if** we substitute the probabilities $p_X(x)\, dx$ with the $Pr(x_i)$ values obtained for the histogram, i.e. $Pr(x_i) = \frac{N(x_i < x \leq x_i + \Delta x)}{N} = \frac{N_i}{N}$, thus transforming the integral into a summation, that is

$$\int_{-\infty}^{\infty} x p_X(x)\, dx \Rightarrow \sum_i x_i Pr(x_i)$$

This point of view motivates the concept of *weighing* the possible values of x with their respective frequencies

Central moment and variance

Central moment

In the case $g(x) = (x - m_X)^n$ the corresponding expected value is called a *central moment* of order n, expressed as

$$\mu_X^{(n)} = E\left\{(x - m_X)^n\right\} = \int_{-\infty}^{\infty} (x - m_X)^n \, p_X(x) \, dx$$

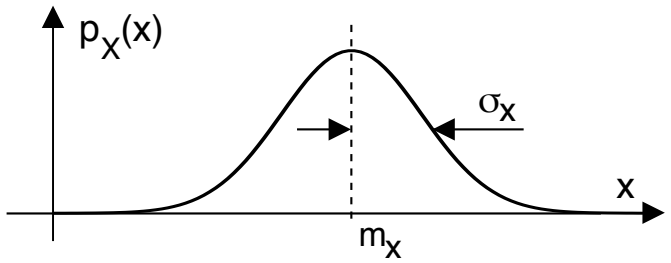
It is straightforward to note that $\mu_X^{(0)} = 1$ and that $\mu_X^{(1)} = 0$

Variance σ^2

It is the name given to the central moment of 2^{nd} order, corresponding to

$$\sigma_X^2 = \mu_X^{(2)} = E\left\{(x - m_X)^2\right\} = \int_{-\infty}^{\infty} (x - m_X)^2 \, p_X(x) \, dx$$

The variance square root σ_X is called *standard deviation*. While the mean m_X indicates where the "statistical center" of the probability density is located, σ_X indicates how much the individual realizations of the r.v. are *dispersed* around m_x



A *remarkable* relationship that links the first two moments (central and not) is

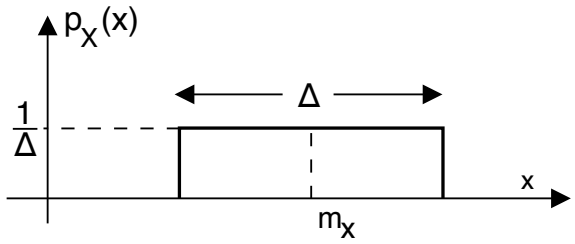
$$\sigma_X^2 = m_X^{(2)} - (m_X)^2$$

Uniform and Gaussian random variable

Often the true p.d.f. of the data under examination can only be postulated or hypothesized, taking advantage of the fact that the parameters that define it can be estimated starting from the experimental data

Uniform random variable

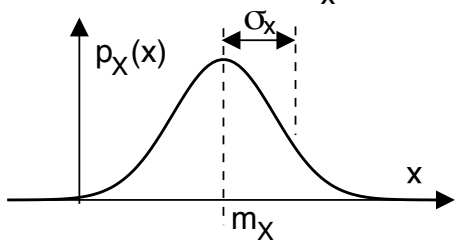
$$p_X(x) = \frac{1}{\Delta} rect_\Delta (x - m_X)$$



It uses m_X as *centering* parameter and its variance *is related* to the dynamic range Δ as (check the book) $\sigma_X^2 = \frac{\Delta^2}{12}$

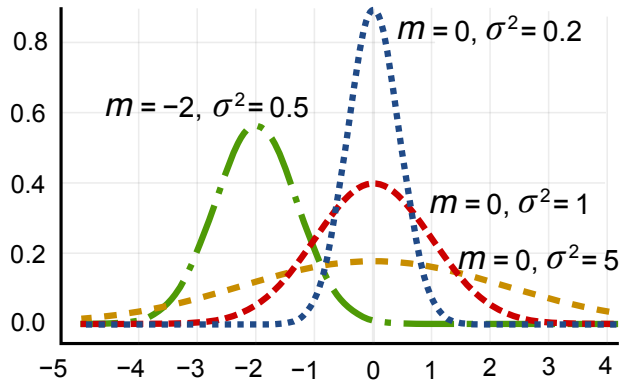
Gaussian random variable

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{ -\frac{(x-m_X)^2}{2\sigma_x^2} \right\}$$



completely defined by the first order and second order central moments m_x and $\sigma_x^2 = m_X^{(2)} - (m_X)^2$, which can be *estimated* from the experimental data as

$$\widehat{m}_x = \frac{1}{N} \sum_{n=1}^{N} x_n \quad \textit{and} \quad \widehat{m}_x^{(2)} = \frac{1}{N} \sum_{n=1}^{N} x_n^2$$



The Gaussian r.v. emerges in many natural phenomena where a very large number of random causes add up, all statistically independent and with the same p.d.f., as can be experimented at https://www.randomservices.org/random/apps/Dice.html.

Multivariate random variable

Here $\boldsymbol{X}$ represents an $N-$dimensional r.v. whose realizations $\boldsymbol{x} = [x_1, x_2, \cdots, x_N]^T$ are made by N r.v.s, described by a *joint* $N-$dim p.d.f. $p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N)$

Marginal probability density function For each *single* r.v. x_i a one-dimensional p.d.f. $p_{X_i}(x_i)$ can be obtained by *saturation* of $p_{\boldsymbol{X}}(\boldsymbol{x})$ as

$$p_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{N-1 \text{ integrals}} p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_N)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N$$

Conditional probability density function: is the p.d.f. of a $\boldsymbol{x}_a = (x_1, x_2, \cdots, x_a)$ sub-set of r.v. if the other $\boldsymbol{x}_b = (x_{a+1}, x_{a+2}, \cdots, x_N)$ are known, obtained by dividing the joint p.d.f. $p_{\boldsymbol{X}}(\boldsymbol{x})$ for the marginal $p_{\boldsymbol{X}}(\boldsymbol{x}_b)$ of the *conditioning* r.v., that is $p_{\boldsymbol{X}}(\boldsymbol{x}_a/\boldsymbol{x}_b) = \frac{p_{\boldsymbol{X}}(x)}{p_{\boldsymbol{X}}(\boldsymbol{x}_b)}$

Expected value and moments The marginal $p_{X_i}(x_i)$ allows us to find the moments of each marginal r.v. as $m_{x_i} = E_{\boldsymbol{X}}\{x_i\} = \int x_i p_{X_i}(x_i)\, dx_i$, and to obtain a *mean vector* for $\boldsymbol{X}$ as $m_{\boldsymbol{X}} = (m_{x_1}, m_{x_2,}, \cdots, m_{x_N})$

Mixed moments They refer to two or more elements of $\boldsymbol{X}$: for example, the mixed moment of order (n, m) is defined as $m_{x_i x_j}^{(n,m)} = E_{\boldsymbol{X}}\left\{x_i^n x_j^m\right\} = \int\int x_i^n x_j^m p_{X_i X_j}(x_i, x_j)\, dx_i dx_j$ and the central mixed moment of order (n, m) as

$$\mu_{x_i x_j}^{(n,m)} = E_{\boldsymbol{X}}\left\{(x_i - m_{x_i})^n \left(x_j - m_{x_j}\right)^m\right\} = \int\int (x_i - m_{x_i})^n \left(x_j - m_{x_j}\right)^m p_{X_i X_j}(x_i, x_j)\, dx_i dx_j$$

where the two-dimensional $p_{X_i X_j}(x_i, x_j)$ is obtained by saturating $p_{\boldsymbol{X}}(\boldsymbol{x})$ over dimensions other than i and j

Now we talk about...

Random processes for computing *the unknown*

While the spectral properties of deterministic signals can be evaluated by calculating an inner product, either

- as an integral when an analytic expression of the signal is known, or
- as a finite summation when it is a discrete-time sequence

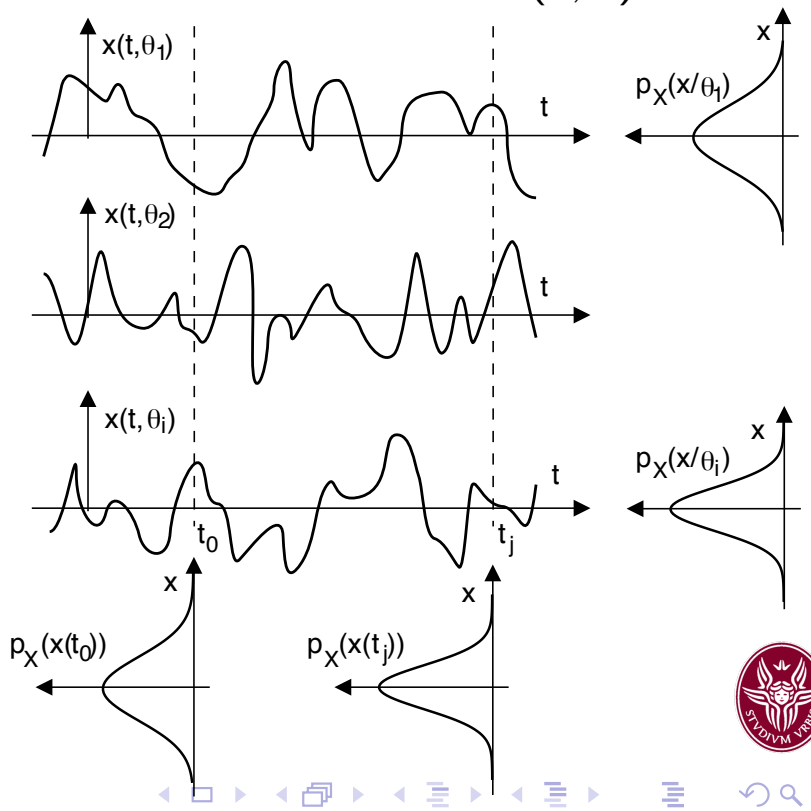
on the other hand if the signal is *not known in advance*, we have to resort to some *probabilistic description* of it, and work on it

Such a signal is called *a member* of a *random process* and is indicated as $x(t, \theta)$, where

- $t \in \mathcal{T}$ is some time instant, and
- $\theta \in \Theta$ is a r.v. identifying the process member

Therefore a specific member $x(t, \theta_i)$ is known only after the knowledge of $\theta_i \in \Theta$, and can be regarded as a deterministic signal for which a conditional pdf $p_X(x/\theta_i)$ can be defined (at the rigth in figure)

If, on the other hand, we fix a time instant t_j, the value $x(t_j, \theta)$ is a random variable, whose realization depends on that of $\theta \in \Theta$; therefore, the density $p_X(x(t_j))$ (independent of θ) is defined (at the bottom in figure)



Ensamble and time averages

Ensamble averages With reference to any (specific) time instant t_j we can compute moments of the process *as a function of time* by operating an ensamble average, which can be written as

$$m_X^{(n)}(t_j) = E_\Theta\left\{x^n(t_j,\theta)\right\} = \int_{-\infty}^{\infty} x^n(t_j,\theta)\, p_\Theta(\theta)\, d\theta = \int_{-\infty}^{\infty} x^n p_X(x(t_j))\, dx$$

where the last equality indicates how the statistical variability of x^n is fully described by the p.d.f. $p_X(x(t_j))$ of $x(t_j,\theta)$ as $\theta \in \Theta$ varies

Time average Alternatively, we can fix a particular θ_i realization of Θ, such that time averages can be calculated for any a single member $x(t,\theta_i)$, noted by a line above $\overline{(.)}$ the averaged quantity:

$$\overline{x^n(t,\theta_i)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^n(t,\theta_i)\, dt$$

In particular, we find

- the *mean value* $\overline{x(t,\theta_i)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t,\theta_i)\, dt$ and
- the *power* (or *quadratic mean*) $\overline{x^2(t,\theta_i)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t,\theta_i)\, dt$

We note that a generic time average:

- does not depend on time;
- is a random variable, as it depends on the realization of Θ.

Time average calculated as ensemble average

The extraction from $x(t, \theta_i)$ of a value at a random instant $t \in \mathcal{T}$ defines a further random variable, described by the (conditional) p.d.f. $p_X(x/\theta_i)$, which we drewn to the right of the previous figure

If the $p_X(x/\theta_i)$ is known, the temporal averages of order n can be calculated (for that member) as the respective moments:

$$\overline{x^n(t, \theta_i)} = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^n(t, \theta_i)\, dt = \int_{-\infty}^{\infty} x^n p_X(x/\theta_i)\, dx =$$
$$= E_{X/\Theta=\theta_i}\{x^n\} = m_X^{(n)}(\theta_i)$$

This is in fact equivalent to carrying out a weighted average, in which each possible value of x is weighted by its probability $p_X(x/\theta_i)\, dx$

It is the same concept applied to the equivalence given at pag. 8

Stationary and ergodic processes

Stationarity Is when $p_X(x(t_j))$ does not depend on t_j, i.e. $p_X(x(t_j)) = p_X^{\mathcal{T}}(x)$ for any $t_j \in \mathcal{T}$. Then all ensemble averages are time independent, that is $m_X^{(n)}(t) = m_X^{(n)}$ for $\forall t \in \mathcal{T}$, and the $p_X(x(t_j))$ at the bottom of the drawing are all equal

- Suppose now to divide $x(t, \theta_i)$ into time intervals and to calculate the time averages for each interval:
  - if they are all (almost) equal to each other, and equal to time average $m_X^{(n)}(\theta_i)$ of the entire member, then the member is (*individually*) stationary
  - obviously, if all members are individually stationary, so is the process to which they belong

Egodicity Is when *each member of a stationary process is statistically representative of all the others*

- This occurs when the p.d.f. $p_X(x/\theta_i)$ for any member does not depend on θ_i, so that $p_X(x/\theta_i) = p_X^{\Theta}(x)$
- Since $p_X(x/t_j) = p_X^{\mathcal{T}}(x)$ also holds for the stationarity, then $p_X^{\Theta}(x) = p_X^{\mathcal{T}}(x) = p_X(x)$, so that the time averages $m_X^{(n)}(\theta_i)$ (calculated as moments) are identical for all θ_i, and also identical to the ensemble averages $m_X^{(n)}(t_j)$ calculated for any instant

We can therefore state the definition of ergodicity:

> *A stationary process is ergodic if the time average calculated on any realization of the process* ***coincides*** *with the ensemble average relative to a random variable extracted at any instant (due to stationarity) from the set of its members*

Examples

We are now able to calculate time averages *without knowing* the waveform of the processes's members:

Power of a signal For ergodic processes it is equal to the moment of 2^{nd} order:

$$\mathcal{P}_X(\theta) = \overline{x^2(\theta)} = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x^2(t,\theta)\,dt = \int_{-\infty}^{\infty} x^2 p_X(x/\theta)\,dx =$$
$$= \int_{-\infty}^{\infty} x^2 p_X(x)\,dx = m_X^{(2)} = E\left\{x^2\right\} = \mathcal{P}_X$$

Mean value It can be calculated as the *first-order* moment m_x of the r.v. x extracted from the process:

$$\bar{x}(\theta) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t,\theta)\,dt = \int_{-\infty}^{\infty} x p_X(x/\theta)\,dx =$$
$$= \int_{-\infty}^{\infty} x p_X(x)\,dx = E\{x\} = m_X$$

Power, variance, quadratic mean and effective value From the above results and that of page 9 we can write

$$\mathcal{P}_X = m_X^{(2)} = \sigma_x^2 + (m_x)^2$$

- for signals with zero mean ($m_x = 0$) we obtain $\mathcal{P}_X = \sigma_x^2$
- the effective value $\sqrt{\mathcal{P}_X}$ coincides with the standard deviation σ_x
- the root of the power is also often referred to as the *root mean square* (RMS) value, defined as $x_{RMS} = \sqrt{\mathcal{P}_X} = \sqrt{\overline{x^2(t)}}$, that is the root of the *square mean* (over time)
- If the signal has zero mean, then x_{RMS} coincides with the effective value; if $x(t)$ is a member of an ergodic process with zero mean, x_{RMS} coincides with the standard deviation

Summarizing

- If a process is ergodic, it is also stationary, but not the other way around
  - Example: if $x(t,\theta) = C_\theta$ is equal to a (random) constant, then it is certainly stationary, but as $p_X(x/\theta) = \delta(x - C_\theta)$, it is not ergodic
- If a process is ergodic, then it is possible to:
  - calculate the ensemble averages in the form of time averages on a single actual realization, *or*
  - obtain the time averages of any realization starting from the ensemble averages, having the statistics $p_X(x)$, *and also*
  - estimate the p.d.f. starting from the histogram of the values extracted from any member.
- If the equality between ensemble and temporal averages exists only up to to a certain order and not beyond, the process *is not* ergodic *in strict sense*
  - as far as telecommunications are concerned, the ergodicity property in a *wide sense* is often sufficient, i.e. limited to the 2^{nd} order, which guarantees

$$\overline{x(t)} = E\{x\} = m_x \qquad \text{and} \qquad \overline{x^2(t)} = E\left\{x^2\right\} = m_x^{(2)}.$$

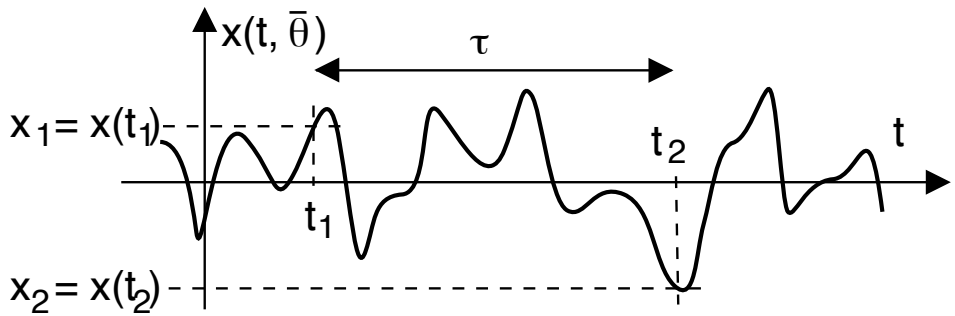
Now we talk about...

1. **Probability and statistics**
   - Building a ground of common terms
   - Stationary and ergodic processes
   - Correlation and covariance for signals
2. **A common point of view**
   - Autocorrelation and intercorrelation
   - Geometry and adaptation
3. **Power density spectrum**
   - Wiener's theorem
   - Multidimensional Gaussian and process
   - Spectral estimation
4. **Filtering and combining of signals and processes**
   - Spectral density at the output of a filter
   - Statistical characteristics at the output of a filter
   - Sum and product of random and deterministic signals

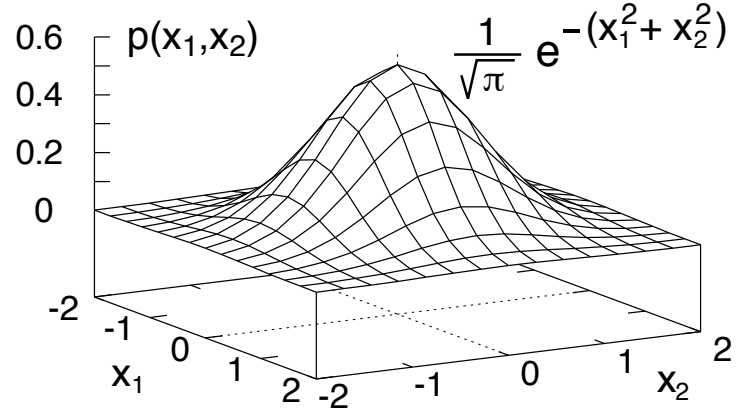
From a process to a 2D random variable

Probability and statistics can be further used in signal processing after the definition of a new function of a signal called *autocorrelation*, used to evaluate the power spectral density for deterministic and random signals in a unified way

To this aim we consider two instants t_1 and $t_2 = t_1 + \tau$ and the random variables $x_1 = x(t_1)$, $x_2 = x(t_2)$ extracted from a θ member of process $x(t, \theta)$, or their values taken from a specific member $x(t, \bar{\theta})$



As $\theta \in \Theta$ varies, the *two-dimensional* r.v. (x_1, x_2) is described by a *joint* probability density $p_{X_1 X_2}(x_1 x_2; t_1 t_2)$, which also depends on the instants t_1 and t_2, subtending a unitary volume (i.e. $\iint p(x_1, x_2)\, dx_1 dx_2 = 1$), and whose 3D graph describes where in the $x_1 x_2$ plane any couple of values is more or less probable



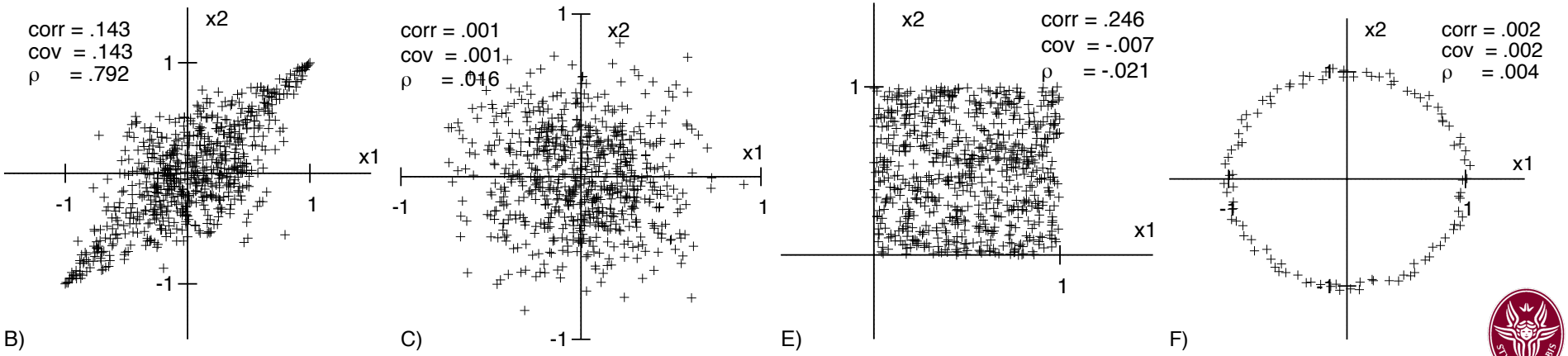
Correlation between random variables

The joint p.d.f. $p_{X_1X_2}(x_1x_2; t_1t_2)$ is now used in the computation of the *mixed moments* of the 2D r.v. (x_1, x_2), that is, an expected value in which the pairs of values are weighted according to their probability of occurring *together*

The mixed moment *of order (1,1)* (page 11) $m_{XX}^{(1,1)}(t_1, t_2)$ between the r.v. is called *correlation* and is defined as

$$m_{XX}^{(1,1)}(t_1, t_2) = E_{X_1X_2}\{x_1x_2\} = \iint x_1x_2 \cdot p_{X_1X_2}(x_1x_2; t_1t_2)\, dx_1\, dx_2$$

Statisticians use the same term to define a value computed from a finite set of joint observations, such as $corr(x, y) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i$, which is formally equivalent to a time average rather than an ensemble average; in this case instead of the histogram for 1D variables there are *scattering diagrams*



whose meaning could possibly be discussed verbally

Covariance, statistical independence and uncorrelatedness

If the two r.v. are *statistically independent* (i.e. $p_{X_1X_2}(x_1, x_2; t_1, t_2) = p(x_1)\, p(x_2)$), then the integral defining the correlation *factors*, and $m_{XX}^{(1,1)}(t_1, t_2)$ reduces to the product of two mean values:

$$\begin{aligned} m_{XX}^{(1,1)}(t_1, t_2) &= E\{x_1, x_2\} = \iint x_1 x_2 p(x_1)\, p(x_2)\, dx_1 dx_2 = \\ &= \int x_1 p(x_1)\, dx_1 \cdot \int x_2 p(x_2)\, dx_2 = E\{x_1\} E\{x_2\} = m_{X_1} m_{X_2} \end{aligned}$$

Covariance Is indicated as $\sigma(x_1, x_2)$ and is equal to the correlation $m_{XX}^{(1,1)}(t_1, t_2)$ minus the term $m_{X_1} m_{X_2}$, corresponding to the *central* mixed moment between the two r.v.:

$$\begin{aligned} \sigma(x_1, x_2) &= E\{(x_1 - m_{X_1})(x_2 - m_{X_2})\} = \\ &= E\{x_1 x_2\} - E\{x_1 m_{X_2}\} - E\{m_{X_1} x_2\} + E\{m_{X_1} m_{X_2}\} = \\ &= E\{x_1 x_2\} - m_{X_1} m_{X_2} = m_{XX}^{(1,1)}(t_1, t_2) - m_{X_1} m_{X_2} \end{aligned}$$

Uncorrelatedness By combining the above results we can verify that

> *If two random variables x_1 and x_2 are statistically independent, their covariance $\sigma(x_1, x_2)$ is null, and are therefore said to be* UNCORRELATED

This is valid in *one direction* only, since if $\sigma(x_1, x_2) = 0$ the two r.v. *are not necessarily* statistically independent. Uncorrelatedness *implies* independence only for jointly Gaussian r.v.

Correlation of an ergodic stationary process

If the process from which x_1 and x_2 are extracted is *stationary*

- then the joint p.d.f. depends only on the difference $\tau = t_2 - t_1$ between the instants t_2 and t_1 and therefore also the correlation depends only on τ:

$$m_{XX}^{(1,1)}(t_1, t_2) = E\{x_1 x_2\} = \iint x_1 x_2 \cdot p_{X_1 X_2}(x_1 x_2; \tau)\, dx_1\, dx_2 = m_{XX}^{(1,1)}(\tau)$$

If the process is also *ergodic*

- then $m_{XX}^{(1,1)}(\tau)$ has the same values of the corresponding time average
  On the other hand, if $p_{X_1 X_2}(x_1 x_2; \tau)$ is not known, but we have some realization of the process instead
  - then the correlation can be obtained by the *time average* $\overline{x(t, \theta_i)\, x(t + \tau, \theta_i)}$ calculated for any θ_i realization, is indicated as $\mathcal{R}_x(\tau)$, and corresponds to

$$\mathcal{R}_x(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t, \theta_i)\, x(t + \tau, \theta_i)\, dt \quad \forall \theta_i \in \Theta$$

Since for stationary and ergodic processes the above formulas provide the same result, for both of them the correlation will be indicated as $\mathcal{R}_x(\tau)$

Now we talk about...

1. **Probability and statistics**
   - Building a ground of common terms
   - Stationary and ergodic processes
   - Correlation and covariance for signals
2. **A common point of view**
   - Autocorrelation and intercorrelation
   - Geometry and adaptation
3. **Power density spectrum**
   - Wiener's theorem
   - Multidimensional Gaussian and process
   - Spectral estimation
4. **Filtering and combining of signals and processes**
   - Spectral density at the output of a filter
   - Statistical characteristics at the output of a filter
   - Sum and product of random and deterministic signals

Autocorrelation and intercorrelation

of deterministic signals

Autocorrelation Is the result of evaluating the latest *time average* for a *deterministic* power signal $x(t)$, and is again indicated by $\mathcal{R}_x(\tau)$

$$\mathcal{R}_x(\tau) = \overline{x(t)\,x(t+\tau)} = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,x(t+\tau,)\,dt$$

while for an *energy* signal the autocorrelation is defined as

$$\mathcal{R}_x(\tau) = \int_{-\infty}^{\infty} x^*(t)\,x(t+\tau)\,dt$$

where the conjugate generalizes the expression to the case of complex signals

- note that the autocorrelation is a form of *dot product* that calculates the *mutual* energy (or power) of signal $x(t)$ with respect to its *anticipated copy* (speech example)
- a high value of $\mathcal{R}_x(\tau)$ for some τ indicates that the two copies of the signal are *similar*

Intercorrelation The same concepts are even more valid when operating on two *different* signals $x(t)$ and $y(t)$. For *energy* signals we write

$$\mathcal{R}_{xy}(\tau) = \int_{-\infty}^{\infty} x^*(t)\,y(t+\tau)\,dt$$

and for *power* signals $\mathcal{R}_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x^*(t)\,y(t+\tau)\,dt$.

- If we find $\mathcal{R}_{xy}(\tau) = 0$ for some τ then the signals are said to be
  - *orthogonal*, with reference to the signal space for which $\mathcal{R}_{xy}(\tau)$ is a dot product, but also
  - *uncorrelated*, with reference to the statistical aspect for signals with zero mean

Links with the convolution

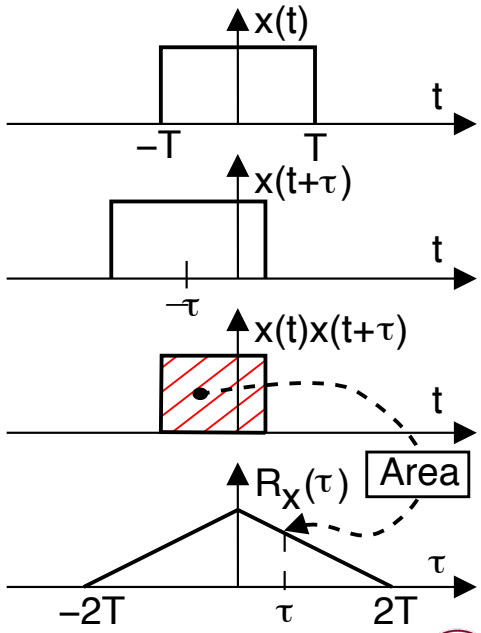
Link with convolution $\mathcal{R}_x(\tau)$ and $\mathcal{R}_{xy}(\tau)$ can also be seen as time-dependent *signals*, in a way similar to the convolution integral whose result is a function of time. The two operators are interconnected, as by means of a change of variables it can be shown that

$$\mathcal{R}_{xy}(\tau) = \int_{-\infty}^{\infty} x^*(t)\, y(t+\tau)\, dt = x^*(-t) * y(t)$$

Graphic construction

The last observation invites us to draw the graphic construction shown, illustrating the calculus of the autocorrelation for $x(t) = rect_{2T}(t)$, in a similar way to how convolution operates, BUT:

- *no axis inversion* is now performed, and the translation is *backward* (time advance) rather than forward
- for a real rectangle $x(t) = x^*(-t)$ holds, and therefore the operation is equivalent to calculating $x(t) * x(t)$, but unlike the convolution, in the second line of the graph the term $x(t+\tau)$ for $\tau > 0$ is shifted to the *left*
- the third line shows the product of the signals above of it, whose integral calculates the area, providing the value of $\mathcal{R}_x(\tau)$ on the right, as in the picture



Properties of the autocorrelation

Invariance with respect to time shifts If $y(t) = x(t+\theta)$ then $\mathcal{R}_y(\tau) = \mathcal{R}_x(\tau)$, or

Autocorrelation does not take into account the phase spectrum

Actually, $x(t)$ and $y(t)$ also have the same energy density $\mathcal{E}_x(f) = \mathcal{E}_y(f)$

Temporal extension If $x(t)$ has a finite duration then $\mathcal{R}_x(\tau)$ has twice the duration

– if $x(t)$ is an energy signal with unlimited duration then $\lim_{t\to\infty} x(t) = 0$ and $\mathcal{R}_x(\tau)$ also

– if $x(t)$ is an ergodic process member (a power signal), then $\lim_{\tau\to\infty} \mathcal{R}_x(\tau) = 0$, but

- **Periodic signals** Their $\mathcal{R}_x(\tau)$ is also periodic with the same period, and defined as $\mathcal{R}_x(\tau) = \sum_{n=-\infty}^{\infty} \mathcal{R}_x^T(\tau - nT)$ where $\mathcal{R}_x^T(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} x^*(t)\, x(t+\tau)\, dt$
- **Continuous component** If $x(t) = x_0(t) + a$ with $E\{x_0(t)\} = 0$ then $m_x = a$, and $\mathcal{R}_x(\tau) = \mathcal{R}_{x_0}(\tau) + a^2$, so that $\mathcal{R}_x(\tau)$ does not vanishes but $\lim_{\tau\to\infty} \mathcal{R}_x(\tau) = m_x^2$

Maximum in the origin It results $\mathcal{R}_x(0) = \max_\tau \{\mathcal{R}_x(\tau)\}$ and

$$\mathcal{R}_x(0) = \begin{cases} \int_{-\infty}^{\infty} |x(t)|^2\, dt = \mathcal{E}_x > |\mathcal{R}_x(\tau \neq 0)| & \text{if } x(t) \text{ is of energy} \\ \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} |x(t)|^2\, dt = \mathcal{P}_x \geq |\mathcal{R}_x(\tau \neq 0)| & \text{if } x(t) \text{ is of power} \end{cases}$$

Conjugate symmetry It turns out $\mathcal{R}_x(\tau) = \mathcal{R}_x^*(-\tau)$, so $\mathcal{F}\{\mathcal{R}_x(\tau)\}$ is real

– if $x(t)$ is real then $\mathcal{R}_x(\tau) = \mathcal{R}_x(-\tau)$ i.e. the $\mathcal{R}_x(\tau)$ of a real signal is *real even*

– Finally, $\mathcal{R}_{xy}(\tau) = \mathcal{R}_{yx}^*(-\tau)$

Now we talk about...

Pearson correlation coefficient

It is defined as

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

normalizing the σ_{xy} covariance of two r.v. x and y with respect to their standard deviations σ_x and σ_y, thus producing a limited range of values $-1 \leq \rho_{xy} \leq 1$. For it there is a geometric interpretation:

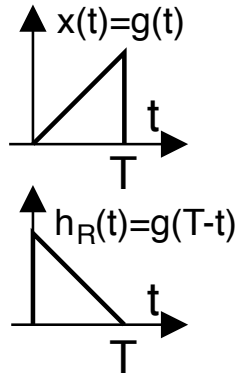
- the standard deviation σ_x can be thought of as to *the norm* $\|\overline{x}\|$ of x, and
- σ_{xy} can be seen as the *dot product* $\langle \overline{x}, \overline{y} \rangle$ between x and y; indeed
  - if x comes from an ergodic process with zero mean then σ_x^2 equals the *power* of its members, and
  - if x and y are extracted from jointly ergodic processes then the σ_{xy} equals the *intercorrelation* function (evaluated in zero), i.e. their *mutual power*
- two r.v. can be said to be *orthogonal* when $\rho_{xy} = 0$, while $\rho_{xy} = \pm 1$ means that x and y are always proportional each other
  - orthogonality $\rho_{xy} = 0$ only expresses the absence of a *linear* relationship between x and y, as for the case F) at page 21

We also mention the formal extension of *Schwartz's inequality*, if

- the concept of cosine between x and y is associated to ρ_{xy}
- the condition $-1 < \rho_{xy} < 1$ allows us to state that $|\sigma_{xy}| \leq \sigma_x \sigma_y$

The matched filter

- A filter is said to be *matched* when its impulse response $h(t)$ is defined starting from a *specific* signal $g(t)$ with limited duration T, which one wishes to identify even when immersed in a particularly intense noise
- by setting $h(t) = g(T - t)$ the convolution operation evaluates the *intercorrelation* between $g(t)$ and the input $x(t)$ to the filter
  - if $x(t) = g(t)$ then the filter output computes the *autocorrelation* $\mathcal{R}_g(\tau)$ with a maximum at the origin (actually, at time $t = T$)
  - if $x(t) = g(t) + n(t)$ then the very small intercorrelation $\mathcal{R}_{gn}(\tau)$ is also added to the output
  - if $x(t) = n(t)$ then only intercorrelation $\mathcal{R}_{gn}(\tau)$ is present at the output



- the functionality is best explained by means of *a simulation*, and for those who dare to use *Octave*, this is the code
  - then if someone was able to try it also with Matlab and could tell me the outcome, I would be very grateful
    - ★ it could be enough to replace the # at the beginning of the comments with a %, who knows why Matlab wants it like this!
- an in-depth analysis is provided in the book, demonstrating how the matched filter is *optimal* from the point of view of the reduction of the signal-to-noise ratio, with a reasoning based on the *Schwartz inequality*

Now we talk about...

How to get the spectral density of unknown signals

Wiener's theorem allows to characterize the spectral density for both the case of random processes as well as deterministic signals

> *The power density spectrum $\mathcal{P}_x(f)$ (or energy density $\mathcal{E}_x(f)$) of a deterministic or random signal $x(t)$ is equal to the Fourier transform of its autocorrelation function, namely $\mathcal{P}_x(f) = \mathcal{F}\{\mathcal{R}_x(\tau)\}$*

Proof for energy signals

$$\begin{aligned}\mathcal{R}_x(\tau) &= \int_{-\infty}^{\infty} x^*(t)\, x(t+\tau)\, dt = \int_{-\infty}^{\infty} X^*(f)\, X(f)\, e^{j2\pi f\tau}\, df = \\ &= \mathcal{F}^{-1}\{X^*(f)\, X(f)\} = \mathcal{F}^{-1}\{\mathcal{E}_x(f)\}\end{aligned}$$

where we first applied Parseval's theorem, then the transform property for time shift, and finally recognized $X^*(f)\, X(f)$ as the $\mathcal{E}_x(f)$ energy density

As anticipated, the theorem also holds (demonstrations) for

- power signals, for which the autocorrelation function $\mathcal{R}_x(\tau)$ to use is that given by $\mathcal{R}_x(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t)\, x(t+\tau)\, dt$
- ergodic processes, as every member of it has the same $\mathcal{P}_x(f)$, obtained from
  - the above $\mathcal{R}_x(\tau)$ of any of them, or
  - the mixed moment $m_{XX}^{(1,1)}(\tau) = E\{x(t)\, x(t+\tau)\}$ calculated as an ensemble mean

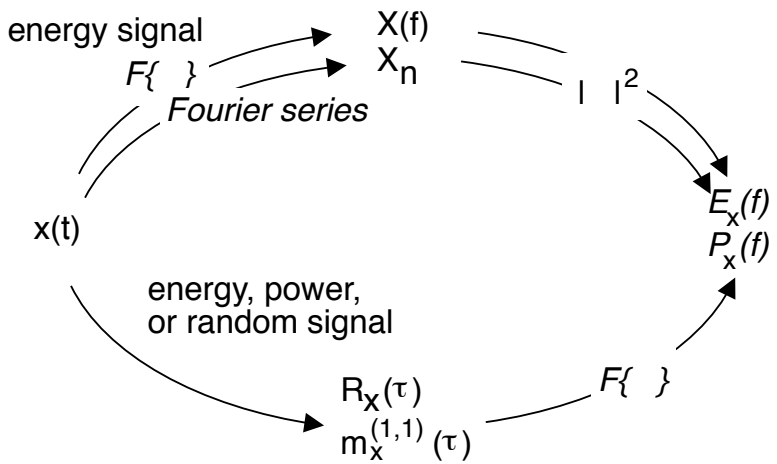
Discussion

Wiener's theorem is a very powerful tool which opens an *alternative way* to verify and extend known results

- for example the property $\mathcal{R}_x(0) = \mathcal{P}_x$ can now be derived from the one of the initial value

$$\mathcal{R}_x(0) = \mathcal{F}^{-1}\{\mathcal{P}_x(f)\}\big|_{\tau=0} = \int_{-\infty}^{\infty} \mathcal{P}_x(f)\, e^{j2\pi f\tau}\, df\Big|_{\tau=0} = \int_{-\infty}^{\infty} \mathcal{P}_x(f)\, df = \mathcal{P}_x$$

- when applied to periodic or energy signals it allows to follow an *alternative paths* for the calculation of the corresponding power (or energy) density, as shown
- being able to define a $\mathcal{P}_x(f)$ also for random processes and power signals allows us to analyze the filtering effects also for this class of signals



Let us now apply the relationship between $\mathcal{P}_x(f)$ and $\mathcal{R}_x(\tau)$ expressed by the Wiener's theorem to some notable cases

Periodic signals

In this case expressing $x(t)$ with period T by its Fourier series $x(t) = \sum_{n=-\infty}^{\infty} X_n e^{j2\pi nFt}$ and substituting it into $\mathcal{R}_x(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} x^*(t)\, x(t+\tau)\, dt$ we arrive at

$$\mathcal{R}_x(\tau) = \sum_{n=-\infty}^{\infty} |X_n|^2 e^{j2\pi nF\tau}$$

that is, also $\mathcal{R}_x(\tau)$ is expressed as a Fourier series, so that it is periodic in turn, as already noted. Thus the power density $\mathcal{P}_x(f)$ is equal to

$$\mathcal{P}_x(f) = \mathcal{F}\{\mathcal{R}_x(\tau)\} = \sum_{n=-\infty}^{\infty} |X_n|^2 \delta(f - nF)$$

confirming the Parseval's theorem

Continuous component
if the signal can be written as $x(t) = x_0(t) + a$ with $E\{x_0(t)\} = 0$ then we get $\mathcal{R}_x(\tau) = \mathcal{R}_{x_0}(\tau) + a^2$ and therefore

$$\mathcal{P}_x(f) = \mathcal{F}\{\mathcal{R}_x(\tau)\} = \mathcal{P}_{x_0}(f) + a^2\delta(f)$$

that is, the spectral density has a pulse with area a^2 in the origin

- or, from the opposite point of view, a pulse in the origin for $\mathcal{P}_x(f)$ reveals a continuous component in $x(t)$

Band-limited white process

A *noise* process $n(t)$ is called *white* when it has a *constant* power density, and bandwidth limitation in between $\pm W$ allows to write

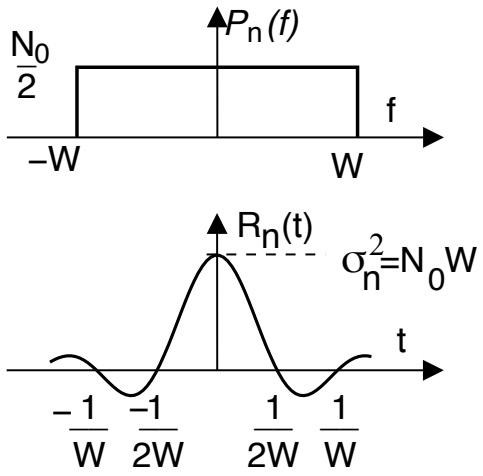
$$\mathcal{P}_n(f) = \frac{N_0}{2}\mathrm{rect}_{2W}(f)$$

with autocorrelation

$$\mathcal{R}_n(t) = \mathcal{F}^{-1}\{\mathcal{P}_n(f)\} = N_0 W \mathrm{sinc}(2Wt)$$

so that it can be verified that

$$\mathcal{R}_n(0) = \mathcal{P}_n = \int_{-\infty}^{\infty} \mathcal{P}_n(f)\,df = \int_{-W}^{W} \frac{N_0}{2}\,df = N_0 W = \sigma_n^2$$

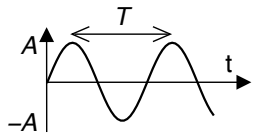
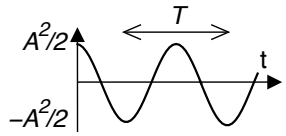
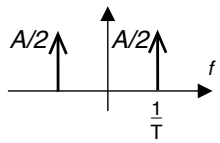
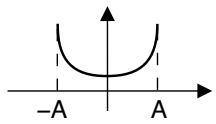
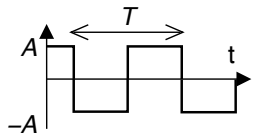
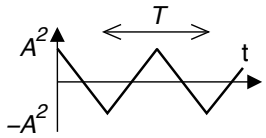
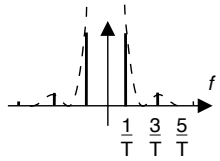
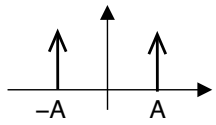
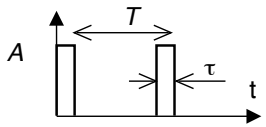
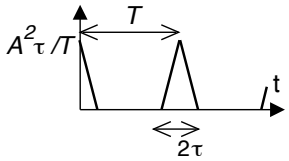
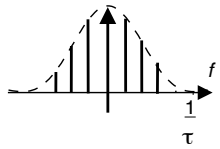
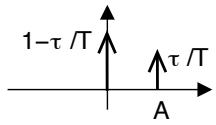
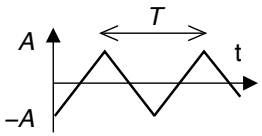
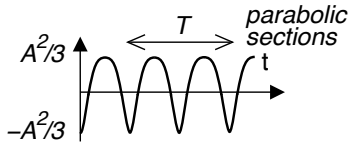
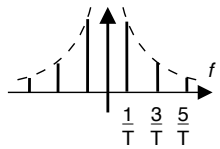
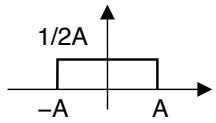
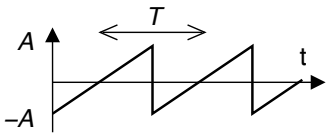
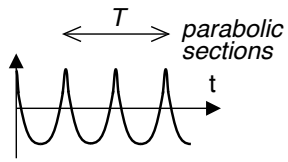
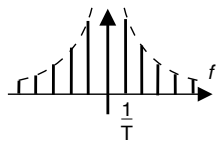
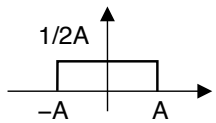
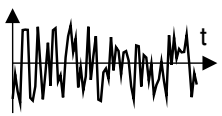
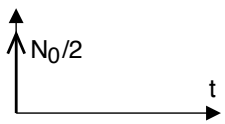
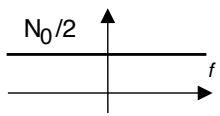
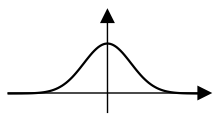


- As $\mathcal{R}_n(t)$ zeroes for $t = k/2W$, samples of $n(t)$ spaced by $k/2W$ are *uncorrelated*
  - if $n(t)$ is also *Gaussian* these samples are also statistically independent (2 page 40)
- As W increases $\mathcal{R}_n(t)$ goes to zero *faster*, so that
  - noise samples remains correlated for a shorter time, or
  - the correlation between samples spaced by a fixed time interval becomes smaller

Thinking now at $n(t)$ as the output of an ideal low pass filter with $H(f) = rect_{2W}(f)$ when an infinite bandwidth process $p(t)$ (with $\mathcal{R}_p(t) = \delta(t)$) is put at its input

- it is evident that correlation $\mathcal{R}_n(\tau)$ between $n(t)$ samples spaced by $\tau \neq 0$ is a consequence of the memory introduced by the filter impulse response, since
- convolution between $p(t)$ and $h(t)$ makes the output values to be a *linear combination* of the (past) input values

Other examples

signal	waveform	autocorrelation	spectral density	probability density func.
sinusoid				
square wave				
rectangular pulses				
triangular wave				
sawtooth				
white gaussian noise				

Now we talk about...

Multidimensional Gaussian r.v.

A realization of an $n-$dimensional gaussian r.v. $\boldsymbol{X}$ is given by a column vector $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$ whose elements are n different uni-dimensional marginal r.v. x_i, $i = 1, 2, \cdots, n$, all Gaussian. Their *joint* p.d.f. is expressed as

$$p_X(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma}_x)}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m}_x)^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x} - \boldsymbol{m}_x)\right\}$$

where $\boldsymbol{m}_x$ is the vector of mean values of the marginals and $\boldsymbol{\Sigma}_x$ is their *covariance matrix* whose $n \times n$ elements are equal to $\sigma_{ij} = E\left\{\left(x_i - m_i\right)\left(x_j - m_j\right)\right\}$, that is

$$\boldsymbol{m}_x = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_x = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

- For $\boldsymbol{x} = \boldsymbol{m}_x$ the exponent vanishes and the term $1/\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma}_x)}$ is the *maximum height* of $p_X(\boldsymbol{x})$, and
- for $\boldsymbol{x} \neq \boldsymbol{m}_x$ the exponent itself is a *quadratic form* (actually, a paraboloid) always positive, which increases as $|\boldsymbol{x} - \boldsymbol{m}_x|$ increases, so that $p_X(\boldsymbol{x})$ decreases

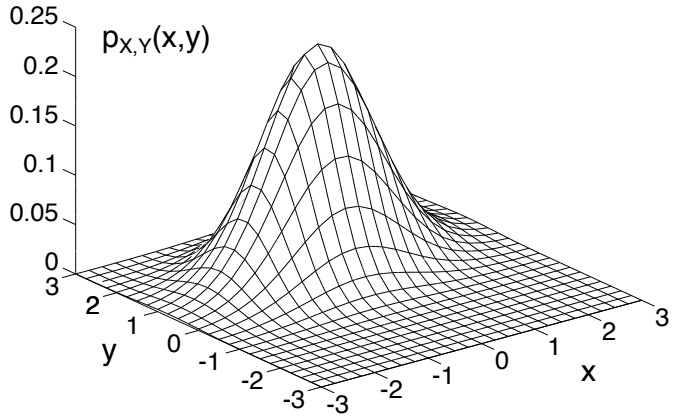
An example of a bidimensional Gaussian r.v.

- Figure a) shows a 3D plot for a two-dimensional Gaussian p.d.f. $p_{XY}(x, y)$, with

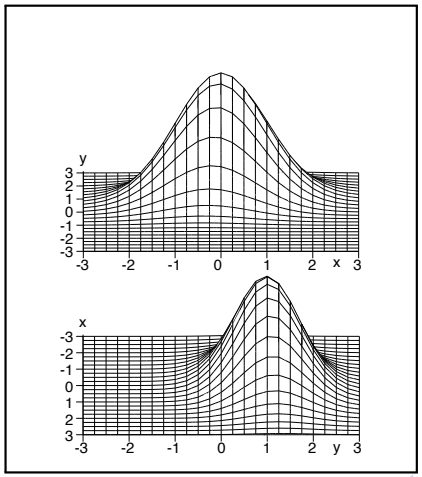
$$\boldsymbol{m} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$$

  where the *asymmetry* is due to the values $\sigma_x^2 \neq \sigma_y^2$, and the centering due to $m_y \neq 0$
- fig. b) shows the same p.d.f. from a point of view *parallel* to the axes
- fig. c) reports the *level curves*, showing that
  - the quadratic form in the exponent determine *elliptical* contours for the surface of $p_{XY}(x, y)$, and
  - the length of the axes of the ellipses is related to σ_x and σ_y, while the inclination depends on the covariance σ_{xy} (check it by graphics with Octave)

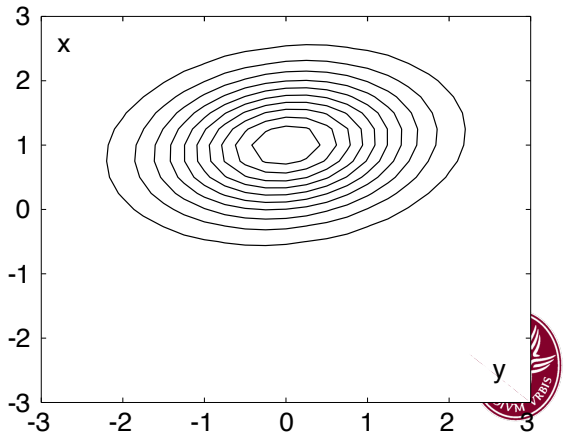
a)



b)



c)



Statistical independence for uncorrelated Gaussian

We want to demonstrate what was stated at the end of page 22, that is,

uncorrelatedness implies statistical independence only for jointly Gaussian r.v.

Indeed, if the marginal gaussian r.v. are uncorrelated ($\sigma_{x_i x_j} = 0$ with $i \neq j$) then

- the covariance matrix $\mathbf{\Sigma}_x$ turns out to be *diagonal*
- the same for its inverse, whose elements are now equal to $1/\sigma^2_{x_i}$
- we obtain $\det(\mathbf{\Sigma}_x) = \prod_{i=1}^n \sigma^2_{x_i}$

Thus

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_{x_i}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n \frac{(x_i - m_{x_i})^2}{\sigma^2_{x_i}}\right]\right\}$$

which is equal to the product of all the marginals $p(x_i) = \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp\left\{-\frac{1}{2}\frac{(x_i - m_{x_i})^2}{\sigma^2_{x_i}}\right\}$

- since the factorization of a joint p.d.f. in terms of its marginals is precisely the condition of statistical independence, the claim is true

---

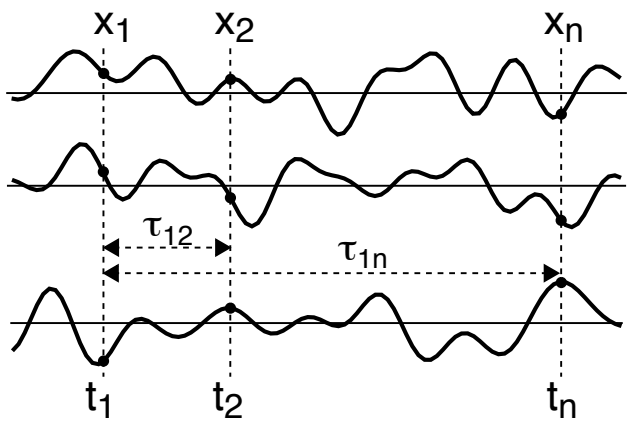
Let us verify the result for an (x, y) *pair* of statistically independent Gaussian r.v., with zero mean and variance σ_x^2 and σ_y^2 respectively. We get:

$$p_X(x)\, p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right]$$

Gaussian process

A fundamental class of random signals is that of

- ergodic processes whose first order p.d.f. is Gaussian, and
- a set of x_i r.v. extracted from its members at different instants t_i are jointly Gaussian r.v., so that
- if a random vector $\mathbf{x}$ is made from the x_i, their joint p.d.f. is that at page 38



Stationarity guarantees that

- all the elements of the vector $\boldsymbol{m}_x$ of mean values are equal to $m_x = E\{x(t)\}$, and
- the elements of the $\boldsymbol{\Sigma}_x$ covariance matrix are equal to the values $\sigma_x(\tau) = E\{(x(t) - m_x)(x(t+\tau) - m_x)\}$ of the process covariance, evaluated at the time intervals τ_{ij} between the instants in which the marginal r.v. x_i and x_j are extracted
- In other words, the values σ_{ij} appearing in $\boldsymbol{\Sigma}_x$ are obtained as $\sigma_{ij} = \sigma_x(\tau_{ij})$, and the variance $\sigma_x^2 = \sigma_x(0)$ appears on the entire diagonal *(example in the book)*

Since the process is Gaussian, the two quantities $\boldsymbol{m}_x$ and $\boldsymbol{\Sigma}_x$ describe it completely, and for an ergodic process they can be *estimated* by any member of the process

Now we talk about...

Spectral estimation

Wiener's theorem allow us to know the power spectral density for an ergodic process when its autocorrelation $m_X^{(1,1)}(\tau) = \mathcal{R}_X(\tau)$ is known

- if the ensemble statistics are unknown, but the process is ergodic, $\mathcal{R}_X(\tau)$ can be *estimated* from some of its available members
  - by means of a *short-time* time average computed on a signal window, or
  - as a *running* real-time estimate as shown here

Another solution is to estimate $\mathcal{P}_x(f)$ *without going through autocorrelation*

- using instead the squared Fourier transform $|X_T(f)|^2$ of a time-limited signal segment, as described in the *next slide*

But: some signals, although representative of many others, *cannot* be considered members of a stationary process *(for example, speech)*, and we are more interested on its short-time *spectral changes*

- in this case (in addition to performing the DFT on signal windows) we can use parametric techniques such as Linear Predictive Coding or LPC, which also uses temporal segments of the signal to evaluate a short-time autocorrelation function

Periodogram

Given a $x(t, \theta_i)$ realization of a process

- we select from it a window of length T and we define a signal of limited duration
$$x_T(t) = x(t, \theta_i) \operatorname{rect}_T(t)$$
  - now $x_T(t)$ is an *energy signal*, with transform $X_T(f)$ and energy density
$$\mathcal{E}_{x_T}(f) = |X_T(f)|^2$$
- under the assumption of stationarity an estimate $\widehat{\mathcal{P}}_x(f)$ of $\mathcal{P}_x(f)$ of the whole member is simply obtained as
$$\widehat{\mathcal{P}}_x(f) = \mathcal{P}_{x_T}(f) = \frac{|X_T(f)|^2}{T}$$
which takes the name of ***periodogram***, due to its original use for the discovery of periodicity within of a *noisy* signal
- as T tends to ∞, $\widehat{\mathcal{P}}_x(f)$ tends to the *true* power density
$$\lim_{T\to\infty} \frac{|X_T(f)|^2}{T} = \mathcal{P}_x(f)$$
of the process member $x_T(t, \theta_i)$ and, if it belongs to an ergodic process, to that of any other member

Bias, resolution and variance of the periodogram

Bias Having T not tending to ∞, $\widehat{\mathcal{P}}_x(f)$ is a *biased* estimate of $\mathcal{P}_x(f)$ *(the rect window on* $x(t)$ *is equivalent to a tri window applied to the* $\mathcal{R}_X(\tau)$ *estimate)*, giving

$$\widehat{\mathcal{P}}_x(f) = \mathcal{P}_x(f) * T\left(\operatorname{sinc}(fT)\right)^2$$

that is, a *temporal windowing* distortion effect happens, and $\widehat{\mathcal{P}}_x(f)$ for any value of f got *a bias*, which disappears if $T \to \infty$, because in this case $T\left(\operatorname{sinc}(fT)\right)^2 \to \delta(f)$

Spectral resolution It depends on the main lobe width of $\left(\operatorname{sinc}(fT)\right)^2$, i.e. $1/T$, and also disappears if $T \to \infty$, for the same reason

Variance of the estimate

Although $\widehat{\mathcal{P}}_x(f)$ reduces its bias as T increases, it still is a r.v. (it depends on the member θ), and its variance... *does not decrease* with increasings T, making the estimator *inconsistent*!

- It can be proven that the variance σ_T^2 of the estimate is equal to the value of $\mathcal{P}_x(f)$ itself, that is for each frequency value, the standard deviation of the value of $\widehat{\mathcal{P}}_x(f)$ is equal to $\sqrt{\mathcal{P}_x(f)}$, regardless of how large T is
- in other words, the estimator variance does not decreases as available data increases, and the reason is that in a numerical implementation using DFT, as T increases the number of frequency values that are calculated also increases

Solutions to this problem introduce a reduction in spectral resolution

- one can *smooth* the $\widehat{\mathcal{P}}_x(f)$ obtained, averaging the values on nearby frequencies: this operation corresponds to a *filtering in frequency*
- a different method involves dividing the observation interval in different sub-intervals, calculating the periodogram on each of them, and averaging the results

Example in the book

Now we talk about...

1. **Probability and statistics**
   - Building a ground of common terms
   - Stationary and ergodic processes
   - Correlation and covariance for signals
2. **A common point of view**
   - Autocorrelation and intercorrelation
   - Geometry and adaptation
3. **Power density spectrum**
   - Wiener's theorem
   - Multidimensional Gaussian and process
   - Spectral estimation
4. **Filtering and combining of signals and processes**
   - Spectral density at the output of a filter
   - Statistical characteristics at the output of a filter
   - Sum and product of random and deterministic signals

A review of results

In this last section of the set of slides, a series of results relating to the passage of signals, both deterministic and random, through a filter are listed

These concern

- the determination of the power (or energy) density at the output of a filter
  - where Wiener's theorem plays a great role in unifying the treatment for any type of signal
- the calculation of statistical indicators for the same output signal
- the result for the sum and product of random signals

You probably won't find the passages very interesting, which in fact we mostly leave inside the book: however, the series of results shown can constitute a sort of handbook to consult in case of doubt

Energy signals through a filter

We first evaluate the result for the output energy density $\mathcal{E}_y(f)$ (and the respective energy $\mathcal{E}_y$) when an energy signal $x(t)$ is input

We know that by Parseval's theorem it results $\mathcal{E}_y(f) = Y(f)\,Y^*(f)$; then since $Y(f) = X(f)\,H(f)$ we can write

$$\mathcal{E}_y(f) = \mathcal{E}_x(f)\,|H(f)|^2$$

By executing the Fourier antitransform of both sides we obtain

$$\mathcal{R}_y(\tau) = \mathcal{F}^{-1}\{\mathcal{E}_y(f)\} = \mathcal{F}^{-1}\left\{\mathcal{E}_x(f)\,|H(f)|^2\right\} = \mathcal{R}_x(\tau) * \mathcal{R}_h(\tau)$$

- as will be shown later, this result is valid (in the respective terms) also for the cases of a periodic or random signal
- $|H(f)|^2$ can also be seen as the *energy density of the filter*, i.e. $|H(f)|^2 = \mathcal{E}_h(f) = \mathcal{F}\{\mathcal{R}_h(\tau)\}$

As a corollary we have the following results, all equivalent for the purpose of calculating the total energy:

$$\mathcal{E}_y = \mathcal{R}_y(0) = \int_{-\infty}^{\infty}\mathcal{E}_y(f)\,df = \int_{-\infty}^{\infty}\mathcal{E}_x(f)\,|H(f)|^2\,df = \int_{-\infty}^{\infty}\mathcal{R}_x(\tau)\,\mathcal{R}_h(\tau)\,d$$

Periodic signals through a filter

Here the input signal $x(t)$ can be expressed as $x(t) = \sum_n X_n e^{j2\pi nFt}$ to which it corresponds

- a transform $X(f) = \sum_n X_n \delta(f - nF)$ and
- a power density $\mathcal{P}_x(f) = \sum |X_n|^2 \delta(f - nF)$

The output signal $y(t)$ is also periodic whith Fourier coefficients given by

- $Y_n = X_n H(nF)$, i.e.

$$|Y_n| = |X_n| \, |H(nF)| \, ; \quad \textit{and} \quad \arg(Y_n) = \arg(X_n) + \arg(H(nF))$$

Since the power density of $y(t)$ is equal to $\mathcal{P}_y(f) = \sum_n |Y_n|^2 \delta(f - nF)$, we get

$$\mathcal{P}_y(f) = \sum_n |X_n|^2 \, |H(nF)|^2 \delta(f - nF) = |H(f)|^2 \mathcal{P}_x(f)$$

Again, anti-transformation gives $\mathcal{R}_y(\tau) = \mathcal{R}_x(\tau) * \mathcal{R}_h(\tau)$

Ergodic processes and power signals

Also in this case it can be verified that $m_Y^{(1,1)}(\tau) = m_X^{(1,1)}(\tau) * \mathcal{R}_h(\tau)$, and so

$$\mathcal{P}_y(f) = \mathcal{P}_x(f)\,|H(f)|^2$$

The result obviously applies to any member of the process, for which as known it results $m_X^{(1,1)}(\tau) = \mathcal{R}_x(\tau)$, and therefore the upper relation is also valid for any power signal

Power gain
It is the name by which the ratio

$$|H(f)|^2 = \frac{\mathcal{P}_y(f)}{\mathcal{P}_x(f)}$$

is most often indicated, or $|H(f)|^2 = \frac{\mathcal{E}_y(f)}{\mathcal{E}_x(f)}$ in the case of energy signals

- $|H(f)|^2$ re-proposes in power or energy terms the input-output signal relation given by $H(f)$
- Otherwise in literature $|H(f)|^2$ is also referred to as
  - the *power response*, or even
  - *spectral density of the power response*

Now we talk about...

Mean value, variance, and amplitude density

It may be of interest to investigate about the statistical characterization of a filter output when an ergodic process is put at its input

Mean value Equal to the input one, times the *zero frequency gain* $H(0)$, i.e.

$$\begin{aligned} m_y &= E\{y(t)\} = E\{x(t) * h(t)\} = E\{x(t)\} * h(t) = \\ &= m_x \int_{-\infty}^{\infty} h(\tau)\, d\tau = m_x H(0) \end{aligned}$$

Variance By ergodicity $\mathcal{P}_y = \overline{y^2(t)} = E\left\{y^2\right\} = m_y^{(2)}$, thus we have

$$\sigma_y^2 = m_y^{(2)} - (m_y)^2 = \mathcal{P}_y - (m_y)^2$$

so that for $\mathcal{P}_y$ these alternative formulas holds

$$\begin{aligned} \mathcal{P}_y &= \mathcal{R}_y(0) = \int_{-\infty}^{\infty} \mathcal{P}_y(f)\, df = \int_{-\infty}^{\infty} \mathcal{P}_x(f) |H(f)|^2\, df = \\ &= \int_{-\infty}^{\infty} \mathcal{R}_x(\tau)\, \mathcal{R}_h(\tau)\, d\tau \end{aligned}$$

Probability density function Nothing general can be said about $p_Y(y)$, and its exact expression must be determined from time to time. **BUT**:

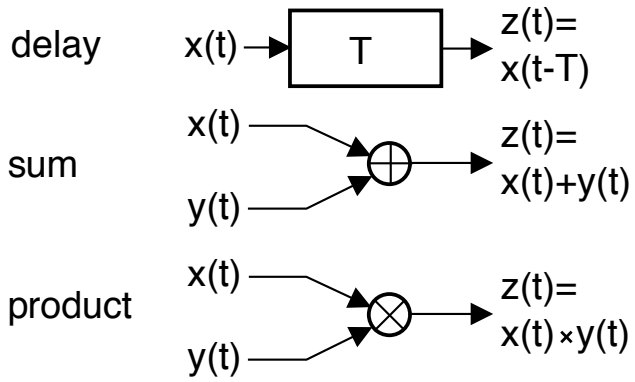
- it the filter input is a Gaussian processes, also the output will be Gaussian
  - This is a consequence of the invariance feature of the Gaussian processes with respect to linear transformations

Now we talk about...

1. **Probability and statistics**
   - Building a ground of common terms
   - Stationary and ergodic processes
   - Correlation and covariance for signals
2. **A common point of view**
   - Autocorrelation and intercorrelation
   - Geometry and adaptation
3. **Power density spectrum**
   - Wiener's theorem
   - Multidimensional Gaussian and process
   - Spectral estimation
4. **Filtering and combining of signals and processes**
   - Spectral density at the output of a filter
   - Statistical characteristics at the output of a filter
   - Sum and product of random and deterministic signals

Combination of different kind of signals

We refer again to the three basic operations already exploited as building blocks for digital filter architecture design



Delay - as well known, there is only a tilt in the phase spectrum, or $Z(f) = X(f) \cdot e^{-j2\pi fT}$

Sum and product of a deterministic and a random signal

- $x(t)$ deterministic, $y(t)$ random process
  - $z(t)$ is (*in general*) a *non-stationary* process, since the ensemble averages of $z(t)$ depend at each instant on the value of $x(t)$
- $x(t)$ periodic, $y(t)$ a random process
  - $z(t)$ is a *cyclostationary* process, as its statistics vary *cyclically* over time, assuming *identical* values with the same period of $x(t)$
- $x(t)$ is a constant equal to a, and $y(t)$ is a random process
  - $z(t)$ is a process with mean $m_z = m_y + a$ (or $m_z = m_y \cdot a$ for the product), with power $\mathcal{P}_z = \mathcal{P}_y + a^2$ (or $\cdot a^2$), and autocorrelation $\mathcal{R}_z(\tau) = \mathcal{R}_y(\tau) +$ (or $\cdot a^2$)

Sum of statistically independent random signals

Mean value - also valid in the absence of statistical independence

$$m_z = E\{x(t) + y(t)\} = E\{x(t)\} + E\{y(t)\} = m_x + m_y$$

Total power

$$\begin{aligned} \mathcal{P}_z &= E\left\{(x(t) + y(t))^2\right\} = E\left\{x^2(t)\right\} + E\left\{y^2(t)\right\} + 2E\{x(t) \cdot y(t)\} = \\ &= \mathcal{P}_x + \mathcal{P}_y + 2m_x m_y \end{aligned}$$

Variance

$$\begin{aligned} \sigma_z^2 &= E\left\{(z(t) - m_z)^2\right\} = \mathcal{P}_z - (m_z)^2 = \mathcal{P}_x + \mathcal{P}_y + 2m_x m_y - (m_x + m_y)^2 = \\ &= \mathcal{P}_x - (m_x)^2 + \mathcal{P}_y - (m_y)^2 = \sigma_x^2 + \sigma_y^2 \end{aligned}$$

Autocorrelation

$$\begin{aligned} \mathcal{R}_z(\tau) &= E\{z(t)\, z(t+\tau)\} = E\{(x(t) + y(t))(x(t+\tau) + y(t+\tau))\} = \\ &= E\{x(t)\, x(t+\tau)\} + E\{y(t)\, y(t+\tau)\} + E\{x(t)\, y(t+\tau)\} + E\{x(t+\tau) \\ &= \mathcal{R}_x(\tau) + \mathcal{R}_y(\tau) + 2m_x m_y \end{aligned}$$

Power density spectrum

$$\mathcal{P}_z(f) = \mathcal{F}\{\mathcal{R}_z(\tau)\} = \mathcal{P}_x(f) + \mathcal{P}_y(f) + 2m_x m_y \delta(f)$$

Probability density function

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(\theta)\, p_Y(z - \theta)\, d\theta = p_X(x) * p_Y(y)$$

BUT: the sum of two Gaussian processes is still Gaussian: in fact the convolution between Gaussian functions is a Gaussian, with mean equal to the sum of the means, and variance equal to the sum of variances

Product of statistically independent random signals

Mean value

$$m_z = E\{z(t)\} = E\{x(t)\,y(t)\} = E\{x(t)\}\,E\{y(t)\} = m_x \cdot m_y$$

Total power

$$\mathcal{P}_z = E\left\{z^2(t)\right\} = E\left\{x^2(t)\,y^2(t)\right\} = E\left\{x^2(t)\right\}E\left\{y^2(t)\right\} = \mathcal{P}_x \cdot \mathcal{P}_y$$

Variance

$$\sigma_z^2 = E\left\{(z(t) - m_z)^2\right\} = \mathcal{P}_z - (m_z)^2 = \mathcal{P}_x \cdot \mathcal{P}_y - (m_x \cdot m_y)^2$$

Autocorrelation function

$$\begin{aligned}\mathcal{R}_z(\tau) &= E\{z(t)\,z(t+\tau)\} = E\{x(t)\,y(t)\,x(t+\tau)\,y(t+\tau)\} = \\ &= E\{x(t)\,x(t+\tau)\}\,E\{y(t)\,y(t+\tau)\} = \mathcal{R}_x(\tau) \cdot \mathcal{R}_y(\tau)\end{aligned}$$

Note that the uncorrelatedness of one of the two processes for a certain value of τ causes the uncorrelatedness of the product, at the same time τ

Power density spectrum

$$\mathcal{P}_z(f) = \mathcal{F}\{\mathcal{R}_z(\tau)\} = \mathcal{F}\{\mathcal{R}_x(\tau) \cdot \mathcal{R}_y(\tau)\} = \mathcal{P}_x(f) * \mathcal{P}_y(f)$$

Probability density function

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(\theta)\,p_Y\left(\tfrac{z}{\theta}\right)\frac{d\theta}{|\theta|}$$