

AN INTRODUCTION TO THE Z-TRANSFORM AND ITS APPLICATION TO SAMPLED-DATA SYSTEMS

1. INTRODUCTION

This Data Item presents an introduction to the Z-transform and the basic principles relating to its application to linear sampled-data systems.

A linear sampled-data system is a system containing both continuous and sampled signals, where the sampled signals are known at only discrete, equally spaced intervals of time. The Z-transform, which is analogous to the Laplace transform used for entirely continuous systems, is the principal analytical tool used for the study of such systems. The Z-transform variable, z , is associated with time-shifting in the difference equations that describe the sampled signals and is analogous to the Laplace transform variable, s , which is associated with differentiation in the differential equations of a continuous system.

For systems in which the signals are in discrete form throughout, the use of difference equations is sufficient to specify completely the system dynamics. However, for systems containing a mixture of continuous and discrete signals, a model is needed for analytical purposes to represent the relation between the input and output variables of the system and the processes of digital-to-analogue, and analogue-to-digital, signal conversions.

This Item considers, as a preliminary to the Z-transform, the representation of the sampling process, *i.e.* the relationship of a continuous analogue signal to a sequence of pulses equally spaced in time, each of which provides the value of the original signal only at the particular sampling instants. The relation of this pulse sequence to the Z-transform is given together with its properties. This Z-transform is also shown to be related to the Laplace transform with the use of an impulse sequence and the relationships between poles and zeros in the complex s -plane to those in the complex z -plane are provided. The necessary relationship between the sampled input and output of a system element, the pulse transfer function, is found with the use of the Z-transform and this is complemented by a description of the algebraic manipulation required to obtain the pulse transfer function for cascaded elements of a system in both open and closed loop form.

For design purposes, it is useful to obtain information in the interval between sampling instants and two such methods that achieve this, the Modified Z-transform and multirate sampling, are described.

A necessary element of all sampled data systems is the filter which is used to regain, as far as is possible, the original signal from the pulse sequence. This element is known as a data hold and is usually based on a polynomial extrapolation. The simplest form of data hold, one that is commonly employed in sampled systems, is the zero-order polynomial hold and the transfer function of this element is derived. Finally it is shown how the discrete time function may be obtained from the Z-transform of a system using the inverse Z-transform. Three alternative methods are presented.

2. NOTATION

a, b	general constants
C_n	coefficients of a Fourier series
$D(s)$	denominator of a Laplace transform
$D'(s)$	first derivative of D with respect to s
$\bar{f}(i\omega), \bar{f}^*(i\omega)$	Fourier transforms of $f(t)$ and $f^*(t)$, respectively
$\bar{f}_\gamma^*(i\omega)$	Fourier transform of $f^*(t)$
$\bar{f}(s), \bar{f}^*(s)$	Laplace transforms of $f(t)$ and $f^*(t)$, respectively
$F(z)$	Z-transform of $f^*(t)$
$Fo\{f(t)\}$	Fourier transform of $f(t)$, $= \bar{f}(i\omega)$
$f(t)$	a continuous arbitrary function of t
$f^*(t)$	an impulse sampled function of t
$f_0^*(t)$	an ideal sampled function of t
$\dot{f}_\gamma^*(t)$	a sampled function of t obtained from a sampler of width γ
$f(nT)$	$= f_0^*(t)$ for $t = nT, n = 0, 1, 2, \dots$
$G(z)$	pulse transfer function of an element
$\bar{g}(s)$	Laplace transfer function of an element
$\bar{h}_0(i\omega)$	Laplace transform of zero-order polynomial data hold
K	control system gain
$L\{f(t)\}$	Laplace transform of $f(t)$, $= \bar{f}(s)$
$L^{-1}\{\bar{f}(s)\}$	inverse Laplace transform of $\bar{f}(s)$, $= f(t)$
$N(s)$	numerator of a Laplace transform
n, m	general integer variables
$p(t)$	an impulse train
$p_\gamma(t)$	a unit amplitude pulse train of width γ
s	Laplace transform variable
T	sampling interval

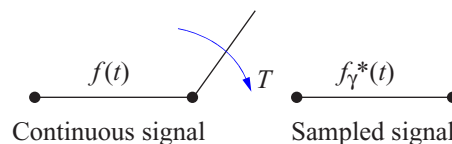
t	time
x_k	roots of characteristic equation, $D(s) = 0$, where D is of order p
z	Z-transform variable
$Z\{f(t)\}$	Z-transform of $f(t)$, $= F(z)$
$Z^{-1}\{F(z)\}$	inverse Z-transform of $F(z)$
α	angle between the real axis and a line from the origin to a complex pole on an Argand diagram
γ	pulse width
Δ	negative delay or time advance
$\delta_T(T)$	unit area impulse train
ζ	damping ratio, $= \cos(\alpha)$
μ	real part of complex number
ω, ω_N	general arbitrary frequency and natural frequency, respectively
ω_S	sampling frequency, $= 2\pi/T$

3. THE SAMPLING PROCESS

As a preliminary step the nature of sampled signals and the sampling process are discussed.

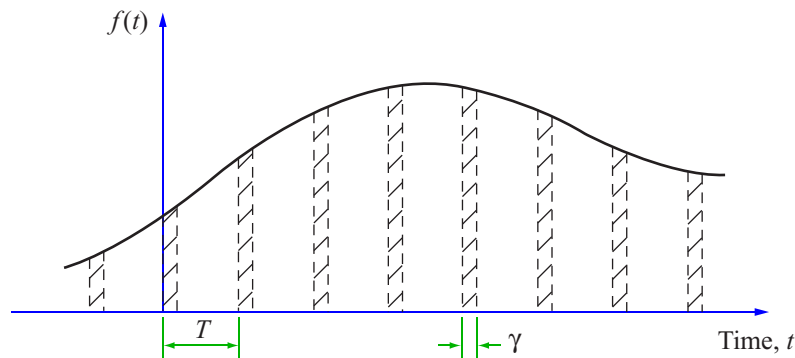
The sampling process is the means by which a continuous signal is converted into a sequence of pulses or numbers, where the signal is available only at the times when the sampling is made. In the majority of practical situations the sampling intervals between pulses are equal and only this case will be considered in this Item.

The sampling process is schematically represented by Sketch 3.1 in which a continuous signal, $f(t)$, is sampled every T seconds by opening and closing a sampling switch or sampler.



Sketch 3.1 Sampling process

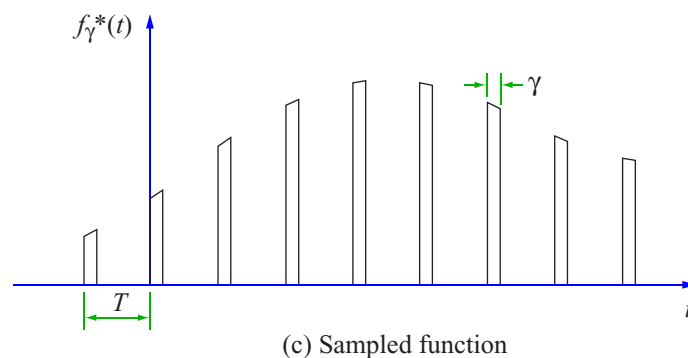
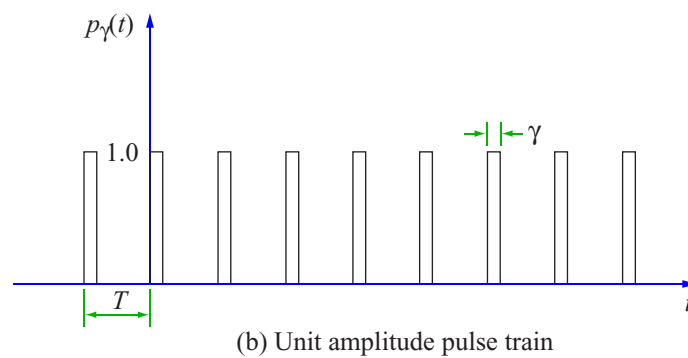
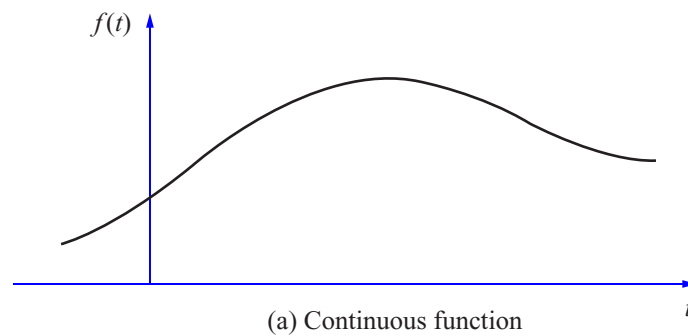
The output from the sampler, $f_\gamma^*(t)$, is available as pulses, of width γ , only at times during which the switch is closed and is diagrammatically represented by the shaded areas of Sketch 3.2, where the tops of the pulses are formed by the appropriate section of the curve $f(t)$. The variation of the pulse signal, $f_\gamma^*(t)$, varies as $f(t)$ during the time interval nT to $nT + \gamma$, where n is the set of integers 0, 1, 2, ... The equivalent digital value of this signal is then the number occurring at the discrete times $nT + \gamma$.



Sketch 3.2 Sampled pulses

Sampling may be regarded as an amplitude modulation process, depicted in Sketch 3.3 in which a pulse train, $p_\gamma(t)$, of unit amplitude pulses at intervals T is modulated by a continuous time function, $f(t)$, to produce a sampled function, $f_\gamma^*(t)$. This may be expressed mathematically as

$$f_\gamma^*(t) = f(t) p_\gamma(t). \quad (3.1)$$



Sketch 3.3 Amplitude modulation process

The unit amplitude pulse train, $p_\gamma(t)$, is a periodic function, since T is constant, which can be represented by a Fourier series of the form

$$p_\gamma(t) = \sum_{n=-\infty}^{\infty} C_n \exp\{in\omega_S t\}, \quad (3.2)$$

where the sampling frequency, ω_S , is $2\pi/T$ and the Fourier coefficients, C_n , are given by

$$C_n = \frac{1}{T} \int_0^T p(t) \exp\{-in\omega_S t\} dt = \frac{\gamma}{T} \frac{\sin(n\omega_S \gamma/2)}{n\omega_S \gamma/2} \exp\{-in\omega_S \gamma/2\}. \quad (3.3)$$

It is important to note that $0 < \gamma < T$ in any real physical system.

Using Equation (3.2) in Equation (3.1), the sampled function may be written as

$$\bar{f}_\gamma^*(t) = f(t) \sum_{n=-\infty}^{\infty} C_n \exp\{in\omega_S t\}. \quad (3.4)$$

If the Fourier transform of the continuous function, $f(t)$, is $\bar{f}(i\omega)$, the Fourier transform of the sampled function, $f_\gamma^*(t)$, is

$$f_\gamma^*(i\omega) = Fo\{f_\gamma^*(t)\} = \int_{-\infty}^{\infty} f_\gamma^*(t) \exp\{-i\omega t\} dt, \quad (3.5)$$

or, using Equation (3.4)

$$\bar{f}_\gamma^*(i\omega) = \int_{-\infty}^{\infty} f(t) \sum_{n=-\infty}^{\infty} C_n \exp\{in\omega_S t\} \exp\{-i\omega t\} dt. \quad (3.6)$$

By the use of the frequency shifting theorem (see Reference 1, page 37) which states that

$$Fo\{\exp(at) f(t)\} = \int_{-\infty}^{\infty} f(t) \exp\{at\} \exp\{-i\omega t\} dt = \bar{f}(i\omega - a), \quad (3.7)$$

the sampled function of Equation (3.6) may be rewritten as

$$\bar{f}_\gamma^*(i\omega) = \sum_{n=-\infty}^{\infty} C_n \bar{f}(i\omega - in\omega_S). \quad (3.8)$$

Sketch 3.4(a) shows the amplitude spectrum for the chosen continuous function

$$\begin{aligned} f(t) &= e^{-0.5t}, & t &\geq 0 \\ &= 0, & t &< 0, \end{aligned}$$

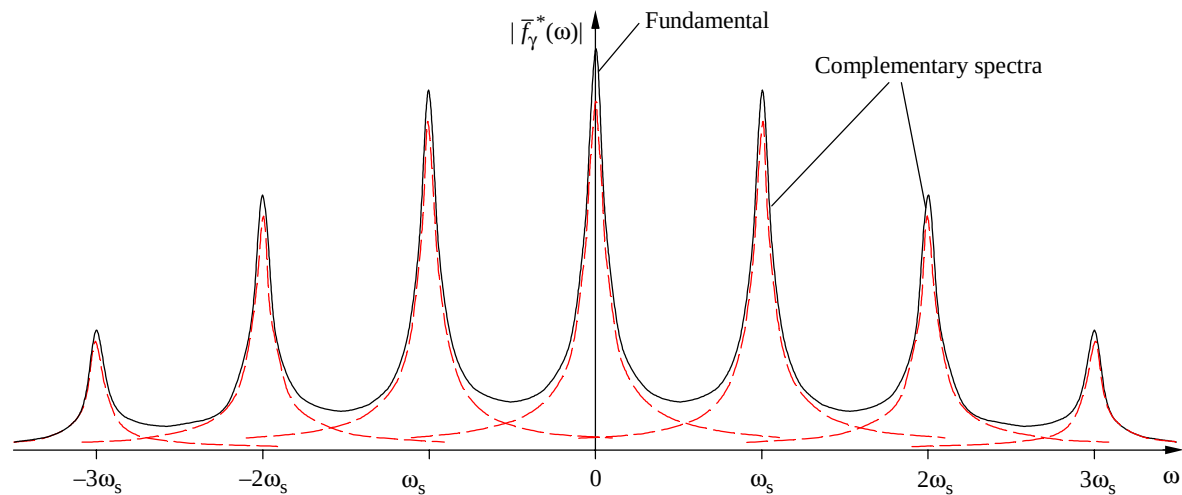
and Sketch 3.4(b) shows the corresponding spectrum for the sampled function, $f_\gamma^*(t)$, where the chosen values of sample width, γ , and sampling period, T , are 0.13 and 0.5, respectively.

The sampling process, as depicted by Equation (3.8), produces a fundamental spectrum similar in shape to the spectrum of the continuous signal but also produces an infinite succession of unwanted complementary spectra which are shifted by a separation frequency of ω_S and weighted by a factor of the coefficients, C_n . These complementary spectra are known as aliases and result from the Fourier representation of a square wave which has an infinite number of terms, each term representing a harmonic of higher frequency.

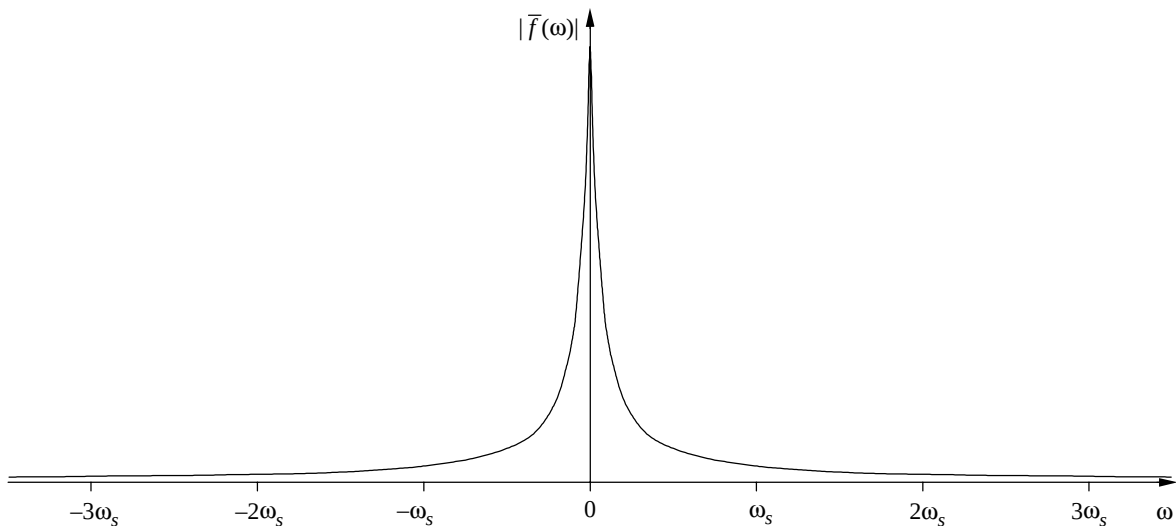
It can be seen in Sketch 3.4 that adjacent complementary spectra overlap both with each other and with the fundamental and in such a case it would not be possible to extract the complete information contained in

the spectrum of the continuous function from the spectrum of the sampled function. There are two features that contribute to this overlap of adjacent spectra. The first involves the sampling interval, T , which when reduced (equivalent to an increase of the sampling frequency, ω_s , since $\omega_s = 2\pi/T$) will tend to separate each adjacent complementary spectrum both from one another and from the fundamental. The second concerns the frequency content of the original signal, the higher the range of frequencies present in the original signal the higher will be the overlap of complementary spectra for a given sampling period.

It thus follows that it would never be possible to recover completely the original signal unless the spectrum is band-limited, *i.e.* it contains no components above a certain frequency.



b. Pulse sampled function



a. Continuous function

Sketch 3.4 Amplitude spectrum of a continuous function and a pulse sampled function

These considerations have been formalised in a theorem by Shannon (see Reference 9) that states that for signals having a finite bandwidth, *i.e.* signals that have frequency components up to, but not exceeding, a frequency ω_B rad/s, then it is theoretically possible to recover completely the original signal using a

sampling interval equal to, or less than, (π/ω_B) , i.e.

$$\omega_S > 2\omega_B. \quad (3.9)$$

To recover the original signal from the sampled pulses it is necessary to extract the fundamental spectrum from the infinite sequence of complementary spectra using a low pass filter. If it were possible to construct an ideal filter which would produce zero attenuation in the frequency range $-\omega_B$ to ω_B with perfect cut-off outside this band, then complete reconstruction of the original signal would be possible. However, practical filters do not have perfect response characteristics and continuous signals do not, in general, have a finite (band-limited) spectrum and this gives rise to a certain amount of distortion in the recovered signal. Consequently, it is rare in practice to work near the limit as given by Shannon's theorem. It is common practice to filter an analogue signal prior to sampling to produce a band-limited signal. Even with this pre-sampling filter it is usual to sample a signal at five to ten times the highest frequency known to be present.

3.1 Ideal Sampling

It is generally true of digital systems, which provide pulse sequences of information only at times when the sampler is closed, that the time interval during which a sample is taken, γ , is very small compared with the periodic sampling interval, T . Thus the assumption that a pulse has an infinitely small width, $\gamma \rightarrow 0$, represents a good approximation to the physical sampling process. Ideal sampling consequently produces a sequence of numbers, analogue-to-digital conversion, which represent the original signal at the discrete times, nT , given by

$$f_0^*(t) = \lim_{\gamma \rightarrow 0} \{f_\gamma^*(t)\} = f(nT), \quad n = 0, 1, 2, \dots \quad (3.10)$$

4. THE Z-TRANSFORM

As already mentioned in the Introduction a linear sampled-data system may contain both continuous and discrete signals. Analysis of such systems can be effected through the use of the Z-transform. A natural preliminary, therefore, of such an analysis is to define a Z-transform, to set out its algebraic properties and to consider its relation to the Laplace transform.

Consider a continuous function, $f(t)$, which is assumed to be zero for negative time. The Z-transform of $f(t)$, $F(z)$, is defined by

$$Z\{f(t)\} = F(z) = \sum_{n=0}^{\infty} f(nT)z^{-n}, \quad (4.1)$$

where for convergence $|z| < 1$. The lower case symbol z is the complex Z-transform variable, and the upper case symbol Z , used as $Z\{f(t)\}$, denotes the Z-transform of $f(t)$.

This is the fundamental equation for the Z-transform of a time function, and for its calculation the sampled function, $f^*(t)$, is used, since the continuous signal, $f(t)$, is undefined except at the sampling instant.

Example

As an example of the use of Equation (4.1) consider the following continuous time function

$$f(t) = \frac{K}{ab} + \frac{K}{a-b} \left\{ \frac{e^{-at}}{a} - \frac{e^{-bt}}{b} \right\}, \quad (4.2)$$

for which the corresponding Z-transform is required. The Z-transform is found to be

$$F(z) = \sum_{n=0}^{\infty} \left(\frac{K}{ab} + \frac{K}{a-b} \left(\frac{e^{-ant}}{a} - \frac{e^{-bnT}}{b} \right) \right) z^{-n},$$

or

$$F(z) = \frac{K}{ab} \left(1 + z^{-1} + z^{-2} + \dots \right) + \frac{K}{(a-b)a} \left(1 + e^{-aT} z^{-1} + e^{-2aT} z^{-2} + \dots \right) - \frac{K}{(a-b)b} \left(1 + e^{-bT} z^{-1} + e^{-2bT} z^{-2} + \dots \right),$$

i.e.

$$F(z) = \frac{K}{ab} \left\{ \frac{z}{z-1} \right\} + \frac{K}{(a-b)} \left\{ \frac{z}{a(z-e^{-aT})} - \frac{z}{b(z-e^{-bT})} \right\}. \quad (4.3)$$

4.1 Properties of Z-Transforms

The following properties and theorems of the Z-transform are presented without proof. Reference 3 may be consulted for the proofs of a number of the properties or theorems detailed. In all cases, the following notation will be used,

$$Z\{f(t)\} = F(z).$$

(a) Linearity

$$\begin{aligned} Z\{a_1 f_1(t) + a_2 f_2(t)\} &= a_1 Z\{f_1(t)\} + a_2 Z\{f_2(t)\}, \\ &= a_1 F_1(z) + a_2 F_2(z), \end{aligned} \quad (4.4)$$

where a_1 and a_2 are constants and the functions, $f_1(t)$ and $f_2(t)$, are assumed Z-transformable.

(b) Shift property

$$Z\{f(t + mT)\} = z^m \left\{ F(z) - \sum_{k=0}^{m-1} f(kT) z^{-k} \right\}, \quad (4.5)$$

where m is a constant. In particular, for $m = 1$, the following result is obtained

$$Z\{f(t + T)\} = z\{F(z) - f(0)\}. \quad (4.6)$$

(c) Complex scale change

$$Z\{e^{-at} f(t)\} = F(e^{aT} z). \quad (4.7)$$

(d) Finite summation

$$Z \left\{ \sum_{k=0}^n f(kT) \right\} = \frac{z}{z-1} F(z), \quad |z| > 1. \quad (4.8)$$

(e) Initial value

$$f(0) = \lim_{z \rightarrow \infty} \{F(z)\}. \quad (4.9)$$

(f) Final value

$$\lim_{n \rightarrow \infty} \{f(nT)\} = \lim_{z \rightarrow 1} \{(z-1)F(z)\}. \quad (4.10)$$

(g) Convolution (complex multiplication)

$$F_1(z)F_2(z) = Z \left\{ \sum_{k=0}^n f_1(kT) f_2(n-kT) \right\}. \quad (4.11)$$

(h) Multiplication by t (complex differentiation)

$$Z\{t f(t)\} = -T z \frac{d}{dz} \{F(z)\}. \quad (4.12)$$

4.2 Relationship to the Laplace Transform

As already mentioned the Z-transform analysis of sampled-data systems is analogous to the Laplace transform analysis of continuous systems where the symbol z is the analogue of the symbol s . To develop a relationship between the Z-transform and the Laplace transform yet another mathematical representation or idealisation of the sampling process, separate to that of Section 3.1, is required.

If the function $f^*(t)$ is defined as the sequence of rectangular pulses of area $f(nT)$, this sequence can be replaced by an impulse modulation process given by

$$f^*(t) = \sum_{n=0}^{\infty} f(nT) \delta(t-nT) = \sum_{n=0}^{\infty} f(nT) \delta(t-nT), \quad (4.13)$$

where $\delta(t-nT)$, the Dirac-delta function, can be considered as the limit of a pulse of unit area. The sampler can be regarded as

$$p(t) = \delta_T(t) = \sum_{n=-\infty}^{\infty} \delta(t-nT), \quad (4.14)$$

and is an impulse train which is illustrated in Sketch 4.1.