

# Analog and digital worlds: Part 1. Signal sampling and Fourier Transform

Renato Seeber<sup>1</sup>  · Alessandro Ulrici<sup>2</sup> 

Received: 16 August 2016 / Accepted: 8 November 2016 / Published online: 19 November 2016  
© Springer International Publishing Switzerland 2016

**Abstract** Proper data sampling of a continuously varying quantity that describes the occurrence of a natural process leads to a simpler equivalent representation. This representation consists of a sequence of discrete data, which is more suitable to be mathematically handled and allows one to maintain essential information provided by the original signal. The discrete values of the sequence obtained by sampling differ from one another by finite quantities. In the ideal case, the original representation should be perfectly reconstructed by a backward procedure. The rules that should be respected to satisfy this basic condition are very simple, but require the decomposition of the signal into a suitable set of elementary components. This may be performed by applying to the sequence the algorithm called

Discrete Fourier Transform (DFT). The direct result of the transformation consists of the spectrum of the signal, which can be analyzed in many ways. The mathematics to make the FT algorithm work, eventually in an as fast as possible way (FFT—Fast Fourier Transform) is perhaps of less interest to the chemists and will not be treated here. Rather, the aim is to address the reader about what FT allows to obtain.

**Keywords** Sampling · Discretisation · Analog-to-digital conversion · Windowing · Fourier analysis

## Introduction

Several books [1–3] are devoted to the theory and practice of *signal sampling*. Proper data sampling procedures allow to extract information of an *event* or a *signal* in the form of individual data. Those data can be arranged in one or more columns, which are usually named vectors or matrices. Signal sampling constitutes a basic issue in a variety of situations. Correct sampling is necessary for obtaining the sequence of data that correctly account for whatever phenomenon we want to examine. In the simplest case, a *continuous trace* reports the values of a given quantity that depends on a single variable. The *dependent variable* depends on a single *independent variable*, which is the case of a *univariate signal*.

Whatever variable is appropriate to describe the occurrence of a dynamic natural process (change in temperature at passing time or at varying location, change in volume at varying external pressure, change in distance from the ground of a falling body, etc.), the relevant values change progressively by increments that are so small to be hardly conceivable.

---

Signals are broadly defined as a sequence of measurements that can either be continuous or discrete. Often, sequences are captured at varying intervals in time, frequency, wave numbers, or distance.

---

From D. A. Donald, Y. L. Everingham, L. W. McKinna, and D. Coomans, Feature Selection in the Wavelet Domain: Adaptive Wavelets, in COMPREHENSIVE CHEMOMETRICS. Chemical and Biochemical Data Analysis, S. D. Brown, R. Tauler, B. Walczak Eds., Elsevier, Amsterdam, 2009, pp. 357–376.

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s40828-016-0037-1) contains supplementary material, which is available to authorized users.

---

✉ Renato Seeber  
renato.seeber@unimore.it

<sup>1</sup> Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Via G.Campi, 103, 41125 Modena, Italy

<sup>2</sup> Department of Life Sciences, University of Modena and Reggio Emilia, Padiglione Besta, Via Amendola, 2, 42122 Reggio Emilia, Italy

The most part of the natural processes, as well as most of the human-driven ones, occur *continuously*. This means that subsequent steps are different from each other to such a small extent not to be perceivable or directly measurable by any human senses and even instruments, in agreement with the term ‘conceivable’ used above. Mathematicians define the whole matter as follows: the *infinitesimal* change in the quantity  $y$  describing the event is indicated by  $dy$  and corresponds to an *infinitesimal* change of the independent variable (e.g., time),  $dt$ .<sup>1</sup> For this reason,  $dy$  and  $dt$  are the general symbols used for infinitesimal changes in  $y$  and  $t$  quantities, respectively. To account for what happens ‘at a given instant’, rather than over a finite interval, they should be preferred to  $\Delta y$  and  $\Delta t$ , that indicate finite variations. This notwithstanding, the rate at which the event occurs is a finite quantity, because the ratio between the two infinitesimal changes expresses a finite value:  $dy/dt$ . The concept of *instantaneous rate*, mathematically corresponding to the value of the local tangent to a function expressing the dependence of  $y$  on  $t$ , is well known to express a (finite) rate. On the contrary, the  $\Delta y/\Delta t$  incremental ratio represents the *medium rate* within the  $\Delta t$  interval, more or less close to the instantaneous rate in the mid point of  $\Delta t$ . Noteworthy, the expression ‘at a given instant’ may refer to a ‘time instant’, to a ‘space instant’ or whatsoever ‘independent variable instant’, represented by  $t$ , chosen here to describe the progress of any event under observation.

The most direct way to monitor the changes occurring in the course of an event consists in ‘recording’ it by our own senses, that follow the movement of an object, the increase or decrease of loudness of a sound, the increase or decrease of the temperature ( $\Theta$ ) in the surrounding environment, etc., at flowing time or at changing the location, or whatsoever variable. By considering common instruments used to follow the evolution of the last exemplified event, the temperature may be measured on the basis of the ‘standardized’ change in volume of a liquid or of a gas: the *continuous* changes in temperature over time are transduced into the corresponding continuous changes in volume, which may be read or even recorded. In other words, two subsequent  $\Theta$  temperature values differ from each other by an infinitesimal amount,  $d\Theta$ , similarly to the corresponding change in volume,  $dV$  of the liquid or the gas used, and are recorded at infinitesimal time difference,  $dt$ .

## Data sampling

### Simple signals

Once we choose to read the temperature with a thermometer at given constant  $\Delta t$  intervals, every minute, every hour, or every day, we are sampling  $\Theta$  on the independent variable time. It will be clear from the following how crucial the sampling operation is.

Criteria for correct sampling are mandatory. Correct sampling means picking up any significant changes of the monitored quantity: we fill in a matrix consisting of two columns (for instance, time,  $t$ , and temperature,  $\Theta$ ) and of as many rows as many readings are made. No meaningful change in  $\Theta$  over time is lost only if sampling is correctly performed. On the other hand, incorrect sampling implies to miss information about what we are observing with respect to the event going on. For instance, we can choose to collect air samples at a crossroad, to determine the  $\text{NO}_x$  content due to the emission of the cars. Let sampling occur manually every hour, the time interval being conditioned by the lack of an in situ suitable gas sensor. No possibility to shift the workers’ shifts, due to untouchable time of lunch! However, let the rush hour occur just within the interval between two subsequent sampling operations: the peak of air poisoning is lost.... Similarly, in a space-located example, a source of poisoning may be unidentified if the water sampling sites within a basin are not properly chosen.

A visual example of how a critical issue proper sampling can be illustrated by sequences of pictures. Figure 1S in *Supplementary Material* describes the rotation of a needle at a velocity such that two subsequent pictures are taken at constant time intervals (*sampling interval* or *sampling period*,  $\Delta t$ ), lower than that required to cover half a round ( $\Delta t < T_p/2$ , where  $T_p$  is the rotation *period*, i.e., the time of a complete round). The needle rotates clockwise and looking at the sequence of pictures the perception of the direction of rotation is correct. On the other hand, if  $\Delta t$  is not lower than the time requested by half a round ( $\Delta t \geq T_p/2$ ), the rotation appears to occur counter clockwise: *aliasing* has occurred. In the same Fig. 1S, a point corresponding to the extremity of the needle is plotted vs. the time corresponding to the different positions: a sine (or cosine, depending on the starting point) function is depicted.<sup>2</sup> The difference in the independent variable,  $t$ , between two subsequent maxima, i.e., the *period* of the trigonometric function, coincides with that of rotation, namely the time difference between two subsequent equal positions of the needle. Correct or incorrect sampling of the needle

<sup>1</sup> In the present article, apart from the actual meaning (time, space or whatever else) we choose to indicate the independent variable with  $t$ .

<sup>2</sup> It may be useful to remind the basics of sine and cosine functions, expressed as follows in the general terms:  $y(t) = A \cos(2\pi v t + \phi)$  and  $y'(t) = B \sin(2\pi v t + \phi)$  where  $A$  and  $B$  represent the *amplitude* of the relevant functions.  $v$  is the *circular frequency*, expressed in  $\text{s}^{-1}$ , and  $\phi$  angle is the *phase*, in rad, such that  $0 \leq \phi < 2\pi$ .

rotation corresponds to correct or incorrect sampling in the corresponding trigonometric functions in Fig. 1S: the effect of aliasing is also shown with respect to the period of the sine and cosine functions.

In the all-day life, a practical example of incorrect sampling is found in old-fashioned western movies, where the recorded individual pictures (the individual samples) of the film are taken at constant time intervals,  $\Delta t$ , that are longer than that required to avoid aliasing: it may happen that the spokes in the wheels of a coach seem curiously to rotate in the opposite sense with respect to that predictable on the basis of the direction of the coach itself. An analogous effect can be observed in videos of propellers of an aircraft or of vanes of a fan.

The condition under which a sequence of pictures, taken at constant  $\Delta t$  sampling intervals, is suitable to represent the process properly, is expressed by the so-called *Nyquist<sup>3</sup> criterion* in terms of  $\Delta t$ ,  $T_P$ , and the relevant *frequencies*  $\nu_{\text{sampl}}$  and  $\nu$ , respectively. Making reference to the example of the rotating needle of Fig. 1S, the well known Eq. (1) describes the relation between rotation period,  $T_P$ , and frequency of rotation,  $\nu$ :

$$\nu = \frac{1}{T_P} \quad (1)$$

Similarly, the relation between the sampling interval,  $\Delta t$ , and the corresponding sampling frequency,  $\nu_{\text{sampl}}$ , is given by:

$$\nu_{\text{sampl}} = \frac{1}{\Delta t} \quad (2)$$

The *Nyquist criterion* expresses the condition for a correct sampling that can be written in different forms, all expressing the same condition:

<sup>3</sup>



Harry Theodor Nyquist was born in Nilsby, Sweden in 1889. He was a physicist, electrical and communications engineer, making fundamental contributions to telecommunications, electronics, and information theory. He developed theoretical background for analog-to-digital conversion of signals, which is closely related to the Nyquist criterion discussed here. He also developed the Nyquist diagrams, which are familiar to those working in impedance spectroscopy.

$$\Delta t < \frac{T_P}{2} \quad (3)$$

i.e.:

$$\Delta t < \frac{1}{2\nu} \quad (4)$$

or:

$$\nu_{\text{sampl}} > 2\nu \quad (5)$$

Summarizing and integrating:  $T_P$ , in seconds (s), the *period*, is the time required by the sine or cosine function to assume equal values in subsequent cycles, i.e., the time between two identical positions of the needle. The sampling period  $\Delta t$  is expressed in s, and corresponds to the time elapsed between two subsequent samplings, i.e., in the example of a film, between two subsequent pictures. Frequencies,  $\nu$  and  $\nu_{\text{sampl}}$ , are expressed in  $\text{s}^{-1}$  (also referred to as Hertz, Hz).  $\nu$  corresponds to the number of maxima of the sine or the cosine function in 1 s; in the cited example it corresponds to the circular frequency that is the number of rotations of the needle per time units. The sampling frequency,  $\nu_{\text{sampl}}$ , corresponds here to the number of samplings, i.e., of pictures, per s.

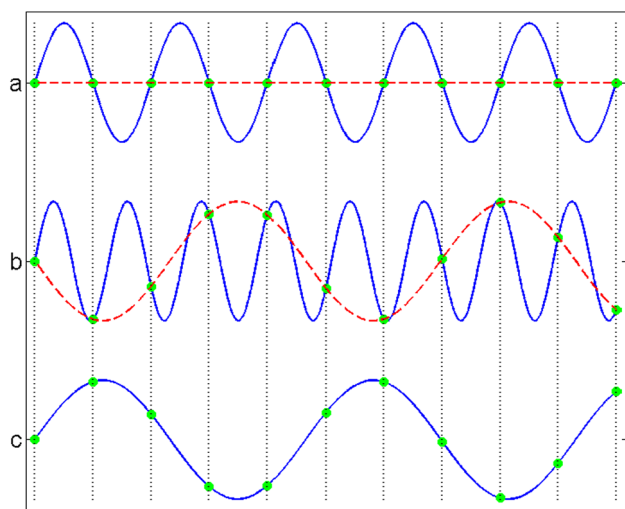
In Fig. 1 examples are shown of sine functions with different frequencies, sampled in agreement or in disagreement with the Nyquist criterion.

Equations (3–5) are valid whenever an event may be described by a single sine or cosine function, depending on the initial condition chosen. However, it very often happens that a signal is too complex to be accounted for by a single sine or a cosine function. Whenever the signal is a complex one, the correct sampling period,  $\Delta t$ , is given quantitative meaning by defining the *frequency* in the frame of the *analysis of the signal*, as it will be explained in the following sections. This requires the transformation of the signal from the original domain (for instance, *time* or *space domain*), into the corresponding *frequency domain*.

### Complex signals

In a complex periodic signal more individual sines and cosines at different frequencies have to be summed to each other to account for the whole process. Nothing is lost in the relevant description, only once  $\Delta t$  is lower than half the shortest period, corresponding to the highest frequency of the individual signals.<sup>4</sup> The sine and cosine individual functions are called *components* that, altogether, describe the overall periodic event. Equations (3) and (4) become, respectively:

<sup>4</sup> It is evident that in the real world it is not always clear or easy to say what the shortest period is. In this case, oversampling is necessary. The issue is treated in the following.



**Fig. 1** The same sampling period on simple sine signals (blue lines) possessing different frequency values. Sampling at  $\Delta t = \frac{1}{2v}$  (a);  $\Delta t > \frac{1}{2v}$  (b);  $\Delta t < \frac{1}{2v}$  (c). Aliasing occurs in (a) and in (b) (see red-dashed trace)

$$\Delta t < \frac{T_{P\min}}{2} \quad (6)$$

and

$$\Delta t < \frac{1}{2v_{\max}} \quad (7)$$

where  $T_{P\min}$  and  $v_{\max}$  are referred to the individual signal with the shortest period, corresponding to the highest frequency.

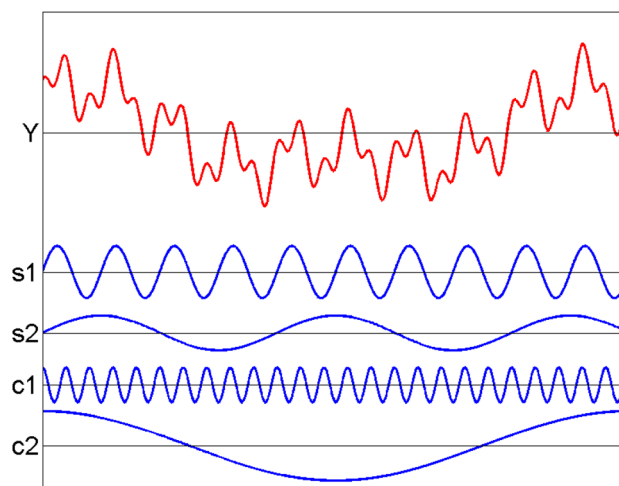
Noteworthy, the Nyquist criterion is often expressed directly with reference to a signal consisting of different components, in the form of Eq. (7). It follows that the definition of the useful lower limit for the sampling frequency,  $v_{\text{sampl}}$ , is:

$$v_{\text{sampl}} > 2v_{\max} \quad (8)$$

As an example, in Fig. 2, four trigonometric functions (individual signals) are used to build up the complex signal [4].

Actually, in the frame of an experiment it is often requested to follow the opposite path, i.e., to identify the sine and cosine linear combination suitable to reconstruct a complex signal.

As it was mentioned at the beginning of the article, similar to what happens for many natural events, an instrumental signal may vary continuously: it may be recorded by a system suitable to register *continuous* changes of the measured quantity as a function of a suitable independent variable. This is an *analog recorder*. As an example, the changes in atmosphere composition at varying the location or in agreement with seasonal variations, which is an analog process, may be monitored by an analog recorder (see Fig. 3).



**Fig. 2** The complex signal in the top part of the figure (Y, in red color) is given by the sum of two sine functions (s1 and s2) and two cosine functions (c1 and c2), each one with different amplitude and frequency values

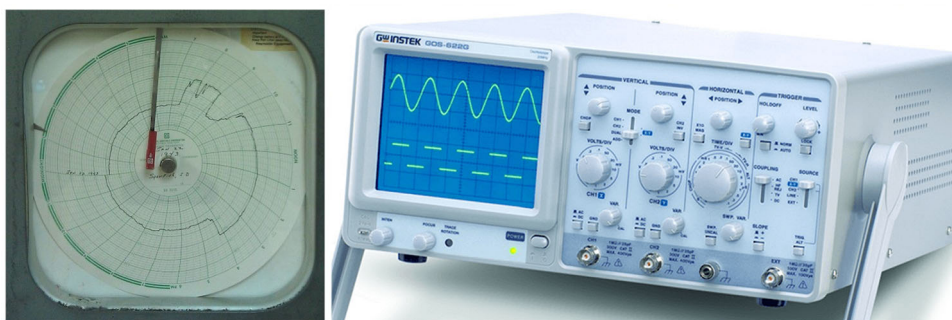
Sampling the analog signal of the recorder is necessary to obtain discrete data, i.e., numbers, suitable for computer handling: the computer is a *digital device*<sup>5</sup> that manipulates *digital data*. Proper sampling of the continuous signal is necessary to perform proper analog-to-digital (ADC) conversion: computers should be fed with vectors or sequences of data.<sup>6</sup> As an example, Fig. 4 shows how the continuous signal constituting a chromatogram is actually sampled, stored and displayed as a sequence of discrete data.

Recalling that the original continuous signal should be reconstructed from the sequences of discrete points obtained by sampling, some practical ways to operate should be exposed. Aiming at a good reconstruction of the continuous signal by interpolation of the points constituting the sampled sequence, it is strongly advisable to follow a general rule of thumb, i.e., to pick up 10 points for every cycle of the highest frequency present in a signal. This practical rule, also referred to as ‘Real World Sampling Theorem’ [5], implies a suggested sampling frequency

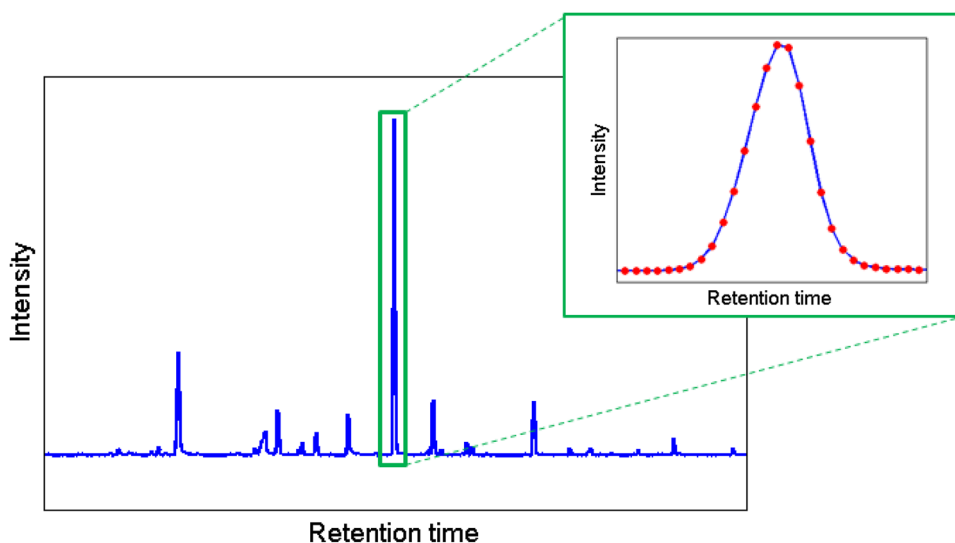
<sup>5</sup> Actually, analog computers also exist. An analog computer uses continuously varying electrical, mechanical, or hydraulic quantities to model the problem to solve. However, in 1950s and 1960s the advent of digital computers made analog computers obsolete, except for very specific applications like the flight computer in aircraft, which is still in widespread use in the 21st century.

<sup>6</sup> Numerical Analysis has progressively ‘substituted’ for Analysis, formulating algorithms that may be translated into computer programs. For the last forty years numerical analysis has shown an impressive growth, in coincidence with increasing diffusion and computational power of small computers. Numerical Analysis algorithms are much simpler for non-mathematicians. It happens that iterative procedures, often applied in numerical solution of mathematical problems, require many subsequent steps, are hence repetitive, long and even boring, which is however far from constituting a problem for a modern personal computer.

**Fig. 3** Some examples of analog recorders. *Left* analog recording thermometer (from <https://www.weather.gov/unr/1943-01-22>); *right* XY analog oscilloscope (from <http://www.gwinstek.com/upload/website/8d7367f7d206f3d63218396ab8a9a178.jpg>)



**Fig. 4** Zooming shows how a chromatographic peak, collected by an analog recorder, may be converted into a sequence of discrete data (highlighted in red). They can be displayed as such or interpolated by a proper algorithm to lead back to a continuous trace



much higher than the minimum value fixed by the Nyquist criterion. By following this choice, a piecewise interpolation procedure [6] consisting of a series of quadratic or even linear equations will give back the original signal with quite good accuracy. On the contrary, by adopting the minimum sampling frequency fixed by the Nyquist criterion, interpolation to reconstruct the original signal is much more demanding and computationally intensive.

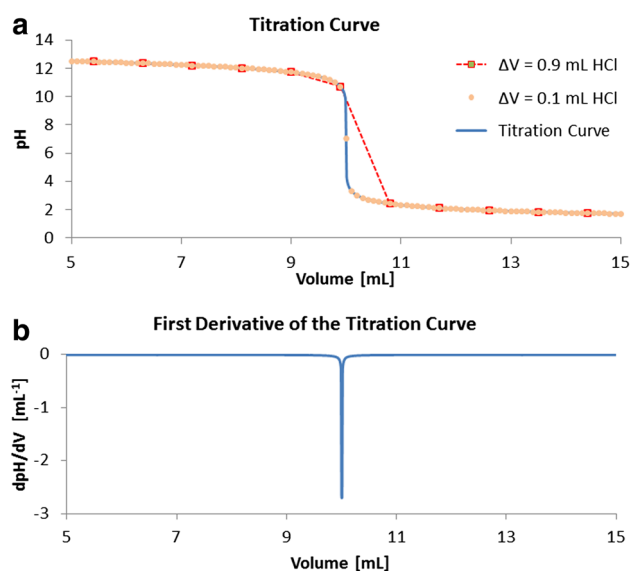
A further complication arises whenever a signal is an even more complex one, the highest frequency present being not always the same along the occurrence of the event. It is evident that changes of the sampling period may be suitable to avoid sampling, storing and elaborating redundant, meaningless data. *Sectioning* in some way different portions of the signal, looking at it through a *window*, may be advantageous. This operation (*windowing*), will be described in a further contribution, in a more general frame. An effective approach to economically handle a small number of data, as that described hereafter, may be followed in specific situations.

Any technicians, (skilful) chemistry students, and researchers, are well aware of a practical case of proper (or improper) sampling in the course of a trivial laboratory operation. When performing a pH-metric titration by hand, using a graduated glass burette, they read pH values different from

each other by discrete quantities, in correspondence to subsequent additions of discrete volumes  $V$  of titrant. It is very important, to represent the titration curve accurately in the most critical part, i.e., close to the endpoint, that the larger the change in the measured pH,  $\Delta\text{pH}$ , the smaller the corresponding finite volume of titrant added,  $\Delta V$ . The sampling frequency necessary to correctly describe the titration curve has to increase with increasing the rate at which pH changes, expressed by  $\Delta\text{pH}/\Delta V$ . If we do not ‘sample the event’ properly, the description of the titration procedure is unsatisfactory: we may lose the most significant points of the relevant pH vs.  $V$  curve.

Figure 5 shows an example of a similar titration curve and of the relevant derivative, which well accounts for the occurrence of smooth and of rapid changes in correspondence to different parts of the titration curve. Different sampling frequencies are suitable to adequately describe the curve ‘locally’: the Nyquist criterion requires different sampling frequencies for different segments of the signal.

The way how, in general, automatic titration systems work is quite interesting in this respect. If the additions are made by finite volume increments, the computer managing the titration decreases each titrant addition when pH changes rapidly, in order not to lose data in these parts of the pH vs.  $V$  curve close to the endpoint.



**Fig. 5** Titration curve of 10 mL NaOH 0.1 M with HCl 0.1 M, considering different sampling periods (a); derivative of the titration curve (b)

The case of a continuously varying signal may be represented once more by an automatic titrator in which the titrant volume addition is realised by a continuous flow. A procedure in which pH is sampled at discrete times may be adopted: the titrant flow is slowed down when  $\Delta\text{pH}$  increases. The two-column matrix of pH and corresponding  $V$  values is efficiently managed by the computer. The procedure is quite different from what usually happens in nature: we can not slow down the occurrence of a natural event so easily, as we slow down the titrant flux!

It is often of basic importance to know the rate at which the measured quantity changes, possibly ‘before’ the changes occur. In this case, we must have, in a way or another, a good idea about the ‘shape’ of the signal: it is not trivial to plan an ‘online’ effective way to proceed, unless we have access to a signal that is representative of what we are going to sample. Otherwise, the event has to proceed smoothly enough not to present sudden, unexpected changes. The trend of the potentiometric pH vs.  $V$  titration is once more helpful to illustrate what happens in this respect. The curve sends us ‘messages’ about its mind to change more rapidly. It does not exhibit discontinuities, so that the extent of the subsequent additions and of the parallel readings may be modulated ‘online’: the higher the  $\Delta\text{pH}/\Delta V$  change at the  $n$ -th titrant addition, the lower will be the  $\Delta V$  at the  $(n + 1)$ th addition.

### More about sampling: more about discretization or Analog-to-Digital Conversion (ADC)

An instrumental signal is often represented by a more or less complex function of one (or even more) independent

variable(s). The sequence of sampled data sent to the computer should bear the whole information content of the continuous trace; in principle, proper sampling implies that a suitable interpolation of the sampled points is capable to fit the continuous signal.

It is obvious that whenever no indication is possible with respect to a variation of the sampling frequency in correspondence to different portions of the signal (as it is the case of the automatic titrant), it is necessary to fix a constant suitable sampling frequency that obeys the Nyquist criterion over the whole signal.

To account for how the sampling operation on a *continuous* (univariate) signal is usually dealt with in dedicated textbooks, the definition of the *delta function*, also called *impulse*, or *Dirac delta function*,  $\delta$ , is necessary [1, 2]. It is defined as an infinitely high, infinitely thin spike, subtending a unitary area: it constitutes the limit of the sequence of zero-centred normal distribution as the standard deviation tends to 0. These properties make the  $\delta$  function suitable to pick up the value of a *continuous function*,  $f(t)$ , at a given  $t$  value.<sup>7</sup> An array of  $\delta$  functions shifted from 0 to be spaced from each other by  $\Delta t$  units, extending from  $-\infty$  to  $+\infty$ , describes the so-called *comb function*: a sampling function extended over the whole possible range of the signal (see Fig. 6).

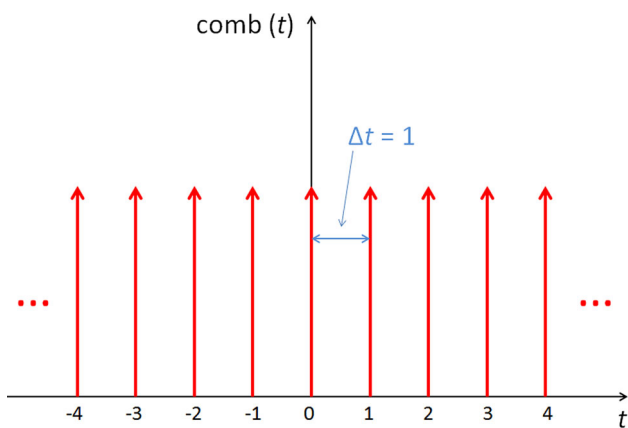
Sampling of a continuous signal by adopting  $\Delta t$  sampling period may be then represented by the scheme in Fig. 7, which is expressed by the following notation:

$$x(k) = f(t)\text{comb}_{\Delta t}(t) = \sum_{k=-\infty}^{+\infty} [f(t)\delta(t - k\Delta t)] \quad (9)$$

In conclusion, the operation leading to a *sequence*  $x(k)$  is performed by sampling the continuous signal, represented here by  $f(t)$ , through a train of equally spaced unit impulses, as expressed by the comb function in Fig. 6, with  $\Delta t$  sampling period (Fig. 7), computed in agreement with Nyquist criterion. As a result, the continuous signal  $f(t)$  is converted into a sequence  $x$  with index  $k$ ,  $x(k)$ , which corresponds to the discretization of  $f(t)$ , i.e., to the conversion of the analog signal  $f(t)$  into a digital (discrete) signal  $x(k)$ . Figure 8 a reports a continuous signal, which is correctly sampled in Fig. 8 b and undersampled in Fig. 8c, leading to the clearly misleading sequence in Fig. 8d. The already discussed aliasing phenomenon has occurred in this case.

On the other hand, the inverse path is possible by the interpolation operation, which performs Digital-to-Analog Conversion. Scheme 1 sketches the operations that have been described in the foregoing.

<sup>7</sup> The  $\delta(t)$  function only exists at  $t = 0$  and is defined as follows:  $\delta(t - t_0) = 0$  when  $t \neq t_0$ , and  $\int_{t_1}^{t_2} f(t)\delta(t - t_0)dt = f(t_0)$  for  $t_1 < t_0 < t_2$ . The impulse function can be used to ‘extract’ from a function  $f(t)$  only the value at time  $t_0$ .



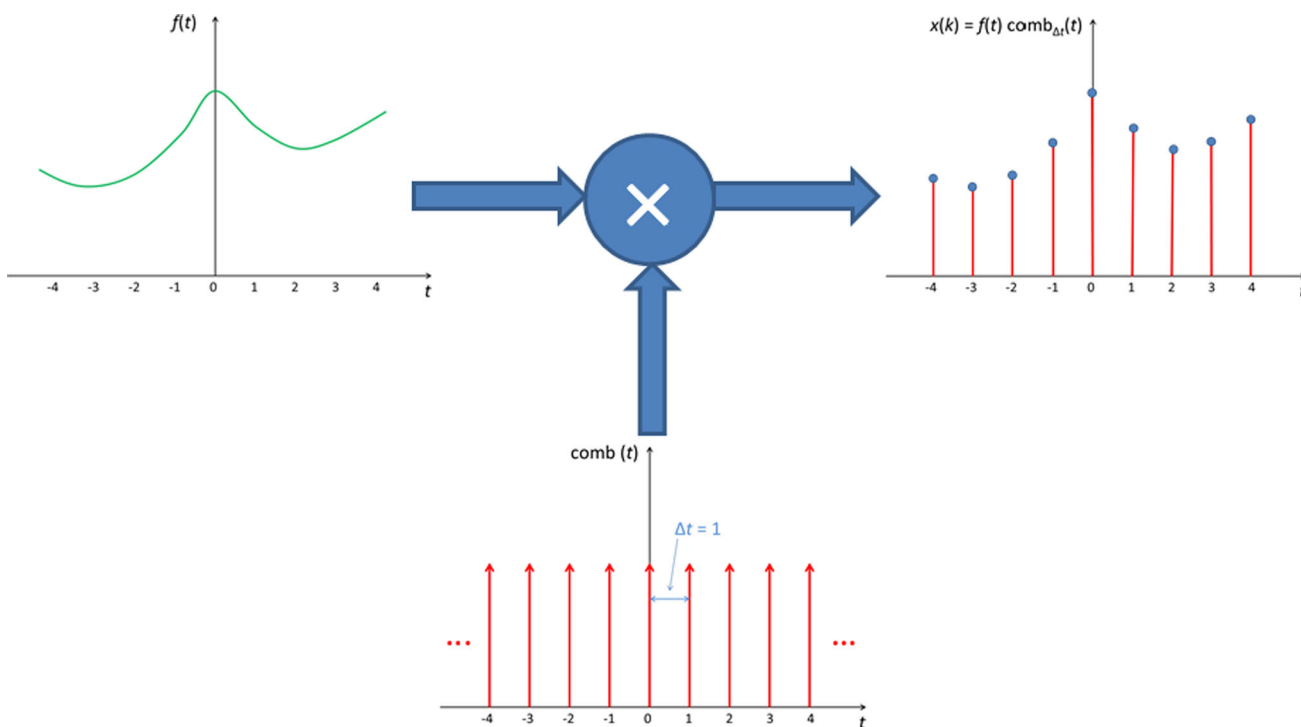
**Fig. 6** The comb function, defined in Eq. (9) for  $\Delta t = 1$

### Transform techniques

Generally, in mathematics, a transform means a mathematical operation to be applied on a function. We are familiar with some *transform* procedures, such as the *logarithmic transform*: some calculations are made easier by transforming the variables into their logarithms. For example, multiplications and divisions are substituted by sums and subtractions, respectively. Furthermore, in some cases logarithms allow more compact and effective representation of data over a wide range, as it happens in the

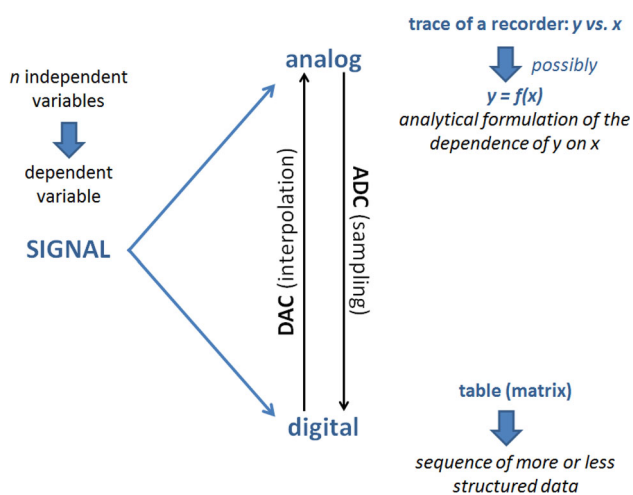
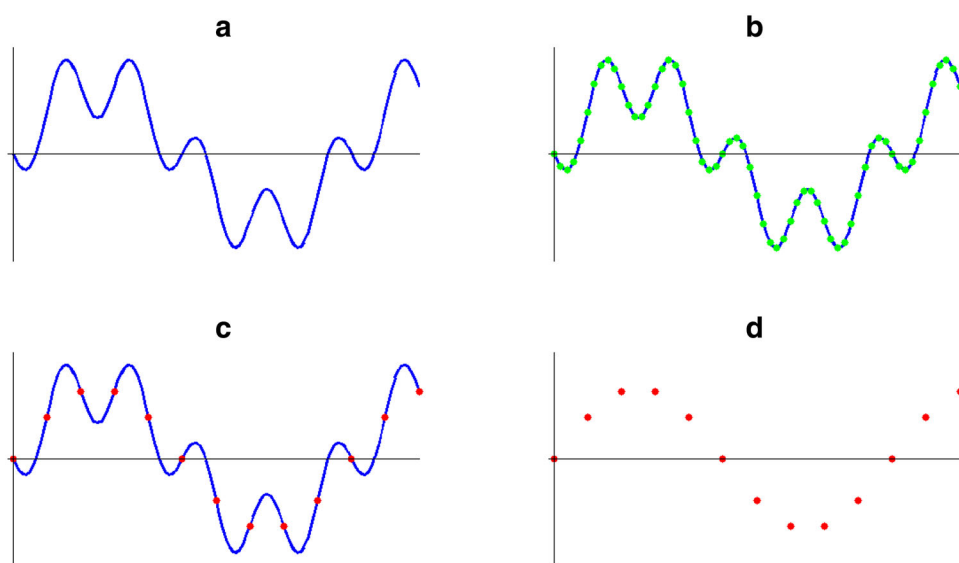
case of the acidity of a solution, expressed by pH, which is the logarithm of the reciprocal of the hydrogen ion activity.

When dealing with a signal, either a continuous signal described by a function  $f(t)$  or a discrete signal represented by a sequence  $x(k)$ , the meaning of transform generally assumes quite a different meaning. Essentially, calculating the transform of a signal leads to an alternative representation of the information contained in the signal itself, using an alternative domain. To give the reader an intuitive explanation of this obscure definition, it could be useful to recall Fig. 2: the red signal (Y) represents the values of a dependent variable measured as a function of an independent variable, e.g., time. In other words, Y represents the measured values in the time domain. However, Y can be also represented as the sum of the four blue signals (s1, s2, c1 and c2), that correspond to two sine and two cosine functions with different frequencies and amplitudes. Therefore, an alternative representation of the same signal Y could consist in reporting the amplitudes of sines and cosines as a function of their frequencies. This is a possible example of transform of the signal, from the original domain (time) to an alternative domain (frequency). In this case, the sine and cosine functions are used as building blocks to fit the original signal: each sine and cosine function with a given frequency contributes to the original signal with a weight equal to its amplitude. In other words,



**Fig. 7** Sampling of a continuous signal by adopting a  $\Delta t$  sampling period equal to 1

**Fig. 8** Continuous signal (a); correct sampling (b); undersampling (c); sequence of undersampled data (d)



**Scheme 1** Analog-to-digital conversion (ADC) and digital-to-analog conversion (DAC)

sines and cosines with different frequencies and with unitary amplitude constitute an *orthogonal set of functions*<sup>8</sup> that can be used as a *basis set*: a set of *basis functions* constituting the building blocks to reconstruct the original

<sup>8</sup> Sine and cosine functions with different frequencies are an orthogonal set of functions since their inner product over any closed interval of length  $2\pi$  is equal to zero, i.e.:  $\int_{0+I\pi}^{2\pi+I\pi} \cos(nt) \cos(mt) dt = \int_{0+I\pi}^{2\pi+I\pi} \sin(nt) \sin(mt) dt = \int_{0+I\pi}^{2\pi+I\pi} \cos(nt) \sin(mt) dt = 0$ , where  $l$  is a real number,  $n$  and  $m$  are positive integers  $> 0$ , and  $n \neq m$ . The last one of the three integrals (product of sine and cosine functions) is equal to 0 also for  $m = n$ . Orthogonality implies that no overlap of information is present between the components of the set of functions, therefore leading to non-redundancy.

signal. Each one of these functions can be properly weighted by multiplying it by a coefficient (the amplitude), and the sum of all the weighted basis functions gives the original signal.

In the real world, however, generally it is not possible to represent exactly the original signal through the combination of a finite number of functions, conversely to what is reported in the example of Fig. 2. Therefore, a linear combination of the basis functions will be calculated, in a way that the resulting  $g(t)$  function approximates, as close as possible the  $f(t)$  function representing the original signal. While doing this,  $g(t)$  will be calculated in a way to keep as low as possible the number of basis functions, and to achieve at the same time best approximation of  $f(t)$ .

Let  $u'(h)$  be the vector bearing the proper coefficients of the basis set, i.e., the weights of each basis function  $s'_h(t)$ . If only the first  $N$  coefficients are different from zero or if the first  $N$  coefficients are sufficient to obtain a satisfactory representation of the function  $f(t)$ , then  $g(t)$  will be defined as the linear combination given by:

$$g(t) = \sum_{h=0}^N u'(h) s'_h(t) \tag{10}$$

Similarly, when passing from the continuous function  $f(t)$  to the sequence  $x(k)$ , the sequence  $\hat{x}(k)$  approximating  $x(k)$  is defined according to:

$$\hat{x}(k) = \sum_{h=0}^{N-1} u(h) s_h(k) \tag{11}$$

where  $u(h)$  is the *coefficient vector* and the  $s_h(k)$  are the sequences that constitute the *basis set*.

Once the basis set is chosen, the coefficients of the different elements are found by the so-called *transform techniques* [7, 8]



## The Fourier<sup>9</sup> analysis

### Functions

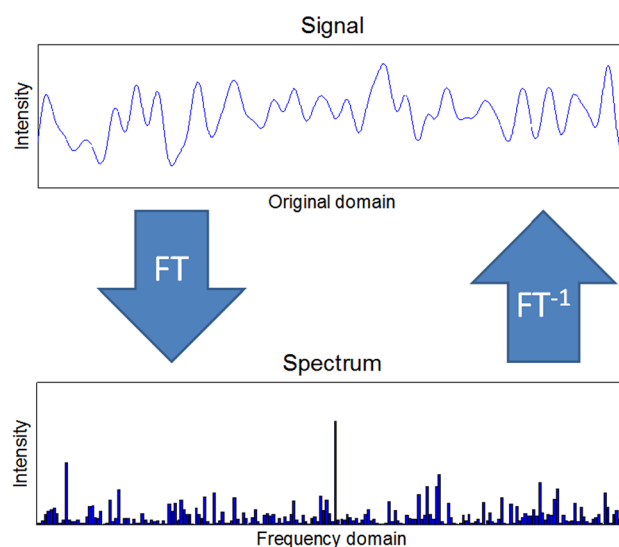
The inverse property exhibited by the logarithmic transform is also a property of the *Fourier Transform* (FT) [5, 9, 10]. At first, let's consider FT applied to a signal represented by a function: *inverse Fourier Transform* ( $FT^{-1}$ ) reconstructs back the original signal (see Fig. 9). Noteworthy, *perfect reconstruction* requires that the frequencies of the trigonometric functions are all multiples of a single *fundamental frequency*.

It has been mentioned above that the transform of a complex signal consists of the calculation of individual signals whose sum gives back the complex signal itself or its approximation,  $g(t)$ , according to Eq. (10). One possible approach is to choose, as individual signals, sine and cosine functions: sine and cosine constitute the basis set of FT. As defined in footnote <sup>h</sup> and sketched in Fig. 10, they constitute an orthogonal basis set. The sum of a proper number of these trigonometric functions with suitable angular frequencies and amplitudes reconstructs the complex signal, avoiding overlapping information, thanks to the reciprocal orthogonality. It does not seem redundant to once more underline that orthogonality constitutes a fundamental request for best effectiveness of the representation: the information content of each one of the two functions of a couple with the same argument does not bear anything redundant. Sines and cosines constitute the most widely used set of elementary functions. An operator, specifically the *Fourier Transform algorithm*, constitutes the tool for computing the amplitudes to assign to each individual sine and cosine component.

<sup>9</sup>



Jean-Baptiste Joseph Fourier was born in Auxerre, France, in 1768 and died in Paris, in 1830. He was a mathematician and physicist, especially known for elaborating the mathematic tools of Fourier series and, in general, of Fourier analysis. He also developed relevant applications to problems of heat transfer and vibrations.



**Fig. 9** Fourier Transform (FT) of a signal and inverse Fourier Transform ( $FT^{-1}$ ) of the corresponding spectrum

Being  $f(t)$  a continuous function of the  $t$  variable, the FT operator applied to the  $f(t)$  function generates the  $r(v)$  function in the so-called *alternative* (to  $t$ ) *frequency* ( $v$ ) *domain*;  $r(v)$  is in turn a continuous function of the  $v$  variable. On the other hand, the inverse transformation operator,  $FT^{-1}$ , regenerates  $f(t)$  from  $r(v)$ :

$$r(v) = FT[f(t)] \quad (12)$$

$$f(t) = FT^{-1}[r(v)] \quad (13)$$

$v$  is a variable expressing frequency:  $r(v)$  gives quantitative account for the *frequency content* of  $f(t)$ . The procedure that allows the determination of the frequency content of a signal, namely the *Fourier analysis*, is suitable to draw the *spectrum of the signal*, i.e., to represent the signal by the relevant frequencies. In other words, a signal  $f(t)$  in the *original domain* is transformed in  $r(v)$ , defined in the relevant *frequency domain*, through the FT algorithm.

As already cited, Fig. 2 reports a continuous signal, expressed by the  $f(t)$  function, together with the sine and cosine components necessary to draw it. It is a relatively simple function: only two sine and two cosine functions are necessary to account for it (see Table 1). The relevant spectra, in terms of cosine and sine coefficients, for four different frequency values, are reported in Fig. 11.

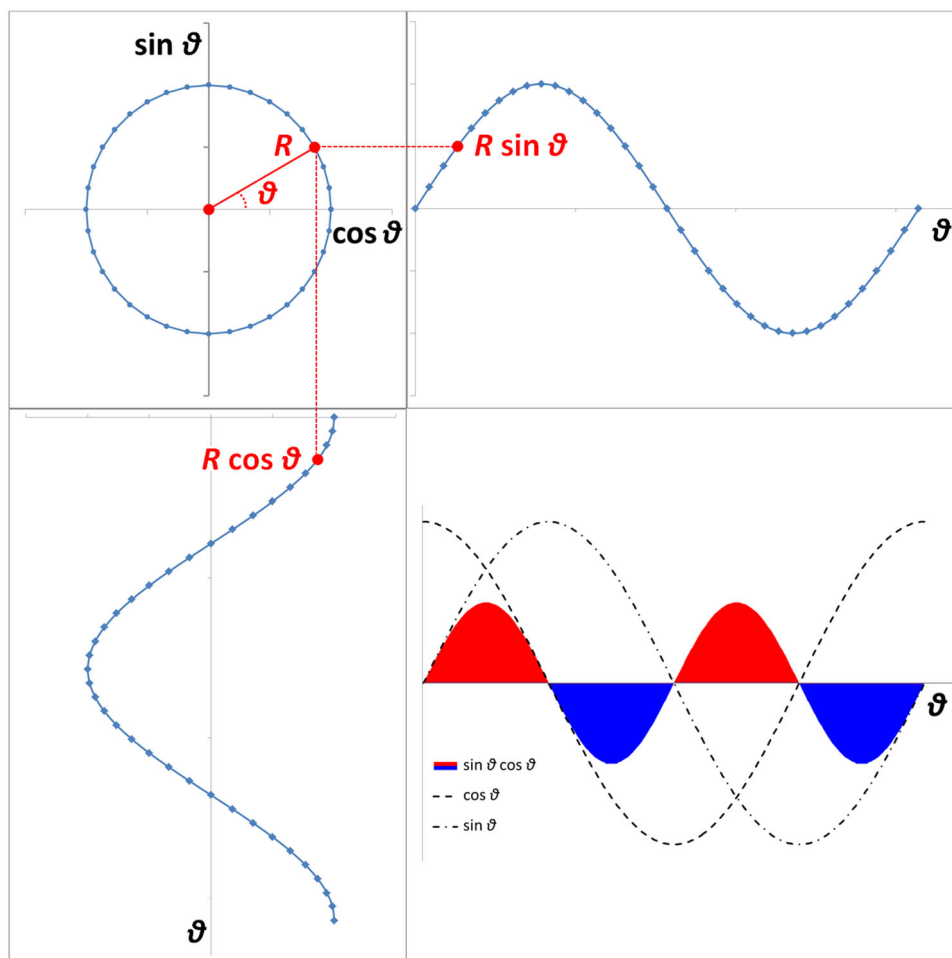
The complex signal in Fig. 2 can be expressed by:

$$g(t) = 3\sin(2\pi 10t) + 2\sin(2\pi 2.5t) + 2\cos(2\pi 25t) + 4\cos(2\pi 1t) \quad (14)$$

In the present case,  $g(t)$  coincides with  $f(t)$ .

Summarizing, we can list the amplitudes of the cosine functions ( $A_i$ ) and of the sine functions ( $B_i$ ) together with the corresponding frequencies,  $v_i$ , as reported in Table 1.

**Fig. 10** Sines and cosines are orthogonal to each other: plot of  $\sin \vartheta$  vs.  $\cos \vartheta$  (upper left), with  $\vartheta$  varying from 0 to  $2\pi$ , together with the corresponding values of  $\sin \vartheta$  and of  $\cos \vartheta$  as a function of  $\vartheta$ . In the lower right plot, the area corresponding to positive values of  $\sin \vartheta \cos \vartheta$  (red) is equal to the area corresponding to negative values of  $\sin \vartheta \cos \vartheta$  (blue); hence, the integral of  $\sin \vartheta \cos \vartheta$  over the interval from 0 to  $2\pi$  is equal to 0



**Table 1** Cosine (*A*) and sine (*B*) coefficients of the basis functions used to generate the complex continuous signal of Fig. 2

Basis functions	$v$	<i>A</i>	<i>B</i>
c2	1	4	0
s2	2.5	0	2
s1	10	0	3
c1	25	2	0

The names of the basis functions are the same as those used in Fig. 2

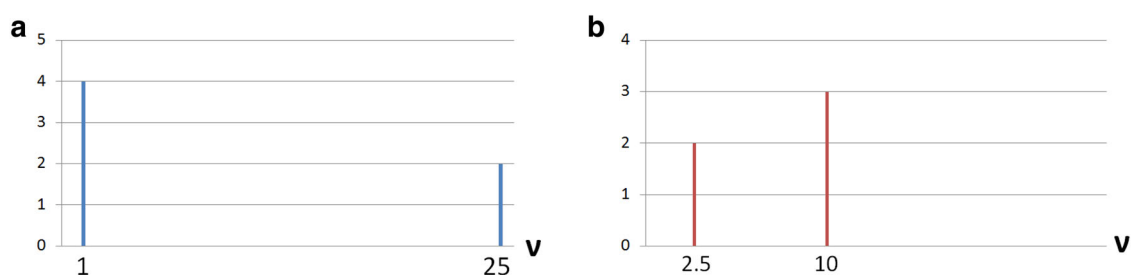
The spectra reported in Fig. 11 are, therefore, the plots of the *A* and of the *B* amplitudes vs.  $v$ .

Quite important operations are easier to perform on spectra than on the corresponding original signals. Furthermore, the analysis of the spectra is often more user-friendly than the signal itself: the spectra contain all information that is often more hidden in the original signals. Figure 12 reports some examples of instrumental signals (left-hand side) that are generally examined by looking at the corresponding spectra (right-hand side).

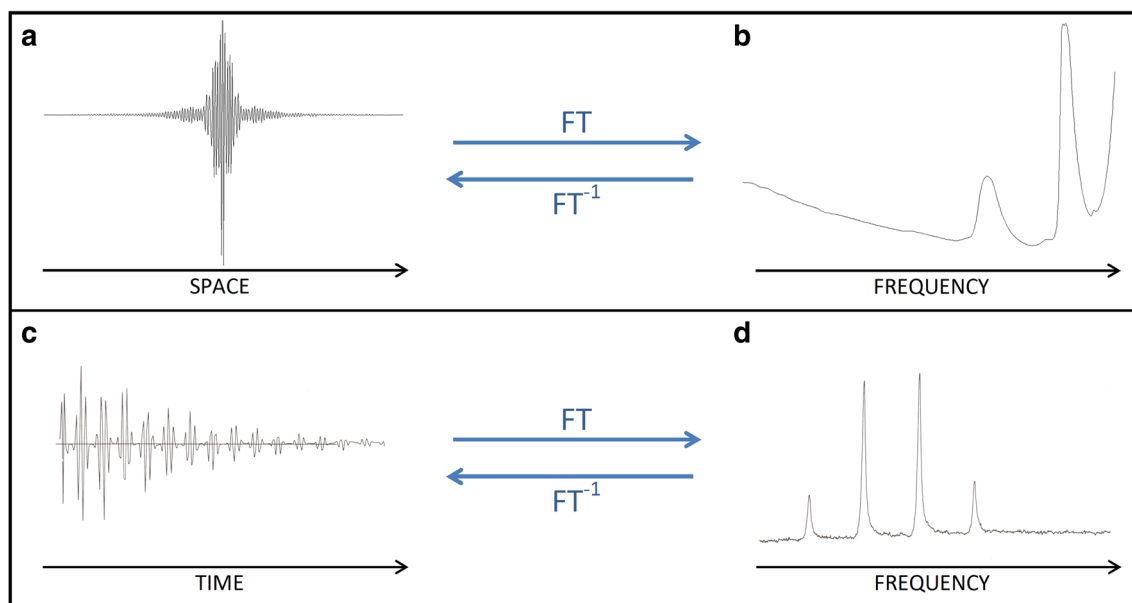
## Sequences

Let's now consider a signal represented by a sequence  $x(k)$ , consisting of  $N$  equally spaced points (where  $N$  is an even number), so that  $0 \leq k < N$ . Whenever we analyze a *digital signal*, i.e., on sequences of data, eventually obtained by proper sampling an originally continuous signal as defined in Eq. (9), we have to use the *Discrete Fourier Transform* (DFT) algorithm. The spectrum of the original signal is accounted for by representations that will be described hereafter: the analysis of the spectrum of the signal is actually necessary to perform correct sampling, according to the Nyquist criterion illustrated above [see Eq. (8)].

At this point, the reader is supposedly worried about an evident *dilemma*, which requires inserting an important consideration. The dilemma is: in front of the necessity to properly sample a continuous signal, we must calculate its spectrum. However, to calculate its spectrum, we should properly sample the signal. Which came first, the chicken or the egg? By reporting exactly from Ref. [5]: "The way out from this *dilemma* is to use common sense and a little intuition. That is, to sample a function when we do not



**Fig. 11** Spectra of the complex continuous signal reported in Fig. 2: cosine (A) and sine (B) coefficients



**Fig. 12** Near infrared interferogram (a) and corresponding spectrum (b); NMR free induction decay (FID) signal (c) and corresponding spectrum (D)

know his bandwidth (the difference between the upper and lower frequencies) a priori, we should choose the sampling rate such that a smooth curve drawn through the samples will ‘strongly resemble’ the original function. If we follow this original rule of thumb, then for most practical applications the function will be sampled fine enough.”.

Once this issue has been pointed out, let us go back to FT of sequences. The general expression for the sines and cosines suitable to account for the  $x(k)$  sequence,<sup>10</sup> in agreement with what has been discussed for the Fourier analysis, makes use of an integer  $h$  such that  $0 \leq h \leq N/2$ . Therefore,  $h$  accounts for discrete values of  $v$ ; the terms with different frequency are also called *harmonics*, or simply *frequencies*. If  $N$  points constitute the sequence in

the original domain,  $N$  harmonics exactly describe the signal in the frequency domain. The product  $2\pi h$  describes, at increasing  $h$  within the interval given:

1. a null frequency term, i.e., a constant term, accounting for the offset of the signal;
2. a unitary frequency, i.e., a frequency for a period as long as the whole  $x$  sequence;
3. a frequency corresponding to a period of one half the length of  $x$ , and so on, until, for  $h = N/2$ , the highest possible frequency.

The knowledge of  $N$  values of the  $x(k)$  sequence allows the solution of a system consisting of  $N$  equations with  $N$  unknowns

FT algorithm furnishes the values for  $A_h$  and  $B_h$  coefficients, constituting the sine and cosine amplitudes at different frequencies, respectively, that are necessary to account for every point of the  $x(k)$  sequence. The series of cosines accounts for the *symmetric portion* of the sequence, and that of sines the *antisymmetric* one (see the sine and

<sup>10</sup> For simplicity, in this paper we made use of sines and cosines to describe how FT works. Actually, representation through sines and cosines is equivalent to alternative representations, such as through modulus and phase of vectors or through complex numbers, making use of the Euler’s formulas. For the relationships existing among the different representations, the reader is referred to Ref. [5].

cosine functions in Fig. 10). The value of the generic  $x(k)$  point of the sequence is given by:

$$x(k) = \sum_{h=0}^{N/2} \left[ A_h \cos \frac{2\pi hk}{N} + B_h \sin \frac{2\pi hk}{N} \right] \quad (15)$$

where, for  $h = 0$  the cosine term equals 1 and the sine term equals 0, so that the  $A_0$  term is multiplied by 1, i.e., is constant, and  $B_0$  is absent. Similarly, for  $h = N/2$ ,  $B_{N/2}$  is absent. Therefore, the term  $A_0$  is a constant term, often referred to as continuous component.<sup>11</sup>

The cosine and sine coefficients, constituting the spectrum of the signal, are given by:

$$A_h = \frac{2}{N} \sum_{k=0}^{N-1} x(k) \cos \frac{2\pi hk}{N} \quad 0 \leq h \leq N/2 \quad (16)$$

and

$$B_h = \frac{2}{N} \sum_{k=0}^{N-1} x(k) \sin \frac{2\pi hk}{N} \quad 0 < h < N/2 \quad (17)$$

Therefore, the spectrum of the  $x(k)$  sequence consists of  $N/2 + 1$  cosine terms (actually,  $N/2$  true cosine terms + the continuous component) and  $N/2 - 1$  sine terms.

Summarizing, in Eq. (15) all coefficients  $A_h$  and  $B_h$  defined in Eqs. (16) and (17) are multiplied by the corresponding trigonometric functions with a given  $k$  value, leading to the corresponding  $x(k)$  point of the sequence.

It is noteworthy that, for given treatments of the signal that will be considered in the Part 2 of this contribution, such as the application of low-pass filters, the plot of the vector modulus  $\left[ M_h = \sqrt{(A_h^2 + B_h^2)} \right]$  vs. frequency,  $\nu$ , is sufficient to apply the filter properly.

## Conclusions and perspectives

This contribution, devoted to the treatment of analog and digital signals, discusses how the analog representation of a natural event or of a response of an instrument can be

reversibly translated into digital form, i.e., into a sequence, which brings the same information content. Practical indications are given with this respect. One of the 'transform techniques', the Discrete Fourier Transform, is especially popular and the characteristics of the frequency spectrum, obtained by this algorithm, are discussed.

In an upcoming part of this contribution, additional advantages and tools made available once a signal is 'Fourier transformed' will be discussed, with specific attention to situations that are most familiar to chemists. Mathematic operations, such as *convolution* and *deconvolution* of sequences, are much less known by chemists than the four basic arithmetic operations. This notwithstanding, they are more or less 'hidden' inside many instrumental signals (spectra, chromatograms, etc.), which chemists deal with every day and can constitute a powerful tool for effective treatment of data.

## References

1. Av Oppenheim, Schafer RW (1975) Digital signal processing. Prentice-Hall, Englewood Cliffs
2. Poularikas AD, Seely S (1994) Signals and systems, 2nd edn. PWS-KENT, Boston
3. Hamming RW (1989) Digital filters, 3rd edn. Dover Publ, inc., Mineola
4. Griffiths PR, de Haseth JA (2007) Transform techniques in chemistry, 2nd edn. Wiley, New York
5. Weaver HJ (1983) Applications of discrete and continuous Fourier analysis. Wiley, New York
6. Norris AC (1981) Computational chemistry. An introduction to numerical methods. Wiley, New York
7. Marshall AG (ed) (1982) Fourier, Hadamard, and Hilbert transform in chemistry. Plenum Press, New York
8. Seeber R, Ulrici A (1999) Transform methods in the synthesis and elaboration of signals. Quim Anal 18:11–27
9. Brigham EO (1974) The fast Fourier transform. Prentice-Hall, Englewood Cliffs
10. Bracewell RN (1978) The Fourier transform and its applications, 2nd edn. McGraw-Hill, New York

<sup>11</sup> For  $h = 0$  the first and the last terms of Eq. (15) are, respectively:  $A_0 \cos(2\pi 0 k/N) = A_0 \cos(0) = A_0$ , and  $B_0 \sin(2\pi 0 k/N) = B_0 \sin(0) = 0$ . Similarly, for  $h = N/2$ :  $B_{N/2} \sin(2\pi Nk/2 N) = B_{N/2} \sin(\pi k) = 0$ .