Signals, systems, and Z transforms

F. A. Ramponi
Rev. 0.2, 2019-11-03

0.1 Discrete-time signals and systems

0.1.1 Signals

A discrete-time signal is a function $u : \mathbb{Z} \to \mathbb{C}$ (complex-valued) or $u : \mathbb{Z} \to \mathbb{R}$ (real-valued), depending on the context. Sometimes we deal with *vector-valued* signals, e.g. $u : \mathbb{Z} \to \mathbb{C}^n$, but in what follows the substance doesn't change, for a vector-valued signal can always be interpreted also as a n-tuple of scalar signals (viz. a "column of scalar signals"). For the sake of clarity, we may denote a signal u more explicitly as $u(\cdot)$; but the reader is invited to think at the function u as a sequence of numbers infinite in both directions, $\ldots, u(-2), u(-1), u(0), u(1), u(2), \ldots$; hence, with sloppy notation, we may denote it also $\{u(t)\}$, it being understood that the variable t takes values in $\mathbb{Z}$.

The set $\mathcal{S}(\mathbb{Z})$ (or $\mathcal{S}$ for short) of all discrete-time signals, i.e. sequences of numbers, equipped with the component-wise sum and the component-wise multiplication by a scalar α

$$(u_1 + u_2)(t) = u_1(t) + u_2(t)$$
$$(\alpha u_2)(t) = \alpha u(t),$$

is of course a vector space.

The signal $\{u(t)\}$ is called *bounded* if there exists a constant $K_u \in \mathbb{R}$ such that

$$|u(t)| \leq K_u \quad \text{for all } t \in \mathbb{Z}. \tag{1}$$

and this fact is denoted $u \in \ell^\infty(\mathbb{Z})$, or $u \in \ell^\infty$ for short. Indeed ℓ^∞ is defined as the set of all the bounded sequences in $\mathbb{Z}$; endowed with the component-wise sum and multiplication by a scalar, it becomes a vector space, or more precisely a subspace of $\mathcal{S}$: indeed it is straightforward to show that if two sequences u_1 and u_2 are bounded then the sequence $\alpha u_1 + \beta u_2$ is also bounded for any two scalars α, β. More than that, $\ell^\infty(\mathbb{Z})$ is a *normed* vector space, where the norm of a sequence u is $\|u\|_\infty =$ the minimum number $K_u \in \mathbb{R}$ such that (1) is true[1].

The signal $\{u(t)\}$ is called *summable* if

$$\sum_{t=-\infty}^{+\infty} |u(t)| = \lim_{t\to+\infty} \sum_{t=-T}^{T} |u(t)| = M < +\infty; \tag{2}$$

this fact is denoted $u \in \ell^1(\mathbb{Z})$, or $u \in \ell^1$ for short. Indeed ℓ^1 is defined as the set of all the summable sequences in $\mathbb{Z}$; endowed with the component-wise sum and multiplication by a scalar it becomes a vector space; and it becomes a normed vector space if we define the norm of a sequence u to be the number $\|u\|_1 := M$ appearing in (2).[2]

Any summable signal is also bounded. In other terms,

$$\ell^1(\mathbb{Z}) \subset \ell^\infty(\mathbb{Z}) \subset \mathcal{S};$$

[1] Even more: it is a *complete* normed vector space, or a so-called *Banach space*.

[2] The vector space $\ell^1(\mathbb{Z})$ is also a Banach space.

and in fact $\ell^1(\mathbb{Z})$ is a proper subspace of $\ell^\infty(\mathbb{Z})$, which in turn is a proper subspace of $\mathcal{S}(\mathbb{Z})$.

The summable (and of course bounded) signal

$$\delta(t) = \begin{cases} 1, & \text{if } t = 0, \\ 0, & \text{otherwise,} \end{cases}$$

is called the *impulse*, or the "discrete-time delta".

The *convolution* of two signals $\{u(t)\}$, $\{v(t)\}$ is the signal $\{u * v(t)\}$ defined as follows:

$$u * v(t) := \sum_{\tau=-\infty}^{+\infty} u(t-\tau)v(\tau) = \lim_{T\to\infty} \sum_{\tau=-T}^{T} u(t-\tau)v(\tau); \qquad (3)$$

the above expression is well-defined only if the series converges. The convolution '$*$' is an *operation* between sequences: it maps a pair of sequences into another sequence; as an operation, it has both the associative property ($(u * v) * w = u * (v * w)$) and the commutative property ($u * v = v * u$). Moreover, it possesses an *identity element*, which is precisely the impulse:

$$\delta * u(t) = u * \delta(t) = \sum_{\tau=-\infty}^{+\infty} u(t-\tau)\delta(\tau) = \sum_{\tau=0} u(t-\tau) = u(t) \quad \text{for all } t,$$

so that $\delta * u = u * \delta = u$.

0.1.2 Systems

In full generality, a *discrete-time system* is a map φ from a vector space $\mathcal{U}$ of discrete-time signals, called the *input* set, to another vector space $\mathcal{Y}$ of discrete-time signals, called the *output* set; to fix ideas, let $\mathcal{U} = \mathcal{Y} = \mathcal{S}$ and $\varphi : \mathcal{S} \to \mathcal{S}$. To any sequence $\{u(t)\} \in \mathcal{S}$, φ associates another sequence $\{y(t)\} \in \mathcal{S}$:

$$\varphi : \{\cdots, u(-1), u(0), u(1), \cdots\} \mapsto \{\cdots, y(-1), y(0), y(1), \cdots\}$$

The definition of some systems impose that the output sample $y(t)$ is a function of the sole input sample $u(t)$, and possibly of the time t, for all t:

$$y(t) = f(t, u(t)).$$

These systems are called *instantaneous transformations*. Despite the fact that they are useful models for a lot of phenomena considered in control engineering and telecommunications (e.g. quantizers, saturations, "dead zones", etc.), they are not much interesting from our point of view.

0.1.3 Linearity and time invariance

In the cases that we are going to consider, the output sample $y(t)$ is instead a function of the whole input signal $\{u(t)\}$, or of part of it, and possibly of the time t; for instance, with loose notation,

$$y(t) = f(t, \{\cdots, u(t-1), u(t), u(t+1), \cdots\}) \quad \text{("time-varying, non-causal system")},$$
$$y(t) = f(\{\cdots, u(t-1), u(t)\}) \quad \text{("time-invariant causal system")}.$$

These are called *dynamical* systems (from the Greek word *dynamis*, *force*), because they are suitable models of physical systems subject to forces, accelerations and so on.

Among these, *linear dynamical systems* are of paramount importance in practically every branch of science and engineering. Recall that linearity means that the *superposition principle* holds: if

$$\varphi : \{u_1(t)\} \mapsto \{y_1(t)\} \qquad \text{and} \qquad \varphi : \{u_2(t)\} \mapsto \{y_2(t)\},$$

then for any constants a and b,

$$\varphi : \{au_1(t) + bu_2(t)\} \mapsto \{ay_1(t) + by_2(t)\}$$

("to the sum of the causes corresponds the sum of the effects"). The system is called *time-invariant* if to the translation in time of an input corresponds the translation in time of the output, with the same time lag $\tau \in \mathbb{Z}$: that is, if

$$\varphi : \{u(t)\} \mapsto \{y(t)\}$$

then for any $\tau \in \mathbb{Z}$

$$\varphi : \{u(t+\tau)\} \mapsto \{y(t+\tau)\}$$

A time-invariant system has the following property: the output $\{y(t)\}$ corresponding to a certain input $\{u(t)\}$ is the convolution between $\{u(t)\}$ and the output $\{w(t)\} = \varphi\left[\{\delta(t)\}\right]$ that corresponds to the impulse $\{\delta(t)\}$. Indeed, with loose notation,

$$\begin{aligned} y(t) = \varphi\left[\{u(t)\}\right] = \varphi\left[\{\delta * u(t)\}\right] = \varphi\left[\left\{\sum_{\tau=-\infty}^{+\infty} \delta(t-\tau)u(\tau)\right\}\right] \\ = \sum_{\tau=-\infty}^{+\infty} \varphi\left[\{\delta(t-\tau)\}\right] u(\tau) = \sum_{\tau=-\infty}^{+\infty} w(t-\tau)u(\tau) = w * u(t), \end{aligned} \tag{4}$$

where the fourth equality is an application of linearity that deliberately ignores convergence details (the sum is infinite!), and the fifth one is due to time-invariance. The sequence $\{w(t)\}$ is called the *impulse response* of the system. In what follows we will always refer to discrete-time systems that are both linear and time-invariant, and we will call them LTI systems for short.

0.1.4 Stability

An LTI system is called *externally stable*, or *BIBO-stable*, if any bounded input signal is mapped to an output signal which is also bounded[3]. More precisely, φ is BIBO-stable if $\varphi[u] \in \ell^\infty$ for all $u \in \ell^\infty$. It is easy to show that if the impulse response is summable ($w \in \ell^1$), then the system is BIBO-stable[4]. Indeed, suppose that $\{w(t)\}$ is such that $\sum_{t=-\infty}^{+\infty} |w(t)| = M < \infty$, that $|u(t)| \leq K_u$ for all t, and that $\{y(t)\} = \varphi[\{u(t)\}]$; then

$$\begin{aligned} |y(t)| = \left| \sum_{\tau=-\infty}^{+\infty} u(\tau)w(t-\tau) \right| \leq \sum_{\tau=-\infty}^{+\infty} |u(\tau)|\ |w(t-\tau)| \\ \leq K_u \sum_{\tau=-\infty}^{+\infty} |w(t-\tau)| = K_u M := K_y \end{aligned}$$

for all t, so that $\{y(t)\}$ is a bounded signal as well.

[3] BIBO stands for Bounded Input $\Rightarrow$ Bounded Output.

[4] The converse is also true, but details are omitted here.

0.1.5 Causality

An LTI system is called *causal* if whenever two input signals satisfy

$$u_1(\tau) = u_2(\tau), \quad \tau = \cdots, t-2, t-1, t;$$

the corresponding output signals satisfy

$$y_1(\tau) = y_2(\tau), \quad \tau = \cdots, t-2, t-1, t.$$

This means that the output $y(t)$ at a given time t depends on the past samples of the input signal $\{u(\tau)\}, \tau = \cdots, t-2, t-1, t$, but not on its future samples $\{u(\tau)\}, \tau = t+1, t+2, \cdots$.

This definition implies, in particular, that the output corresponding to an input signal $\{u(t)\}$ such that $u(\tau) = 0$ tor all $\tau < t_0$ also satisfies $y(\tau) = 0$ for all $\tau < t_0$. Indeed if $u(\tau) = 0$ for $\tau < t_0$ then also $u(\tau) = -u(\tau)$ for $\tau < t_0$; then, by causality and linearity, $y(\tau) = -y(\tau)$ for $\tau < t_0$, i.e. $y(\tau) = 0$ for $\tau < t_0$.

Thus, if the input "starts" at t_0, so does the output. In view of time-invariance, it is common practice to let $t_0 = 0$, and to call *causal* also those signals that "start at 0". Thus, *a causal LTI system has a causal impulse response* $\{w(t)\}$, because the impulse is causal in the first place ($\delta(t) = 0$ for all $t < 0$). The response of a causal system to an arbitrary input $\{u(t)\}$ is

$$y(t) = w * u(t) = \sum_{\tau=-\infty}^{+\infty} w(t-\tau)u(\tau) = \sum_{\tau=-\infty}^{t} w(t-\tau)u(\tau), \tag{5}$$

because $w(t-\tau) = 0$ for $\tau > t$. Here $y(t)$ is still well defined if and only if the series converges. But if, moreover, the input signal $\{u(t)\}$ is also causal, which is typically the case for signals and systems considered in deterministic control theory, then

$$y(t) = \sum_{\tau=-\infty}^{t} w(t-\tau)u(\tau) = \sum_{\tau=0}^{t} w(t-\tau)u(\tau), \tag{6}$$

because $u(\tau) = 0$ for $\tau < 0$. Differently from (3), (4), and (5), the convolution (6) is *always* well-defined for all t, because the sum is finite. It goes without saying that $y(t) = 0$ for all $t < 0$, so that the output signal is also causal.

0.2 Transforms and transfer functions

0.2.1 Fourier transforms

The Fourier transform of a signal $\{u(t)\}_{-\infty}^{+\infty}$ is the power series:

$$\begin{aligned} \hat{U}(\omega) = \mathcal{F}\left[\{u(t)\}\right](\omega) &:= \sum_{t=-\infty}^{+\infty} u(t)e^{-j\omega t} \\ &= \lim_{T\to\infty} \sum_{t=-T}^{T} u(t)e^{-j\omega t}, \end{aligned} \tag{7}$$

where $\omega \in [-\pi, \pi]$. Such series may very well not converge, and the Fourier transform may not exist; a sufficient condition for the existence of $\hat{U}(\omega)$ for all ω is that $\{u(t)\}$ is summable: if $\sum_{t=-\infty}^{+\infty} |u(t)| < \infty$, then

$$\sum_{t=-\infty}^{+\infty} |u(t)e^{-j\omega t}| = \sum_{t=-\infty}^{+\infty} |u(t)|\ |e^{-j\omega t}| = \sum_{t=-\infty}^{+\infty} |u(t)|,$$

and (7) converges at all ω, since it converges also absolutely.

A Fourier transform, even that of a real signal, is in general a complex function of ω. It can therefore be expressed as $\hat{U}(\omega) = |\hat{U}(\omega)|\ e^{j\angle\hat{U}(\omega)}$. However, if $\{u(t)\}$ is real its transform enjoys the following property,

$$\begin{aligned}\hat{U}(-\omega) &= \sum_{t=-\infty}^{+\infty} u(t)e^{j\omega t} = \sum_{t=-\infty}^{+\infty} \overline{u(t)}\ \overline{e^{-j\omega t}} \\ &= \overline{\sum_{t=-\infty}^{+\infty} u(t)e^{-j\omega t}} = \overline{\hat{U}(\omega)},\end{aligned}$$

called *Hermitian symmetry.* It follows at once that

$$\begin{aligned}|\hat{U}(-\omega)| &= |\hat{U}(\omega)|; \\ \angle\hat{U}(-\omega) &= -\angle\hat{U}(\omega).\end{aligned} \tag{8}$$

In words, the absolute value of the Fourier transform of a real signal is an even function, and its phase is an odd one. Also, it is immediate to show that its real part is even, and its imaginary part is odd.

0.2.2 $\mathcal{Z}$-transforms

The so-called $\mathcal{Z}$-transform of a signal $\{u(t)\}$ is the power series:

$$U(z) = \mathcal{Z}\left[\{u(t)\}\right](z) := \sum_{t=-\infty}^{+\infty} u(t)z^{-t}, \tag{9}$$

where $z \in \mathbb{C}$. As happens for the Fourier transform, the $\mathcal{Z}$-transform may not converge for any $z \in \mathbb{C}$ (take for example the sequence $u(t) = |t|!$); if, however, $\{u(t)\}$ is summable, then (9) converges at least on the *unit circle* $\{z \in \mathbb{C} \text{ s.t. } |z| = 1\} = \{e^{j\omega} \mid \omega \in [-\pi, \pi]\}$, and there it coincides with the Fourier transform, i.e. $\hat{U}(\omega) = U\left(e^{j\omega}\right)$.

0.2.3 Unilateral $\mathcal{Z}$-transforms

The $\mathcal{Z}$-transform of a *causal* signal $\{u(t)\}$ reads

$$U(z) = \mathcal{Z}\left[\{u(t)\}\right](z) = \sum_{t=0}^{+\infty} u(t)z^{-t}. \tag{10}$$

In deterministic control theory systems are usually assumed to be causal (because is causal in the first place any realistic physical system: the effects always "start after" the causes, so the output $y(t)$ of a physical system at time t cannot depend on future samples of the input $u(t+1), u(t+2), u(t+3), \ldots$). So, impulse responses are causal. It is also common practice to start the analysis of the behavior of systems always at a finite time t_0, and in view of time-invariance it is customary to set $t_0 = 0$. Furthermore, the values of inputs and outputs before $t_0 = 0$ are often irrelevant to the analysis.

To make the long story short, in control theory it is customary to assume that *all* the signals that come into play are causal (inputs, outputs, impulse responses, state trajectories, and so on), and to assume that (10), always starting from $t = 0$, is "the" $\mathcal{Z}$-transform. At least, it is the most used version of the $\mathcal{Z}$-transform in this branch of engineering, in the same way as $\int_0^{+\infty} f(t)e^{-st}dt$, starting from 0, is the common version of the Laplace transform. We will call it *unilateral*.

The $\mathcal{Z}$-transform of the impulse response of an LTI causal system is called the *transfer function* of that system. Its Fourier transform is also called sometimes the transfer function or, depending on the context, the *frequency response* of the system.

0.3 Some properties of the unilateral $\mathcal{Z}$-transform

0.3.1 Convergence region

If the series (10) converges for a certain $\bar{z} \in \mathbb{C}$, then it converges *absolutely* for all $z \in \mathbb{C}$ such that $|z| > |\bar{z}|$. Indeed, if $\sum_{t=0}^{+\infty} u(t)\bar{z}^{-t}$ converges, then the sequence $\{u(t)\bar{z}^{-t}\}$ must be bounded, that is $|u(t)\bar{z}^{-t}| \leq K$ for all t. But then, for all $|z| > |\bar{z}|$,

$$\sum_{t=0}^{+\infty} |u(t)z^{-t}| = \sum_{t=0}^{+\infty} |u(t)\bar{z}^{-t}| \left|\frac{z^{-t}}{\bar{z}^{-t}}\right| \leq K \sum_{t=0}^{+\infty} \left|\frac{\bar{z}}{z}\right|^t = \frac{K}{1 - |\bar{z}/z|} < \infty.$$

Hence, either the series does not converge for any $z \in \mathbb{C}$ (example: $u(t) = t!$), or it converges at least on an open region outside a disc, i.e. on a set of the form $\{z \in \mathbb{C} \text{ s.t. } |z| > R\}$. The minimum R for which this happens is called *convergence radius*. If, in particular, $\{u(t)\}$ is summable, then $R < 1$, the convergence region includes the unit circle and the Fourier transform can be recovered as $\hat{U}(\omega) = U\left(e^{j\omega}\right)$.

0.3.2 Linearity

The operator $\mathcal{Z}$ that maps sequences to transforms is linear: if $\{u_1(t)\}$ and $\{u_2(t)\}$ are signals, and a_1, a_2 are real constants, then

$$\begin{aligned}
\mathcal{Z}\left[\{a_1u_1(t) + a_2u_2(t)\}\right](z) &= \sum_{t=-\infty}^{+\infty} (a_1u_1(t) + a_2u_2(t))z^{-t} \\
&= a_1 \sum_{t=-\infty}^{+\infty} u_1(t)z^{-t} + a_2 \sum_{t=-\infty}^{+\infty} u_2(t)z^{-t} \\
&= a_1U_1(z) + a_2U_2(z),
\end{aligned}$$

provided that both $U_1(z)$ and $U_2(z)$ exist for some $z \in \mathbb{C}$; the transform of the linear combination exists at all such z. More generally,

$$\mathcal{Z}\left[\left\{\sum_{\tau=0}^{T} a_\tau u_\tau(t)\right\}\right](z) = \sum_{\tau=0}^{T} a_\tau U_\tau(z)$$

for all $z \in \mathbb{C}$ such that $U_\tau(z)$ exists for $\tau = 0, \cdots, T$, that is at least in the intersection of the convergence regions of $U_1(z), \ldots, U_T(z)$.

0.3.3 Transform of a convolution

The transform of a convolution is the product of the respective transforms. Consider for example the input-output relation of a causal LTI system with impulse response $\{w(t)\}$. To a causal input signal $\{u(t)\}$ corresponds an output $y(t) = w * u(t)$ which is also causal. It holds:

$$\begin{aligned}
y(t) &= \sum_{\tau=0}^{t} w(t-\tau)u(\tau) \\
Y(z) &= \sum_{t=0}^{+\infty}\left(\sum_{\tau=0}^{t} w(t-\tau)u(\tau)\right) z^{-t} = \sum_{t=0}^{+\infty}\left(\sum_{\tau=0}^{+\infty} w(t-\tau)u(\tau)\right) z^{-t} \\
&= \sum_{\tau=0}^{+\infty}\sum_{t=0}^{+\infty} w(t-\tau)u(\tau)z^{-(t-\tau+\tau)} = \sum_{\tau=0}^{+\infty} u(\tau)z^{-\tau}\sum_{t=0}^{+\infty} w(t-\tau)z^{-(t-\tau)};
\end{aligned}$$

since $\{w(t)\}$ is causal all the terms in the inner sum for which $t-\tau<0$ vanish, hence the inner sum starts from τ, and

$$\begin{aligned}
Y(z) &= \sum_{\tau=0}^{+\infty} u(\tau)z^{-\tau}\sum_{t=\tau}^{+\infty} w(t-\tau)z^{-(t-\tau)} = \left(\sum_{\tau=0}^{+\infty} u(\tau)z^{-\tau}\right)\left(\sum_{t'=0}^{+\infty} w(t')z^{-t'}\right) \\
&= W(z)U(z).
\end{aligned} \tag{11}$$

0.3.4 Delay

If a signal is delayed by one time step its transform gets multiplied by z^{-1}. Let $\{u(t)\}$ be a causal signal and let $\{\bar{u}(t)\}$ be its *delayed* version, defined by $\bar{u}(t) := u(t-1)$ for all t. Then:

$$\begin{aligned}
\bar{U}(z) &= \sum_{t=0}^{+\infty} \bar{u}(t)z^{-t} = \sum_{t=0}^{+\infty} u(t-1)z^{-t} = z^{-1}\sum_{t=0}^{+\infty} u(t-1)z^{-(t-1)} \\
&= z^{-1}\sum_{t'=0}^{+\infty} u(t')z^{-(t')} = z^{-1}U(z),
\end{aligned}$$

where the fourth equality holds because $u(-1) = 0$. This fact tells us that, despite being a complex number in the original definition, z^{-1} can be interpreted as a *delay operator* acting on $\mathcal{Z}$-transforms. With a slight abuse of notation, we will write "z^{-1}" to denote a delay also when dealing with sequences, e.g. $u(t-1) = z^{-1}u(t)$.

0.3.5 Anticipation

Let $\{u(t)\}$ be a causal signal and let $\{\bar{u}(t)\}$ be its *anticipated* version, defined by $\bar{u}(t) := u(t+1)$ for all t. Then:

$$\begin{aligned}
\bar{U}(z) &= \sum_{t=0}^{+\infty} \bar{u}(t)z^{-t} = \sum_{t=0}^{+\infty} u(t+1)z^{-(t+1)+1} = z\sum_{t=0}^{+\infty} u(t+1)z^{-(t+1)} \\
&= z\sum_{t'=1}^{+\infty} u(t')z^{-t'} = z\left(U(z) - u(0)\right).
\end{aligned}$$

Note the asymmetry between delay and anticipation (the term $u(0)$): this is the only price we have to pay for having chosen the transform to be *unilateral*.

In some respects, if $u(0)$ is not relevant or if it can disregarded, one can loosely interpret z as an *anticipation operator* acting on sequences, in the same sense as z^{-1} is the delay operator; but strictly speaking they are not the inverses of each other (if we adopt the full-fledged *bilateral* $\mathcal{Z}$-transform (9), instead, they are). This is the discrete-time counterpart of the customary interpretation, in continuous-time models, of the complex variable s of Laplace transforms as a representative of the *derivative* operator and of $1/s$ as the representative of integration.

0.3.6 Some useful transforms

In this section we suppose that all signals are causal. $\mathbb{1}(t)$ will denote the unit step:

$$\mathbb{1}(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

Here follows a table of unilateral $\mathcal{Z}$-transforms:

signal	**transform**	**convergence region**
$\delta(t)$	1	$\mathbb{C}$
$\mathbb{1}(t)$	$\frac{z}{z-1}$	$\|z\| > 1$
$\mathbb{1}(t) \cdot t$	$\frac{z}{(z-1)^2}$	$\|z\| > 1$
$\mathbb{1}(t) \cdot t^2$	$\frac{z(z-1)}{(z-1)^3}$	$\|z\| > 1$
$\mathbb{1}(t) \cdot a^t$	$\frac{z}{(z-a)}$	$\|z\| > \|a\|$
$\mathbb{1}(t) \cdot \cos(\omega t)$	$\frac{z(z-\cos\omega)}{z^2-2z\cos\omega+1}$	$\|z\| > 1$
$\mathbb{1}(t) \cdot \sin(\omega t)$	$\frac{z\sin\omega}{z^2-2z\cos\omega+1}$	$\|z\| > 1$

Example. The transform of $u(t) = \mathbb{1}(t)$ is

$$U(z) = \sum_{t=0}^{+\infty} z^{-t} = \sum_{t=0}^{+\infty} \left(z^{-1}\right)^t = \frac{1}{1 - z^{-1}} = \frac{z}{z-1},$$

where of course the series converges if $|z^{-1}| < 1$, i.e. $|z| > 1$; this is rigorous enough. Here's instead a true example of "sporting club" mathematics. The transform of $u(t) = \mathbb{1}(t) \cdot t$ is

$$\begin{aligned} U(z) &= \sum_{t=0}^{+\infty} t z^{-t} \qquad (\text{let } \zeta = z^{-1}) \\ &= \sum_{t=0}^{+\infty} t\zeta^t = \zeta \cdot \sum_{t=0}^{+\infty} t\zeta^{t-1} = \zeta \cdot \sum_{t=0}^{+\infty} \frac{\partial \zeta^t}{\partial \zeta} = \zeta \cdot \frac{\partial}{\partial \zeta} \sum_{t=0}^{+\infty} \zeta^t = \zeta \cdot \frac{\partial}{\partial \zeta} \frac{1}{1-\zeta} \\ &= \zeta \cdot \frac{1}{(1-\zeta)^2} = \frac{z^{-1}}{(1-z^{-1})^2} = \frac{z}{(z-1)^2}. \end{aligned}$$

0.3.7 Final value theorem

I provide this result without proof:

Theorem 0.3.1 *Let $\{u(t)\}$ be a signal and $U(z)$ be its $\mathcal{Z}$-transform; suppose that $\lim_{t\to+\infty} u(t) = L$ exists and that $(u(t) - L) \in O(a^t)$ for some $a \in (0, 1)$. Then it holds*

$$L = \lim_{t\to+\infty} u(t) = \lim_{z\to 1} (z-1)U(z), \tag{12}$$

implying in particular that the second limit also exists.

For example, let's check the limit of a causal signal whose transform is a rational transfer function:

$$\begin{aligned} W(z) &= \frac{b_0 z^n + b_1 z^{n-1} + \ldots + b_{n-1} z + b_0}{z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_0} \\ &= b_0 \frac{(z - \bar{z}_1)(z - \bar{z}_2) \ldots (z - \bar{z}_n)}{(z - \bar{p}_1)(z - \bar{p}_2) \ldots (z - \bar{p}_n)}. \end{aligned}$$

If the roots $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_n$ of the denominator belong to the open unit circle ($|\bar{p}_i| < 1$) then, as we shall see, this is the transfer function of an asymptotically stable system. It is also the transform of its (causal) impulse response $\{w(t)\}$. We know that the modes of the system are decaying exponentials or damped oscillations, so the limit L exists and $w(t) - L$ decays exponentially. Then

$$\lim_{t \to +\infty} w(t) = \lim_{z \to 1} (z-1) W(z) = 0.$$

The response of the same system to the unit step $u(t) = \mathbb{1}(t)$ is instead some causal signal $\{y(t)\}$. It holds

$$\lim_{t \to +\infty} y(t) = \lim_{z \to 1} (z-1) W(z) U(z) = \lim_{z \to 1} (z-1) W(z) \frac{z}{z-1} = W(1).$$

0.4 The response to harmonic signals

0.4.1 Non-causal systems

Consider a BIBO-stable system, not necessarily causal. Its impulse response $\{w(t)\}$ must be summable ($\sum_{t=-\infty}^{+\infty} |w(t)| < \infty$), hence its frequency response $\hat{W}(\omega) = W\left(e^{j\omega}\right)$ must exist for every $\omega \in [-\pi, \pi]$.
A fundamental property of such a systems is that its response to *harmonic* signals (i.e. sinusoids, or sums of sinusoids) are also harmonic. To see this, consider first the response $\{y(t)\}_{-\infty}^{+\infty}$ to a complex input $\{u(t)\}_{-\infty}^{+\infty}$ of the form $u(t) = e^{j\omega t}$:

$$\begin{aligned} y(t) = w * u(t) &= \sum_{\tau=-\infty}^{+\infty} w(t-\tau) e^{j\omega\tau} \\ &= \sum_{\tau=-\infty}^{+\infty} w(t-\tau) e^{-j\omega(t-\tau)} e^{j\omega t} \\ &= e^{j\omega t} \sum_{\tau'=-\infty}^{+\infty} w(\tau') e^{-j\omega\tau'} = \hat{W}(\omega)\ e^{j\omega t}. \end{aligned}$$

This property is too important to let it pass without re-stating it in the proper, magnificent, *linear-algebraic* language. A BIBO-stable LTI system is a linear operator φ mapping sequences to sequences. For all the sequences $u(t) = e^{j\omega t}$ it holds, with loose notation,

$$\varphi\left[e^{j\omega t}\right] = \hat{W}(\omega)\ e^{j\omega t};$$

and here is the statement:

any harmonic sequence of the form $e^{j\omega t}$ is an eigenvector (or "eigenfunction") of φ, having $\hat{W}(\omega)$ as the corresponding eigenvalue.

Since the property holds for any $\omega \in [-\pi, \pi]$, it also does for $-\omega$; if $u(t) = e^{-j\omega t} = e^{j(-\omega)t}$, then:

$$y(t) = \hat{W}(-\omega) e^{-j\omega t}.$$

Now let $\{u(t)\}_{-\infty}^{+\infty}$ be a sinusoidal signal with frequency ω:

$$\begin{aligned} u(t) &= A \cos(\omega t + \varphi) \\ &= A \frac{e^{j(\omega t + \varphi)} + e^{-j(\omega t + \varphi)}}{2} = \frac{A}{2} e^{j\varphi} e^{j\omega t} + \frac{A}{2} e^{-j\varphi} e^{-j\omega t}; \end{aligned}$$

in view of linearity, the corresponding response is

$$
\begin{aligned}
y(t) &= \frac{A}{2} e^{j\varphi}\ \hat{W}(\omega)\ e^{j\omega t} + \frac{A}{2} e^{-j\varphi}\ \hat{W}(-\omega)\ e^{-j\omega t} \\
&= |\hat{W}(\omega)|\ \frac{A}{2} \left(e^{j\varphi}\ e^{j\angle \hat{W}(\omega)}\ e^{j\omega t} + A e^{-j\varphi}\ e^{-j\angle \hat{W}(\omega)}\ e^{-j\omega t} \right) \\
&= |\hat{W}(\omega)|\ A\ \frac{e^{j(\omega t + \varphi + \angle \hat{W}(\omega))} + e^{-j(\omega t + \varphi + \angle \hat{W}(\omega))}}{2} \\
&= |\hat{W}(\omega)|\ A \cos(\omega t + \varphi + \angle \hat{W}(\omega)).
\end{aligned}
$$

Thus, to a sinusoidal signal with frequency ω (non-causal, i.e. infinite in both directions), a BIBO-stable system responds with the same sinusoidal signal, amplified by $|\hat{W}(\omega)|$ and anticipated by $\angle \hat{W}(\omega)$.

0.4.2 Causal systems

What happens if the system is causal and the sinusoid is causal too, i.e. it "starts at 0"? Consider now a truncated exponential signal,

$$
u(t) = \mathbb{1}(t) \cdot e^{j\omega t}
$$

It holds

$$
\begin{aligned}
y(t) = w * u(t) &= \sum_{\tau=0}^{t} w(t-\tau) e^{j\omega\tau} = \sum_{\tau=0}^{t} w(t-\tau) e^{-j\omega(t-\tau)} e^{j\omega t} \\
&= e^{j\omega t} \sum_{\tau'=0}^{t} w(\tau') e^{-j\omega\tau'} = e^{j\omega t} \left(\hat{W}(\omega) - \sum_{\tau'=t+1}^{+\infty} w(\tau') e^{-j\omega\tau'} \right);
\end{aligned}
$$

note that since $\hat{W}(\omega) = \sum_{\tau'=0}^{+\infty} w(\tau') e^{-j\omega\tau'}$ must exist finite, the infinite sum within parentheses must tend to 0 as $t \to \infty$; hence

$$
y(t) = \hat{W}(\omega) e^{j\omega t} + s(t),
$$

where $s(t)$ is a "transient" term, which tends to 0 as $t \to \infty$. The same reasoning can be done for the truncated version of $u(t) = e^{-j\omega t}$, hence the response to a sinusoidal signal "starting at 0"

$$
u(t) = \begin{cases} 0, & t < 0, \\ A\cos(\omega t + \varphi), & t \geq 0, \end{cases}
$$

is, by similar computations to the above ones,

$$
y(t) = \begin{cases} 0, & t < 0; \\ |\hat{W}(\omega)|\ A\cos(\omega t + \varphi + \angle \hat{W}(\omega)) + \bar{s}(t), & t \geq 0, \end{cases}
$$

where $\bar{s}(t)$ is another transient term.

In conclusion, now we have four distinct interpretations for the frequency response of a LTI BIBO-stable system:

1. $\hat{W}(\omega)$ is the Fourier transform of the impulse response of the system; provided that the transfer function $W(z)$ converges in a region of the complex plane that includes the unit circle, the frequency response is $\hat{W}(\omega) = W\left(e^{j\omega}\right)$;

2. provided that the $\mathcal{Z}$-transforms of both the input and the output of the system converge in a region of the complex plane including the unit circle (this happens if they are summable), in that region the transfer function is a proportionality factor linking them:

$$Y(z) = W(z)U(z),$$

and similarly

$$\hat{Y}(\omega) = \hat{W}(\omega)\hat{U}(\omega);$$

3. the values $\hat{W}(\omega)$ are the eigenvalues of the system φ, corresponding to the "eigenfunctions" $e^{j\omega t}$;

4. the response of the system to a sinusoid with frequency ω is the same sinusoid, amplified by the modulus $|\hat{W}(\omega)|$ of the frequency response and anticipated by its phase $\angle\hat{W}(\omega)$; if the system is causal and the sinusoid is fed at the input only starting from a certain time, the response "starts" at that time and approaches the amplified and anticipated sinusoid after a transient.

0.5 Difference equations

The causal LTI systems that are used in practice to model filters, sampled version of continuous-time systems etc., are usually denoted by so-called *difference equations*. These are equalities written in one of the following equivalent forms, depending on which is more convenient for ease of notation:

$$\begin{aligned}
a_0y(t) + a_1y(t-1) + \cdots + a_ny(t-n) &= b_0u(t) + b_1u(t-1) + \cdots + b_mu(t-m) \\
a_0y(t) - a_1y(t-1) - \cdots - a_ny(t-n) &= b_0u(t) + b_1u(t-1) + \cdots + b_mu(t-m),
\end{aligned}$$

where $m \leq n$ and $a_0 \neq 0$. Without loss of generality, it is also customary to divide everything by a_0, obtaining a model where the first coefficient is 1; this being convenient for our purposes, we will work with the following model:

$$y(t) - a_1y(t-1) - \cdots - a_ny(t-n) = b_0u(t) + b_1u(t-1) + \cdots + b_mu(t-m). \tag{13}$$

The linear model (13) "represents" a causal system if the system imposes that (13) holds at all times t. In particular, given a *causal* input, the corresponding output can be defined recursively:

$$\begin{aligned}
&y(t) = 0 \quad \text{for all } t < 0 \text{ by assumption (causality);} \\
&y(t) = a_1y(t-1) + \cdots + a_ny(t-n) + b_0u(t) + \cdots + b_mu(t-m), \quad t \geq 0.
\end{aligned}$$

Here, $y(t)$ is well defined because we *assume* that the system is causal; hence $y(t)$ is a function of "past" samples $y(t-1), \cdots, y(t-n), u(t), \cdots, u(t-m)$, and not of future ones; but *nowhere* does an equation like (13) imply that the represented system is indeed *causal*. In other words, (13) is *not* "the system": it is a *property* of the system, and causality is another; the equation per se, taken alone, could very well represent a non-causal system.

Supposing that the output $\{y(t)\}$ possesses a $\mathcal{Z}$-transform as well as the input, $\{u(t)\}$, let transform both sides:

$$\begin{aligned}
Y(z) - a_1z^{-1}Y(z) - \cdots - a_nz^{-n}Y(z) &= b_0U(z) + b_1z^{-1}U(z) + \cdots + b_mz^{-m}U(z), \\
(1 - a_1z^{-1} - \cdots - a_nz^{-n})Y(z) &= (b_0 + b_1z^{-1} + \cdots + b_mz^{-m})U(z), \\
Y(z) &= \frac{b_0 + b_1z^{-1} + \cdots + b_mz^{-m}}{1 - a_1z^{-1} - \cdots - a_nz^{-n}}\, U(z).
\end{aligned} \tag{14}$$

The function

$$\bar{W}(z) = \frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{1 - a_1 z^{-1} - \cdots - a_n z^{-n}} = \frac{b_0 z^n + b_1 z^{n-1} + \cdots + b_m z^{n-m}}{z^n - a_1 z^{n-1} - \cdots - a_n} \tag{15}$$

exists in the region where both the transforms $U(z)$ and $Y(z)$ converge (the open region outside a disc), and comparing with (11), it must coincide there with the transfer function $W(z)$ of the system. It can be rewritten as follows:

$$\bar{W}(z) = \frac{B(z)}{A(z)} = \frac{b_0(z - z_1)(z - z_2)\cdots(z - z_n)}{(z - p_1)(z - p_2)\cdots(z - p_n)}, \tag{16}$$

The (complex) roots $p_1, \cdots, p_n$ of the polynomial $A(z) = z^n - a_1 z^{n-1} - \cdots - a_n$ are called the *poles* of $\bar{W}(z) = W(z)$, and the roots $z_1, \cdots, z_n$ of the polynomial $B(z) = b_0 z^n + b_1 z^{n-1} + \cdots + b_m z^{n-m}$ are called its *zeros*. It so happens that the region in which $W(z)$ converges is the set $\{z \in \mathbb{C} \mid |z| > |p_i|\}$, where p_i is the pole with maximum modulus; in other words, the maximum modulus among poles is the convergence radius. Consequently, the system is BIBO-stable if and only if the poles of $\bar{W}(z)$ belong to the open unit disc, that is, if all the poles p_i satisfy $|p_i| < 1$.
Again, the expression of $W(z)$ as a function of a complex variable is a *property* of the system, and it is not sufficient to describe it completely unless causality is assumed apart. Indeed, it is causality that dictates the region where $\bar{W}(z)$ is defined (and where $W(z)$ converges)[5].

Models like (13) are called *finite-dimensional*, because they can be realized with a recursive algorithm using only a finite amount of memory. In view of the reasoning above, it is obvious that any causal LTI system satisfying (13) admits a *rational* transfer function. However,

- it is not at all true that every LTI causal system can be described by a finite-dimensional model. This is rather intuitive, because the true representative of an LTI system is its impulse response, an infinite sequence which cannot, in general, be reconstructed by an algorithm with a finite amount of memory. Stated another way, the impulse response may very well have a non-rational $\mathcal{Z}$-transform. Anyway, finite-dimensional models are of paramount importance in engineering, because they are far handier than other models for computation, estimation, prediction, identification, and closed-loop control. Moreover, the transfer function of any BIBO-stable system can be approximated by a rational one with arbitrary accuracy.

- to any finite-dimensional model there corresponds a unique rational transfer function, but the converse is false. Consider indeed $\bar{W}(z)$ written in the form (16): zero/pole cancellations may happen, and they correspond to a "hidden" dynamics of the system that either is not affected by the input, or is not visible at the output, or both. The proper way to understand such dynamics is through state-space system theory; however, the point here is that adding, so to say, "a zero and a pole at the same arbitrary position", two at a time, one leaves the transfer function unchanged, but obtaining larger and larger models (13): hence, to the same transfer function there correspond infinitely many models.

[5] All of these concepts have their counterpart in the world of *causal* continuous-time systems, which you should remember from control courses: the difference equation (13) corresponds to an ordinary linear differential equation with constant coefficients; the left- and right-hand sides of the differential equation can be transformed according to Laplace, and the rational function that one obtains dividing the right-hand polynomial by the left-hand one happens to be the transfer function of the system. Such transfer function exists on a right half-plane *assuming that the system is causal*; and the continuous-time system is BIBO stable if and only if the poles of its transfer function lie in the open left-hand plane having the imaginary axis as its boundary. Indeed the unit circle plays for discrete-time systems the role that the imaginary axis plays for continuous-time ones, i.e. a sort of frontier between stable systems and unstable ones.