ELEC 3004 Quiz 1 Review

Contents

Articles

References

Article Licenses

Signals

Signal (electronics)

A **signal** is any stream of quantities in time or spatial sequence. *Signals* categorizes to the fields of communications, signal processing, and to electrical engineering more generally.

Signals may contain and transport information in coded form like modulation. In that case the actual quantities are finally not used, but are decoded in a detector or demodulator.

In the physical world, any quantity measurable through time or over space can be taken as a signal. Within a complex society, any set of human information or machine data can also be taken as a signal. Such information or machine data (for example, the dots on a screen, the ink making up text on a paper page, or the words now flowing into the reader's mind) must all be part of systems existing in the physical world – either living or non-living.

Despite the complexity of such systems, their outputs and inputs can often be represented as simple quantities measurable through time or across space. In the latter half of the 20th century, electrical engineering itself separated into several disciplines, specializing in the design and analysis of physical signals and systems, on the one hand, and in the functional behavior and conceptual structure of the complex human and machine systems, on the other. These engineering disciplines have led the way in the design, study, and implementation of systems that take advantage of signals as simple measurable quantities in order to facilitate the transmission, storage, and manipulation of information.

Some definitions

Definitions specific to subfields are common. For example, in information theory, a *signal* is a codified message, that is, the sequence of states in a communication channel that encodes a message.

In the context of signal processing, arbitrary binary data streams are not considered as signals, but only analog and digital signals that are representations of analog physical quantities.

In a *communication system*, a *transmitter* encodes a *message* into a signal, which is carried to a *receiver* by the communications *channel*. For example, the words "Mary had a little lamb" might be the message spoken into a telephone. The telephone transmitter converts the sounds into an electrical voltage signal. The signal is transmitted to the receiving telephone by wires; and at the receiver it is reconverted into sounds.

In telephone networks, signalling, for example common-channel signaling, refers to phone number and other digital control information rather than the actual voice signal.

Signals can be categorized in various ways. The most common distinction is between discrete and continuous spaces that the functions are defined over, for example discrete and continuous time domains. Discrete-time signals are often referred to as *time series* in other fields. Continuous-time signals are often referred to as *continuous signals* even when the signal functions are not continuous; an example is a square-wave signal.

A second important distinction is between discrete-valued and continuous-valued. Digital signals are sometimes defined as discrete-valued sequencies of quantified values, that may or may not be derived from an underlying continuous-valued physical process. In other contexts, digital signals are defined as the continuous-time waveform signals in a digital system, representing a bit-stream. In the first case, a signal that is generated by means of a digital modulation method is considered as converted to an analog signal, while it is considered as a digital signal in the second case.signal transfers information

Discrete-time and continuous-time signals

If for a signal, the quantities are defined only on a discrete set of times, we call it a discrete-time signal. In other words, a discrete-time real (or complex) signal can be seen as a function from (a subset of) the set of integers to the set of real (or complex) numbers.

A continuous-time real (or complex) signal is any real-valued (or complex-valued) function which is defined for all time *t* in an interval, most commonly an infinite interval.

Analog and digital signals

Less formally than the theoretical distinctions mentioned above, two main types of signals encountered in practice are *analog* and *digital*. In short, the difference between them is that digital signals are *discrete* and *quantized*, as defined below, while analog signals possess neither property.

Discretization

One of the fundamental distinctions between different types of signals is between continuous and discrete time. In the mathematical abstraction, the domain of a continuous-time (CT) signal is the set of real numbers (or some interval thereof), whereas the domain of a discrete-time (DT) signal is the set of integers (or some interval). What these integers represent depends on the nature of the signal.

DT signals often arise via sampling of CT signaexample, consists of a continually fluctuating voltage on a line that can be digitized by an ADC circuit, wherein the circuit will read the voltage level on the line, say, every 50 microseconds. The resulting stream of numbers is stored as digital data on a discrete-time signal. Computers and other digital devices are restricted to discrete time.

Quantization

If a signal is to be represented as a sequence of numbers, it is impossible to maintain arbitrarily high precision - each number in the sequence must have a finite number of digits. As a result, the values of such a signal are restricted to belong to a finite set; in other words, it is quantized.

Examples of signals

- *Motion*. The motion of a particle through some space can be considered to be a signal, or can be represented by a signal. The domain of a motion signal is one-dimensional (time), and the range is generally three-dimensional. Position is thus a 3-vector signal; position and orientation is a 6-vector signal.
- *Sound*. Since a sound is a vibration of a medium (such as air), a sound signal associates a pressure value to every value of time and three space coordinates. A microphone converts sound pressure at some place to just a function of time, generating a voltage signal as an analog of the sound signal. Sound signals can be sampled to on a discrete set of time points; for example, compact discs (CDs) contain discrete signals representing sound, recorded at 44,100 samples per second; each sample contains data for a left and right channel, which may be considered to be a 2-vector signal (since CDs are recorded in stereo).
- *Images*. A picture or image consists of a brightness or color signal, a function of a two-dimensional location. A 2D image can have a continuous spatial domain, as in a traditional photograph or painting; or the image can be discretized in space, as in a raster scanned digital image. Color images are typically represented as a combination of images in three primary colors, so that the signal is vector-valued with dimension three.
- *Videos*. A video signal is a sequence of images. A point in a video is identified by its two-dimensional position and by the time at which it occurs, so a video signal has a three-dimensional domain. Analog video has one continuous domain dimension (across a scan line) and two discrete dimensions (frame and line).

- Biological *membrane potentials*. The value of the signal is a straightforward electric potential ("voltage"). The domain is more difficult to establish. Some cells or organelles have the same membrane potential throughout; neurons generally have different potentials at different points. These signals have very low energies, but are enough to make nervous systems work; they can be measured in aggregate by the techniques of electrophysiology.

Entropy

Another important property of a signal (actually, of a statistically defined class of signals) is its entropy or *information content*.

References

- Hsu, P. H. *Schaum's Theory and Problems: Signals and Systems*, McGraw-Hill 1995, ISBN 0-07-030641-9
- Lathi, B.P., *Signal Processing & Linear Systems*, Berkeley-Cambridge Press, 1998, ISBN 00941413357
- Shannon, C. E., 2005 [1948], "A Mathematical Theory of Communication," (corrected reprint [1]), accessed Dec. 15, 2005. Orig. 1948, *Bell System Technical Journal*, vol. 27, pp. 379–423, 623-656.

References

[1] http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html

Even and odd functions

In mathematics, **even functions** and **odd functions** are functions which satisfy particular symmetry relations, with respect to taking additive inverses. They are important in many areas of mathematical analysis, especially the theory of power series and Fourier series. They are named for the parity of the powers of the power functions which satisfy each condition: the function $f(x) = x^n$ is an even function if n is an even integer, and it is an odd function if n is an odd integer.

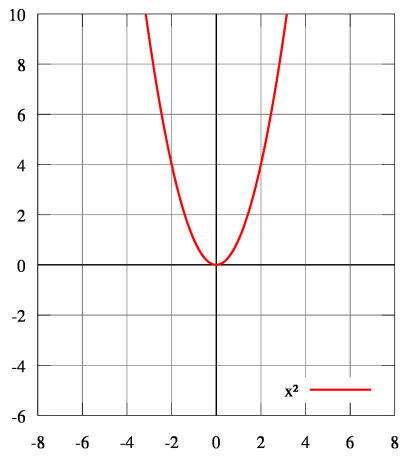
Even functions

Let $f(x)$ be a real-valued function of a real variable. Then f is **even** if the following equation holds for all x in the domain of f:

$$f(x) = f(-x).$$

Geometrically speaking, the graph face of an even function is symmetric with respect to the y-axis, meaning that its graph remains unchanged after reflection about the y-axis.

Examples of even functions are $|x|$, x^2, x^4, $\cos(x)$, and $\cosh(x)$.



$f(x) = x^2$ is an example of an even function.

Odd functions

Again, let $f(x)$ be a real-valued function of a real variable. Then f is **odd** if the following equation holds for all x in the domain of f:

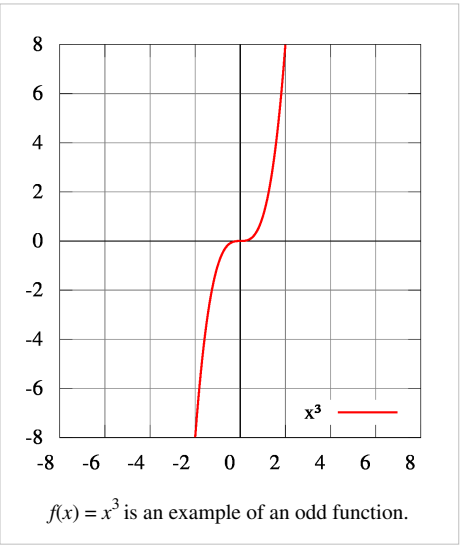
$$-f(x) = f(-x)\,,$$

or

$$f(x) + f(-x) = 0\,.$$

Geometrically, the graph of an odd function has rotational symmetry with respect to the origin, meaning that its graph remains unchanged after rotation of 180 degrees about the origin.

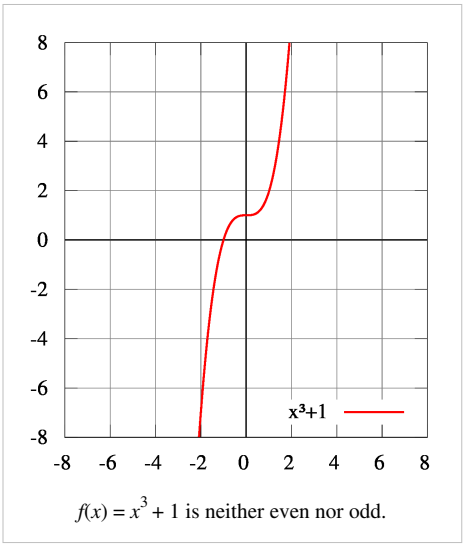
Examples of odd functions are x, x^3, $\sin(x)$, $\sinh(x)$, and $\operatorname{erf}(x)$.



$f(x) = x^3$ is an example of an odd function.

Some facts

A function's being odd or even does not imply differentiability, or even continuity. For example, the Dirichlet function is even, but is nowhere continuous. Properties involving Fourier series, Taylor series, derivatives and so on may only be used when they can be assumed to exist.



$f(x) = x^3 + 1$ is neither even nor odd.

Basic properties

- The only function which is *both* even and odd is the constant function which is equal to zero (i.e., $f(x) = 0$ for all x).
- The sum of an even and odd function is neither even nor odd, unless one of the functions is equal to zero over the given domain.
- The sum of two even functions is even, and any constant multiple of an even function is even.
- The sum of two odd functions is odd, and any constant multiple of an odd function is odd.
- The product of two even functions is an even function.
- The product of two odd functions is an even function.
- The product of an even function and an odd function is an odd function.
- The quotient of two even functions is an even function.
- The quotient of two odd functions is an even function.
- The quotient of an even function and an odd function is an odd function.
- The derivative of an even function is odd.
- The derivative of an odd function is even.
- The composition of two even functions is even, and the composition of two odd functions is odd.
- The composition of an even function and an odd function is even.
- The composition of any function with an even function is even (but not vice versa).

- The integral of an odd function from $-A$ to $+A$ is zero (where A is finite, and the function has no vertical asymptotes between $-A$ and A).
- The integral of an even function from $-A$ to $+A$ is twice the integral from 0 to $+A$ (where A is finite, and the function has no vertical asymptotes between $-A$ and A).

Series

- The Maclaurin series of an even function includes only even powers.
- The Maclaurin series of an odd function includes only odd powers.
- The Fourier series of a periodic even function includes only cosine terms.
- The Fourier series of a periodic odd function includes only sine terms.

Algebraic structure

- Any linear combination of even functions is even, and the even functions form a vector space over the reals. Similarly, any linear combination of odd functions is odd, and the odd functions also form a vector space over the reals. In fact, the vector space of *all* real-valued functions is the direct sum of the subspaces of even and odd functions. In other words, every function $f(x)$ can be written uniquely as the sum of an even function and an odd function:

  $$f(x) = f_{\mathrm{e}}(x) + f_{\mathrm{o}}(x)\,,$$

  where

  $$f_{\mathrm{e}}(x) = \tfrac{1}{2}[f(x) + f(-x)]$$

  is even and

  $$f_{\mathrm{o}}(x) = \tfrac{1}{2}[f(x) - f(-x)]$$

  is odd. For example, if f is exp, then f_{e} is cosh and f_{o} is sinh.
- The even functions form a commutative algebra over the reals. However, the odd functions do *not* form an algebra over the reals.

Harmonics

In signal processing, harmonic distortion occurs when a sine wave signal is sent through a memoryless nonlinear system, that is, a system whose output at time t only depends on the input at time t and does not depend on the input at any previous times. Such a system is described by a response function $V_{\mathrm{out}}(t) = f(V_{\mathrm{in}}(t))$. The type of harmonics produced depend on the response function f:[1]

- When the response function is even, the resulting signal will consist of only even harmonics of the input sine wave; $2f, 4f, 6f, \ldots$
  - The fundamental is also an odd harmonic, so will not be present.
  - A simple example is a full-wave rectifier.
- When it is odd, the resulting signal will consist of only odd harmonics of the input sine wave; $1f, 3f, 5f, \ldots$
  - The output signal will be half-wave symmetric.
  - A simple example is clipping in a symmetric push-pull amplifier.
- When it is asymmetric, the resulting signal may contain either even or odd harmonics; $1f, 2f, 3f, \ldots$
  - Simple examples are a half-wave rectifier, and clipping in an asymmetrical class A amplifier.

Notes

[1] Ask the Doctors: Tube vs. Solid-State Harmonics (http://www.uaudio.com/webzine/2005/october/content/content2.html)

Linear system

A **linear system** is a mathematical model of a system based on the use of a linear operator. Linear systems typically exhibit features and properties that are much simpler than the general, nonlinear case. As a mathematical abstraction or idealization, linear systems find important applications in automatic control theory, signal processing, and telecommunications. For example, the propagation medium for wireless communication systems can often be modeled by linear systems.

A general deterministic system can be described by operator, H , that maps an input, $x(t)$, as a function of t to an output, $y(t)$, a type of black box description. Linear systems satisfy the properties of superposition and scaling or homogeneity. Given two valid inputs

$x_1(t)$

$x_2(t)$

as well as their respective outputs

$$y_1(t) = H\{x_1(t)\}$$
$$y_2(t) = H\{x_2(t)\}$$

then a linear system must satisfy

$$\alpha y_1(t) + \beta y_2(t) = H\{\alpha x_1(t) + \beta x_2(t)\}$$

for any scalar values α and β .

The behavior of the resulting system subjected to a complex input can be described as a sum of responses to simpler inputs. In nonlinear systems, there is no such relation. This mathematical property makes the solution of modelling equations simpler than many nonlinear systems. For time-invariant systems this is the basis of the impulse response or the frequency response methods (see LTI system theory), which describe a general input function $x(t)$ in terms of unit impulses or frequency components.

Typical differential equations of linear time-invariant systems are well adapted to analysis using the Laplace transform in the continuous case, and the Z-transform in the discrete case (especially in computer implementations).

Another perspective is that solutions to linear systems comprise a system of functions which act like vectors in the geometric sense.

A common use of linear models is to describe a nonlinear system by linearization. This is usually done for mathematical convenience.

Time-varying impulse response

The **time-varying impulse response** $h(t_2,t_1)$ of a linear system is defined as the response of the system at time $t = t_2$ to a single impulse applied at time $t = t_1$. In other words, if the input $x(t)$ to a linear system is

$$x(t) = \delta(t - t_1)$$

where $\delta(t)$ represents the Dirac delta function, and the corresponding response $y(t)$ of the system is

$$y(t)|_{t=t_2} = h(t_2, t_1)$$

then the function $h(t_2,t_1)$ is the time-varying impulse response of the system.

Time-varying convolution integral

Continuous time

The output of any continuous time linear system is related to the input by the time-varying convolution integral:

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

or, equivalently,

$$y(t) = \int_{-\infty}^{\infty} h(t,t-\tau)x(t-\tau)d\tau$$

Discrete time

The output of any discrete time linear system is related to the input by the time-varying convolution sum:

$$y[n] = \sum_{k=-\infty}^{\infty} h[n,k]x[k]$$

or equivalently,

$$y[n] = \sum_{m=-\infty}^{\infty} h[n,n-m]x[n-m]$$

where

$$k = n - m$$

represents the lag time between the stimulus at time *m* and the response at time *n*.

Causality

A linear system is **causal** if and only if the system's time varying impulse response is identically zero whenever the time *t* of the response is earlier than the time *s* of the stimulus. In other words, for a causal system, the following condition must hold:

$$h(t,s) = 0 \text{ for } t < s$$

Time-invariant system

A **time-invariant** (TIV) system is one whose output does not depend explicitly on time.

> If the input signal $x(t)$ produces an output $y(t)$ then any time shifted input, $x(t+\delta)$, results in a time-shifted output $y(t+\delta)$

This property can be satisfied if the transfer function of the system is not a function of time except expressed by the input and output. This property can also be stated in another way in terms of a schematic

> If a system is time-invariant then the system block is commutative with an arbitrary delay.

Simple example

To demonstrate how to determine if a system is time-invariant then consider the two systems:

- System A: $y(t) = t\,x(t)$
- System B: $b(t) = 10x(t)$

Since system A explicitly depends on *t* outside of $x(t)$ and $y(t)$ then it is not time-invariant. System B, however, does not depend explicitly on *t* so it is time-invariant.

Formal example

A more formal proof of why system A & B from above differ is now presented. To perform this proof, the second definition will be used.

System A:

> Start with a delay of the input $x_d(t) = x(t+\delta)$
>
> $$y(t) = t\,x_d(t)$$
> $$y_1(t) = t\,x_d(t) = t\,x(t+\delta)$$
>
> Now delay the output by δ
>
> $$y(t) = t\,x_d(t)$$
> $$y_2(t) = y(t+\delta) = (t+\delta)x(t+\delta)$$
>
> Clearly $y_1(t) \neq y_2(t)$, therefore the system is not time-invariant.

System B:

> Start with a delay of the input $x_d(t) = x(t+\delta)$
>
> $$y(t) = 10\,x_d(t)$$
> $$y_1(t) = 10\,x_d(t) = 10\,x(t+\delta)$$
>
> Now delay the output by δ
>
> $$y(t) = 10\,x_d(t)$$
> $$y_2(t) = y(t+\delta) = 10\,x(t+\delta)$$
>
> Clearly $y_1(t) = y_2(t)$, therefore the system is time-invariant. Although there are many other proofs, this is the easiest.

Abstract example

We can denote the **shift operator** by $\mathbb{T}_r$ where r is the amount by which a vector's index set should be shifted. For example, the "advance-by-1" system

$$x(t+1) = \delta(t+1) * x(t)$$

can be represented in this abstract notation by

$$\tilde{x}_1 = \mathbb{T}_1 \, \tilde{x}$$

where $\tilde{x}$ is a function given by

$$\tilde{x} = x(t) \, \forall \, t \in \mathbb{R}$$

with the system yielding the shifted output

$$\tilde{x}_1 = x(t+1) \, \forall \, t \in \mathbb{R}$$

So $\mathbb{T}_1$ is an operator that advances the input vector by 1.

Suppose we represent a system by an operator $\mathbb{H}$. This system is **time-invariant** if it commutes with the shift operator, i.e.,

$$\mathbb{T}_r \, \mathbb{H} = \mathbb{H} \, \mathbb{T}_r \; \forall \, r$$

If our system equation is given by

$$\tilde{y} = \mathbb{H} \, \tilde{x}$$

then it is time-invariant if we can apply the system operator $\mathbb{H}$ on $\tilde{x}$ followed by the shift operator $\mathbb{T}_r$, or we can apply the shift operator $\mathbb{T}_r$ followed by the system operator $\mathbb{H}$, with the two computations yielding equivalent results.

Applying the system operator first gives

$$\mathbb{T}_r \, \mathbb{H} \, \tilde{x} = \mathbb{T}_r \, \tilde{y} = \tilde{y}_r$$

Applying the shift operator first gives

$$\mathbb{H} \, \mathbb{T}_r \, \tilde{x} = \mathbb{H} \, \tilde{x}_r$$

If the system is time-invariant, then

$$\mathbb{H} \, \tilde{x}_r = \tilde{y}_r$$

Causal system

A **causal system** (also known as a physical or **nonanticipative system**) is a system where the output depends on past/current inputs but not future inputs i.e. the output $y(t_0)$ only depends on the input $x(t)$ for values of $t \leq t_0$.

The idea that the output of a function at any time depends only on past and present values of input is defined by the property commonly referred to as causality. A system that has *some* dependence on input values from the future (in addition to possible dependence on past or current input values) is termed a non-causal or acausal system, and a system that depends *solely* on future input values is an anticausal system. Note that some authors have defined an anticausal system as one that depends solely on future *and present* input values or, more simply, as a system that does not depend on past input values.

Classically, nature or physical reality has been considered to be a causal system. Physics involving special relativity or general relativity require more careful definitions of causality, as described elaborately in causality (physics).

The causality of systems also plays an important role in digital signal processing, where filters are often constructed so that they are causal. For more information, see causal filter. For a causal system, the impulse response of the system must be 0 for all t<0. That is the sole necessary as well as sufficient condition for causality of a system, linear or non-linear.

Note that the systems may be discrete or continuous. Similar rules apply to both kind of systems.

Mathematical definitions

Definition 1: A system mapping x to y is causal if and only if, for any pair of input signals $x_1(t)$ and $x_2(t)$ such that

$$x_1(t) = x_2(t), \quad \forall\ t \leq t_0,$$

the corresponding outputs satisfy

$$y_1(t) = y_2(t), \quad \forall\ t \leq t_0.$$

Definition 2: Suppose $h(t)$ is the impulse response of the system H . (only fully accurate for a system described by linear constant coefficient differential equation)

$$h(t) = 0, \quad \forall\ t < 0$$

then the system H is causal, otherwise it is non-causal.<tr>

Examples

The following examples are for systems with an input x and output y .

Examples of causal systems

- Memoryless system

$$y\left(t\right) = 1 + x\left(t\right)\cos\left(\omega t\right)$$

- Autoregressive filter

$$y\left(t\right) = \int_0^\infty x(t-\tau)e^{-\beta\tau}\,d\tau$$

Examples of non-causal (acausal) systems

$$y(t) = \int_{-\infty}^{\infty} \sin(t+\tau)x(\tau)\,d\tau$$

- Central moving average

$$y_n = \frac{1}{2}\,x_{n-1} + \frac{1}{2}\,x_{n+1}$$

- For coefficients of t

$$y\,(t) = x(at)$$

Examples of anti-causal systems

$$y(t) = \int_{0}^{\infty} \sin(t+\tau)x(\tau)\,d\tau$$

- Time reversal

$$y\,(t) = x(-t)$$

- Look-ahead

$$y_n = x_{n+1}$$

References

- Oppenheim, Alan V.; Willsky, Alan S.; Nawab, Hamid; with S. Hamid (1998). *Signals and Systems*. Pearson Education. ISBN 0-13-814757-4.

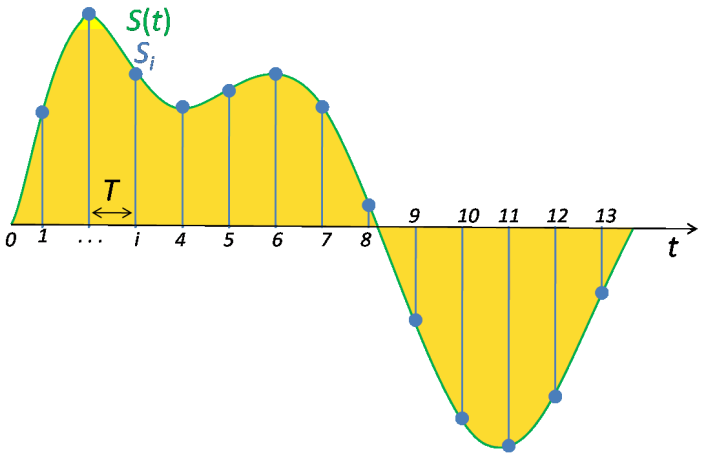
Sampling

Sampling (signal processing)

In signal processing, **sampling** is the reduction of a continuous signal to a discrete signal. A common example is the conversion of a sound wave (a continuous signal) to a sequence of samples (a discrete-time signal).

A **sample** refers to a value or set of values at a point in time and/or space.

A **sampler** is a subsystem or operation that extracts samples from a continuous signal. A theoretical ideal sampler produces samples equivalent to the instantaneous value of the continuous signal at the desired points.



Signal sampling representation. The continuous signal is represented with a green color whereas the discrete samples are in blue.

Theory

See also: Nyquist–Shannon sampling theorem

Sampling can be done for functions varying in space, time, or any other dimension, and similar results are obtained in two or more dimensions.

For functions that vary with time, let $s(t)$ be a continuous function to be sampled, and let sampling be performed by measuring the value of the continuous function every T seconds, which is called the sampling interval. Thus, the sampled function is given by the sequence**:**

$s(nT)$, for integer values of n.

The sampling frequency or sampling rate f_s is defined as the number of samples obtained in one second (samples per second), thus $f_s = 1/T$.

Although most of the signal is discarded by the sampling process, it is still generally possible to accurately reconstruct a signal from the samples if the signal is band-limited. A sufficient condition for perfect reconstruction is that the non-zero portion of the signal's Fourier transform be contained within the interval $[-f_s/2, f_s/2]$.

The frequency $f_s/2$ is called the Nyquist frequency of the sampling system. Without an anti-aliasing filter, frequencies higher than the Nyquist frequency will influence the samples in a way that is misinterpreted by the Whittaker–Shannon interpolation formula, the typical reconstruction formula. For details, see Aliasing.

Practical implications

In practice, the continuous signal is sampled using an analog-to-digital converter (ADC), a device with various physical limitations. This results in deviations from the theoretically perfect reconstruction, collectively referred to as distortion.

Various types of distortion can occur, including:

- Aliasing. A precondition of the sampling theorem is that the signal be bandlimited. However, in practice, no time-limited signal can be bandlimited. Since signals of interest are almost always time-limited (e.g., at most spanning the lifetime of the sampling device in question), it follows that they are not bandlimited. However, by designing a sampler with an appropriate guard band, it is possible to obtain output that is as accurate as necessary.
- Integration effect or aperture effect. This results from the fact that the sample is obtained as a time average within a sampling region, rather than just being equal to the signal value at the sampling instant. The integration effect is readily noticeable in photography when the exposure is too long and creates a blur in the image. An ideal camera would have an exposure time of zero. In a capacitor-based sample and hold circuit, the integration effect is introduced because the capacitor cannot instantly change voltage thus requiring the sample to have non-zero width.
- Jitter or deviation from the precise sample timing intervals.
- Noise, including thermal sensor noise, analog circuit noise, etc.
- Slew rate limit error, caused by an inability for an ADC output value to change sufficiently rapidly.
- Quantization as a consequence of the finite precision of words that represent the converted values.
- Error due to other non-linear effects of the mapping of input voltage to converted output value (in addition to the effects of quantization).

The conventional, practical digital-to-analog converter (DAC) does not output a sequence of dirac impulses (such that, if ideally low-pass filtered, result in the original signal before sampling) but instead output a sequence of piecewise constant values or rectangular pulses. This means that there is an inherent effect of the zero-order hold on the effective frequency response of the DAC resulting in a mild roll-off of gain at the higher frequencies (a 3.9224 dB loss at the Nyquist frequency). This zero-order hold effect is a consequence of the *hold* action of the DAC and is **not** due to the sample and hold that might precede a conventional ADC as is often misunderstood. The DAC can also suffer errors from jitter, noise, slewing, and non-linear mapping of input value to output voltage.

Jitter, noise, and quantization are often analyzed by modeling them as random errors added to the sample values. Integration and zero-order hold effects can be analyzed as a form of low-pass filtering. The non-linearities of either ADC or DAC are analyzed by replacing the ideal linear function mapping with a proposed nonlinear function.

Applications

Audio sampling

Digital audio uses pulse-code modulation and digital signals for sound reproduction. This includes analog-to-digital conversion (ADC), digital-to-analog conversion (DAC), storage, and transmission. In effect, the system commonly referred to as digital is in fact a discrete-time, discrete-level analog of a previous electrical analog. While modern systems can be quite subtle in their methods, the primary usefulness of a digital system is the ability to store, retrieve and transmit signals without any loss of quality.

Sampling rate

When it is necessary to capture audio covering the entire 20–20,000 Hz range of human hearing, such as when recording music or many types of acoustic events, audio waveforms are typically sampled at 44.1 kHz (CD), 48 kHz (professional audio), or 96 kHz. The approximately double-rate requirement is a consequence of the Nyquist theorem.

There has been an industry trend towards sampling rates well beyond the basic requirements; 96 kHz and even 192 kHz are available.[1] This is in contrast with laboratory experiments, which have failed to show that ultrasonic frequencies are audible to human observers; however in some cases ultrasonic sounds do interact with and modulate the audible part of the frequency spectrum (intermodulation distortion). It is noteworthy that intermodulation distortion is not present in the live audio and so it represents an artificial coloration to the live sound.[2]

One advantage of higher sampling rates is that they can relax the low-pass filter design requirements for ADCs and DACs, but with modern oversampling sigma-delta converters this advantage is less important.

Bit depth (quantization)

Audio is typically recorded at 8-, 16-, and 20-bit depth, which yield a theoretical maximum signal to quantization noise ratio (SQNR) for a pure sine wave of, approximately, 49.93 dB, 98.09 dB and 122.17 dB.[3] Eight-bit audio is generally not used due to prominent and inherent quantization noise (low maximum SQNR), although the A-law and u-law 8-bit encodings pack more resolution into 8 bits while increase total harmonic distortion. CD quality audio is recorded at 16-bit. In practice, not many consumer stereos can produce more than about 90 dB of dynamic range, although some can exceed 100 dB. Thermal noise limits the true number of bits that can be used in quantization. Few analog systems have signal to noise ratios (SNR) exceeding 120 dB; consequently, few situations will require more than 20-bit quantization.

For playback and not recording purposes, a proper analysis of typical programme levels throughout an audio system reveals that the capabilities of well-engineered 16-bit material far exceed those of the very best hi-fi systems, with the microphone noise and loudspeaker headroom being the real limiting factors.

Speech sampling

Speech signals, i.e., signals intended to carry only human speech, can usually be sampled at a much lower rate. For most phonemes, almost all of the energy is contained in the 5Hz-4 kHz range, allowing a sampling rate of 8 kHz. This is the sampling rate used by nearly all telephony systems, which use the G.711 sampling and quantization specifications.

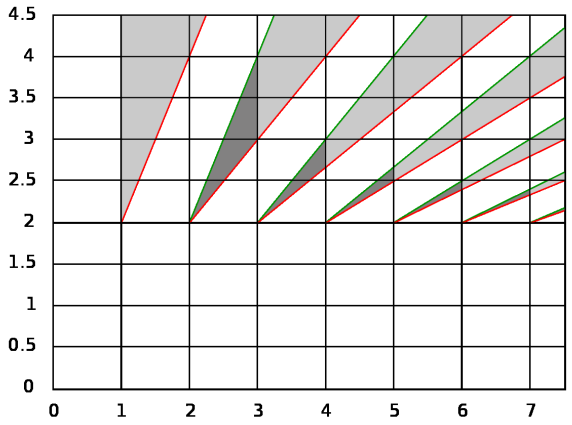
Video sampling

Standard-definition television (SDTV) uses either 720 by 480 pixels (US NTSC 525-line) or 704 by 576 pixels (UK PAL 625-line) for the visible picture area.

High-definition television (HDTV) is currently moving towards three standards referred to as 720p (progressive), 1080i (interlaced) and 1080p (progressive, also known as Full-HD) which all 'HD-Ready' sets will be able to display.

Undersampling

When one samples a bandpass signal at a rate lower than the Nyquist rate, the samples are equal to samples of a low-frequency alias of the high-frequency signal; the original signal will still be uniquely represented and recoverable if the spectrum of its alias does not cross over half the sampling rate. Such undersampling is also known as *bandpass sampling*, *harmonic sampling*, *IF sampling*, and *direct IF to digital conversion.*[4]



Plot of sample rates (y axis) versus the upper edge frequency (x axis) for a band of width 1; grays areas are combinations that are "allowed" in the sense that no two frequencies in the band alias to same frequency. The darker gray areas correspond to undersampling with the lowest allowable sample rate.

Oversampling

Oversampling is used in most modern analog-to-digital converters to reduce the distortion introduced by practical digital-to-analog converters, such as a zero-order hold instead of idealizations like the Whittaker–Shannon interpolation formula.

Complex sampling

Complex sampling refers to the simultaneous sampling of two different, but related, waveforms, resulting in pairs of samples that are subsequently treated as complex numbers. Usually one waveform , $\hat{s}(t)$, is the Hilbert transform of the other waveform , $s(t)$, and the complex-valued function, $s_a(t) \ \stackrel{\text{def}}{=}\ s(t) + j \cdot \hat{s}(t)$, is called an analytic signal, whose Fourier transform is zero for all negative values of frequency. In that case, the Nyquist rate for a waveform with no frequencies **≥ B** can be reduced to just *B* (complex samples/sec), instead of *2B* (real samples/sec).[5] More apparently, the equivalent baseband waveform, $s_a(t) \cdot e^{-j2\pi \frac{B}{2} t}$, also has a Nyquist rate of *B*, because all of its non-zero frequency content is shifted into the interval [-B/2, B/2).

Although complex-valued samples can be obtained as described above, they are much more commonly created by manipulating samples of a real-valued waveform. For instance, the equivalent baseband waveform can be created without explicitly computing $\hat{s}(t)$, by processing the product sequence , $\left[s(nT) \cdot e^{-j2\pi \frac{B}{2} Tn}\right]$,[6] through a digital lowpass filter whose cutoff frequency is B/2.[7] Computing only every other sample of the output sequence reduces the sample-rate commensurate with the reduced Nyquist rate. The result is half as many complex-valued samples as the original number of real samples. No information is lost, and the original s(t) waveform can be recovered, if necessary.

Notes

[1] Digital Pro Sound (http://www.digitalprosound.com/Htm/SoapBox/soap2_Apogee.htm)

[2] http://world.std.com/~griesngr/intermod.ppt

[3] MT-001: Taking the Mystery out of the Infamous Formula, "SNR=6.02N + 1.76dB," and Why You Should Care (http://www.analog.com/static/imported-files/tutorials/MT-001.pdf)

[4] Walt Kester (2003). *Mixed-signal and DSP design techniques* (http://books.google.com/books?id=G8XyNItpy8AC&pg=PA20). Newnes. p. 20. ISBN 978-0-7506-7611-3. .

[5] When the complex sample-rate is B, a frequency component at 0.6 B, for instance, will have an alias at -0.4 B, which is unambiguous because of the constraint that the pre-sampled signal was analytic. Also see Aliasing#Complex_signal_representation

[6] When s(t) is sampled at the Nyquist frequency (1/T = 2B), the product sequence simplifies to $\left[s(nT)\cdot(-j)^n\right]$.

[7] The sequence of complex numbers is convolved with the impulse response of a filter with real-valued coefficients. That is equivalent to separately filtering the sequences of real parts and imaginary parts and reforming complex pairs at the outputs.

References

- Matt Pharr and Greg Humphreys, *Physically Based Rendering: From Theory to Implementation*, Morgan Kaufmann, July 2004. ISBN 0-12-553180-X. The chapter on sampling (available online (http://graphics.stanford.edu/~mmp/chapters/pbrt_chapter7.pdf)) is nicely written with diagrams, core theory and code sample.
- Shannon, Claude E., Communications in the presence of noise, Proc. IRE, vol. 37, pp. 10–21, Jan. 1949.

External links

- Nyquist sampling in digital microscopy (http://www.vanosta.be/pcrnyq.htm)
- Journal devoted to Sampling Theory (http://www.stsip.org)

Nyquist–Shannon sampling theorem

The **Nyquist–Shannon sampling theorem**, after Harry Nyquist and Claude Shannon, is a fundamental result in the field of information theory, in particular telecommunications and signal processing. Sampling is the process of converting a signal (for example, a function of continuous time or space) into a numeric sequence (a function of discrete time or space). Shannon's version of the theorem states:[1]

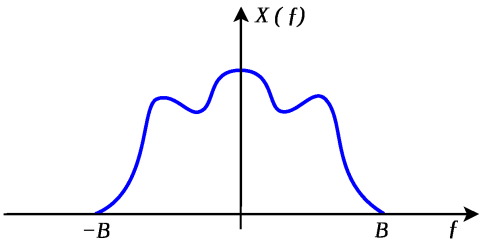


Fig. 1: Hypothetical spectrum of a bandlimited signal as a function of frequency

> If a function $x(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart.

The theorem is commonly called the **Nyquist sampling theorem**; since it was also discovered independently by E. T. Whittaker, by Vladimir Kotelnikov, and by others, it is also known as **Nyquist–Shannon–Kotelnikov**, **Whittaker–Shannon–Kotelnikov**, **Whittaker–Nyquist–Kotelnikov–Shannon**, **WKS**, etc., sampling theorem, as well as the **Cardinal Theorem of Interpolation Theory**. It is often referred to simply as **the sampling theorem**.

In essence, the theorem shows that a bandlimited analog signal can be perfectly reconstructed from an infinite sequence of samples if the sampling rate exceeds $2B$ samples per second, where B is the highest frequency of the original signal. If a signal contains a component at exactly B hertz, then samples spaced at exactly $1/(2B)$ seconds do not completely determine the signal, Shannon's statement notwithstanding. This sufficient condition can be weakened, as discussed at Sampling of non-baseband signals below.

More recent statements of the theorem are sometimes careful to exclude the equality condition; that is, the condition is if $x(t)$ contains no frequencies higher than *or equal to* B; this condition is equivalent to Shannon's except when the

function includes a steady sinusoidal component at exactly frequency *B*.

The theorem assumes an idealization of any real-world situation, as it only applies to signals that are sampled for infinite time; any time-limited *x*(*t*) cannot be perfectly bandlimited. Perfect reconstruction is mathematically possible for the idealized model but only an approximation for real-world signals and sampling techniques, albeit in practice often a very good one.

The theorem also leads to a formula for reconstruction of the original signal. The constructive proof of the theorem leads to an understanding of the aliasing that can occur when a sampling system does not satisfy the conditions of the theorem.

The sampling theorem provides a sufficient condition, but not a necessary one, for perfect reconstruction. The field of compressed sensing provides a stricter sampling condition when the underlying signal is known to be sparse. Compressed sensing specifically yields a sub-Nyquist sampling criterion.

Introduction

A signal or function is bandlimited if it contains no energy at frequencies higher than some bandlimit or bandwidth *B*. The sampling theorem asserts that, given such a bandlimited signal, the uniformly spaced discrete samples are a complete representation of the signal as long as the sampling rate is larger than twice the bandwidth *B*. To formalize these concepts, let *x*(*t*) represent a continuous-time signal and *X*(*f*) be the continuous Fourier transform of that signal:

$$X(f) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x(t)\ e^{-i2\pi ft}\ \mathrm{d}t.$$

The signal *x*(*t*) is said to be bandlimited to a one-sided baseband bandwidth, *B*, if

$$X(f) = 0 \ \text{ for all } \ |f| > B, \quad \text{or, equivalently, } \operatorname{supp}(X) \subseteq [-B, B].^{[2]}$$

Then the sufficient condition for exact reconstructability from samples at a uniform sampling rate f_s (in samples per unit time) is:

$$f_s > 2B.$$

The quantity 2*B* is called the *Nyquist rate* and is a property of the bandlimited signal, while $f_s/2$ is called the *Nyquist frequency* and is a property of this sampling system.

The time interval between successive samples is referred to as the *sampling interval*:

$$T \stackrel{\text{def}}{=} \frac{1}{f_s},$$

and the samples of *x*(*t*) are denoted by *x*(*nT*) for integer values of *n*. The sampling theorem leads to a procedure for reconstructing the original *x*(*t*) from the samples and states sufficient conditions for such a reconstruction to be exact.

The sampling process

The theorem describes two processes in signal processing: a sampling process, in which a continuous time signal is converted to a discrete time signal, and a reconstruction process, in which the original continuous signal is recovered from the discrete time signal.

The continuous signal varies over *time* (or *space* in a digitized image, or another independent variable in some other application) and the sampling process is performed by measuring the continuous signal's value every *T* units of time (or space), which is called the *sampling interval*. Sampling results in a sequence of numbers, called *samples*, to represent the original signal. Each sample value is associated with the instant in time when it was measured. The reciprocal of the sampling interval (1/*T*) is the sampling frequency denoted f_s, which is measured in samples per unit of time. If *T* is expressed in seconds, then f_s is expressed in hertz.

Reconstruction

Reconstruction of the original signal is an interpolation process that mathematically defines a continuous-time signal $x(t)$ from the discrete samples $x(nT)$ and at times in between the sample instants nT.

- **The procedure:** Each sample value is multiplied by the sinc function scaled so that the zero-crossings of the sinc function occur at the sampling instants and that the sinc function's central point is shifted to the time of that sample, nT. All of these shifted and scaled functions are then added together to recover the original signal. The scaled and time-shifted sinc functions are continuous making the sum of these also continuous, so the result of this operation is a continuous signal. This procedure is represented by the Whittaker–Shannon interpolation formula.

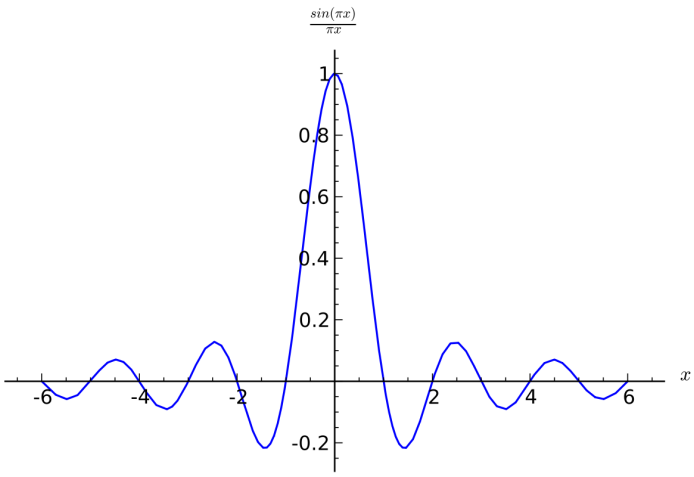


Fig. 2: The normalized sinc function: sin(πx) / (πx) ... showing the central peak at x= 0, and zero-crossings at the other integer values of x.

- **The condition:** The signal obtained from this reconstruction process cannot have any frequencies higher than one-half the sampling frequency. According to the theorem, the reconstructed signal will match the original signal provided that the original signal contains no frequencies at or above this limit. This condition is called the *Nyquist criterion*, or sometimes the *Raabe condition.*

If the original signal contains a frequency component equal to one-half the sampling rate, the condition is not satisfied. The resulting reconstructed signal may have a component at that frequency, but the amplitude and phase of that component generally will not match the original component.

This reconstruction or interpolation using sinc functions is not the only interpolation scheme. Indeed, it is impossible in practice because it requires summing an infinite number of terms. However, it is the interpolation method that in theory exactly reconstructs *any* given bandlimited $x(t)$ with *any* bandlimit $B < 1/(2T)$; any other method that does so is formally equivalent to it.

Practical considerations

A few consequences can be drawn from the theorem:

- If the highest frequency B in the original signal is known, the theorem gives the lower bound on the sampling frequency for which perfect reconstruction can be assured. This lower bound to the sampling frequency, $2B$, is called the Nyquist rate.
- If instead the sampling frequency is known, the theorem gives us an upper bound for frequency components, $B<f_s/2$, of the signal to allow for perfect reconstruction. This upper bound is the Nyquist frequency, denoted f_N.
- Both of these cases imply that the signal to be sampled must be bandlimited; that is, any component of this signal which has a frequency above a certain bound should be zero, or at least sufficiently close to zero to allow us to neglect its influence on the resulting reconstruction. In the first case, the condition of bandlimitation of the sampled signal can be accomplished by assuming a model of the signal which can be analysed in terms of the frequency components it contains; for example, sounds that are made by a speaking human normally contain very small frequency components at or above 10 kHz and it is then sufficient to sample such an audio signal with a sampling frequency of at least 20 kHz. For the second case, we have to assure that the sampled signal is

bandlimited such that frequency components at or above half of the sampling frequency can be neglected. This is usually accomplished by means of a suitable low-pass filter; for example, if it is desired to sample speech waveforms at 8 kHz, the signals should first be lowpass filtered to below 4 kHz.

- In practice, neither of the two statements of the sampling theorem described above can be completely satisfied, and neither can the reconstruction formula be precisely implemented. The reconstruction process that involves scaled and delayed sinc functions can be described as *ideal*. It cannot be realized in practice since it implies that each sample contributes to the reconstructed signal at almost all time points, requiring summing an infinite number of terms. Instead, some type of approximation of the sinc functions, finite in length, has to be used. The error that corresponds to the sinc-function approximation is referred to as *interpolation error*. Practical digital-to-analog converters produce neither scaled and delayed sinc functions nor ideal impulses (that if ideally low-pass filtered would yield the original signal), but a sequence of scaled and delayed rectangular pulses. This practical piecewise-constant output can be modeled as a zero-order hold filter driven by the sequence of scaled and delayed dirac impulses referred to in the mathematical basis section below. A shaping filter is sometimes used after the DAC with zero-order hold to make a better overall approximation.
- Furthermore, in practice, a signal can never be perfectly bandlimited, since ideal "brick-wall" filters cannot be realized. All practical filters can only attenuate frequencies outside a certain range, not remove them entirely. In addition to this, a "time-limited" signal can never be bandlimited. This means that even if an ideal reconstruction could be made, the reconstructed signal would not be exactly the original signal. The error that corresponds to the failure of bandlimitation is referred to as *aliasing*.
- The sampling theorem does not say what happens when the conditions and procedures are not exactly met, but its proof suggests an analytical framework in which the non-ideality can be studied. A designer of a system that deals with sampling and reconstruction processes needs a thorough understanding of the signal to be sampled, in particular its frequency content, the sampling frequency, how the signal is reconstructed in terms of interpolation, and the requirement for the total reconstruction error, including aliasing, sampling, interpolation and other errors. These properties and parameters may need to be carefully tuned in order to obtain a useful system.

Aliasing

The Poisson summation formula shows that the samples, $x(nT)$, of function $x(t)$ are sufficient to create a periodic summation of function $X(f)$. The result is:

$$X_s(f) \ \stackrel{\text{def}}{=} \ \sum_{k=-\infty}^{\infty} X\left(f - kf_s\right) = \sum_{n=-\infty}^{\infty} \underbrace{T \cdot x(nT)}_{x[n]} \ e^{-i2\pi nTf}, \qquad \textbf{(Eq.1)}$$

which is a periodic function and its equivalent representation as a Fourier series, whose coefficients are $x[n]$. This function is also known as the discrete-time Fourier transform (DTFT). As depicted in Figures 3, 4, and 8, copies of $X(f)$ are shifted by multiples of f_s and combined by addition.

If the Nyquist sampling condition is not satisfied, adjacent copies overlap, and it is not possible in general to discern an unambiguous $X(f)$. Any frequency component above $f_s/2$ is indistinguishable from a lower-frequency component, called an *alias*, associated with one of the

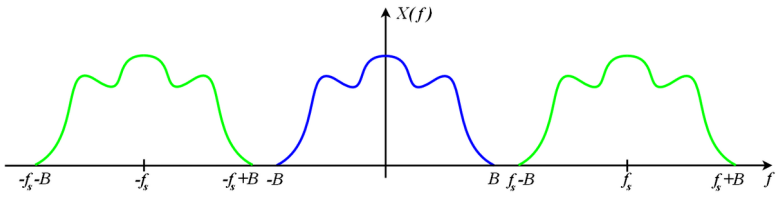


Fig. 3: Illustration of the spectrum (blue) of a properly sampled bandlimited function and the adjacent DTFT images (green) that do not overlap. A "*brick-wall*" low-pass filter can remove the images and leave the original spectrum, X(f), thus also recovering the original x(t) from just its samples.

copies. In such cases, the reconstruction technique described below produces the alias, rather than the original component.

For a sinusoidal component of exactly half the sampling frequency, the component will in general alias to another sinusoid of the same frequency, but with a different phase and amplitude.

To prevent or reduce aliasing, two things can be done:

1. Increase the sampling rate, to above twice some or all of the frequencies that are aliasing.
2. Introduce an anti-aliasing filter or make the anti-aliasing filter more stringent.

The anti-aliasing filter restricts the bandwidth of $x(t)$ to satisfy the Nyquist sampling criterion. Such a restriction works in theory but is not precisely realizable, because realizable filters will always allow some *leakage* of high frequencies. However, the leakage energy can be made small enough so that the aliasing effects are negligible.

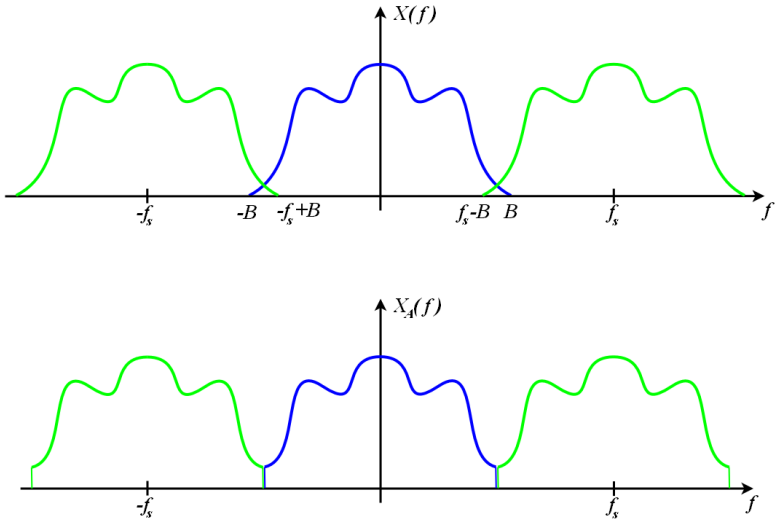


Fig. 4, top: Illustration of the spectrum (blue) of an under-sampled, bandlimited function, $x(t)$, where the adjacent DTFT images (green) overlap. These overlapping edges or "*tails*" of the images add, creating a DTFT from which $X(f)$ is no longer exactly discernible. Bottom: Illustration of the spectrum (blue) of a critically sampled bandlimited function, $x_A(t)$, where the DTFT images (green) narrowly do not overlap. But the resultant DTFT is the same as the top figure, because the sum of baseband and images are the same in both cases. The samples of both functions are also indistinguishable. When they are used to reconstruct a continuous-time function, in the normal way, the result will be $x_A(t)$, not $x(t)$. And that function is called an *alias* of $x(t)$, for the particular sampling rate, f_s.

Application to multivariable signals and images

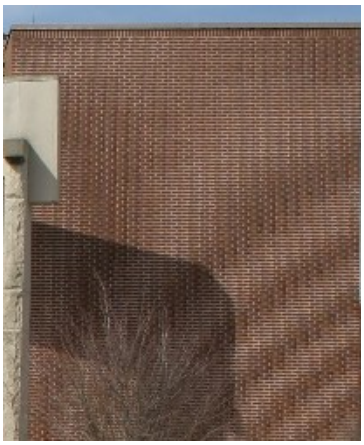


Fig. 5: Subsampled image showing a Moiré pattern

The sampling theorem is usually formulated for functions of a single variable. Consequently, the theorem is directly applicable to time-dependent signals and is normally formulated in that context. However, the sampling theorem can be extended in a straightforward way to functions of arbitrarily many variables. Grayscale images, for example, are often represented as two-dimensional arrays (or matrices) of real numbers representing the relative intensities of pixels (picture elements) located at the intersections of row and column sample locations. As a result, images require two independent variables, or indices, to specify each pixel uniquely — one for the row, and one for the column.

Color images typically consist of a composite of three separate grayscale images, one to represent each of the three primary colors — red, green, and blue, or *RGB* for short. Other colorspaces using 3-vectors for colors include HSV, LAB, XYZ, etc. Some colorspaces such as cyan, magenta, yellow, and black (CMYK) may represent color by four dimensions. All of these are treated as vector-valued functions over a two-dimensional sampled domain.

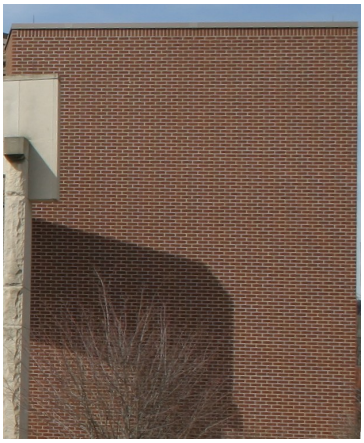


Fig. 6

Similar to one-dimensional discrete-time signals, images can also suffer from aliasing if the sampling resolution, or pixel density, is inadequate. For example, a digital photograph of a striped shirt with high frequencies (in other words, the distance between the stripes is small), can cause aliasing of the shirt when it is sampled by the camera's image sensor. The aliasing appears as a moiré pattern. The "solution" to higher sampling in the spatial domain for this case would be to move closer to the shirt, use a higher resolution sensor, or to optically blur the image before acquiring it with the sensor.

Another example is shown to the left in the brick patterns. The top image shows the effects when the sampling theorem's condition is not satisfied. When software rescales an image (the same process that creates the thumbnail shown in the lower image) it, in effect, runs the image through a low-pass filter first and then downsamples the image to result in a smaller image that does not exhibit the moiré pattern. The top image is what happens when the image is downsampled without low-pass filtering: aliasing results.

The application of the sampling theorem to images should be made with care. For example, the sampling process in any standard image sensor (CCD or CMOS camera) is relatively far from the ideal sampling which would measure the image intensity at a single point. Instead these devices have a relatively large sensor area at each sample point in order to obtain sufficient amount of light. In other words, any detector has a finite-width point spread function. The analog optical image intensity function which is sampled by the sensor device is not in general bandlimited, and the non-ideal sampling is itself a useful type of low-pass filter, though not always sufficient to remove enough high frequencies to sufficiently reduce aliasing. When the area of the sampling spot (the size of the pixel sensor) is not large enough to provide sufficient anti-aliasing, a separate anti-aliasing filter (optical low-pass filter) is typically included in a camera system to further blur the optical image. Despite images having these problems in relation to the sampling theorem, the theorem can be used to describe the basics of down and up sampling of images.

Downsampling

When a signal is downsampled, the sampling theorem can be invoked via the artifice of resampling a hypothetical continuous-time reconstruction. The Nyquist criterion must still be satisfied with respect to the new lower sampling frequency in order to avoid aliasing. To meet the requirements of the theorem, the signal must usually pass through a low-pass filter of appropriate cutoff frequency as part of the downsampling operation. This low-pass filter, which prevents aliasing, is called an anti-aliasing filter.

Critical frequency

To illustrate the necessity of $f_s > 2B$, consider the sinusoid:

$$x(t) = \cos(2\pi Bt + \theta) \ = \ \cos(2\pi Bt)\cos(\theta) - \sin(2\pi Bt)\sin(\theta).$$

With $f_s = 2B$ or equivalently $T = 1/(2B)$, the samples are given by:

$$x(nT) = \cos(\pi n)\cos(\theta) - \underbrace{\sin(\pi n)}_{0}\sin(\theta) = \cos(\pi n)\cos(\theta).$$

Those samples cannot be distinguished from the samples of:

$$x_A(t) = \cos(2\pi Bt)\cos(\theta).$$

But for any θ such that |cos(θ)| < 1, $x(t)$ and $x_A(t)$ have different amplitudes and different phase. Ambiguities such as that are the reason for the *strict* inequality of the sampling theorem's condition.

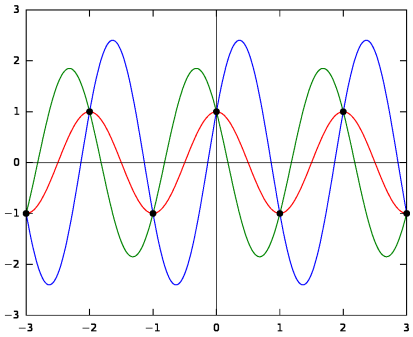


Fig. 7: A family of sinusoids at the critical frequency, all having the same sample sequences of alternating +1 and −1. That is, they all are aliases of each other, even though their frequency is not above half the sample rate.

Mathematical reasoning for the theorem

From Figures 3 and 8, it is apparent that when there is no overlap of the copies (aka "images") of $X(f)$, the $k = 0$ term of $X_s(f)$ can be recovered by the product:

$$X(f) = H(f) \cdot X_s(f),$$

where:

$$H(f) = \begin{cases} 1 & |f| < B \\ 0 & |f| > f_s - B. \end{cases}$$

$H(f)$ need not be precisely defined in the region $[B, f_s - B]$ because $X_s(f)$ is zero in that region. However, the worst case is when $B = f_s/2$, the Nyquist frequency. A function that is sufficient for that and all less severe cases is:

$$H(f) = \operatorname{rect}\left(\frac{f}{f_s}\right) = \begin{cases} 1 & |f| < \frac{f_s}{2} \\ 0 & |f| > \frac{f_s}{2}, \end{cases}$$

where rect(u) is the rectangular function.

Therefore:

$$\begin{aligned} X(f) &= \operatorname{rect}\left(\frac{f}{f_s}\right) \cdot X_s(f) \\ &= \operatorname{rect}(Tf) \cdot T \sum_{n=-\infty}^{\infty} x(nT)\ e^{-i2\pi nTf} \qquad \text{(from } \textbf{Eq.1}\text{, above).} \\ &= T \sum_{n=-\infty}^{\infty} x(nT) \cdot \operatorname{rect}(Tf) \cdot e^{-i2\pi nTf}. \end{aligned}$$

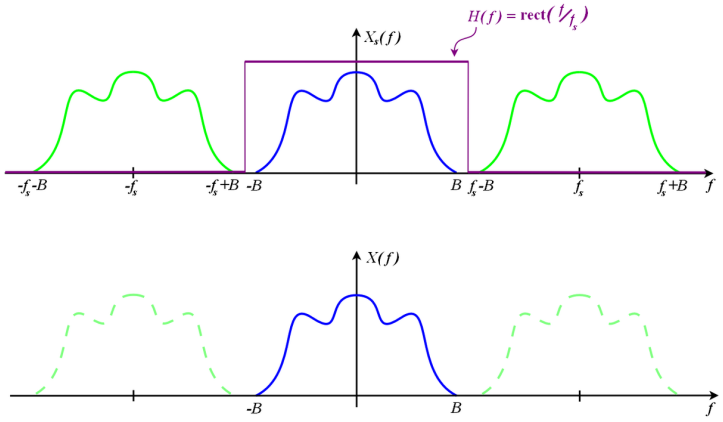


Fig. 8: Spectrum, $X_s(f)$, of a properly sampled bandlimited signal (blue) and images (green) that do not overlap. A *brick-wall* low-pass filter, $H(f)$, removes the images, leaves the original spectrum, $X(f)$, and recovers the original signal from the samples.

The original function that was sampled can be recovered by an inverse Fourier transform:

$$\begin{aligned} x(t) &= \mathcal{F}^{-1}\left\{ T \sum_{n=-\infty}^{\infty} x(nT) \cdot \operatorname{rect}(Tf) \cdot e^{-i2\pi nTf} \right\} \\ &= T \sum_{n=-\infty}^{\infty} x(nT) \cdot \underbrace{\mathcal{F}^{-1}\left\{ \operatorname{rect}(Tf) \cdot e^{-i2\pi nTf} \right\}}_{\frac{1}{T}\cdot\operatorname{sinc}\left(\frac{t-nT}{T}\right)} \quad \text{[3]} \\ &= \sum_{n=-\infty}^{\infty} x(nT) \cdot \operatorname{sinc}\left(\frac{t-nT}{T}\right), \end{aligned}$$

which is the Whittaker–Shannon interpolation formula. It shows explicitly how the samples, $x(nT)$, can be combined to reconstruct $x(t)$.

- From Figure 8, it is clear that larger-than-necessary values of f_s (smaller values of T), called *oversampling*, have no effect on the outcome of the reconstruction and have the benefit of leaving room for a *transition band* in which $H(f)$ is free to take intermediate values. Undersampling, which causes aliasing, is not in general a reversible operation.
- Theoretically, the interpolation formula can be implemented as a low pass filter, whose impulse response is $\operatorname{sinc}(t/T)$ and whose input is $\sum_{n=-\infty}^{\infty} x(nT) \cdot \delta(t-nT)$, which is a Dirac comb function modulated by the signal samples. Practical digital-to-analog converters (DAC) implement an approximation like the zero-order hold. In that case, oversampling can reduce the approximation error.

Shannon's original proof

The original proof presented by Shannon is elegant and quite brief, but it offers less intuitive insight into the subtleties of aliasing, both unintentional and intentional. Quoting Shannon's original paper, which uses f for the function, F for the spectrum, and W for the bandwidth limit:

> Let $F(\omega)$ be the spectrum of $f(t)$. Then
>
> $$\begin{aligned} f(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega) e^{i\omega t}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} F(\omega) e^{i\omega t}\,\mathrm{d}\omega \end{aligned}$$
>
> since $F(\omega)$ is assumed to be zero outside the band W. If we let
>
> $$t = \frac{n}{2W}$$
>
> where n is any positive or negative integer, we obtain
>
> $$f\left(\frac{n}{2W}\right) = \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} F(\omega) e^{i\omega \frac{n}{2W}}\,\mathrm{d}\omega.$$
>
> On the left are values of $f(t)$ at the sampling points. The integral on the right will be recognized as essentially the n^{th} coefficient in a Fourier-series expansion of the function $F(\omega)$, taking the interval $-W$ to W as a fundamental period. This means that the values of the samples $f(n/2W)$ determine the Fourier coefficients in the series expansion of $F(\omega)$. Thus they determine $F(\omega)$, since $F(\omega)$ is zero for frequencies greater than W, and for lower frequencies $F(\omega)$ is determined if its Fourier coefficients are determined. But $F(\omega)$ determines the original function $f(t)$ completely, since a function is determined if its spectrum is known. Therefore the original samples determine the function $f(t)$ completely.

Shannon's proof of the theorem is complete at that point, but he goes on to discuss reconstruction via sinc functions, what we now call the Whittaker–Shannon interpolation formula as discussed above. He does not derive or prove the

properties of the sinc function, but these would have been familiar to engineers reading his works at the time, since the Fourier pair relationship between rect (the rectangular function) and sinc was well known. Quoting Shannon:

> Let x_n be the n^{th} sample. Then the function $f(t)$ is represented by:
>
> $$f(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)}.$$

As in the other proof, the existence of the Fourier transform of the original signal is assumed, so the proof does not say whether the sampling theorem extends to bandlimited stationary random processes.

Sampling of non-baseband signals

As discussed by Shannon:[1]

> A similar result is true if the band does not start at zero frequency but at some higher value, and can be proved by a linear translation (corresponding physically to single-sideband modulation) of the zero-frequency case. In this case the elementary pulse is obtained from sin(*x*)/*x* by single-side-band modulation.

That is, a sufficient no-loss condition for sampling signals that do not have baseband components exists that involves the *width* of the non-zero frequency interval as opposed to its highest frequency component. See *Sampling (signal processing)* for more details and examples.

A bandpass condition is that $X(f) = 0$, for all nonnegative f outside the open band of frequencies:

$$\left(\frac{N}{2} f_{\mathrm{s}}, \frac{N+1}{2} f_{\mathrm{s}}\right),$$

for some nonnegative integer N. This formulation includes the normal baseband condition as the case N=0.

The corresponding interpolation function is the impulse response of an ideal brick-wall bandpass filter (as opposed to the ideal brick-wall lowpass filter used above) with cutoffs at the upper and lower edges of the specified band, which is the difference between a pair of lowpass impulse responses:

$$(N+1)\operatorname{sinc}\left(\frac{(N+1)t}{T}\right) - N\operatorname{sinc}\left(\frac{Nt}{T}\right).$$

Other generalizations, for example to signals occupying multiple non-contiguous bands, are possible as well. Even the most generalized form of the sampling theorem does not have a provably true converse. That is, one cannot conclude that information is necessarily lost just because the conditions of the sampling theorem are not satisfied; from an engineering perspective, however, it is generally safe to assume that if the sampling theorem is not satisfied then information will most likely be lost.

Nonuniform sampling

The sampling theory of Shannon can be generalized for the case of nonuniform samples, that is, samples not taken equally spaced in time. The Shannon sampling theory for non-uniform sampling states that a band-limited signal can be perfectly reconstructed from its samples if the average sampling rate satisfies the Nyquist condition.[4] Therefore, although uniformly spaced samples may result in easier reconstruction algorithms, it is not a necessary condition for perfect reconstruction.

The general theory for non-baseband and nonuniform samples was developed in 1967 by Landau.[5] He proved that, to paraphrase roughly, the average sampling rate (uniform or otherwise) must be twice the *occupied* bandwidth of the signal, assuming it is *a priori* known what portion of the spectrum was occupied. In the late 1990s, this work was partially extended to cover signals of when the amount of occupied bandwidth was known, but the actual occupied portion of the spectrum was unknown.[6] In the 2000s, a complete theory was developed (see the section Beyond Nyquist below) using compressed sensing. In particular, the theory, using signal processing language, is described in

this 2009 paper.[7] They show, among other things, that if the frequency locations are unknown, then it is necessary to sample at least at twice the Nyquist criteria; in other words, you must pay at least a factor of 2 for not knowing the location of the spectrum. Note that minimum sampling requirements do not necessarily guarantee stability.

Beyond Nyquist

The Nyquist–Shannon sampling theorem provides a sufficient condition for the sampling and reconstruction of a band-limited signal. When reconstruction is done via the Whittaker–Shannon interpolation formula, the Nyquist criterion is also a necessary condition to avoid aliasing, in the sense that if samples are taken at a slower rate than twice the band limit, then there are some signals that will not be correctly reconstructed. However, if further restrictions are imposed on the signal, then the Nyquist criterion may no longer be a necessary condition.

A non-trivial example of exploiting extra assumptions about the signal is given by the recent field of compressed sensing, which allows for full reconstruction with a sub-Nyquist sampling rate. Specifically, this applies to signals that are sparse (or compressible) in some domain. As an example, compressed sensing deals with signals that may have a low over-all bandwidth (say, the *effective* bandwidth *EB*), but the frequency locations are unknown, rather than all together in a single band, so that the passband technique doesn't apply. In other words, the frequency spectrum is sparse. Traditionally, the necessary sampling rate is thus 2*B*. Using compressed sensing techniques, the signal could be perfectly reconstructed if it is sampled at a rate slightly greater than the 2*EB*. The downside of this approach is that reconstruction is no longer given by a formula, but instead by the solution to a convex optimization program which requires well-studied but nonlinear methods.

Historical background

The **sampling theorem** was implied by the work of Harry Nyquist in 1928 ("Certain topics in telegraph transmission theory"), in which he showed that up to 2*B* independent pulse samples could be sent through a system of bandwidth *B*; but he did not explicitly consider the problem of sampling and reconstruction of continuous signals. About the same time, Karl Küpfmüller showed a similar result,[8] and discussed the sinc-function impulse response of a band-limiting filter, via its integral, the step response *Integralsinus*; this bandlimiting and reconstruction filter that is so central to the sampling theorem is sometimes referred to as a *Küpfmüller filter* (but seldom so in English).

The sampling theorem, essentially a dual of Nyquist's result, was proved by Claude E. Shannon in 1949 ("Communication in the presence of noise"). V. A. Kotelnikov published similar results in 1933 ("On the transmission capacity of the 'ether' and of cables in electrical communications", translation from the Russian), as did the mathematician E. T. Whittaker in 1915 ("Expansions of the Interpolation-Theory", "Theorie der Kardinalfunktionen"), J. M. Whittaker in 1935 ("Interpolatory function theory"), and Gabor in 1946 ("Theory of communication").

Other discoverers

Others who have independently discovered or played roles in the development of the sampling theorem have been discussed in several historical articles, for example by Jerri[9] and by Lüke.[10] For example, Lüke points out that H. Raabe, an assistant to Küpfmüller, proved the theorem in his 1939 Ph.D. dissertation; the term *Raabe condition* came to be associated with the criterion for unambiguous representation (sampling rate greater than twice the bandwidth).

Meijering[11] mentions several other discoverers and names in a paragraph and pair of footnotes:

> As pointed out by Higgins [135], the sampling theorem should really be considered in two parts, as done above: the first stating the fact that a bandlimited function is completely determined by its samples, the second describing how to reconstruct the function using its samples. Both parts of the sampling theorem were given in a somewhat different form by J. M. Whittaker [350, 351, 353] and before him also by Ogura [241, 242]. They were probably not aware of the fact that the first part of the theorem had been

stated as early as 1897 by Borel [25].[27] As we have seen, Borel also used around that time what became known as the cardinal series. However, he appears not to have made the link [135]. In later years it became known that the sampling theorem had been presented before Shannon to the Russian communication community by Kotel'nikov [173]. In more implicit, verbal form, it had also been described in the German literature by Raabe [257]. Several authors [33, 205] have mentioned that Someya [296] introduced the theorem in the Japanese literature parallel to Shannon. In the English literature, Weston [347] introduced it independently of Shannon around the same time.[28]

[27] Several authors, following Black [16], have claimed that this first part of the sampling theorem was stated even earlier by Cauchy, in a paper [41] published in 1841. However, the paper of Cauchy does not contain such a statement, as has been pointed out by Higgins [135].

[28] As a consequence of the discovery of the several independent introductions of the sampling theorem, people started to refer to the theorem by including the names of the aforementioned authors, resulting in such catchphrases as "the Whittaker-Kotel'nikov-Shannon (WKS) sampling theorem" [155] or even "the Whittaker-Kotel'nikov-Raabe-Shannon-Someya sampling theorem" [33]. To avoid confusion, perhaps the best thing to do is to refer to it as the sampling theorem, "rather than trying to find a title that does justice to all claimants" [136].

Why Nyquist?

Exactly how, when, or why Harry Nyquist had his name attached to the sampling theorem remains obscure. The term *Nyquist Sampling Theorem* (capitalized thus) appeared as early as 1959 in a book from his former employer, Bell Labs,[12] and appeared again in 1963,[13] and not capitalized in 1965.[14] It had been called the *Shannon Sampling Theorem* as early as 1954,[15] but also just *the sampling theorem* by several other books in the early 1950s.

In 1958, Blackman and Tukey[16] cited Nyquist's 1928 paper as a reference for *the sampling theorem of information theory*, even though that paper does not treat sampling and reconstruction of continuous signals as others did. Their glossary of terms includes these entries:

> *Sampling theorem* (*of information theory*)
>
> Nyquist's result that equi-spaced data, with two or more points per cycle of highest frequency, allows reconstruction of band-limited functions. (See *Cardinal theorem.*)
>
> *Cardinal theorem* (*of interpolation theory*)
>
> A precise statement of the conditions under which values given at a doubly infinite set of equally spaced points can be interpolated to yield a continuous band-limited function with the aid of the function
>
> $$\frac{\sin(x - x_i)}{x - x_i}.$$

Exactly what "Nyquist's result" they are referring to remains mysterious.

When Shannon stated and proved the sampling theorem in his 1949 paper, according to Meijering[11] "he referred to the critical sampling interval $T = 1/(2W)$ as the *Nyquist interval* corresponding to the band W, in recognition of Nyquist's discovery of the fundamental importance of this interval in connection with telegraphy." This explains Nyquist's name on the critical interval, but not on the theorem.

Similarly, Nyquist's name was attached to *Nyquist rate* in 1953 by Harold S. Black:[17]

> "If the essential frequency range is limited to B cycles per second, $2B$ was given by Nyquist as the maximum number of code elements per second that could be unambiguously resolved, assuming the peak interference is less half a quantum step. This rate is generally referred to as **signaling at the Nyquist rate** and $1/(2B)$ has been termed a *Nyquist interval*." (bold added for emphasis; italics as in the original)

According to the OED, this may be the origin of the term *Nyquist rate*. In Black's usage, it is not a sampling rate, but a signaling rate.

Notes

[1] C. E. Shannon, "Communication in the presence of noise", Proc. Institute of Radio Engineers, vol. 37, no. 1, pp. 10–21, Jan. 1949. Reprint as classic paper in: *Proc. IEEE*, vol. 86, no. 2, (Feb. 1998) (http://www.stanford.edu/class/ee104/shannonpaper.pdf)

[2] supp(X) means the support of X.

[3] The time-domain form follows from rows 202 and 102 of the transform tables

[4] Nonuniform Sampling, Theory and Practice (ed. F. Marvasti), Kluwer Academic/Plenum Publishers, New York, 2000

[5] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," Acta Math., vol. 117, pp.37–52, Feb. 1967.

[6] see, e.g., P. Feng, "Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1997.

[7] Blind Multiband Signal Reconstruction: Compressed Sensing for Analog Signals (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.154.4255), Moshe Mishali and Yonina C. Eldar, in **IEEE Trans. Signal Processing**, March 2009, Vol 57 Issue 3

[8] K. Küpfmüller, "Über die Dynamik der selbsttätigen Verstärkungsregler", *Elektrische Nachrichtentechnik*, vol. 5, no. 11, pp. 459-467, 1928. (German)
K. Küpfmüller, On the dynamics of automatic gain controllers (http://ict.open.ac.uk/classics/2.pdf), *Elektrische Nachrichtentechnik*, vol. 5, no. 11, pp. 459-467, 1928 (English translation 2005).

[9] Abdul Jerri, The Shannon Sampling Theorem—Its Various Extensions and Applications: A Tutorial Review (http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1455040), *Proceedings of the IEEE*, 65:1565–1595, Nov. 1977. See also Correction to "The Shannon sampling theorem—Its various extensions and applications: A tutorial review" (http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1455576), Proceedings of the IEEE, 67:695, April 1979

[10] Hans Dieter Lüke, The Origins of the Sampling Theorem (doi:10.1109/35.755459), *IEEE Communications Magazine,* pp.106–108, April 1999.

[11] Erik Meijering, A Chronology of Interpolation From Ancient Astronomy to Modern Signal and Image Processing (doi:10.1109/5.993400), *Proc. IEEE,* 90, 2002.

[12] Members of the Technical Staff of Bell Telephone Lababoratories (1959). *Transmission Systems for Communications*. AT&T. pp. 26–4 (Vol.2).

[13] Ernst Adolph Guillemin (1963). *Theory of Linear Physical Systems* (http://books.google.com/books?id=jtI-AAAAIAAJ&q=Nyquist-sampling-theorem+date:0-1965&dq=Nyquist-sampling-theorem+date:0-1965&as_brr=0&ei=ief5Ru-DJqegowLyiJTpBw&pgis=1). Wiley. .

[14] Richard A. Roberts and Ben F. Barton, *Theory of Signal Detectability: Composite Deferred Decision Theory*, 1965.

[15] Truman S. Gray, *Applied Electronics: A First Course in Electronics, Electron Tubes, and Associated Circuits*, 1954.

[16] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra : From the Point of View of Communications Engineering*, New York: Dover, 1958.

[17] Harold S. Black, *Modulation Theory,* 1953

References

- J. R. Higgins: *Five short stories about the cardinal series*, Bulletin of the AMS 12(1985)
- V. A. Kotelnikov, "On the carrying capacity of the ether and wire in telecommunications", Material for the First All-Union Conference on Questions of Communication, Izd. Red. Upr. Svyazi RKKA, Moscow, 1933 (Russian). (english translation, PDF) (http://ict.open.ac.uk/classics/1.pdf)
- Karl Küpfmüller, "Utjämningsförlopp inom Telegraf- och Telefontekniken", ("Transients in telegraph and telephone engineering"), Teknisk Tidskrift, no. 9 pp. 153–160 and 10 pp. 178–182, 1931. (http://runeberg.org/tektid/1931e/0157.html) (http://runeberg.org/tektid/1931e/0182.html)
- R.J. Marks II: *Introduction to Shannon Sampling and Interpolation Theory* (http://marksmannet.com/RobertMarks/REPRINTS/1999_IntroductionToShannonSamplingAndInterpolationTheory.pdf), Spinger-Verlag, 1991.
- R.J. Marks II, Editor: Advanced Topics in Shannon Sampling and Interpolation Theory (http://marksmannet.com/RobertMarks/REPRINTS/1993_AdvancedTopicsOnShannon.pdf), Springer-Verlag, 1993.
- R.J. Marks II, *Handbook of Fourier Analysis and Its Applications,* Oxford University Press, (2009), Chapters 5-8. Google books (http://books.google.com/books?id=Sp7O4bocjPAC).

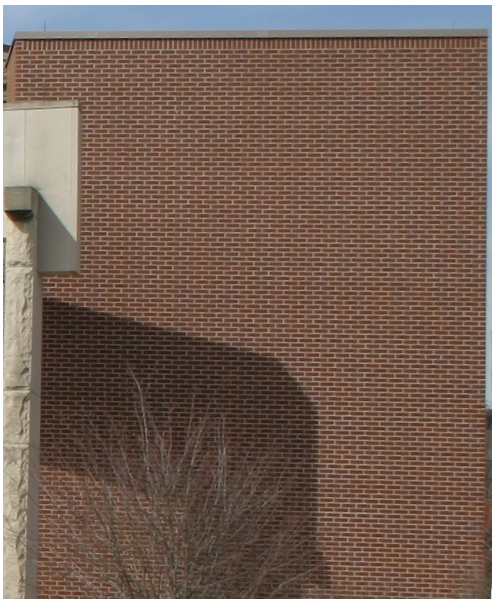
- H. Nyquist, "Certain topics in telegraph transmission theory", Trans. AIEE, vol. 47, pp. 617–644, Apr. 1928 Reprint as classic paper in: *Proc. IEEE, Vol. 90, No. 2, Feb 2002* (http://replay.web.archive.org/20060706192816/http://www.loe.ee.upatras.gr/Comes/Notes/Nyquist.pdf).
- Press, WH; Teukolsky, SA; Vetterling, WT; Flannery, BP (2007), "Section 13.11. Numerical Use of the Sampling Theorem" (http://apps.nrbook.com/empanel/index.html#pg=717), *Numerical Recipes: The Art of Scientific Computing* (3rd ed.), New York: Cambridge University Press, ISBN 978-0-521-88068-8
- C. E. Shannon, "Communication in the presence of noise", Proc. Institute of Radio Engineers, vol. 37, no.1, pp. 10–21, Jan. 1949. Reprint as classic paper in: *Proc. IEEE*, Vol. 86, No. 2, (Feb 1998) (http://www.stanford.edu/class/ee104/shannonpaper.pdf)
- Michael Unser: *Sampling-50 Years after Shannon* (http://bigwww.epfl.ch/publications/unser0001.html), Proc. IEEE, vol. 88, no. 4, pp. 569–587, April 2000
- E. T. Whittaker, "On the Functions Which are Represented by the Expansions of the Interpolation Theory", Proc. Royal Soc. Edinburgh, Sec. A, vol.35, pp. 181–194, 1915
- J. M. Whittaker, *Interpolatory Function Theory*, Cambridge Univ. Press, Cambridge, England, 1935.

External links

- Learning by Simulations (http://www.vias.org/simulations/simusoft_nykvist.html) Interactive simulation of the effects of inadequate sampling
- Undersampling and an application of it (http://spazioscuola.altervista.org/UndersamplingAR/UndersamplingARnv.htm)
- Sampling Theory For Digital Audio (http://www.lavryengineering.com/documents/Sampling_Theory.pdf)
- Journal devoted to Sampling Theory (http://www.stsip.org)
- "The Origins of the Sampling Theorem" by Hans Dieter Lüke published in "IEEE Communications Magazine" April 1999 (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.2887&rep=rep1&type=pdf)

Aliasing

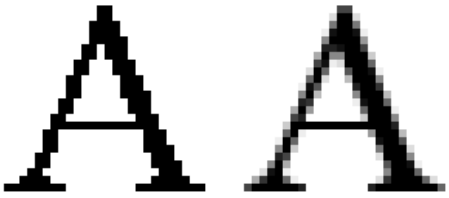
In signal processing and related disciplines, **aliasing** refers to an effect that causes different signals to become indistinguishable (or *aliases* of one another) when sampled. It also refers to the distortion or artifact that results when the signal reconstructed from samples is different from the original continuous signal.



Properly sampled image of brick wall.

Spatial aliasing in the form of a Moiré pattern.

Description



Aliasing example of the A letter in Times New Roman. Left: aliased image, right: *antialiased* image.

When a digital image is viewed, a reconstruction—also known as an interpolation—is performed by a display or printer device, and by the eyes and the brain. If the resolution is too low, the reconstructed image will differ from the original image, and an alias is seen. An example of **spatial aliasing** is the Moiré pattern one can observe in a poorly pixelized image of a brick wall. Techniques that avoid such poor pixelizations are called anti-aliasing. Aliasing can be caused either by the sampling stage or the reconstruction stage; these may be distinguished by calling sampling aliasing *prealiasing* and reconstruction aliasing *postaliasing*.[1]

Temporal aliasing is a major concern in the sampling of video and audio signals. Music, for instance, may contain high-frequency components that are inaudible to humans. If a piece of music is sampled at 32000 samples per second (sps), any frequency components above 16000 Hz (the Nyquist frequency) will cause aliasing when the music is reproduced by a digital to analog converter (DAC). To prevent that, it is customary to remove components above the Nyquist frequency (with an anti-aliasing filter) prior to sampling. But any realistic filter or DAC will also affect (attenuate) the components just below the Nyquist frequency. Therefore, it is also customary to choose a higher Nyquist frequency by sampling faster.

In video or cinematography, temporal aliasing results from the limited frame rate, and causes the wagon-wheel effect, whereby a spoked wheel appears to rotate too slowly or even backwards. Aliasing has changed its apparent frequency of rotation. A reversal of direction can be described as a negative frequency. Temporal aliasing frequencies in video and cinematography are determined by the frame rate of the camera, but the relative intensity of the aliased frequencies is determined by the shutter timing (exposure time) or the use of a temporal aliasing reduction filter during filming.[2]

Like the video camera, most sampling schemes are periodic; that is they have a characteristic sampling frequency in time or in space. Digital cameras provide a certain number of samples (pixels) per degree or per radian, or samples per mm in the focal plane of the camera. Audio signals are sampled (digitized) with an analog-to-digital converter, which produces a constant number of samples per second. Some of the most dramatic and subtle examples of aliasing occur when the signal being sampled also has periodic content.

Bandlimited functions

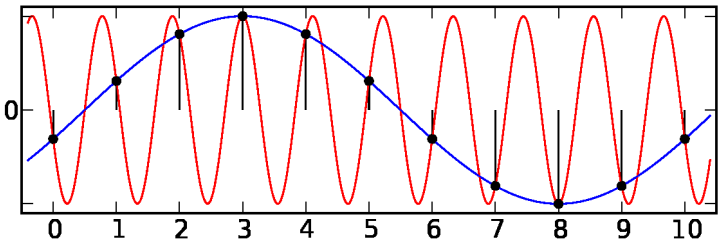
Actual signals have finite duration and their frequency content, as defined by the Fourier transform, has no upper bound. Some amount of aliasing always occurs when such functions are sampled. Functions whose frequency content is bounded (*bandlimited*) have infinite duration. If sampled at a high enough rate, determined by the *bandwidth*, the original function can in theory be perfectly reconstructed from the infinite set of samples.

Bandpass signals

Sometimes aliasing is used intentionally on signals with no low-frequency content, called *bandpass* signals. Undersampling, which creates low-frequency aliases, can produce the same result, with less effort, as frequency-shifting the signal to lower frequencies before sampling at the lower rate. Some digital channelizers[3] exploit aliasing in this way for computational efficiency. See Sampling (signal processing) and Nyquist rate (relative to sampling).

Sampling sinusoidal functions

Sinusoids are an important type of periodic function, because realistic signals are often modeled as the summation of many sinusoids of different frequencies and different amplitudes (with a Fourier series or transform). Understanding what aliasing does to the individual sinusoids is useful in understanding what happens to their sum.

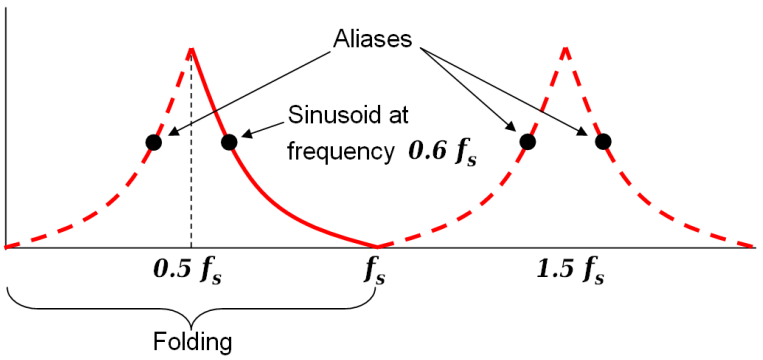


Two different sinusoids that fit the same set of samples.

Here a plot depicts a set of samples whose sample-interval is 1, and two (of many) different sinusoids that could have produced the samples. The sample-rate in this case is $f_s = 1$. For instance, if the interval is 1 second, the rate is 1 sample per second. Nine cycles of the red sinusoid and 1 cycle of the blue sinusoid span an interval of 10. The respective sinusoid frequencies are $f_{\text{red}} = 0.9$ and $f_{\text{blue}} = 0.1$. In general, when a sinusoid of frequency f is sampled with frequency f_s, the resulting samples are indistinguishable from those of another sinusoid of frequency $(f - N f_s)$, for any integer N. The values corresponding to N ≠ 0 are called *images* or *aliases* of frequency f. In our example, the N=±1 aliases of $f = f_{\text{red}} = 0.9$ are $0.9 + 1.0 = 1.9$ and $0.9 - 1.0 = -0.1$. A negative frequency is equivalent to its absolute value, because sin(-wt+θ)=sin(wt-θ+π), and cos(-wt+θ)=cos(wt-θ). Therefore we can express the positive-valued image frequencies as $f_{\text{image}}(N) = |f - N f_s|$, for any integer N (with $f_{\text{image}}(0) = f$ being the actual signal frequency). Then the N=1 alias of f_{red} is f_{blue}, (and vice versa).

Aliasing matters when one attempts to reconstruct the original waveform from its samples. The most common reconstruction technique produces the smallest of the $f_{\text{image}}(N)$ frequencies. So it is usually important that $f_{\text{image}}(0)$ be the unique minimum. A necessary and sufficient condition for that is $f_s/2 > |f|$, where $f_s/2$ is commonly called the Nyquist frequency of a system that samples at rate f_s. In our example, the Nyquist condition is satisfied if the original signal is the blue sinusoid ($f = f_{\text{blue}}$). But if $f = f_{\text{red}} = 0.9$, the usual reconstruction method will produce the blue sinusoid instead of the red one.

Folding



The black dots are aliases of each other. The solid red line is an example of adjusting amplitude vs frequency. The dashed red lines are the corresponding paths of the aliases.

As f increases from 0 to $f_s/2$, $f_{\text{image}}(1)$ goes from f_s to $f_s/2$. Similarly, as f increases from $f_s/2$ to f_s, $f_{\text{image}}(1)$ continues decreasing from $f_s/2$ to 0.

A graph of amplitude vs frequency for a single sinusoid at frequency $0.6 f_s$ and some of its aliases at $0.4 f_s$, $1.4 f_s$, and $1.6 f_s$ would look like the 4 black dots in the adjacent figure. The red lines depict the paths (loci) of the 4 dots if we were to adjust the frequency and amplitude of the sinusoid along the solid red segment (between $f_s/2$ and f_s). No matter what function we choose to change the amplitude vs frequency, the graph will exhibit symmetry between 0 and f_s. This symmetry is commonly referred to as **folding**, and another name for $f_s/2$ (the Nyquist frequency) is **folding frequency**. Folding is most often observed in practice when viewing the frequency spectrum of real-valued samples using a discrete Fourier transform.

Complex sinusoids

Complex sinusoids are waveforms whose samples are complex numbers, and the concept of negative frequency is necessary to distinguish them. In that case, the frequencies of the aliases are given by just: $f_{\text{image}}(N) = f - Nf_s$. Therefore, as f increases from $f_s/2$ to f_s, $f_{\text{image}}(1)$ goes from $-f_s/2$ up to 0. Consequently, complex sinusoids do not exhibit *folding*. Complex samples of real-valued sinusoids have zero-valued imaginary parts and do exhibit folding.

Sample frequency

When the condition $f_s/2 > f$ is met for the highest frequency component of the original signal, then it is met for all the frequency components, a condition known as the Nyquist criterion. That is typically approximated by filtering the original signal to attenuate high frequency components before it is sampled. They still generate low-frequency aliases, but at very low amplitude levels, so as not to cause a problem. A filter chosen in anticipation of a certain sample frequency is called an anti-aliasing filter. The filtered signal can subsequently be reconstructed without significant additional distortion, for example by the Whittaker–Shannon interpolation formula.

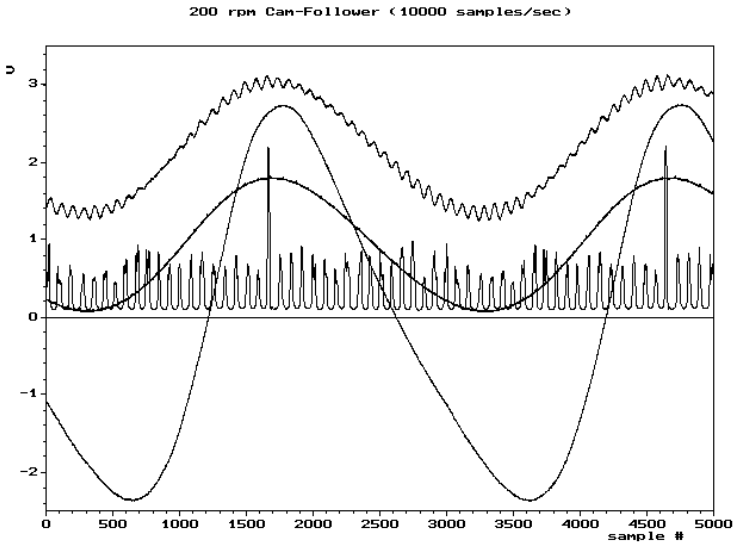


Illustration of 4 waveforms reconstructed from samples taken at 6 different rates. Two of the waveforms are sufficiently sampled to avoid aliasing at all 6 rates. The other two illustrate increasing distortion (aliasing) at the lower rates.

The Nyquist criterion presumes that the frequency content of the signal being sampled has an upper bound. Implicit in that assumption is that the signal's duration has *no* upper bound. Similarly, the Whittaker–Shannon interpolation formula represents an interpolation filter with an unrealizable frequency response. These assumptions make up a mathematical model that is an idealized approximation, at best, to any realistic situation. The conclusion, that perfect reconstruction is possible, is mathematically correct for the model, but only an approximation for actual samples of an actual signal.

Historical usage

Historically the term *aliasing* evolved from radio engineering because of the action of superheterodyne receivers. When the receiver shifts multiple signals down to lower frequencies, from RF to IF by heterodyning, an unwanted signal, from an RF frequency equally far from the local oscillator (LO) frequency as the desired signal, but on the wrong side of the LO, can end up at the same IF frequency as the wanted one. If it is strong enough it can interfere with reception of the desired signal. This unwanted signal is known as an *image* or *alias* of the desired signal.

Angular aliasing

Aliasing occurs whenever the use of discrete elements to capture or produce a continuous signal causes frequency ambiguity.

Spatial aliasing, particular of angular frequency, can occur when reproducing a light field[4] or sound field with discrete elements, as in 3D displays or wave field synthesis of sound.

This aliasing is visible in images such as posters with lenticular printing: if they have low angular resolution, then as one moves past them, say from left-to-right, the 2D image does not initially change (so it appears to move left), then as one moves to the next angular image, the image suddenly changes (so it jumps right) – and the frequency and amplitude of this side-to-side movement corresponds to the angular resolution of the image (and, for frequency, the speed of the viewer's lateral movement), which is the angular aliasing of the 4D light field.

The lack of parallax on viewer movement in 2D images and in 3-D film produced by stereoscopic glasses (in 3D films the effect is called "yawing", as the image appears to rotate on its axis) can similarly be seen as loss of angular resolution, all angular frequencies being aliased to 0 (constant).

More examples

Online "live" example

The qualitative effects of aliasing can be heard in the following audio demonstration. Six sawtooth waves are played in succession, with the first two sawtooths having a fundamental frequency of 440 Hz (A4), the second two having fundamental frequency of 880 Hz (A5), and the final two at 1760 Hz (A6). The sawtooths alternate between bandlimited (non-aliased) sawtooths and aliased sawtooths and the sampling rate is 22.05 kHz. The bandlimited sawtooths are synthesized from the sawtooth waveform's Fourier series such that no harmonics above the Nyquist frequency are present.

The aliasing distortion in the lower frequencies is increasingly obvious with higher fundamental frequencies, and while the bandlimited sawtooth is still clear at 1760 Hz, the aliased sawtooth is degraded and harsh with a buzzing audible at frequencies lower than the fundamental.

Direction finding

A form of spatial aliasing can also occur in antenna arrays or microphone arrays used to estimate the direction of arrival of a wave signal, as in geophysical exploration by seismic waves. Waves must be sampled at more than two points per wavelength, or the wave arrival direction becomes ambiguous.

Notes

[1] Mitchell, Don P.; Netravali, Arun N. (August 1988). "Reconstruction filters in computer-graphics" (http://www.mentallandscape.com/Papers_siggraph88.pdf). **22**. ACM SIGGRAPH International Conference on Computer Graphics and Interactive Techniques (http://portal.acm.org/toc.cfm?id=54852&type=proceeding&coll=GUIDE&dl=GUIDE,ACM&CFID=30538218&CFTOKEN=95411512). pp. 221–228. doi:10.1145/54852.378514. ISBN 0-89791-275-6. .

[2] Tessive, LLC (2010). "Time Filter Technical Explanation" (http://www.tessive.com/home/time-filter-faq-1/time-filter-technical-explanation)

[3] harris, frederic j. (Aug 2006). *Multirate Signal Processing for Communication Systems*. Upper Saddle River, NJ: Prentice Hall PTR. ISBN 0-13-146511-2.

[4] The (New) Stanford Light Field Archive (http://lightfield.stanford.edu/lfs.html)

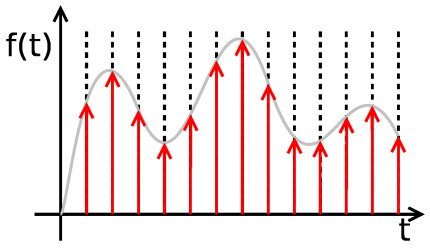
References

External links

- Anti-Aliasing Filter Primer (http://lavidaleica.com/content/anti-aliasing-filter-primer) by La Vida Leica discusses its purpose and effect on the image recorded.
- Frequency Aliasing Demonstration (http://burtonmackenzie.com/2006/07/i-cant-drive-55.html) by Burton MacKenZie using stop frame animation and a clock.
- Your Calculator is Wrong Video (http://www.youtube.com/watch?v=UDfR6GVFFSc) from YouTube, includes some information about aliasing toward the end.

Quantization (signal processing)

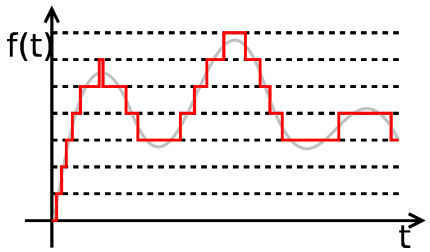
Quantization, in mathematics and digital signal processing, is the process of mapping a large set of input values to a smaller set – such as rounding values to some unit of precision. A device or algorithmic function that performs quantization is called a **quantizer**. The error introduced by quantization is referred to as **quantization error** or **round-off error**. Quantization is involved to some degree in nearly all digital signal processing, as the process of representing a signal in digital form ordinarily involves rounding. Quantization also forms the core of essentially all lossy compression algorithms.



Sampled signal (discrete signal): discrete time, continuous values.

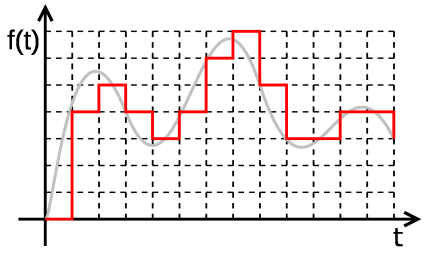
Because quantization is a many-to-few mapping, it is an inherently non-linear and irreversible process (i.e., because the same output value is shared by multiple input values, it is impossible in general to recover the exact input value when given only the output value).

The set of possible input values may be infinitely large, and may possibly be continuous and therefore uncountable (such as the set of all real numbers, or all real numbers within some limited range). The set of possible output values may be finite or countably infinite. The input and output sets involved in quantization can be defined in rather general way. For example, *vector quantization* is the application of quantization to multi-dimensional (vector-valued) input data.[1]



Quantized signal: continuous time, discrete values.

There are two substantially different classes of applications where quantization is used:



Digital signal (sampled, quantized): discrete time, discrete values.

- The first type, which may simply be called *rounding* quantization, is the one employed for many applications, to enable the use of a simple approximate representation for some quantity that is to be measured and used in other calculations. This category includes the simple rounding approximations used in everyday arithmetic. This category also includes analog-to-digital conversion of a signal for a digital signal processing system (e.g., using a sound card of a personal computer to capture an audio signal) and the calculations performed within most digital filtering processes. Here the purpose is primarily to retain as much signal fidelity as possible while eliminating unnecessary precision and keeping the dynamic range of the signal within practical limits (to avoid signal clipping

or arithmetic overflow). In such uses, substantial loss of signal fidelity is often unacceptable, and the design often centers around managing the approximation error to ensure that very little distortion is introduced.

- The second type, which can be called *rate–distortion optimized* quantization, is encountered in source coding for "lossy" data compression algorithms, where the purpose is to manage distortion within the limits of the bit rate supported by a communication channel or storage medium. In this second setting, the amount of introduced distortion may be managed carefully by sophisticated techniques, and introducing some significant amount of distortion may be unavoidable. A quantizer designed for this purpose may be quite different and more elaborate in design than an ordinary rounding operation. It is in this domain that substantial rate–distortion theory analysis is likely to be applied. However, the same concepts actually apply in both use cases.

The analysis of quantization involves studying the amount of data (typically measured in digits or bits or bit *rate*) that is used to represent the output of the quantizer, and studying the loss of precision that is introduced by the quantization process (which is referred to as the *distortion*). The general field of such study of rate and distortion is known as *rate–distortion theory*.

Scalar quantization

The most common type of quantization is known as *scalar quantization*. Scalar quantization, typically denoted as $y = Q(x)$, is the process of using a quantization function $Q(\ )$ to map a scalar (one-dimensional) input value x to a scalar output value y. Scalar quantization can be as simple and intuitive as rounding high-precision numbers to the nearest integer, or to the nearest multiple of some other unit of precision (such as rounding a large monetary amount to the nearest thousand dollars). Scalar quantization of continuous-valued input data that is performed by an electronic sensor is referred to as *analog-to-digital conversion*. Analog-to-digital conversion often also involves sampling the signal periodically in time (e.g., at 44.1 kHz for CD-quality audio signals).

Rounding example

As an example, rounding a real number x to the nearest integer value forms a very basic type of quantizer – a *uniform* one. A typical (*mid-tread*) uniform quantizer with a quantization *step size* equal to some value Δ can be expressed as

$$Q(x) = \operatorname{sgn}(x) \cdot \Delta \cdot \left\lfloor \frac{|x|}{\Delta} + \frac{1}{2} \right\rfloor,$$

where the function $\operatorname{sgn}(\ )$ is the sign function (also known as the *signum* function). For simple rounding to the nearest integer, the step size Δ is equal to 1. With $\Delta = 1$ or with Δ equal to any other integer value, this quantizer has real-valued inputs and integer-valued outputs, although this property is not a necessity – a quantizer may also have an integer input domain and may also have non-integer output values. The essential property of a quantizer is that it has a countable set of possible output values that has fewer members than the set of possible input values. The members of the set of output values may have integer, rational, or real values (or even other possible values as well, in general – such as vector values or complex numbers).

When the quantization step size is small (relative to the variation in the signal being measured), it is relatively simple to show[2][3][4][5][6][7] that the mean squared error produced by such a rounding operation will be approximately $\Delta^2/12$.

Because the set of possible output values of a quantizer is countable, any quantizer can be decomposed into two distinct stages, which can be referred to as the *classification* stage (or *forward quantization* stage) and the *reconstruction* stage (or *inverse quantization* stage), where the classification stage maps the input value to an integer *quantization index* k and the reconstruction stage maps the index k to the *reconstruction value* y_k that is the output approximation of the input value. For the example uniform quantizer described above, the forward quantization stage can be expressed as

$$k = \operatorname{sgn}(x) \cdot \left\lfloor \frac{|x|}{\Delta} + \frac{1}{2} \right\rfloor,$$

and the reconstruction stage for this example quantizer is simply

$$y_k = k \cdot \Delta.$$

This decomposition is useful for the design and analysis of quantization behavior, and it illustrates how the quantized data can be communicated over a communication channel – a *source encoder* can perform the forward quantization stage and send the index information through a communication channel (possibly applying entropy coding techniques to the quantization indices), and a *decoder* can perform the reconstruction stage to produce the output approximation of the original input data. In more elaborate quantization designs, both the forward and inverse quantization stages may be substantially more complex. In general, the forward quantization stage may use any function that maps the input data to the integer space of the quantization index data, and the inverse quantization stage can conceptually (or literally) be a table look-up operation to map each quantization index to a corresponding reconstruction value. This two-stage decomposition applies equally well to vector as well as scalar quantizers.

Mid-riser and mid-tread uniform quantizers

Most uniform quantizers for signed input data can be classified as being of one of two types: **mid-riser** and **mid-tread**. The terminology is based on what happens in the region around the value 0, and uses the analogy of viewing the input-output function of the quantizer as a stairway. Mid-tread quantizers have a zero-valued reconstruction level (corresponding to a *tread* of a stairway), while mid-riser quantizers have a zero-valued classification threshold (corresponding to a *riser* of a stairway).[8]

The formulas for mid-tread uniform quantization are provided above.

The input-output formula for a mid-riser uniform quantizer is given by:

$$Q(x) = \Delta \cdot \left(\left\lfloor \frac{x}{\Delta} \right\rfloor + \frac{1}{2} \right),$$

where the classification rule is given by

and the reconstruction rule is

$$y_k = \Delta \cdot \left(k + \tfrac{1}{2}\right).$$

Note that mid-riser uniform quantizers do not have a zero output value – their minimum output magnitude is half the step size. When the input data can be modeled as a random variable with a probability density function (pdf) that is smooth and symmetric around zero, mid-riser quantizers also always produce an output *entropy* of at least 1 bit per sample.

In contrast, mid-tread quantizers do have a zero output level, and can reach arbitrarily low bit rates per sample for input distributions that are symmetric and taper off at higher magnitudes. For some applications, having a zero output signal representation or supporting low output entropy may be a necessity. In such cases, using a mid-tread uniform quantizer may be appropriate while using a mid-riser one would not be.

In general, a mid-riser or mid-tread quantizer may not actually be a *uniform* quantizer – i.e., the size of the quantizer's classification intervals may not all be the same, or the spacing between its possible output values may not all be the same. The distinguishing characteristic of a mid-riser quantizer is that it has a classification threshold value that is exactly zero, and the distinguishing characteristic of a mid-tread quantizer is that is it has a reconstruction value that is exactly zero.[8]

Another name for a mid-tread quantizer is **dead-zone quantizer**, and the classification region around the zero output value of such a quantizer is referred to as the *dead zone*. The dead zone can sometimes serve the same purpose as a

noise gate or squelch function.

Granular distortion and overload distortion

Often the design of a quantizer involves supporting only a limited range of possible output values and performing clipping to limit the output to this range whenever the input exceeds the supported range. The error introduced by this clipping is referred to as *overload* distortion. Within the extreme limits of the supported range, the amount of spacing between the selectable output values of a quantizer is referred to as its *granularity*, and the error introduced by this spacing is referred to as *granular* distortion. It is common for the design of a quantizer to involve determining the proper balance between granular distortion and overload distortion. For a given supported number of possible output values, reducing the average granular distortion may involve increasing the average overload distortion, and vice-versa. A technique for controlling the amplitude of the signal (or, equivalently, the quantization step size Δ) to achieve the appropriate balance is the use of *automatic gain control* (AGC). However, in some quantizer designs, the concepts of granular error and overload error may not apply (e.g., for a quantizer with a limited range of input data or with a countably infinite set of selectable output values).

The additive noise model for quantization error

A common assumption for the analysis of quantization error is that it affects a signal processing system in a similar manner to that of additive white noise – having negligible correlation with the signal and an approximately flat power spectral density.[3][9][10][7] The additive noise model is commonly used for the analysis of quantization error effects in digital filtering systems, and it can be very useful in such analysis. It has been shown to be a valid model in cases of high resolution quantization (small Δ relative to the signal strength) with smooth probability density functions.[3][11] However, additive noise behaviour is not always a valid assumption, and care should be taken to avoid assuming that this model always applies. In actuality, the quantization error (for quantizers defined as described here) is deterministically related to the signal rather than being independent of it,[7] and in some cases it can even cause limit cycles to appear in digital signal processing systems.[10]

One way to ensure effective independence of the quantization error from the source signal is to perform *dithered quantization* (sometimes with *noise shaping*), which involves adding random (or pseudo-random) noise to the signal prior to quantization.[10][7] This can sometimes be beneficial for such purposes as improving the subjective quality of the result, however it can increase the total quantity of error introduced by the quantization process.

Rate–distortion quantizer design

A scalar quantizer, which performs a quantization operation, can ordinarily be decomposed into two stages:

- **Classification:** A process that classifies the input signal range into M non-overlapping **intervals** $\{I_k\}_{k=1}^{M}$, by defining $M-1$ **boundary (decision)** values $\{b_k\}_{k=1}^{M-1}$, such that $I_k = [b_{k-1}\ ,\ b_k)$ for $k = 1, 2, \ldots, M$, with the extreme limits defined by $b_0 = -\infty$ and $b_M = \infty$. All the inputs x that fall in a given interval range I_k are associated with the same quantization index k.
- **Reconstruction:** Each interval I_k is represented by a **reconstruction value** y_k which implements the mapping $x \in I_k \Rightarrow y = y_k$.

These two stages together comprise the mathematical operation of $y = Q(x)$.

Entropy coding techniques can be applied to communicate the quantization indices from a source encoder that performs the classification stage to a decoder that performs the reconstruction stage. One way to do this is to associate each quantization index k with a binary codeword c_k. An important consideration is the number of bits used for each codeword, denoted here by $\mathrm{length}(c_k)$.

As a result, the design of an M-level quantizer and an associated set of codewords for communicating its index values requires finding the values of $\{b_k\}_{k=1}^{M-1}$, $\{c_k\}_{k=1}^{M}$ and $\{y_k\}_{k=1}^{M}$ which optimally satisfy a selected set of

design constraints such as the **bit rate** R and **distortion** D.

Assuming that an information source S produces random variables X with an associated probability density function $f(x)$, the probability p_k that the random variable falls within a particular quantization interval I_k is given by

$$p_k = P[x \in I_k] = \int_{b_{k-1}}^{b_k} f(x)dx \ .$$

The resulting bit rate R, in units of average bits per quantized value, for this quantizer can be derived as follows:

$$R = \sum_{k=1}^{M} p_k \cdot \text{length}(c_k) = \sum_{k=1}^{M} \text{length}(c_k) \int_{b_{k-1}}^{b_k} f(x)dx \ .$$

If it is assumed that distortion is measured by mean squared error, the distortion **D**, is given by:

$$D = E[(x - Q(x))^2] = \int_{-\infty}^{\infty} (x - Q(x))^2 f(x)dx = \sum_{k=1}^{M} \int_{b_{k-1}}^{b_k} (x - y_k)^2 f(x)dx \ .$$

Note that other distortion measures can also be considered, although mean squared error is a popular one.

A key observation is that rate R depends on the decision boundaries $\{b_k\}_{k=1}^{M-1}$ and the codeword lengths $\{\text{length}(c_k)\}_{k=1}^{M}$, whereas the distortion D depends on the decision boundaries $\{b_k\}_{k=1}^{M-1}$ and the reconstruction levels $\{y_k\}_{k=1}^{M}$.

After defining these two performance metrics for the quantizer, a typical Rate–Distortion formulation for a quantizer design problem can be expressed in one of two ways:

1. Given a maximum distortion constraint $D \leq D_{\max}$, minimize the bit rate R
2. Given a maximum bit rate constraint $R \leq R_{\max}$, minimize the distortion D

Often the solution to these problems can be equivalently (or approximately) expressed and solved by converting the formulation to the unconstrained problem

where the Lagrange multiplier λ is a non-negative constant that establishes the appropriate balance between rate and distortion. Solving the unconstrained problem is equivalent to finding a point on the convex hull of the family of solutions to an equivalent constrained formulation of the problem. However, finding a solution – especially a closed-form solution – to any of these three problem formulations can be difficult. Solutions that do not require multi-dimensional iterative optimization techniques have been published for only three probability distribution functions: the uniform,[12] exponential,[13] and Laplacian[13] distributions. Iterative optimization approaches can be used to find solutions in other cases.[14][15][7]

Note that the reconstruction values $\{y_k\}_{k=1}^{M}$ affect only the distortion – they do not affect the bit rate – and that each individual y_k makes a separate contribution d_k to the total distortion as shown below:

where

This observation can be used to ease the analysis – given the set of $\{b_k\}_{k=1}^{M-1}$ values, the value of each y_k can be optimized separately to minimize its contribution to the distortion D.

For the mean-square error distortion criterion, it can be easily shown that the optimal set of reconstruction values $\{y_k^*\}_{k=1}^{M}$ is given by setting the reconstruction value y_k within each interval I_k to the conditional expected value (also referred to as the *centroid*) within the interval, as given by:

$$y_k^* = \frac{1}{p_k} \int_{b_{k-1}}^{b_k} x f(x)dx \ .$$

The use of sufficiently well-designed entropy coding techniques can result in the use of a bit rate that is close to the true information content of the indices $\{k\}_{k=1}^{M}$, such that effectively

and therefore

The use of this approximation can allow the entropy coding design problem to be separated from the design of the quantizer itself. Modern entropy coding techniques such as arithmetic coding can achieve bit rates that are very close to the true entropy of a source, given a set of known (or adaptively estimated) probabilities $\{p_k\}_{k=1}^{M}$.

In some designs, rather than optimizing for a particular number of classification regions M, the quantizer design problem may include optimization of the value of M as well. For some probabilistic source models, the best performance may be achieved when M approaches infinity.

Neglecting the entropy constraint: Lloyd–Max quantization

In the above formulation, if the bit rate constraint is neglected by setting λ equal to 0, or equivalently if it is assumed that a fixed-length code (FLC) will be used to represent the quantized data instead of a variable-length code (or some other entropy coding technology such as arithmetic coding that is better than an FLC in the rate–distortion sense), the optimization problem reduces to minimization of distortion D alone.

The indices produced by an M-level quantizer can be coded using a fixed-length code using $R = \lceil \log_2 M \rceil$ bits/symbol. For example when $M = 256$ levels, the FLC bit rate R is 8 bits/symbol. For this reason, such a quantizer has sometimes been called an 8-bit quantizer. However using an FLC eliminates the compression improvement that can be obtained by use of better entropy coding.

Assuming an FLC with M levels, the Rate–Distortion minimization problem can be reduced to distortion minimization alone. The reduced problem can be stated as follows: given a source X with pdf $f(x)$ and the constraint that the quantizer must use only M classification regions, find the decision boundaries $\{b_k\}_{k=1}^{M-1}$ and reconstruction levels $\{y_k\}_{k=1}^{M}$ to minimize the resulting distortion

Finding an optimal solution to the above problem results in a quantizer sometimes called a MMSQE (minimum mean-square quantization error) solution, and the resulting pdf-optimized (non-uniform) quantizer is referred to as a *Lloyd–Max* quantizer, named after two people who independently developed iterative methods[16][17][7] to solve the two sets of simultaneous equations resulting from $\partial D/\partial b_k = 0$ and $\partial D/\partial y_k = 0$, as follows:

$$\frac{\partial D}{\partial b_k} = 0 \Rightarrow b_k = \frac{y_k + y_{k+1}}{2},$$

which places each threshold at the mid-point between each pair of reconstruction values, and

which places each reconstruction value at the centroid (conditional expected value) of its associated classification interval.

Lloyd's Method I algorithm, originally described in 1957, can be generalized in a straighforward way for application to vector data. This generalization results in the Linde–Buzo–Gray (LBG) or k-means classifier optimization methods. Moreover, the technique can be further generalized in a straightforward way to also include an entropy constraint for vector data.[18]

Uniform quantization and the 6 dB/bit approximation

The Lloyd–Max quantizer is actually a uniform quantizer when the input pdf is uniformly distributed over the range $[y_1 - \Delta/2,\ y_M + \Delta/2)$. However, for a source that does not have a uniform distribution, the minimum-distortion quantizer may not be a uniform quantizer.

The analysis of a uniform quantizer applied to a uniformly distributed source can be summarized in what follows:

A symmetric source X can be modelled with $f(x) = \frac{1}{2X_{max}}$, for $x \in [-X_{max}, X_{max}]$ and 0 elsewhere. The step size $\Delta = \frac{2X_{max}}{M}$ and the *signal to quantization noise ratio* (SQNR) of the quantizer is

$$\text{SQNR} = 10\log_{10}\frac{\sigma_x^2}{\sigma_q^2} = 10\log_{10}\frac{(M\Delta)^2/12}{\Delta^2/12} = 10\log_{10} M^2 = 20\log_{10} M.$$

For a fixed-length code using N bits, $M = 2^N$, resulting in

$$\text{SQNR} = 20\log_{10} 2^N = N \cdot (20\log_{10} 2) = N \cdot 6.0206\,\text{dB},$$

or approximately 6 dB per bit. For example, for N =8 bits, M =256 levels and SQNR = 8*6 = 48 dB; and for N =16 bits, M =65536 and SQNR = 16*6 = 96 dB. The property of 6 dB improvement in SQNR for each extra bit used in quantization is a well-known figure of merit. However, it must be used with care: this derivation is only for a uniform quantizer applied to a uniform source.

For other source pdfs and other quantizer designs, the SQNR may be somewhat different than predicted by 6 dB/bit, depending on the type of pdf, the type of source, the type of quantizer, and the bit rate range of operation.

However, it is common to assume that for many sources, the slope of a quantizer SQNR function can be approximated as 6 dB/bit when operating at a sufficiently high bit rate. At asymptotically high bit rates, cutting the step size in half increases the bit rate by approximately 1 bit per sample (because 1 bit is needed to indicate whether the value is in the left or right half of the prior double-sized interval) and reduces the mean squared error by a factor of 4 (i.e., 6 dB) based on the $\Delta^2/12$ approximation.

At asymptotically high bit rates, the 6 dB/bit approximation is supported for many source pdfs by rigorous theoretical analysis.[3][4][6][7] Moreover, the structure of the optimal scalar quantizer (in the rate–distortion sense) approaches that of a uniform quantizer under these conditions.[6][7]

Companding quantizers

Companded quantization is the combination of three functional building blocks – namely, a (continuous-domain) signal dynamic range *compressor*, a limited-range uniform quantizer, and a (continuous-domain) signal dynamic range *expander* that basically inverts the compressor function. This type of quantization is frequently used in older speech telephony systems. The compander function of the compressor is key to the performance of such a quantization system. In principle, the compressor function can be designed to exactly map the boundaries of the optimal intervals of any desired scalar quantizer function to the equal-size intervals used by the uniform quantizer and similarly the expander function can exactly map the uniform quantizer reconstruction values to any arbitrary reconstruction values. Thus, with arbitrary compressor and expander functions, any possible non-uniform scalar quantizer can be equivalently implemented as a companded quantizer.[3][7] In practice, companders are designed to operate according to relatively-simple dynamic range compressor functions that are designed to be suitable for implementation using simple analog electronic circuits. The two most popular compander functions used for telecommunications are the A-law and μ-law functions.

Notes

[1] Allen Gersho and Robert M. Gray, *Vector Quantization and Signal Compression*, Springer, ISBN 978-0-7923-9181-4, 1991.

[2] William Fleetwood Sheppard, "On the Calculation of the Most Probable Values of Frequency Constants for data arranged according to Equidistant Divisions of a Scale", *Proceedings of the London Mathematical Society*, Vol. 29, pp. 353–80, 1898.

[3] W. R. Bennett, "Spectra of Quantized Signals", *Bell System Technical Journal*, Vol. 27, pp. 446–472, July 1948.

[4] B. M. Oliver, J. R. Pierce, and Claude E. Shannon, "The Philosophy of PCM", *Proceedings of the IRE*, Vol. 36, pp. 1324–1331, Nov. 1948.

[5] Seymour Stein and J. Jay Jones, *Modern Communication Principles*, McGraw–Hill, ISBN 978-0-07-061003-3, 1967 (p. 196).

[6] Herbert Gish and John N. Pierce, "Asymptotically Efficient Quantizing", *IEEE Transactions on Information Theory*, Vol. IT-14, No. 5, pp. 676–683, Sept. 1968.

[7] Robert M. Gray and David L. Neuhoff, "Quantization", *IEEE Transactions on Information Theory*, Vol. IT-44, No. 6, pp. 2325–2383, Oct. 1998.

[8] Allen Gersho, "Quantization", *IEEE Communications Society Magazine*, pp. 16–28, Sept. 1977.

[9] Bernard Widrow, "A study of rough amplitude quantization by means of Nyquist sampling theory", *IRE Trans. Circuit Theory*, Vol. CT-3, pp. 266–276, 1956.

[10] Bernard Widrow, "Statistical analysis of amplitude quantized sampled data systems", *Trans. AIEE Pt. II: Appl. Ind.*, Vol. 79, pp. 555–568, Jan. 1961.

[11] Daniel Marco and David L. Neuhoff, "The Validity of the Additive Noise Model for Uniform Scalar Quantizers", *IEEE Transactions on Information Theory*, Vol. IT-51, No. 5, pp. 1739–1755, May 2005.

[12] Nariman Farvardin and James W. Modestino, "Optimum Quantizer Performance for a Class of Non-Gaussian Memoryless Sources", *IEEE Transactions on Information Theory*, Vol. IT-30, No. 3, pp. 485–497, May 1982 (Section VI.C and Appendix B).

[13] Gary J. Sullivan, "Efficient Scalar Quantization of Exponential and Laplacian Random Variables", *IEEE Transactions on Information Theory*, Vol. IT-42, No. 5, pp. 1365–1374, Sept. 1996.

[14] Toby Berger, "Optimum Quantizers and Permutation Codes", *IEEE Transactions on Information Theory*, Vol. IT-18, No. 6, pp. 759–765, Nov. 1972.

[15] Toby Berger, "Minimum Entropy Quantizers and Permutation Codes", *IEEE Transactions on Information Theory*, Vol. IT-28, No. 2, pp. 149–157, Mar. 1982.

[16] Stuart P. Lloyd, "Least Squares Quantization in PCM", *IEEE Transactions on Information Theory*, Vol. IT-28, pp. 129–137, No. 2, March 1982 (work documented in a manuscript circulated for comments at Bell Laboratories with a department log date of 31 July 1957 and also presented at the 1957 meeting of the Institute of Mathematical Statistics, although not formally published until 1982).

[17] Joel Max, "Quantizing for Minimum Distortion", *IRE Transactions on Information Theory*, Vol. IT-6, pp. 7–12, March 1960.

[18] Philip A. Chou, Tom Lookabaugh, and Robert M. Gray, "Entropy-Constrained Vector Quantization", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-37, No. 1, Jan. 1989.

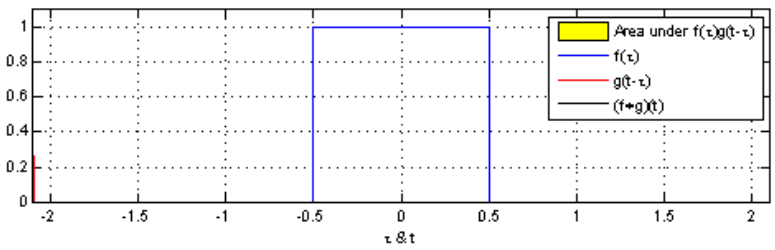
References

- Sayood, Khalid (2005), *Introduction to Data Compression, Third Edition*, Morgan Kaufmann, ISBN 978-0-12-620862-7
- Jayant, Nikil S.; Noll, Peter (1984), *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice–Hall, ISBN 978-0-13-211913-9
- Gregg, W. David (1977), *Analog & Digital Communication*, John Wiley, ISBN 978-0-471-32661-8
- Stein, Seymour; Jones, J. Jay (1967), *Modern Communication Principles*, McGraw–Hill, ISBN 978-0-07-061003-3

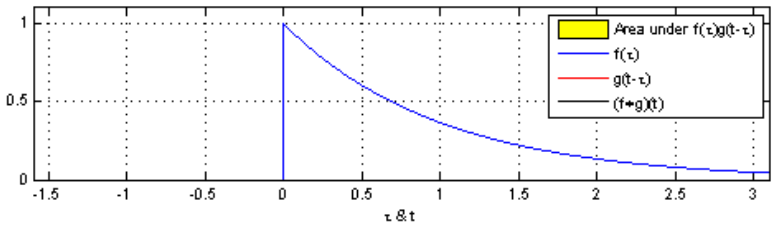
Convolution

Convolution

In mathematics and, in particular, functional analysis, **convolution** is a mathematical operation on two functions f and g, producing a third function that is typically viewed as a modified version of one of the original functions, giving the area overlap between the two functions as a function of the amount that one of the original functions is translated. Convolution is similar to cross-correlation. It has applications that include probability, statistics, computer vision, image and signal processing, electrical engineering, and differential equations.



Convolution of two square pulses: the resulting waveform is a triangular pulse. One of the functions (in this case g) is first reflected about $\tau = 0$ and then offset by t, making it $g(t-\tau)$. The area under the resulting product gives the convolution at t. The horizontal axis is τ for f and g, and t for $f * g$.



Convolution of a square pulse (as input signal) with the impulse response of an RC circuit to obtain the output signal waveform. The integral of their product is the area of the yellow region. In both animations the function g is symmetric, and so is unchanged under reflection.

The convolution can be defined for functions on groups other than Euclidean space. In particular, the circular convolution can be defined for periodic functions (that is, functions on the circle), and the discrete convolution can be defined for functions on the set of integers. These generalizations of the convolution have applications in the field of numerical analysis and numerical linear algebra, and in the design and implementation of finite impulse response filters in signal processing.

Computing the inverse of the convolution operation is known as deconvolution.

History

The operation

$$\int_0^t \varphi(s)\psi(t-s)\,ds, \qquad 0 \le t < \infty,$$

is a particular case of composition products considered by the Italian mathematician Vito Volterra in 1913.[1]

Convolution is also sometimes called "Faltung" (which means *folding* in German); both *Faltung* and *convolution* were used as early as 1903, though the definition is rather unfamiliar in older uses.[2][3] The term *Faltung* was sometimes used in English through the 1940s, before the notion of convolution became widely used, along with other terms such as *composition product*, *superposition integral*, and *Carson's integral*.[4]

Definition

The convolution of f and g is written $f*g$, using an asterisk or star. It is defined as the integral of the product of the two functions after one is reversed and shifted. As such, it is a particular kind of integral transform:

$$\begin{aligned}(f*g)(t) &\ \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau \\ &= \int_{-\infty}^{\infty} f(t-\tau)\, g(\tau)\, d\tau. \qquad \text{(commutativity)}\end{aligned}$$

While the symbol t is used above, it need not represent the time domain. But in that context, the convolution formula can be described as a weighted average of the function $f(\tau)$ at the moment t where the weighting is given by $g(-\tau)$ simply shifted by amount t. As t changes, the weighting function emphasizes different parts of the input function.

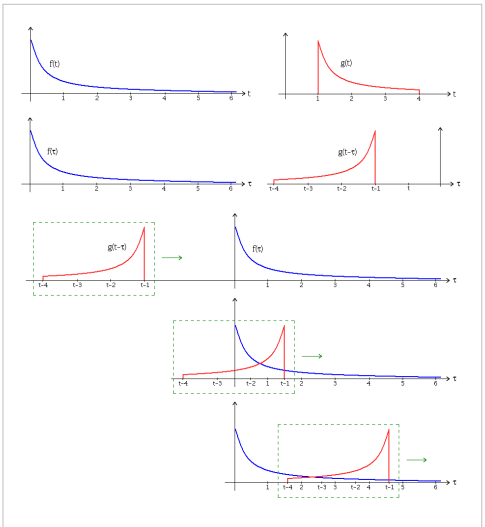
More generally, if f and g are complex-valued functions on $\mathbf{R}^d$, then their convolution may be defined as the integral:

$$(f*g)(x) = \int_{\mathbf{R}^d} f(y)g(x-y)\,dy = \int_{\mathbf{R}^d} f(x-y)g(y)\,dy.$$

Visual explanation of convolution.

1. Express each function in terms of a dummy variable τ.
2. Reflect one of the functions: $g(\tau) \rightarrow g(-\tau)$.
3. Add a time-offset, t, which allows $g(t-\tau)$ to slide along the τ-axis.
4. Start t at -∞ and slide it all the way to +∞. Wherever the two functions intersect, find the integral of their product. In other words, compute a sliding, weighted-average of function $f(\tau)$, where the weighting function is $g(-\tau)$.

The resulting waveform (not shown here) is the convolution of functions f and g. If $f(t)$ is a unit impulse, the result of this process is simply $g(t)$, which is therefore called the impulse response.



Circular convolution

When a function g_T is periodic, with period T, then for functions, f, such that $f*g_T$ exists, the convolution is also periodic and identical to:

$$(f*g_T)(t) \equiv \int_{t_0}^{t_0+T} \left[\sum_{k=-\infty}^{\infty} f(\tau+kT)\right] g_T(t-\tau)\,d\tau,$$

where t_0 is an arbitrary choice. The summation is called a periodic summation of the function f.

If g_T is a periodic summation of another function, g, then $f*g_T$ is known as a *circular*, *cyclic*, or *periodic* convolution of f and g.

Discrete convolution

For complex-valued functions *f*, *g* defined on the set **Z** of integers, the **discrete convolution** of *f* and *g* is given by:

$$(f * g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f[m]\, g[n-m]$$
$$= \sum_{m=-\infty}^{\infty} f[n-m]\, g[m]. \qquad \text{(commutativity)}$$

When multiplying two polynomials, the coefficients of the product are given by the convolution of the original coefficient sequences, extended with zeros where necessary to avoid undefined terms; this is known as the Cauchy product of the coefficients of the two polynomials.

Circular discrete convolution

When a function g_{N} is periodic, with period *N*, then for functions, *f*, such that $f*g_{\mathrm{N}}$ exists, the convolution is also periodic and identical to:

$$(f * g_N)[n] \equiv \sum_{m=0}^{N-1} \left(\sum_{k=-\infty}^{\infty} f[m+kN] \right) g_N[n-m].$$

The summation on *k* is called a periodic summation of the function *f*.

If g_N is a periodic summation of another function, *g*, then $f*g_N$ is known as a circular convolution of *f* and *g*.

When the non-zero durations of both *f* and *g* are limited to the interval [0, *N* – 1], $f*g_N$ reduces to these common forms:

$$\begin{aligned}(f * g_N)[n] &= \sum_{m=0}^{N-1} f[m]\ g_N[n-m] \\ &= \sum_{m=0}^{n} f[m]\ g[n-m] + \sum_{m=n+1}^{N-1} f[m]\ g[N+n-m] \\ &= \sum_{m=0}^{N-1} f[m]\ g[(n-m)_{\bmod N}] \equiv (f *_N g)[n]\end{aligned} \qquad \textbf{(Eq.1)}$$

The notation $(f *_N g)$ for *cyclic convolution* denotes convolution over the cyclic group of integers modulo *N*.

Circular convolution is frequently used to characterized systems analyzed through the lens of the Discrete Fourier Transform.

Fast convolution algorithms

In many situations, discrete convolutions can be converted to circular convolutions so that fast transforms with a convolution property can be used to implement the computation. For example, convolution of digit sequences is the kernel operation in multiplication of multi-digit numbers, which can therefore be efficiently implemented with transform techniques (Knuth 1997, §4.3.3.C; von zur Gathen & Gerhard 2003, §8.2).

Eq.1 requires *N* arithmetic operations per output value and N^2 operations for *N* outputs. That can be significantly reduced with any of several fast algorithms. Digital signal processing and other applications typically use fast convolution algorithms to reduce the cost of the convolution to O(*N* log *N*) complexity.

The most common fast convolution algorithms use fast Fourier transform (FFT) algorithms via the circular convolution theorem. Specifically, the circular convolution of two finite-length sequences is found by taking an FFT of each sequence, multiplying pointwise, and then performing an inverse FFT. Convolutions of the type defined above are then efficiently implemented using that technique in conjunction with zero-extension and/or discarding portions of the output. Other fast convolution algorithms, such as the Schönhage–Strassen algorithm, use fast Fourier

transforms in other rings.

Domain of definition

The convolution of two complex-valued functions on $\mathbf{R}^d$

$$(f*g)(x) = \int_{\mathbf{R}^d} f(y)g(x-y)\,dy$$

is well-defined only if f and g decay sufficiently rapidly at infinity in order for the integral to exist. Conditions for the existence of the convolution may be tricky, since a blow-up in g at infinity can be easily offset by sufficiently rapid decay in f. The question of existence thus may involve different conditions on f and g.

Compactly supported functions

If f and g are compactly supported continuous functions, then their convolution exists, and is also compactly supported and continuous (Hörmander). More generally, if either function (say f) is compactly supported and the other is locally integrable, then the convolution $f*g$ is well-defined and continuous.

Integrable functions

The convolution of f and g exists if f and g are both Lebesgue integrable functions (in $\mathrm{L}^1(\mathbf{R}^d)$), and in this case $f*g$ is also integrable (Stein & Weiss 1971, Theorem 1.3). This is a consequence of Tonelli's theorem. Likewise, if $f \in L^1(\mathbf{R}^d)$ and $g \in L^p(\mathbf{R}^d)$ where $1 \le p \le \infty$, then $f*g \in L^p(\mathbf{R}^d)$ and

$$\|f*g\|_p \le \|f\|_1 \|g\|_p.$$

In the particular case $p = 1$, this shows that L^1 is a Banach algebra under the convolution (and equality of the two sides holds if f and g are non-negative almost everywhere).

More generally, Young's inequality implies that the convolution is a continuous bilinear map between suitable L^p spaces. Specifically, if $1 \le p,q,r \le \infty$ satisfy

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1,$$

then

$$\|f*g\|_r \le \|f\|_p \|g\|_q, \quad f \in L^p,\ g \in L^q,$$

so that the convolution is a continuous bilinear mapping from $L^p \times L^q$ to L^r.

Functions of rapid decay

In addition to compactly supported functions and integrable functions, functions that have sufficiently rapid decay at infinity can also be convolved. An important feature of the convolution is that if f and g both decay rapidly, then $f*g$ also decays rapidly. In particular, if f and g are rapidly decreasing functions, then so is the convolution $f*g$. Combined with the fact that convolution commutes with differentiation (see **Properties**), it follows that the class of Schwartz functions is closed under convolution.

Distributions

Under some circumstances, it is possible to define the convolution of a function with a distribution, or of two distributions. If *f* is a compactly supported function and *g* is a distribution, then *f*∗*g* is a smooth function defined by a distributional formula analogous to

$$\int_{\mathbf{R}^d} f(y)g(x-y)\,dy.$$

More generally, it is possible to extend the definition of the convolution in a unique way so that the associative law

$$f*(g*\varphi)=(f*g)*\varphi$$

remains valid in the case where *f* is a distribution, and *g* a compactly supported distribution (Hörmander 1983, §4.2).

Measures

The convolution of any two Borel measures μ and ν of bounded variation is the measure λ defined by

$$\int_{\mathbf{R}^d} f(x)d\lambda(x)=\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} f(x+y)\,d\mu(x)\,d\nu(y).$$

This agrees with the convolution defined above when μ and ν are regarded as distributions, as well as the convolution of L^1 functions when μ and ν are absolutely continuous with respect to the Lebesgue measure.

The convolution of measures also satisfies the following version of Young's inequality

$$\|\mu*\nu\|\leq\|\mu\|\|\nu\|$$

where the norm is the total variation of a measure. Because the space of measures of bounded variation is a Banach space, convolution of measures can be treated with standard methods of functional analysis that may not apply for the convolution of distributions.

Properties

Algebraic properties

The convolution defines a product on the linear space of integrable functions. This product satisfies the following algebraic properties, which formally mean that the space of integrable functions with the product given by convolution is a commutative algebra without identity (Strichartz 1994, §3.3). Other linear spaces of functions, such as the space of continuous functions of compact support, are closed under the convolution, and so also form commutative algebras.

Commutativity

$$f*g=g*f$$

Associativity

$$f*(g*h)=(f*g)*h$$

Distributivity

$$f*(g+h)=(f*g)+(f*h)$$

Associativity with scalar multiplication

$$a(f*g)=(af)*g=f*(ag)$$

for any real (or complex) number a.

Multiplicative identity

No algebra of functions possesses an identity for the convolution. The lack of identity is typically not a major inconvenience, since most collections of functions on which the convolution is performed can be convolved with a

delta distribution or, at the very least (as is the case of L^1) admit approximations to the identity. The linear space of compactly supported distributions does, however, admit an identity under the convolution. Specifically,

$$f * \delta = f$$

where δ is the delta distribution.

Inverse element

Some distributions have an inverse element for the convolution, $S^{(-1)}$, which is defined by

$$S^{(-1)} * S = \delta.$$

The set of invertible distributions forms an abelian group under the convolution.

Complex conjugation

$$\overline{f * g} = \overline{f} * \overline{g}$$

Integration

If *f* and *g* are integrable functions, then the integral of their convolution on the whole space is simply obtained as the product of their integrals:

$$\int_{\mathbf{R}^d} (f * g)(x)\,dx = \left(\int_{\mathbf{R}^d} f(x)\,dx\right)\left(\int_{\mathbf{R}^d} g(x)\,dx\right).$$

This follows from Fubini's theorem. The same result holds if *f* and *g* are only assumed to be nonnegative measurable functions, by Tonelli's theorem.

Differentiation

In the one-variable case,

$$\frac{d}{dx}(f * g) = \frac{df}{dx} * g = f * \frac{dg}{dx}$$

where *d*/*dx* is the derivative. More generally, in the case of functions of several variables, an analogous formula holds with the partial derivative:

$$\frac{\partial}{\partial x_i}(f * g)(x) = \frac{\partial f}{\partial x_i} * g = f * \frac{\partial g}{\partial x_i}.$$

A particular consequence of this is that the convolution can be viewed as a "smoothing" operation: the convolution of *f* and *g* is differentiable as many times as *f* and *g* are together.

These identities hold under the precise condition that *f* and *g* are absolutely integrable and at least one of them has an absolutely integrable (L^1) weak derivative, as a consequence of Young's inequality. For instance, when *f* is continuously differentiable with compact support, and *g* is an arbitrary locally integrable function,

$$\frac{d}{dx}(f * g) = \frac{df}{dx} * g.$$

These identities also hold much more broadly in the sense of tempered distributions if one of *f* or *g* is a compactly supported distribution or a Schwartz function and the other is a tempered distribution. On the other hand, two positive integrable and infinitely differentiable functions may have a nowhere continuous convolution.

In the discrete case, the difference operator $D\,f(n) = f(n + 1) - f(n)$ satisfies an analogous relationship:

$$D(f * g) = (Df) * g = f * (Dg).$$

Convolution theorem

The convolution theorem states that

$$\mathcal{F}\{f * g\} = k \cdot \mathcal{F}\{f\} \cdot \mathcal{F}\{g\}$$

where $\mathcal{F}\{f\}$ denotes the Fourier transform of f, and k is a constant that depends on the specific normalization of the Fourier transform (see "Properties of the Fourier transform"). Versions of this theorem also hold for the Laplace transform, two-sided Laplace transform, Z-transform and Mellin transform.
See also the less trivial Titchmarsh convolution theorem.

Translation invariance

The convolution commutes with translations, meaning that

$$\tau_x(f * g) = (\tau_x f) * g = f * (\tau_x g)$$

where $\tau_x f$ is the translation of the function f by x defined by

$$(\tau_x f)(y) = f(y - x).$$

If f is a Schwartz function, then $\tau_x f$ is the convolution with a translated Dirac delta function $\tau_x f = f * \tau_x \delta$. So translation invariance of the convolution of Schwartz functions is a consequence of the associativity of convolution.

Furthermore, under certain conditions, convolution is the most general translation invariant operation. Informally speaking, the following holds

- Suppose that S is a linear operator acting on functions which commutes with translations: $S(\tau_x f) = \tau_x(Sf)$ for all x. Then S is given as convolution with a function (or distribution) g_S; that is $Sf = g_S * f$.

Thus any translation invariant operation can be represented as a convolution. Convolutions play an important role in the study of time-invariant systems, and especially LTI system theory. The representing function g_S is the impulse response of the transformation S.

A more precise version of the theorem quoted above requires specifying the class of functions on which the convolution is defined, and also requires assuming in addition that S must be a continuous linear operator with respect to the appropriate topology. It is known, for instance, that every continuous translation invariant continuous linear operator on L^1 is the convolution with a finite Borel measure. More generally, every continuous translation invariant continuous linear operator on L^p for $1 \leq p < \infty$ is the convolution with a tempered distribution whose Fourier transform is bounded. To wit, they are all given by bounded Fourier multipliers.

Convolutions on groups

If G is a suitable group endowed with a measure λ, and if f and g are real or complex valued integrable functions on G, then we can define their convolution by

$$(f * g)(x) = \int_G f(y) g(y^{-1}x)\, d\lambda(y).$$

In typical cases of interest G is a locally compact Hausdorff topological group and λ is a (left-) Haar measure. In that case, unless G is unimodular, the convolution defined in this way is not the same as $\int f(xy^{-1})g(y)\, d\lambda(y)$. The preference of one over the other is made so that convolution with a fixed function g commutes with left translation in the group:

$$L_h(f * g) = (L_h f) * g = f * (L_h g).$$

Furthermore, the convention is also required for consistency with the definition of the convolution of measures given below. However, with a right instead of a left Haar measure, the latter integral is preferred over the former.

On locally compact abelian groups, a version of the convolution theorem holds: the Fourier transform of a convolution is the pointwise product of the Fourier transforms. The circle group **T** with the Lebesgue measure is an

immediate example. For a fixed g in $L^1(\mathbf{T})$, we have the following familiar operator acting on the Hilbert space $L^2(\mathbf{T})$:

$$Tf(x) = \frac{1}{2\pi}\int_{\mathbf{T}} f(y)g(x-y)\,dy.$$

The operator T is compact. A direct calculation shows that its adjoint T^* is convolution with

$$\bar{g}(-y).$$

By the commutativity property cited above, T is normal: $T^*T = TT^*$. Also, T commutes with the translation operators. Consider the family S of operators consisting of all such convolutions and the translation operators. Then S is a commuting family of normal operators. According to spectral theory, there exists an orthonormal basis $\{h_k\}$ that simultaneously diagonalizes S. This characterizes convolutions on the circle. Specifically, we have

$$h_k(x) = e^{ikx}, \quad k \in \mathbb{Z},$$

which are precisely the characters of $\mathbf{T}$. Each convolution is a compact multiplication operator in this basis. This can be viewed as a version of the convolution theorem discussed above.

A discrete example is a finite cyclic group of order n. Convolution operators are here represented by circulant matrices, and can be diagonalized by the discrete Fourier transform.

A similar result holds for compact groups (not necessarily abelian): the matrix coefficients of finite-dimensional unitary representations form an orthonormal basis in L^2 by the Peter–Weyl theorem, and an analog of the convolution theorem continues to hold, along with many other aspects of harmonic analysis that depend on the Fourier transform.

Convolution of measures

Let G be a topological group. If μ and ν are finite Borel measures on a group G, then their convolution μ∗ν is defined by

$$(\mu * \nu)(E) = \iint 1_E(xy)\,d\mu(x)\,d\nu(y)$$

for each measurable subset E of G. The convolution is also a finite measure, whose total variation satisfies

$$\|\mu * \nu\| \le \|\mu\|\|\nu\|.$$

In the case when G is locally compact with (left-)Haar measure λ, and μ and ν are absolutely continuous with respect to a λ, so that each has a density function, then the convolution μ∗ν is also absolutely continuous, and its density function is just the convolution of the two separate density functions.

If μ and ν are probability measures, then the convolution μ∗ν is the probability distribution of the sum $X + Y$ of two independent random variables X and Y whose respective distributions are μ and ν.

Bialgebras

Let $(X, \Delta, \nabla, \varepsilon, \eta)$ be a bialgebra with comultiplication Δ, multiplication ∇, unit η, and counit ε. The convolution is a product defined on the endomorphism algebra End(X) as follows. Let $\varphi, \psi \in \mathrm{End}(X)$, that is, $\varphi,\psi : X \to X$ are functions that respect all algebraic structure of X, then the convolution φ∗ψ is defined as the composition

$$X \xrightarrow{\Delta} X \otimes X \xrightarrow{\phi\otimes\psi} X \otimes X \xrightarrow{\nabla} X.$$

The convolution appears notably in the definition of Hopf algebras (Kassel 1995, §III.3). A bialgebra is a Hopf algebra if and only if it has an antipode: an endomorphism S such that

$$S * \mathrm{id}_X = \mathrm{id}_X * S = \eta \circ \varepsilon.$$

Applications

Convolution and related operations are found in many applications of engineering and mathematics.

Blurring of an image using the Gaussian function, implemented using a sequence of one-dimensional convolutions; see Gaussian blur

- In electrical engineering, the convolution of one function (the input signal) with a second function (the impulse response) gives the output of a linear time-invariant system (LTI). At any given moment, the output is an accumulated effect of all the prior values of the input function, with the most recent values typically having the most influence (expressed as a multiplicative factor). The impulse response function provides that factor as a function of the elapsed time since each input value occurred.
  - In digital signal processing and image processing applications, the entire input function is often available for computing every sample of the output function. In that case, the constraint that each output is the effect of only prior inputs can be relaxed.
  - Convolution amplifies or attenuates each frequency component of the input independently of the other components.
- In statistics, as noted above, a weighted moving average is a convolution.
- In probability theory, the probability distribution of the sum of two independent random variables is the convolution of their individual distributions.
- In optics, many kinds of "blur" are described by convolutions. A shadow (e.g., the shadow on the table when you hold your hand between the table and a light source) is the convolution of the shape of the light source that is casting the shadow and the object whose shadow is being cast. An out-of-focus photograph is the convolution of the sharp image with the shape of the iris diaphragm. The photographic term for this is bokeh.
- Similarly, in digital image processing, convolutional filtering plays an important role in many important algorithms in edge detection and related processes.
- In linear acoustics, an echo is the convolution of the original sound with a function representing the various objects that are reflecting it.
- In artificial reverberation (digital signal processing, pro audio), convolution is used to map the impulse response of a real room on a digital audio signal (see previous and next point for additional information).
- In time-resolved fluorescence spectroscopy, the excitation signal can be treated as a chain of delta pulses, and the measured fluorescence is a sum of exponential decays from each delta pulse.
- In radiotherapy treatment planning systems, most part of all modern codes of calculation applies a convolution-superposition algorithm.
- In physics, wherever there is a linear system with a "superposition principle", a convolution operation makes an appearance. For instance, given a function that describes an electric charge distribution and the function that gives the electric potential of a point charge, then the potential of the charge distribution is the convolution of these two functions.
- In kernel density estimation, a distribution is estimated from sample points by convolution with a kernel, such as an isotropic Gaussian. (Diggle 1995).

- In computational fluid dynamics, the large eddy simulation (LES) turbulence model uses the convolution operation to lower the range of length scales necessary in computation thereby reducing computational cost.

Notes

[1] According to [Lothar von Wolfersdorf (2000), "Einige Klassen quadratischer Integralgleichungen", *Sitzungsberichte der Sächsischen Akademie der Wissenschaften zu Leipzig*, *Mathematisch-naturwissenschaftliche Klasse*, volume **128**, number 2, 6–7], the source is Volterra, Vito (1913), "Leçons sur les fonctions de linges". Gauthier-Villars, Paris 1913.

[2] John Hilton Grace and Alfred Young (1903), *The algebra of invariants* (http://books.google.com/books?id=NIe4AAAAIAAJ&pg=PA40), Cambridge University Press, p. 40,

[3] Leonard Eugene Dickson (1914), *Algebraic invariants* (http://books.google.com/books?id=LRGoAAAAIAAJ&pg=PA85), J. Wiley, p. 85,

[4] R. N. Bracewell (2005), "Early work on imaging theory in radio astronomy" (http://books.google.com/books?id=v2SqL0zCrwcC&pg=PA172), in W. T. Sullivan, *The Early Years of Radio Astronomy: Reflections Fifty Years After Jansky's Discovery*, Cambridge University Press, p. 172, ISBN 9780521616027,

References

- Bracewell, R. (1986), *The Fourier Transform and Its Applications* (2nd ed.), McGraw–Hill, ISBN 0071160434.
- Hewitt, Edwin; Ross, Kenneth A. (1979), *Abstract harmonic analysis. Vol. I*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **115** (2nd ed.), Berlin, New York: Springer-Verlag, ISBN 978-3-540-09434-0, MR551496.
- Hewitt, Edwin; Ross, Kenneth A. (1970), *Abstract harmonic analysis. Vol. II: Structure and analysis for compact groups. Analysis on locally compact Abelian groups*, Die Grundlehren der mathematischen Wissenschaften, Band 152, Berlin, New York: Springer-Verlag, MR0262773.
- Hörmander, L. (1983), *The analysis of linear partial differential operators I*, Grundl. Math. Wissenschaft., **256**, Springer, ISBN 3-540-12104-8, MR0717035.
- Kassel, Christian (1995), *Quantum groups*, Graduate Texts in Mathematics, **155**, Berlin, New York: Springer-Verlag, ISBN 978-0-387-94370-1, MR1321145.
- Knuth, Donald (1997), *Seminumerical Algorithms* (3rd. ed.), Reading, Massachusetts: Addison–Wesley, ISBN 0-201-89684-2.
- Rudin, Walter (1962), *Fourier analysis on groups*, Interscience Tracts in Pure and Applied Mathematics, No. 12, Interscience Publishers (a division of John Wiley and Sons), New York–London, ISBN 047152364X, MR0152834.
- Sobolev, V.I. (2001), "Convolution of functions" (http://www.encyclopediaofmath.org/index.php?title=C/c026430), in Hazewinkel, Michiel, *Encyclopedia of Mathematics*, Springer, ISBN 978-1556080104.
- Stein, Elias; Weiss, Guido (1971), *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, ISBN 0-691-08078-X.
- Strichartz, R. (1994), *A Guide to Distribution Theory and Fourier Transforms*, CRC Press, ISBN 0849382734.
- Titchmarsh, E (1948), *Introduction to the theory of Fourier integrals* (2nd ed.), New York, N.Y.: Chelsea Pub. Co. (published 1986), ISBN 978-0828403245.
- Uludag, A. M. (1998), "On possible deterioration of smoothness under the operation of convolution", *J. Math. Anal. Appl. 227 no. 2, 335–358*
- Treves, François (1967), *Topological Vector Spaces, Distributions and Kernels*, Academic Press, ISBN 0486453529.
- von zur Gathen, J.; Gerhard, J. (2003), *Modern Computer Algebra*, Cambridge University Press, ISBN 0-521-82646-2.
- Diggle, P. J. (1995), "A kernel method for smoothing point process data", *Journal of the Royal Statistical Society, Series C) 34 (1985) 138–147*

External links

- Earliest Uses: The entry on Convolution has some historical information. (http://jeff560.tripod.com/c.html)
- Convolution (http://rkb.home.cern.ch/rkb/AN16pp/node38.html#SECTION000380000000000000000), on The Data Analysis BriefBook (http://rkb.home.cern.ch/rkb/titleA.html)
- http://www.jhu.edu/~signals/convolve/index.html Visual convolution Java Applet
- http://www.jhu.edu/~signals/discreteconv2/index.html Visual convolution Java Applet for discrete-time functions
- Lectures on Image Processing: A collection of 18 lectures in pdf format from Vanderbilt University. Lecture 7 is on 2-D convolution. (http://www.archive.org/details/Lectures_on_Image_Processing), by Alan Peters
  - http://www.vuse.vanderbilt.edu/~rap2/EECE253/EECE253_01_Intro.pdf
- Convolution Kernel Mask Operation Interactive tutorial (http://micro.magnet.fsu.edu/primer/java/digitalimaging/processing/kernelmaskoperation/)
- Convolution (http://mathworld.wolfram.com/Convolution.html) at MathWorld
- GNU C-Graph (http://www.gnu.org/software/c-graph): Free Software for visualizing convolution
- Freeverb3 Impulse Response Processor (http://freeverb3.sourceforge.net/): Opensource zero latency impulse response processor with VST plugins
- Stanford University CS 178 interactive Flash demo (http://graphics.stanford.edu/courses/cs178/applets/convolution.html) showing how spatial convolution works.

Appendix: Systems & Laplace

Ordinary differential equation

In mathematics, an **ordinary differential equation** (**ODE**) is an equation in which there is only one independent variable and one or more derivatives of a dependent variable with respect to the independent variable, so that all the derivatives occurring in the equation are ordinary derivatives.[1][2]

A simple example is Newton's second law of motion—the relationship between the displacement and the time of the object under the force—which leads to the differential equation

$$m\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} = F(x(t)),$$

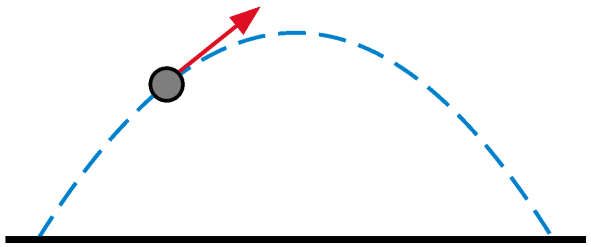
for the motion of a particle of constant mass *m*. In general, the force *F* depends upon the position *x(t)* of the particle at time *t*, and thus the unknown function *x(t)* appears on both sides of the differential equation, as is indicated in the notation $F(x(t))$.[3][4][5][6]

Ordinary differential equations are distinguished from partial differential equations, which involve partial derivatives of functions of several variables.

Ordinary differential equations arise in many different contexts including geometry, mechanics, astronomy and population modelling. Many mathematicians have studied differential equations and contributed to the field, including Newton, Leibniz, the Bernoulli family, Riccati, Clairaut, d'Alembert and Euler.

Much study has been devoted to the solution of ordinary differential equations. In the case where the equation is linear, it can be solved by analytical methods. Unfortunately, most of the interesting differential equations are non-linear and, with a few exceptions, cannot be solved exactly. Approximate solutions are arrived at using computer approximations (see numerical ordinary differential equations).

Definitions



The trajectory of a projectile launched from a cannon follows a curve determined by an ordinary differential equation that is derived from Newton's second law.

Ordinary differential equation

Let *y* be an unknown function

$$y : \mathbb{R} \to \mathbb{R}$$

in *x* with $y^{(n)}$ the *n*th derivative of *y*, and let *F* be a given function

$$F : \mathbb{R}^{n+1} \to \mathbb{R},$$

then an equation of the form

$$F\left(x, y, y', \ \ldots, \ y^{(n-1)}\right) = y^{(n)}$$

is called an **ordinary differential equation** of **order** *n*.[7][8] If *y* is an unknown vector valued function

$$y : \mathbb{R} \to \mathbb{R}^m,$$

it is called a **system of ordinary differential equations** of **dimension** *m* (in this case, $F : \mathbb{R}^{m(n+1)} \to \mathbb{R}^m$).

More generally, an **implicit** ordinary differential equation of order *n* takes the form

$$F\left(x, y, y', y'', \ \ldots, \ y^{(n)}\right) = 0$$

where $F : \mathbb{R}^{n+2} \to \mathbb{R}$ depends on $y^{(n)}$.[9] To distinguish the above case from this one, an equation of the form

$$F\left(x, y, y', y'', \ \ldots, \ y^{(n-1)}\right) = y^{(n)}$$

is called an **explicit** differential equation.

A differential equation not depending on x is called **autonomous**.

A differential equation is said to be **linear** if F can be written as a linear combination of the derivatives of y together with a constant term, all possibly depending on x:

$$y^{(n)} = \sum_{i=0}^{n-1} a_i(x) y^{(i)} + r(x)$$

with $a_i(x)$ and $r(x)$ continuous functions in x.[10][11][12] The function $r(x)$ is called the **source term**; if $r(x)=0$ then the linear differential equation is called **homogeneous**, otherwise it is called **non-homogeneous** or **inhomogeneous**.[13][14]

Solutions

Given a differential equation

$$F(x, y, y', \ldots, y^{(n)}) = 0$$

a function $u: I \subset \mathbf{R} \to \mathbf{R}$ is called the **solution** or integral curve for F, if u is n-times differentiable on I, and

$$F(x, u, u', \ \ldots, \ u^{(n)}) = 0 \quad x \in I.$$

Given two solutions $u: J \subset \mathbf{R} \to \mathbf{R}$ and $v: I \subset \mathbf{R} \to \mathbf{R}$, u is called an **extension** of v if $I \subset J$ and

$$u(x) = v(x) \quad x \in I.$$

A solution which has no extension is called a **maximal solution**. A solution defined on all of **R** is called a **global solution**.

A **general solution** of an n-th order equation is a solution containing n arbitrary independent constants of integration. A **particular solution** is derived from the general solution by setting the constants to particular values, often chosen to fulfill set 'initial conditions or boundary conditions'.[15] A singular solution is a solution which cannot be obtained by assigning definite values to the arbitrary constants in the general solution.[16]

Applications

Ordinary differential equations describe the basic mathematical theory and methods of the natural sciences and social sciences which govern objects and phenomena, evolution and variation. Many principles and rules in physical, chemical, biological, engineering, aerospace, medical, economic and financial fields of study can be described by the appropriate ordinary differential equations, such as Newtons laws of motion, Newton's law of universal gravitation, the law of conservation of energy, the law of population growth, ecological population competition, infectious diseases, genetic variation, stock trends, interest rates and the market equilibrium price changes. People attribute the understanding and analysis of these problems to the study of the corresponding ordinary differential equations to describe the mathematical model. Therefore, the theory and methods of ordinary differential equations are widely used in various fields of social science.

Existence and uniqueness of solutions

There are several theorems that establish existence and uniqueness of solutions to initial value problems involving ODEs both locally and globally. The two main theorems are

theorem	assumption	conclusion
Peano existence theorem	F continuous	local existence only
Picard–Lindelöf theorem	F Lipschitz continuous	local existence and uniqueness

which are both local results.

Global uniqueness and maximum domain of solution

When the hypothesis of the Picard–Lindelöf theorem are satisfied, then local existence and uniqueness can be extended to a global result. More precisely:

Theorem[17] For each initial condition (x_0, y_0)there exists an unique maximum (possibly infinite) open interval

$I_{max} =]x_-, x_+[, x_\pm \in \mathbb{R} \cup \pm\infty, x_0 \in I_{max}$

such that any solution which satisfies this initial condition is a restriction of the solution which satisfies this initial condition with domain I_{max}.

In the case that $x_\pm \neq \pm\infty$, there are exactly two possibilities

- explosion in finite time: $\lim_{x \to x_\pm} ||y(x)|| = \infty$
- leaves domain of definition: $\lim_{x \to x_\pm} \in \partial\bar{\Omega}$

where $\Omega$is the open set in which F is defined, and $\partial\bar{\Omega}$is its boundary.

Note that the maximum domain of the solution

- is always an interval (to have uniqueness)
- may be smaller than $\mathbb{R}$
- may depend on the specific choice of (x_0, y_0)

Example $y' = y^2$

This means that $F(x, y) = y^2$, which is $C^1$and therefore Lipschitz continuous for all y, satisfying the Picard–Lindelöf theorem.

Even in such a simple setting, the maximum domain of solution cannot be all $\mathbb{R}$, since the solution is

$$y(x) = \frac{y_0}{(x_0 - x)y_0 + 1}$$

which has maximum domain:

$$\begin{cases} \mathbb{R} & y_0 = 0 \\ ]-\infty, x_0 + \frac{1}{y_0}, [& y_0 > 0 \\ ]x_0 + \frac{1}{y_0}, +\infty[& y_0 < 0 \end{cases}$$

This shows clearly that the maximum interval may depend on the initial conditions.

We could take the domain of y as being $\mathbb{R} \setminus x_0 + \frac{1}{y_0}$, but this would lead to a domain that is not an interval, so that the side opposite to the initial condition would be disconnected from the initial condition, and therefore not uniquely determined by it.

The maximum domain is not $\mathbb{R}$ because $\lim_{x \to x_\pm} ||y(x)|| = \infty$, which is one of the two possible cases according to the above theorem.

Reduction to a first order system

Any differential equation of order *n* can be written as a system of *n* first-order differential equations. Given an explicit ordinary differential equation of order *n* (and dimension 1),

$$F\left(x, y, y', y'', \ \ldots, \ y^{(n-1)}\right) = y^{(n)}$$

define a new family of unknown functions

$$y_i := y^{(i-1)}.$$

for *i* from 1 to *n*.

The original differential equation can be rewritten as the system of differential equations with order 1 and dimension *n* given by

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= y_3 \\ &\vdots \\ y_{n-1}' &= y_n \\ y_n' &= F(x, y_1, \ldots, y_n). \end{aligned}$$

which can be written concisely in vector notation as

$$\mathbf{y}' = \mathbf{F}(x, \mathbf{y})$$

with

$$\mathbf{y} := (y_1, \ldots, y_n)$$

and

$$\mathbf{F}(x, y_1, \ldots, y_n) = (y_2, \ldots, y_n, F(x, y_1, \ldots, y_n)).$$

Linear ordinary differential equations

A well understood particular class of differential equations is linear differential equations. We can always reduce an explicit linear differential equation of any order to a system of differential equations of order 1

$$y_i'(x) = \sum_{j=1}^{n} a_{i,j}(x) y_j + b_i(x), \quad i = 1, \ldots, n$$

which we can write concisely using matrix and vector notation as

$$\mathbf{y}'(x) = \mathbf{A}(x)\mathbf{y}(x) + \mathbf{b}(x)$$

with

$$\begin{aligned} \mathbf{y}(x) &:= (y_1(x), \ldots, y_n(x)) \\ \mathbf{b}(x) &:= (b_1(x), \ldots, b_n(x)) \\ \mathbf{A}(x) &:= (a_{i,j}(x)), \quad i, j = 1, \ldots, n. \end{aligned}$$

Homogeneous equations

The set of solutions for a system of homogeneous linear differential equations of order 1 and dimension *n*

$$\mathbf{y}'(x) = \mathbf{A}(x)\mathbf{y}(x)$$

forms an *n*-dimensional vector space. Given a basis for this vector space $\mathbf{z}_1(x), \ldots, \mathbf{z}_n(x)$, which is called a **fundamental system**, every solution $\mathbf{s}(x)$ can be written as

$$\mathbf{s}(x) = \sum_{i=1}^{n} c_i \mathbf{z}_i(x).$$

The $n \times n$ matrix

$$\mathbf{Z}(x) := (\mathbf{z}_1(x), \ldots, \mathbf{z}_n(x))$$

is called **fundamental matrix**. In general there is no method to explicitly construct a fundamental system, but if one solution is known d'Alembert reduction can be used to reduce the dimension of the differential equation by one.

Nonhomogeneous equations

The set of solutions for a system of inhomogeneous linear differential equations of order 1 and dimension *n*

$$\mathbf{y}'(x) = \mathbf{A}(x)\mathbf{y}(x) + \mathbf{b}(x)$$

can be constructed by finding the fundamental system $\mathbf{z}_1(x), \ldots, \mathbf{z}_n(x)$ to the corresponding homogeneous equation and one particular solution $\mathbf{p}(x)$ to the inhomogeneous equation. Every solution $\mathbf{s}(x)$ to nonhomogeneous equation can then be written as

$$\mathbf{s}(x) = \sum_{i=1}^{n} c_i \mathbf{z}_i(x) + \mathbf{p}(x).$$

A particular solution to the nonhomogeneous equation can be found by the method of undetermined coefficients or the method of variation of parameters.

Concerning second order linear ordinary differential equations, it is well known that

$$y = e^{\int s\,dx} \Rightarrow y'' + Py' + \left(-s' - s^2 - sP\right) y = 0.$$

So, if y_h is a solution of: $y'' + Py' + Qy = 0$, then $\exists s = \frac{y_h'}{y_h}$ such that: $Q = -s' - s^2 - sP.$

So, if y_h is a solution of: $y'' + Py' + Qy = 0$; then a particular solution y_p of $y'' + Py' + Qy = W$, is given by:

$$y_p = y_h \int \left(\frac{1}{y_h^2} \int W y_h e^{\int P\,dx}\,dx \right) e^{-\int P\,dx}\,dx\,.$$ [18]

Fundamental systems for homogeneous equations with constant coefficients

If a system of homogeneous linear differential equations has *constant* coefficients

$$\mathbf{y}'(x) = \mathbf{A}\mathbf{y}(x)$$

then we *can* explicitly construct a fundamental system. The fundamental system can be written as a matrix differential equation

$$\mathbf{Y}' = \mathbf{A}\mathbf{Y}$$

with solution as a matrix exponential

$$e^{x\mathbf{A}}$$

which is a fundamental matrix for the original differential equation. To explicitly calculate this expression we first transform **A** into Jordan normal form

$$e^{x\mathbf{A}} = e^{x\mathbf{C}^{-1}\mathbf{J}\mathbf{C}^{1}} = \mathbf{C}^{-1}e^{x\mathbf{J}}\mathbf{C}^{1}$$

and then evaluate the Jordan blocks

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

of *J* separately as

$$e^{x\mathbf{J_i}} = e^{\lambda_i x} \begin{bmatrix} 1 & x & \frac{x^2}{2} & \cdots & \frac{x^{n-1}}{(n-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{x^2}{2} \\ & & & \ddots & x \\ & & & & 1 \end{bmatrix}.$$

General Case

To solve

$\mathbf{y}'(x) = A(x)\mathbf{y}(x)+\mathbf{b}(x)$ with $\mathbf{y}(x_0) = \mathbf{y_0}$ (here $\mathbf{y}(x)$ is a vector or matrix, and $A(x)$ is a matrix),

let $U(x)$ be the solution of $\mathbf{y}'(x) = A(x)\mathbf{y}(x)$ with $U(x_0) = I$ (the identity matrix). After substituting $\mathbf{y}(x) = U(x)\mathbf{z}(x)$, the equation $\mathbf{y}'(x) = A(x)\mathbf{y}(x)+\mathbf{b}(x)$ simplifies to $U(x)z'(x) = \mathbf{b}(x)$. Thus,

$$\mathbf{y}(x) = U(x)\mathbf{y_0} + U(x)\int_{x_0}^{x} U^{-1}(x)\mathbf{b}(x)\,dx$$

If $A(x_1)$ commutes with $A(x_2)$ for all x_1 and x_2, then $U(x) = e^{\int_{x_0}^{x} A(x)\,dx}$ (and thus $U^{-1}(x) = e^{-\int_{x_0}^{x} A(x)\,dx}$), but in the general case there is no closed form solution, and an approximation method such as Magnus expansion may have to be used.

Theories of ODEs

Singular solutions

The theory of singular solutions of ordinary and partial differential equations was a subject of research from the time of Leibniz, but only since the middle of the nineteenth century did it receive special attention. A valuable but little-known work on the subject is that of Houtain (1854). Darboux (starting in 1873) was a leader in the theory, and in the geometric interpretation of these solutions he opened a field which was worked by various writers, notably Casorati and Cayley. To the latter is due (1872) the theory of singular solutions of differential equations of the first order as accepted circa 1900.

Reduction to quadratures

The primitive attempt in dealing with differential equations had in view a reduction to quadratures. As it had been the hope of eighteenth-century algebraists to find a method for solving the general equation of the n th degree, so it was the hope of analysts to find a general method for integrating any differential equation. Gauss (1799) showed, however, that the differential equation meets its limitations very soon unless complex numbers are introduced. Hence analysts began to substitute the study of functions, thus opening a new and fertile field. Cauchy was the first to appreciate the importance of this view. Thereafter the real question was to be, not whether a solution is possible by means of known functions or their integrals, but whether a given differential equation suffices for the definition of a function of the independent variable or variables, and if so, what are the characteristic properties of this function.

Fuchsian theory

Two memoirs by Fuchs (*Crelle*, 1866, 1868), inspired a novel approach, subsequently elaborated by Thomé and Frobenius. Collet was a prominent contributor beginning in 1869, although his method for integrating a non-linear system was communicated to Bertrand in 1868. Clebsch (1873) attacked the theory along lines parallel to those followed in his theory of Abelian integrals. As the latter can be classified according to the properties of the fundamental curve which remains unchanged under a rational transformation, so Clebsch proposed to classify the transcendent functions defined by the differential equations according to the invariant properties of the corresponding surfaces $f = 0$ under rational one-to-one transformations.

Lie's theory

From 1870 Sophus Lie's work put the theory of differential equations on a more satisfactory foundation. He showed that the integration theories of the older mathematicians can, by the introduction of what are now called Lie groups, be referred to a common source; and that ordinary differential equations which admit the same infinitesimal transformations present comparable difficulties of integration. He also emphasized the subject of transformations of contact.

Lie's group theory of differential equations, has been certified, namely: (1) that it unifies the many ad hoc methods known for solving differential equations, and (2) that it provides powerful new ways to find solutions. The theory has applications to both ordinary and partial differential equations.[19]

A general approach to solve DE's uses the symmetry property of differential equations, the continuous infinitesimal transformations of solutions to solutions (Lie theory). Continuous group theory, Lie algebras and differential geometry are used to understand the structure of linear and nonlinear (partial) differential equations for generating integrable equations, to find its Lax pairs, recursion operators, Bäcklund transform and finally finding exact analytic solutions to the DE.

Symmetry methods have been recognized to study differential equations arising in mathematics, physics, engineering, and many other disciplines.

Sturm–Liouville theory

Sturm–Liouville theory is a theory of eigenvalues and eigenfunctions of linear operators defined in terms of second-order homogeneous linear equations, and is useful in the analysis of certain partial differential equations.

Software for ODE solving

- FuncDesigner (free license: BSD, uses Automatic differentiation, also can be used online via Sage-server [20])
- odeint [21] - A C++ library for solving ordinary differential equations numerically
- VisSim [22] - a visual language for differential equation solving
- Mathematical Assistant on Web [23] online solving first order (linear and with separated variables) and second order linear differential equations (with constant coefficients), including intermediate steps in the solution.
- DotNumerics: Ordinary Differential Equations for C# and VB.NET [24] Initial-value problem for nonstiff and stiff ordinary differential equations (explicit Runge-Kutta, implicit Runge-Kutta, Gear's BDF and Adams-Moulton).
- Online experiments with JSXGraph [25]
- Maxima computer algebra system (GPL)
- COPASI a free (Artistic License 2.0) software package for the integration and analysis of ODEs.
- MATLAB a matrix-programming software (MATrix LABoratory)

Notes

[1] Kreyszig (1972, p. 1)
[2] Simmons (1972, p. 2)
[3] Kreyszig (1972, p. 64)
[4] Simmons (1972, pp. 1,2)
[5] Halliday & Resnick (1977, p. 78)
[6] Tipler (1991, pp. 78-83)
[7] Harper (1976, p. 127)
[8] Kreyszig (1972, p. 2)
[9] Simmons (1972, p. 3)
[10] Harper (1976, p. 127)
[11] Kreyszig (1972, p. 24)
[12] Simmons (1972, p. 47)
[13] Harper (1976, p. 128)
[14] Kreyszig (1972, p. 24)
[15] Kreyszig (1972, p. 78)
[16] Kreyszig (1972, p. 4)
[17] Boscain; Chitour 2011, p.21
[18] Polyanin, Andrei D.; Valentin F. Zaitsev (2003). *Handbook of Exact Solutions for Ordinary Differential Equations, 2nd. Ed.*. Chapman & Hall/CRC. ISBN 1-58488-297-2.
[19] Lawrence (1999, p. 9)
[20] http://sage.openopt.org/welcome
[21] https://github.com/headmyshoulder/odeint-v2
[22] http://www.vissim.us
[23] http://user.mendelu.cz/marik/maw/index.php?lang=en&form=ode
[24] http://www.dotnumerics.com/NumericalLibraries/DifferentialEquations/
[25] http://jsxgraph.uni-bayreuth.de/wiki/index.php/Differential_equations

References

- Halliday, David; Resnick, Robert (1977), *Physics* (3rd ed.), New York: Wiley
- Harper, Charlie (1976), *Introduction to Mathematical Physics*, New Jersey: Prentice-Hall
- Kreyszig, Erwin (1972), *Advanced Engineering Mathematics* (3rd ed.), New York: Wiley.
- Polyanin, A. D. and V. F. Zaitsev, *Handbook of Exact Solutions for Ordinary Differential Equations (2nd edition)", Chapman & Hall/CRC Press, Boca Raton, 2003. ISBN 1-58488-297-2*
- Simmons, George F. (1972), *Differential Equations with Applications and Historical Notes*, New York: McGraw-Hill
- Tipler, Paul A. (1991), *Physics for Scientists and Engineers: Extended version* (3rd ed.), New York: Worth Publishers
- Boscain, Ugo; Chitour, Yacine (2011) (in french), *Introduction à l'automatique* (http://www.cmapx.polytechnique.fr/~boscain/poly2011.pdf)
- Lawrence, Dresner (1999), *Applications of Lie's Theory of Ordinary and Partial Differential Equations*, Bristol and Philadelphia: Institute of Physics Publishing

Bibliography

- Coddington, Earl A.; Levinson, Norman (1955). *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.
- Hartman, Philip, *Ordinary Differential Equations, 2nd Ed.*, Society for Industrial & Applied Math, 2002. ISBN 0-89871-510-5.
- W. Johnson, *A Treatise on Ordinary and Partial Differential Equations* (http://www.hti.umich.edu/cgi/b/bib/bibperm?q1=abv5010.0001.001), John Wiley and Sons, 1913, in University of Michigan Historical Math Collection (http://hti.umich.edu/u/umhistmath/)
- E.L. Ince, *Ordinary Differential Equations*, Dover Publications, 1958, ISBN 0-486-60349-0
- Witold Hurewicz, *Lectures on Ordinary Differential Equations*, Dover Publications, ISBN 0-486-49510-8
- Ibragimov, Nail H (1993). *CRC Handbook of Lie Group Analysis of Differential Equations Vol. 1-3*. Providence: CRC-Press. ISBN 0-8493-4488-3.
- Teschl, Gerald. *Ordinary Differential Equations and Dynamical Systems* (http://www.mat.univie.ac.at/~gerald/ftp/book-ode/). Providence: American Mathematical Society.
- A. D. Polyanin, V. F. Zaitsev, and A. Moussiaux, Handbook of First Order Partial Differential Equations*, Taylor & Francis, London, 2002. ISBN 0-415-27267-X*
- D. Zwillinger, *Handbook of Differential Equations (3rd edition)*, Academic Press, Boston, 1997.

External links

- Differential Equations (http://www.dmoz.org/Science/Math/Differential_Equations//) at the Open Directory Project (includes a list of software for solving differential equations).
- EqWorld: The World of Mathematical Equations (http://eqworld.ipmnet.ru/index.htm), containing a list of ordinary differential equations with their solutions.
- Online Notes / Differential Equations (http://tutorial.math.lamar.edu/classes/de/de.aspx) by Paul Dawkins, Lamar University.
- Differential Equations (http://www.sosmath.com/diffeq/diffeq.html), S.O.S. Mathematics.
- A primer on analytical solution of differential equations (http://numericalmethods.eng.usf.edu/mws/gen/08ode/mws_gen_ode_bck_primer.pdf) from the Holistic Numerical Methods Institute, University of South Florida.
- Ordinary Differential Equations and Dynamical Systems (http://www.mat.univie.ac.at/~gerald/ftp/book-ode/) lecture notes by Gerald Teschl.
- Notes on Diffy Qs: Differential Equations for Engineers (http://www.jirka.org/diffyqs/) An introductory textbook on differential equations by Jiri Lebl of UIUC.

Laplace transform applied to differential equations

The Laplace transform is a powerful integral transform used to switch a function from the time domain to the S-domain. The use of Laplace transform makes it much easier to solve linear differential equations with given initial conditions.

First consider the following relations:

$$\mathcal{L}\{f'\} = s\mathcal{L}\{f\} - f(0)$$
$$\mathcal{L}\{f''\} = s^2\mathcal{L}\{f\} - sf(0) - f'(0)$$
$$\mathcal{L}\{f^{(n)}\} = s^n\mathcal{L}\{f\} - \Sigma_{i=1}^{n} s^{n-i} f^{(i-1)}(0).$$

Consider the following differential equation:

$$\sum_{i=0}^{n} a_i f^{(i)}(t) = \phi(t).$$

This equation is equivalent to

$$\sum_{i=0}^{n} a_i \mathcal{L}\{f^{(i)}(t)\} = \mathcal{L}\{\phi(t)\}$$

which is equivalent to

$$\mathcal{L}\{f(t)\} = \frac{\mathcal{L}\{\phi(t)\} + \sum_{i=1}^{n} a_i \sum_{j=1}^{i} s^{i-j} f^{(j-1)}(0)}{\sum_{i=0}^{n} a_i s^i}.$$

Note that the $f^{(k)}(0)$ are initial conditions.

The solution for *f*(*t*) will be given by applying the inverse Laplace transform to $\mathcal{L}\{f(t)\}$.

An example

We want to solve

$$f''(t) + 4f(t) = \sin(2t)$$

with initial conditions f(0) = 0 and $f'(0)$=0.

We note that

$$\phi(t) = \sin(2t)$$

and we get

$$\mathcal{L}\{\phi(t)\} = \frac{2}{s^2 + 4}$$

So this is equivalent to

$$s^2\mathcal{L}\{f(t)\} - sf(0) - f'(0) + 4\mathcal{L}\{f(t)\} = \mathcal{L}\{\phi(t)\}$$

We deduce

$$\mathcal{L}\{f(t)\} = \frac{2}{(s^2 + 4)^2}$$

So we apply the Laplace inverse transform and get

$$f(t) = \frac{1}{8}\sin(2t) - \frac{t}{4}\cos(2t)$$

Bibliography

- A. D. Polyanin, *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, Boca Raton, 2002. ISBN 1-58488-299-9

External links

Examples:http://www.esr.ruhr-uni-bochum.de/rt1/syscontrol/node11.html

Fourier transform

The **Fourier transform** is a mathematical operation with many applications in physics and engineering that expresses a mathematical function of time as a function of frequency, known as its frequency spectrum; Fourier's theorem guarantees that this can always be done. For instance, the transform of a musical chord made up of pure notes (without overtones) expressed as amplitude as a function of time, is a mathematical representation of the amplitudes and phases of the individual notes that make it up. The function of time is often called the *time domain* representation, and the frequency spectrum the *frequency domain* representation. The inverse Fourier transform expresses a frequency domain function in the time domain. Each value of the function is usually expressed as a complex number (called *complex amplitude*) that can be interpreted as a magnitude and a phase component. The term "Fourier transform" refers to both the transform operation and to the complex-valued function it produces.

In the case of a **periodic function**, such as a continuous, but not necessarily sinusoidal, musical tone, the Fourier transform can be simplified to the calculation of a discrete set of complex amplitudes, called Fourier series coefficients. Also, when a time-domain function is **sampled** to facilitate storage or computer-processing, it is still possible to recreate a version of the original Fourier transform according to the Poisson summation formula, also known as discrete-time Fourier transform. These topics are addressed in separate articles. For an overview of those and other related operations, refer to **Fourier analysis** or **List of Fourier-related transforms**.

Definition

There are several common conventions for defining the Fourier transform $\hat{f}$ of an integrable function $f : \mathbf{R} \to \mathbf{C}$ (Kaiser 1994). This article will use the definition:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \xi}\, dx, \quad \text{for every real number } \xi.$$

When the independent variable x represents *time* (with SI unit of seconds), the transform variable ξ represents frequency (in hertz). Under suitable conditions, f can be reconstructed from $\hat{f}$ by the **inverse transform**:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)\, e^{2\pi i \xi x}\, d\xi, \quad \text{for every real number } x.$$

For other common conventions and notations, including using the angular frequency ω instead of the frequency ξ, see Other conventions and Other notations below. The Fourier transform on Euclidean space is treated separately, in which the variable x often represents position and ξ momentum.

Introduction

The motivation for the Fourier transform comes from the study of Fourier series. In the study of Fourier series, complicated functions are written as the sum of simple waves mathematically represented by sines and cosines. Due to the properties of sine and cosine, it is possible to recover the amplitude of each wave in the sum by an integral. In many cases it is desirable to use Euler's formula, which states that $e^{2\pi i\theta} = \cos 2\pi\theta + i \sin 2\pi\theta$, to write Fourier series in terms of the basic waves $e^{2\pi i\theta}$. This has the advantage of simplifying many of the formulas involved, and provides a formulation for Fourier series that more closely resembles the definition followed in this article. Re-writing sines and cosines as complex exponentials makes it necessary for the Fourier coefficients to be complex valued. The usual interpretation of this complex number is that it gives both the amplitude (or size) of the wave present in the function and the phase (or the initial angle) of the wave. These complex exponentials sometimes contain negative "frequencies". If θ is measured in seconds, then the waves $e^{2\pi i\theta}$ and $e^{-2\pi i\theta}$ both complete one cycle per second, but they represent different frequencies in the Fourier transform. Hence, frequency no longer measures the number of cycles per unit time, but is still closely related.

There is a close connection between the definition of Fourier series and the Fourier transform for functions f which are zero outside of an interval. For such a function, we can calculate its Fourier series on any interval that includes the points where f is not identically zero. The Fourier transform is also defined for such a function. As we increase the length of the interval on which we calculate the Fourier series, then the Fourier series coefficients begin to look like the Fourier transform and the sum of the Fourier series of f begins to look like the inverse Fourier transform. To explain this more precisely, suppose that T is large enough so that the interval $[-T/2,T/2]$ contains the interval on which f is not identically zero. Then the n-th series coefficient c_n is given by:

$$c_n = \int_{-T/2}^{T/2} f(x)\ e^{-2\pi i(n/T)x} dx.$$

Comparing this to the definition of the Fourier transform, it follows that $c_n = \hat{f}(n/T)$ since $f(x)$ is zero outside $[-T/2,T/2]$. Thus the Fourier coefficients are just the values of the Fourier transform sampled on a grid of width $1/T$. As T increases the Fourier coefficients more closely represent the Fourier transform of the function.

Under appropriate conditions, the sum of the Fourier series of f will equal the function f. In other words, f can be written:

$$f(x) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \hat{f}(n/T)\ e^{2\pi i(n/T)x} = \sum_{n=-\infty}^{\infty} \hat{f}(\xi_n)\ e^{2\pi i\xi_n x} \Delta\xi,$$

where the last sum is simply the first sum rewritten using the definitions $\xi_n = n/T$, and $\Delta\xi = (n + 1)/T - n/T = 1/T$.

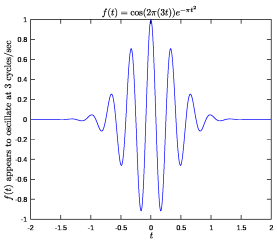
This second sum is a Riemann sum, and so by letting $T \to \infty$ it will converge to the integral for the inverse Fourier transform given in the definition section. Under suitable conditions this argument may be made precise (Stein & Shakarchi 2003).

In the study of Fourier series the numbers c_n could be thought of as the "amount" of the wave present in the Fourier series of f. Similarly, as seen above, the Fourier transform can be thought of as a function that measures how much of each individual frequency is present in our function f, and we can recombine these waves by using an integral (or "continuous sum") to reproduce the original function.

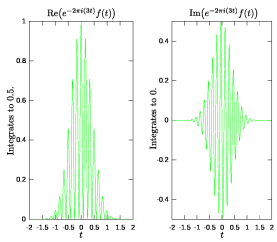
Example

The following images provide a visual illustration of how the Fourier transform measures whether a frequency is present in a particular function. The function depicted $f(t) = \cos(6\pi t)e^{-\pi t^2}$ oscillates at 3 hertz (if t measures seconds) and tends quickly to 0. (Note: the second term in this equation is an envelope function that shapes the continuous sinusoid into a short pulse. Its general form is a Gaussian function). This function was specially chosen to have a real Fourier transform which can easily be plotted. The first image contains its graph. In order to calculate $\hat{f}(3)$ we must integrate $e^{-2\pi i(3t)}f(t)$. The second image shows the plot of the real and imaginary parts of this

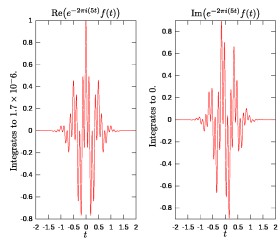
function. The real part of the integrand is almost always positive, this is because when $f(t)$ is negative, then the real part of $e^{-2\pi i(3t)}$ is negative as well. Because they oscillate at the same rate, when $f(t)$ is positive, so is the real part of $e^{-2\pi i(3t)}$. The result is that when you integrate the real part of the integrand you get a relatively large number (in this case 0.5). On the other hand, when you try to measure a frequency that is not present, as in the case when we look at $\hat{f}(5)$, the integrand oscillates enough so that the integral is very small. The general situation may be a bit more complicated than this, but this in spirit is how the Fourier transform measures how much of an individual frequency is present in a function $f(t)$.



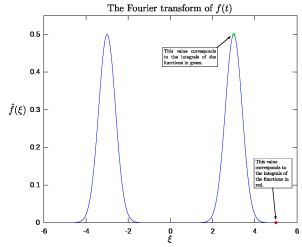
Original function showing oscillation 3 hertz.



Real and imaginary parts of integrand for Fourier transform at 3 hertz



Real and imaginary parts of integrand for Fourier transform at 5 hertz



Fourier transform with 3 and 5 hertz labeled.

Properties of the Fourier transform

Here we assume $f(x)$, $g(x)$, and $h(x)$ are *integrable functions*, are Lebesgue-measurable on the real line, and satisfy:

$$\int_{-\infty}^{\infty} |f(x)|\,dx < \infty.$$

We denote the Fourier transforms of these functions by $\hat{f}(\xi)$, $\hat{g}(\xi)$, and $\hat{h}(\xi)$ respectively.

Basic properties

The Fourier transform has the following basic properties: (Pinsky 2002).

Linearity

For any complex numbers a and b, if $h(x) = af(x) + bg(x)$, then $\hat{h}(\xi) = a \cdot \hat{f}(\xi) + b \cdot \hat{g}(\xi)$.

Translation

For any real number x_0, if $h(x) = f(x - x_0)$, then $\hat{h}(\xi) = e^{-2\pi i x_0 \xi}\hat{f}(\xi)$.

Modulation

For any real number ξ_0, if $h(x) = e^{2\pi i x \xi_0} f(x)$, then $\hat{h}(\xi) = \hat{f}(\xi - \xi_0)$.

Scaling

For a non-zero real number a, if $h(x) = f(ax)$, then $\hat{h}(\xi) = \frac{1}{|a|}\hat{f}\left(\frac{\xi}{a}\right)$. The case $a = -1$ leads to the *time-reversal* property, which states: if $h(x) = f(-x)$, then $\hat{h}(\xi) = \hat{f}(-\xi)$.

Conjugation

If $h(x) = \overline{f(x)}$, then $\hat{h}(\xi) = \overline{\hat{f}(-\xi)}$.

In particular, if f is real, then one has the *reality condition* $\hat{f}(-\xi) = \overline{\hat{f}(\xi)}$.

And if f is purely imaginary, then $\hat{f}(-\xi) = -\overline{\hat{f}(\xi)}$.

Duality

If $h(x) = \hat{f}(x)$ then $\hat{h}(\xi) = f(-\xi)$.

Convolution

If $h(x) = (f * g)(x)$, then $\hat{h}(\xi) = \hat{f}(\xi) \cdot \hat{g}(\xi)$.

Uniform continuity and the Riemann–Lebesgue lemma

The Fourier transform may be defined in some cases for non-integrable functions, but the Fourier transforms of integrable functions have several strong properties.

The Fourier transform $\hat{f}$ of any integrable function f is uniformly continuous and $\|\hat{f}\|_\infty \leq \|f\|_1$ (Katznelson 1976). By the *Riemann–Lebesgue lemma* (Stein & Weiss 1971),

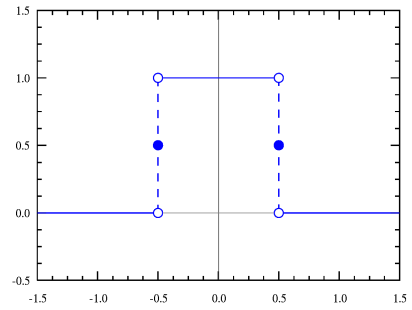
$$\hat{f}(\xi) \to 0 \text{ as } |\xi| \to \infty.$$

Furthermore, $\hat{f}$ is bounded and continuous, but need not be integrable. For example, the Fourier transform of the rectangular function, which is integrable, is the sinc function, which is not Lebesgue integrable, because its improper integrals behave analogously to the alternating harmonic series, in converging to a sum without being absolutely convergent.

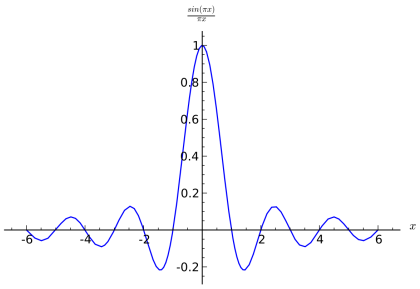
It is not generally possible to write the *inverse transform* as a Lebesgue integral. However, when both f and $\hat{f}$ are integrable, the inverse equality

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2i\pi x\xi}\, d\xi$$

holds almost everywhere. That is, the Fourier transform is injective on $L^1(\mathbf{R})$. (But if f is continuous, then equality holds for every x.)



The rectangular function is Lebesgue integrable.



The sinc function, which is the Fourier transform of the rectangular function, is bounded and continuous, but not Lebesgue integrable.

Plancherel theorem and Parseval's theorem

Let $f(x)$ and $g(x)$ be integrable, and let $\hat{f}(\xi)$ and $\hat{g}(\xi)$ be their Fourier transforms. If $f(x)$ and $g(x)$ are also square-integrable, then we have Parseval's theorem (Rudin 1987, p. 187):

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)}\,\mathrm{d}x = \int_{-\infty}^{\infty} \hat{f}(\xi)\overline{\hat{g}(\xi)}\,\mathrm{d}\xi,$$

where the bar denotes complex conjugation.

The Plancherel theorem, which is equivalent to Parseval's theorem, states (Rudin 1987, p. 186):

$$\int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x = \int_{-\infty}^{\infty} \left|\hat{f}(\xi)\right|^2\,\mathrm{d}\xi.$$

The Plancherel theorem makes it possible to define the Fourier transform for functions in $L^2(\mathbf{R})$, as described in Generalizations below. The Plancherel theorem has the interpretation in the sciences that the Fourier transform preserves the energy of the original quantity. It should be noted that depending on the author either of these theorems might be referred to as the Plancherel theorem or as Parseval's theorem.

See Pontryagin duality for a general formulation of this concept in the context of locally compact abelian groups.

Poisson summation formula

The Poisson summation formula (PSF) is an equation that relates the Fourier series coefficients of the periodic summation of a function to values of the function's continuous Fourier transform. It has a variety of useful forms that are derived from the basic one by application of the Fourier transform's scaling and time-shifting properties. The frequency-domain dual of the standard PSF is also called discrete-time Fourier transform, which leads directly to**:**

- a popular, graphical, frequency-domain representation of the phenomenon of aliasing, and
- a proof of the Nyquist-Shannon sampling theorem.

Convolution theorem

The Fourier transform translates between convolution and multiplication of functions. If *f*(*x*) and *g*(*x*) are integrable functions with Fourier transforms $\hat{f}(\xi)$ and $\hat{g}(\xi)$ respectively, then the Fourier transform of the convolution is given by the product of the Fourier transforms $\hat{f}(\xi)$ and $\hat{g}(\xi)$ (under other conventions for the definition of the Fourier transform a constant factor may appear).
This means that if**:**

$$h(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)\,dy,$$

where $*$ denotes the convolution operation, then**:**

$$\hat{h}(\xi) = \hat{f}(\xi) \cdot \hat{g}(\xi).$$

In linear time invariant (LTI) system theory, it is common to interpret *g*(*x*) as the impulse response of an LTI system with input *f*(*x*) and output *h*(*x*), since substituting the unit impulse for *f*(*x*) yields *h*(*x*) = *g*(*x*). In this case, $\hat{g}(\xi)$ represents the frequency response of the system.
Conversely, if *f*(*x*) can be decomposed as the product of two square integrable functions *p*(*x*) and *q*(*x*), then the Fourier transform of *f*(*x*) is given by the convolution of the respective Fourier transforms $\hat{p}(\xi)$ and $\hat{q}(\xi)$.

Cross-correlation theorem

In an analogous manner, it can be shown that if *h*(*x*) is the cross-correlation of *f*(*x*) and *g*(*x*):

$$h(x) = (f \star g)(x) = \int_{-\infty}^{\infty} \overline{f(y)}\, g(x+y)\,dy$$

then the Fourier transform of *h*(*x*) is:

$$\hat{h}(\xi) = \overline{\hat{f}(\xi)}\, \hat{g}(\xi).$$

As a special case, the autocorrelation of function *f*(*x*) is:

$$h(x) = (f \star f)(x) = \int_{-\infty}^{\infty} \overline{f(y)} f(x+y)\,dy$$

for which

$$\hat{h}(\xi) = \overline{\hat{f}(\xi)}\, \hat{f}(\xi) = |\hat{f}(\xi)|^2.$$

Eigenfunctions

One important choice of an orthonormal basis for $L^2(\mathbf{R})$ is given by the Hermite functions

$$\psi_n(x) = \frac{2^{1/4}}{\sqrt{n!}} e^{-\pi x^2} He_n(2x\sqrt{\pi}),$$

where $He_n(x)$ are the "probabilist's" Hermite polynomials, defined by $He_n(x) = (-1)^n \exp(x^2/2)\, \mathrm{D}^n \exp(-x^2/2)$. Under this convention for the Fourier transform, we have that

$$\hat{\psi}_n(\xi) = (-i)^n \psi_n(\xi).$$

In other words, the Hermite functions form a complete orthonormal system of eigenfunctions for the Fourier transform on $L^2(\mathbf{R})$ (Pinsky 2002). However, this choice of eigenfunctions is not unique. There are only four different eigenvalues of the Fourier transform (±1 and ±*i*) and any linear combination of eigenfunctions with the same eigenvalue gives another eigenfunction. As a consequence of this, it is possible to decompose $L^2(\mathbf{R})$ as a direct sum of four spaces H_0, H_1, H_2, and H_3 where the Fourier transform acts on He_k simply by multiplication by i^k. This approach to define the Fourier transform is due to N. Wiener (Duoandikoetxea 2001). Among other properties, Hermite functions decrease exponentially fast in both frequency and time domains and they are used to define a generalization of the Fourier transform, namely the fractional Fourier transform used in time-frequency analysis (Boashash 2003).

Fourier transform on Euclidean space

The Fourier transform can be in any arbitrary number of dimensions *n*. As with the one-dimensional case, there are many conventions. For an integrable function *f*(*x*), this article takes the definition**:**

$$\hat{f}(\xi) = \mathcal{F}(f)(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot \xi}\, dx$$

where *x* and ξ are *n*-dimensional vectors, and $x \cdot \xi$ is the dot product of the vectors. The dot product is sometimes written as $\langle x, \xi \rangle$.

All of the basic properties listed above hold for the *n*-dimensional Fourier transform, as do Plancherel's and Parseval's theorem. When the function is integrable, the Fourier transform is still uniformly continuous and the Riemann–Lebesgue lemma holds. (Stein & Weiss 1971)

Uncertainty principle

Generally speaking, the more concentrated *f*(*x*) is, the more spread out its Fourier transform $\hat{f}(\xi)$ must be. In particular, the scaling property of the Fourier transform may be seen as saying: if we "squeeze" a function in *x*, its Fourier transform "stretches out" in ξ. It is not possible to arbitrarily concentrate both a function and its Fourier transform.

The trade-off between the compaction of a function and its Fourier transform can be formalized in the form of an **uncertainty principle** by viewing a function and its Fourier transform as conjugate variables with respect to the symplectic form on the time–frequency domain: from the point of view of the linear canonical transformation, the Fourier transform is rotation by 90° in the time–frequency domain, and preserves the symplectic form.

Suppose *f*(*x*) is an integrable and square-integrable function. Without loss of generality, assume that *f*(*x*) is normalized:

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx = 1.$$

It follows from the Plancherel theorem that $\hat{f}(\xi)$ is also normalized.

The spread around *x* = 0 may be measured by the *dispersion about zero* (Pinsky 2002) defined by

$$D_0(f) = \int_{-\infty}^{\infty} x^2 |f(x)|^2 \, dx.$$

In probability terms, this is the second moment of $|f(x)|^2$ about zero.

The Uncertainty principle states that, if $f(x)$ is absolutely continuous and the functions $x \cdot f(x)$ and $f'(x)$ are square integrable, then

$$D_0(f) D_0(\hat{f}) \geq \frac{1}{16\pi^2}$$ (Pinsky 2002).

The equality is attained only in the case $f(x) = C_1 \, e^{-\pi x^2/\sigma^2}$ (hence $\hat{f}(\xi) = \sigma C_1 \, e^{-\pi \sigma^2 \xi^2}$) where $\sigma > 0$ is arbitrary and C_1 is such that f is L^2–normalized (Pinsky 2002). In other words, where f is a (normalized) Gaussian function with variance σ^2, centered at zero, and its Fourier transform is a Gaussian function with variance $1/\sigma^2$.

In fact, this inequality implies that:

$$\left(\int_{-\infty}^{\infty} (x - x_0)^2 |f(x)|^2 \, dx \right) \left(\int_{-\infty}^{\infty} (\xi - \xi_0)^2 |\hat{f}(\xi)|^2 \, d\xi \right) \geq \frac{1}{16\pi^2}$$

for any $x_0, \ \xi_0$ in $\mathbf{R}$ (Stein & Shakarchi 2003).

In quantum mechanics, the momentum and position wave functions are Fourier transform pairs, to within a factor of Planck's constant. With this constant properly taken into account, the inequality above becomes the statement of the Heisenberg uncertainty principle (Stein & Shakarchi 2003).

A stronger uncertainty principle is the Hirschman uncertainty principle which is expressed as:

$$H(|f|^2) + H(|\hat{f}|^2) \geq \log(e/2)$$

where $H(p)$ is the differential entropy of the probability density function $p(x)$:

$$H(p) = -\int_{-\infty}^{\infty} p(x) \log(p(x)) dx$$

where the logarithms may be in any base which is consistent. The equality is attained for a Gaussian, as in the previous case.

Spherical harmonics

Let the set of homogeneous harmonic polynomials of degree k on $\mathbf{R}^n$ be denoted by $\mathbf{A}_k$. The set $\mathbf{A}_k$ consists of the solid spherical harmonics of degree k. The solid spherical harmonics play a similar role in higher dimensions to the Hermite polynomials in dimension one. Specifically, if $f(x) = e^{-\pi |x|} 2P(x)$ for some $P(x)$ in $\mathbf{A}_k$, then $\hat{f}(\xi) = i^{-k} f(\xi)$. Let the set $\mathbf{H}_k$ be the closure in $L^2(\mathbf{R}^n)$ of linear combinations of functions of the form $f(|x|)P(x)$ where $P(x)$ is in $\mathbf{A}_k$. The space $L^2(\mathbf{R}^n)$ is then a direct sum of the spaces $\mathbf{H}_k$ and the Fourier transform maps each space $\mathbf{H}_k$ to itself and is possible to characterize the action of the Fourier transform on each space $\mathbf{H}_k$ (Stein & Weiss 1971). Let $f(x) = f_0(|x|)P(x)$ (with $P(x)$ in $\mathbf{A}_k$), then $\hat{f}(\xi) = F_0(|\xi|)P(\xi)$ where

$$F_0(r) = 2\pi i^{-k} r^{-(n+2k-2)/2} \int_0^{\infty} f_0(s) J_{(n+2k-2)/2}(2\pi r s) s^{(n+2k)/2} \, ds.$$

Here $J_{(n+2k-2)/2}$ denotes the Bessel function of the first kind with order $(n + 2k - 2)/2$. When $k = 0$ this gives a useful formula for the Fourier transform of a radial function (Grafakos 2004).

Restriction problems

In higher dimensions it becomes interesting to study *restriction problems* for the Fourier transform. The Fourier transform of an integrable function is continuous and the restriction of this function to any set is defined. But for a square-integrable function the Fourier transform could be a general *class* of square integrable functions. As such, the restriction of the Fourier transform of an $L^2(\mathbf{R}^n)$ function cannot be defined on sets of measure 0. It is still an active area of study to understand restriction problems in L^p for $1 < p < 2$. Surprisingly, it is possible in some cases to define the restriction of a Fourier transform to a set S, provided S has non-zero curvature. The case when S is the unit

sphere in $\mathbf{R}^n$ is of particular interest. In this case the Tomas-Stein restriction theorem states that the restriction of the Fourier transform to the unit sphere in $\mathbf{R}^n$ is a bounded operator on L^p provided $1 \le p \le (2n + 2) / (n + 3)$.

One notable difference between the Fourier transform in 1 dimension versus higher dimensions concerns the partial sum operator. Consider an increasing collection of measurable sets E_R indexed by $R \in (0,\infty)$: such as balls of radius R centered at the origin, or cubes of side $2R$. For a given integrable function f, consider the function f_R defined by:

$$f_R(x) = \int_{E_R} \hat{f}(\xi) e^{2\pi i x \cdot \xi}\, d\xi, \quad x \in \mathbb{R}^n.$$

Suppose in addition that f is in $L^p(\mathbf{R}^n)$. For $n = 1$ and $1 < p < \infty$, if one takes $E_R = (-R, R)$, then f_R converges to f in L^p as R tends to infinity, by the boundedness of the Hilbert transform. Naively one may hope the same holds true for $n > 1$. In the case that E_R is taken to be a cube with side length R, then convergence still holds. Another natural candidate is the Euclidean ball $E_R = \{\xi : |\xi| < R\}$. In order for this partial sum operator to converge, it is necessary that the multiplier for the unit ball be bounded in $L^p(\mathbf{R}^n)$. For $n \ge 2$ it is a celebrated theorem of Charles Fefferman that the multiplier for the unit ball is never bounded unless $p = 2$ (Duoandikoetxea 2001). In fact, when $p \ne 2$, this shows that not only may f_R fail to converge to f in L^p, but for some functions $f \in L^p(\mathbf{R}^n)$, f_R is not even an element of L^p.

Fourier transform on other function spaces

The definition of the Fourier transform by the integral formula

$$\hat{f}(\xi) = \int_{\mathbf{R}} f(x) e^{-2\pi i \xi \cdot x}\, dx$$

is valid for Lebesgue integrable functions f; that is, f in $L^1(\mathbf{R})$. The image of L^1 a subset of the space $C_0(\mathbf{R})$ of continuous functions that tend to zero at infinity (the Riemann–Lebesgue lemma), although it is not the entire space. Indeed, there is no simple characterization of the image.

It is possible to extend the definition of the Fourier transform to other spaces of functions. Since compactly supported smooth functions are integrable and dense in $L^2(\mathbf{R})$, the Plancherel theorem allows us to extend the definition of the Fourier transform to general functions in $L^2(\mathbf{R})$ by continuity arguments. Further $\mathcal{F}: L^2(\mathbf{R}) \to L^2(\mathbf{R})$ is a unitary operator (Stein & Weiss 1971, Thm. 2.3). In particular, the image of $L^2(\mathbf{R})$ is itself under the Fourier transform. The Fourier transform in $L^2(\mathbf{R})$ is no longer given by an ordinary Lebesgue integral, although it can be computed by an improper integral, here meaning that for an L^2 function f,

$$\hat{f}(\xi) = \lim_{R \to \infty} \int_{-R}^{R} f(x) e^{-2\pi i x \xi}\, dx$$

where the limit is taken in the L^2 sense. Many of the properties of the Fourier transform in L^1 carry over to L^2, by a suitable limiting argument.

The definition of the Fourier transform can be extended to functions in $L^p(\mathbf{R})$ for $1 \le p \le 2$ by decomposing such functions into a fat tail part in L^2 plus a fat body part in L^1. In each of these spaces, the Fourier transform of a function in $L^p(\mathbf{R})$ is in $L^q(\mathbf{R})$, where $q = p/(p-1)$ is the Hölder conjugate of p. by the Hausdorff–Young inequality. However, except for $p = 2$, the image is not easily characterized. Further extensions become more technical. The Fourier transform of functions in L^p for the range $2 < p < \infty$ requires the study of distributions (Katznelson 1976). In fact, it can be shown that there are functions in L^p with $p>2$ so that the Fourier transform is not defined as a function (Stein & Weiss 1971).

Tempered distributions

The Fourier transform maps the space of Schwartz functions to itself, and gives a homeomorphism of the space to itself (Stein & Weiss 1971). Because of this it is possible to define the Fourier transform of tempered distributions. These include all the integrable functions mentioned above, as well as well-behaved functions of polynomial growth and distributions of compact support, and have the added advantage that the Fourier transform of any tempered distribution is again a tempered distribution.

The following two facts provide some motivation for the definition of the Fourier transform of a distribution. First let f and g be integrable functions, and let $\hat{f}$ and $\hat{g}$ be their Fourier transforms respectively. Then the Fourier transform obeys the following multiplication formula (Stein & Weiss 1971),

$$\int_{\mathbb{R}^n} \hat{f}(x) g(x)\,dx = \int_{\mathbb{R}^n} f(x) \hat{g}(x)\,dx.$$

Secondly, every integrable function f defines (induces) a distribution T_f by the relation

$$T_f(\varphi) = \int_{\mathbb{R}^n} f(x)\varphi(x)\,dx \quad \text{for all Schwartz functions } \varphi.$$

In fact, given a distribution T, we define the Fourier transform by the relation

$$\hat{T}(\varphi) = T(\hat{\varphi}) \quad \text{for all Schwartz functions } \varphi.$$

It follows that

$$\hat{T}_f = T_{\hat{f}}.$$

Distributions can be differentiated and the above mentioned compatibility of the Fourier transform with differentiation and convolution remains true for tempered distributions.

Generalizations

Fourier–Stieltjes transform

The Fourier transform of a finite Borel measure μ on $\mathbf{R}^n$ is given by (Pinsky 2002):

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^n} \mathrm{e}^{-2\pi i x\cdot\xi}\,d\mu.$$

This transform continues to enjoy many of the properties of the Fourier transform of integrable functions. One notable difference is that the Riemann–Lebesgue lemma fails for measures (Katznelson 1976). In the case that $d\mu = f(x)\,dx$, then the formula above reduces to the usual definition for the Fourier transform of f. In the case that μ is the probability distribution associated to a random variable X, the Fourier-Stieltjes transform is closely related to the characteristic function, but the typical conventions in probability theory take $e^{ix\cdot\xi}$ instead of $e^{-2\pi i x\cdot\xi}$ (Pinsky 2002). In the case when the distribution has a probability density function this definition reduces to the Fourier transform applied to the probability density function, again with a different choice of constants.

The Fourier transform may be used to give a characterization of continuous measures. Bochner's theorem characterizes which functions may arise as the Fourier–Stieltjes transform of a measure (Katznelson 1976).

Furthermore, the Dirac delta function is not a function but it is a finite Borel measure. Its Fourier transform is a constant function (whose specific value depends upon the form of the Fourier transform used).

Locally compact abelian groups

The Fourier transform may be generalized to any locally compact abelian group. A locally compact abelian group is an abelian group which is at the same time a locally compact Hausdorff topological space so that the group operations are continuous. If G is a locally compact abelian group, it has a translation invariant measure μ, called Haar measure. For a locally compact abelian group G it is possible to place a topology on the set of characters $\hat{G}$ so that $\hat{G}$ is also a locally compact abelian group. For a function f in $L^1(G)$ it is possible to define the Fourier transform by (Katznelson 1976):

$$\hat{f}(\xi) = \int_G \xi(x) f(x)\, d\mu \qquad \text{for any } \xi \in \hat{G}.$$

Locally compact Hausdorff space

The Fourier transform may be generalized to any locally compact Hausdorff space, which recovers the topology but loses the group structure.

Given a locally compact Hausdorff topological space X, the space $A=C_0(X)$ of continuous complex-valued functions on X which vanish at infinity is in a natural way a commutative C*-algebra, via pointwise addition, multiplication, complex conjugation, and with norm as the uniform norm. Conversely, the characters of this algebra A, denoted Φ_A, are naturally a topological space, and can be identified with evaluation at a point of x, and one has an isometric isomorphism $C_0(X) \to C_0(\Phi_A)$. In the case where X=**R** is the real line, this is exactly the Fourier transform.

Non-abelian groups

The Fourier transform can also be defined for functions on a non-abelian group, provided that the group is compact. Unlike the Fourier transform on an abelian group, which is scalar-valued, the Fourier transform on a non-abelian group is operator-valued (Hewitt & Ross 1971, Chapter 8). The Fourier transform on compact groups is a major tool in representation theory (Knapp 2001) and non-commutative harmonic analysis.

Let G be a compact Hausdorff topological group. Let Σ denote the collection of all isomorphism classes of finite-dimensional irreducible unitary representations, along with a definite choice of representation $U^{(\sigma)}$ on the Hilbert space H_σ of finite dimension d_σ for each σ ∈ Σ. If μ is a finite Borel measure on G, then the Fourier–Stieltjes transform of μ is the operator on H_σ defined by

$$\langle \hat{\mu}\xi, \eta\rangle_{H_\sigma} = \int_G \langle \overline{U}_g^{(\sigma)}\xi, \eta\rangle \, d\mu(g)$$

where $\overline{U}^{(\sigma)}$ is the complex-conjugate representation of $U^{(\sigma)}$ acting on H_σ. As in the abelian case, if μ is absolutely continuous with respect to the left-invariant probability measure λ on G, then it is represented as

$$d\mu = f d\lambda$$

for some $f \in L^1(\lambda)$. In this case, one identifies the Fourier transform of f with the Fourier–Stieltjes transform of μ.

The mapping $\mu \mapsto \hat{\mu}$ defines an isomorphism between the Banach space $M(G)$ of finite Borel measures (see rca space) and a closed subspace of the Banach space $\mathbf{C}_\infty(\Sigma)$ consisting of all sequences $E = (E_\sigma)$ indexed by Σ of (bounded) linear operators $E_\sigma : H_\sigma \to H_\sigma$ for which the norm

$$\|E\| = \sup_{\sigma\in\Sigma} \|E_\sigma\|$$

is finite. The "convolution theorem" asserts that, furthermore, this isomorphism of Banach spaces is in fact an isomorphism of C^* algebras into a subspace of $\mathbf{C}_\infty(\Sigma)$, in which $M(G)$ is equipped with the product given by convolution of measures and $\mathbf{C}_\infty(\Sigma)$ the product given by multiplication of operators in each index σ.

The Peter-Weyl theorem holds, and a version of the Fourier inversion formula (Plancherel's theorem) follows: if $f \in L^2(G)$, then

$$f(g) = \sum_{\sigma \in \Sigma} d_\sigma \mathrm{tr}(\hat{f}(\sigma) U_g^{(\sigma)})$$

where the summation is understood as convergent in the L^2 sense.

The generalization of the Fourier transform to the noncommutative situation has also in part contributed to the development of noncommutative geometry. In this context, a categorical generalization of the Fourier transform to noncommutative groups is Tannaka-Krein duality, which replaces the group of characters with the category of representations. However, this loses the connection with harmonic functions.

Alternatives

In signal processing terms, a function (of time) is a representation of a signal with perfect *time resolution,* but no frequency information, while the Fourier transform has perfect *frequency resolution,* but no time information: the magnitude of the Fourier transform at a point is how much frequency content there is, but location is only given by phase (argument of the Fourier transform at a point), and standing waves are not localized in time – a sine wave continues out to infinity, without decaying. This limits the usefulness of the Fourier transform for analyzing signals that are localized in time, notably transients, or any signal of finite extent.

As alternatives to the Fourier transform, in time-frequency analysis, one uses time-frequency transforms or time-frequency distributions to represent signals in a form that has some time information and some frequency information – by the uncertainty principle, there is a trade-off between these. These can be generalizations of the Fourier transform, such as the short-time Fourier transform or fractional Fourier transform, or can use different functions to represent signals, as in wavelet transforms and chirplet transforms, with the wavelet analog of the (continuous) Fourier transform being the continuous wavelet transform. (Boashash 2003).

Applications

Analysis of differential equations

Fourier transforms and the closely related Laplace transforms are widely used in solving differential equations. The Fourier transform is compatible with differentiation in the following sense: if $f(x)$ is a differentiable function with Fourier transform $\hat{f}(\xi)$, then the Fourier transform of its derivative is given by $2\pi i\xi\hat{f}(\xi)$. This can be used to transform differential equations into algebraic equations. Note that this technique only applies to problems whose domain is the whole set of real numbers. By extending the Fourier transform to functions of several variables partial differential equations with domain $\mathbf{R}^n$ can also be translated into algebraic equations.

Fourier transform spectroscopy

The Fourier transform is also used in nuclear magnetic resonance (NMR) and in other kinds of spectroscopy, e.g. infrared (FTIR). In NMR an exponentially-shaped free induction decay (FID) signal is acquired in the time domain and Fourier-transformed to a Lorentzian line-shape in the frequency domain. The Fourier transform is also used in magnetic resonance imaging (MRI) and mass spectrometry.

Other notations

Other common notations for $\hat{f}(\xi)$ include:

$$\tilde{f}(\xi),\ \tilde{f}(\omega),\ F(\xi),\ \mathcal{F}(f)(\xi),\ (\mathcal{F}f)(\xi),\ \mathcal{F}(f),\ \mathcal{F}(\omega),\ F(\omega),\ \mathcal{F}(j\omega),\ \mathcal{F}\{f\},\ \mathcal{F}(f(t)),\ \mathcal{F}\{f(t)\}.$$

Denoting the Fourier transform by a capital letter corresponding to the letter of function being transformed (such as $f(x)$ and $F(\xi)$) is especially common in the sciences and engineering. In electronics, the omega (ω) is often used instead of ξ due to its interpretation as angular frequency, sometimes it is written as $F(j\omega)$, where j is the imaginary

unit, to indicate its relationship with the Laplace transform, and sometimes it is written informally as $F(2\pi f)$ in order to use ordinary frequency.

The interpretation of the complex function $\hat{f}(\xi)$ may be aided by expressing it in polar coordinate form

$$\hat{f}(\xi) = A(\xi)e^{i\varphi(\xi)}$$

in terms of the two real functions $A(\xi)$ and $\varphi(\xi)$ where:

$$A(\xi) = |\hat{f}(\xi)|,$$

is the amplitude and

$$\varphi(\xi) = \arg\left(\hat{f}(\xi)\right),$$

is the phase (see arg function).

Then the inverse transform can be written:

$$f(x) = \int_{-\infty}^{\infty} A(\xi)\ e^{i(2\pi\xi x + \varphi(\xi))}\, d\xi,$$

which is a recombination of all the **frequency components** of $f(x)$. Each component is a complex sinusoid of the form $e^{2\pi i x\xi}$ whose amplitude is $A(\xi)$ and whose initial phase angle (at $x = 0$) is $\varphi(\xi)$.

The Fourier transform may be thought of as a mapping on function spaces. This mapping is here denoted $\mathcal{F}$ and $\mathcal{F}(f)$ is used to denote the Fourier transform of the function f. This mapping is linear, which means that $\mathcal{F}$ can also be seen as a linear transformation on the function space and implies that the standard notation in linear algebra of applying a linear transformation to a vector (here the function f) can be used to write $\mathcal{F}f$ instead of $\mathcal{F}(f)$. Since the result of applying the Fourier transform is again a function, we can be interested in the value of this function evaluated at the value ξ for its variable, and this is denoted either as $\mathcal{F}(f)(\xi)$ or as $(\mathcal{F}f)(\xi)$. Notice that in the former case, it is implicitly understood that $\mathcal{F}$ is applied first to f and then the resulting function is evaluated at ξ, not the other way around.

In mathematics and various applied sciences it is often necessary to distinguish between a function f and the value of f when its variable equals x, denoted $f(x)$. This means that a notation like $\mathcal{F}(f(x))$ formally can be interpreted as the Fourier transform of the values of f at x. Despite this flaw, the previous notation appears frequently, often when a particular function or a function of a particular variable is to be transformed.

For example, $\mathcal{F}(\mathrm{rect}(x)) = \mathrm{sinc}(\xi)$ is sometimes used to express that the Fourier transform of a rectangular function is a sinc function,

or $\mathcal{F}(f(x + x_0)) = \mathcal{F}(f(x))e^{2\pi i\xi x_0}$ is used to express the shift property of the Fourier transform.

Notice, that the last example is only correct under the assumption that the transformed function is a function of x, not of x_0.

Other conventions

The Fourier transform can also be written in terms of angular frequency: $\omega = 2\pi\xi$ whose units are radians per second.

The substitution $\xi = \omega/(2\pi)$ into the formulas above produces this convention:

$$\hat{f}(\omega) = \int_{\mathbb{R}^n} f(x)e^{-i\omega\cdot x}\, dx.$$

Under this convention, the inverse transform becomes:

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\omega)e^{i\omega\cdot x}\, d\omega.$$

Unlike the convention followed in this article, when the Fourier transform is defined this way, it is no longer a unitary transformation on $L^2(\mathbf{R}^n)$. There is also less symmetry between the formulas for the Fourier transform and its

inverse.

Another convention is to split the factor of $(2\pi)^n$ evenly between the Fourier transform and its inverse, which leads to definitions:

$$\hat{f}(\omega) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\omega \cdot x}\, dx$$
$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}(\omega) e^{i\omega \cdot x}\, d\omega.$$

Under this convention, the Fourier transform is again a unitary transformation on $L^2(\mathbf{R}^n)$. It also restores the symmetry between the Fourier transform and its inverse.

Variations of all three conventions can be created by conjugating the complex-exponential kernel of both the forward and the reverse transform. The signs must be opposites. Other than that, the choice is (again) a matter of convention.

Summary of popular forms of the Fourier transform

ordinary frequency ξ (hertz)	**unitary**	$\hat{f}_1(\xi) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot \xi}\, dx = \hat{f}_2(2\pi\xi) = (2\pi)^{n/2} \hat{f}_3(2\pi\xi)$ $f(x) = \int_{\mathbb{R}^n} \hat{f}_1(\xi) e^{2\pi i x \cdot \xi}\, d\xi$
angular frequency ω (rad/s)	**non-unitary**	$\hat{f}_2(\omega) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} f(x) e^{-i\omega \cdot x}\, dx = \hat{f}_1\left(\frac{\omega}{2\pi}\right) = (2\pi)^{n/2}\, \hat{f}_3(\omega)$ $f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}_2(\omega) e^{i\omega \cdot x}\, d\omega$
	unitary	$\hat{f}_3(\omega) \stackrel{\text{def}}{=} \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x)\, e^{-i\omega \cdot x}\, dx = \frac{1}{(2\pi)^{n/2}} \hat{f}_1\left(\frac{\omega}{2\pi}\right) = \frac{1}{(2\pi)^{n/2}} \hat{f}_2(\omega)$ $f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}_3(\omega) e^{i\omega \cdot x}\, d\omega$

As discussed above, the characteristic function of a random variable is the same as the Fourier–Stieltjes transform of its distribution measure, but in this context it is typical to take a different convention for the constants. Typically characteristic function is defined $E(e^{it \cdot X}) = \int e^{it \cdot x} d\mu_X(x)$.

As in the case of the "non-unitary angular frequency" convention above, there is no factor of 2π appearing in either of the integral, or in the exponential. Unlike any of the conventions appearing above, this convention takes the opposite sign in the exponential.

Tables of important Fourier transforms

The following tables record some closed form Fourier transforms. For functions $f(x)$, $g(x)$ and $h(x)$ denote their Fourier transforms by $\hat{f}$, $\hat{g}$, and $\hat{h}$ respectively. Only the three most common conventions are included. It may be useful to notice that entry 105 gives a relationship between the Fourier transform of a function and the original function, which can be seen as relating the Fourier transform and its inverse.

Functional relationships

The Fourier transforms in this table may be found in (Erdélyi 1954) or the appendix of (Kammler 2000).

	Function	Fourier transform unitary, ordinary frequency	Fourier transform unitary, angular frequency	Fourier transform non-unitary, angular frequency	Remarks
	$f(x)$	$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi}\,dx$	$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x}\,dx$	$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(x) e^{-i\nu x}\,dx$	Definition
101	$a \cdot f(x) + b \cdot g(x)$	$a \cdot \hat{f}(\xi) + b \cdot \hat{g}(\xi)$	$a \cdot \hat{f}(\omega) + b \cdot \hat{g}(\omega)$	$a \cdot \hat{f}(\nu) + b \cdot \hat{g}(\nu)$	Linearity
102	$f(x-a)$	$e^{-2\pi i a \xi} \hat{f}(\xi)$	$e^{-ia\omega} \hat{f}(\omega)$	$e^{-ia\nu} \hat{f}(\nu)$	Shift in time domain
103	$e^{2\pi i a x} f(x)$	$\hat{f}\,(\xi - a)$	$\hat{f}(\omega - 2\pi a)$	$\hat{f}(\nu - 2\pi a)$	Shift in frequency domain, dual of 102
104	$f(ax)$	$\frac{1}{\lvert a\rvert} \hat{f}\left(\frac{\xi}{a}\right)$	$\frac{1}{\lvert a\rvert} \hat{f}\left(\frac{\omega}{a}\right)$	$\frac{1}{\lvert a\rvert} \hat{f}\left(\frac{\nu}{a}\right)$	Scaling in the time domain. If $\lvert a\rvert$ is large, then $f(ax)$ is concentrated around 0 and $\frac{1}{\lvert a\rvert} \hat{f}\left(\frac{\omega}{a}\right)$ spreads out and flattens.
105	$\hat{f}(x)$	$f(-\xi)$	$f(-\omega)$	$2\pi f(-\nu)$	Duality. Here $\hat{f}$ needs to be calculated using the same method as Fourier transform column. Results from swapping "dummy" variables of x and ξ or ω or ν.
106	$\frac{d^n f(x)}{dx^n}$	$(2\pi i \xi)^n \hat{f}(\xi)$	$(i\omega)^n \hat{f}(\omega)$	$(i\nu)^n \hat{f}(\nu)$	
107	$x^n f(x)$	$\left(\frac{i}{2\pi}\right)^n \frac{d^n \hat{f}(\xi)}{d\xi^n}$	$i^n \frac{d^n \hat{f}(\omega)}{d\omega^n}$	$i^n \frac{d^n \hat{f}(\nu)}{d\nu^n}$	This is the dual of 106
108	$(f * g)(x)$	$\hat{f}(\xi)\hat{g}(\xi)$	$\sqrt{2\pi}\hat{f}(\omega)\hat{g}(\omega)$	$\hat{f}(\nu)\hat{g}(\nu)$	The notation $f * g$ denotes the convolution of f and g — this rule is the convolution theorem
109	$f(x)g(x)$	$(\hat{f} * \hat{g})(\xi)$	$\frac{(\hat{f} * \hat{g})(\omega)}{\sqrt{2\pi}}$	$\frac{1}{2\pi}(\hat{f} * \hat{g})(\nu)$	This is the dual of 108
110	For $f(x)$ a purely real	$\hat{f}(-\xi) = \overline{\hat{f}(\xi)}$	$\hat{f}(-\omega) = \overline{\hat{f}(\omega)}$	$\hat{f}(-\nu) = \overline{\hat{f}(\nu)}$	Hermitian symmetry. $\overline{z}$ indicates the complex conjugate.
111	For $f(x)$ a purely real even function	$\hat{f}(\omega)$, $\hat{f}(\xi)$ and $\hat{f}(\nu)$ are purely real even functions.			
112	For $f(x)$ a purely real odd function	$\hat{f}(\omega)$, $\hat{f}(\xi)$ and $\hat{f}(\nu)$ are purely imaginary odd functions.			

Square-integrable functions

The Fourier transforms in this table may be found in (Campbell & Foster 1948), (Erdélyi 1954), or the appendix of (Kammler 2000).

	Function	Fourier transform unitary, ordinary frequency	Fourier transform unitary, angular frequency	Fourier transform non-unitary, angular frequency	Remarks
	$f(x)$	$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi}\,dx$	$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x}\,dx$	$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(x) e^{-i\nu x}\,dx$	
201	$\operatorname{rect}(ax)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{sinc}\left(\frac{\xi}{a}\right)$	$\frac{1}{\sqrt{2\pi a^2}} \cdot \operatorname{sinc}\left(\frac{\omega}{2\pi a}\right)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{sinc}\left(\frac{\nu}{2\pi a}\right)$	The rectangular pulse and the *normalized* sinc function, here defined as sinc(*x*) = sin(*πx*)/(*πx*)
202	$\operatorname{sinc}(ax)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{rect}\left(\frac{\xi}{a}\right)$	$\frac{1}{\sqrt{2\pi a^2}} \cdot \operatorname{rect}\left(\frac{\omega}{2\pi a}\right)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{rect}\left(\frac{\nu}{2\pi a}\right)$	Dual of rule 201. The rectangular function is an ideal low-pass filter, and the sinc function is the non-causal impulse response of such a filter.
203	$\operatorname{sinc}^2(ax)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{tri}\left(\frac{\xi}{a}\right)$	$\frac{1}{\sqrt{2\pi a^2}} \cdot \operatorname{tri}\left(\frac{\omega}{2\pi a}\right)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{tri}\left(\frac{\nu}{2\pi a}\right)$	The function tri(*x*) is the triangular function
204	$\operatorname{tri}(ax)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{sinc}^2\left(\frac{\xi}{a}\right)$	$\frac{1}{\sqrt{2\pi a^2}} \cdot \operatorname{sinc}^2\left(\frac{\omega}{2\pi a}\right)$	$\frac{1}{\lvert a\rvert} \cdot \operatorname{sinc}^2\left(\frac{\nu}{2\pi a}\right)$	Dual of rule 203.
205	$e^{-ax}u(x)$	$\frac{1}{a + 2\pi i \xi}$	$\frac{1}{\sqrt{2\pi}(a + i\omega)}$	$\frac{1}{a + i\nu}$	The function *u*(*x*) is the Heaviside unit step function and *a*>0.
206	$e^{-\alpha x^2}$	$\sqrt{\frac{\pi}{\alpha}} \cdot e^{-\frac{(\pi\xi)^2}{\alpha}}$	$\frac{1}{\sqrt{2\alpha}} \cdot e^{-\frac{\omega^2}{4\alpha}}$	$\sqrt{\frac{\pi}{\alpha}} \cdot e^{-\frac{\nu^2}{4\alpha}}$	This shows that, for the unitary Fourier transforms, the Gaussian function $\exp(-\alpha x^2)$ is its own Fourier transform for some choice of α. For this to be integrable we must have Re(α)>0.
207	$\mathrm{e}^{-a\lvert x\rvert}$	$\frac{2a}{a^2 + 4\pi^2\xi^2}$	$\sqrt{\frac{2}{\pi}} \cdot \frac{a}{a^2 + \omega^2}$	$\frac{2a}{a^2 + \nu^2}$	For *a*>*0*. That is, the Fourier transform of a decaying exponential function is a Lorentzian function.
208	$\operatorname{sech}(ax)$	$\frac{\pi}{a}\operatorname{sech}\left(\frac{\pi^2}{a}\xi\right)$	$\frac{1}{a}\sqrt{\frac{\pi}{2}}\operatorname{sech}\left(\frac{\pi}{2a}\omega\right)$	$\frac{\pi}{a}\operatorname{sech}\left(\frac{\pi}{2a}\nu\right)$	Hyperbolic secant is its own Fourier transform
209	$e^{-\frac{a^2x^2}{2}} H_n(ax)$	$\frac{\sqrt{2\pi}(-i)^n}{a} \cdot e^{-\frac{2\pi^2\xi^2}{a^2}} H_n\left(\frac{2\pi\xi}{a}\right)$	$\frac{(-i)^n}{a} \cdot e^{-\frac{\omega^2}{2a^2}} H_n\left(\frac{\omega}{a}\right)$	$\frac{(-i)^n\sqrt{2\pi}}{a} \cdot e^{-\frac{\nu^2}{2a^2}} H_n\left(\frac{\nu}{a}\right)$	H_n is the Hermite's polynomial. If $a = 1$ then the Gauss-Hermite functions are eigenfunctions of the Fourier transform operator. For a derivation, see Hermite polynomial. The formula reduces to 206 for $n = 0$.

Distributions

The Fourier transforms in this table may be found in (Erdélyi 1954) or the appendix of (Kammler 2000).

	Function	Fourier transform unitary, ordinary frequency	Fourier transform unitary, angular frequency	Fourier transform non-unitary, angular frequency	Remarks
	$f(x)$	$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi}\,dx$	$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x}\,dx$	$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(x) e^{-i\nu x}\,dx$	
301	1	$\delta(\xi)$	$\sqrt{2\pi} \cdot \delta(\omega)$	$2\pi\delta(\nu)$	The distribution $\delta(\xi)$ denotes the Dirac delta function.
302	$\delta(x)$	1	$\frac{1}{\sqrt{2\pi}}$	1	Dual of rule 301.
303	e^{iax}	$\delta\left(\xi - \frac{a}{2\pi}\right)$	$\sqrt{2\pi} \cdot \delta(\omega - a)$	$2\pi\delta(\nu - a)$	This follows from 103 and 301.
304	$\cos(ax)$	$\frac{\delta\left(\xi - \frac{a}{2\pi}\right) + \delta\left(\xi + \frac{a}{2\pi}\right)}{2}$	$\sqrt{2\pi} \cdot \frac{\delta(\omega - a) + \delta(\omega + a)}{2}$	$\pi\left(\delta(\nu - a) + \delta(\nu + a)\right)$	This follows from rules 101 and 303 using Euler's formula: $\cos(ax) = (e^{iax} + e^{-iax})/2$.
305	$\sin(ax)$	$\frac{\delta\left(\xi - \frac{a}{2\pi}\right) - \delta\left(\xi + \frac{a}{2\pi}\right)}{2i}$	$\sqrt{2\pi} \cdot \frac{\delta(\omega - a) - \delta(\omega + a)}{2i}$	$-i\pi\left(\delta(\nu - a) - \delta(\nu + a)\right)$	This follows from 101 and 303 using $\sin(ax) = (e^{iax} - e^{-iax})/(2i)$.
306	$\cos(ax^2)$	$\sqrt{\frac{\pi}{a}} \cos\left(\frac{\pi^2\xi^2}{a} - \frac{\pi}{4}\right)$	$\frac{1}{\sqrt{2a}} \cos\left(\frac{\omega^2}{4a} - \frac{\pi}{4}\right)$	$\sqrt{\frac{\pi}{a}} \cos\left(\frac{\nu^2}{4a} - \frac{\pi}{4}\right)$	
307	$\sin(ax^2)$	$-\sqrt{\frac{\pi}{a}} \sin\left(\frac{\pi^2\xi^2}{a} - \frac{\pi}{4}\right)$	$\frac{-1}{\sqrt{2a}} \sin\left(\frac{\omega^2}{4a} - \frac{\pi}{4}\right)$	$-\sqrt{\frac{\pi}{a}} \sin\left(\frac{\nu^2}{4a} - \frac{\pi}{4}\right)$	
308	x^n	$\left(\frac{i}{2\pi}\right)^n \delta^{(n)}(\xi)$	$i^n\sqrt{2\pi}\delta^{(n)}(\omega)$	$2\pi i^n \delta^{(n)}(\nu)$	Here, n is a natural number and $\delta^{(n)}(\xi)$is the n-th distribution derivative of the Dirac delta function. This rule follows from rules 107 and 301. Combining this rule with 101, we can transform all polynomials.
309	$\frac{1}{x}$	$-i\pi\,\operatorname{sgn}(\xi)$	$-i\sqrt{\frac{\pi}{2}}\operatorname{sgn}(\omega)$	$-i\pi\,\operatorname{sgn}(\nu)$	Here sgn(ξ) is the sign function. Note that $1/x$ is not a distribution. It is necessary to use the Cauchy principal value when testing against Schwartz functions. This rule is useful in studying the Hilbert transform.
310	$\frac{1}{x^n} := \frac{(-1)^{n-1}}{(n-1)!}\frac{d^n}{dx^n}\log\lvert x\rvert$	$-i\pi\frac{(-2\pi i\xi)^{n-1}}{(n-1)!}\operatorname{sgn}(\xi)$	$-i\sqrt{\frac{\pi}{2}}\cdot\frac{(-i\omega)^{n-1}}{(n-1)!}\operatorname{sgn}(\omega)$	$-i\pi\frac{(-i\nu)^{n-1}}{(n-1)!}\operatorname{sgn}(\nu)$	$1/x^n$ is the homogeneous distribution defined by the distributional derivative $\frac{(-1)^{n-1}}{(n-1)!}\frac{d^n}{dx^n}\log\lvert x\rvert$

311	$\|x\|^\alpha$	$-2\frac{\sin(\pi\alpha/2)\Gamma(\alpha+1)}{\|2\pi\xi\|^{\alpha+1}}$	$\frac{-2}{\sqrt{2\pi}}\frac{\sin(\pi\alpha/2)\Gamma(\alpha+1)}{\|\omega\|^{\alpha+1}}$	$-2\frac{\sin(\pi\alpha/2)\Gamma(\alpha+1)}{\|\nu\|^{\alpha+1}}$	This formula is valid for $0 > \alpha > -1$. For $\alpha > 0$ some singular terms arise at the origin that can be found by differentiating 318. If Re $\alpha > -1$, then $\|x\|^\alpha$ is a locally integrable function, and so a tempered distribution. The function $\alpha \mapsto \|x\|^\alpha$ is a holomorphic function from the right half-plane to the space of tempered distributions. It admits a unique meromorphic extension to a tempered distribution, also denoted $\|x\|^\alpha$ for $\alpha \neq -2, -4, \ldots$ (See homogeneous distribution.)
312	$\operatorname{sgn}(x)$	$\frac{1}{i\pi\xi}$	$\sqrt{\frac{2}{\pi}}\frac{1}{i\omega}$	$\frac{2}{i\nu}$	The dual of rule 309. This time the Fourier transforms need to be considered as Cauchy principal value.
313	$u(x)$	$\frac{1}{2}\left(\frac{1}{i\pi\xi}+\delta(\xi)\right)$	$\sqrt{\frac{\pi}{2}}\left(\frac{1}{i\pi\omega}+\delta(\omega)\right)$	$\pi\left(\frac{1}{i\pi\nu}+\delta(\nu)\right)$	The function $u(x)$ is the Heaviside unit step function; this follows from rules 101, 301, and 312.
314	$\sum_{n=-\infty}^{\infty}\delta(x-nT)$	$\frac{1}{T}\sum_{k=-\infty}^{\infty}\delta\left(\xi-\frac{k}{T}\right)$	$\frac{\sqrt{2\pi}}{T}\sum_{k=-\infty}^{\infty}\delta\left(\omega-\frac{2\pi k}{T}\right)$	$\frac{2\pi}{T}\sum_{k=-\infty}^{\infty}\delta\left(\nu-\frac{2\pi k}{T}\right)$	This function is known as the Dirac comb function. This result can be derived from 302 and 102, together with the fact that $\sum_{n=-\infty}^{\infty} e^{inx} = 2\pi\sum_{k=-\infty}^{\infty}\delta(x+2\pi k)$ as distributions.
315	$J_0(x)$	$\frac{2\operatorname{rect}(\pi\xi)}{\sqrt{1-4\pi^2\xi^2}}$	$\sqrt{\frac{2}{\pi}}\cdot\frac{\operatorname{rect}\left(\frac{\omega}{2}\right)}{\sqrt{1-\omega^2}}$	$\frac{2\operatorname{rect}\left(\frac{\nu}{2}\right)}{\sqrt{1-\nu^2}}$	The function $J_0(x)$ is the zeroth order Bessel function of first kind.
316	$J_n(x)$	$\frac{2(-i)^n T_n(2\pi\xi)\operatorname{rect}(\pi\xi)}{\sqrt{1-4\pi^2\xi^2}}$	$\sqrt{\frac{2}{\pi}}\frac{(-i)^n T_n(\omega)\operatorname{rect}\left(\frac{\omega}{2}\right)}{\sqrt{1-\omega^2}}$	$\frac{2(-i)^n T_n(\nu)\operatorname{rect}\left(\frac{\nu}{2}\right)}{\sqrt{1-\nu^2}}$	This is a generalization of 315. The function $J_n(x)$ is the n-th order Bessel function of first kind. The function $T_n(x)$ is the Chebyshev polynomial of the first kind.
317	$\log\|x\|$	$-\frac{1}{2}\frac{1}{\|\xi\|}-\gamma\delta(\xi)$	$-\frac{\sqrt{\pi/2}}{\|\omega\|}-\sqrt{2\pi}\gamma\delta(\omega)$	$-\frac{\pi}{\|\nu\|}-2\pi\gamma\delta(\nu)$	γ is the Euler–Mascheroni constant.

318	$(\mp ix)^{-\alpha}$	$\frac{(2\pi)^{\alpha}}{\Gamma(\alpha)} u(\pm\xi)(\pm\xi)^{\alpha-1}$	$\frac{\sqrt{2\pi}}{\Gamma(\alpha)} u(\pm\omega)(\pm\omega)^{\alpha-1}$	$\frac{2\pi}{\Gamma(\alpha)} u(\pm\nu)(\pm\nu)^{\alpha-1}$	This formula is valid for $1 > \alpha > 0$. Use differentiation to derive formula for higher exponents. u is the Heaviside function.

Two-dimensional functions

	Function	Fourier transform unitary, ordinary frequency	Fourier transform unitary, angular frequency	Fourier transform non-unitary, angular frequency
400	$f(x, y)$	$\hat{f}(\xi_x, \xi_y) = \iint f(x, y) e^{-2\pi i(\xi_x x + \xi_y y)}\, dx dy$	$\hat{f}(\omega_x, \omega_y) = \frac{1}{2\pi} \iint f(x, y) e^{-i(\omega_x x + \omega_y y)}\, dx dy$	$\hat{f}(\nu_x, \nu_y) = \iint f(x, y) e^{-i(\nu_x x + \nu_y y)}\, dx dy$
401	$e^{-\pi\left(a^2x^2 + b^2y^2\right)}$	$\frac{1}{\lvert ab\rvert} e^{-\pi\left(\xi_x^2/a^2 + \xi_y^2/b^2\right)}$	$\frac{1}{2\pi \cdot \lvert ab\rvert} e^{\frac{-\left(\omega_x^2/a^2 + \omega_y^2/b^2\right)}{4\pi}}$	$\frac{1}{\lvert ab\rvert} e^{\frac{-\left(\nu_x^2/a^2 + \nu_y^2/b^2\right)}{4\pi}}$
402	$\operatorname{circ}\left(\sqrt{x^2 + y^2}\right)$	$\frac{J_1\left(2\pi\sqrt{\xi_x^2 + \xi_y^2}\right)}{\sqrt{\xi_x^2 + \xi_y^2}}$	$\frac{J_1\left(\sqrt{\omega_x^2 + \omega_y^2}\right)}{\sqrt{\omega_x^2 + \omega_y^2}}$	$\frac{2\pi J_1\left(\sqrt{\nu_x^2 + \nu_y^2}\right)}{\sqrt{\nu_x^2 + \nu_y^2}}$

Remarks

To 400: The variables ξ_x, ξ_y, ω_x, ω_y, ν_x and ν_y are real numbers. The integrals are taken over the entire plane.

To 401: Both functions are Gaussians, which may not have unit volume.

To 402: The function is defined by $\operatorname{circ}(r)=1$ $0 \le r \le 1$, and is 0 otherwise. This is the Airy distribution, and is expressed using J_1 (the order 1 Bessel function of the first kind). (Stein & Weiss 1971, Thm. IV.3.3)

Formulas for general *n*-dimensional functions

	Function	Fourier transform unitary, ordinary frequency	Fourier transform unitary, angular frequency	Fourier transform non-unitary, angular frequency
500	$f(x)$	$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot \xi}\, d^n x$	$\hat{f}(\omega) = \frac{1}{(2\pi)^{(n/2)}} \int_{\mathbb{R}^n} f(x) e^{-i\omega \cdot x}\, d^n x$	$\hat{f}(\nu) = \int_{\mathbb{R}^n} f(x) e^{-ix \cdot \nu}\, d^n x$
501	$\chi_{[0,1]}(\lvert x\rvert)(1 - \lvert x\rvert^2)^{\delta}$	$\pi^{-\delta}\Gamma(\delta+1)\lvert\xi\rvert^{-(n/2)-\delta} \cdot J_{n/2+\delta}(2\pi\lvert\xi\rvert)$	$2^{-\delta}\Gamma(\delta+1)\,\lvert\omega\rvert^{-(n/2)-\delta} \cdot J_{n/2+\delta}(\lvert\omega\rvert)$	$\pi^{-\delta}\Gamma(\delta+1)\left\lvert\frac{\nu}{2\pi}\right\rvert^{-(n/2)-\delta} \cdot J_{n/2+\delta}(\lvert\nu\rvert)$
502	$\lvert x\rvert^{-\alpha}, \quad 0 < \operatorname{Re}\alpha < n.$	$c_\alpha \lvert\xi\rvert^{-(n-\alpha)}$		

Remarks

To 501: The function $\chi_{[0,1]}$ is the indicator function of the interval [0, 1]. The function $\Gamma(x)$ is the gamma function. The function $J_{n/2+\delta}$ is a Bessel function of the first kind, with order $n/2 + \delta$. Taking $n = 2$ and $\delta = 0$ produces 402. (Stein & Weiss 1971, Thm. 4.15)

To 502: See Riesz potential. The formula also holds for all $\alpha \neq -n, -n-1, \ldots$ by analytic continuation, but then the function and its Fourier transforms need to be understood as suitably regularized tempered distributions. See homogeneous distribution.

References

- Boashash, B., ed. (2003), *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Oxford: Elsevier Science, ISBN 0-08-044335-4
- Bochner S., Chandrasekharan K. (1949), *Fourier Transforms*, Princeton University Press
- Bracewell, R. N. (2000), *The Fourier Transform and Its Applications* (3rd ed.), Boston: McGraw-Hill, ISBN 0-07-116043-4.
- Campbell, George; Foster, Ronald (1948), *Fourier Integrals for Practical Applications*, New York: D. Van Nostrand Company, Inc..
- Duoandikoetxea, Javier (2001), *Fourier Analysis*, American Mathematical Society, ISBN 0-8218-2172-5.
- Dym, H; McKean, H (1985), *Fourier Series and Integrals*, Academic Press, ISBN 978-0-12-226451-1.
- Erdélyi, Arthur, ed. (1954), *Tables of Integral Transforms*, **1**, New Your: McGraw-Hill
- Fourier, J. B. Joseph (1822), *Théorie Analytique de la Chaleur* [1], Paris
- Grafakos, Loukas (2004), *Classical and Modern Fourier Analysis*, Prentice-Hall, ISBN 0-13-035399-X.
- Hewitt, Edwin; Ross, Kenneth A. (1970), *Abstract harmonic analysis. Vol. II: Structure and analysis for compact groups. Analysis on locally compact Abelian groups*, Die Grundlehren der mathematischen Wissenschaften, Band 152, Berlin, New York: Springer-Verlag, MR0262773.
- Hörmander, L. (1976), *Linear Partial Differential Operators, Volume 1*, Springer-Verlag, ISBN 978-3-540-00662-6.
- James, J.F. (2011), *A Student's Guide to Fourier Transforms* (3rd ed.), New York: Cambridge University Press, ISBN 978-0-521-17683-5.
- Kaiser, Gerald (1994), *A Friendly Guide to Wavelets*, Birkhäuser, ISBN 0-8176-3711-7
- Kammler, David (2000), *A First Course in Fourier Analysis*, Prentice Hall, ISBN 0-13-578782-3
- Katznelson, Yitzhak (1976), *An introduction to Harmonic Analysis*, Dover, ISBN 0-486-63331-4
- Knapp, Anthony W. (2001), *Representation Theory of Semisimple Groups: An Overview Based on Examples* [2], Princeton University Press, ISBN 978-0-691-09089-4
- Pinsky, Mark (2002), *Introduction to Fourier Analysis and Wavelets*, Brooks/Cole, ISBN 0-534-37660-6
- Polyanin, A. D.; Manzhirov, A. V. (1998), *Handbook of Integral Equations*, Boca Raton: CRC Press, ISBN 0-8493-2876-4.
- Rudin, Walter (1987), *Real and Complex Analysis* (Third ed.), Singapore: McGraw Hill, ISBN 0-07-100276-6.
- Stein, Elias; Shakarchi, Rami (2003), *Fourier Analysis: An introduction*, Princeton University Press, ISBN 0-691-11384-X.
- Stein, Elias; Weiss, Guido (1971), *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton, N.J.: Princeton University Press, ISBN 978-0-691-08078-9.
- Wilson, R. G. (1995), *Fourier Series and Optical Transform Techniques in Contemporary Optics*, New York: Wiley, ISBN 0-471-30357-7.
- Yosida, K. (1968), *Functional Analysis*, Springer-Verlag, ISBN 3-540-58654-7.

External links

- The Discrete Fourier Transformation (DFT): Definition and numerical examples [3] - A Matlab tutorial
- Fourier Series Applet [4] (Tip: drag magnitude or phase dots up or down to change the wave form).
- Stephan Bernsee's FFTlab [5] (Java Applet)
- Stanford Video Course on the Fourier Transform [6]
- Weisstein, Eric W., "Fourier Transform [7]" from MathWorld.
- The DFT “à Pied”: Mastering The Fourier Transform in One Day [8] at The DSP Dimension
- An Interactive Flash Tutorial for the Fourier Transform [9]

References

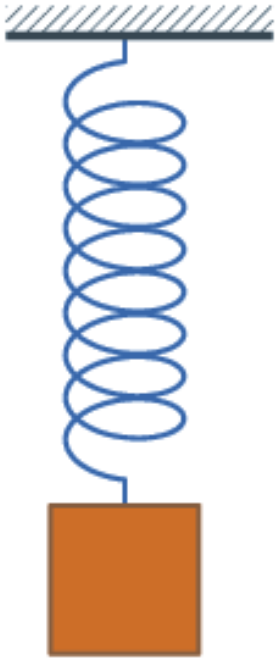
[1] http://books.google.com/?id=TDQJAAAAIAAJ&printsec=frontcover&dq=Th%C3%A9orie+analytique+de+la+chaleur&q
[2] http://books.google.com/?id=QCcW1h835pwC
[3] http://www.nbtwiki.net/doku.php?id=tutorial:the_discrete_fourier_transformation_dft
[4] http://www.westga.edu/~jhasbun/osp/Fourier.htm
[5] http://www.dspdimension.com/fftlab/
[6] http://www.academicearth.com/courses/the-fourier-transform-and-its-applications
[7] http://mathworld.wolfram.com/FourierTransform.html
[8] http://www.dspdimension.com/admin/dft-a-pied/
[9] http://www.fourier-series.com/f-transform/index.html

Harmonic oscillator

In classical mechanics, a **harmonic oscillator** is a system that, when displaced from its equilibrium position, experiences a restoring force, *F*, proportional to the displacement, *x*:

$$\vec{F} = -k\vec{x}$$

where *k* is a positive constant.



An undamped spring-mass system is a simple harmonic oscillator.

If *F* is the only force acting on the system, the system is called a **simple harmonic oscillator**, and it undergoes simple harmonic motion: sinusoidal oscillations about the equilibrium point, with a constant amplitude and a constant frequency (which does not depend on the amplitude).

If a frictional force (damping) proportional to the velocity is also present, the harmonic oscillator is described as a **damped oscillator**. Depending on the friction coefficient, the system can:

- Oscillate with a frequency smaller than in the non-damped case, and an amplitude decreasing with time (*underdamped* oscillator).
- Decay to the equilibrium position, without oscillations (*overdamped* oscillator).

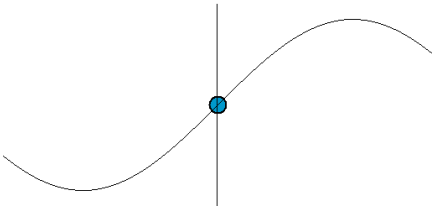
The boundary solution between an underdamped oscillator and an overdamped oscillator occurs at a particular value of the friction coefficient and is called "critically damped".

If an external time dependent force is present, the harmonic oscillator is described as a **driven oscillator**.

Mechanical examples include pendula (with small angles of displacement), masses connected to springs, and acoustical systems. Other analogous systems include electrical harmonic oscillators such as RLC circuits. The harmonic oscillator model is very important in physics, because any mass subject to a force in stable equilibrium acts as a harmonic oscillator for small vibrations. Harmonic oscillators occur widely in nature and are exploited in many manmade devices, such as clocks and radio circuits. They are the source of virtually all sinusoidal vibrations and waves.

Simple harmonic oscillator

A simple harmonic oscillator is an oscillator that is neither driven nor damped. It consists of a mass *m*,which experiences a single force, *F*, which pulls the mass in the direction of the point *x*=0 and depends only on the mass's position *x* and a constant *k*. Newton's second law for the system is



Simple harmonic motion.

$$F = ma = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -kx.$$

Solving this differential equation, we find that the motion is described by the function

$$x(t) = A\cos\left(2\pi ft + \phi\right),$$

where

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{T}.$$

The motion is periodic— repeating itself in a sinusoidal fashion with constant amplitude, A. In addition to its amplitude, the motion of a simple harmonic oscillator is characterized by its period T, the time for a single oscillation or its frequency $f = {}^{1}/_{T}$, the number of cycles per unit time. The position at a given time t also depends on the phase, φ, which determines the starting point on the sine wave. The period and frequency are determined by the size of the mass m and the force constant k, while the amplitude and phase are determined by the starting position and velocity.

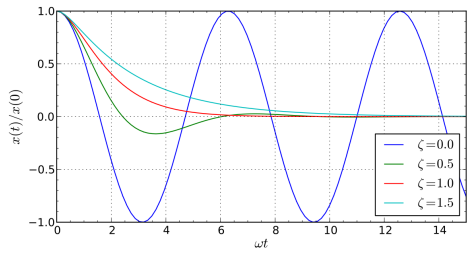
The velocity and acceleration of a simple harmonic oscillator oscillate with the same frequency as the position but with shifted phases. The velocity is maximum for zero displacement, while the acceleration is in the opposite direction as the displacement.

The potential energy stored in a simple harmonic oscillator at position x is

$$U = \frac{1}{2}kx^2.$$

Damped harmonic oscillator

In real oscillators, friction, or damping, slows the motion of the system. In many vibrating systems the frictional force F_f can be modeled as being proportional to the velocity v of the object: $F_f = -cv$, where c is called the *viscous damping coefficient*.



Dependence of the system behavior on the value of the damping ratio ζ

Newton's second law for damped harmonic oscillators is then

$$F = -kx - c\frac{\mathrm{d}x}{\mathrm{d}t} = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}.$$

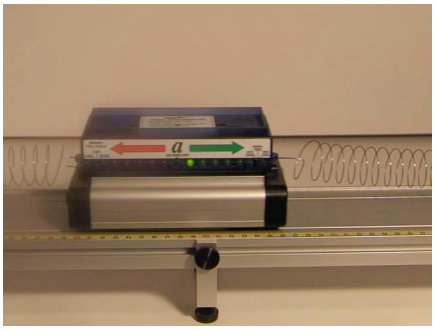
This is rewritten into the form

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\zeta\omega_0\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2x = 0,$$

where

$\omega_0 = \sqrt{\frac{k}{m}}$ is called the 'undamped angular frequency of the oscillator' and

$\zeta = \frac{c}{2m\omega_0}$ is called the 'damping ratio'.

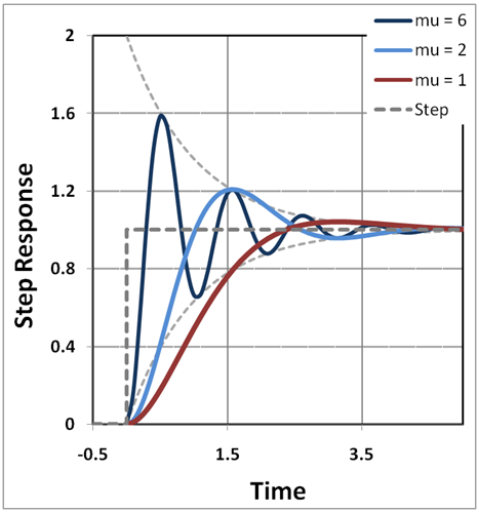


A damped harmonic oscillator, which slows down due to friction

The value of the damping ratio ζ critically determines the behavior of the system. A damped harmonic oscillator can be:

- *Overdamped* ($\zeta > 1$): The system returns (exponentially decays) to equilibrium without oscillating. Larger values of the damping ratio ζ return to equilibrium slower.
- *Critically damped* ($\zeta = 1$): The system returns to equilibrium as quickly as possible without oscillating. This is often desired for the damping of systems such as doors.
- *Underdamped* ($\zeta < 1$): The system oscillates (with a slightly different frequency than the undamped case) with the amplitude gradually decreasing to zero. The angular frequency of the underdamped harmonic oscillator is given by

$$\omega_1 = \omega_0\sqrt{1-\zeta^2}.$$



Step-response of a damped harmonic oscillator; curves are plotted for three values of $\mu = \omega_1 = \omega_0\sqrt{1-\zeta^2}$. Time is in units of the decay time $\tau = 1/(\zeta\omega_0)$.

The Q factor of a damped oscillator is defined as

$$Q = 2\pi \times \frac{\text{Energy stored}}{\text{Energy lost per cycle}}.$$

Q is related to the damping ratio by the equation $Q = \dfrac{1}{2\zeta}$.

Driven harmonic oscillators

Driven harmonic oscillators are damped oscillators further affected by an externally applied force $F(t)$.

Newton's second law takes the form

$$F(t) - kx - c\frac{\mathrm{d}x}{\mathrm{d}t} = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}.$$

It is usually rewritten into the form

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\zeta\omega_0\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2x = \frac{F(t)}{m}.$$

This equation can be solved exactly for any driving force using the solutions $z(t)$ to the unforced equation, which satisfy

$$\frac{\mathrm{d}^2z}{\mathrm{d}t^2} + 2\zeta\omega_0\frac{\mathrm{d}z}{\mathrm{d}t} + \omega_0^2z = 0,$$

and which can be expressed as damped sinusoidal oscillations,

$$z(t) = A\mathrm{e}^{-\zeta\omega_0 t}\ \sin\left(\sqrt{1-\zeta^2}\ \omega_0 t + \varphi\right),$$

in the case where $\zeta \leq 1$. The amplitude A and phase φ determine the behavior needed to match the initial conditions.

Step input

In the case $\zeta < 1$ and a unit step input with $x(0) = 0$:

$$F(t) = \begin{cases} \omega_0^2 & t \geq 0 \\ 0 & t < 0 \end{cases}$$

the solution is:

$$x(t) = 1 - \mathrm{e}^{-\zeta\omega_0 t} \frac{\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t + \varphi\right)}{\sin(\varphi)},$$

with phase φ given by

$$\cos\varphi = \zeta.$$

This behavior is found in (for example) feedback amplifiers, where the amplifier design is adjusted to obtain the fastest step response possible without undue overshoot or undershoot and with an adequate settling time.

The time an oscillator needs to adapt to changed external conditions is of the order $\tau = 1/(\zeta\omega_0)$. In physics, the adaptation is called relaxation, and τ is called the relaxation time.

In electrical engineering, a multiple of τ is called the *settling time*, i.e. the time necessary to insure the signal is within a fixed departure from final value, typically within 10%. The term *overshoot* refers to the extent the maximum response exceeds final value, and *undershoot* refers to the extent the response falls below final value for times following the maximum response.

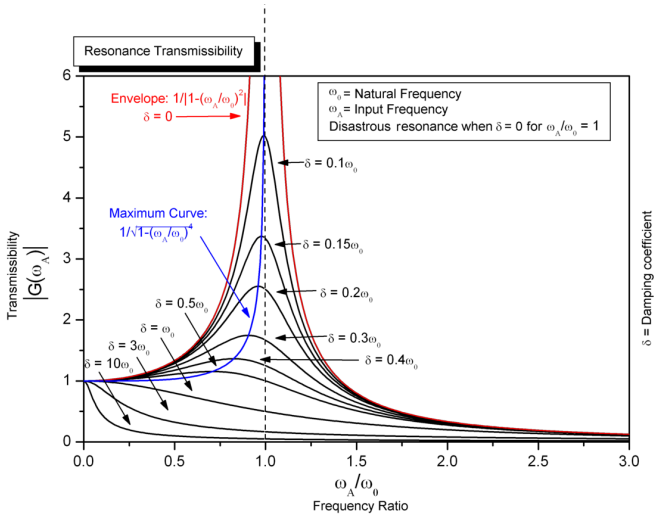
Sinusoidal driving force

In the case of a sinusoidal driving force:

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\zeta\omega_0\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x = \frac{1}{m}F_0\sin(\omega t),$$

where F_0 is the driving amplitude and ω is the driving frequency for a sinusoidal driving mechanism. This type of system appears in AC driven RLC circuits (resistor-inductor-capacitor) and driven spring systems having internal mechanical resistance or external air resistance.

The general solution is a sum of a transient solution that depends on initial conditions, and a steady state that is independent of initial conditions and depends only on the driving amplitude F_0, driving frequency, ω, undamped angular frequency ω_0, and the damping ratio ζ.



Steady state variation of amplitude with frequency and damping of a driven simple harmonic oscillator.[1][2]

The steady-state solution is proportional to the driving force with an induced phase change of ϕ:

$$x(t) = \frac{F_0}{mZ_m\omega}\sin(\omega t + \phi)$$

where

$$Z_m = \sqrt{(2\omega_0\zeta)^2 + \frac{1}{\omega^2}\left(\omega_0^2 - \omega^2\right)^2}$$

is the absolute value of the impedance or linear response function and

$$\phi = \arctan\left(\frac{2\omega\omega_0\zeta}{\omega^2 - \omega_0^2}\right)$$

is the phase of the oscillation relative to the driving force.

For a particular driving frequency called the resonance, or resonant frequency $\omega_r = \omega_0\sqrt{1 - 2\zeta^2}$, the amplitude (for a given F_0) is maximum. This resonance effect only occurs when $\zeta < 1/\sqrt{2}$, i.e. for significantly underdamped systems. For strongly underdamped systems the value of the amplitude can become quite large near the resonance frequency.
The transient solutions are the same as the unforced ($F_0 = 0$) damped harmonic oscillator and represent the systems response to other events that occurred previously. The transient solutions typically die out rapidly enough that they can be ignored.

Parametric oscillators

A parametric oscillator is a harmonic oscillator whose parameters oscillate in time. A familiar example of both parametric and driven oscillation is playing on a swing.[3][4][5] Rocking back and forth pumps the swing as a driven harmonic oscillator, but once moving, the swing can also be parametrically driven by alternately standing and squatting at key points in the swing. The varying of the parameters drives the system. Examples of parameters that may be varied are its resonance frequency ω and damping β .

Parametric oscillators are used in many applications. The classical varactor parametric oscillator oscillates when the diode's capacitance is varied periodically. The circuit that varies the diode's capacitance is called the "pump" or "driver". In microwave electronics, waveguide/YAG based parametric oscillators operate in the same fashion. The designer varies a parameter periodically to induce oscillations.

Parametric oscillators have been developed as low-noise amplifiers, especially in the radio and microwave frequency range. Thermal noise is minimal, since a reactance (not a resistance) is varied. Another common use is frequency conversion, e.g., conversion from audio to radio frequencies. For example, the Optical parametric oscillator converts an input laser wave into two output waves of lower frequency (ω_s, ω_i).

Parametric resonance occurs in a mechanical system when a system is parametrically excited and oscillates at one of its resonant frequencies. Parametric excitation differs from forcing, since the action appears as a time varying modification on a system parameter. This effect is different from regular resonance because it exhibits the instability phenomenon.

Universal oscillator equation

The equation

$$\frac{\mathrm{d}^2q}{\mathrm{d}\tau^2} + 2\zeta\frac{\mathrm{d}q}{\mathrm{d}\tau} + q = 0$$

is known as the **universal oscillator equation** since all second order linear oscillatory systems can be reduced to this form. This is done through nondimensionalization.

If the forcing function is $f(t) = \cos(\omega t) = \cos(\omega t_c\tau) = \cos(\omega\tau)$, where $\omega = \omega t_c$, the equation becomes

$$\frac{\mathrm{d}^2q}{\mathrm{d}\tau^2} + 2\zeta\frac{\mathrm{d}q}{\mathrm{d}\tau} + q = \cos(\omega\tau).$$

The solution to this differential equation contains two parts, the "transient" and the "steady state".

Transient solution

The solution based on solving the ordinary differential equation is for arbitrary constants c_1 and c_2

$$q_t(\tau) = \begin{cases} \mathrm{e}^{-\zeta\tau}\left(c_1\mathrm{e}^{\tau\sqrt{\zeta^2-1}} + c_2\mathrm{e}^{-\tau\sqrt{\zeta^2-1}}\right) & \zeta > 1 \text{ (overdamping)} \\ \mathrm{e}^{-\zeta\tau}(c_1 + c_2\tau) = \mathrm{e}^{-\tau}(c_1 + c_2\tau) & \zeta = 1 \text{ (critical damping)} \\ \mathrm{e}^{-\zeta\tau}\left[c_1\cos\left(\sqrt{1-\zeta^2}\tau\right) + c_2\sin\left(\sqrt{1-\zeta^2}\tau\right)\right] & \zeta < 1 \text{(underdamping)} \end{cases}$$

The transient solution is independent of the forcing function.

Steady-state solution

Apply the "complex variables method" by solving the auxiliary equation below and then finding the real part of its solution:

$$\frac{\mathrm{d}^2q}{\mathrm{d}\tau^2} + 2\zeta\frac{\mathrm{d}q}{\mathrm{d}\tau} + q = \cos(\omega\tau) + \mathrm{i}\sin(\omega\tau) = \mathrm{e}^{\mathrm{i}\omega\tau}.$$

Supposing the solution is of the form

$$q_s(\tau) = A\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)}.$$

Its derivatives from zero to 2nd order are

$$q_s = A\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)},\ \frac{\mathrm{d}q_s}{\mathrm{d}\tau} = \mathrm{i}\omega A\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)},\ \frac{\mathrm{d}^2q_s}{\mathrm{d}\tau^2} = -\omega^2 A\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)}.$$

Substituting these quantities into the differential equation gives

$$-\omega^2 A\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)} + 2\zeta\mathrm{i}\omega A\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)} + A\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)} = (-\omega^2 A + 2\zeta\mathrm{i}\omega A + A)\mathrm{e}^{\mathrm{i}(\omega\tau+\phi)} = \mathrm{e}^{\mathrm{i}\omega\tau}.$$

Dividing by the exponential term on the left results in

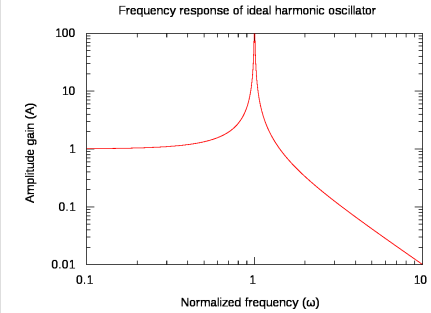
$$-\omega^2 A + 2\zeta\mathrm{i}\omega A + A = \mathrm{e}^{-\mathrm{i}\phi} = \cos\phi - \mathrm{i}\sin\phi.$$

Equating the real and imaginary parts results in two independent equations

$$A(1-\omega^2) = \cos\phi \qquad 2\zeta\omega A = -\sin\phi.$$

Amplitude part

Squaring both equations and adding them together gives



Bode plot of the frequency response of an ideal harmonic oscillator.

$$\left.\begin{aligned} A^2(1-\omega^2)^2 &= \cos^2\phi \\ (2\zeta\omega A)^2 &= \sin^2\phi \end{aligned}\right\} \Rightarrow A^2[(1-\omega^2)^2 + (2\zeta\omega)^2] = 1.$$

Therefore,

$$A = A(\zeta, \omega) = sign\left(\frac{-\sin\phi}{2\zeta\omega}\right)\frac{1}{\sqrt{(1-\omega^2)^2 + (2\zeta\omega)^2}}.$$

Compare this result with the theory section on resonance, as well as the "magnitude part" of the RLC circuit. This amplitude function is particularly important in the analysis and understanding of the frequency response of second-order systems.

Phase part

To solve for φ, divide both equations to get

$$\tan\phi = -\frac{2\zeta\omega}{1-\omega^2} = \frac{2\zeta\omega}{\omega^2-1} \Rightarrow \phi \equiv \phi(\zeta,\omega) = \arctan\left(\frac{2\zeta\omega}{\omega^2-1}\right).$$

This phase function is particularly important in the analysis and understanding of the frequency response of second-order systems.

Full solution

Combining the amplitude and phase portions results in the steady-state solution

$$q_s(\tau) = A(\zeta,\omega)\cos(\omega\tau + \phi(\zeta,\omega)) = A\cos(\omega\tau+\phi).$$

The solution of original universal oscillator equation is a superposition (sum) of the transient and steady-state solutions

$$q(\tau) = q_t(\tau) + q_s(\tau).$$

For a more complete description of how to solve the above equation, see linear ODEs with constant coefficients.

Equivalent systems

Harmonic oscillators occurring in a number of areas of engineering are equivalent in the sense that their mathematical models are identical (see universal oscillator equation above). Below is a table showing analogous quantities in four harmonic oscillator systems in mechanics and electronics. If analogous parameters on the same line in the table are given numerically equal values, the behavior of the oscillators—their output waveform, resonant frequency, damping factor, etc.—are the same.

Translational Mechanical	**Torsional Mechanical**	**Series RLC Circuit**	**Parallel RLC Circuit**
Position x	Angle θ	Charge q	Voltage e
Velocity $\frac{\mathrm{d}x}{\mathrm{d}t}$	Angular velocity $\frac{\mathrm{d}\theta}{\mathrm{d}t}$	Current $\frac{\mathrm{d}q}{\mathrm{d}t}$	$\frac{\mathrm{d}e}{\mathrm{d}t}$
Mass M	Moment of inertia I	Inductance L	Capacitance C
Spring constant K	Torsion constant μ	Elastance $1/C$	Susceptance $1/L$
Friction γ	Rotational friction Γ	Resistance R	Conductance $1/R$
Drive force $F(t)$	Drive torque $\tau(t)$	e	$\mathrm{d}i/\mathrm{d}t$
Undamped resonant frequency f_n:			
$\frac{1}{2\pi}\sqrt{\frac{K}{M}}$	$\frac{1}{2\pi}\sqrt{\frac{\mu}{I}}$	$\frac{1}{2\pi}\sqrt{\frac{1}{LC}}$	$\frac{1}{2\pi}\sqrt{\frac{1}{LC}}$
Differential equation:			
$M\ddot{x} + \gamma\dot{x} + Kx = F$	$I\ddot{\theta} + \Gamma\dot{\theta} + \mu\theta = \tau$	$L\ddot{q} + R\dot{q} + q/C = e$	$C\ddot{e} + \dot{e}/R + e/L = \dot{i}$

Application to a conservative force

The problem of the simple harmonic oscillator occurs frequently in physics, because a mass at equilibrium under the influence of any conservative force, in the limit of small motions, behaves as a simple harmonic oscillator.

A conservative force is one that has a potential energy function. The potential energy function of a harmonic oscillator is:

$$V(x) = \frac{1}{2}kx^2$$

Given an arbitrary potential energy function $V(x)$, one can do a Taylor expansion in terms of x around an energy minimum ($x = x_0$) to model the behavior of small perturbations from equilibrium.

$$V(x) = V(x_0) + (x - x_0)V'(x_0) + \frac{1}{2}(x - x_0)^2V^{(2)}(x_0) + O(x - x_0)^3$$

Because $V(x_0)$ is a minimum, the first derivative evaluated at x_0 must be zero, so the linear term drops out:

$$V(x) = V(x_0) + \frac{1}{2}(x - x_0)^2V^{(2)}(x_0) + O(x - x_0)^3$$

The constant term $V(x_0)$ is arbitrary and thus may be dropped, and a coordinate transformation allows the form of the simple harmonic oscillator to be retrieved:

$$V(x) \approx \frac{1}{2}x^2V^{(2)}(0) = \frac{1}{2}kx^2$$

Thus, given an arbitrary potential energy function $V(x)$ with a non-vanishing second derivative, one can use the solution to the simple harmonic oscillator to provide an approximate solution for small perturbations around the equilibrium point.

Examples

Simple pendulum

Assuming no damping and small amplitudes, the differential equation governing a simple pendulum is

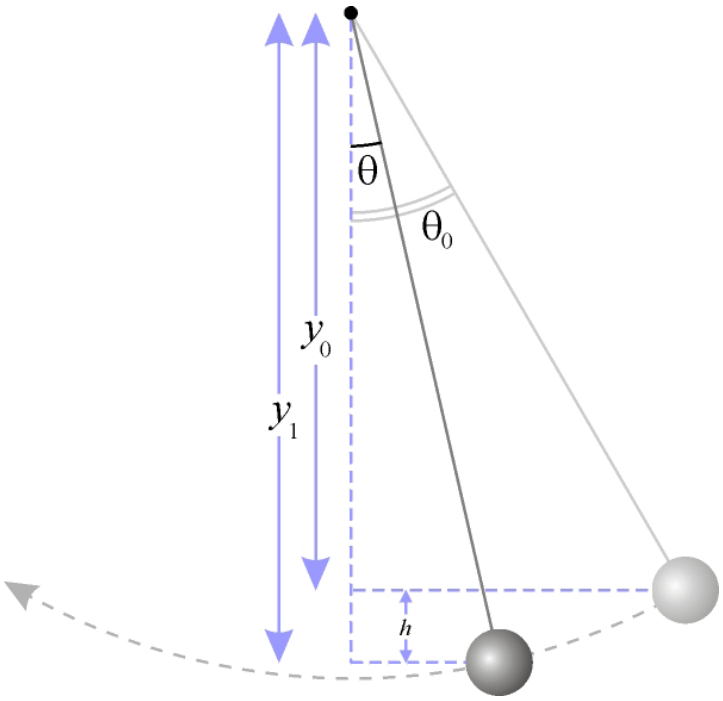
$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \frac{g}{\ell}\theta = 0.$$

The solution to this equation is given by:

$$\theta(t) = \theta_0 \cos\left(\sqrt{\frac{g}{\ell}}t\right) \qquad |\theta_0| \ll 1$$

where θ_0 is the largest angle attained by the pendulum. The period, the time for one complete oscillation, is given by 2π divided by whatever is multiplying the time in the argument of the cosine ($\sqrt{\frac{g}{\ell}}$ here).

$$T_0 = 2\pi\sqrt{\frac{\ell}{g}} \qquad |\theta_0| \ll 1.$$



A simple pendulum exhibits simple harmonic motion under the conditions of no damping and small amplitude.

Pendulum swinging over turntable

Simple harmonic motion can in some cases be considered to be the one-dimensional projection of two-dimensional circular motion. Consider a long pendulum swinging over the turntable of a record player. On the edge of the turntable there is an object. If the object is viewed from the same level as the turntable, a projection of the motion of the object seems to be moving backwards and forwards on a straight line orthogonal to the view direction, sinusoidally like the pendulum.

Spring–mass system

When a spring is stretched or compressed by a mass, the spring develops a restoring force. Hooke's law gives the relationship of the force exerted by the spring when the spring is compressed or stretched a certain length:

$$F\left(t\right) = -kx\left(t\right)$$

where F is the force, k is the spring constant, and x is the displacement of the mass with respect to the equilibrium position. This relationship shows that the distance of the spring is always opposite to the force of the spring.

By using either force balance or an energy method, it can be readily shown that the motion of this system is given by the following differential equation:

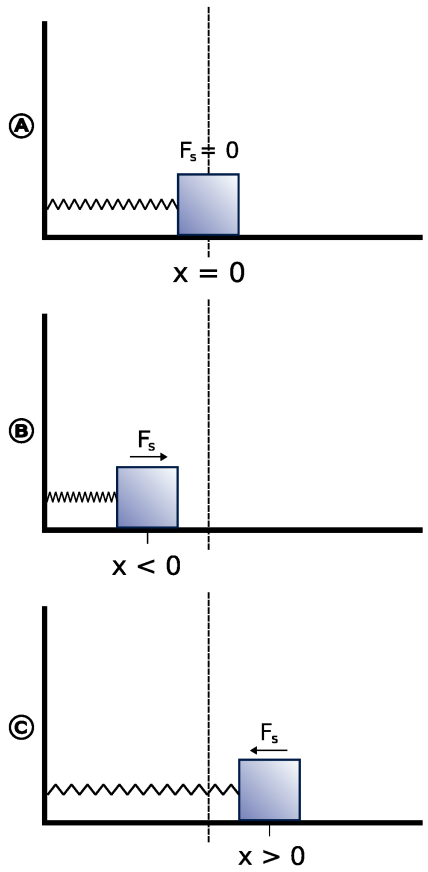
$$F(t) = -kx(t) = m\frac{\mathrm{d}^2}{\mathrm{d}t^2}x\left(t\right) = ma.$$

...the latter evidently being Newton's second law of motion.

If the initial displacement is A, and there is no initial velocity, the solution of this equation is given by:

$$x\left(t\right) = A\cos\left(\sqrt{\frac{k}{m}}t\right).$$

Given an ideal massless spring, m is the mass on the end of the spring. If the spring itself has mass, its effective mass must be included in m .



Spring–mass system in equilibrium (A), compressed (B) and stretched (C) states.

Energy variation in the spring–damping system

In terms of energy, all systems have two types of energy, potential energy and kinetic energy. When a spring is stretched or compressed, it stores elastic potential energy, which then is transferred into kinetic energy. The potential energy within a spring is determined by the equation $U = kx^2/2.$

When the spring is stretched or compressed, kinetic energy of the mass gets converted into potential energy of the spring. By conservation of energy, assuming the datum is defined at the equilibrium position, when the spring reaches its maximum potential energy, the kinetic energy of the mass is zero. When the spring is released, it tries to return to equilibrium, and all its potential energy converts to kinetic energy of the mass.

Notes

[1] Katsuhiko Ogata (2005). *System Dynamics* (4th ed.). University of Minnesota. p. 617.

[2] Ajoy Ghatak (2005). *Optics, 3E* (http://books.google.com/books?id=jStDc2LmU5IC&pg=PT97&dq=damping-decreases+resonance+amplitude#v=onepage&q=damping-decreases resonance amplitude&f=false) (3rd ed.). Tata McGraw-Hill. p. 6.10. ISBN 9780070585836. .

[3] Case, William. "Two ways of driving a child's swing" (http://www.grinnell.edu/academic/physics/faculty/case/swing/). . Retrieved 27 November 2011.

[4] Case, W. B. (1996). "The pumping of a swing from the standing position". *American Journal of Physics* **64**: 215–220.

[5] Roura, P.; Gonzalez, J.A. (2010). "Towards a more realistic description of swing pumping due to the exchange of angular momentum". *European Journal of Physics* **31**: 1195–1207.

References

- Serway, Raymond A.; Jewett, John W. (2003). *Physics for Scientists and Engineers*. Brooks/Cole. ISBN 0-534-40842-7.
- Tipler, Paul (1998). *Physics for Scientists and Engineers: Vol. 1* (4th ed.). W. H. Freeman. ISBN 1-57259-492-6.
- Wylie, C. R. (1975). *Advanced Engineering Mathematics* (4th ed.). McGraw-Hill. ISBN 0-07-072180-7.

External links

- Harmonic Oscillator (http://hypertextbook.com/chaos/41.shtml) from The Chaos Hypertextbook
- A Java applet of harmonic oscillator with damping proportional to velocity or damping caused by dry friction (http://phy.hk/wiki/englishhtm/Damped.htm)

Article Sources and Contributors

Signal (electronics) *Source*: http://en.wikipedia.org/w/index.php?oldid=483914857 *Contributors*: Aaronbrick, Adoniscik, Ancheta Wis, Andrus Kallastu, Andycjp, Bomac, Borgx, BorzouBarzegar, Causa sui, Cburnett, Cruccone, David Shay, Dicklyon, Dreadstar, Ensign beedrill, Espetkov, Firefly322, Grubber, Guy Macon, HoodedMan, Ian Pitchford, Jeff3000, Jogloran, JordoCo, Jorge Stolfi, Kku, Mange01, Martynas Patasius, Michael Hardy, Mild Bill Hiccup, Misteror, Moostang, Msadaghd, Nbarth, Nilfanion, Num43, PATLORICK, Pgan002, PieRRoMaN, Purgatory Fubar, Radiojon, Remindmelater, Rjwilmsi, Ro8269, Romanm, Shanel, Sikory, Skizzik, Smack, Steve Pucci, Swagato Barman Roy, Tagremover, The Photon, Unomano, Volfy, Vrenator, Wam junkie, Yves-Laurent, Zvika, 91 anonymous edits

Even and odd functions *Source*: http://en.wikipedia.org/w/index.php?oldid=481199718 *Contributors*: 16@r, Albedo, Anskas, BiT, CALR, CecilBlade, Charles Matthews, Cremepuff222, Dbfirs, Dicklyon, Dirac1933, Doctormatt, Dogah, Download, Dysprosia, El aprendelenguas, Fayenatic london, Gandalf61, GeordieMcBain, Giftlite, Golgofrinchian, Ih8evilstuff, Johnuniq, KPH2293, KYN, KYPark, Kilva, Laurentius, Lixy, MagneticFlux, Marek69, Marino-slo, Melchoir, Mets501, Michael Devore, Michael Hardy, Mochi, Mormegil, Nutster, Old Death, Oleg Alexandrov, Omegatron, PV=nRT, Paul August, Pgan002, Plastikspork, Ploppery555, Rbj, Revolver, RobHar, Salgueiro, Sideways713, Six.oh.six, Skamecrazy123, Spoon!, Staka, Taejo, TeleComNasSprVen, Tobias Bergemann, Tomruen, WOSlinker, Yoshigev, Zvika, გიგა, 145 anonymous edits

Linear system *Source*: http://en.wikipedia.org/w/index.php?oldid=474379310 *Contributors*: Allan McInnes, Athkalani, Ben pcc, BenFrantzDale, Berland, Btyner, Cburnett, Charles Matthews, Cihan, Creidieki, DR04, Digfarenough, El C, First Harmonic, FocalPoint, Frozenport, Gwaihir, Hadal, Icairns, Jgreeter, Jitse Niesen, K-UNIT, Krsti, Lommer, Lowercase Sigma, MarkSweep, Mathmanta, Maurice Carbonaro, Michael Hardy, Mks004, Nixdorf, Nvrmnd, Pak21, Paolo.dL, Rbj, Reyk, Sanfranman59, Sapphic, Srleffler, Treesmill, Vanakaris, Vanish2, VolvoMan, WhiteHatLurker, سعی, 40 anonymous edits

Time-invariant system *Source*: http://en.wikipedia.org/w/index.php?oldid=474097810 *Contributors*: Arroww, Belizefan, Cburnett, Giftlite, HatlessAtlas, HenningThielemann, JHunterJ, Jamelan, Jmaliakal, Michael Hardy, Mwilde, Nbarth, Oleg Alexandrov, OlivierMiR, Prinsen, R'n'B, Rinconsoleao, Ro8269, Siddhant, Sterrys, 老陳, 46 anonymous edits

Causal system *Source*: http://en.wikipedia.org/w/index.php?oldid=465467725 *Contributors*: 90 Auto, Dja25, Dominic, Francisco Del Piero, Halfridge, Jpbowen, LachlanA, Lankiveil, LiDaobing, Mdd, Michael Hardy, Nixdorf, Pg8p, RJFJR, Rafy, Rbj, SimonP, Theresa knott, Tomer shalev, Zvika, 37 anonymous edits

Sampling (signal processing) *Source*: http://en.wikipedia.org/w/index.php?oldid=483786350 *Contributors*: ANDROBETA, Adoniscik, Andres, Barking Mad42, Binksternet, Bjbarman18, Bob K, BorgQueen, Chemako0606, Colonies Chris, ContinueWithCaution, Cosmin1595, DaBler, Dan Polansky, Deineka, Dicklyon, Edcolins, Email4mobile, Epbr123, Gene Nygaard, Ggbgggb, Giftlite, Gionnico, Heron, History2007, ImEditingWiki, Infvwl, JLaTondre, JaGa, Jan Arkesteijn, Jheald, Jk2q3jrklse, John Holmes II, JulienBourdon, Kimchi.sg, Kjoonlee, Kvng, LilHelpa, Lorem Ip, LutzL, MagnusA, Mattbr, Miym, Nanren888, Nbarth, Oleg Alexandrov, OverlordQ, Patriarch, Philip Trueman, Pvosta, Rami R, Ravn, Rbj, Selket, Sigmundg, Smack, Soliloquial, Spinningspark, StefanosNikolaou, Teapeat, Thiseye, Trusilver, Txomin, Unused0022, Wolfkeeper, Zvika, 114 anonymous edits

Nyquist–Shannon sampling theorem *Source*: http://en.wikipedia.org/w/index.php?oldid=484521172 *Contributors*: .:Ajvol:., 195.149.37.xxx, 24.7.91.xxx, A4, Acdx, Ash, AxelBoldt, Bdmy, BenFrantzDale, Bender235, BeteNoir, Bob K, Brad7777, Brion VIBBER, Cabe6403, Camembert, Cameronc, Cburnett, Cihan, Colonies Chris, Conversion script, CosineKitty, Courcelles, Cryptonaut, DARTH SIDIOUS 2, Damian Yerrick, Dbagnall, Default007, Deodar, Dicklyon, Dougmcgeehan, Drhex, Duagloth, Dysprosia, Editor at Large, Enochlau, ErickOrtiz, Fang Aili, Finell, First Harmonic, Fleem, Floatjon, Frazzydee, Gah4, Gaius Cornelius, Geekmug, Gene Nygaard, Giftlite, Graham87, Heron, Hogno, Hrafn, Ian Pitchford, IanOfNorwich, IgorCarron, IstvanWolf, Jac16888, Jacobolus, Jannex, Jasón, Jesse Viviano, Jheald, Jncraton, Juansempere, KYN, Karn, Kenyon, Ketil3, Krash, Larry Sanger, Lavaka, LenoreMarks, Ligand, Light current, Lightmouse, Loisel, LouScheffer, Lowellian, Lupin, LutzL, M4701, Maansi.t, Maghnus, Mange01, Manscher, Marcoracer, Mav, Mboverload, Metacomet, Mets501, Michael Hardy, Michaelmestre, Minochang, Miym, Movado73, Mperrin, Mrand, Mwilde, Nafeun, Nbarth, Neckro, Niteowlneils, Nuwewsco, Odo Benus, Oli Filth, Omegatron, Oskar Sigvardsson, OverlordQ, Patrick, Perelaar, Pfalstad, PhotoBox, Pr103, Psychonaut, Pvosta, Qef, RAE, RDBury, Rbj, Reedy, Robert K S, Sam Hocevar, Sander123, Sawf, SciCompTeacher, SebastianHelm, Shelbymoore3, Skittleys, SlimDeli, Slipstream, Smack, Solarapex, Stupefaction, TedPavlic, The Thing That Should Not Be, Thenub314, Tiddly Tom, Tomas e, Tompagenet, Waveguy, Wolfkeeper, ZeroOne, Zueignung, Zundark, Zvika, 286 anonymous edits

Aliasing *Source*: http://en.wikipedia.org/w/index.php?oldid=481836645 *Contributors*: ANDROBETA, Ablewisuk, Aeons, Anomalocaris, Arconada, Awotter, BenFrantzDale, BernardH, Blacklemon67, Bob K, Bwmodular, Cburnett, Ccwelt, Chodorkovskiy, Cindy141, Closetsingle, Commander Keane, Dicklyon, Diftil, Dino, Dlincoln, DonDiego, Dude1818, Dysprosia, ENeville, ElectronicsEnthusiast, Eloquence, Furrykef, GRAHAMUK, Gene Nygaard, Giftlite, Gopalkrishnan83, Grimey, Heron, Hfastedge, Hymek, Ike9898, Jamelan, Jeepday, Jorge Stolfi, Josh Cherry, JulesH, Kshieh, Kusma, Leon math, Lmatt, Loisel, Lowellian, Martin Kozák, Michael Hardy, Moxfyre, Nbarth, Nova77, Oli Filth, Patrick, Pawinsync, Phil Boswell, RTC, Rbj, RedWolf, Revolving Bugbear, RexNL, Roadrunner, Robert P. O'Shea, Sade, Shanes, Simiprof, SlipperyHippo, Smack, Snoyes, Srittau, Ssola, Stevegrossman, Stevertigo, Stillwaterising, Straker, TaintedMustard, The Anome, Tlotoxl, Tomic, Verijax, Wapcaplet, WhiteOak2006, WikiSlasher, Wolfkeeper, Wsmarz, Zoicon5, 132 anonymous edits

Quantization (signal processing) *Source*: http://en.wikipedia.org/w/index.php?oldid=482512763 *Contributors*: Aadri, Beetelbug, Bulser, C xong, Cat5nap, Cburnett, Charles Matthews, Chris the speller, Constant314, DeadTotoro, DmitriyV, DopefishJustin, Engelec, Fernando.iglesiasg, Giftlite, Glenn, Hadiyana, History2007, JordiTost, Kashyap.cherukuri, Kku, Krash, Longhair, Matt Gies, Metacomet, Michael Hardy, Mobeets, Nayuki, Nbarth, Necromancer44, Oarih, Omegatron, OverlordQ, Pak21, Paolo.dL, Petr.adamek, Pol098, Radagast83, Rbj, RedWolf, Sapphic, Shmlchr, Slady, Smack, SudoMonas, Sylenius, The Anome, The Seventh Taylor, Topbanana, WhiteHatLurker, 93 anonymous edits

Convolution *Source*: http://en.wikipedia.org/w/index.php?oldid=484111693 *Contributors*: 16@r, 1exec1, AhmedFasih, Aitias, Ajb, Akella, Almwi, Anonymous Dissident, Antandrus, Anthony Appleyard, Ap, Arcenciel, Armin Straub, Auclairde, AxelBoldt, Bdiscoe, Bdmy, Belizefan, Ben pcc, BenFrantzDale, Berland, BigHaz, BjarteSorensen, Bob K, Boing! said Zebedee, Bombshell, Bracchesimo, Btyner, CRGreathouse, Carolus m, CaseInPoint, Cdnc, Charles Matthews, Chato, Chowbok, Coco, Coffee2theorems, Comech, Conversion script, Crisófilax, Cronholm144, CryptoDerk, Cwkmail, Cyp, D1ma5ad, Daniele.tampieri, DaveWF, Dbfirs, Dicklyon, DirkOliverTheis, Dmmaus, Dmn, Dpbsmith, Drew335, Dysprosia, Dzordzm, Edward Z. Yang, Eleassar, Emc2, Emote, Esoth, Feraudyh, Flavio Guitian, Frozenport, Furrykef, GabrielEbner, Gareth Owen, Gene Ward Smith, Giftlite, Giuscarl, Greenec3, Greg newhouse, H2g2bob, Hankwang, Headbomb, Helder.wiki, Hellbus, HenningThielemann, Hippasus, Hurricane111, Hyacinth, IMSoP, Inductiveload, Intelliproject, InverseHypercube, JSKau, Jeff G., JethroElfman, Jevinsweval, Jschuler16, Jshadias, Justin545, Kallikanzarid, Karada, Kazkaskazkasako, Kelly Martin, Kenyon, Kielhorn, Kku, Konradek, Kurykh, Lamoidfl, Larvy, Lautarocarmona, Liotier, Lklundin, Loisel, Loodog, Looxix, Loren Rosen, Lovibond, Lwalt, MFH, Madmath789, MarcLevoy, Matteo.Bertini, Mckee, Mct mht, Melchoir, Michael Hardy, Mild Bill Hiccup, Minna Sora no Shita, Mircha, Mitrut, Miym, Mmendis, Modster, Mojodaddy, Mormegil, Mpatel, MrOllie, Mschlindwein, Mwilde, N.MacInnes, Naddy, Nbarth, Neffk, Ngchen, Nuwewsco, Odedee, Oleg Alexandrov, Oli Filth, Omegatron, Orderud, Ortho, Oski, P4lm0r3, PMajer, Patrick, Paul Rako, Perey, Phys, Pinethicket, Pion, PowerWill500, Rade Kutil, RapidR, Rlcook.unm, Robert K S, RobertStar20, Rogerbrent, Rubybrian, SD5, Schneelocke, Seabhcan, SebastianHelm, Shanes, Silly rabbit, Slawekb, Smack the donkey, Smithwn, Steve2011, Stevenj, SuburbaniteFury, Sullivan.t.j, Sławomir Biały, Tarquin, Tblackma222, TheObtuseAngleOfDoom, Thenub314, Tinos, Tomash, User A1, UtherSRG, Visionat, Vuo, Waqas946, Wayne Slam, Wikomidia, WojciechSwiderski, Wwoods, Wxm29, Xe7al, Yaris678, Zoltan808, Zvika, 303 anonymous edits

Ordinary differential equation *Source*: http://en.wikipedia.org/w/index.php?oldid=484567993 *Contributors*: 192.115.22.xxx, 48v, A. di M., Absurdburger, AdamSmithee, After Midnight, Ahadley, Ahoerstemeier, Alfy Alf, Alll, Amixyue, Andrei Polyanin, Anetode, Anita5192, Ap, Arthena, Arthur Rubin, BL, BMF81, Baccala@freesoft.org, Bemoeial, BenFrantzDale, Benjamin.friedrich, Berean Hunter, Bernhard Bauer, Beve, Bloodshedder, Bo Jacoby, Bogdangiusca, Bryan Derksen, Charles Matthews, Chenyu, Chilti, Chris in denmark, ChrisUK, Christian List, Ciro.santilli, Cloudmichael, Cmdrjameson, Cmprince, Conversion script, Cpuwhiz11, Cutler, Danger, Delaszk, Dicklyon, DiegoPG, Dmitrey, Dmr2, Dmytro, DominiqueNC, Dominus, Don4of4, Donludwig, Doradus, Dysprosia, Ed Poor, Ekotkie, Emperorbma, Enochlau, Fintor, Fruge, Fzix info, Gauge, Gene s, Gerbrant, Giftlite, Gombang, Graham87, HappyCamper, Heuwitt, Hongsichuan, Ht686rg90, Icairns, Isilanes, Iulianu, Jack in the box, Jak86, Jao, JeLuF, Jitse Niesen, Jni, JoanneB, John C PI, JohnBlackburne, Jokes Free4Me, JonMcLoone, Josevellezcaldas, Juansempere, Kawautar, Kdmckale, Klaas van Aarsen, Krakhan, Kwantus, L-H, LachlanA, Lethe, Linas, Lingwitt, Liquider, Lupo, MarkGallagher, Math.geek3.1415926, MathMartin, Mathuvw, Matusz, Melikamp, Michael Hardy, Mikez, Moskvax, MrOllie, Msh210, Mtness, Niteowlneils, Oleg Alexandrov, Patrick, Paul August, Paul Matthews, PaulTanenbaum, Pdenapo, PenguiN42, Phil Bastian, PizzaMargherita, Pm215, Poonamjadhav, Poor Yorick, Pt, Randomguess, Rasterfahrer, Raven in Orbit, Recentchanges, RedWolf, Rich Farmbrough, Rl, RobHar, Rogper, Romanm, Rpm, Ruakh, Salix alba, Sbyrnes321, Sekky, Shandris, Shirt58, SilverSurfer314, Sofia karampataki, Ssd, Starlight37, Stevertigo, Stw, Superlaza, Susvolans, Sverdrup, Tarquin, Tbsmith, Technopilgrim, Telso, Template namespace initialisation script, The Anome, Thenub314, Tobias Hoevekamp, TomyDuby, Tot12, TotientDragooned, Tristanreid, Twin Bird, Tyagi, Ulner, Vadimvadim, Waltpohl, Wclxlus, Whommighter, Wideofthemark, WriterHound, Xrchz, Yhkhoo, 今古庸龍, 186 anonymous edits

Laplace transform applied to differential equations *Source*: http://en.wikipedia.org/w/index.php?oldid=465766352 *Contributors*: Andrei Polyanin, Cjpuffin, Dysprosia, Giftlite, Jdclevenger, Jko, Lindberg G Williams Jr, Mariussimonsen, MaxEnt, Michael Hardy, Pglenn77, Sodin, Starx, Topbanana, VMS Mosaic, Wmahan, 30 anonymous edits

Fourier transform *Source*: http://en.wikipedia.org/w/index.php?oldid=483356505 *Contributors*: 9258fahsflkh917fas, A Doon, A. Pichler, Abecedare, Adam.stinchcombe, Admartch, Adoniscik, Ahoerstemeier, Akbg, Alejo2083, AliceNovak, Alipson, Amaher, AnAj, Andrei Polyanin, Andres, Angalla, Anna Lincoln, Ap, Army1987, Arondals, Asmeurer, Astronautameya, Avicennasis, AxelBoldt, Barak Sh, Bci2, Bdmy, BehzadAhmadi, BenFrantzDale, BigJohnHenry, Bo Jacoby, Bob K, Bobblewik, Bobo192, BorisG, Bugnot, Burhem, Butala, CSTAR, Caio2112, Cassandra B, Catslash, Cburnett, CecilWard, Ch mad, Charles Matthews, Chris the speller, ClickRick, Cmghim925, Complexica, Compsonheir, Coppertwig, CrisKatz, Crisófilax, DX-MON, Da nuke, DabMachine, DavidCBryant, Demosta, Dhabih, Discospinster, DmitTrix, Dmmaus, Dougweller, Dr.enh, DrBob, Drew335, Drilnoth, Dysprosia, EconoPhysicist, Ed g2s, Eliyak, Elkman, Enochlau, Epzcaw, Favonian, Feline Hymnic, Feraudyh, Fizyxnrd, Forbes72, Fr33kman, Fred Bradstadt, Fropuff, Futurebird, Gaius Cornelius, Gareth Owen, Geekdiva, Giftlite, Glenn, GuidoGer,

GyroMagician, H2g2bob, HappyCamper, Heimstern, HenningThielemann, Herr Lip, Hesam7, HirsuteSimia, Hrafeiro, Ht686rg90, I am a mushroom, Igny, Iihki, Ivan Shmakov, Iwfyita, Jaakobou, Jdorje, Jhealy, Jko, Joerite, JohnOFL, JohnQPedia, Joriki, Justwantedtofixonething, KHamsun, KYN, Keenan Pepper, Kevmitch, Kostmo, Kunaporn, Larsobrien, Linas, LokiClock, Looxix, Lovibond, Luciopaiva, Lupin, M1ss1ontomars2k4, Manik762007, MathKnight, Maxim, Mckee, Metacomet, Michael Hardy, Mikeblas, Mikiemike, Millerdl, Moxfyre, Mr. PIM, NTUDISP, Naddy, NathanHagen, Nbarth, Nihil, Nishantjr, Njerseyguy, Nk, Nmnogueira, NokMok, Od Mishehu, Offsure, Oleg Alexandrov, Oli Filth, Omegatron, Oreo Priest, Ouzel Ring, PAR, Pak21, Papa November, Paul August, Pedrito, Pete463251, Petergans, Phasmatisnox, Phils, PhotoBox, PigFlu Oink, Poincarecon, Pol098, PsiEpsilon, PtDw832, Publichealthguru, Quietbritishjim, Quintote, Qwfp, R.e.b., Rainwarrior, Rbj, Red Winged Duck, Riesz, Rifleman 82, Rijkbenik, Rjwilmsi, RobertHannah89, Rror, Rs2, Rurz2007, SKvalen, Safenner1, Sai2020, Sandb, Sbyrnes321, SchreiberBike, SebastianHelm, Sepia tone, Sgoder, Sgreddin, Shreevatsa, Silly rabbit, Slawekb, SlimDeli, Snigbrook, Snoyes, Sohale, Soulkeeper, SpaceFlight89, Spanglej, Sprocedato, Stevan White, Stevenj, Stpasha, StradivariusTV, Sunev, Sverdrup, Sylvestersteele, Sławomir Biały, THEN WHO WAS PHONE?, TYelliot, Tabletop, Tahome, TakuyaMurata, TarryWorst, Tetracube, The Thing That Should Not Be, Thenub314, Thermochap, Thinking of England, Tim Goodwyn, Tim Starling, Tinos, Tobias Bergemann, TranceThrust, Tunabex, Ujjalpatra, User A1, Vadik wiki, Vasil', Verdy p, VeryNewToThis, VictorAnyakin, Vidalian Tears, Vnb61, Voronwae, WLior, Wavelength, Wiki Edit Testing, WikiDao, Wile E. Heresiarch, Writer130, Wwheaton, Ybhatti, YouRang?, Zoz, Zvika, 547 anonymous edits

Harmonic oscillator *Source*: http://en.wikipedia.org/w/index.php?oldid=480913027 *Contributors*: 2help, A. di M., AManWithNoPlan, Abdullah Köroğlu, Abtract, Aclohesy, Arch dude, Audriusa, AugPi, BWHB, Benjamin.friedrich, Bh3u4m, Bobblewik, Breawycker, Brews ohare, Bryan Derksen, CYD, Caltas, Chetvorno, Chopchopwhitey, ChristophE, Chtirrell, Clemwang, Conversion script, Curtdbz, Cutler, Cyrius, Daano15, Dai bach, Dannown, Davehi1, David C Bailey, David s graff, Db099221, Dchristle, Dcirovic, Deklund, Delauxs, Derek Ross, Dhayes12, Dicklyon, Doraemonpaul, DrBob, Draeco, Drils, Eequor, El C, Elert, Enochlau, FF2010, Fermat618, Fresheneesz, Giftlite, Gilliam, Glenn, GliderMaven, Gonfer, Grafen, Hal4, HappyCamper, Headbomb, Helix84, Heron, HiraV, Hqb, Hu12, Hutcher, Ignatzmice, InverseHypercube, JabberWok, JacksonVance, Jimp, Jitse Niesen, Joeljkp, John, Jrdioko, Kaifeng, KasugaHuang, Keenan Pepper, Ken g6, Kieff, Kku, KrakatoaKatie, Lambiam, Laurascudder, Lee, Lseixas, Luke321321, MTC, Maatghandi, Marie Poise, Mat-C, MathKnight, Maurice Carbonaro, Mcapdevila, Mgummess, Michael Hardy, Mikiemike, Mlewis000, Murison, Muu-karhu, MystRivenExile, Nmnogueira, NorwegianBlue, NotALizard, Oleg Alexandrov, Osharia, Ozuma, Passw0rd, Patrick, PeteX, Phil Boswell, PhySusie, Pieter Kuiper, Pip2andahalf, Poeloq, PsiStar, R'n'B, RG2, Reddi, RekishiEJ, Roberta F., Ronhjones, Sanpaz, SeventyThree, Spike Wilbury, Spp143, Srbaroach, StarLight, Stevenj, StradivariusTV, Superdelocalizable, Suspekt, TStein, Tardis, Tesi1700, The Anome, The Cunctator, The Original Wildbear, TheEgyptian, Timb66, Tristanreid, Ultimaga, UncleDouggie, Van helsing, Vegard, Veledan, Voxon, Vql, Waveguy, Wewoodruff, WillowW, Wiwalsh00, Woodshed, Woohookitty, XJamRastafire, Yakeyglee, Yevgeny Kats, Youandme, Zoicon5, 184 anonymous edits

Image Sources, Licenses and Contributors

Image:Function x^2.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Function_x^2.svg *License*: Creative Commons Attribution-ShareAlike 3.0 Unported *Contributors*: Qualc1

Image:Function-x3.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Function-x3.svg *License*: unknown *Contributors*: -

Image:Function-x3plus1.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Function-x3plus1.svg *License*: unknown *Contributors*: -

Image:Signal Sampling.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Signal_Sampling.png *License*: Public Domain *Contributors*: Email4mobile (talk)

File:Samplerates.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Samplerates.svg *License*: Public Domain *Contributors*: Dick Lyon (original) ANDROBETA (vector)

Image:Bandlimited.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Bandlimited.svg *License*: Public Domain *Contributors*: User:Editor at Large, User:Rbj

Image:Sinc function (normalized).svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Sinc_function_(normalized).svg *License*: GNU Free Documentation License *Contributors*: Bender235, Juiced lemon, Krishnavedala, Omegatron, Pieter Kuiper

Image:NonoverlappedSpectrum.png *Source*: http://en.wikipedia.org/w/index.php?title=File:NonoverlappedSpectrum.png *License*: Public Domain *Contributors*: Deadstar, Rbj, 2 anonymous edits

Image:AliasedSpectrum.png *Source*: http://en.wikipedia.org/w/index.php?title=File:AliasedSpectrum.png *License*: Public Domain *Contributors*: Bdamokos, Bob K, Rbj, WikipediaMaster, 1 anonymous edits

Image:Moire pattern of bricks small.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Moire_pattern_of_bricks_small.jpg *License*: GNU Free Documentation License *Contributors*: Jesse Viviano, Maksim, Man vyi, Teofilo

Image:Moire pattern of bricks.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Moire_pattern_of_bricks.jpg *License*: GNU Free Documentation License *Contributors*: Jesse Viviano, Maksim, Man vyi, Mdd, Teofilo

Image:CriticalFrequencyAliasing.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:CriticalFrequencyAliasing.svg *License*: Public Domain *Contributors*: Qef

Image:ReconstructFilter.png *Source*: http://en.wikipedia.org/w/index.php?title=File:ReconstructFilter.png *License*: Public Domain *Contributors*: Pline, Rbj, 1 anonymous edits

File:Moire pattern of bricks.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Moire_pattern_of_bricks.jpg *License*: GNU Free Documentation License *Contributors*: Jesse Viviano, Maksim, Man vyi, Mdd, Teofilo

File:Moire pattern of bricks small.jpg *Source*: http://en.wikipedia.org/w/index.php?title=File:Moire_pattern_of_bricks_small.jpg *License*: GNU Free Documentation License *Contributors*: Jesse Viviano, Maksim, Man vyi, Teofilo

File:Aliasing a.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Aliasing_a.png *License*: Creative Commons Attribution-Sharealike 3.0,2.5,2.0,1.0 *Contributors*: Mwyann

File:AliasingSines.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:AliasingSines.svg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Moxfyre

File:Aliasing-folding.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Aliasing-folding.png *License*: Public Domain *Contributors*: Bob K

File:Aliasing.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Aliasing.gif *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: Bob K, Simiprof, 2 anonymous edits

Image:Sampled.signal.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Sampled.signal.svg *License*: Public Domain *Contributors*: User:Petr.adamek, User:Rbj

Image:Quantized.signal.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Quantized.signal.svg *License*: Public Domain *Contributors*: User:Petr.adamek, User:Rbj

Image:Digital.signal.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Digital.signal.svg *License*: Public Domain *Contributors*: User:Petr.adamek, User:Rbj

File:Convolution of box signal with itself2.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Convolution_of_box_signal_with_itself2.gif *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Convolution_of_box_signal_with_itself.gif: Brian Amberg derivative work: Tinos (talk)

File:Convolution of spiky function with box2.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Convolution_of_spiky_function_with_box2.gif *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Convolution_of_spiky_function_with_box.gif: Brian Amberg derivative work: Tinos (talk)

File:Convolution3.PNG *Source*: http://en.wikipedia.org/w/index.php?title=File:Convolution3.PNG *License*: Public Domain *Contributors*: Original uploader was Emote at en.wikipedia Later version(s) were uploaded by McLoaf at en.wikipedia.

File:Sunflower convolution.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Sunflower_convolution.png *License*: GNU Free Documentation License *Contributors*: Sunflower_sky_backdrop.jpg: Fir0002 derivative work: InverseHypercube

Image:Parabolic trajectory.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Parabolic_trajectory.svg *License*: Public Domain *Contributors*: Oleg Alexandrov

Image:Function ocsillating at 3 hertz.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Function_ocsillating_at_3_hertz.svg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Thenub314

Image:Onfreq.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Onfreq.svg *License*: GNU Free Documentation License *Contributors*: Original: Nicholas Longo, SVG conversion: DX-MON (Richard Mant)

Image:Offfreq.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Offfreq.svg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Thenub314

Image:Fourier transform of oscillating function.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Fourier_transform_of_oscillating_function.svg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Thenub314

File:Rectangular function.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Rectangular_function.svg *License*: GNU Free Documentation License *Contributors*: Axxgreazz, Bender235, Darapti, Omegatron

File:Sinc function (normalized).svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Sinc_function_(normalized).svg *License*: GNU Free Documentation License *Contributors*: Bender235, Juiced lemon, Krishnavedala, Omegatron, Pieter Kuiper

Image:Simple harmonic oscillator.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Simple_harmonic_oscillator.gif *License*: Public Domain *Contributors*: Oleg Alexandrov

Image:Simple harmonic motion animation.gif *Source*: http://en.wikipedia.org/w/index.php?title=File:Simple_harmonic_motion_animation.gif *License*: Public Domain *Contributors*: User:Evil_saltine

File:Damping 1.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Damping_1.svg *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: Nuno Nogueira (Nmnogueira). Original uploader was Nmnogueira at en.wikipedia. Later version(s) were uploaded by Dicklyon at en.wikipedia.

File:Oscillatory motion acceleration.ogv *Source*: http://en.wikipedia.org/w/index.php?title=File:Oscillatory_motion_acceleration.ogv *License*: Attribution *Contributors*: Pieter Kuiper, Linnæus University, Sweden

Image:Step response for two-pole feedback amplifier.PNG *Source*: http://en.wikipedia.org/w/index.php?title=File:Step_response_for_two-pole_feedback_amplifier.PNG *License*: GNU Free Documentation License *Contributors*: Brews ohare

File:Resonance.PNG *Source*: http://en.wikipedia.org/w/index.php?title=File:Resonance.PNG *License*: Public Domain *Contributors*: "MasterHD" Original uploader was MasterHD at en.wikipedia

Image:Harmonic oscillator gain.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Harmonic_oscillator_gain.svg *License*: Public Domain *Contributors*: Keenan Pepper

Image:Simple pendulum height.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Simple_pendulum_height.png *License*: Public Domain *Contributors*: User:Helix84

Image:Harmonic oscillator.svg *Source*: http://en.wikipedia.org/w/index.php?title=File:Harmonic_oscillator.svg *License*: Public Domain *Contributors*: Kieff

License