Lesson 1

Signals and Systems

1.1 General considerations

The purpose of signal theory is to study signals and the systems that transmit them. The notion of *signal* is extensive. The observation of some phenomenon yields certain quantities that depend on time (on space, on frequency, or on something else!). These quantities, which are assumed to be measurable, will be called signals. They correspond in mathematics to the notion of function (of one or more variables of time, space, etc.), and thus signals are modeled by functions. We will see later that the notion of distribution provides a model for signals that is both more general and more satisfactory than that of function.

Examples of signals:

- Intensity of an electric current
- Potential difference between two points in a circuit
- Position of an object, located with respect to time, $M = M(t)$, or with respect to space, $M = M(x, y, z)$
- Gray levels of the points of an image $g(i, j)$
- Components of a field $\mathbf{V}(x, y, z)$
- A sound

There are different ways to think about a signal:

(i) It can be modeled deterministically or statistically. The deterministic point of view will be the only one used here.

(ii) The variable can be continuous; one is then said to have an *analog signal* $x = x(t)$. If the variable is discrete, one is said to have a *discrete signal* $x = (x_n)_{n\in\mathbb{Z}}$. A discrete signal will most often result from *sampling* (also called *discretizing*) an analog signal. (See Figure 1.1.)

(iii) Finally, we will consider the values $x = x(t)$ of a signal to be exact real or complex numbers. However, for computer processing, it is necessary to store these numbers in some finite form, for example, as multiples of an elementary quantity q. This approximation of the exact values is called

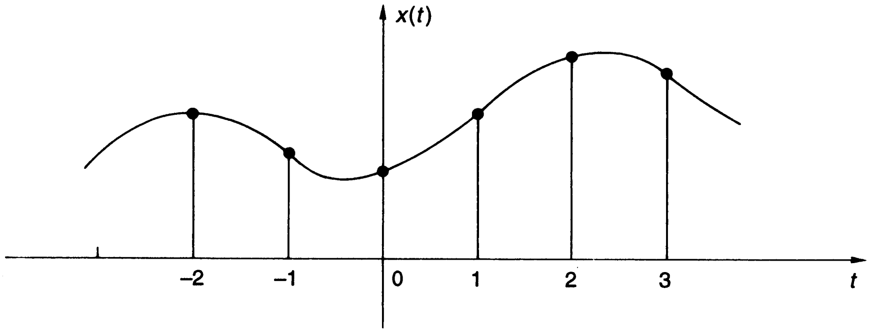

FIGURE 1.1. Sampling an analog signal.

quantization. We will not examine the effects of this operation. A discrete, quantized signal is called a *digital signal.*

Any entity, or apparatus, where one can distinguish input signals and output signals will be called a (transmission) *system* (Figure 1.2). The input and output signals are not necessarily of the same kind (see, for example, Section 1.3.7).

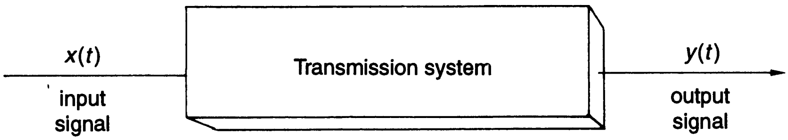

FIGURE 1.2. Diagram of a system.

When there are several input or output signals, the functions $x(t)$ and $y(t)$ are vectors. We will limit our discussion to the scalar case, where there is a single input signal and a single output signal.

In signal theory, one is not necessarily interested in the system's components, but rather in the way it transforms the input signal into the output signal. It is a "black box." It will be modeled by an operator acting on functions, and we write

$$y = Ax,$$

where $x \in X$, the set of input signals, and $y \in Y$, the set of output signals.

Examples of systems:

- An electric circuit
- An amplifier
- The telephone
- The Internet

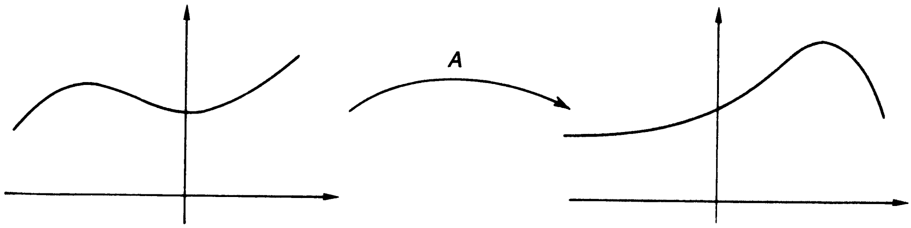

FIGURE 1.3. Analog system.

One distinguishes:

- *Analog systems* that transform an analog signal into another analog signal (Figure 1.3)
- *Discrete systems* that transform a discrete signal into another discrete signal (Figure 1.4)

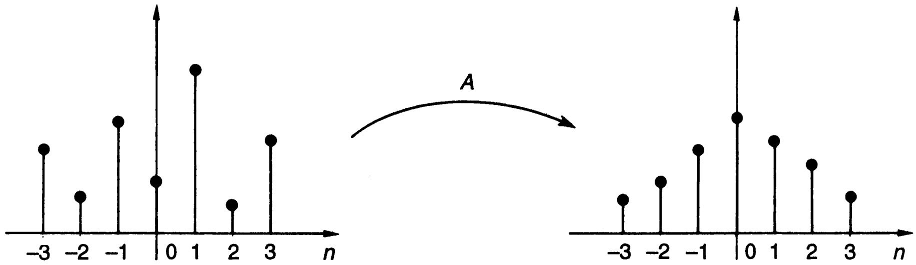

FIGURE 1.4. Discrete system.

One can go from a discrete signal to an analog signal, or conversely, using *converters* that are called *hybrid systems*:

- An analog-to-digital converter, like a sampler, for example;
- A digital-to-analog converter, which produces an analog signal from a digital signal. We mention as an example the *clamper*, or clamping circuit (Figure 1.5). This device yields the last value of the digital signal until the point when the next value arrives.

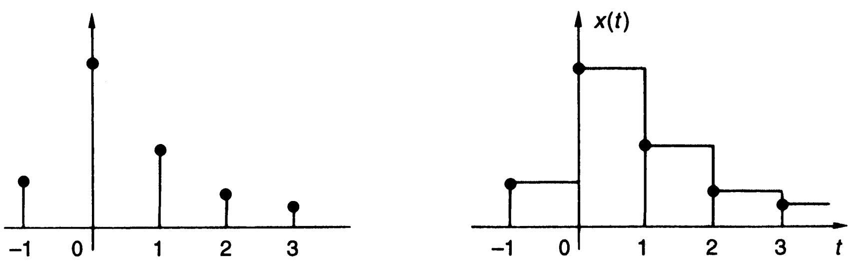

FIGURE 1.5. The clamper.

1.2 Some elementary signals

1.2.1 The Heaviside function

The Heaviside function is the signal, denoted throughout the book by $u(t)$, defined by

$$u(t) = \begin{cases} 0 & \text{if} \quad t < 0, \\ 1 & \text{if} \quad t > 0. \end{cases}$$

(See Figure 1.6.) The value at $t = 0$ can be specified or not. This value is not important for integration. The Heaviside signal models the instantaneous establishment of a steady state.

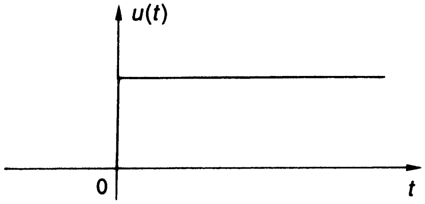

FIGURE 1.6. The Heaviside function.

1.2.2 A rectangular window

The (centered) rectangular signal $r(t)$ (Figure 1.7) is defined, for $a > 0$, by

$$r(t) = \begin{cases} 1 & \text{if} \quad |t| < a, \\ 0 & \text{if} \quad |t| > a. \end{cases}$$

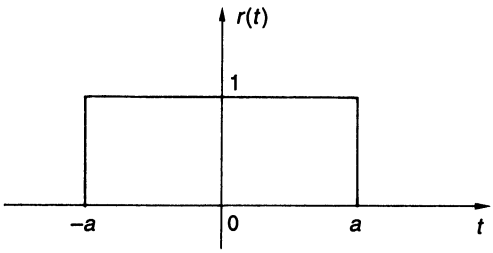

FIGURE 1.7. A rectangular window.

1.2.3 A pure sinusoidal, or monochromatic, signal

A sinusoidal signal is of the form

$$x(t) = \alpha \cos(\omega t + \varphi),$$

where the parameters have the following interpretations:

$$\begin{aligned}
|\alpha| &= \max|x(t)| \text{ is the amplitude of the signal;}\\
\omega &= \text{the angular rate;}\\
a &= 2\pi/\omega \text{ is the (smallest) period;}\\
\lambda &= 1/a \text{ is the frequency;}\\
\varphi &= \text{the initial phase.}
\end{aligned}$$

Signal values are, in principle, real numbers, and the frequency is a positive number. However, for reasons of convenience (Fresnel representation, derivation, multiplication, ...) a complex-valued function

$$z(t) = \alpha e^{i(\omega t+\varphi)}$$

is often used, and one has

$$x(t) = \mathrm{Re}(z(t)) = \frac{1}{2}(z(t) + \overline{z}(t)),$$

where $\overline{z}$ is the complex conjugate of z. Writing the signal this way involves negative frequencies, which make no sense physically. Nevertheless, this is a useful convention. It is always understood that the frequencies of opposite sign will be combined to reproduce the real signal.

A signal of the form $z(t) = ce^{2i\pi\lambda t}$, where $c = |c|e^{i\varphi}$ and where $\lambda \in \mathbb{R}$ — but where c is can be complex and can thus include a phase φ — is often represented by plotting the modulus and argument of c in frequency space, that is, as a function of the frequency λ. This is illustrated in Figures 7.1 and 7.2 for more general functions of the form

$$z(t) = \sum_{n=-\infty}^{\infty} c_n e^{2i\pi\lambda_n t}.$$

1.3 Examples of systems

1.3.1 Ideal amplifier

$$y(t) = kx(t), \text{ where } k \text{ is a fixed constant.}$$

1.3.2 Delay line

$$y(t) = x(t-a), \text{ where } a \text{ is a real constant.}$$

1.3.3 Differentiator

$$y(t) = x'(t), \text{ where } x' \text{ is the derivative of } x.$$

Lesson 2

Filters and Transfer Functions

Systems have properties, at least sometimes. We are going to review several of the more standard properties of systems.

2.1 Algebraic properties of systems

The set of input signals X and the set of output signals Y are assumed to be vector spaces (real or complex). A system A can have several properties:

2.1.1 Linearity

Consider the system

$$A : X \to Y.$$

A is said to be linear if

$$A(x + u) = A(x) + A(u)$$

and

$$A(\lambda x) = \lambda A(x)$$

for all $x, u \in X$ and all $\lambda \in \mathbb{R}$ (or $\mathbb{C}$ if X is complex). This is also called the *principle of superposition.* The systems in Section 1.3 are all linear, which is easily verified by examining the governing equations (1.1) through (1.4).

2.1.2 Causality

A is said to be *realizable* (or *causal*) if the equality of any two input signals up to time $t = t_0$ implies the equality of the two output signals at least to time t_0:

$$x_1(t) = x_2(t) \text{ for } t < t_0 \quad \Longrightarrow \quad Ax_1(t) = Ax_2(t) \text{ for } t < t_0.$$

This property is completely natural for a physical system in which the variable is time. It says that the response at time t depends only on what has happened before t. In particular, the system does not respond before there is an input. Thus causality is a necessary condition for the system to be physically realizable.

2.1.3 Invariance

A is said to be *invariant*, or *stationary*, if a translation in time of the input leads to the same translation of the output; that is,

$$x(t) \to y(t) \quad \Longrightarrow \quad x(t-a) \to y(t-a).$$

Let τ_a be the delay operator defined by

$$\tau_a x(t) = x(t-a).$$

If the system A is invariant, then

$$A(\tau_a x) = \tau_a(Ax)$$

for all $x \in X$ and $a \in \mathbb{R}$. Thus, for all $a \in \mathbb{R}$,

$$A\tau_a = \tau_a A,$$

which says that A commutes with all translations. For discrete systems, one considers only a that are multiples of the sampling interval.

2.2 Continuity of a system

The system

$$A : X \to Y$$

is said to be *continuous* if the sequence Ax_n $(= y_n)$ tends to Ax $(= y)$ when the sequence x_n tends to x. *This concept assumes that there exists some notion of sequential limit for signals in both X and Y.*

Continuity is a natural hypothesis; it expresses that idea that if two input signals are close, then the output signals are also close.

2.2.1 The analog case

When the signals are functions, the notion of limit is often defined in terms of a *norm* $\|\cdot\|$ defined on each of the vector spaces X and Y. In this case

$$x_n \to x \quad \text{means that} \quad \|x_n - x\| \to 0.$$

These are the three most frequently used norms:

(i) The norm for *uniform convergence*:

$$\|x\|_\infty = \sup_{t \in I} |x(t)|.$$

(ii) The norm for *mean convergence*:

$$\|x\|_1 = \int_I |x(t)|\, dt.$$

(iii) The norm for *convergence "in energy"* (*mean quadratic convergence*):

$$\|x\|_2 = \Big(\int_I |x(t)|^2\, dt \Big)^{1/2}.$$

In all cases, I is the interval of interest.

This last norm has the advantage over the other two of being derived from a *scalar product*,

$$(x, y) = \int_I x(t)\overline{y}(t)\, dt,$$

where $\overline{y}(t)$ denotes the complex conjugate of $y(t)$. Thus, $\|x\|_2 = \sqrt{(x, x)}$. Such a structure allows one to introduce the notion of orthogonality between two signals. This generalizes the concept of orthogonality in $\mathbb{R}^n$ and is expressed by the relation

$$(x, y) = 0.$$

One often uses a less restrictive notion of continuity, namely, *continuity in the sense of distributions.* This concept will be studied Chapter VIII.

2.2.2 The discrete case

When the signals are discrete, one can use the analogous norms:

$$\|x\|_\infty = \sup_{n \in \mathbb{Z}} |x_n|; \quad \|x\|_1 = \sum_{n=-\infty}^{+\infty} |x_n|; \quad \|x\|_2 = \Big(\sum_{n=-\infty}^{+\infty} |x_n|^2 \Big)^{1/2}.$$

The *simple convergence* of a sequence of signals,

$$x_n = (x_{nk})_{k \in \mathbb{Z}},$$

is also used. By this we mean that the sequence x_n tends to x if the limit exists for each of the components:

$$x_n \to x \quad \Longleftrightarrow \quad x_{nk} \to x_k \text{ for each } k \in \mathbb{Z} \text{ as } n \to +\infty.$$

Example: A differentiator is not a continuous system in the uniform convergence norm. Indeed, if we take $x_n = (1/n)\sin(nt)$, then $x_n \to 0$ uniformly in t. But $y_n(t) = x_n'(t) = \cos(nt)$ does not tend to zero as $n \to +\infty$. On the other hand, we will show that the integrator

$$y(t) = \int_{-\infty}^{t} e^{-(t-s)} x(s)\, ds$$

is continuous with respect to uniform convergence.

2.3 The filter and its transfer function

The term *filter* refers both to a physical system having certain properties and to its mathematical model defined in terms of the following objects:

(i) two vector spaces X and Y of input and output signals, respectively, that are endowed with a notion of convergence;

(ii) a linear operator $A : X \mapsto Y$ that is continuous and translation-invariant.

We will say informally that *a filter is a continuous, translation-invariant linear system.*

Such a system satisfies the *principle of superposition*, which is another name for linearity. Thus,

$$A\Big(\sum_{n=0}^{k} a_n x_n\Big) = \sum_{n=0}^{k} a_n A x_n,$$

and by continuity, one can pass to the limit when the infinite sums converge:

$$A\Big(\sum_{n=0}^{+\infty} a_n x_n\Big) = \sum_{n=0}^{+\infty} a_n A x_n.$$

Later we will see that a periodic signal (under rather general conditions) can be written as an infinite sum of *monochromatic* signals in the form

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n e^{2i\pi\lambda n t}.$$

Hence, at the output of a filter we will have

$$y = Ax = \sum_{n=-\infty}^{+\infty} c_n A(e_\lambda^n), \quad \text{where} \quad e_\lambda(t) = e^{2i\pi\lambda t}.$$

It is thus sufficient to know the outputs for each of the inputs $(e_\lambda^n)_{n\in\mathbb{Z}}$ to know the image of an arbitrary periodic signal. Furthermore, it is easy to determine the image f_λ of the signal e_λ, assuming that the latter belongs to the space of input signals for the filter. Indeed, for all values of t and u,

$$e_\lambda(t+u) = e_\lambda(t)e_\lambda(u) = \tau_{-t}e_\lambda(u),$$

where t is considered to be a parameter and u to be the variable. The image of this signal is $f_\lambda(t+u)$. As a result, we see that

$$f_\lambda(t+u) = A(e_\lambda(t)e_\lambda)(u) = e_\lambda(t)f_\lambda(u)$$

for all $u \in \mathbb{R}$. Hence, for $u = 0$,

$$f_\lambda(t) = e_\lambda(t)f_\lambda(0),$$

which we write as

$$A(e_\lambda) = H(\lambda)e_\lambda, \quad \text{where} \quad H(\lambda) = f_\lambda(0).$$

This result can be expressed as follows:

2.3.1 Proposition *Assume that e_λ is an admissible input function for the filter A, which is otherwise arbitrary. Then e_λ is an eigenfunction of the filter A. That is to say, there exists a scalar function $H(\lambda)$ such that for $\lambda \in \mathbb{R}$,*

$$A(e_\lambda) = H(\lambda)e_\lambda.$$

The function $H : \mathbb{R} \mapsto \mathbb{C}$ is called the *transfer function* of the filter A. There will be many occasions in the rest of the book where we will see the essential role that this function plays in the action of a filter on the spectrum of an input signal.

2.4 A standard analog filter: the *RC* cell

We will illustrate the general ideas presented above with the *RC* circuit shown in Section 1.3.5.

2.4.1 System response

Writing the unknown function v as $v(t) = w(t)e^{-\frac{t}{RC}}$ reduces the initial equation

$$RCv'(t) + v(t) = x(t)$$

to

$$w'(t) = \frac{1}{RC}e^{\frac{t}{RC}}x(t).$$

Assuming that the input signal $x(t)$ is such that the second member is integrable on every interval $(-\infty, t)$, we have

$$w(t) = \frac{1}{RC}\int_{-\infty}^{t} e^{\frac{s}{RC}}\, x(s)\, ds + K$$

and

$$v(t) = \frac{1}{RC}\int_{-\infty}^{t} e^{-\frac{t-s}{RC}}\, x(s)\, ds + Ke^{-\frac{t}{RC}}.$$

The constant K is determined by an auxiliary condition. For example, if we assume that the response to the zero input is zero, we see that $K = 0$.

One can *define* the response of the system A to the input x to be

$$v(t) = Ax(t) = \frac{1}{RC}\int_{-\infty}^{t} e^{-\frac{t-s}{RC}}\, x(s)\, ds. \tag{2.1}$$

It is clear from this expression that A is linear, realizable, and invariant. It is also continuous, for example, in the uniform norm, since

$$|Ax(t)| \le \|x\|_\infty \frac{1}{RC}\int_{-\infty}^{t} e^{-\frac{t-s}{RC}}\, ds = \|x\|_\infty,$$

and thus

$$\|Ax\|_\infty \le \|x\|_\infty.$$

This shows that the RC cell is a filter.

2.4.2 An expression for the output

If we write

$$h(t) = \frac{1}{RC} e^{-\frac{t}{RC}}\, u(t),$$

where u is the Heaviside function, then we can express (2.1) as

$$Ax(t) = \int_{-\infty}^{+\infty} h(t-s)x(s)\, ds = (h * x)(t). \tag{2.2}$$

This operation is, by definition, the *convolution* of the two signals h and x. It is denoted by $h * x$, and we have

$$Ax = h * x.$$

In this situation, one is said to have a *convolution system*. The function h, called the *impulse response* of the system, characterizes the filter because knowing h implies that the output of the filter is known for any input x. Throughout the book, we will use h to denote the impulse response of a system. A companion notion, the response of a system to the unit step function $u(t)$, will be defined in Lesson 24.

2.4.3 The transfer function of the *RC* filter

The response to the input $x(t) = e_\lambda(t)$ is $v(t) = H(\lambda)e_\lambda(t)$. Substitution in equation (1.2) gives

$$(2i\pi\lambda RC + 1)H(\lambda)e_\lambda(t) = e_\lambda(t),$$

and we have

$$H(\lambda) = \frac{1}{1 + 2i\pi\lambda RC}.$$

We see that signals for which $|\lambda|$ is small, the low-frequency signals, are transmitted by the filter almost as if it were the identity mapping (see Figure 2.1). On the other hand, the high-frequency signals, for which $|\lambda|$ is large, are almost completely attenuated. This explains why this filter is called a *low-pass filter*. The action of the filter on different frequencies is clearly apparent from the graph of the function

$$|H(\lambda)|^2 = \frac{1}{1 + 4\pi^2\lambda^2R^2C^2},$$

which is called the *energy spectrum of the filter*. The function $|H(\lambda)|$ is called the *spectral amplitude.*

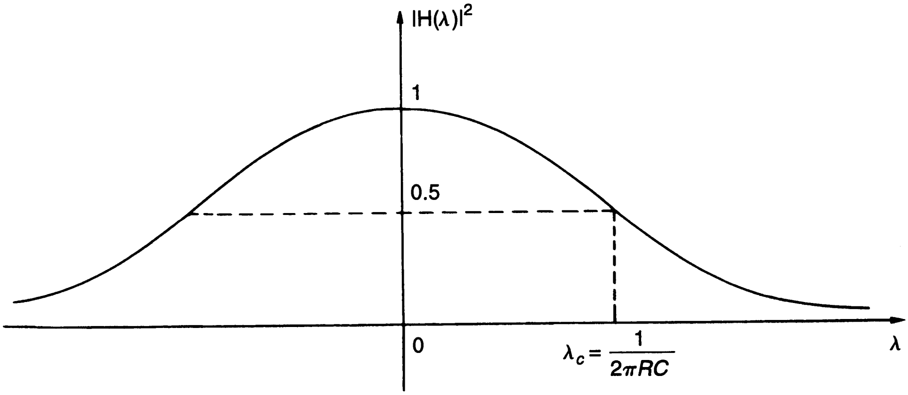

FIGURE 2.1. Energy spectrum of the low-pass *RC* filter.

The frequency $\lambda_C = 1/(2\pi RC)$, beyond which the amplitudes of the input frequencies are reduced by more than the factor $1/\sqrt{2}$, is considered to be the cutoff frequency.

We will return to the *RC* filter in Lesson 25. In fact, the analysis of this filter and of the systems described by generalizing the equation that governs the *RC* filter will be the main application of the mathematical tools that are developed in the book.