Convolution and deconvolution: two mathematical tools to help performing tests in research and industry

Jean-Pierre Fanton^{1,2,*}

¹ Electricity & Magnetism, LNE, Paris, France

² ECAM-EPMI, Cergy-Pontoise, France

Received: 26 June 2020 / Accepted: 10 February 2021

Abstract. The concepts of convolution and deconvolution are well known in the field of physical measurement. In particular, they are of interest in the field of metrology, since they can positively influence the performance of the measurement. Numerous mathematical models and computer developments dedicated to convolution and deconvolution have emerged, enabling a more efficient use of experimental data; this in sectors as different as biology, astronomy, manufacturing and energy industries. The subject finds today a new topicality because it has been made accessible to a large public for applications such as processing photographic images. The purpose of this paper is to take into account some recent evolutions such as the introduction of convolution methods in international test standards. Thus, its first part delivers a few reminders of some associated definitions. They concern linear systems properties, and integral transforms. If convolution, in most cases, does not create major calculation problems, deconvolution on the contrary is an inverse problem, and as such needs more attention. The principles of some of the methods available today are exposed. In the third part, illustrations are given on recent examples of applications, belonging to the domain of electrical energy networks and photographic enhancement.

Keywords: convolution / deconvolution / metrology / electrical networks

1 Introduction

The purpose of this paper is to discuss the two mathematical concepts of convolution and deconvolution.

Although these already appear as well-known, we observe today some reasons to reconsider them in the light of recent changes.

We shall focus on the application of these concepts in the context of technical laboratories, practising either research, or accredited conformity tests.

A significant evolution to take into account is the apparition, in international tests standards, of recommended calculation methods, which make use either of convolution or deconvolution.

Moreover, the applicability and efficiency of these methods, generally working on time signals, have to be questioned with respect to other currently used methods based on Fourier analysis.

We wish here to introduce the physical concept of convolution, before presenting the formal mathematical definition which covers it. The “phenomenon” of convolution is indeed an extremely common reality in all the branches of Physics, or more exactly in the activity of the physicist, who tries to describe reality by the means of measurements.

When an optical, electronic, mechanical observation instrument is used, there is an almost systematic deterioration of the input information, whether it is mono or multidimensional: signal, image, ... This distortion reflects the necessarily non-ideal nature of physical systems: the finite nature of these makes it impossible to transmit infinite information, hence limitations with regard to frequency; the same cause creates similar limitations in the domain of amplitudes [1–3].

The alteration of information in the frequency domain corresponds to the phenomenon of convolution of the input by an intrinsic property of the system encountered, or of the device used, which is its impulse response; this answer depends, in practice, on the physical principle, and also on the technical quality of the device. The restitution of the

* Corresponding author: jean-pierre.fanton@wanadoo.fr

Fig. 1. Symbolic diagram of a linear multiple-input multiple-output system.

initial information is not necessarily impractical, but all the more complex as there are intermediate phenomena, often of random nature such as noise.

2 Theoretical reminders

2.1 Systems, signals, fundamental properties

The concept of system initially comes from the desire to dispose of an abstract model to study, with a high degree of generality, any set of components receiving and restoring information (signals). A system can be physical or not (automatic, mechanical, optical, economical, etc.) continuous or discontinuous, as well as the quantities which constitute the associated signals (Fig. 1).

In practice, the intrinsic properties and the constitution of the system are not always known, and it is in fact characterized by the properties of the signals which make it possible to observe it.

How general this model may be, one must accept significant restrictions, so that the signal processing methods which we are going to talk about are applicable to it:

- **invariance:** the properties of the system must not change over time

$$\text{if : } e(t) \Rightarrow s(t) \text{ then : } e(t-\tau) \Rightarrow s(t-\tau) \quad (1)$$

- **linearity:** proportionality between respective variations at the input and the output of the system

$$\text{if : } e(t) \Rightarrow s(t) \text{ then : } k \cdot e(t) \Rightarrow k \cdot s(t) \quad (2)$$

2.2 Important definitions

- a) Dirac impulse or function $\delta(t)$ (Fig. 2)

It is a function (actually a distribution) defined by:

$$\begin{aligned} \delta(t) &= 0 \quad \forall t \neq 0 \\ \delta(0) &= \infty \end{aligned} \quad (3)$$

or:

$$\int_{-\infty}^{+\infty} \delta(t) \cdot \varphi(t) dt = \varphi(0) \forall \varphi \quad (4)$$

The physicist interprets $\delta(t)$ as the limit of an increasingly brief and intense pulse, knowing also that:

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1 \quad (5)$$

- b) Heaviside function, or step function $\psi(t)$

Fig. 2. Dirac impulse.

Fig. 3. Step function.

Fig. 4. Effect of a delay.

It is the integral of the previous one, much easier to approach, in practice (Fig. 3):

$$\begin{cases} \psi(t) = 0 & t < 0 \\ \psi(t) = 1 & t \geq 0 \end{cases} \quad (6)$$

- c) causal function

this can be any function such that $f(t) = 0$ for $t < 0$; $\delta(t)$ and $\psi(t)$ belong to this class, as well as any response of a physical system to an input (which the output cannot indeed precede).

The computation of certain integrals relating to causal functions can be significantly simplified by change of integration boundaries.

2.3 Convolution equation

This essential equation will make it possible to establish how the response of a system to any excitation can be determined from the response to a canonical input (impulse input or step input).

We use the properties of invariance (1) and linearity (2) to write (Fig. 4):

$$e(t) = \delta(t) \Rightarrow s(t) = h(t) \quad (7)$$

impulse response

$$e(t) = \delta(t-\tau) \Rightarrow s(t) = h(t-\tau) \quad (8)$$

$$e(t) = k \cdot \delta(t) \Rightarrow s(t) = k \cdot h(t), k \in \mathbb{R} \quad (9)$$

Fig. 5. The Duhamel integral.

we cut any input signal $e(t)$ into incremental steps of step Δt (small) (Fig. 5):

$$e(0), e(\Delta t), e(2\Delta t), \dots, e(k\Delta t), \dots$$

The additional pulse $e(k\Delta t)$ produces an additive output signal; using (8) and (9):

$$s_{k\Delta t} = e(k\Delta t) \cdot h(t - k\Delta t) \quad (10)$$

we can write the following:

$$\begin{aligned} s_0 &= e \cdot h(t) \\ s_1 &= e(\Delta t) \cdot h(t - \Delta t) \\ &\dots \dots \dots \\ &\dots \dots \dots \\ s_{k\Delta t} &= e(k\Delta t) \cdot h(t - k\Delta t) \end{aligned} \quad (11)$$

due to linearity¹:

$$s(t) = \sum_k e(k\Delta t) \cdot h(t - k\Delta t) \quad (12)$$

we go from (12) to the limit when Δt tends towards 0

$$\Delta t \rightarrow 0 \quad s(t) = \int_0^t e(\tau) \cdot h(t - \tau) d\tau \quad (13)$$

(Duhamel integral)

The corresponding equation is the convolution equation, noted:

$$s(t) = e(t) * h(t) \quad (14)$$

Mathematicians write:

$$s(t) = (e * h)(t) \quad (15)$$

Indeed, it is important to note that we do not have iteration on 2 operators, but a single operator; the calculation of a point of $s(t)$ does not depend only on

the instantaneous values $e(t)$ and $h(t)$, but on all the values of e in particular on $[0, t]$ (past of the input)¹.

Writing the reverse operation under the form:

$$e(t) = \frac{s(t)}{h(t)} \quad \text{or:} \quad h(t) = \frac{s(t)}{e(t)}$$

is therefore **entirely conventional**.

We can show that the convolution operator has the properties of commutativity, associativity, distributivity, from those of simple integrals.

In reality, the physicist's problem is rarely to calculate a convolution product, since this generally occurs "all by itself", despite himself, or even without him knowing it. It is much more generally the resolution of one of the two possible forms of the inverse problem, that is to say:

- knowing $e(t)$ and $s(t)$, calculate $h(t)$, i.e. the system response; this is the so-called **identification** problem.
- knowing correctly the system, that is to say $h(t)$ and also its output $s(t)$, go back to a non accessible input $e(t)$; this is the **deconvolution** problem.

Given the commutative nature of the operator:

$$(s(t) = e(t) * h(t)) \quad (16)$$

one might think that the two problems are identical.

In fact, the hypotheses that we know how to make on the supposed known quantity $e(t)$ or $h(t)$ lead us to consider often different methods.

The **mathematical** notion of convolution product does not refer to a particular physical nature of a variable; it has a meaning in the frequency domain as well as in the time domain. We will see that in particular certain consequences of the duality between these two quantities can make it possible to operate in the field where the calculation is most convenient.

2.4 Integral transforms

The reason for introducing this class of transforms is to show properties common to some of them, in particular with regard to convolution problems [4-6].

- **definition:** let $f(t)$ be a function, and $g(s, t)$ be an auxiliary function

The function $F(s)$ is obtained by an integral transformation on $f(t)$:

$$F(s) = \int_a^b f(t) \cdot g(s, t) dt \quad (17)$$

There are a large number of such transformations, each having an optimal field of use. They make it possible to constitute powerful mathematical tools, for example for

¹ The impulse response of a system being said "with bounded support", that is to say with null value outside a finite interval, in particular for $t < 0$, (principle of causality), we can extend this sum to $s(t) = \sum_{-\infty}^{+\infty} e(k\Delta t) \cdot h(t - k\Delta t)$.

solving ordinary differential or partial differential equations.

The Laplace and Fourier transforms, which are especially important in signal processing, belong to this class, and have certain essential properties in common.

2.4.1 Transform of a convolution

We will limit ourselves to integral transforms for which the auxiliary function $g(s, t)$ will be of exponential nature:

$$\begin{cases} F_1(\alpha) = \mathcal{T}[f_1(x)] = \int e^{-\alpha x} f_1(x) dx \\ F_2(\alpha) = \mathcal{T}[f_2(x)] = \int e^{-\alpha x} f_2(x) dx \end{cases} \quad (18)$$

we can write:

$$F_1(\alpha) \cdot F_2(\alpha) = \int f_1(v) \cdot e^{-\alpha v} dv \cdot \int f_2(u) \cdot e^{-\alpha u} du \quad (19)$$

(simple product)

we set: $w = u + v$ and $dw = du$

$$\begin{aligned} F_1 \cdot F_2 &= \int \left[\int e^{-\alpha(u+v)} f_1(w-u) \cdot f_2(u) du \right] dw \\ &= \int e^{-\alpha w} \left[\int f_1(w-u) \cdot f_2(u) du \right] dw \\ &= \int (f_1 * f_2) e^{-\alpha w} dw \\ &= \mathcal{T}(f_1 * f_2) \end{aligned} \quad (20)$$

The simple product of the transforms $F_1 \cdot F_2$ is the transform of the convolution product $f_1 * f_2$ of the functions f_1 and f_2 .

This result applies:

- to all integral transforms whose auxiliary function is exponential, therefore, as we will see, so in particular to the Laplace and Fourier transforms, by introducing in each case the integration bounds and the appropriate multiplicative coefficients.
- whatever the variables considered, in particular whether the starting variable is time or frequency; the correspondence between product and convolution is valid in both directions.

2.5 Laplace transform

This particular transformation, of very ancient introduction (Analytical theory of probabilities – 1812) played a fundamental theoretical role by facilitating the use of the notions of **transfer function** and **symbolic calculation**.

It is defined by:

$$\mathcal{L}[f(t)] = F(p) = \int_0^{+\infty} f(t) \cdot e^{-pt} dt \quad (21)$$

where p is a complex variable:

$$p \in \mathbb{C} \text{ with } p = \sigma + j\omega; \sigma, \omega \in \mathbb{R} \text{ and } \sigma > \sigma_F$$

σ_F is said convergence abscissa; indeed, for any function $f(t)$, there exists σ_F such that the Laplace Transform is holomorphic in the open half-plane $\Re(p) > \sigma_F$.

One speaks of **original** $f(t)$ and **image** $F(p)$.

note: in signal processing, p frequently has the dimension of a frequency.

2.5.1 Symbolic calculation

Symbolic computing was developed for solving electricity problems (Heaviside); it is based on intuitive rules.

Thus, to formally simplify the resolution of differential systems, we replace the derivation operator (d/dt) with a simple algebraic product.

$$\text{example: } E = Ri + L \frac{di}{dt} + \frac{1}{C} \int i dt \quad (22)$$

$$\text{becomes: } E = i \left(R + Lp + \frac{1}{Cp} \right) \quad (23)$$

This calculation technique can only partially justify itself. On the other hand, it can be totally justified, and moreover considerably extended, by means of the Laplace transform.

♦ examples:

❶ by integration by parts

$$I = \int_0^{\infty} f'(t) \cdot e^{-pt} dt = \underbrace{\left[f(t) \cdot e^{-pt} \right]_0^{\infty}}_{-f(0)} + p \underbrace{\int_0^{\infty} f(t) \cdot e^{-pt} dt}_{p \mathcal{L}[f(t)]} \quad (24)$$

$$\text{hence } \mathcal{L}[f'(t)] = p \mathcal{L}[f(t)] - f(0) \quad (25)$$

❷ by derivation under the sum sign

$$\underbrace{\frac{d}{dt} [F(p)]}_{=0 \text{ when } f(0)=0} = \underbrace{\int_0^{\infty} f'(t) \cdot e^{-pt} dt}_{\mathcal{L}[f'(t)]} - p \underbrace{\int_0^{\infty} f(t) \cdot e^{-pt} dt}_{p \mathcal{L}[f(t)]} \quad (26)$$

$$\text{hence } \mathcal{L}[f'(t)] = p \mathcal{L}[f(t)] \quad \text{when } f(0)=0$$

$$\text{❸ } \frac{d}{dp} [F(p)] = -t \int_0^{\infty} f(t) \cdot e^{-pt} dt \quad (27)$$

$$= -t F(p) = -\mathcal{L}[t f(t)]$$

$$\text{(duality) (duality)}$$

etc.

Fig. 6. Time signals.

Fig. 7. Complex frequency signals.

2.5.2 Transfer function

The transfer function is the Laplace transform of the impulse response of a system (Figs. 6 and 7).

The transition from the time formulation (convolution) to the operational formulation (algebraic product) makes it possible to considerably simplify the theoretical study of systems, in particular for the study of problems of:

- filtering
- closed-loop control systems
- **Filtering** is a (common) operation consisting in operating a convolution by a function whose transformation in the dual domain is often simple, for example of the window type. Thus frequency filtering (low pass, high pass, band pass, or other) is equivalent to a time convolution.

In the same way, the time windowing of a signal, necessary to carry out its digital analysis on a finite time sample, is equivalent to a frequency convolution.

By the way, it can be seen that deconvolution may in particular consist in de-filtering, that is to say filtering by a function with opposite properties to that of the system which one has passed through.

We intuitively understand the limits of such an approach (noise amplification, etc.)

- by the functional decomposition of the transfer functions that it allows, Laplace formalism particularly facilitates the analysis of closed-loop systems (for stability, etc.)

2.6 Fourier transform

This transformation, like the Laplace transform, is very anciently known (Analytical heat theory – 1822); it still constitutes today an irreplaceable element of the spectral description of the signals. It finds itself its origin in the Fourier series.

2.6.1 Series and Fourier coefficients

The basic idea is to decompose a periodic function, on the basis of simple functions, orthogonal with each other, such

as complex sinusoidal or exponential functions (with in view to solve these equations by decomposition) for example:

$$f(x) = a_0 + a_1 \cos x + \dots + a_n \cos nx + \dots + b_1 \sin x + \dots + b_n \sin nx + \dots$$

or also:

$$f(x) = c_0 + c_1 e^{+ix} + \dots + c_n e^{+inx} + \dots + c_{-1} e^{-ix} + \dots + c_{-n} e^{-inx} + \dots$$

the period here being $T_0 = 2\pi$.

The formula of the average allows to write, for any function:

$$\overline{f(x)} = \frac{\int_a^b f(x) dx}{b - a} \quad (30)$$

This formula makes it possible to determine the Fourier coefficients.

note: if f is a real function, $c_{-n} = c_n$

2.6.2 Fourier transform

This transform is obtained when the period τ of the function is indefinitely increased; c_n becomes a continuous function $C(n)$; the discrete sum in equation ③ becomes an integral:

$$\tau \rightarrow \infty \begin{cases} C(n) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) \cdot e^{-inx} dx & (31) \\ f(x) = \int_{-\infty}^{+\infty} C(n) \cdot e^{+inx} dn & (32) \end{cases}$$

(31) and (32) respectively define the direct and the reverse Fourier transforms.

We immediately note a certain number of characteristics of these transforms:

- a formal analogy with the Laplace transform, despite the different integration limits
- the analogy between the two transforms themselves
- the possibility of limiting the calculation, in general to the one of real integrals for the Fourier transform

All these properties are used on the level of numerical computation, to render the algorithms efficient (Fast Fourier Transform or F.F.T.); also to re-use them in multiple ways (direct transform + inverse transform), and to apply them to the problem of deconvolution.

2.6.3 Reverse Laplace transform

(or Mellin-Fourier transform)

The expression of the reverse Laplace transform is deduced from that of the reverse Fourier transform,

The reverse Laplace transform is written:

$$\mathcal{L}^{-1}[F(p)] = f(t) = \frac{1}{2\pi i} \int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} F(p) \cdot e^{+pt} dt \quad (33)$$

It is a complex integral. Its analytical calculation is only possible in a very limited number of cases.

For any function $E(p)$, there does not necessarily exist a function $e(t)$ which corresponds to it:

$$E(p) = \frac{S(p)}{H(p)} \Rightarrow e(t) = ?$$

2.7 Laplace and Fourier transforms; relationships, comparisons

2.7.1 Relationships

The Laplace and Fourier transform are in practice two distinct particular cases of the same integral transformation, the bilateral Laplace transform:

$$X(p) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-pt} dt \quad (34)$$

holomorphic in a vertical band of the complex plane $\sigma_1 < \Re(p) < \sigma_2$

- a)
- for the causal signals, $X(p)$ is reduced to the usual Laplace transform
- for non-causal signals:

$$X(p) = [\mathcal{L}(x)](p) + [\mathcal{L}(\underline{x})](-p) \text{ with } \underline{x}(t) = x(-t) \quad (35)$$

- b) if $f(t)$ is a causal function:

$$\mathcal{L}[f(t)] = \int_0^{\infty} f(t) \cdot e^{-\sigma t} \cdot e^{-i\omega t} dt = \mathcal{F}[e^{-\sigma t} \cdot f(t)] \quad (36)$$

if $\sigma = 0$, we have: $p = i\omega$; the analytical expressions of the transforms are formally the same.

2.7.2 Comparison

The Laplace and Fourier transforms are two tools whose development has been contemporary to each other, for the study of linear systems. The Fourier transform has for long been little used, for theoretical reasons: the pure imaginary exponential, unlike that of the Laplace transform, does not let the integrals converge.

The Laplace transform has been widely used by electricians for the study of transient circuit conditions; then electronics developed, and Fourier transform then became generalized, for several reasons:

- ease of calculation by discretized elementary operations (summations/products) $\Rightarrow$ Discrete Fourier Transform (D.F.T.), or even by accelerated algorithms Fast Fourier Transform (F.F.T.)
- the great faculty of **data representation** that it authorizes, for the properties of signals and systems: by switching to the auto-spectrum or to the power cross-spectrum and plotting in module/phase form (Bode diagram, Nyquist diagram), possible use in

three-dimensional (time/frequency, etc.). The Laplace transform on the contrary cannot be so easily represented (it is a holomorphic function of a complex variable)

note: when, with a usual signal analyzer (“F.F.T. analyser”), we make the quotient of the power spectra of the output by the input of a system, we do NOT obtain the transfer function $H(p)$, but simply the “frequency response” $H(j\omega)$, from which we can however deduce $H(p)$.

- finally, in practice and at least qualitatively, the Fourier transform allows frequency analysis of all types of signals, including transients.

The Laplace Transform will nevertheless remain a precious tool, at least for theoretical calculations; in addition to its convenience of use (for electrical and automatic control engineers, if necessary in a form more suited to digital signals, the Z transform), it offers, contrary to the Fourier Transform, the advantage that convergence can always be obtained.

Recent technical applications will be presented below. They show the topicality of the Laplace transform.

2.8 Deconvolution

2.8.1 Position of the problem

Without further description, in general terms, the properties and the extended applications of Fourier and Laplace transforms, we will examine their possible use in the field of deconvolution.

Deconvolution is the reverse of convolution. It belongs to inverse problems, and as such, leads to much more difficulties than its associated direct problem [4-6].

Let us remember that a physical signal that passes through a measuring instrument can be altered in many ways; it can be: filtered, delayed, clipped, altered with noise, ...

Deconvolution only corrects the effects of linear disturbances, may they be certain or random. This leads to say that there is an increase of the observation bandwidth.

The presence of noise in the signal reduces the accuracy of deconvolution. In some circumstances, the knowledge of some *a priori* information on the noise may allow to limit this degradation.

In general, deconvolution is par excellence an operation which can take advantage of possibilities of digital signal processing, and for which methods have recently developed and diversified: use of Kalman filtering, Wiener filtering, Mendel estimator, ...

Unlike **identification**, where we can have a large number of signal pairs $e(t)/s(t)$, and where we thus minimize estimation errors, we only have in deconvolution one signal $s(t)$, and the expression of $h(t)$ or $H(v)$.

Deconvolution is therefore not a systematically possible operation; it can, if H is not “favourable” (position of the poles in the complex plane, very small module, ...) be related to the class of so-called ill-posed problems, i.e. very sensitive to small errors on the input ; this *a fortiori* if the input is altered by a significant noise.

3 Particular methods

3.1 Frequency methods

These methods, which take advantage of the time/frequency duality, and therefore product/convolution duality, lead to widely used and very convenient solutions:

$$x(t) = \int \frac{Y(v)}{H(v)} \cdot e^{+2\pi i v t} dv \quad (37)$$

the ratio (Y/H) must be integrable, or at least of summable square on $\mathbb{R}$

$$x(t) = \lim x_n(t) = \lim \int_{-n}^{+n} \frac{Y(v)}{H(v)} \cdot e^{+2\pi i v t} dv \quad (38)$$

It is in this form that we frequently operate from digital signal analyzers:

$$\boxed{\text{F.T.}} \quad \boxed{\text{F.T.}^{-1}} \\ y(t) \rightarrow Y(v) \quad \Leftrightarrow \quad y(v)/H(v) \rightarrow x(t)$$

Despite the apparent ease of the process, the basic problems already mentioned remain, and should be used with adequate precautions or reservations about results validity.

3.2 Real time calculation

Reference [7] explains that most deconvolution problems can be expressed under the form of a linear system

$$[y] = [H] \cdot [x] + [b] \quad (39)$$

where $[y]$, $[H]$, $[x]$, and $[b]$ are matrices with the following meanings: $[x]$ is the unknown signal, $[y]$ is a vector of observations, $[H]$ is a known matrix, and $[b]$ a vector of random errors.

As one knows, methods for solving linear systems are many. But a new difficulty occurs when one has to carry out a **real-time** deconvolution.

The constraint in this case is not so much the fact that a fast calculation is needed, but the fact that we ignore a part of the signal (its “future”), that we could use in deferred time for more accurate results.

Thus reference [7] describes a method for raising this difficulty. This method is based on a Kalman filter approach, it provides a so-called “suboptimal estimator”, to accelerate the convergence of the classical Kaczmarz algorithm; but it also needs, for satisfactory results, that the noise and input signal have stationary properties.

It is also possible to re-introduce the signal resulting from a deconvolution into a convolution product with the system impulse response, and then compare the computed output with the true output. A deconvolution problem is then turned into an **optimisation** problem.

Reduction of non-deterministic disturbances, if any, may be only obtained through correlation techniques.

4 Applications

In this paragraph, we shall first describe some typical applications belonging to the domain of electrical energy networks.

The study of high voltage or high current signals can indeed make an efficient use of both convolution and deconvolution techniques.

4.1 The context

In the area of electrical networks, many types of technical problems can arise.

First, one can have to deal with very high voltages or currents values, which can render experimentation unpractical or even dangerous; hence the need to proceed to preliminary calculations to validate the proper behaviour of a newly designed equipment.

Secondly, many transient phenomena are likely to occur on networks. Their origin can be either natural (severe weather, thunderstorm, ...), technical (source switching, line switching, short-circuits, overloads, ...) or even human (false manoeuvre, error, ...).

Possible consequences of the transients are also to be studied in advance.

Many books have already thoroughly explored these subjects [8–10].

4.1.1 The needs

As said above, needs in the field of electrical networks are twofold; it may refer to either normal or accidental situations

- predictive analyzes of the behaviour of installations and equipment under specified stresses
- experimental validation of this behaviour in the purpose of equipment qualification or certification

These validations can be the subject of preliminary calculations, so as to predetermine an impulse or a step response, or even to a particular characteristic wave, such as a lightning strike.

The qualification of equipments results from the validation of their compliance to requirements defined by international standards. These equipments may be very diverse: appliances, circuit breakers, network components, ...

Testing laboratories, working in the framework of accreditation, are currently in charge of proceeding to such validations. They perform the experimental tests prescribed by the standards and establish official test reports.

Although the theoretical response of an equipment under test can also be predicted by a calculation, only the experimental results are considered in the framework of accreditation as a proof of the compliance.

From the metrological point of view, the need is generally not for an extreme measurement accuracy, but for the reliability of the results, which may require that the uncertainties attached to the measurement be precisely determined.

Actual stresses on networks are often of random nature. Their study can then be approached by simulations, whose hypotheses and statistical models are sometimes codified by international standards.

Standards corresponding to our examples below will be indicated.

4.1.2 The possibilities

Both methods, convolution and deconvolution, can be of valuable efficiency to operate predictive analyses in the domain of energy.

In fact, the convolution/deconvolution operations carry out a filtering of a particular type, better suited to the study of transients than the one achievable by Fourier transform: the Fourier transform is carried out on digital time signals over a necessarily finite duration (hence the effect of time windowing).

It is important to remember that both transforms – L.T. and F.T. – can only be applied to linear systems; and especially in the field of networks, they cannot apply to the so-called ferroresonance phenomenon, caused by the non-linearity of magnetic origin presented by some equipments such as transformers.

4.2 The cases

The different examples that we will examine are:

- the test of a switcher of electrical sources under a rated R/L (resistive/inductive) load
- the correction by deconvolution of the response of a voltage divider
- an example of propagation of a transient wave along an overhead power line

4.2.1 Endurance test of electrical equipment

Let us consider the case of a particular industrial product: a source switcher, transferring a specified R-L electrical load from one energy source to another. The test may be defined as more or less severe, depending on the power factor of the load (L/R ratio).

This type of test is described for instance by the international standard IEC 60947-6-1.

The source provides a sinusoidal voltage wave at industrial frequency f ; so here $\omega = 2\pi f = 314 \text{ rd/s}$, the load time constant $\tau = L/R$ and the power factor $\cos \varphi$ are closely linked together: $\varphi = \arctan(\omega\tau)$ (Fig. 8)

The load has an analytical definition in the complex frequency domain

$$Y = \frac{1}{Z} = \frac{1}{L} \cdot \frac{\tau}{1 + \tau p} \quad (40)$$

The following figure represents the predicted response to the transient (R and L are known), which can be compared with the corresponding experimental results (Fig. 9).

Fig. 8. General provisions for the endurance test of a switching device S.

Fig. 9. Intensity $i(t)$ time signal during the test.

The voltage transient e applied to the switcher has for expression

$$e = E_0 \frac{\omega \cos(\theta - \varphi) + p \sin(\theta - \varphi)}{\omega^2 + p^2} \left(1 - e^{-\frac{2k\pi p}{\omega}}\right) \quad (41)$$

in this expression of the voltage e :

– the factor $\omega/(\omega^2 + p^2)$

represents the sinusoidal source waveform

– the factor $\omega \cos(\theta - \varphi) + p \sin(\theta - \varphi)$

represents the phase lag at the time of switching

– the factor $(1 - e^{-2k\pi p/\omega})$

represents the signal windowing in the time domain (k periods)

The current intensity $I(p)$ is the product of the two quantities $e(p)$ and $Y(p)$ in the complex frequency domain, and thus the time image of this current $i(t)$ may be calculated by deconvolution.

What is made apparent in Figure 9 is the impact of switching conditions on the resulting overvoltage. The load power factor has been given here for visibility an unusually low value, $\cos \varphi = 0.1$

4.2.1.1 Algorithm implementation

The calculation here has been made by use of an ancient but efficient deconvolution algorithm, described in detail by [11].

This algorithm is quite simple in itself, but requires a high accuracy of intermediate calculations.

It determines the numerical time response $f(t)$ by inversion of an analytical function $F(p)$

The algorithm is based on the Crump algorithm using Fourier series, but strongly simplifies it.

This has been made possible after many numerical experimentations. Hence a formula to compute separately each point in the time domain:

$$f(t) = \frac{e^{aT}}{T} \left[\frac{1}{2} F(a) + \sum_{k=1}^{\infty} \rho \cos \left(\theta + \frac{k\pi t}{T} \right) \right] \quad (42)$$

with

$$\rho_{[\theta]} = F \left(a + j \frac{k\pi}{T} \right) \quad (43)$$

In this formula, a and T are defined experimentally from constants by $a = 9/t$ and $T = 1.5 t$

The algorithm has been validated by both comparisons on elementary cases for which the analytical solution is known, and by comparisons with other software results, on more complex test cases.

In the present example, the algorithm has been implemented in PYTHON 3 by the author.

This last language presents significant advantages on former programming languages, and in particular:

- numerical values, either complex or real, are represented in 64-bit format, which provides the required accuracy ($2^{64} \sim 10^{20}$); this corresponds, in integer format, to 20 significant figures.
- although it might also be compiled, with CYTHON or equivalent, PYTHON is basically an interpreted language, which allows easy bug fixing and parameter adjustment.

4.2.2 Calibration errors determination based on standards

International standards define models of wave shapes for the various types of voltage impulses, such as lightning impulses.

These wave shapes are based on statistical models.

So does for example IEC 60230.

Wave shapes are often based on a combination of two exponential functions (whose exponents are sometimes difficult to find out)

Typical couples of values for rise time t_r and fall time t_f according to Figure 10 are $0.84/50 \mu s$, $1.25/50 \mu s$, and $1.56/50 \mu s$.

To perform the tests on High Voltage pulses, the test laboratories include in their equipment some specific devices such as voltage dividers.

A typical voltage divider may have a high division ratio such as 100, i.e. 1000 V output for 100 kV input (Fig. 11).

Fig. 10. Sample of a lightning standardized wave shape.

Fig. 11. Stand test for the calibration of a High Voltage divider; with permission of GE Grid Solutions – Cerda (France).

It is frequently not possible to calibrate a divider on its whole measurement range, but only for the lower voltages, hence a need for extrapolation.

The international standard IEC 60060, describes – without imposing it – a method based on convolution to determine the measurement errors due to equipments like dividers.

The divider response $v_{out}(t)$ to a given standardised input signal $v_{in}(t)$ is determined using its experimental step response $g(t)$.

The calculation process is the convolution

$$v_{out}(t) = \int_0^t v'_{in}(\tau) \cdot g(t - \tau) d\tau \quad (44)$$

The step response $g(t)$ must be discretised with enough accuracy, whereas the input signal $v_{in}(t)$, having an analytical definition, may be as accurate as necessary.

Why do we use here the first derivative $v'_{in}(t)$ instead of $v_{in}(t)$?

The step response $g(t)$ is much easier to measure than the associated impulse response $h(t)$.

In the complex frequency domain, it is equivalent to write

$$[p \cdot V_{int}(p)] \cdot G(p) = V_{int}(p) \cdot [p \cdot G(p)] = V_{int}(p) \cdot H(p) \quad (45)$$

i.e. in the time domain:

$$v'_{int}(t) * g(t) \equiv v_{int}(t) * g'(t) = v_{int}(t) * h(t) \quad (46)$$

this is exposed as property ❶ of the Laplace Transform, at Section 2.5.1 above.

The standard IEC 60060 has been applied by Laboratoire National d'Essais (LNE) – France, to determine dividers calibration errors, on the basis of three parameters, namely the voltage peak value V_p , the rise time t_r , and the fall time t_f , as defined above on standardised wave shapes for lightning shocks.

Tests results showed that in all cases the errors remain inferior to 1%

The same tests were executed for switching impulses (250/2500 μ s) and led to equivalent error rates [12].

4.2.3 Wave propagation on an overhead line

A third example of a case where deconvolution can provide fast and reliable information is the one of overhead electrical line resonance.

In large electrical networks, the frequency f of the transmitted energy is generally low, such as 50 or 60 Hz. Networks lines and components are, generally speaking, protected against overvoltages by dedicated devices such as surge arresters.

Propagation phenomena on the lines may result either of natural or artificial transients. These are rare, they nevertheless have to be anticipated especially for the case of long lines. “Long” means here that the line length L is not negligible compared to the wavelength λ associated with the network frequency f .

The determination of the electrical regime of an overhead line during a transient is a complex problem. The goal of our deconvolution computation of the line voltage waveform is here to make a predictive analysis. This in order to insure that there is no unacceptable danger in any point of the line during a possible transient.

A voltage transient on a line may be caused by the switching of a source at one line end.

Fig. 12. Voltage and intensity wave reflections along a line after a transient shock.

This transient can give birth to a travelling wave which will provoke reflections on impedance discontinuities and especially on the extremities of the line. In particular circumstances, unlikely to happen in general, the reflections can result in a resonance phenomenon.

This problem has been studied with the algorithm mentioned above [13].

We make use of a line model, based on the length unit properties of the line:

- longitudinal resistance r
- longitudinal inductance l
- transverse conductance g
- transverse capacitance c (Fig. 12)

the expression of the voltage at any coordinate x is

$$V(p, x) = e_0 \frac{\exp(n \cdot x) + \exp(-n \cdot x)}{\exp(n \cdot L) + \exp(-n \cdot L)} \quad (47)$$

with

$$n = \sqrt{(r + lp) \cdot (g + cp)} \quad (48)$$

The most severe case would correspond to a lossless line, i.e. to $r = g = 0$

Figure 13 below presents a calculation result for a particular theoretical case, also unfavourable, but nevertheless more realistic than the lossless line; this case is called “no deformation line”, it includes some damping due to resistive losses along the line.

It is defined by

$$r/l = g/c = v \quad (49)$$

with

$$n = (p + v)\sqrt{lc} \quad (50)$$

The length unit properties of the line are here (all parameters in S.I.):

$r = 9.e-5$, $l = 3.00e-6$, $g = 1.e-9$, $c = 3.33e-11$
line length $L = 500000$.

$$\omega = 314 \text{ rad/s}$$

Fig. 13. Calculated resonance voltage at the end of the line.

The quantity $\tau = L\sqrt{lc}$ is the wave journey time for length L ; a round trip of the wave corresponds to $2L$, thus in time to 2τ . In our conditions, τ has a value close to $T/4$, period of the network frequency, the effects of the reflections will be additive ($2\tau \sim T/2$).

Finally, the calculation process thus allows to predetermine the possible overvoltage at any point of the line, and allows all relevant analyses.

4.3 Graphic applications

Beyond 1D signals, deconvolution can also be efficiently applied in the domain of 2D or 3D signals. This has been used for long in the fields of microscopy and astronomy.

Reference [14] compares two main classes of methods used in 2D deconvolution, i.e. filtering methods based on the application of the Fourier transform, and variational methods.

In the first class, the Wiener filter is the most commonly used. The numerical tool associated is the Fast Fourier Transform (or F.F.T.). The variational methods aim at minimising an adequate energetic function. Different numerical schemes are used in this framework (explicit schemes and stabilised explicit schemes) as methods for gradient descent.

This comparison of the diverse methods concludes that filtering methods are superior to variational methods, but this in fact under particular conditions, so that finally both method classes are satisfactory.

Today performing software modules are available at affordable conditions, and have thus widely extended the use of deconvolution techniques to large audiences researching photographic enhancement, by means of de-noising or de-blurring functions (Fig. 14).

Results obtained by deconvolution in this field are frequently much better than those obtained by conventional sharpening algorithms.

Fig. 14. Image of Noisette, the author's cat, voluntarily blurred with Gaussian noise, and de-blurred by deconvolution.

According to reference [15], the need for deconvolution in the field of image arises from a major problem which is diffraction from optical origin. High spatial frequencies correspond to the finest details to observe.

STimulated Emission Depletion microscopy (STED) can be a remedy to this problem.

In the STED technique, the noise level is unfortunately high due to the weak number of photons per pixel. An additional constraint in the field of biology is that the observation technique should not be invasive (light up of the object) since the observed object may be alive (cell, ...)

The spatial resolution of an optical microscope is currently around 200 nm, whereas the smallest biological objects to observe are only a few nm or tens of nm. The improvement in resolution obtainable with a STED software can be as high as 8 times.

So the process used consists in blurring the observed object with noise, and then correct the effect by deconvolution. In fact the perturbations of the pictures are twofold: there are at the same time blurring (deterministic) and noise (stochastic). The deconvolution hence needs the use of advanced algorithms.

An iterative process takes thus place, where the measured object is compared to an estimate. From the comparison, a new image is computed by convolution with a function called PSF (Point Spread Function); and then again compared with the estimate until the difference reaches a given criterion.

Today the STED microscopy has proven to be a valuable image enhancement technique, allowing to overcome the theoretical diffraction limit. It also enables researchers to widen the field and methods of their experiments, by reducing of the impact of the observation [15]

5 Conclusion

We presented in this paper a reminder about integral transforms, especially Fourier and Laplace transforms, which we compared. Some deconvolution methods were presented, and then a few application examples.

Deconvolution methods can contribute to improve the metrological performance of measurements. But it is also important to remember some of their limitations

- they may correctly restore certain properties of the input signals, but they do not correct all alterations; the reconstruction of useful information located far outside the useful frequency band, therefore containing minimal or zero energy, can only be defective.
- the elimination of the parasitic signals located inside the useful frequency band may be undertaken using correlation methods
- these methods cannot be efficient when signals concern non-linear phenomena, and in particular in the domain of electrical networks they cannot be used to study ferroresonance phenomena

Finally, even though new efficient processing methods appear, and despite their extended possibilities, it always remains safer whenever possible, to take great care of the information collect phase, i.e. to the measurement techniques. used, to an adequate sensors choice, and to the limitation of undesirable noise sources.

Nomenclature

t	Time
$s(t)$	Time-dependent signal
f	Frequency ($f = \omega/2\pi$)
p	Operational variable; $p = \sigma + j\omega$
$H(p)$	Transfer function of a system
$\delta(t)$	Dirac impulse
$\psi(t)$	Heaviside step function
$*$	Convolution operator
$\cdot$	Algebraic product operator
$\mathbb{R}$	Set of reals
$\mathbb{C}$	Set of complexes
$\Re, \Im$	Real and imaginary parts of a complex variable
$\mathcal{I}()$	Integral transform
$\mathcal{L}()$	Laplace transform, or L.T.
$\mathcal{F}()$	Fourier transform, or F.T.

References

1. R.G. Lyons, *Understanding digital signal processing*, 3rd Edition (Prentice Hall, 2011)
2. L.R. Rabiner, B. Gold, *Theory and application of digital signal processing* (Prentice Hall, 1975)
3. A.V. Oppenheim, R.W. Schaffer, *Digital signal processing* (Pearson, 2015)
4. J.H. Heinbockel, *Mathematics reference book for scientists and engineers* (2009)
5. K.F. Riley, M.P. Hobson, S.J. Bence, *Mathematical methods for physics and engineering* (Cambridge University Press, 2006)
6. J. Bird, *Higher engineering mathematics: for scientists and engineers* (2020)
7. G. Demoment, A. Segalen, *Rapid suboptimal estimation for real-time deconvolution*, (Gretsi Nice, 1983), pp. 205–210
8. A. Greenwood, *Electrical Transients in power systems*, ed. (Wiley, 1991)
9. L. Van der Sluis, *Transients in power systems*, ed. (Wiley, 2001)
10. R. Rudenberg, *Transient performance of electric power systems*, ed. (MIT Press, 1969)
11. P. Johannet et al., *Numerical inverse laplace transform: an efficient algorithm* (Bulletin de la DER EDF série B, déc. 1986), pp. 5–35
12. M. Agazar et al., *New reference systems for the calibration of HV impulses at LNE*, 19th international congress of metrology, <https://doi.org/10.1051/metrology/201911004>
13. P. Johannet, *Propagation des surtensions sur les réseaux électriques* (Bulletin de la DER EDF série B, jun. 1987), pp. 37–47
14. Janik M. Hager, *Dekonvolution mit Variationsansätzen* (Universität Stuttgart, Abschlussarbeit, 2015)
15. V. Schoonderwoert et al., *Huygens STED deconvolution increases signal-to-noise and image resolution*, *Micros. Today* **21**, 38–44 (2013)

Cite this article as: Jean-Pierre Fanton, Convolution and deconvolution: two mathematical tools to help performing tests in research and industry, Int. J. Metrol. Qual. Eng. **12**, 6 (2021)