جامعة الانبار

كلية العلوم -قسم الجيولوجيا التطبيقية

اسم المحاضر باللغة الانكليزية: Ali Mishaal Abed

اسم المحاضر باللغة العربية: علي مشعل عبد

اسم المادة بالإنكليزي: Gravity Methods

اسم المادة بالعربي: الطرق الجذبية

عنوان المحاضرة بالإنكليزي: Geophysical data processing

2 Geophysical data processing

2.1 Introduction

Geophysical surveys measure the variation of some physical quantity, with respect either to position or to time. The quantity may, for example, be the strength of the Earth's magnetic field along a profile across an igneous intrusion. It may be the motion of the ground surface as a function of time associated with the passage of seismic waves. In either case, the simplest way to present the data is to plot a graph (Fig. 2.1) showing the variation of the measured quantity with respect to distance or time as appropriate. The graph will show some more or less complex waveform shape, which will reflect physical variations in the underlying geology, superimposed on unwanted variations from non-geological features (such as the effect of electrical power cables in the magnetic example, or vibration from passing traffic for the seismic case), instrumental inaccuracy and data collection errors. The detailed shape of the waveform may be uncertain due to the difficulty in interpolating the curve between widely spaced stations. The geophysicist's task is to separate the 'signal' from the 'noise' and interpret the signal in terms of ground structure.

Analysis of waveforms such as these represents an essential aspect of geophysical data processing and interpretation. The fundamental physics and mathematics of such analysis is not novel, most having been discovered in the 19th or early 20th centuries. The use of these ideas is also widespread in other technological areas such as radio, television, sound and video recording, radio-astronomy, meteorology and medical imaging, as well as military applications such as radar, sonar and satellite imaging. Before the general availability of digital computing, the quantity of data and the complexity of the processing severely restricted the use of the known techniques. This no longer applies and nearly all the techniques described in this chapter may be implemented in standard computer spreadsheet programs.

The fundamental principles on which the various methods of data analysis are based are brought together in this chapter. These are accompanied by a discussion of the techniques of digital data processing by computer that are routinely used by geophysicists. Throughout this chapter, waveforms are referred to as functions of time, but all the principles discussed are equally applicable to functions of distance. In the latter case, frequency (number of waveform cycles per unit time) is replaced by spatial frequency or *wavenumber* (number of waveform cycles per unit distance).

2.2 Digitization of geophysical data

Waveforms of geophysical interest are generally continuous (analogue) functions of time or distance. To apply the power of digital computers to the task of analysis, the data need to be expressed in digital form, whatever the form in which they were originally recorded.

A continuous, smooth function of time or distance can be expressed digitally by sampling the function at a fixed interval and recording the instantaneous value of the function at each sampling point. Thus, the analogue function of time $f(t)$ shown in Fig. 2.2(a) can be represented as the digital function $g(t)$ shown in Fig. 2.2(b) in which the continuous function has been replaced by a series of discrete values at fixed, equal, intervals of time. This process is inherent in many geophysical surveys, where readings are taken of the value of some parameter (e.g. magnetic field strength) at points along survey lines. The extent to which the digital values faithfully represent the original waveform will depend on the accuracy of the amplitude measurement and the intervals between measured samples. Stated more formally, these two parameters of a digitizing system are the sampling precision (dynamic range) and the sampling frequency.

Dynamic range is an expression of the ratio of the largest measurable amplitude A_{max} to the smallest measurable amplitude A_{min} in a sampled function. The higher the

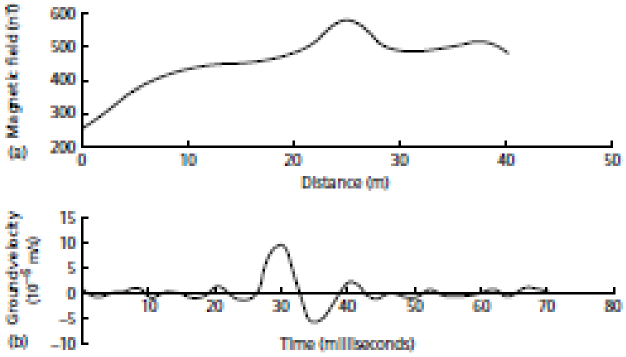

Fig. 2.1 (a) A graph showing a typical magnetic field strength variation which may be measured along a profile. (b) A graph of a typical seismogram, showing variation of particle velocities in the ground as a function of time during the passage of a seismic wave.

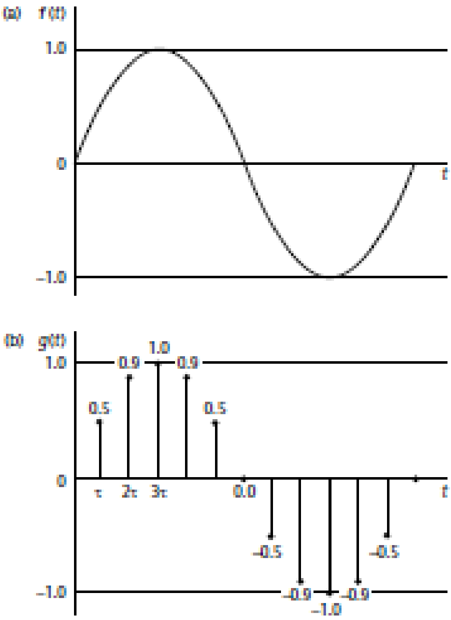

Fig. 2.2 (a) Analogue representation of a sinusoidal function. (b) Digital representation of the same function.

dynamic range, the more faithfully the amplitude variations in the analogue waveform will be represented in the digitized version of the waveform. Dynamic range is normally expressed in the *decibel* (dB) scale used to define electrical power ratios: the ratio of two power values P_1 and P_2 is given by $10\log_{10}(P_1/P_2)$ dB. Since *power* is proportional to the square of *signal amplitude A*

$$10\log_{10}(P_1/P_2) = 10\log_{10}(A_1/A_2)^2 = 20\log_{10}(A_1/A_2) \tag{2.1}$$

Thus, if a digital sampling scheme measures amplitudes over the range from 1 to 1024 units of amplitude, the dynamic range is given by

$$20\log_{10}(A_{max}/A_{min}) = 20\log_{10}1024 \approx 60\,\text{dB}$$

In digital computers, digital samples are expressed in binary form (i.e. they are composed of a sequence of digits that have the value of either 0 or 1). Each binary digit is known as a *bit* and the sequence of bits representing the sample value is known as a *word*. The number of bits in each word determines the dynamic range of a digitized waveform. For example, a dynamic range of 60 dB requires 11-bit words since the appropriate amplitude ratio of 1024 ($= 2^{10}$) is rendered as 10000000000 in binary form. A dynamic range of 84 dB represents an amplitude ratio of 2^{14} and, hence, requires sampling with 15-bit words. Thus, increasing the number of bits in each word in digital sampling increases the dynamic range of the digital function.

Sampling frequency is the number of sampling points in unit time or unit distance. Intuitively, it may appear that the digital sampling of a continuous function inevitably leads to a loss of information in the resultant digital function, since the latter is only specified by discrete values at a series of points. Again intuitively, there will be no significant loss of information content as long as the frequency of sampling is much higher than the highest frequency component in the sampled function. Mathematically, it can be proved that, if the waveform is a sine curve, this can always be reconstructed provided that there are a minimum of two samples per period of the sine wave.

Thus, if a waveform is sampled every two milliseconds (sampling interval), the sampling frequency is 500 samples per second (or 500 Hz). Sampling at this rate will preserve all frequencies up to 250 Hz in the sampled function. This frequency of half the sampling frequency is known as the *Nyquist frequency* (f_N) and the *Nyquist interval* is the frequency range from zero up to f_N

$$f_N = 1/(2\Delta t) \tag{2.2}$$

where Δt = sampling interval.

If frequencies above the Nyquist frequency are present in the sampled function, a serious form of distortion results known as *aliasing*, in which the higher frequency components are 'folded back' into the Nyquist interval. Consider the example illustrated in Fig. 2.3 in which sine waves at different frequencies are sampled. The lower frequency wave (Fig. 2.3(a)) is accurately reproduced, but the higher frequency wave (Fig. 2.3(b), solid line) is rendered as a fictitious frequency, shown by the dashed line, within the Nyquist interval. The relationship between input and output frequencies in the case of a sampling frequency of 500 Hz is shown in Fig. 2.3(c). It is apparent that an input frequency of 125 Hz, for example, is retained in the output but that an input frequency of 625 Hz is folded back to be output at 125 Hz also.

To overcome the problem of aliasing, the sampling frequency must be at least twice as high as the highest frequency component present in the sampled function. If the function does contain frequencies above the Nyquist frequency determined by the sampling, it must be passed through an *antialias filter* prior to digitization. The antialias filter is a low-pass frequency filter with a sharp cut-off that removes frequency components above the Nyquist frequency, or attenuates them to an insignificant amplitude level.

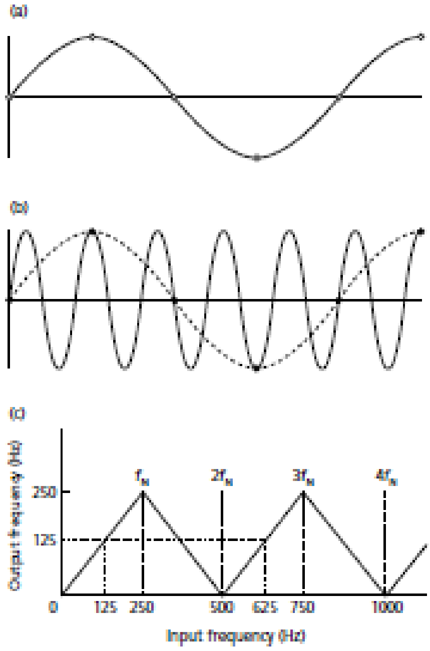

Fig. 2.3 (a) Sine wave frequency less than Nyquist frequency. (b) Sine wave frequency greater than Nyquist frequency (solid line) showing the fictitious frequency that is generated by aliasing (dashed line). (c) Relationship between input and output frequencies for a sampling frequency of 500 Hz (Nyquist frequency f_N = 250 Hz).

2.3 Spectral analysis

An important mathematical distinction exists between *periodic waveforms* (Fig. 2.4(a)), that repeat themselves at a fixed time period T, and *transient waveforms* (Fig. 2.4(b)), that are non-repetitive. By means of the mathematical technique of *Fourier analysis* any periodic waveform, however complex, may be decomposed into a series of sine (or cosine) waves whose frequencies are integer multiples of the basic repetition frequency $1/T$, known as the *fundamental frequency*. The higher frequency components, at frequencies of n/T (n = 1, 2, 3, . . .), are known as harmonics. The complex waveform of Fig. 2.5(a) is built up from the addition of the two individual sine wave components shown. To express any waveform in terms of its constituent sine wave components, it is necessary to define not only the frequency of each component but also its amplitude and phase. If in the above example the relative amplitude and phase relations of the individual sine waves are altered, summation can

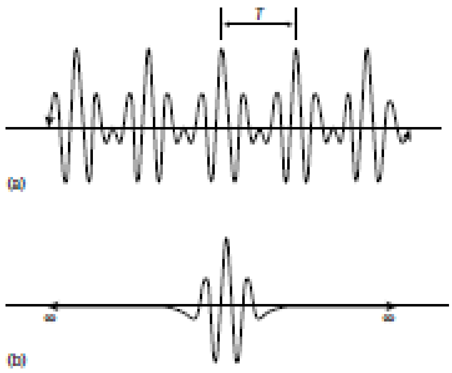

Fig. 2.4 (a) Periodic and (b) transient waveforms.

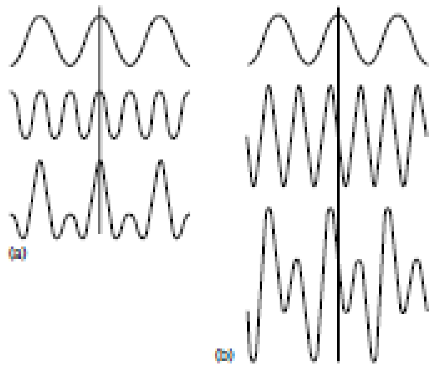

Fig. 2.5 Complex waveforms resulting from the summation of two sine wave components of frequency f and $2f$. (a) The two sine wave components are of equal amplitude and in phase. (b) The higher frequency component has twice the amplitude of the lower frequency component and is $\pi/2$ out of phase. (After Anstey 1965.)

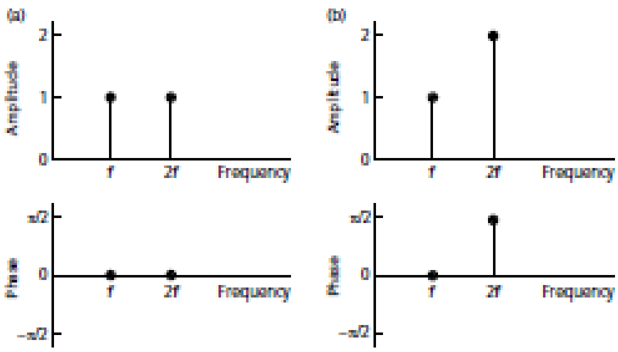

Fig. 2.6 Representation in the frequency domain of the waveforms illustrated in Fig. 2.5, showing their amplitude and phase spectra.

produce the quite different waveform illustrated in Fig. 2.5(b).

From the above it follows that a periodic waveform can be expressed in two different ways: in the familiar *time domain*, expressing wave amplitude as a function of time, or in the *frequency domain*, expressing the amplitude and phase of its constituent sine waves as a function of frequency. The waveforms shown in Fig. 2.5(a) and (b) are represented in Fig. 2.6(a) and (b) in terms of their amplitude and phase spectra. These spectra, known as line spectra, are composed of a series of discrete values of the amplitude and phase components of the waveform at set frequency values distributed between 0 Hz and the Nyquist frequency.

Transient waveforms do not repeat themselves; that is, they have an infinitely long period. They may be regarded, by analogy with a periodic waveform, as having an infinitesimally small fundamental frequency ($1/T \rightarrow 0$) and, consequently, harmonics that occur at infinitesimally small frequency intervals to give continuous amplitude and phase spectra rather than the line spectra of periodic waveforms. However, it is impossible to cope analytically with a spectrum containing an infinite number of sine wave components. Digitization of the waveform in the time domain (Section 2.2) provides a means of dealing with the continuous spectra of transient waveforms. A digitally sampled transient waveform has its amplitude and phase spectra subdivided into a number of

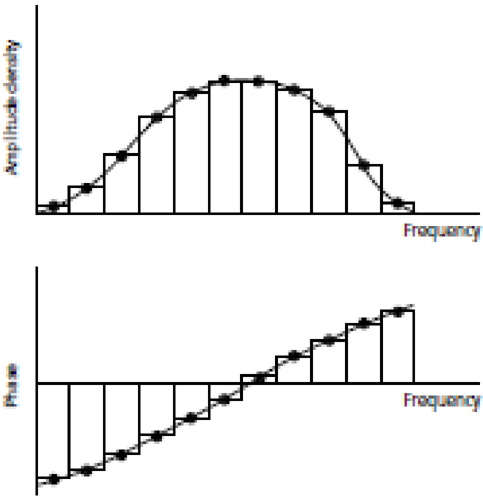

Fig. 2.7 Digital representation of the continuous amplitude and phase spectra associated with a transient waveform.

thin frequency slices, with each slice having a frequency equal to the mean frequency of the slice and an amplitude and phase proportional to the area of the slice of the appropriate spectrum (Fig. 2.7). This digital expression of a continuous spectrum in terms of a finite number of discrete frequency components provides an approximate representation in the frequency domain of a transient waveform in the time domain. Increasing the sampling frequency in the time domain not only improves the time-domain representation of the waveform, but also increases the number of frequency slices in the frequency domain and improves the accuracy of the approximation here too.

Fourier transformation may be used to convert a time function $g(t)$ into its equivalent amplitude and phase spectra $A(f)$ and $\phi(f)$, or into a complex function of frequency $G(f)$ known as the *frequency spectrum*, where

$$G(f) = A(f)e^{i\phi(f)} \tag{2.3}$$

The time- and frequency-domain representations of a waveform, $g(t)$ and $G(f)$, are known as a *Fourier pair*, represented by the notation

$$g(t) \leftrightarrow G(f) \tag{2.4}$$

Components of a Fourier pair are interchangeable, such that, if $G(f)$ is the Fourier transform of $g(t)$, then $g(t)$ is the Fourier transform of $G(f)$. Figure 2.8 illustrates Fourier pairs for various waveforms of geophysical significance. All the examples illustrated have *zero phase spectra*; that is, the individual sine wave components of the waveforms are in phase at zero time. In this case $\phi(f)$ = 0 for all values of ϕ. Figure 2.8(a) shows a spike function (also known as a *Dirac function*), which is the shortest possible transient waveform. Fourier transformation shows that the spike function has a continuous frequency spectrum of constant amplitude from zero to infinity; thus, a spike function contains all frequencies from zero to infinity at equal amplitude. The 'DC bias' waveform of Fig. 2.8(b) has, as would be expected, a line spectrum comprising a single component at zero frequency. Note that Fig. 2.8(a) and (b) demonstrate the principle of interchangeability of Fourier pairs stated above (equation (2.4)). Figures 2.8(c) and (d) illustrate transient waveforms approximating the shape of seismic pulses, together with their amplitude spectra. Both have a band-limited amplitude spectrum, the spectrum of narrower bandwidth being associated with the longer transient waveform. In general, the shorter a time pulse the wider is its frequency bandwidth and in the limiting case a spike pulse has an infinite bandwidth.

Waveforms with zero phase spectra such as those illustrated in Fig. 2.8 are symmetrical about the time axis and, for any given amplitude spectrum, produce the maximum peak amplitude in the resultant waveform. If phase varies linearly with frequency, the waveform remains unchanged in shape but is displaced in time; if the phase variation with frequency is non-linear the shape of the waveform is altered. A particularly important case in seismic data processing is the phase spectrum associated with *minimum delay* in which there is a maximum concentration of energy at the front end of the waveform. Analysis of seismic pulses sometimes assumes that they exhibit minimum delay (see Chapter 4).

Fourier transformation of digitized waveforms is readily programmed for computers, using a '*fast Fourier transform*' (FFT) algorithm as in the Cooley–Tukey method (Brigham 1974). FFT subroutines can thus be routinely built into data processing programs in order to carry out spectral analysis of geophysical waveforms. Fourier transformation is supplied as a function to standard spreadsheets such as Microsoft Excel. Fourier transformation can be extended into two dimensions (Rayner 1971), and can thus be applied to areal distributions of data such as gravity and magnetic contour maps.

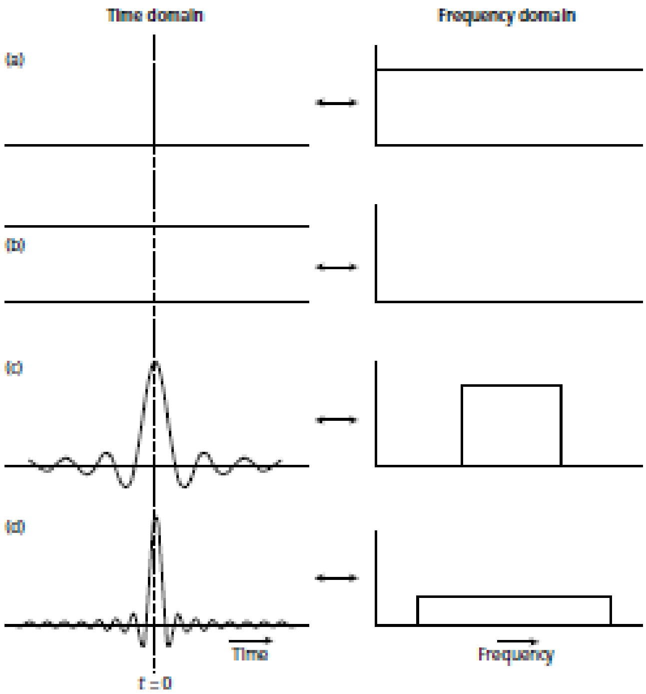

Fig. 2.8 Fourier transform pairs for various waveforms. (a) A spike function. (b) A 'DC bias'. (c) and (d) Transient waveforms approximating seismic pulses.

In this case, the time variable is replaced by horizontal distance and the frequency variable by wavenumber (number of waveform cycles per unit distance). The application of two-dimensional Fourier techniques to the interpretation of potential field data is discussed in Chapters 6 and 7.

2.4 Waveform processing

The principles of convolution, deconvolution and correlation form the common basis for many methods of geophysical data processing, especially in the field of seismic reflection surveying. They are introduced here in general terms and are referred to extensively in later chapters. Their importance is that they quantitatively describe how a waveform is affected by a filter. Filtering modifies a waveform by discriminating between its constituent sine wave components to alter their relative amplitudes or phase relations, or both. Most audio systems are provided with simple filters to cut down on high-frequency 'hiss', or to emphasize the low-frequency 'bass'. Filtering is an inherent characteristic of any system through which a signal is transmitted.

2.4.1 Convolution

Convolution (Kanasewich 1981) is a mathematical operation defining the change of shape of a waveform resulting from its passage through a filter. Thus, for example, a seismic pulse generated by an explosion is altered in shape by filtering effects, both in the ground and in the recording system, so that the seismogram (the filtered output) differs significantly from the initial seismic pulse (the input).

As a simple example of filtering, consider a weight suspended from the end of a vertical spring. If the top of the spring is perturbed by a sharp up-and-down movement (the input), the motion of the weight (the filtered output) is a series of damped oscillations out of phase with the initial perturbation (Fig. 2.9).

The effect of a filter may be categorized by its *impulse*

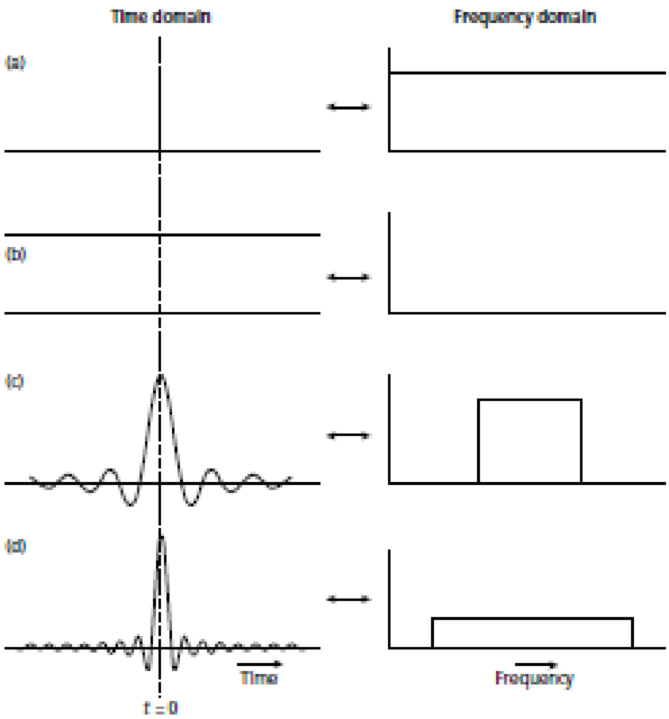

Fig. 2.8 Fourier transform pairs for various waveforms. (a) A spike function. (b) A 'DC bias'. (c) and (d) Transient waveforms approximating seismic pulses.

In this case, the time variable is replaced by horizontal distance and the frequency variable by wavenumber (number of waveform cycles per unit distance). The application of two-dimensional Fourier techniques to the interpretation of potential field data is discussed in Chapters 6 and 7.

2.4 Waveform processing

The principles of convolution, deconvolution and correlation form the common basis for many methods of geophysical data processing, especially in the field of seismic reflection surveying. They are introduced here in general terms and are referred to extensively in later chapters. Their importance is that they quantitatively describe how a waveform is affected by a filter. Filtering modifies a waveform by discriminating between its constituent sine wave components to alter their relative amplitudes or phase relations, or both. Most audio systems are provided with simple filters to cut down on high-frequency 'hiss', or to emphasize the low-frequency 'bass'. Filtering is an inherent characteristic of any system through which a signal is transmitted.

2.4.1 Convolution

Convolution (Kanasewich 1981) is a mathematical operation defining the change of shape of a waveform resulting from its passage through a filter. Thus, for example, a seismic pulse generated by an explosion is altered in shape by filtering effects, both in the ground and in the recording system, so that the seismogram (the filtered output) differs significantly from the initial seismic pulse (the input).

As a simple example of filtering, consider a weight suspended from the end of a vertical spring. If the top of the spring is perturbed by a sharp up-and-down movement (the input), the motion of the weight (the filtered output) is a series of damped oscillations out of phase with the initial perturbation (Fig. 2.9).

The effect of a filter may be categorized by its *impulse*

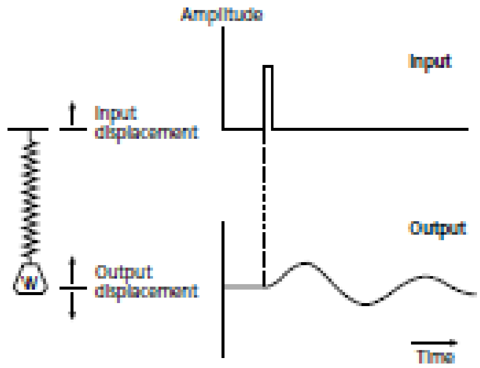

Fig. 2.9 The principle of filtering illustrated by the perturbation of a suspended weight system.

response which is defined as the output of the filter when the input is a spike function (Fig. 2.10). The impulse response is a waveform in the time domain, but may be transformed into the frequency domain as for any other waveform. The Fourier transform of the impulse response is known as the *transfer function* and this specifies the amplitude and phase response of the filter, thus defining its operation completely. The effect of a filter is described mathematically by a *convolution* operation such that, if the input signal $g(t)$ to the filter is convolved with the impulse response $f(t)$ of the filter, known as the convolution operator, the filtered output $y(t)$ is obtained:

$$y(t) = g(t) * f(t) \tag{2.5}$$

where the asterisk denotes the convolution operation.

Figure 2.11(a) shows a spike function input to a filter whose impulse response is given in Fig. 2.11(b). Clearly the latter is also the filtered output since, by definition, the impulse response represents the output for a spike input. Figure 2.11(c) shows an input comprising two separate spike functions and the filtered output (Fig. 2.11(d)) is now the superposition of the two impulse response functions offset in time by the separation of the input spikes and scaled according to the individual spike amplitudes. Since any transient wave can be represented as a series of spike functions (Fig. 2.11(e)), the general form of a filtered output (Fig. 2.11(f)) can be regarded as the summation of a set of impulse responses related to a succession of spikes simulating the overall shape of the input wave.

The mathematical implementation of convolution involves time inversion (or folding) of one of the functions and its progressive sliding past the other function, the individual terms in the convolved output being derived by summation of the cross-multiplication products over the overlapping parts of the two functions. In general, if g_i $(i = 1, 2, \ldots, m)$ is an input function and f_j $(j = 1, 2, \ldots, n)$ is a convolution operator, then the convolution output function y_k is given by

$$y_k = \sum_{i=1}^{m} g_i f_{k-i} \quad (k = 1, 2, \ldots, m+n-1) \tag{2.6}$$

In Fig. 2.12 the individual steps in the convolution process are shown for two digital functions, a double spike function given by $g_i = g_1, g_2, g_3 = 2, 0, 1$ and an impulse response function given by $f_i = f_1, f_2, f_3, f_4 = 4, 3, 2, 1$, where the numbers refer to discrete amplitude values at the sampling points of the two functions. From Fig. 2.11 it can be seen that the convolved output $y_i = y_1, y_2, y_3, y_4, y_5, y_6 = 8, 6, 8, 5, 2, 1$. Note that the convolved output is longer than the input waveforms; if the functions to be convolved have lengths of m and n, the convolved output has a length of $(m + n - 1)$.

The convolution of two functions in the time domain becomes increasingly laborious as the functions become longer. Typical geophysical applications may have functions which are each from 250 to a few thousand samples long. The same mathematical result may be obtained by transforming the functions to the frequency domain, then multiplying together equivalent frequency terms of their amplitude spectra and adding terms of their phase spectra. The resulting output amplitude and phase spectra can then be transformed back to the time domain. Thus, digital filtering can be enacted in either the time

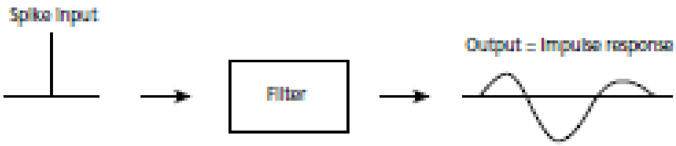

Fig. 2.10 The impulse response of a filter.

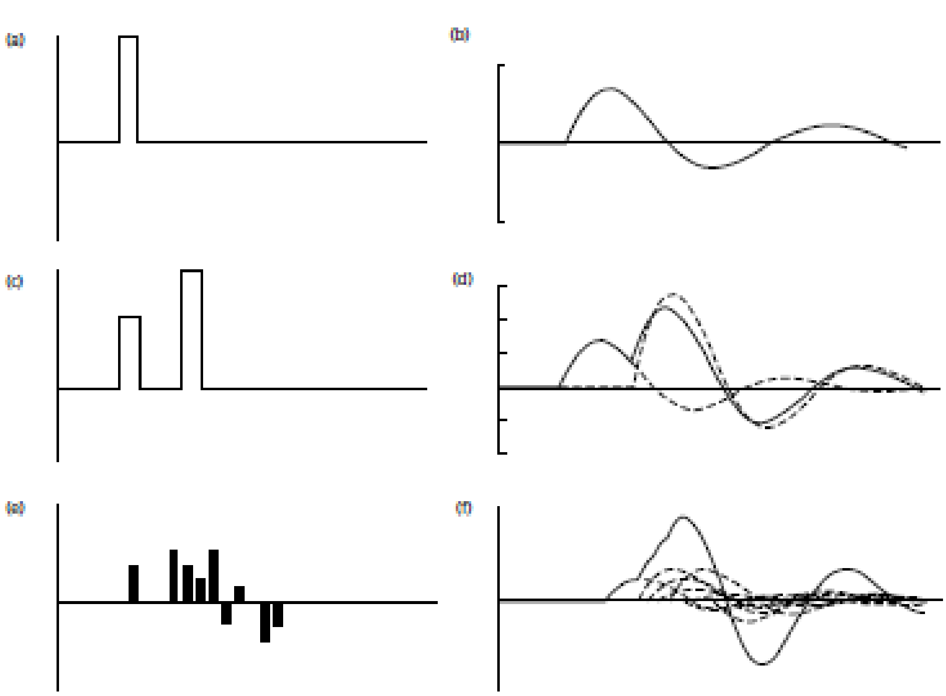

Fig. 2.11 Examples of filtering. (a) A spike input. (b) Filtered output equivalent to impulse response of filter. (c) An input comprising two spikes. (d) Filtered output given by summation of two impulse response functions offset in time. (e) A complex input represented by a series of contiguous spike functions. (f) Filtered output given by the summation of a set of impulse responses.

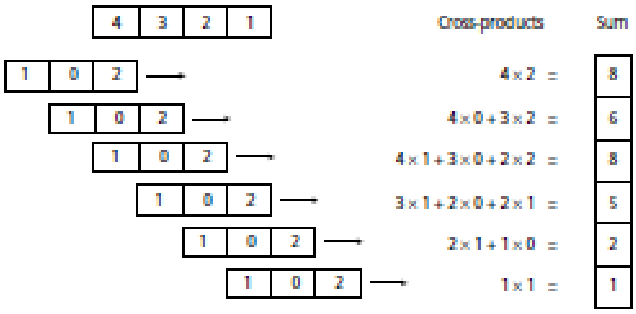

Fig. 2.12 A method of calculating the convolution of two digital functions.

domain or the frequency domain. With large data sets, filtering by computer is more efficiently carried out in the frequency domain since fewer mathematical operations are involved.

Convolution, or its equivalent in the frequency domain, finds very wide application in geophysical data processing, notably in the digital filtering of seismic and potential field data and the construction of synthetic seismograms for comparison with field seismograms (see Chapters 4 and 6).

References

Kearey P. , Brooks M. , Hill I., (2002) An Introduction Geophysical Exploration", 3rd ed. Blackwell Science Ltd., USA, 281 pages