Chapter 1

Random signals, correlation, Wiener's theorem and signal statistics

The degree course to which this edition is dedicated already includes classes in probability and statistics, therefore the basic notions in this regard are given as known and not included from here on. However, some concepts may be repeated, which is never bad. The random variables are then directly defined, characterized by distributions and moments, and the uniform and Gaussian random variables are discussed. We then move on to describe the random processes, the different ways of defining their averages, and the properties of stationarity and ergodicity. The multidimensional Gaussian density is then introduced with its properties. The particularization of these concepts to random signals is performed starting from § 1.3, which is actually a shortened version of an entire chapter of the original text.

1.1 Random variables

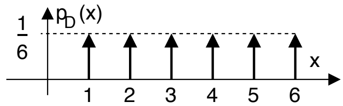
While probability theory talks about *events* in an abstract way, we often find ourselves associating a numerical value to each point of the sample space Ω, which then becomes the *set of numbers* and takes the name of *random variable*, henceforth abbreviated as *r.v.* The occurrence of an event now corresponds to the assignment of a value (among the possible ones) to the r.v.; this "chosen" value therefore takes the name of *realization* of the r.v. We then distinguish between *discrete* and *continuous* random variables, depending on whether the quantity they describe has countable or continuous values[1]. The characterization of the random variable in probabilistic terms is obtained by indicating how the "probability mass" *is distributed* over the set of values that it can assume, by means of the following two functions (of r.v.).

[1] A classic example of discrete r.v. is that of throwing a dice, another is the lotto numbers. A continuos r.v. can be for example an atmospheric pressure value in a location, or the attenuation of a radio transmission due to atmospheric phenomena.

1.1.1 Probability density and distribution functions

Just as the mass of an *inhomogeneous* object is more or less densely distributed in different regions of its overall volume, so the *probability density function* (or *p.d.f.*) indicates on which values of the random variable the probability is concentrated. For example, the density of the discrete r.v. associated with the launch of a dice can be written as

$$p_D(x) = \sum_{n=1}^{6} \frac{1}{6}\delta(x-n) \tag{1.1}$$



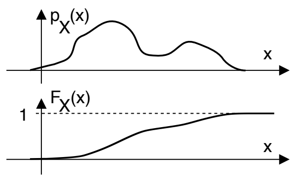
the meaning of which we discuss immediately, with the help of the graph on the side, where D indicates the r.v. (the number that will come out), and x is one of its possible realizations (one of the 6 faces). The 6 pulses centered in $x = n$ represent a *concentration* of probability in the six possible values, and the area of such pulses is exactly equal to the probability of each of the six results. It is easy to verify that

$$\int_{-\infty}^{\infty} p_D(x)\,dx = 1 \quad \text{and that results} \quad \int_{a}^{b} p_D(x)\,dx = Pr\{a < D \leq b\}$$

or equal to the probability that the r.v. D takes a value between a and b. In particular, since we cannot have a negative probability, we obtain $p_D(x) \geq 0$ with $\forall x$.

A function of r.v. closely related to density is the *partition* or *distribution* function[2], defined as

$$F_X(x) = \int_{-\infty}^{x} p_X(\xi)\,d\xi = Pr\{X \leq x\} \tag{1.2}$$



and which is a non-decreasing function of x, limited to a maximum value of 1, and whose graph is shown on the side below that of $p_X(x)$, for which it obviously results $p_X(x) = \frac{d}{dx}F_X(x)$; in the case instead of the discrete r.v. D, its distribution function is discontinuous[3].

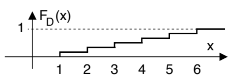
Now it is even more evident that $p_X(x)$ is a *density*, and becomes a probability only when multiplied for an interval of x ([4]).

Histogram If you do not have an analytical expression suitable to represent the way in which the values of a r.v. are distributed, it may be useful to *estimate* it using a *histogram*. It takes on the appearance of a so to speak *quantized* version of the unknown p.d.f., and is obtained starting from a series of realizations[5] of the r.v., dividing the variability field

---

[2]In reality, the historical order is to first define $F_X(x)$ as the probability that X is not greater than a x value, or $F_X(x) = Pr\{X \leq x\}$, and therefore $p_X(x) = \frac{dF_X(x)}{dx}$. The reason of this "priority" lies in the fact that $F_X(x)$ presents fewer "analytical difficulties" of definition (for example, it presents only discontinuities of the first kind, even with discrete r.v.).

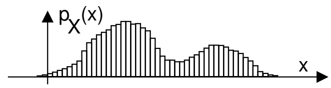
[3]The $F_D(x)$ for the roll of a dice is shown alongside: in fact, remembering that the derivative of a step is an impulse of area equal to the difference in height, we can verify that by applying (1.2) to (1.1) we obtain the staircase.



[4]In fact the probability that X falls between x_0 and $x_0 + \Delta x$ holds $\int_{x_0}^{x_0+\Delta x} p_X(x)\,dx \simeq p_X(x_0)\,\Delta x$.

[5]Obtained, for example, from personal data, health, meteorological or whatever else, or by carrying out a specific *measurement campaign* based on a *statistical sample* of adequate size (see also § ??).

of the quantity X in sub-intervals, and drawing vertical rectangles, each of a height equal to the number of times that (within the statistical sample available) X takes on a value in that interval, as shown in the figure.



By dividing the height of each rectangle by the number of observations N, an approximation of $p_X(x)$ is obtained, gradually more precise with $N \to \infty$, and with a simultaneous reduction in the extension of the intervals.

1.1.2 Expected value, moment and central moment

These are *summaries*, so to speak, of the way in which the values of a r.v. are distributed, and are defined starting from a *generic function of random variable*[6] that we indicate by $g(x)$.

Expected value We define the *expected value* (or *ensemble average*[7]) of $g(x)$ with respect to the random variable X the quantity

$$E_X\{g(x)\} = \int_{-\infty}^{\infty} g(x)\, p_X(x)\, dx \tag{1.3}$$

which corresponds to a *weighted average*, in which the values assumed by $g(x)$ in correspondence to a certain x are *weighed* proportionally to the corresponding probability value $p_X(x)\,dx$; this integral average operation is denoted by the notation $E_X\{.\}$[8], by means of which the r.v. (X) with respect to which to carry out the weighting is indicated as a subscript.

In the case of a function $g(x, y)$ of more than one r.v. the relative expected value is calculated on the basis of the *joint* p.d.f., that is

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

where $p_{XY}(x,y)$ is obtained starting from the conditional and marginal p.d.f., as explained in § **??**. Finally, eq. (1.3) can be calculated using a conditional p.d.f. $p_{X/Y}(x/y)$, and in this case also the expected value $E_{X/Y}\{g(x)\}$ is called *conditional*, resulting in a function of y.

Moment If we set $g(x) = x^n$, that is equal to the *n-th* power of the r.v., the expected value is called *moment of order n*, and is indicated as

$$m_X^{(n)} = E\{x^n\} = \int_{-\infty}^{\infty} x^n p_X(x)\, dx \tag{1.4}$$

[6] An example of a function of r.v. could be the value of the payout associated with the 13 on the football pools coupon, which depends on the r.v. represented by the results of the matches, once the prize pool and the bets are known. In fact, for each possible vector of results, a different number of winning bets is determined, and therefore a different way of dividing the prize pool. Since the unlikely results was played by a reduced number of coupons, these are entitled to a higher value in the event of a win, well above its expected value, indicative instead of *average* win.

[7] By *ensemble* we refer to the *sample space* Ω, consisting of the possible values assumed by the r.v. X.

[8] In fact, the E symbolizes the word *Expectation*.

In the case of discrete random variables, the moments are defined as $m_X^{(n)} = \sum_i x_i^n p_i$, where $p_i = Pr\{x = x_i\}$, thus weighting the possible realizations x_i with the respective probabilities. We immediately note that $m_X^{(0)} = \int_{-\infty}^{\infty} p_X(x)\,dx = 1$. Let us now consider two important moments.

Mean value and quadratic mean The *first-order* moment

$$m_X = m_X^{(1)} = \int_{-\infty}^{\infty} x p_X(x)\,dx \tag{1.5}$$

takes the name of *mean value* of the r.v., sometimes called *centroid*, and coincides with the *arithmetic mean* obtainable starting from the knowledge of the realizations of the r.v. obtained by repeating the aleatory experiment indefinitely. Vice versa the second order moment

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

is referred to as the *square mean.*

Example Suppose that the r.v. X represents the height of individuals: the *average* height m_X can be *estimated* as the arithmetic mean of the relative measurements

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

As $N = \sum_{i=1}^{n} N_i$ tends to ∞, the estimate $\hat{m}_X$ coincides with the result m_X provided by (1.5) if, instead of the probabilities $p_X(x)\,dx$, the values $Pr(x_i)$ obtained through the histogram $Pr(x_i) = \frac{N(x_i < x \le x_i + \Delta x)}{N} = \frac{N_i}{N}$ are replaced, 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 *weighting* the possible values of x with the respective frequencies.

Central moment In case $g(x) = (x - m_X)^n$ the relative expected value is called the *central moment* of order n, and indicated such as

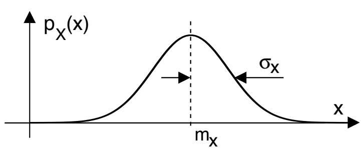
$$\mu_X^{(n)} = E\{(x - m_X)^n\} = \int_{-\infty}^{\infty} (x - m_X)^n p_X(x)\,dx$$

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

Variance 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 square root σ_X of the variance σ_X^2 is called standard deviation, and while the mean m_X indicates where the "statistical center" of the probability density lies, σ_X indicates how much the individual determinations of the r.v. are dispersed around m_x.

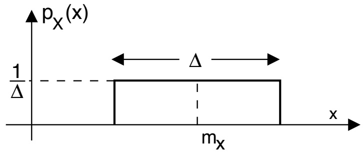


A remarkable relationship that links the first two moments (central and not) is ([9]):

$$\sigma_X^2 = m_X^{(2)} - (m_X)^2 \tag{1.6}$$

1.1.3 Uniform random variable

It is characterized by presenting the same probability value for the whole range of possible realizations, comprised between a minimum and a maximum value, as shown in the figure; therefore the probability density function can be expressed by means of a rectangular function



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

where Δ represents the extension of the existence interval for the random variable, while the parameter m_X, which indicates the abscissa at which the rectangle is centred corresponds exactly to the first order moment of X. The calculation of variance, on the other hand, provides[10] $\sigma_X^2 = \frac{\Delta^2}{12}$.

1.1.4 Gaussian random variable

Unlike the uniform case, the Gaussian r.v. has more probable values close to the mean value m_x, in accordance with p.d.f. with expression

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \mathrm{e}^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \tag{1.7}$$

and whose characteristic *bell-shaped* graph is shown for different values of the parameters m_x and σ_x that appear in (1.7), respectively equal to the mean and standard deviation of the r.v. (see § ??), and which fully describe the p.d.f. from an analytical point of view: therefore the estimate of m_x and σ_x (starting from a good number of realizations[11]) is sufficient to completely describe the random

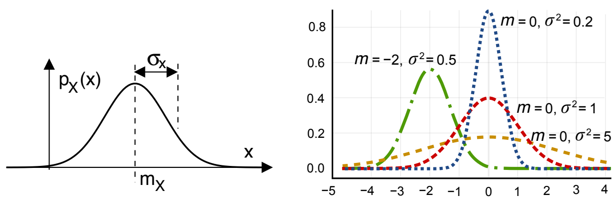


Figure 1.1: Graph of a Gaussian p.d.f.

[9] In fact it turns out that

$$\begin{aligned} \sigma_X^2 &= E\left\{(x - m_X)^2\right\} = E\left\{x^2 + (m_X)^2 - 2x m_X\right\} = E\left\{x^2\right\} + (m_X)^2 - 2m_X E\{x\} = \\ &= m_X^{(2)} + (m_X)^2 - 2(m_X)^2 = m_X^{(2)} - (m_X)^2 \end{aligned}$$

We preferred to use the notation $E\{x\}$, which is more compact than the indication of the integrals involved; the steps carried out are justified by recalling the distributive property of integrals, and by observing that the expected value of a constant is the constant itself.

[10] Instead of calculating σ_X^2 for the given $p_X(x)$, we calculate $m_X^{(2)}$ for a uniform r.v. with zero mean, i.e. with $m_X = 0$, exploiting the fact that according to (1.6) in this case it results $m_X^{(2)} = \sigma_X^2$. We get:

$$m_X^{(2)} = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} x^2 \frac{1}{\Delta} dx = \left.\frac{x^3}{3\Delta}\right|_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} = \frac{1}{3\Delta}\left(\frac{\Delta^3}{8} + \frac{\Delta^3}{8}\right) = \frac{1}{3\Delta} 2\frac{\Delta^3}{8} = \frac{\Delta^2}{12}$$

[11] Having a set $\{x_n\}$ of N realizations of a random variable X, we can make the estimates $\widehat{m}_x =$

phenomenon. The Gaussian r.v. emerges in many natural phenomena, and it is analytically demonstrable that its density is typical[12] for quantities obtained by adding a very large number of random causes, all statistically independent and with the same p.d.f.[13] (*central limit theorem*[14]).

1.1.5 Multivariate random variable

In this case the r.v. jointly represents an entire *vector* $\boldsymbol{x}$ of one-dimensional random variables, i.e. an ordered collection of them, in finite number (e.g. N), in relation or not between them on the basis of probabilistic connections.

Probability density function Indicating with $\boldsymbol{X}$ the vector r.v., and with $\boldsymbol{x}$ a realization thereof consisting of the N components $x_1, x_2, \cdots, x_N$, the multivariate r.v. is described by means of the p.d.f. $p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N)$ function of N variables, for which it must be

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N)\, dx_1 dx_2 \cdots dx_N = 1$$

Distribution Also in the multivariate case an $F_{\boldsymbol{X}}(\boldsymbol{x})$ distribution function can be defined, also it $N-$ dimensional, whose value $F_{\boldsymbol{X}}(\bar{\boldsymbol{x}}) = Pr\{\boldsymbol{x} \le \bar{\boldsymbol{x}}\}$ at the point $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_N)$ is calculated as

$$F_{\boldsymbol{X}}(\bar{\boldsymbol{x}}) = \int_{-\infty}^{\bar{x}_1}\int_{-\infty}^{\bar{x}_2}\cdots\int_{-\infty}^{\bar{x}_N} p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N)\, dx_1 dx_2 \cdots dx_N$$

Marginal probability density function The *marginal* p.d.f. $p_{X_i}(x_i)$ of the single *one-dimensional* r.v. x_i which takes part in the coordinate system on which $\boldsymbol{X}$ is defined, can be calculated starting from the *joint* p.d.f. $p_{\boldsymbol{X}}(\boldsymbol{x})$ by *saturation* of the others r.v., or

$$p_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{N-1\ integrali} p_{\boldsymbol{X}}(x_1, x_2, \cdots, x_N)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N \tag{1.8}$$

Conditional probability density function The p.d.f. of a sub-group of r.v. $\boldsymbol{x}_a = (x_1, x_2, \cdots, x_a)$, if the value of the remaining coordinates $\boldsymbol{x}_b = (x_{a+1}, x_{a+2}, \cdots, x_N)$ of $\boldsymbol{x}$ is to be considered known, is obtained by dividing the joint p.d.f. $p_{\boldsymbol{X}}(\boldsymbol{x})$ for the marginal $p_{\boldsymbol{X}}(\boldsymbol{x}_b)$ which describes the conditioning events, that is

$$p_{\boldsymbol{X}}(\boldsymbol{x}_a/\boldsymbol{x}_b) = \frac{p_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{X}}(\boldsymbol{x}_b)}$$

$\frac{1}{N}\sum_{n=1}^{N} x_n$ e $\widehat{m}_x^{(2)} = \frac{1}{N}\sum_{n=1}^{N} x_n^2$, whose value tends asymptotically to that of the respective ensamble means, as N (the size of the statistical sample) tends a ∞. In this regard, see § ??.

[12] So much so that (1.7) is also called Normal, and for this reason it is also indicated as $N(m, \sigma^2)$.

[13] This condition is also called of independent and identically distributed (or *i.i.d.*) random variables.

[14] The theorem is proved at § ??, but it can be fun and useful to test its validity by recurring to the applet present at `http://www.randomservices.org/random/apps/DiceExperiment.html`. Furthermore, considering that at § ?? it is shown how the p.d.f. of a sum of independent r.v. is equal to the convolution between the respective p.d.f., we observe that the repeated convolution of a same p.d.f. with itself, Gaussianizes it.

wherein $p_{\boldsymbol{X}}(\boldsymbol{x}_b)$ is obtained by saturation (1.8). The *ordinal* separation between the two groups of variables is intended to simplify the notation of this definition; in reality, the r.v. of the two groups can be taken in any order.

Expected value and moments In case in which we are dealing with the expected value of a function of a single marginal r.v., we still use eq. (1.3) in which the p.d.f. is the marginal $p_{X_i}(x_i)$ related to the r.v. with respect to which the ensemble mean is being performed. By this way it is possible to obtain a vector $m_{\boldsymbol{X}} = (m_{x_1}, m_{x_{2,}}, \cdots, m_{x_N})$ which represents the mean value of the multivariate r.v. $\boldsymbol{X}$, whose components m_{x_i} are the first order moments of the marginal r.v., that is

$$m_{x_i} = E_{\boldsymbol{X}}\{x_i\} = \int x_i p_{X_i}(x_i)\, dx_i$$

We note that although the notation $E_{\boldsymbol{X}}\{x_i\}$ indicates an expected value calculated with respect to the variability of all the components of the multivariate r.v. $\boldsymbol{X}$, the calculation is carried out using the marginal p.d.f., since the other r.v. x_j with $j \neq i$ *saturate* (see eq. (1.8)).

On the other hand, it is now possible to evaluate also the so-called mixed moments, in which the ensemble average considers all the possible values of two or more components of $\boldsymbol{X}$, weighting each of these with the relative probability value. For example, a mixed moment of order (n, m) is defined as

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

and a central mixed moment of order (n, m) as

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

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

It is now the turn of extending probabilistic concepts to the signal space.

1.2 Stationary and ergodic processes

After having described how to statistically characterize the values of single or vector r.v., let's deal with the case in which we want to describe from a probabilistic point of view a entire signal, whose real identity is not known a priori[15].

Such a signal is called *a member* (or realization) of a *random process*, and can be indicated as $x(t, \theta)$, through a formal description that includes a couple of sets: the first is the $\mathcal{T}$ set of time instants (typically within an interval) on which the process members are defined, while the second set represents a Θ random variable, whose θ values each identify a particular realization of the process. Therefore, a specific θ_i realization of the Θ r.v. *indexes* the process, whose members $x(t, \theta_i)$, with $t \in \mathcal{T}$, are known only

[15]Clearly, most of the signals transmitted by telecommunication equipments are of this type.

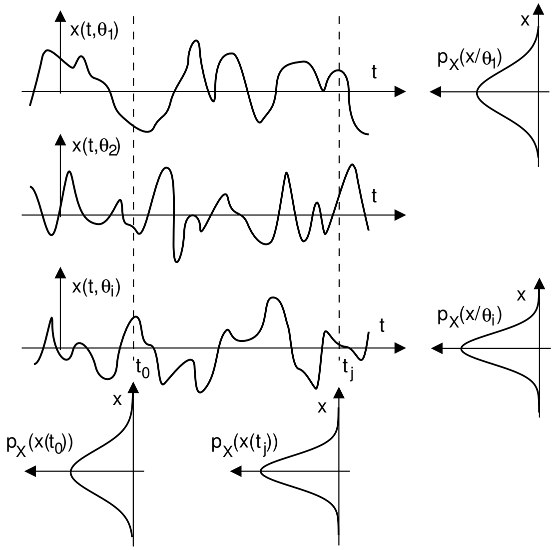


Figure 1.2: A non-ergodic process

after the knowledge of $\theta_i \in \Theta$ ([16]). The random process is therefore defined as the set of signals $\{x(t,\theta)\}$, with $t \in \mathcal{T}$ and $\theta \in \Theta$.

If, on the other hand, we fix a particular 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, which we can draw at the instant t_j in which the sample is taken[17]; in this regard, refer to figure 1.2, which shows the defined probability densities as referred to members of a process.

1.2.1 Moment as an ensemble average

It consists of the *expected value* of an n^{th} *power* of the signal values, performed with respect to the variability due to Θ, and is therefore calculated 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, shown below on fig. 1.2. We note that according to this approach, the ensemble mean depends on the instant t_j in which a value is taken[18].

1.2.2 Time average

Alternatively, we can fix a particular θ_i realization of Θ, and then focus the attention on a single member $x(t,\theta_i)$, which is now a *deterministic* signal[19]: time averages can

[16] To fix the ideas, we lead parallel to the text a "real" an example in which the random process consists of... the musical selection carried out by a DJ. The set $\mathcal{T}$ will then be consistuted by the opening hours of the discos (from 10pm to dawn?), while in θ we will include all types of variability (mood of the DJ, the records he has in his box, the disco where we are, the day of the week ...).

[17] In the example, $x(t_0,\theta)$ is the pressure value sound detectable at a given instant (eg 11.30pm) when it changes of θ (any DJ, disco, day ...).

[18] For example, if in all evenings the volume increases progressively over time, the $p_X(x(t_j))$ will widen for increasing t_j.

[19] $x(t,\theta_i)$ represents, in our example, the whole musical selection (called *night at the disco*) proposed by a specific DJ, in a specific place, on a specific day.

then be calculated for it, which are 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* (page **??**)

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

and the *power*[20] (Eq. (**??**)) (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.

1.2.3 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 draw alongside the single members shown in fig. 1.2. 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 every possible value of x is weighted by its probability $p_X (x/\theta_i)\, dx$ (see the example on page 4).

1.2.4 Stationary process

If $p_X \left(x \left(t_j\right)\right)$ does not depend on t_j, but results $p_X \left(x \left(t_j\right)\right) = p_X^{\mathcal{T}} (x)$ for any $t_j \in \mathcal{T}$, the process $\{x (t, \theta)\}$ is called stationary[21] in the *strict sense*. In this case all the ensemble averages no longer depend on time, that is $m_X^{(n)} (t) = m_X^{(n)}$ for $\forall t \in \mathcal{T}$, and the $p_X \left(x \left(t_j\right)\right)$ at the bottom in fig. 1.2 are all equal.

If, on the other hand, it is only the first two ensemble means $m_X (t)$ and $m_X^{(2)} (t)$ that do not depend on t, the process $\{x (t, \theta)\}$ is called stationary *on mean* and *on quadratic mean*, or even stationary in a *wide sense*[22]. In the case of a Gaussian process (§ 1.6.2),

[20] $m_X^{(2)} (\theta_i)$ in this case represents the average power with which the music is played on the particular evening θ_i.

[21] The stationary "night at the disco" therefore occurs if the genre of music, the volume of the amplification, etc do not change over time... or rather if any variations in some particular disco-realizations are compensated by opposite variations in as many different members of the process.

[22] In this case $p_X (x (t))$ is not known, or it is not stationary, but the major applications of the stationarity property depend only on $m_X (t)$ and $m_X^{(2)} (t)$, which can be measured (or rather *estimated*, see § **??**), and be stationary even if $p_X (x (t))$ is not.

stationarity in the wide sense implies that in the strict sense[23].

Now suppose to divide the member $x(t, \theta_i)$ in several time intervals, and to calculate the temporal averages for each of them, limited to the relative interval. If these are equal to each other, and consequently equal to time average $m_X^{(n)}(\theta_i)$, the member is (individually) stationary[24]. Obviously, if all members are individually stationary, so too is the process to which they belong.

1.2.5 Stationary and ergodic process

This important subclass of stationary processes identifies the circumstance that *each member of the process is statistically representative of all the others.* This occurs when the probability density function (on the right in fig. 1.2) of the values extracted from a single member $p_X(x/\theta_i)$ is always the same, regardless of the particular θ_i, ultimately obtaining $p_X(x/\theta_i) = p_X^{\Theta}(x)$ regardless of the realization and, for the stationarity, also $p_X(x/t_j) = p_X^{\mathcal{T}}(x)$, and therefore $p_X^{\Theta}(x) = p_X^{\mathcal{T}}(x) = p_X(x)$. In this case the time averages $m_X^{(n)}(\theta_i)$, which can be calculated as moments on the single realization as illustrated in § 1.2.3, are identical for all θ_i members[25], and also identical to the ensemble averages $m_X^{(n)}(t_j)$ calculated for any instant. We therefore state the definition:

> *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.*

Example: the power of a signal We show how the computation of the power of a member of an ergodic process is equivalent to that of the moment of 2^{nd} order of the process:

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

This result shows how it is possible to calculate the power of a realization of a process, without knowing the waveform of its members.

Example: the mean value On page ?? it was defined as $\bar{x} = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x(t)\,dt$, that is, as a *time* average of the first order. If $x(t)$ is a member of an ergodic process, this value can also be calculated as the *expected value* of $x(t)$, that is the *first-order* moment m_x of

[23]In fact, the Gaussian p.d.f. is completely defined if the mean and (co)variance values are know, see §§ 1.1.4 and 1.6.

[24]This happens if the musical selection of a particular nigth remains constant (e.g. raggamuffin only) or varied but homogeneously (e.g. without three "slows" in a row).

[25]Therefore wanting to arrive at the definition of an ergodic evening in the disco, we should eliminate those cases that, albeit individually stationary, they are definitely " out of standard " (all metal, only smooth ...).

the r.v. x extracted from the process:

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

Power, variance, quadratic mean and effective value In particular we observe that on the basis of (1.6) we can write

$$\mathcal{P}_X = m_X^{(2)} = \sigma_x^2 + (m_x)^2 \tag{1.9}$$

and for signals with zero mean ($m_x = 0$) we obtain $\mathcal{P}_X = \sigma_x^2$; in this case the effective value (page ??) $\sqrt{\mathcal{P}_X}$ 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 $x_{RMS} = \sqrt{\overline{x^2(t)}}$, 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.

1.2.6 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 $\overline{x(t)} = E\{x\} = m_x$; $\overline{x^2(t)} = E\left\{x^2\right\} = m_x^{(2)}$.

1.3 Correlation, spectral density and filtering

HERE begins the announced shortened version of a different chapter, which intertwines the theory of signals (chapters ?? and ??) with that of probabilities (chap. 1) to describe in a unitary way the passage of both deterministic signals and random processes through a physical system, from the point of view of spectral modifications as well as from that of statistical properties. While for deterministic signals we are already able (see § ??, ?? and ??) to determine the amplitude spectrum and

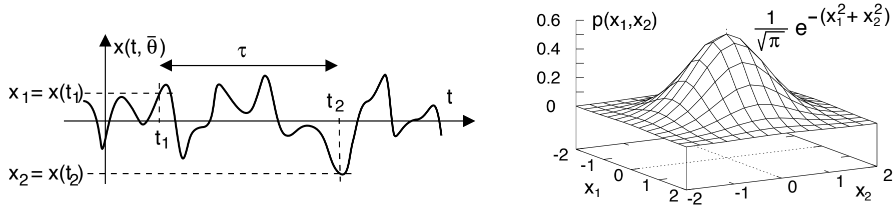


Figure 1.3: Extraction of two r.v. from a random process, and a possible joint p.d.f.

energy density at the output of a filter, if the input signal is a member of a process a new part of theory must be carried out, aimed at defining the *autocorrelation function* $\mathcal{R}_x(\tau)$ as a tool which unifies the world of deterministic signals with the random ones, and whose transform (thanks to the WIENER theorem) provides the power density spectrum sought. The remainder of the chapter proceeds by applying the results obtained to practical cases, determining the power density spectrum for some signals of interest, as well as the relative estimate based on experimental observations. Al § 1.8 we arrive at a unitary setting for the filtering of energy, periodic, and random signals, while at § 1.9 the same scheme is applied to their sum and product.

1.4 Correlation, covariance and autocorrelation

At § 1.2.5 we discussed how for a stationary ed ergodic process $\{x(t,\theta)\}$ the knowledge of the p.d.f. $p_X(x)$ which describes the variability of its values independently of t and θ allows the calculation of the corresponding mean m_X and variance σ_X^2 expected values, as well as the power $\mathcal{P}_X = E_X\left\{x^2\right\} = \sigma_X^2 + (m_x)^2$ of each of its members. These aggregate means are first-order statistical descriptions, as they are related to the p.d.f. of a single extracted value.

In this section we define instead a statistical description of the *second order*, that is a mixed moment (page 7), which as we will see at § 1.5.1 enable us to obtain the *power density spectrum* of the process members. This description is based on the consideration of two instants t_1 and $t_2 = t_1 + \tau$, in correspondence with which the random variables $x_1 = x(t_1)$, $x_2 = x(t_2)$ are extracted starting from any θ realization of the process $x(t,\theta)$, of which the case for a specific member $x(t,\overline{\theta})$ is shown on the left side of fig. 1.3. As the $\theta \in \Theta$ realization varies, all the couples of sampled values have as many determinations of a *two-dimensional* random variable, described by a *joint* probability density $p_{X_1X_2}(x_1x_2;t_1t_2)$, which also depends on the instants t_1 and t_2, and which is exemplified in the right part of fig. 1.3; such *two-dimensional* p.d.f. subtends a unit volume or $\int\int p(x_1,x_2)\,dx_1dx_2 = 1$, and his 3D graph describes the regions of the x_1x_2 plane in which each couple of possible values is more or less probable.

1.4.1 Correlation between random variables

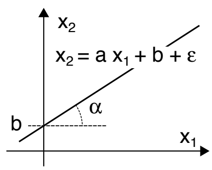
Now that we have the joint p.d.f. $p_{X_1X_2}(x_1x_2;t_1t_2)$ of two r.v. x_1 and x_2 extracted from the process $x(t,\theta)$ after time interval τ, we can calculate their *mixed moment*, that is, an expected value (§ 1.1.2) in which, unlike of the one-dimensional case, the possible values are weighted by the probability that they occur *together*. In particular, the mixed moment *of order (1,1)* (see page 7) $m_{XX}^{(1,1)}(t_1, t_2)$ between the r.v. is called *correlation*, and is defined as

$$m_{XX}^{(1,1)}(t_1, t_2) = E_{X_1X_2}\{x_1x_2\} = \iint x_1x_2 \cdot p_{X_1X_2}(x_1x_2; t_1t_2)\, dx_1dx_2 \qquad (1.10)$$

Before continuing, let's try to deepen the meaning of this new statistical description in its broader context of two r.v. di any type, not necessarily extracted from the same random process, but which describe two somewhat interdependent events[26].

Sign First of all, we observe that the sign of the correlation between two r.v. x_1 and x_2 reflects their *agreement*, in the sense that if $m_{X_1X_2}^{(1,1)} > 0$ the two r.v. frequently have the same sign[27], or opposite if the correlation is negative.

Regression Identify a similar concept[28], but oriented to the problem of *predicting* the expected value of a quantity (eg. x_2) starting from the knowledge of another (in this case x_1): in fact we can think that the quantities depends one another with a relationship of the type $x_2 = f(x_1) + \varepsilon$ where ε represents the random component, with zero mean and statistically independent of both x_1 and x_2. So far as $f(x_1) = a \cdot x_1 + b$ we speak of *linear regression* since $f(x_1)$ is the equation of a line in which $a = \tan\alpha$ is the angular coefficient and b the intercept, and at § ?? we show how it turns out $a = \frac{\sigma_{x_1x_2}}{\sigma_{x_1}}$ and therefore (eq. (1.12)) is related to $m_{X_1X_2}^{(1,1)}$, while $b = m_{x_2} - am_{x_1}$.



Scattering diagram The last reflection before moving on to the analytic side concerns the *scattering* diagrams[29] shown in fig. 1.4, which map the *position* of a large number[30] of pairs of x_1 and x_2 values according to six possible stochastic dependence laws. Together with the clouds, the diagrams also report the *estimated*[31] correlation values $m_{X_1X_2}^{(1,1)}$ (corr), the covariance $\sigma_{x_1x_2}$ (cov) (eq. (1.12)), and the correlation coefficient ρ (§ 1.10.1).

In cases A) and F) the pairs of values are related by a very little random law, but in the second case the correlation is zero since the dependence is *not linear*. In cases B) and D) there is more variability, but a certain dependence between the two r.v. is still noted. In cases C) and E) we are instead in the presence of two *statistically independent* r.v., given that $p_{X_1X_2}(x_1, x_2)$ can be factored as $p_{X_1}(x_1)\, p_{X_2}(x_2)$, and for which it will soon be shown that $m_{X_1X_2}^{(1,1)} = m_{x_1}m_{x_2}$, as in fact we find for the case E) in which the

[26] The term *correlation* dates back to studies on genetic inheritance, and has gradually been adopted by all disciplines (economic, clinical, sociological ...) which analyze from a statistical point of wiew the dependence (*co-relation*) between two or more quantities, see e.g. https://en.wikipedia.org/wiki/Correlation.

[27] As intuitively verifiable by thinking $m_{X_1X_2}^{(1,1)}$ as a probability weighted average of the possible values of the product x_1x_2; terms of equal amplitude and opposite sign can cancel each other out if equiprobable.

[28] The term refers to the concept of *regressing*, that is, from a genetic point of view, seeing remote traits resurface. For further information see https://en.wikipedia.org/wiki/Linear_regression

[29] See e.g. https://en.wikipedia.org/wiki/Scatter_plot

[30] Graphs A, D and F are made with 100 points, while B, C and E with 700.

[31] That is, obtained from the statistical sample, for which for example $\hat{m}_{X_1X_2}^{(1,1)} = \frac{1}{N}\sum_{i=1}^{N} x_1(i)\, x_2(i)$.

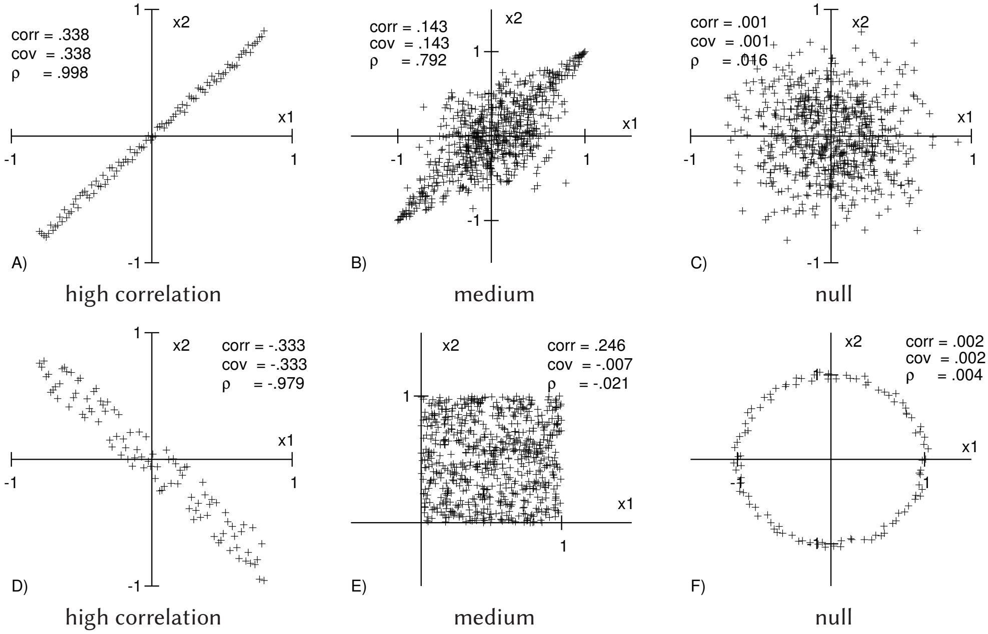


Figure 1.4: Scattering diagram for pairs of random variables

r.v. are independent and the correlation is 0.25, that is equal to the product of the means $m_{x_1} = m_{x_2} = 0.5$. To handle this case, we need to introduce the *covariance*, which we are about to discuss.

1.4.2 Covariance, statistical independence and incorrelation

In the event that the two r.v. are *statistically independent*, i.e. if we can write $p_{X_1X_2}(x_1, x_2; t_1, t_2) = p(x_1)\, p(x_2)$ ([32]), then the integral that defines the correlation *is factored*, providing as a result the product of the mean values of the r.v.:

$$\begin{aligned} m_{XX}^{(1,1)}(t_1, t_2) &= E\{x_1, x_2\} = \iint x_1 x_2 p(x_1)\, p(x_2)\, dx_1 dx_2 = \\ &= \int x_1 p(x_1)\, dx_1 \cdot \int x_2 p(x_2)\, dx_2 = E\{x_1\}\, E\{x_2\} = m_{X_1} m_{X_2} \end{aligned} \quad (1.11)$$

Covariance It is indicated as $\sigma(x_1, x_2)$ and consists of the correlation $m_{XX}^{(1,1)}(t_1, t_2)$ to which the term $m_{X_1} m_{X_2}$ is subtracted, obtaining the *central* mixed moment between the two r.v. In fact:[33]

$$\begin{aligned} \sigma(x_1, x_2) &= E\left\{\left(x_1 - m_{X_1}\right)\left(x_2 - m_{X_2}\right)\right\} = \\ &= E\{x_1 x_2\} - E\left\{x_1 m_{X_2}\right\} - E\left\{m_{X_1} x_2\right\} + E\left\{m_{X_1} m_{X_2}\right\} = \\ &= E\{x_1 x_2\} - m_{X_1} m_{X_2} = m_{XX}^{(1,1)}(t_1, t_2) - m_{X_1} m_{X_2} \end{aligned} \quad (1.12)$$

We are now in a position to state an important consequence of statistical independence:

[32]For the sake of brevity, we omit to indicate the identity of the random variable as subscript of the probability density, as well as the time instants.

[33]Another simplification of notation, to be understood by remembering that an expected value is actually an integral that weighs the argument by the respective p.d.f., to which the distributive property applies for its application to a sum.

Incorrelation Combining the results (1.11) and (1.12) 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[34].

This property is valid in *one direction* only, since if a null $\sigma(x_1, x_2)$ covariance occurs for two r.v., they *do not necessarily* have to be statistically independent[35]. The only circumstance in which the uncorrelation between random variables *implies* their statistical independence is that relating to the *Gaussian case*, as shown in § 1.6.1.

1.4.3 Correlation of an ergodic stationary process

If the process from which x_1 and x_2 are extracted is *stationary* at least in a *wide* sense (§ 1.2.4), the relative joint p.d.f. depends only on the difference $\tau = t_2 - t_1$ between the instants t_2 and t_1 (see fig. 1.3), and therefore also the correlation (1.10) depends only on τ:

$$m_{XX}^{(1,1)}(t_1, t_2) = E\{x_1 x_2\} = \iint x_1 x_2 \cdot p_{X_1 X_2}(x_1 x_2; \tau)\, dx_1 dx_2 = m_{XX}^{(1,1)}(\tau) \qquad (1.13)$$

which is thus now referred to as $m_{XX}^{(1,1)}(\tau)$.

If, besides being stationary, the process is also *ergodic* (§ 1.2.5), then the ensemble mean $m_{XX}^{(1,1)}(\tau)$ takes the same value as the corresponding time average. Therefore, if $p_{X_1 X_2}(x_1 x_2; \tau)$ is not known but you instead have some realization of the process, through (1.13) the correlation can be obtained from the *time average* $\overline{x(t, \theta_i)\, x(t + \tau, \theta_i)}$ (see § 1.2.2) calculated for any θ_i realization. This time average is now indicated as $\mathcal{R}_x(\tau)$, and corresponds to

$$\mathcal{R}_x(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t, \theta_i)\, x(t + \tau, \theta_i)\, dt \quad \forall \theta_i \in \Theta \qquad (1.14)$$

Since for stationary and ergodic processes the (1.13) and (1.14) provide the same result, for them also the correlation (1.13) is indicated with the notation $\mathcal{R}_x(\tau)$ instead of $m_{XX}^{(1,1)}(\tau)$. It being understood that if we do not have process realizations, but the $p_{X_1 X_2}(x_1 x_2; \tau)$ is known, the correlation must be obtained from the expression (1.13).

Before using (at § 1.5.1) the new *correlation* $\mathcal{R}_x(\tau)$ statistical descriptor to arrive at an expression of the power density $\mathcal{P}_x(f)$ for ergodic processes, we start from the point of contact between processes and deterministic signals given by (1.13) and (1.14), to deepen the interpretation of $\mathcal{R}_x(\tau)$ in the *deterministic* context.

1.4.4 Autocorrelation and intercorrelation of deterministic signals

When the time average (1.14) is calculated for a *deterministic* signal $x(t)$, that is

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

---

[34] We immediately note that the more correct term would be "incovarianced"; the use (by now historical and consolidated) of the *uncorrelated* expression probably derives from usually considering quantities with zero mean, for which the two expressions coincide.

[35] See for example the case F) of fig. 1.4, in which the random variables are uncorrelated, but they are not independent at all, since the pairs of values are arranged on a circle. This represents a case of *non-linear* dependence, as the equation describing the circle is quadratic.

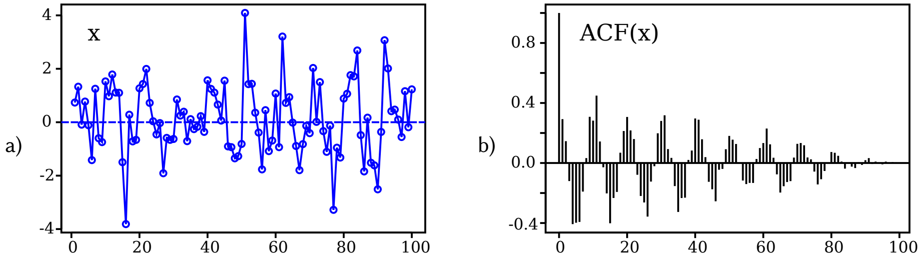


Figure 1.5: a) - sinusoidal sequence immersed in noise; b) - its autocorrelation

we obtain an expression called *autocorrelation function* or *integral*, again indicated by $\mathcal{R}_x(\tau)$ as for (1.14), both valid for *power* signals. In the case of an *energy* signal, on the other hand, the (1.14) would give a null result, so that for energy signals the definition of *autocorrelation* becomes

$$\mathcal{R}_x(\tau) = \int_{-\infty}^{\infty} x^*(t)\, x(t+\tau)\, dt \tag{1.15}$$

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

Comparing (1.15) and (1.14) with (??) on page ??, we note how the autocorrelation evaluates the *mutual* energy (or power) (i.e., a *scalar product*) between an $x(t)$ signal and an *anticipated copy* of it: in this sense, a high value of $\mathcal{R}_x(\tau)$ indicates that for such a value of τ (or advance) the two copies of the signal are *similar*, while its null value is an indication (for that choice of τ) of *orthogonality*.

Example Fig. 1.5-a) shows a numerical sequence x_n with zero mean, obtained from a sinusoid to which noise is superimposed, while on the right the relative autocorrelation is shown, which in the numerical case is evaluated as $\mathcal{R}_x(k) = \frac{1}{N}\sum_{n=1}^{N} x_n x_{n+k}$. We note how $\mathcal{R}_x(k)$ presents maxima for k multiple of the period of the sinusoid, effect of *synchronization* between the signal and its translated copy.

Intercorrelation The same concept of similitude associated to a temporal shift is all the more valid if the scalar product[36] is calculated between two *different* signals $x(t)$ and $y(t)$; in this case the operation is called an *inter*-correlation[37] integral, which for *energy* signals has expression

$$\mathcal{R}_{xy}(\tau) = \int_{-\infty}^{\infty} x^*(t)\, y(t+\tau)\, dt \tag{1.16}$$

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

If we find $\mathcal{R}_{xy}(\tau) = 0$ for any τ then the signals are called *orthogonal*, with reference to the signal space for which $\mathcal{R}_{xy}(\tau)$ is a scalar product, but also *uncorrelated*, with

[36] In fact, according to the definitions given in § ?? it results $\langle \overline{a}(t), \overline{b}(t)\rangle = \int_{-\infty}^{\infty} a(t)\, b^*(t)\, dt$ in which it is the second signal to be conjugated, and not *the first* like for (1.16): therefore this last expression corresponds (in terms of dot product) to

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

equivalent to (1.16) in that instead of anticipating $y(t)$, $x(t)$ is delayed. However, definition (1.16) is preferred for its *formal* similarity to that of a convolution.

[37] Often the same quantity is also called a *cross*-correlation.

reference to the statistical aspect (1.11) for signals with zero mean.

Link with convolution The expressions (1.15) and (1.16) are also referred to as *auto*correlation and *inter*correlation *functions*, and since their argument is a time (the interval between two samples), $\mathcal{R}_x(\tau)$ and $\mathcal{R}_{xy}(\tau)$ can also be seen as *signals* (function of τ instead of t). In the study we have already encountered an integral (of convolution) whose result is a function of time; the similarity between convolution and correlation is deeper than a simple analogy, as it turns out to be[38]

$$\mathcal{R}_{xy}(\tau) = \int_{-\infty}^{\infty} x^*(t)\, y(t+\tau)\, dt = x^*(-t) * y(t) \tag{1.17}$$

where $*$ is the usual convolution symbol.

Graphic construction The last observation invites us to draw the graphic construction of Fig. 1.6, which illustrates the procedure for calculating a value of the autocorrelation integral of $x(t) = rect_{2T}(t)$, very similar to that already illustrated for convolution (see § ??), with the difference that now *no axis inversion* is 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.

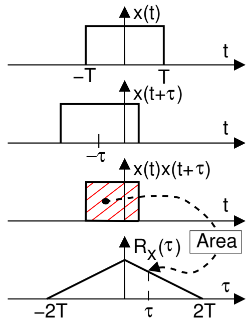


Figure 1.6: Autocorrelation of a rectangle

For an *animated* example, see the last link of footnote ?? on page ??.

1.4.5 Properties of autocorrelation

Let's now dedicate ourselves to deepen some aspects that characterize the autocorrelation function, fundamental for a better understanding the indications that $\mathcal{R}_x(\tau)$ can provide regarding the signal $x(t)$.

Invariance with respect to time shifts Autocorrelation functions $\mathcal{R}_x(\tau)$ and $\mathcal{R}_y(\tau)$ of two signals $x(t)$ and $y(t) = x(t+\theta)$ are identical[39]. Noting now that the two signals have the same magnitude $|X(f)| = |Y(f)|$ and phase spectrum that differs for a linear term (page ??), we observe that the invariance with respect to time shifts is an aspect of a more general result, that is

> *Autocorrelation does not take into account the information associated to the phase spectrum of signals*

[38] The result (1.17) is based on the change of variable $\theta = t + \tau$ that allows you to write

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

[39] In fact we get

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

Effectively $x(t)$ and $y(t)$ also have the same spectral energy density $\mathcal{E}_x(f) = \mathcal{E}_y(f) = |X(f)|^2$, as we will deepen shortly in § 1.5.1.

Temporal extension The autocorrelation of a signal with limited duration is also limited in time, with twice the duration of the original signal, as shown in Fig. 1.6. In the case of an energy signal with unlimited duration, since to obtain $\int_{-\infty}^{\infty} x^2(t)\,dt < \infty$ it is necessary that $\lim_{t\to\infty} x(t) = 0$, $\mathcal{R}_x(\tau)$ tends to zero in the same way.

Finally in the case of a power signal, as for an ergodic process member in which $\mathcal{R}_x(\tau)$ (1.14) equals the ensemble mean $m_{XX}^{(1,1)}(\tau)$, since the latter tends to zero for $\tau \to \infty$, the same happens for $\mathcal{R}_x(\tau)$, with the exception of the following two cases of periodic signal, and of a signal with non-zero mean value.

Periodic signals The autocorrelation of a periodic signal of period T is also periodic, with the same period. In fact for $\tau = nT$ the second integrand factor in (1.14) is shifted by a integer number of periods. Therefore it is not necessary to calculate the integral on the entire time axis, and the autocorrelation of periodic signals is defined as

$$\mathcal{R}_x(\tau) = \sum_{n=-\infty}^{\infty} \mathcal{R}_x^T(\tau - nT) \quad \text{where} \quad \mathcal{R}_x^T(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} x^*(t)\,x(t+\tau)\,dt \tag{1.18}$$

Continuous component If a power signal $x(t)$, or a stationary process, can be written as $x(t) = x_0(t) + a$ where $E\{x_0(t)\} = 0$ and a a constant, we find that $m_x = a$, and that[40] $\mathcal{R}_x(\tau) = \mathcal{R}_{x_0}(\tau) + a^2$: therefore in this case the autocorrelation does not vanish for $t \to \infty$, but tends to the value m_x^2.

Maximum in the origin For an autocorrelation it results $\mathcal{R}_x(0) = \max_\tau\{\mathcal{R}_x(\tau)\}$, that is, its value for $\tau = 0$ is *the maximum* with respect to any other τ. In particular, $\mathcal{R}_x(0)$ is equal to the energy of the signal $x(t)$, or to its power if $x(t)$ is a power signal, that is

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

We also note that if $x(t)$ is periodic, the last sign $\geq$ is an equality for τ multiple of a period.

Conjugate symmetry It is possible to verify[41] that results

$$\mathcal{R}_x(\tau) = \mathcal{R}_x^*(-\tau) \tag{1.19}$$

[40] Adopting the notation suitable for the case of a process, by virtue of the stationarity we can write

$$\begin{aligned} \mathcal{R}_x(\tau) =& E\{(x_0(t)+a)(x_0(t+\tau)+a)\} = \\ =& E\{x_0(t)\,x_0(t+\tau)\} + aE\{x_0(t)\} + aE\{x_0(t+\tau)\} + a^2 = \\ =& \mathcal{R}_{x_0}(\tau) + 2a\cdot 0 + a^2 = \mathcal{R}_{x_0}(\tau) + a^2 \end{aligned}$$

[41] Let's start by rewriting the expression $\mathcal{R}_x(\tau) = \int_{-\infty}^{\infty} x^*(t)\,x(t+\tau)\,dt$ by changing the variable $t+\tau = \alpha$, from which $t = \alpha - \tau$ and $dt = d\alpha$, obtaining

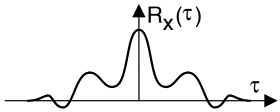
$$\mathcal{R}_x(\tau) = \int_{-\infty}^{\infty} x^*(\alpha-\tau)\,x(\alpha)\,d\alpha = \int_{-\infty}^{\infty} x(\alpha)\,x^*(\alpha-\tau)\,d\alpha = \mathcal{R}^*{}_x(-\tau)$$

while the result for $\mathcal{R}_{xy}(\tau)$ is obtained in a similar way.

and this allows (see § **??**) to state that $\mathcal{F}\{\mathcal{R}_x(\tau)\}$ is real. For the intercorrelation a similar result is obtained, that is

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

If $x(t)$ is real, we get $\mathcal{R}_x(-\tau) = \mathcal{R}_x(\tau)$, that is, the autocorrelation of a real signal is *real even*, like (as we will show now) its Fourier transform.



autocorr. of a real signal

1.5 Power density spectrum

We are finally able to characterize the spectral density for both the case of processes as well as other types of signals, i.e. power, periodic, or energy. The tool that allows this is the...

1.5.1 Wiener's theorem

It is stated without too many complications[42]:

> *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 $\mathcal{P}_x(f) = \mathcal{F}\{\mathcal{R}_x(\tau)\}$*

The proof of the theorem for energy signals is extraordinarily simple:

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

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

As anticipated, the theorem also holds for power signals, for which the autocorrelation function $\mathcal{R}_x(\tau)$ from which to start is the one expressed by (1.14) ([43]). In the case of ergodic processes, each member of the process has the same $\mathcal{P}_x(f)$, which therefore can be calculated from the $\mathcal{R}_x(\tau)$ of any of them. Finally, in the case of stationary processes at least in the wide sense, the autocorrelation from which to start[44] is the

[42] In reality the attributions of this result are many, including also *Khinchin*, *Einstein* and *Kolmogorov* - source `https://en.wikipedia.org/wiki/Wiener-Khinchin_theorem`

[43] In this case an estimate of the power density can be obtained by means of a periodogram (§ 1.7.1) calculated on a signal segment $x_T(t)$ of duration T extracted from $x(t)$ and letting $T \to \infty$, or $\mathcal{P}_x(f) = \lim\limits_{T\to\infty} \frac{1}{T}|X_T(f)|^2$. Since $|X_T(f)|^2$ is precisely the energy density $\mathcal{E}_{x_T}(f)$ of $x_T(t)$, by Wiener's theorem its anti-transform corresponds to autocorrelation function $\mathcal{R}_{x_T}(\tau) = \mathcal{F}^{-1}\left\{\mathcal{E}_{x_T}(f)\right\}$ of $x_T(t)$, as defined by (1.15). By making the passage to the limit, we obtain that

$$\mathcal{F}^{-1}\{\mathcal{P}_x(f)\} = \mathcal{F}^{-1}\lim_{T\to\infty}\left\{\tfrac{1}{T}|X_T(f)|^2\right\} = \mathcal{F}^{-1}\left\{\lim_{T\to\infty}\tfrac{1}{T}\mathcal{E}_{x_T}(f)\right\} = \lim_{T\to\infty}\tfrac{1}{T}\mathcal{R}_{x_T}(\tau)$$

corresponding to autocorrelation $\mathcal{R}_x(\tau)$ of the entire signal, as expressed by (1.14).

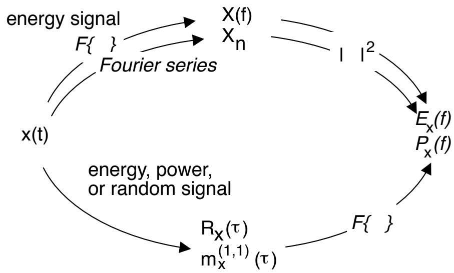
[44] The demonstration of the case of processes is carried out at § **??**; its validity is bound to processes for which $\int |\tau \cdot m^{(1,1)}_{XX}(\tau)|\, d\tau < \infty$, and is based on the consideration that if the $\mathcal{P}^{\theta}_x(f)$ of a particular θ member is evaluable as $\mathcal{P}^{\theta}_x(f) = \lim\limits_{T\to\infty}\frac{1}{T}\left|X^{\theta}_T(f)\right|^2$, then its ensemble mean can be written as $\mathcal{P}_x(f) =$

mixed moment $m_{XX}^{(1,1)}(\tau) = E\{x(t)\,x(t+\tau)\}$ calculated as the ensemble mean from (1.13), and represents the more general way of proceeding, as applied in § ?? for the case of a data signal.

Discussion Wiener's theorem is (in its simplicity) a very powerful tool that can be an *alternative way* to verify and extend known results. For example, the property $\mathcal{R}_x(0) = \mathcal{P}_x$ of maximum autocorrelation in the origin can now be derived from that of the initial value eq. (??):

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

On the other hand, the validity of the theorem also for periodic and energy signals allows to undertake for them *two alternative paths* for the calculation of the corresponding power (or energy) density, as shown in the figure to the side.



Furthermore, since by virtue of Wiener's theorem it is now possible to define a $\mathcal{P}_x(f)$ also for processes and power signals, very often in the text we will refer to power or energy density rather than to the transform of a signal, in order to apply the results to all possible cases.

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 of deterministic signal and random process.

1.5.2 Periodic signal

In this case $x(t)$ with period T can be expressed through the relative Fourier series (??) $x(t) = \sum_{n=-\infty}^{\infty} X_n\,e^{j2\pi nFt}$ referred to the fundamental frequency $F = 1/T$ (or first harmonic), and thus developing[45] the eq. (1.18) we obtain that the autocorrelation of $x(t)$ has expression $\mathcal{R}_x(\tau) = \sum_{n=-\infty}^{\infty} |X_n|^2\,e^{j2\pi nF\tau}$, that is, it can be expressed in Fourier series, and thus it is in turn periodic, as already noted. The relative power density $\mathcal{P}_x(f)$ is therefore equal to

$$\mathcal{P}_x(f) = \mathcal{F}\{\mathcal{R}_x(\tau)\} = \sum_{n=-\infty}^{\infty} |X_n|^2\,\delta(f - nF) \tag{1.20}$$

$\lim_{T\to\infty} \frac{1}{T} E_\Theta\left\{\left|X_T^\theta(f)\right|^2\right\}$.

[45] Starting from (1.18) $\mathcal{R}_x(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} x^*(t)\,x(t+\tau)\,dt$ we can write

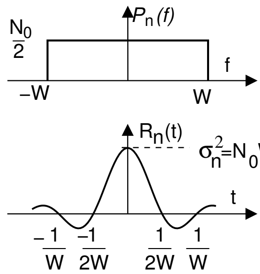
$$\mathcal{R}_x(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} \sum_{n=-\infty}^{\infty} X_n^*\,e^{-j2\pi nFt} \sum_{m=-\infty}^{\infty} X_m\,e^{j2\pi mF(t+\tau)} dt =$$
$$= \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} X_n^* X_m \frac{1}{T} e^{j2\pi mF\tau} \int_{-T/2}^{T/2} e^{j2\pi(m-n)Ft} dt = \sum_{n=-\infty}^{\infty} |X_n|^2\,e^{j2\pi nF\tau}$$

in which the last equality is the application of the property of orthogonality of the exponentials (??).

which constitutes a confirmation of Parseval's theorem (**??**). Some examples of autocorrelation and power density for periodic signals are given on page 39.

Continuous component As observed on page 18, if the signal can be written as $x(t) = x_0(t) + a$ where $E\{x_0(t)\} = 0$ and a is a constant, 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 relative 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 $x(t)$.

1.5.3 Band-limited white Gaussian process



A $n(t)$ process is called *white* when its power density is *constant* in frequency, i.e. when it is expressed as

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

where W is the bandwidth at positive frequencies. For this case the autocorrelation is

$$\mathcal{R}_n(t) = \mathcal{F}^{-1}\{\mathcal{P}_n(f)\} = N_0 W \mathrm{sinc}(2Wt) \qquad (1.21)$$

and we can therefore see that it is obtained

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

in which the last equality holds (see Eq. (1.6) on page 5) as the absence of impulses in the origin for $\mathcal{P}_n(f)$ corresponds to a $n(t)$ with zero mean. Furthermore, since $\mathcal{R}_n(t)$ zeroes for $t = k/2W$, we observe that by sampling $n(t)$ with period $T_c = 1/2W$ *uncorrelated* values are obtained, and if the process is Gaussian, the samples are also statistically independent (see § 1.6.1). This result justifies, at least from a theoretical point of view, a hypothesis that is often made: that of finding *statistically independent* noise samples superimposed on the samples of a limited bandwidth signal.

As W increases, $\mathcal{R}_n(t)$ tends to zero more rapidly, so that the noise remains correlated for an increasingly shorter time, i.e. two samples separated by the same time interval t have an ever lower correlation. A similar result is also valid more generally, since the $\mathcal{R}_x(t)$ autocorrelation of any zero-mean process (except in the periodic case, attributable to a combination of harmonic processes) tends to 0 with $t \to \infty$, i.e. from a certain t onwards the correlation is negligible.

Finally, if we imagine the band-limited white noise as the result of the transit of an infinite band Gaussian process (therefore, with $\mathcal{R}_n(t) = \delta(t)$) through an ideal low pass filter with $H(f) = rect_{2W}(f)$ (see § **??**), we realize that the correlation (1.21) between output noise samples taken at different instants ($t \neq 0$) is a direct consequence of the memory introduced by the impulse response $h(t) = 2W\mathrm{sinc}(2Wt)$ on the signal in transit, since the convolution operation between $n(t)$ and $h(t)$ makes the output values a *linear combination* of the (past) input values (see § **??**).

Exercise Given an ergodic Gaussian process with zero mean and power density $\mathcal{P}_x(f) = P_x\frac{2\beta}{\beta^2+4(\pi f)^2}$ where $\beta = 0.2$ and $P_x = 5$. Express

1. the relative autocorrelation function $\mathcal{R}_x(\tau)$;
2. the p.d.f. $p_X(x)$ of a r.v. x extracted from the process at a random instant;
3. express mean vector $\boldsymbol{m}_x$ and covariance matrix Σ_x of a pair of r.v. x_1 and x_2 extracted with an interval τ, that is $x_1 = x(t)$ and $x_2 = x(t+\tau)$, for a value of $\tau = 10$.

Solution From the table at § **??** it appears that $\mathcal{F}^{-1}\left\{\frac{2\beta}{\beta^2+4(\pi f)^2}\right\} = \mathrm{e}^{-\beta|t|}$, therefore

1. $\mathcal{R}_x(\tau) = \mathcal{F}^{-1}\{\mathcal{P}_x(f)\} = P_x \mathrm{e}^{-\beta|\tau|} = 5 \cdot \mathrm{e}^{-0.2|\tau|}$;
2. we are obviously dealing with a Gaussian r.v., therefore $p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\mathrm{e}^{-\frac{(x-m_x)^2}{2\sigma_x^2}}$ in which $\sigma_x^2 = P_x = 5$ as a zero mean, and equal to the maximum of $\mathcal{R}_x(\tau)$ for $\tau = 0$;
3. the pair $x_1 x_2$ is a two-dimensional Gaussian random vector, whose mean value is zero based on the stationarity and being the process with zero mean, while Σ_x is a 2×2 matrix with diagonal elements $\mathcal{R}_x(0) = \sigma_x^2$, while in the other two positions the covariance $\sigma_{x_1 x_2} = m_{x_1 x_2}^{(1,1)}(\tau = 5) = \mathcal{R}_x(\tau = 5) = 5 \cdot \mathrm{e}^{-0.2 \cdot 10} = 0.68$ appears, hence $m_x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\Sigma_x = \begin{bmatrix} 5 & 0.68 \\ 0.68 & 5 \end{bmatrix}$.

1.6 Multidimensional Gaussian

This term identifies a vector r.v. $\boldsymbol{X}$ obtained starting from n marginal r.v. x_i, $i = 1, 2, \cdots, n$, all Gaussian. The *joint* p.d.f. in this case is expressed in a formally similar way to that of the one-dimensional case, such as

$$p_X(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma_x)}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m}_x)\Sigma_{\boldsymbol{x}}^{-1}(\boldsymbol{x} - \boldsymbol{m}_x)^\top\right\} \tag{1.22}$$

where $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]$ is the row vector representing the n marginal r.v., $\boldsymbol{m}_x$ is the vector of the respective mean values, Σ_x is the *covariance matrix* (see § 1.4.2) whose $n \times n$ elements are equal to $\sigma_{x_i, x_j} = E\left\{(x_i - m_{x_i})(x_j - m_{x_j})\right\}$, and $^\top$ represents the transposition operator. In this case the marginal x_i are called *jointly Gaussian* r.v., and the knowledge of $\boldsymbol{m}_x$ and Σ_x fully defines the probability density. For further information of the analytical properties of Σ_x, see § **??**.

We note that the term $1/\sqrt{(2\pi)^n \det(\Sigma_x)}$ represents the height of $p_X(\boldsymbol{x})$ at $\boldsymbol{x} = \boldsymbol{m}_x$, in which the exponent of (1.22) vanishes. For $\boldsymbol{x} \neq \boldsymbol{m}_x$ the exponent itself is a *quadratic form* (page **??**) always positive, and which increases as $|\boldsymbol{x} - \boldsymbol{m}_x|$ increases.

Example In fig. 1.7-a) the 3D graph of a two-dimensional Gaussian p.d.f. $p_{XY}(x, y)$ is shown, with $\boldsymbol{m} = (0, 1)$ and $\Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}$: we can see the asymmetry due to the values $\sigma_x^2 \neq \sigma_y^2$, and the centering due to $m_y \neq 0$. Fig. 1.7-b) shows the same p.d.f. from a point of view parallel to the axes, while fig. 1.7-c) reports the *level curves*, showing how the quadratic form in the exponent determine *elliptical* contours for the surface of $p_{XY}(x, y)$, with the length of the axes of the ellipses which is related to σ_x and σ_y, while the inclination depends on the covariance σ_{xy}.

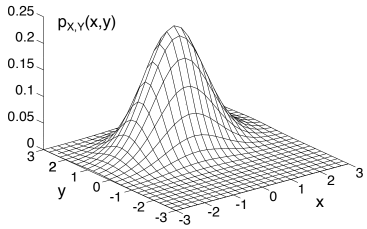
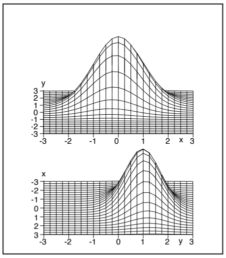
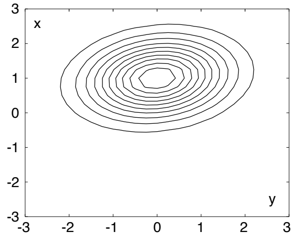
a)  b)  c) 

Figure 1.7: a) - two-dimensional Gaussian p.d.f.; b) - front and side view; c) - level curves

1.6.1 Statistical independence for uncorrelated Gaussians r.v.

We face the demonstration of what is stated at the end of § 1.4.2, namely that, *only in the case of jointly Gaussian r.v.*, their uncorrelation implies statistical independence. In fact, we observe that in the case in which the marginal r.v. are uncorrelated, that is $\sigma_{x_i x_j} = 0$ with $i \neq j$, the covariance matrix Σ_x turns out to be *diagonal*, as well as its inverse, whose elements are in this case equal to $1/\sigma^2_{x_i}$; besides, we obtains that $\det(\Sigma_x) = \prod_{i=1}^n \sigma^2_{x_i}$. Therefore in this case (1.22) is expressed as

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^n \sigma_{x_i}}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n \frac{\left(x_i - m_{x_i}\right)^2}{\sigma^2_{x_i}}\right]\right\} \tag{1.23}$$

which is evidently equivalent to the product of the single marginal p.d.f.[46]

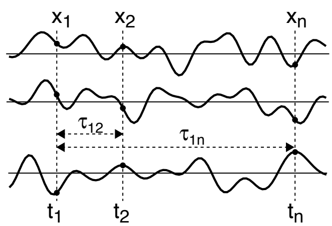
$$p(x_i) = \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp\left\{-\frac{1}{2}\frac{\left(x_i - m_{x_i}\right)^2}{\sigma^2_{x_i}}\right\}$$

But since this result is precisely the definition of statistical independence (§ 1.4.2) between the marginal r.v., we have obtained the the proof sought.

Finally we observe that in the case in which the marginal r.v. are independent, equating the exponent of (1.23) to a constant, we obtain the equation of an ellipse referred to the principal axes, i.e. the level curves of fig. 1.7-c) are arranged with the axes parallel to those of the domain described by the random vector $\boldsymbol{x}$.

1.6.2 Gaussian process

An important class of random signals consists of a stationary process in the wide sense, whose first order p.d.f. is Gaussian, and from whose members it is possible to extract at different instants one or more Gaussian random variables, which we collectively indicate with the random vector **x**, described by the multivariate p.d.f. (1.22).



Stationarity guarantees that the corresponding vector $\boldsymbol{m}_x$ of mean values has all the

[46]Let us verify (for example) that in the case of an (x, y) *pair* of statistically independent Gaussian r.v., with zero mean and variance σ^2_x and σ^2_y respectively, we get

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

elements equal to $m_x = E\{x(t)\}$, and that the Σ_x matrix of covariance has elements obtained by evaluating the $\sigma_x(\tau) = E\{(x(t) - m_x)(x(t+\tau) - m_x)\}$ covariance of the process (see eq. 1.12) in correspondence of the time intervals τ_{ij} between the sampling instants in which the pairs of marginal r.v. x_i and x_j described by the multivariate Gaussian are extracted. In other words, values σ_{ij} appearing in Σ_x are obtained as $\sigma_{ij} = \sigma_x(\tau_{ij})$, while the variance $\sigma_x^2 = \sigma_x(0)$ appears on the whole diagonal.

Being the process Gaussian, the two quantities $\boldsymbol{m}_x$ and Σ_x describe it completely, and if the ergodicity hypothesis is also verified, they can be estimated starting from any realization, see eq. (**??**) on page **??**.

1.7 Spectral estimation

Wiener's theorem (§ 1.5.1) helps us if we want to know the power density for an ergodic process, and its autocorrelation $m_X^{(1,1)}(\tau) = \mathcal{R}_X(\tau)$ is known. But often we are faced with processes whose ensemble statistics are not known, even if the ergodicity hypothesis holds true: one solution can then be to *estimate* $\mathcal{R}_X(\tau)$ starting from a realization, as shown in § 1.9.4, and from that obtain $\mathcal{P}_x(f)$. Another solution can be to directly estimate $\mathcal{P}_x(f)$ without going through autocorrelation, using instead the Fourier transform $X_T(f)$ of a time-limited segment of the signal, as described in § 1.7.1.

A different case is that of a signal which, although representative of many others, cannot be considered member of a stationary process, as its spectral characteristics vary over time, and it is precisely these variations that convey information[47]. There are different techniques to deal with this case, such as the one reported in § 1.7.2, which also uses time segments of the signal to evaluate a short-time autocorrelation function, to be used for estimating a *parametric model* of the time-varing spectral density.

1.7.1 Periodogram

Given a $x(t, \theta_i)$ realization of a process, we identify a time interval T on which to define a signal with limited duration $x_T(t) = x(t, \theta_i)\,\mathrm{rect}_T(t)$. This is an energy signal, with transform $X_T(f)$ and energy density $\mathcal{E}_{x_T}(f) = |X_T(f)|^2$, and under stationarity assumptions, we can obtain an estimate $\widehat{\mathcal{P}}_x(f)$ of the power density $\mathcal{P}_x(f)$ of the entire signal simply by dividing $\mathcal{E}_{x_T}(f)$ by the duration of the segment, or

$$\widehat{\mathcal{P}}_x(f) = \mathcal{P}_{x_T}(f) = \frac{|X_T(f)|^2}{T} \tag{1.24}$$

obtaining a function of the frequency indicated as a *periodogram*, a name linked to the use that was initially made of it, to discover traces of periodicity in a *noisy* signal. As T tends to ∞, eq. (1.24) tends to the *true* power density $\lim_{T\to\infty} \frac{|X_T(f)|^2}{T} = \mathcal{P}_x(f)$ of

[47] An example can be a sound signal, for example a reciting voice, for which we want to study the spectral characteristics of the *different sounds* of the language (the phonemes), to compare them with those of another individual, or to reduce the amount of data necessary to transmit the signal in numerical form (see § **??**), or to make a voice recognition device.

the realization $x_T(t, \theta_i)$ and, if it belongs to an ergodic process[48], to that of any other member.

Bias and spectral resolution In the real case in which T does not tend to infinity, it can be shown[49] that using $\mathcal{P}_{x_T}(f)$ (eq. (1.24)) as an estimate $\widehat{\mathcal{P}}_x(f)$ of the true density $\mathcal{P}_x(f)$ of the process, we get in its place

$$\widehat{\mathcal{P}}_x(f) = \mathcal{P}_x(f) * T(\mathrm{sinc}(fT))^2 \tag{1.25}$$

that is, a distortion of the same nature as that observed in § ?? with regard to the *temporal windowing* procedure, and showing how the estimator is *biased*[50], and characterized by a *spectral resolution*[51] (§ ??) of the order of $1/T$ Hz.

Variance of the estimate As discussed, the estimate $\widehat{\mathcal{P}}_x(f)$ tends to the true $\mathcal{P}_x(f)$ as T increases, improving at the same time the resolving power in frequency; on the other hand, however, the values of $\widehat{\mathcal{P}}_x(f)$ for a given f are still r.v., and their variance... *does not decrease* with increasings T, making the estimator *inconsistent*! Taking up the notation of the note (49), one can in fact prove[52] that the variance σ_T^2 of the estimate (1.24) 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. Even if the theory predicts that the variance of an estimator decreases as the available data increases (see (??) on page ??), this does not happen. The reason can be explained by

[48]In the opposite case in which $x(t, \theta)$ is not ergodic, its spectral density can be defined as $\mathcal{P}_x(f) = \lim_{T\to\infty} E\left\{\frac{|X_T(f)|^2}{T}\right\}$.

[49]For a given frequency f_0, the $\mathcal{P}_{x_T}(f_0) = \frac{|X_T(f_0)|^2}{T}$ value is a random variable (it depends on θ), whose expected value $m_T = E_\theta\left\{\mathcal{P}_{x_T}(f_0)\right\}$ we would like to be equal to the true density $\mathcal{P}_x(f_0)$), and whose variance $\sigma_T^2 = E_\theta\left\{\left(\mathcal{P}_{x_T}(f_0) - \mathcal{P}_x(f_0)\right)^2\right\}$ we would like to be decreasing as T increases. To verify whether these properties are satisfied, we first evaluate the *expected value* of the periodogram, starting from the relations provided by the Wiener theorem applied to $X_T(f)$, namely $|X_T(f)|^2 = \mathcal{E}_{x_T}(f) = \mathcal{F}\left\{\mathcal{R}_{x_T}(\tau)\right\}$:

$$\begin{aligned}
E_\theta\left\{\mathcal{P}_{x_T}(f)\right\} &= E_\theta\left\{\mathcal{F}\left\{\frac{1}{T}\int_{-\infty}^{\infty} x(t,\theta)\, rect_T(t)\, x(t+\tau,\theta)\, rect_T(t+\tau)\, dt\right\}\right\} = \\
&= \mathcal{F}\left\{\frac{1}{T}\int_{-\infty}^{\infty} E_\theta\left\{x(t,\theta)\, x(t+\tau,\theta)\right\} rect_T(t)\, rect_T(t+\tau)\, dt\right\} = \\
&= \mathcal{F}\left\{\mathcal{R}_x(\tau)\frac{1}{T}\int_{-\infty}^{\infty} rect_T(t)\, rect_T(t+\tau)\, dt\right\} = \mathcal{F}\left\{\mathcal{R}_x(\tau)\cdot \mathrm{tri}_{2T}(\tau)\right\} = \\
&= \mathcal{P}_x(f) * T(\mathrm{sinc}(fT))^2
\end{aligned}$$

We therefore observe how a rectangular signal window produces a triangular one on the autocorrelation. But the good thing is that as T increases the estimator tends to the true value, since $\lim_{T\to\infty} T(\mathrm{sinc}(fT))^2$ tends to an impulse.

[50]When the expected value of an estimator tends to the true value it is said (see § ??) that the estimator is *unbiased*; if, by increasing the sample size, the variance of the estimate tends to zero, the estimator is said to be *consistent*. We are consoled to verify that, as commented in the previous note, for $T \to \infty$ the polarization tends to disappear, making the estimate *asymptotically non-polarized*.

[51]The spectral resolution in this case depends on the main lobe width of the energy density of the window function applied to $\mathcal{R}_x(\tau)$ (see the penultimate passage in note 49), which in the case of $\mathrm{tri}_{2T}(\tau)$ results $(\mathrm{sinc}(fT))^2$, whose main lobe is indeed wide $1/T$. Resolution, therefore, also improves as T increases.

[52]See e.g. `http://risorse.dei.polimi.it/dsp/courses/ens_l1/books/libro07secondaparte.pdf`

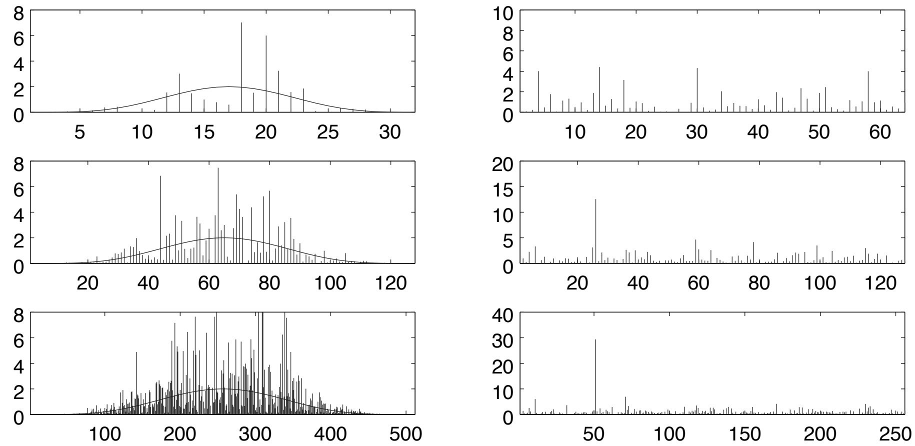


Figure 1.8: Periodogram calculated via FFT for colored noise (left) and for sinusoid immersed in noise (right), on signal intervals of increasing duration (or number of points) (from top to bottom)

considering that in a numerical implementation using DFT (§ ??), as T increases the number of frequency values that are calculated also increases, and therefore no real accumulation of data is determined for the same estimated value.[53].

Example Let's deepen the meaning of the last topic with the help of fig. 1.8, where it is shown the result of the periodogram calculation obtained by a FFT (§ ??) on a variable number of points, indicated on the abscissa. On the left side of the figure, the process $x(t, \theta_i)$ consists of colored noise with a $\mathcal{P}_x(f) = (1 - \cos(4\pi f T_c))^2$, also shown in the figure: as anticipated, the estimated value $\widehat{\mathcal{P}_x}(f)$ differs from that expected $\mathcal{P}_x(f)$ to a grater extent the larger $\mathcal{P}_x(f)$ is, for any duration of the observation.

Conversely, the right column of fig. 1.8 shows the effective usefulness of the periodogram for detecting narrow band signals immersed in noise. In this case a sine wave with frequency $f_0 = \frac{f_c}{10}$ and power $\frac{1}{2}$ was added to a white Gaussian noise with $\sigma_n^2 = 4$, obtaining thus a SNR equal to $\frac{1}{8}$, or -9 dB. Using (in top right) a 128 points FFT (and therefore half of them are shown, see page ??), the tone is difficult to distinguish from the values on which the estimate for white noise can fluctuate. But doubling (and quadrupling) the number of samples is enough to improve the situation: while the height of the line at the frequency of the sinusoid doubles (but not the variance of its estimate), the noise level remains constant (note the different horizontal scale). So here it is explained the reason for his name ☺

1.7.2 Autoregressive spectral estimate

This method is applicable to the case of non-stationary signals such as speech, which can be thought of as the result of the passage of an excitation signal (made by the

[53]There are several solutions to this problem, all of which result in a reduction of the spectral resolution. The first is to smooth the $\widehat{\mathcal{P}_x}(f)$ obtained, averaging the values on nearby frequencies: this operation corresponds to a *filtering in frequency*. A second method involves dividing the observation interval in different sub-intervals, calculating the periodogram on each of them, and averaging the results.

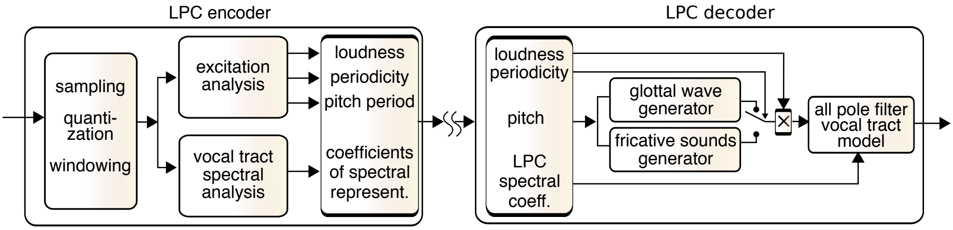


Figure 1.9: Linear Predictive enCoder (LPC) and decoder scheme

vibration of the vocal cords) through a filter consisting of the oral (and nasal) cavity that varies over time depending on the phonemes to be pronounced. Such a filter is characterized by the resonance frequencies it introduces, and can thus be considered an all-pole filter or of the IIR type (§ **??**), whose numeric transfer function takes the form $H(z) = \frac{1}{1-\sum_{i=1}^{p} a_i z^{-i}}$. By determining the coefficients a_i of the filter, smoothed spectral estimates are obtained, such as those shown in fig. 1.11, and the method is called *Linear Predictive Coding* or LPC for the reasons given next.

In this context the short-term spectral analysis is carried out by dividing the speech signal into intervals (or analysis windows) of extension between 10 and 30 msec, during which the signal can be considered practically stationary[54], and on these windows an analysis (or estimate) of the model parameters is performed, that is

- the type of excitation (periodic or chaotic), its fundamental frequency (or pitch) if periodic, and its intensity;
- the parameters that characterize the filtering effect of the vocal tract.

These parameters can therefore be used in the context of the *coding scheme* given in fig. 1.9 where, on the right end, the *synthesis filter* (or vocal tract model) of *recursive* type or IIR (§ **??**) of p order is shown. This filter outputs y_n samples based on a linear combination of p past output samples i.e. $\hat{y}_n = \sum_{i=1}^{p} a_i y_{n-i}$, to which a *prediction error* e_n is added, which in turn represents the excitation input process, that is

$$y_n = \hat{y}_n + e_n = \sum_{i=1}^{p} a_i y_{n-i} + e_n \tag{1.26}$$

For each analysis window the a_i coefficients are obtained as those that *minimize* the expected value of the quadratic error $E\left\{e_n^2\right\} = E\left\{\left(y_n - \sum_{i=1}^{p} a_i y_{n-i}\right)^2\right\}$ (i.e., the energy of the error), and are identified equating to zero the expression of the partial derivatives of $E\left\{e_n^2\right\}$ with respect to the coefficients a_j. So, let's write

$$\frac{\partial}{\partial a_j} E\left\{\left(y_n - \sum_{i=1}^{p} a_i y_{n-i}\right)^2\right\} = 2E\left\{\left(y_n - \sum_{i=1}^{p} a_i y_{n-i}\right) y_{n-j}\right\} = 0$$

from which it is obtained

$$E\left\{y_n y_{n-j}\right\} = \sum_{i=1}^{p} a_i E\left\{y_{n-i} y_{n-j}\right\} \tag{1.27}$$

The expected value $E\left\{y_{n-i} y_{n-j}\right\}$ is *estimated*[55] as that of the discrete autocorrelation

[54] A syllable can extend its duration between 10-15 msec for *reduced* vowels, up to more than 100 msec for *stressed* ones.

[55] Implying a stationarity and ergodicity hypothesis that is not true, but very handy for reaching a

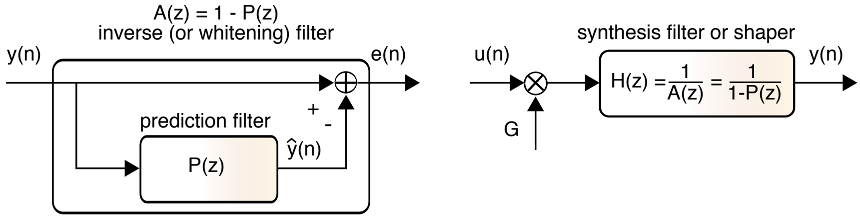


Figure 1.10: Predictor filter, associated reverse filter, and LPC synthesis filter

calculated on the signal samples delimited by the current analysis window, that is

$$\begin{aligned} E\left\{y_{n-i}y_{n-j}\right\} &\simeq \mathcal{R}_{yy}(|i-j|) \qquad \text{and placing } k=|i-j| \\ &= \mathcal{R}_{yy}(k) = \textstyle\sum_{n=1}^{N-k} y_n y_{n+k} \end{aligned} \tag{1.28}$$

where the upper bound of the summation varies so as to include only the samples actually present in the window[56]. Eq. (1.28) allows you to rewrite (1.27) as

$$\mathcal{R}_{yy}(j) = \textstyle\sum_{i=1}^{p} a_i \mathcal{R}_{yy}(|i-j|)$$

which, evaluated for $j = 1, \cdots, p$ identifies a system of p equations[57] with p unknowns

$$\begin{bmatrix} \mathcal{R}(1) \\ \mathcal{R}(2) \\ \vdots \\ \mathcal{R}(p) \end{bmatrix} = \begin{bmatrix} \mathcal{R}(0) & \mathcal{R}(1) & \cdots & \mathcal{R}(p-2) & \mathcal{R}(p-1) \\ \mathcal{R}(1) & \mathcal{R}(0) & \cdots & \mathcal{R}(p-3) & \mathcal{R}(p-2) \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ \mathcal{R}(p-1) & \mathcal{R}(p-2) & \cdots & \cdots & \mathcal{R}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \tag{1.29}$$

which can be solved in terms of the a_i coefficients by particularly efficient methods[58]; and these coefficients can be used by the decoder to apply (1.26).

The autoregressive filter that performs the $\hat{y}_n = \sum_{i=1}^{p} a_i y_{n-i}$ calculation is indicated as a *predictor*, is associated with a polynomial[59] $P(z) = \sum_{i=1}^{p} a_i z^{-i}$, and operates on the decoding side of the codec; vice versa, the FIR filter which evaluates the prediction error (or *residual*) $e_n = y_n - \sum_{i=1}^{p} a_i y_{n-i}$ is indicated as an *inverse* or *whitening filter*, is associated with the polynomial $A(z) = 1 - P(z)$, operates on the coding side of the codec, and is shown on the left side of fig. 1.10. Now indicating with $G \cdot u_n$ a coding of the residue e_n, the original signal can be (almost) re-obtained as shown in the right part of fig. 1.10, ie by passing e_n through the IIR filter $H(z) = \frac{1}{A(z)} = \frac{1}{1-P(z)}$.

---

result.

[56]The (1.28) is effectively an estimate of the *autocorrelation* of the signal with a limited duration that falls within the analysis window, while the inclusion in the sum of a number of terms equal to the number of available samples leads to a different type of result , called the *covariance method*, and a different way of solving the system(1.29).

[57]Sayings of *Yule-Walker*, see e.g. https://it.wikipedia.org/wiki/Equazioni_di_Yule-Walker

[58]On the basis of the adopted assumptions, $R_{yy}(j)$ is an even function of the index j, and the corresponding matrix of the coefficients is called *Toeplix*, allowing its inversion by means of the *Levinson-Durbin* method (see https://en.wikipedia.org/wiki/Levinson_recursion), which presents a complexity $O(n^2)$ instead of $O(n^3)$, as would be necessary to invert the coefficient matrix.

[59]A brief analysis of the relationship between DFT and zeta transform is carried out at § **??**, but see also § **??**.

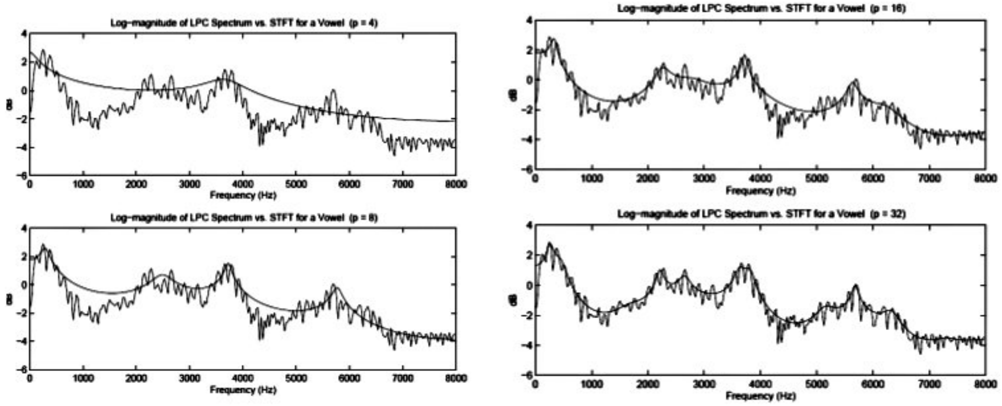
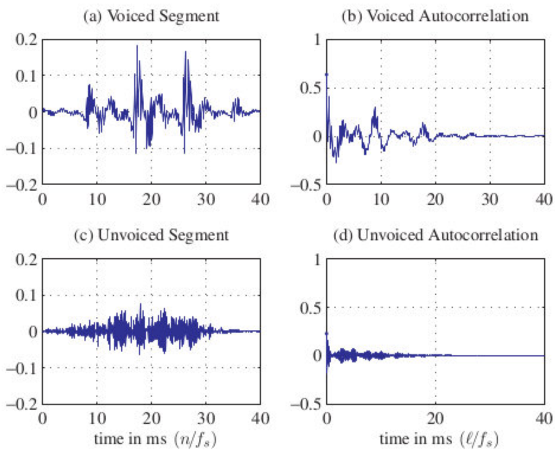


Figure 1.11: LPC spectral approximation by order of prediction p equal to 4, 8, 16 and 32

Since, based on considerations that we do not develop, e_n is characterized by a *white* spectral density, $|H(z)|^2$ (calculated for $z = e^{i\omega}$) represents a real *spectral estimate* of the original signal, as shown in fig. 1.11 for different values of p, verifying that for vowel sounds acceptable results are obtained already for values of p between 8 and 14, while for fricatives the order can be even lower.

1.7.2.1 Estimate of the pitch period

It now remains to illustrate how to decide whether the analysis window contains a fricative or voiced sound, and in the second case, its period. We observe that on average the pitch frequency is equal to about 120 and 210 Hz in the case of male and female voices respectively, with an extension that varies approximately from half to double the average pitch.[60]. The estimate of the pitch period can be realized starting from the short-term autocorrelation function (1.28), shown in the right column of the figure on the side, next to the signal windows on which it was calculated[61], for a vocalic (above) and fricative (below) sound. As evident, in the case of the voiced sound the autocorrelation presents a first peak at 9 msec and a second at 18 msec, corresponding to the pitch period and its double; vice



[60]The pitch varies during the pronunciation of a sentence according to its semantics, language, and the emotional emphasis imparted by the speaker. From a musical point of view, the dynamics of the values (from half to double) therefore extends over an interval of two octaves. The whole range of the registers of the opera differs by 22 semitones, from E2 of the bass to C4 of the soprano, that is a frequency ratio equal to 3.6.

[61]In reality, before the calculation of the autocorrelation the signal segment was multiplied by a Hamming window, which causes the visible smoothing at the edges.

versa in the case of the noise-like sound, no peaks are visible, as to be expected in the case of an uncorrelated signal. Therefore, autocorrelation can be used to indicate the presence or absence of a voiced sound, and if so, estimate its pitch.

In practice, good synthesis results are obtained for fricative sounds using real white noise as excitation; on the other hand, for voiced sounds the use of impulsive waveforms with a period equal to the estimated pitch, although capable of producing a reducible bit rate up to 2.4 kbps, does not provide particularly usable results, producing a rather robotic voice. For this reason, other techniques have been developed.

1.8 Filtering of signals and processes

Wiener's theorem (§ 1.5.1) provide a unifying approach to the definition of the power spectrum by virtue of its connection with the autocorrelation function for all signal types, including processes. Now let's go back to the analysis began at § **??** extending it to the class of random signals, in order to describe from the spectral and statistical point of view the output of a filter with impulse response $h(t)$, that is of $y(t) = x(t) * h(t)$, for the cases of an energy, periodic or random input signal.

1.8.1 Spectral density at the output of 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 is input, or $\mathcal{P}_y(f)$ and $\mathcal{P}_y$ if a periodic, power, or process signal $x(t)$ is given at the input.

Energy signals 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

$$\begin{aligned}\mathcal{E}_y(f) &= X(f)\,H(f)\,X^*(f)\,H^*(f) = |X(f)|^2\,|H(f)|^2 = \\ &= \mathcal{E}_x(f)\,|H(f)|^2\end{aligned}$$

and therefore *the energy density of the output is equal to the product between that in input and* $|H(f)|^2$. At this point, by executing the Fourier antitransform of both sides and remembering (**??**), we obtain

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

or

> *the autocorrelation of the output of a filter is equal to the convolution between the autocorrelation of the input and that of the impulse response*

We anticipate that this result is valid (in the respective terms) also for the cases of a periodic or random signal. Hence, we note that $|H(f)|^2$ 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 there are the following results[62], all equivalent for the purpose of

[62]The fourth equality exists by virtue of the Parseval's theorem associated to Wiener's, while the last is valid if $\mathcal{R}_H(\tau)$ is real, that is, if $h(t)$ is ideally realizable and therefore real, see § **??**.

calculating the total energy:

$$\begin{aligned} \mathcal{E}_y &= \mathcal{R}_y(0) = \int_{-\infty}^{\infty} \mathcal{E}_y(f)\,df = \int_{-\infty}^{\infty} \mathcal{E}_x(f)\,|H(f)|^2\,df = \\ &= \int_{-\infty}^{\infty} \mathcal{R}_x(\tau)\,\mathcal{R}_h(\tau)\,d\tau = \int_{-\infty}^{\infty} \mathcal{R}_x(\tau)\,\mathcal{R}_h^*(\tau)\,d\tau \end{aligned} \qquad (1.30)$$

Periodic signals In this case the input signal $x(t)$ can be expressed in terms of a Fourier series

$$x(t) = \sum_n X_n e^{j2\pi nFt}$$

which corresponds to a transform $X(f) = \sum_n X_n\,\delta(f - nF)$ (see eq. (**??**) on page **??**) and a power density $\mathcal{P}_x(f) = \sum \left|X_n\right|^2 \delta(f - nF)$ (see eq. (1.20)).

The output signal $y(t)$ is also periodic[63], and its Fourier coefficients Y_n can be expressed in terms of those of the input X_n and of the values of the frequency response (see § **??**) as $Y_n = X_n H(nF)$, i.e. in module and phase as

$$|Y_n| = |X_n|\,|H(nF)|\,; \qquad \arg(Y_n) = \arg(X_n) + \arg(H(nF))$$

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

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

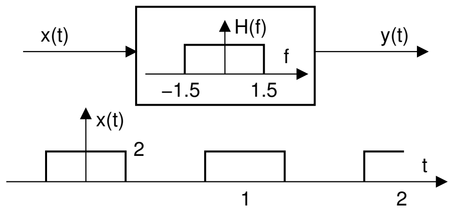
Again, anti-transformation gives $\mathcal{R}_y(\tau) = \mathcal{R}_x(\tau) * \mathcal{R}_h(\tau)$.

Exercise Let the filter in the figure be given with

$$H(f) = rect_3(f)$$

and at whose input is placed the signal

$$x(t) = 2\sum_{n=-\infty}^{\infty} rect_{\frac{1}{2}}(t - n)$$



Calculate:

1) the input power $\mathcal{P}_x$,
2) the output power $\mathcal{P}_y$,
3) the expression of $y(t)$.

Answers

1) Let's calculate the time average: $\mathcal{P}_x = \frac{1}{T}\int_{-T/2}^{T/2} x^2(t)\,dt = \frac{1}{T}\int_{-1/4}^{1/4} 2^2(t)\,dt = \frac{4}{2} = 2$, given that $T = 1$;

2) We know that $\mathcal{P}_y = \int_{-\infty}^{\infty} \mathcal{P}_y(f)\,df$, where $\mathcal{P}_y(f) = \mathcal{P}_x(f)\,|H(f)|^2$, and being $x(t)$ periodic, we have $\mathcal{P}_x(f) = \sum_n |X_n|^2\,\delta(f - nF)$. To determine the coefficients X_n of the series, we calculate

$$X(f) = \mathcal{F}\{2rect_\tau(t) * \sum_{n=-\infty}^{\infty}\delta(t - nT)\} = 2\tau \cdot \mathrm{sinc}(f\tau)\,\frac{1}{T}\sum_{n=-\infty}^{\infty}\delta\left(f - \frac{n}{T}\right)$$

and being $\tau = \frac{1}{2}$ and $T = 1$, it results

$$X(f) = \mathrm{sinc}\left(\frac{f}{2}\right)\sum_{n=-\infty}^{\infty}\delta(f - n) = \sum_{n=-\infty}^{\infty} X_n\delta(f - n)$$

with $X_n = \mathrm{sinc}\left(\frac{n}{2}\right)$. Therefore, since the only impulses that fall within the frequency response $H(f)$ are those for $f = -1$, 0 and 1, we have:

$$\mathcal{P}_y(f) = \mathcal{P}_x(f)\,|H(f)|^2 = \sum_{n=-1}^{1} |X_n|^2\,|H(n)|^2\,\delta(f - n)$$

and therefore we obtain

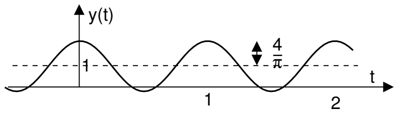
[63] Taking into account the linear and permanent nature of the filter, the output is the combination of the effects of the inputs, which for a periodic signal correspond to harmonics.

$$\mathcal{P}_y = \int_{-\infty}^{\infty} \mathcal{P}_y(f)\, df = \left(\frac{\sin -\frac{\pi}{2}}{-\frac{\pi}{2}}\right)^2 + 1 + \left(\frac{\sin \frac{\pi}{2}}{\frac{\pi}{2}}\right)^2 = 1 + 2\left(\frac{2}{\pi}\right)^2 = 1.811$$

3) Considering again that $T = 1/F = 1$, and which results $X_n = \left\{\frac{2}{\pi}, 1, \frac{2}{\pi}\right\}$, you get

$$y(t) = \sum_{n=-1}^{1} X_n H(n)\, e^{j2\pi nt} = 1 + \frac{2}{\pi}\left(e^{j2\pi t} + e^{-j2\pi t}\right) = 1 + \frac{4}{\pi} cos 2\pi t$$

We note how the filter lets only the direct component and the first harmonic pass through.



Ergodic processes and power signals Also in this case (in the appendix ??) it is verified that $m_Y^{(1,1)}(\tau) = m_X^{(1,1)}(\tau) * \mathcal{R}_h(\tau)$, and so

$$\mathcal{P}_y(f) = \mathcal{P}_x(f)\, |H(f)|^2 \tag{1.31}$$

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 (1.31) is also valid for any power signal.

Exercise A white Gaussian process with power density $\mathcal{P}_n(f) = \frac{N_0}{2}$ passes through a causal filter with $h(t) = \mathrm{e}^{-at}$. Determine the output $\mathcal{P}_y(f)$.

Answer Although in § ?? it is reported that for this case it results $|H(f)|^2 = \frac{1}{a^2+4(\pi f)^2}$ and therefore $\mathcal{P}_y(f) = \mathcal{P}_x(f)\,|H(f)|^2 = \frac{N_0}{2a^2+8(\pi f)^2}$, let's verify to obtain the same result by passing through $\mathcal{R}_h(\tau)$, which is equal[64] to $\frac{1}{2a}\mathrm{e}^{-a|\tau|}$, from which we we get[65] $|H(f)|^2 = \mathcal{F}\{\mathcal{R}_h(\tau)\} = \frac{1}{a^2+4(\pi f)^2}$.

Power gain It is the name by which the ratio

$$|H(f)|^2 = \frac{\mathcal{P}_y(f)}{\mathcal{P}_x(f)} \tag{1.32}$$

is most often indicated (see § ??), 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 $H(f)$ (see § ??). Other times $|H(f)|^2$ is also referred to as the *power response*, or even *spectral density of the power response*, while its identification with the $\mathcal{E}_h(f)$ *energy density of the filter* as done at page 30 it's my *original* definition (to my knowledge).

[64] Since $h(t)$ is real, we know that $\mathcal{R}_h(\tau)$ is real even (page 18), so it is sufficient to calculate $\mathcal{R}_h(\tau)$ only for $\tau \geq 0$; moreover, being $h(t) = 0$ for $t < 0$ the lower integration extreme is zero, obtaining

$$\mathcal{R}_h(\tau)\big|_{\tau>0} = \int_0^{\infty} h(t)\, h(t+\tau)\, dt = \int_0^{\infty} \mathrm{e}^{-at}\mathrm{e}^{-a(t+\tau)} dt = \mathrm{e}^{-a\tau}\int_0^{\infty} \mathrm{e}^{-2at} dt = \mathrm{e}^{-a\tau}\cdot \left.\frac{\mathrm{e}^{-2at}}{-2a}\right|_0^{\infty} = \frac{1}{2a}\mathrm{e}^{-a\tau}$$

and therefore $\mathcal{R}_h(\tau) = \frac{1}{2a}\mathrm{e}^{-a|\tau|}$

[65] Leaving out the term $\frac{1}{2a}$ it turns out

$$\mathcal{F}\left\{\mathrm{e}^{-a|t|}\right\} = \int_{-\infty}^{\infty} \mathrm{e}^{-a|t|}\mathrm{e}^{-j2\pi ft} dt = \int_{-\infty}^{0} \mathrm{e}^{(a-j2\pi f)t} dt + \int_0^{\infty} \mathrm{e}^{-(a+j2\pi f)t} dt = \left.\frac{\mathrm{e}^{(a+j2\pi f)t}}{a-j2\pi f}\right|_{-\infty}^{0} + \left.\frac{\mathrm{e}^{-(a+j2\pi f)t}}{-(a-j2\pi f)}\right|_0^{\infty}$$

$$= \frac{1}{a-j2\pi f} + \frac{1}{a-j2\pi f} = \frac{a+j2\pi f + a - j2\pi f}{(a-j2\pi f)(a+j2\pi f)} = \frac{2a}{a^2 - (j2\pi f)^2} = \frac{2a}{a^2 + 4(\pi f)^2}$$

1.8.2 Statistical characteristics at the output of a filter

Let us now deepen the study of the statistical characterization of the output of a filter when a member of an ergodic process is present at the input, as regards the mean, the variance, and the p.d.f. del outbound process.

Mean value It is equal to that of the input multiplied by the *zero frequency gain* $H(0)$ of the filter, as

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

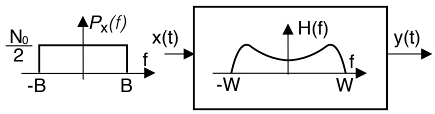
Variance The ergodic condition allows to write $\mathcal{P}_y = \overline{y^2(t)} = E\left\{y^2\right\} = m_y^{(2)}$ and therefore, recalling eq. (1.6), we have

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

where to evaluate $\mathcal{P}_y$ one can use relations analogous to (1.30):

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

Example In input to a filter $H(f)$ a white process $x(t)$ is placed, with $m_x = 0$ and band B, that is $\mathcal{P}_x(f) = \frac{N_0}{2}\mathrm{rect}_{2B}(f)$, and therefore $\mathcal{R}_x(\tau) = N_0 B \cdot \mathrm{sinc}(2Bt)$.



Being the process with zero mean, for the output variance we obtain

$$\sigma_y^2 = \mathcal{P}_y = \int_{-\infty}^{\infty} \mathcal{P}_x(f)\, |H(f)|^2\, df = \frac{N_0}{2}\int_{-B}^{B} |H(f)|^2\, df \le \frac{N_0}{2}\mathcal{R}_h(0)$$

with the equal sign if the bandwidth W of $H(f)$ is less than B[66]. With regard to the power spectral density $\mathcal{P}_y(f)$ output (1.31) applies, obtaining $\mathcal{P}_y(f) = \mathcal{P}_x(f)\, |H(f)|^2 = \frac{N_0}{2}|H(f)|^2$ with $|f| \le B$: therefore the relative output process $y(t)$ is no longer white, and in this case it is said to be *colored*. To this also corresponds to a modification of the autocorrelation function, which no longer is a sinc, but now $\mathcal{R}_y(\tau) = \mathcal{R}_x(\tau) * \mathcal{R}_h(\tau) = N_0 B \cdot \mathrm{sinc}(2Bt) * \mathcal{R}_h(\tau)$ holds. This means that while before the filter (for the white process) two values extracted at multiple instants of $1/2B$ were uncorrelated, the filtering operation introduces *a dependency* between the values extracted at these intervals[67] .

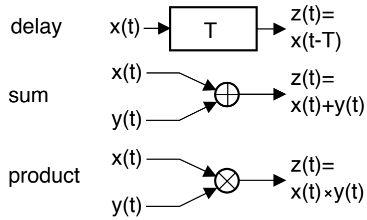
Probability density function Regarding the $p_Y(y)$ of the output process, nothing general can be said, except that it obviously depends on the input $p_X(x)$ and on the operations performed by filter; its exact expression, however, must be determined from

[66] In fact in this case $\int_{-W}^{W} |H(f)|^2\, df$ is just equal to the energy of $h(t)$; if vice versa $W > B$ a part of $|H(f)|^2$ falls outside outside the integration extremes $(-B, B)$, and does not contribute to the result.

[67] This result can be analyzed by remembering that the convolution integral calculates the single values in output from a filter, as dependent on all past inputs, each weighted with value of the impulse response relative to the delay between past input and present output (see § ??). Therefore, even if the single input values are statistically independent, the output ones (distant from each other for less than the duration of the response impulsive) share a piece of common history, and therefore theirs values are no longer uncorrelated.

time to time. For example, in the case of a transversal filter (§ ??) the variable change rules (§ ??) can be applied. A separate case is that of Gaussian processes, which when placed in input of a filter, produce a Gaussian process at the output[68].

1.9 Elementary operations on signals



At § ?? we have seen how the filtering elements can be made by a combination of the three elementary operations shown in the figure, that is *delay*, *sum* and *product*. Before continuing, let's therefore deepen the result of the combination of deterministic and random signals by means of the operators introduced. Let's start by observing that for the two cases of sum and product the following combinations can occur:

- $x(t)$ is a deterministic signal and $y(t)$ a random process: the result (*in general*) is a *non-stationary* process, in fact now the ensemble averages depend, instant by instant, on the value that the deterministic signal assumes in that instant;
- $x(t)$ is a periodic signal and $y(t)$ a random process: a process called *cyclostationary* is obtained, since even if the statistics vary over time, they assume *identical* values with periodicity equal to that of the deterministic signal;
- $x(t)$ is a constant signal equal to a and $y(t)$ is a random process: $z(t)$ is a process of the same nature of $y(t)$, with mean m_z equal to the sum (od the product) between m_y and a, and with power $\mathcal{P}_z = \mathcal{P}_y + a^2$ (or $\cdot a^2$), and autocorrelation $\mathcal{R}_z(\tau) = \mathcal{R}_y(\tau) \cdot a^2$ (or $+a^2$).

We also note that a periodic signal can often be treated as a process, simply by assuming for it a uniform phase over a period, in order to refer to the case of the harmonic process, see § ??. So, in the following we deal only with the case of processes.

1.9.1 Delay

This constitutes a particular case of a perfect channel (page ??), and analytically it corresponds to the convolution with a shifted impulse, i.e. $z(t) = x(t) * \delta(t-T)$. Therefore, the only thing that changes[69] at the output is the phase spectrum, in which a linear term appears, or $Z(f) = X(f) \cdot e^{-j2\pi fT}$, while mean value, autocorrelation and power/energy density remain unchanged, just as the variance and the p.d.f. do not change for the case of processes.

[68]This result is a direct consequence of the invariance property of the Gaussian processes with respect to linear transformations discussed in § ??. In fact, rewriting the convolution operation $y(t) = \int x(\tau) h(t-\tau) d\tau$ in approximate form as a sum of infinite terms $y(t) = \sum_i x(\tau_i) h(t-\tau_i) \Delta\tau_i$ it appears evident that, in the case where $x(t)$ is a Gaussian process, the output is constituted by a linear combination of Gaussian r.v., and therefore it also is Gaussian.

[69]See §§ ?? and ??.

1.9.2 Sum of random signals

Let's proceed in the calculation of the representative quantities making use (apart from the mean value) of the hypothesis of statistical independence between $x(t)$ and $y(t)$:

Mean value

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

We note that this result is also valid in the absence of statistical independence[70].

Total power

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

given that for statistically independent processes the joint p.d.f. $p_{XY}(x,y)$ is factored into the product of the marginals $p_X(x)\,p_Y(y)$, just as the expected value of the product is factored[71]; the next results also are valid only in the hypothesis of statistical independence.

Variance From the two previous relations we get

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

Autocorrelation

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

given that for independent, stationary and jointly ergodic processes[72] it results $E\{x(t)\,y(t+\tau)\} = \mathcal{R}_{xy}(\tau)$, i.e. equal to the product of the means $m_x m_y$, see eq. (1.11). We also observe that for $\tau = 0$ we find again the value of the total power $\mathcal{P}_z$.

Power density spectrum

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

and we emphasize once again that in the case of *non-independent* processes the result *is invalid.*

---

[70] In fact, by virtue of the distributive property, the saturation of a r.v. at a time it is possible, that is

$$\begin{aligned} \int\int (x+y)\,p(x,y)\,dxdy &= \int\int x\cdot p(x,y)\,dxdy + \int\int y\cdot p(x,y)\,dxdy = \\ &= \int x\cdot p(x)\,dx + \int y\cdot p(y)\,dy \end{aligned}$$

[71] In fact it is

$$\begin{aligned} E\{x(t)\cdot y(t)\} &= \int\int x\cdot y\cdot p(x,y)\,dxdy = \int x\cdot p(x)\,dx \cdot \int y\cdot p(y)\,dy = \\ &= E\{x(t)\}\,E\{y(t)\} = m_x m_y \end{aligned}$$

[72] The joint ergodicity property corresponds to verifying the ergodic conditions also for mixed moments $m_{XY}^{(1,1)}(x,y)$ relating to pairs of values extracted from two different processes realizations.

Probability density function In the case of independent $x(t)$ and $y(t)$ we obtain that the sum process is characterized by an *amplitude* density equal to[73]

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(\theta)\, p_Y(z-\theta)\, d\theta = p_X(x) * p_Y(y) \tag{1.33}$$

This result confirms the one already obtained at § **??** and relative to the sum of random variables, i.e. that the p.d.f. of a sum of independent r.v. is obtained by convolution between the densities of the terms of the sum.

Finally, we note that if two Gaussian processes are added, the result is still Gaussian, as discussed in § **??**: in fact the convolution between Gaussian functions is still a Gaussian, with mean equal to the sum of the means, and variance equal to the sum of variances.

1.9.3 Product of random signals

Also for the product between signals the considerations made in § 1.9 apply. If the factors $x(t)$ and $y(t)$ of the product $z(t)$ are statistically independent, stationary and jointly ergodic processes, we have

Mean value

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

since as already observed it results $p_{XY}(x,y) = p_X(x)\, p_Y(y)$ thus allowing the factoring of the expected value of the product, that is $\iint x\, y\, p(x,y)\, dxdy = \int x\, p(x)\, dx \cdot \int y\, p(y)\, dy$.

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

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

Autocorrelation function

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

In particular, we note that the incorrelation of one of the two processes, for a certain value of τ, cause the uncorrelation of the product, at the same time τ.

Power density spectrum

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

[73] We prove (1.33) with a perhaps unorthodox but effective reasoning. From the definition of p.d.f. we have that $z = x + y$ is between z and $z + dz$ with probability $p_Z(z)\, dz$, but for this to happen it is necessary that, for every possible value of x, it results $y = z - x$; for the hypothesis of statistical independence between x and y this happens with joint probability $p_X(x)\, dx \cdot p_Y(z-x)\, dz$. To obtain $p_Z(z)\, dz$ it is therefore necessary to sum the joint probability on all possible values of x, that is $p_Z(z)\, dz = \int_{\Omega_X} p_X(x)\, p_Y(z-x)\, dxdz$ where Ω_X is the sample space for the r.v. x. Therefore in the end we obtain $p_Z(z) = \int_{\Omega_X} p_X(x)\, p_Y(z-x)\, dx$ which corresponds to the convolution expressed in the text.

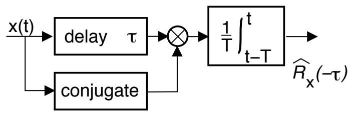
that is, it is equal to the convolution between the spectral densities of the factors. We therefore note that the power density of the product has a bandwidth occupation greater than that of the single factors.

Probability density function It is calculated with the rules for changing a variable, illustrated at § ??. In case the two processes x and y are statistically independent, the result is[74]

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(\theta)\, p_Y\left(\frac{z}{\theta}\right) \frac{d\theta}{|\theta|} \tag{1.34}$$

1.9.4 Estimation of the autocorrelation

As an example of the use of elementary operators, the figure on the side shows the architecture of a processing scheme suitable for estimating[75] the autocorrelation function $\widehat{\mathcal{R}}_x(\tau)$ (§ 1.4.4) of a signal $x(t)$ for an assigned τ advance: in fact it calculates



$$\widehat{\mathcal{R}}_x(-\tau) = \frac{1}{T}\int_{t-T}^{T} x^*(t)\, x(t-\tau)\, dt$$

for $\tau \geq 0$, from which $\widehat{\mathcal{R}}_x(\tau)$ is obtained by applying eq. (1.19). The integrating element is a low-pass filter with an $h(t) = \frac{1}{T} rect_T(t)$. Varying τ we get $\widehat{\mathcal{R}}_x(\tau)$ for the different values of the argument, and if $x(t)$ is a stationary we can get $\widehat{\mathcal{P}}_x(f) = \mathcal{F}\{\widehat{\mathcal{R}}_x(\tau)\}$; finally, if $x(t)$ is a member of an ergodic process, $\widehat{\mathcal{P}}_x(f)$ represents an estimate of the power density for any realization.

1.10 Appendices

1.10.1 Pearson correlation coefficient

The example diagrams presented in fig. 1.4 base the evaluation of how much a pair of r.v. x and y are correlated also on the calculation of the correlation coefficient[76] ρ_{xy}, which takes values between +1 and −1, and is defined as

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

---

[74] We prove (1.34) using the method illustrated in § ??, writing the system (??) as $\left\{\begin{array}{l} z = xy \\ w = y \end{array}\right.$ so that the inverse transformation results in $\left\{\begin{array}{l} x = z/w \\ y = w \end{array}\right.$. At this point we obtain the Jacobian matrix $\mathbf{J} = \begin{bmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{bmatrix}$ as $\begin{bmatrix} 1/w & z/w^2 \\ 0 & 1 \end{bmatrix}$ which corresponds to the modulus of the (Jacobian) determinant $\left|\det(\mathbf{J})\right| = \frac{1}{|w|}$. Therefore the joint p.d.f of z and w is obtained as $p_{ZW}(z,w) = \left|\det(\mathbf{J})\right| \cdot p_{XY}(x, y = f(z,w)) = \frac{1}{|w|} \cdot p_{XY}\left(\frac{z}{w}, w\right) = \frac{1}{|w|} \cdot p_X\left(\frac{z}{w}\right) p_Y(w)$ by virtue of the statistical independence between x and y. All that remains is to saturate the $p_{ZW}(z,w)$ with respect to w, that is $p_Z(z) = \int \frac{1}{|w|} \cdot p_X\left(\frac{z}{w}\right) p_Y(w)\, dw$, which corresponds to (1.34) if we set $\theta = x$ instead of $w = y$.

[75] This is an estimate (see § ??) since the integration interval T is limited.

[76] See e.g. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

In this way, a normalization of the value of the covariance σ_{xy} is carried out, with respect to the standard deviations σ_x e σ_y of the two r.v., thus making the value of ρ independent of the dynamics of the values taken by x and y.

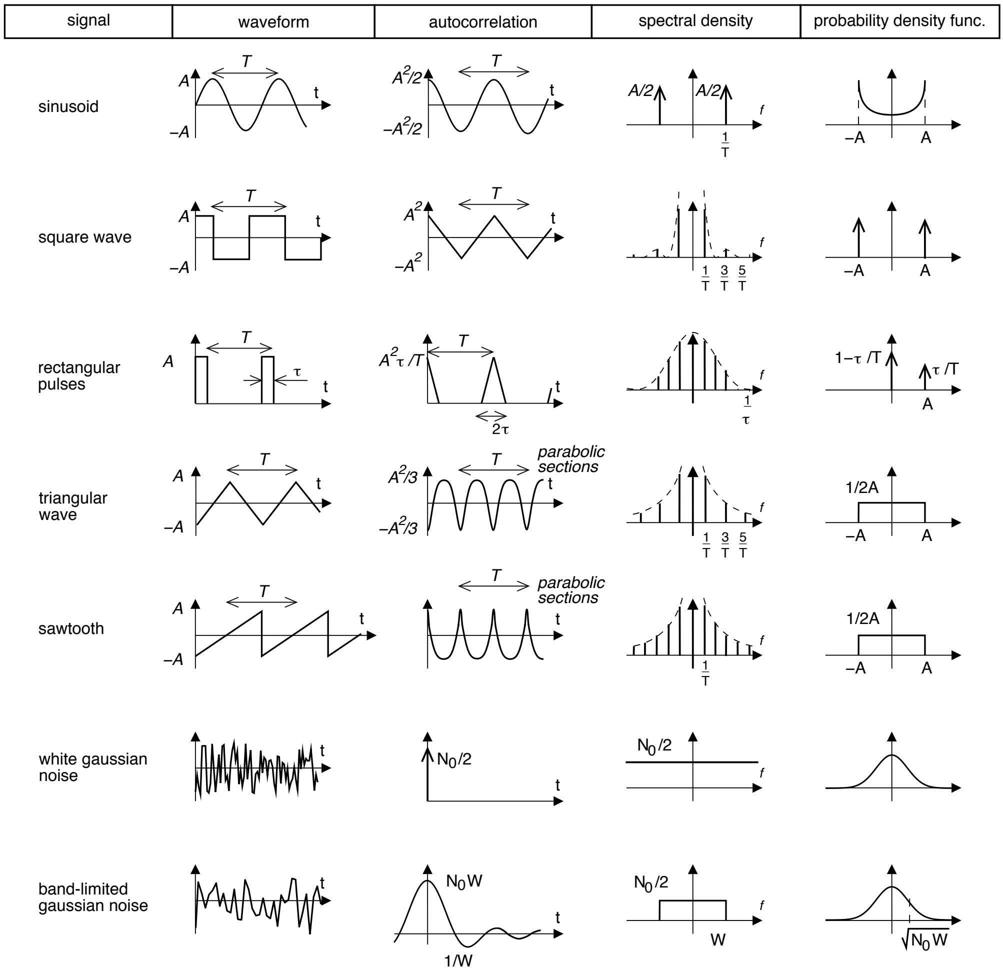
The coefficient ρ also lends itself to an interesting geometric interpretation, once the standard deviation σ_x has been related to the norm $\|\overline{x}\|$ of x (see § **??**), and the σ_{xy} covariance with the dot product $\langle \overline{x}, \overline{y} \rangle$ between x and y[77]. In this context we can define two r.v. as orthogonal if it results $\sigma_{xy} = \rho_{xy} = 0$, while a value $\rho_{xy} = \pm 1$ indicates that one of the two r.v. is always proportional to the other, with a constant factor. We recall that orthogonality $\rho_{xy} = 0$ expresses only the absence of a linear relationship between x and y, as exemplified by the case F) of fig. 1.4.

We also mention the formal extension of the result known as Schwartz's inequality (page **??**), once the concept of cosine between x and y has been associated with the correlation coefficient ρ_{xy}: such an identification derives from being $-1 < \rho_{xy} < 1$, and allows us to assert that $|\sigma_{xy}| \leq \sigma_x \sigma_y$.

1.10.2 Example graphics

Here are are the graphs of the waveform, autocorrelation, spectral density and probability density function, for some typical signals.

[77] The analogy is not too strange, considering that if x is extracted from an ergodic process with zero mean, its variance σ_x^2 coincides with the power of the signal from which it is extracted, while if x and y are extracted from jointly ergodic signals, the covariance σ_{xy} coincides with the intercorrelation function (eq. (1.16)), or with their mutual power.


signal
waveform
autocorrelation
spectral density
probability density func.
sinusoid
square wave
rectangular pulses
triangular wave
sawtooth
white gaussian noise
band-limited gaussian noise
T
A
−A
t
A²/2
−A²/2
A/2
f
1/T
A²
−A²
3/T
5/T
τ
A²τ /T
2τ
1/τ
1−τ /T
τ /T
parabolic sections
A²/3
−A²/3
1/2A
N₀/2
N₀W
1/W
W
√N₀W