- This document is provided without warranty.

- I apologizes for any spelling and grammatical errors.

- The figures have not been changed and present themselves as in the original document.

- Suggestions and comments are wellcome.

IPB-ESTiG
INSTITUTO POLITÉCNICO DE BRAGANÇA
ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO

DIGITAL CONTROL

LECTURE NOTES

João Paulo Coelho

2005/2006

Joao Paulo Coelho
Polytechnic Institute of Bragança
School of Technology and Management
5700 Bragança - Portugal
Web: www.ipb.pt/ ~jpcoelho
E-Mail: jpcoelho@ipb.pt

Index

Chapter

0

Introduction and Objectives

IN 1859, Charles Darwin published his theory on the evolution of species, according to which, the phenotype changes of organisms was due to slow changes of the medium were those organisms live. In other words, when the performance criteria changes, the species tend to physically change in order to adapt to these new conditions. In ecological terms this phenomenon describes a feedback loop between a species and the environment around it.

The regulation by feedback is not exclusive to biological systems. In fact, feedback control is the basic mechanism by which systems, whether mechanical, electrical or biological, maintains its balance. The control actions taken in this context are based on the difference between the desired state and current system state, i.e. the adaptation is made according to the error.

This curricular unit deals with the particularity that, in large part of the control loop, the information is conveyed by an electric signal (analog or digital). This control strategy is used almost everywhere in man-made machinery. The following diagram present the fundamental building blocks of a closed loop digital control system [3] .

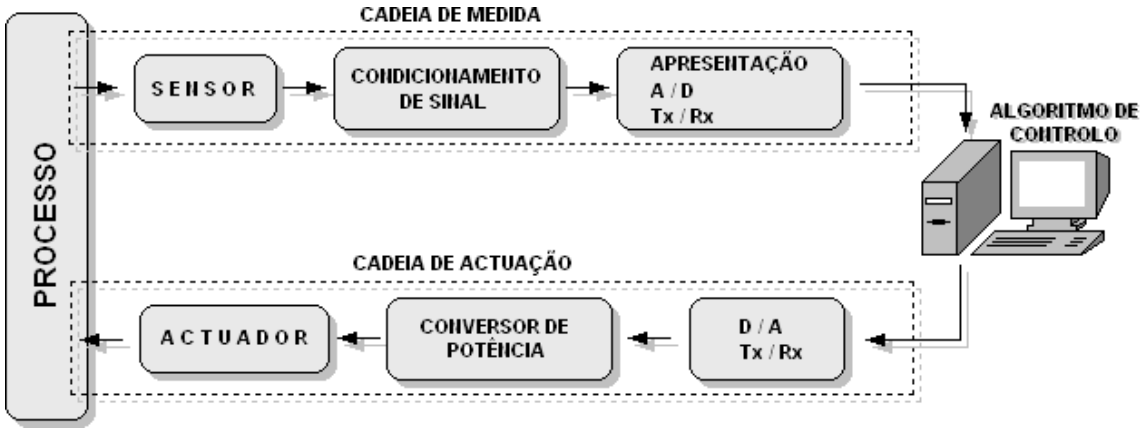


Fig 1. Block diagram of a feedback control system.

It is important to note that every block in this architecture has it proper dynamics. The control algorithm must deal with the overall dynamic behaviour equivalent to cascading all the referred elements. However some of them have almost no impact on the global system dynamic and others may be crucial.

Looking again to figure 1 one can identify three main components. Namely,

- Measurement chain.
- Actuation chain.
- Control Algorithm.

The measuring chain is the subsystem responsible for the acquisition of the control variable. This block in built around a sensor element. In turn, the sensor translate variations of the control variable into variation of any electric property such voltage, impedance or frequency.

The acquired signal, properly conditioned, will be responsible for providing information regarding the present system state. This state is compared with the desired one, resulting in a control signal supplied to the actuation chain. The command order, after being adapted, will be used to excite some type of actuator. The actuator performs the opposite role of the sensor: convert a signal, usually electrical, in another non-electrical. Examples of electric actuators are the electric motor, in which electrical current is converted into rotation energy, or pneumatic cylinders, driven by a solenoid valve, where electrical current is converted, indirectly, into axial displacement.

The block that takes the system state and provides the command signal is called the compensator or controller. In an electrical perspective, this controller may be analog or digital (note that, in general, the controller can be of different nature such as mechanical, pneumatic or hydraulic). Regardless of the controller nature its operation mode is the similar: to perform algebraic operations between signals. For analog controller, mathematical operations are performed using, for example, adders, integrators and differentiators designed around operational amplifiers. On the other hand, in the case of digital controllers, the calculations are carried out by logic gates (more specifically microprocessors).

The operations complexity carried out by the controller can range from a simple hysteresis comparison (on/off control) to a more elaborate control strategy such as the three-term control (Proportional-Integral-Derivative). This last technique will be reviewed in Chapter 1.

In the digital controller field, since it's easier to implement more elaborated numerical computation routines, more advanced control strategies can be found. An example of this is a technique known by adaptive control were the controller adapts to changes in the process dynamics or disturbance. The plant dynamic changes are sensed and the controller degrees-of-freedom are adjusted accordingly. Figures 2 and 3 represent both a simple on/off control strategy and an adaptive one [1].

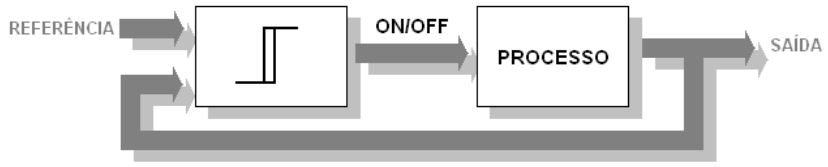


Fig 2. Block diagram of a control system on / off type

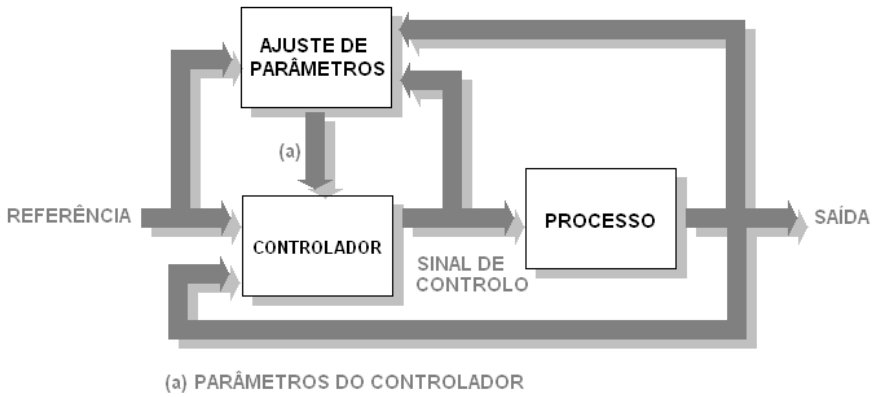


Fig 3. Block diagram of an adaptive control system

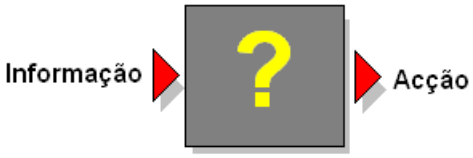
At present time, due to the proliferation and low-cost of digital computer systems and given the enormous benefits associated with it, control systems based on numeric processors such as DSPs, microcontrollers or microprocessors dismissed, almost completely, analog controllers of their functions. For this reason it's important the study of digital control theory on modern technological courses.

0.1 Digital Control: What is it about?

The objective of a control system is to force a system to keep, as close as possible, to the reference state despite possible system disturbances. In order to do this, the controller generates a signal obtained by algebraic manipulation of the system state signal.

In this context, the control design strategy is based on the answer to the following question:

How to establish the relationship between control actions and system information?



Until the nineteenth century, the control systems design involved only empirical knowledge: trial and error and a good dose of intuition. Maxwell in 1868 made the first rigorous presentation on the control system stability. Thereafter the control theory has adopted its formal language: mathematics.

Since Maxwell's stability analysis, and to the present day, numerous people have contributed to the scientific maturity of control theory. Among them are Lyapunov, Nyquist, Bode and Popov, just to name a few. The two great wars, the space race and the telecommunications development were the major engines that drove, without precedence, new development methods for analysis and design of control systems.

Digital control appears as an "upgrade" to analog controllers. Besides the limit on the achievable operations complexity, the limits and tolerances of the physical components used in analog control was a serious disadvantage. More specifically the advent of digital control brought the following advantages:

- Increased performance
- Lower costs
- Reliability

- Flexibility

Regarding the first item, due to compensator complexity, generally the control systems performance increases. Additionally, since the digital processors cost tends to decrease the price of digital controllers are getting lower. Moreover, given that the controller coefficients are not generated by physical components, there are no drifts in the controller parameters. Thus we are witnessing an increase in reliability as well as in the replication capacity of the controllers. The ability to change, by "software", the controller parameters reflects a more agile, and less costly, control strategy.

Returning to the question raised initially, it's actually the study of the theory underlying the analysis and design of control systems that will move us. More specifically, since we are concerned with digital control, the mapping action/information refers to the case where the information feeds a digital processor who, by its turn, produces the action. Due to the usual analog nature of the process the information derived by the sensor is time-sampled before used by the microprocessor. In addition, due to the finite resolution of the computer core, the information is also quantized. Sampling and quantization are two exclusive operations of digital control systems whose effects must be understood. So, the basic objectives that the student should pursue along this course are:

- Understand discrete-time systems.
- Understanding computer-controlled systems.
- Being able to design digital controllers using classical techniques.

0.2 Document Structure

The subjects addressed in this document are condensed in Chapter 1 and Chapter 2. These two chapters are complemented by a set of appendices whose main objective is to make this document more self-contained.

The first Chapter is intended as a review of some basic concepts of continuous control system theory. The understanding of those concepts will be fundamental to fully understand discrete-time control systems. The second Chapter presents the theory of sampling systems and some basic frequency-domain controller

design techniques.

In each chapter, and whenever relevant, text boxes are presented with demonstrations and concepts which, although not contemplated in terms of program content, was found worthwhile to include .

0.3 Pre-requisites

In order to be able to grasp the subjects addressed in this curricular unit, the student must have some knowledge in the following subject areas:

- Differential and integral calculus;
- Complex analysis;
- Systems and Signals;

Some knowledge on data acquisition systems (A/D and D/A converters), under the instrumentation point-of-view, are expected. Additionally it's also expected some experience in using the numeric computation tool MATLAB® [1].

[CHAPTER ◄ 0]

[1] MATLAB is a trademak of The MathWorks Inc.

Chapter

1 Continuous-Time Control

1.1 Basic Concepts for Control Systems

THE main motivation, behind the design of a control system, is to force the system to exhibit a response profile, as consistent as possible, with the one required. This profile should be as independent, as achievable, to disturbances that might affect the system.

In order to do this, the majority of control system deigns procedures are based on a model (usually mathematical) of the process to control. Since the behaviour of a real dynamic system is often too complex to be modelled completely, usually only an approximation is used. In general these approximations rely on a set of assumptions such as linearity and time invariance.

Usually the dynamic systems behaviour, as well as the signals handled by them, are described, in the time domain, through a set of differential equations. For example, in continuous time domain, a signal or deterministic system can be described by a homogeneous differential equation of the type,

$$\frac{d^n x(t)}{dt^n} = \mathcal{F}\left(x(t), \frac{dx(t)}{dt}, \frac{d^2 x(t)}{dt^2}, \cdots, \frac{d^{n-1} x(t)}{dt^n} \right) \tag{1}$$

with initial conditions,

$$\left\{ x(0), \frac{dx(0)}{dt}, \frac{d^2 x(0)}{dt^2}, \cdots, \frac{d^{n-1} x(0)}{dt^{n-1}} \right\} \tag{2}$$

In the particular case of linear and time invariant systems, its dynamic behaviour is usually described by differential equations with constant coefficients of the form:

$$\frac{d^n y(t)}{dt^n} + a_{n-1}\frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + a_0 y(t) = b_m \frac{d^m u(t-T_0)}{dt^m} + \cdots + b_0 u(t-T_0) \qquad (3)$$

where $T_0 \geq 0$ refers to a pure time delay.

[note]

Signals can also be described by differential equations. For example consider the particular case of the 1-D signal $x(t) = A \cdot \sin(\omega t + \varphi)$. Deriving it twice with respect to time,

$$\frac{dx(t)}{dt} = A\omega \cdot \cos(\omega t + \varphi)$$

$$\frac{d^2 x(t)}{dt^2} = -A\omega^2 \cdot \sin(\omega t + \varphi)$$

On the other side since $A \cdot \sin(\omega t + \varphi) = x(t)$ the previous expression takes the form,

$$\frac{d^2 x(t)}{dt^2} = -\omega^2 \cdot x(t) \Rightarrow \frac{d^2 x(t)}{dt^2} + \omega^2 \cdot x(t) = 0$$

The solution of this differential equation is of type:

$$x(t) = C_1 \cdot e^{j\omega t} + C_2 \cdot e^{-j\omega t}$$

Taking into considerations the initial signal and the Euler relation leads to,

$x(t) = A \cdot \sin(\omega t + \varphi) = \frac{A}{2j} e^{j\varphi} \cdot e^{j\omega t} - \frac{A}{2j} e^{-j\varphi} \cdot e^{-j\omega t}$. Hence,

$$C_1 = \frac{A}{2j} e^{j\varphi} \ , \ C_2 = -\frac{A}{2j} e^{-j\varphi}$$

Thus it is easy to see that

$x(0) = C_1 \cdot e^{j\omega t}\Big|_{t=0} + C_2 \cdot e^{-j\omega t}\Big|_{t=0} = C_1 + C_2 = A \cdot \sin(\varphi)$ and

$$\frac{dx(0)}{dt} = j\omega C_1 \cdot e^{j\omega t}\Big|_{t=0} - j\omega C_2 \cdot e^{-j\omega t}\Big|_{t=0} = j\omega C_1 - j\omega C_2 = A\omega \cdot \cos(\varphi)$$

Thus, it can be concluded that $x(t) = A \cdot \sin(\omega t + \varphi)$ can be represented by the differential equation:

$\frac{d^2 x(t)}{dt^2} + \omega^2 \cdot x(t) = 0$ subject to initial conditions

$$\left\{ x(0) = A \cdot \sin(\varphi), \frac{dx(0)}{dt} = A\omega \cdot \cos(\varphi) \right\}$$

An alternative way of representing a system modelled by differential equations arises from the application of the Laplace transform. Thus, for a given set of initial conditions, the generic differential equation presented in (3) is replaced by the following expression in the Laplace domain[2] .

$$Y(s) = \frac{N(s)}{D(s)} U(s) + \frac{CI(s)}{D(s)} \tag{4}$$

where

$$\frac{N(s)}{D(s)} = e^{-sT_0} \cdot \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_0}{s^n + a_{n-1} s^{n-1} + \cdots + a_0} \tag{5}$$

and $CI(s)$ refers to a polynomial in s associated with the initial conditions of the system. Considering only the forced response, i.e. considering the initial conditions as zero, the relationship between $Y(s)$ and $U(s)$ is called the transfer function (TF) and has the shape of the ratio of two polynomials in s as shown in the following equation.

$$\frac{Y(s)}{U(s)} = \frac{N(s)}{D(s)} = e^{-sT_0} \cdot \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_0}{s^n + a_{n-1} s^{n-1} + \cdots + a_0} = G(s) \tag{6}$$

To ensure system causality, the degree of the denominator polynomial must be greater than, or equal, to the polynomial degree of the numerator, i.e. $n \geq m$. Causality is, of course, closely linked to the system physical existence.

[note **]** A system is said to be causal if its response does not depend on future values of the input signals.

The values of s that turn the ratio (6) equal to zero are called the system **zeros**. On the other hand, the values of s that make $G(s)$ infinite are designated by system **poles** system. A system with n poles is called a system of order n If it has l poles at the origin ($s = 0$) then it's a **type** l system.

[note **]** As one would see further ahead, the system type is closely related to the order of the polynomial, associated with the input signal, that the system can follow with finite steady-state error.

Depending on the ratio between the number of poles and the number of zeros the transfer function can be designated by:

[2] It is advised a previous study of Chapter A1 of this document

- **Proper** if $\lim_{s\to\infty} G(s) = C < \infty$ - In this case there are equal number of finite poles and zeros;
- **Improper** if $\lim_{s\to\infty} G(s) = \infty$ - An improper transfer function have more zeros than poles;
- **Strictly proper** if $\lim_{s\to\infty} G(s) = 0$ - This is the case where there are more finite poles than zeros.

[note **]** Most physical systems are modelled by strictly proper transfer functions. Moreover, as already said, they all require that the number of zeros is less than or equal to the number of poles in order to ensure causality.

Finally note that, in Laplace transform, s is a complex variable of the form $s = \sigma + j\omega$ where ω refers to the angular frequency (in radians per second) and σ is a damping coefficient whose value is related to the convergence region of the Laplace transform. Hence the system poles and zeros can be geometrically represented on a pair of orthogonal axis: one associated to the real part of the singularities and other to the imaginary part. This plot is designated by pole-zero map and, in reality, is just the representation of complex numbers in the Argand plane.

In a stable system (more on this subject ahead) there is a tight relationship between the Laplace transform and the Fourier transform. One can say that the Fourier transform is equal to the Laplace transform if $\sigma = 0$. In this case $s = j\omega$ and $G(j\omega)$ as a function of the frequency ω provides what is known as frequency response. Since $G(j\omega)$ is complex, it can be represented by magnitude and phase plots. The graphs of magnitude and phase of $G(j\omega)$, as a function of ω, are frequently designated by Bode plots.

1.1.1 Control System Stability

In control system design, the system stability is a major concern topic and must always be kept at line-of-sight. The stability of a causal, linear and time-invariant system can be evaluated from the solution of the characteristic equation. The characteristic equation is the mathematical equality obtained as $D(s) = 0$. The roots of $D(s)$, are poles of $G(s)$.

A causal, linear and time-invariant system is said asymptotically stable if all its poles, sometimes called modes, have negative real part. On the other hand, if there is at least, one pole with positive real part, the system is asymptotically unstable. In the case a system has one pole with real part equal to zero then one said he is marginal stable.

[note **]**

A system to be stable it's necessary that its impulse response is absolutely integrable (absolutely summable on discrete-time systems) [11] [12]. Mathematically this definition is expressed as:

$$\int_{-\infty}^{+\infty} |h(t)|\,dt < \infty \text{ or } \sum_{k=-\infty}^{+\infty} |h[k]| < \infty$$

Where $h(t)$ and $h[k]$ denotes de impulse response of continuous and discrete-time systems respectively. This is also one of the conditions necessary for a system to admit representation in the Fourier because:

$$H(j\omega) = \int_{-\infty}^{+\infty} h(t) \cdot e^{-j\omega t} dt$$

and for convergence one needs to have:

$$\int_{-\infty}^{+\infty} |h(t)|\,dt < \infty \text{ (one of the Dirichlet conditions [11] [12])}$$

Thus, if a linear time-invariant system with impulse response $h(t)$ admits representation in Laplace, $H(s)$, the convergence region must includes the $j\omega$ axis in order to admits Fourier representation and, by inherence, for the system to be stable. Note also that, if the system is causal, the region of convergence is the entire plane to the right of the rightmost pole. So, if a system is linear time-invariant and causal, it's necessary that all the poles are at the left side of the $j\omega$ axis for the system to be asymptotically stable (all the poles must have negative real parts). Obviously, if a system is not causal, to be stable all the poles must lay down in the right half-plane!

An alternative stability analysis derives from the system forced response. In that perspective, a system is said to be bounded input/ bounded output (BIBO) stable, if his response to a bounded input is bounded. Hence a linear time-invariant system is BIBO stable if, regardless of the signal profile, an amplitude-

limited input lead always to an amplitude-limited output.

[note **]** Bounded Input /Bounded Output Stability

Let us assume a linear time-invariant system governed by the equation,

$$y = T\{x(t)\}$$

where $T\{\cdot\}$ designates a transformation operation over the input signal $x(t)$. This system is stable in the BIBO sense if, after ensuring that $x(t)$ is limited in amplitude by a generic finite value, say B_x, the response $y(t)$ is also limited in amplitude by an arbitrary finite value B_y. So if $|x(t)| \le B_x < \infty \rightarrow |y(t)| \le B_y < \infty$, $\forall t$ then the system is BIBO stable.

Note that an asymptotically stable system is BIBO stable but the converse is not true. Consider, for example, a reducible second order system (one pole and one zero at the same point) with transfer function,

$$G(s) = \frac{s+a}{(s+a)(s+b)}$$

if a is negative and b positive the system is BIBO stable but it is not asymptotic stable since the characteristic equation has a pole with positive real part.

[note **]** The zeros location in the s plane does not contribute to the system stability. However there are different designations for systems with all zeros in the right half-plane and for systems with, at least, one zero in the left half-plane. The first type are called minimum phase systems and the latter non-minimum phase systems.

1.1.2 Control systems performance evaluation

The design procedure of a control system is relates to the fulfilment, by the system under closed-loop, of a set of performance specifications. Those specifications can be made over two different domains: time domain and frequency domain. In the former the figures of merit are expressed in terms of time constraints and in the later, as the name implies, the constraints are established in terms of frequency. The characteristics that a given system should exhibit can be defined in one or both domains. Usually, in overall, they

impose lower and upper limits to the following system characteristics:

- Speed of response and bandwidth;
- Stability (relative);
- Maximum allowed steady-state error.

In the following three subsections one presents and discusses the most common performance criteria in control systems.

1.1.2.1 Steady-State performance criteria

The steady-state error performance index is a measure of system accuracy when referred to a specific excitation signal. Normally three types of input are considered:

- Unit step (zero-order excitation signal)
- Ramp (first-order excitation signal)
- Parabola (second-order excitation signal)

The response to the first input signal measures the system's ability to react to rapid changes of the reference signal, and the remaining the system capacity to follow trajectories. In the time and frequency domain the above signals have the following mathematical representation:

	Step	**Ramp (gradient m)**	**Parabola**
Time	$r(t) = u(t)$	$r(t) = m \cdot t \cdot u(t)$	$r(t) = \frac{1}{2} t^2 u(t)$
Frequency	$R(s) = \frac{1}{s}$	$R(s) = \frac{m}{s^2}$	$R(s) = \frac{1}{s^3}$

The steady-state error (e_{ss}) is the difference between the instantaneous system response and his steady-state value. For stable systems this value can be analytically determined using the final-value theorem. This topic will be discussed in section § 1.1.8.2.

[note **]** Final-value theorem

The final value of the function $f(t)$, whose Laplace transform is $F(s)$, is:

$$\lim_{t \to \infty} f(t) = \lim_{s \to 0} sF(s)$$

Example: The unity step steady-state error of the first order system,

$$G(s) = \frac{\alpha}{s+a}$$

can be computed by the following set of algebraic operations:

$\frac{Y(s)}{U(s)} = \frac{\alpha}{s+a}$, $U(s) = \frac{1}{s}$

and the error is given by $E(s) = U(s) - Y(s)$

Since $Y(s) = \frac{\alpha}{s(s+a)}$ and

$$E(s) = \frac{1}{s} - \frac{\alpha}{s(s+a)} = \frac{1}{s}\left(1 - \frac{\alpha}{s+a}\right)$$

then the steady-state error is given by:

$e_{ss} = \lim_{s\to 0} sE(s) = 1 - \frac{\alpha}{s+a} = \lim_{s\to 0}\left(\frac{s+a-\alpha}{s+a}\right) = \frac{a-\alpha}{a}$ (If $a = \alpha$ then e_{ss} is zero)

1.1.2.2 Time Domain specifications

The time domain specifications are usually defined in terms of system response to a unit step. Among other, the following performance criteria are highlighted:

- *Rise Time* (T_R) - Time required for the unit step system response to raise from 10% to 90% of its value in steady state.
- *Time Delay* (T_D) - Time required for the system unit step response to reaches 50% of its value in steady state.
- *Settling Time* (T_S) - Time required for the unit step system response to reach, and stay, within a specified percentage of its value in steady state (typically ± 1%, ± 2% or ± 5%).
- *Time Constant (Predominant)* (τ) - Refers to an alternative measure of settling time. For a stable system of order greater than one, the time constant refers to the time required for the transient response envelope to reaches 63% of its value in steady state.
- *Overshoot* (δ_S) - Is the maximum difference between the transient and steady state system response to a step input. This criterion is representative of system relative stability and is usually presented as a percentage relative to the steady-state value.

1.1.2.3 Frequency Domain specifications

The most common specifications in the frequency domain are:

- *Gain Margin* (G_m) and *Phase Margin* (P_m) - Define a criterion to measure the system relative stability.
- *Bandwidth* (BW) - It is a measure of the system response speed and is often defined as the range of frequencies over which the gain does not differ by more than 3dB of its value at a specified frequency.
- *Resonance peak* (M_r) - It is also a measure of relative stability and refers to the maximum magnitude value of the closed-loop frequency response. This criterion is closely related to phase margin and often both quantities are related to the following approximation [6]:

$$M_r \approx \frac{1}{2 \cdot \sin\left(P_m/2\right)}$$

[note **]**

Because the models used in control system design are only approximations, it is not sufficient, to guarantee system stability, that the closed-loop poles are in the right half-plane. Thus, even if the system is stable, we want to know how near is from instability. A system with low stability margin is closer to instability than a system with larger stability margin. Stable systems with low stability margins only work in simulation (most likely, in practice, the system is unstable). Usually the systems are destabilized when the gain exceeds a certain threshold or there is too much phase lag. The gain and phase tolerances are referred as gain and phase margins.

The *gain margin* (G_m)is defined as the magnitude of the inverse open-loop transfer function evaluated at frequency ω_π: the frequency in which the phase angle is -180 ° (phase crossover frequency),

$$G_m = \left|G(j\omega)\right|^{-1}\Big|_{\omega=\omega_\pi}$$

On the other hand, *phase margin* (P_m) is defined as 180 ° plus the phase angle of the open-loop transfer function at frequency ω_{gc}: the frequency at which the gain is unity (gain crossover frequency),

$$P_m = 180 + \arg\left(G(j\omega)\right)\Big|_{\omega=\omega_{gc}\,(\omega:\,|G(j\omega)|=1)}$$

Empirically it is desirable that the system phase margin is between 45 ° and 60° and the gain margin between 2 and 4 (6 to 12 dB),

$$45^{\circ} < P_m < 60^{\circ} \quad \text{e} \quad 6dB < G_m < 12dB$$

When the open-loop frequency response produces a phase shift of 180 ° there is a risk of instability if the gain is above unity. More specifically the system is closed-loop unstable if:

$$G_m < 0 \text{ and } P_m \leq 0$$

In some circumstances, the gain and phase margins cannot be used as system instability indicators. For example in first and second-order systems phase never crosses the line of 180 degrees so the gain margin is always infinite.

1.1.3 Open-loop first-order systems

Understanding the behaviour, both in time and frequency domain, of first and second order systems is very important in analysis and design of control systems. This is because many physical systems have dynamics that can be approximated to the ones of first or second order systems.

A first-order system has only one pole and has the following generic transfer function:

$$G(s) = \frac{a}{s+a} \tag{7}$$

where the pole is located at $s = -a$.

The impulse system response has the following mathematical formulation:

$$h(t) = a \cdot e^{-at}u(t) = a \cdot e^{-\frac{1}{\tau}t}u(t) \tag{8}$$

From this last relation one concludes that the first order system time constant is equal to the inverse of the poles's absolute value. On the other hand, its bandwidth is equal to the magnitude of the pole: $BW = |a|$.

Thus in a first-order system, $\tau = BW^{-1}$ which means that the higher the

bandwidth, the smaller the time constant and the faster is the system response.

For some of the performance criteria discussed in section § 1.1.2.2 is easy to see that, in a first-order system, the rise time is approximately equal to:

$$T_R = 2.2 \cdot \tau$$

and the delay time approximately equal to:

$$T_D \approx 0.69 \cdot \tau$$

1.1.4 Open-loop second order systems

The transfer functions of second order are of vital importance in control systems design since the specifications (performance criteria to be met) are normally provied assuming the system is second order. The canonical transfer function for a second-order system has the following aspect:

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{9}$$

where ω_n is called the undamped natural frequency and ζ (zeta) the damping ratio.

It is easy to verify that the two poles of this transfer function are located in:

$$s = -\zeta\omega_n \pm j \cdot \omega_n\sqrt{1-\zeta^2} = \sigma \pm j\omega_d \tag{10}$$

where ω_d is called the damped natural frequency.

Depending on the damping ratio, the system may have:

- Two distinct pure real poles ($\zeta > 1$) - *Over-damped system*
- Two identical real poles ($\zeta = 1$) - *Critically damped system*
- Two complex conjugate poles ($0 < \zeta < 1$) - *System under-damped.*

The figure below illustrates the position of the poles of a canonical second order system, as a function of the damping factor. It should be noted that for values ζ below zero, the poles of the system occur in the right half plane indicating an unstable system.

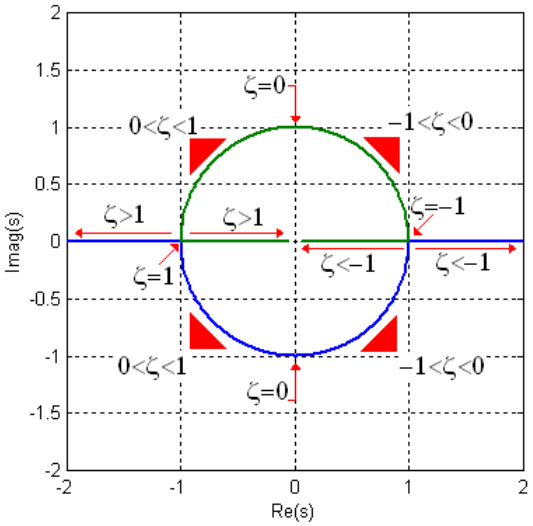


Fig 4. Location of a second-order system poles as a function of ζ

[note] As we will see later, there is a close relationship between the phase margin and the damping factor in closed loop (ζ_{cl}). Thus, if $\zeta_{cl} < 0$ the phase margin is negative. This implies instability of the closed-loop system.

Below is a set of functional relationships that allow to directly compute the values of some of the performance criteria established in § 1.1.2.2 and § 1.1.2.3:

- **Percentage of overshoot**

$$\delta_S = 100 \cdot e^{\left(-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}\right)} \Rightarrow \zeta = \left(\pi^2 \left[\ln\left(100/\delta_S\right)\right]^{-2} + 1\right)^{-1/2} \tag{11}$$

- **Settling time**

$$T_S(\pm 1\%) = 4.6/|\sigma| \text{ or } T_S(\pm 2\%) = 4/|\sigma| \tag{12}$$

- **Rise Time**

$$T_R = 1.8/\omega_n \tag{13}$$

- **Time Constant**

$$\tau = 1/|\sigma| \tag{14}$$

- **Bandwidth**

  The bandwidth depends on the natural frequency and ζ:

$$BW = \omega_n \left[1 - 2\zeta^2 + \left(2 - 4\zeta^2 + 4\zeta^4\right)^{1/2}\right]^{1/2} \tag{15}$$

  However, for $0.3 < \zeta < 1$, $BW \approx \omega_n \left[1.85 - 1.19 \cdot \zeta\right]$.

[note **]** Often the design of control systems assumes that the bandwidth can be approximated by ω_n, i.e. $BW = \omega_n$.

- **Resonance peak**

  In second-order systems the resonance peak is strongly connected to the damping coefficient. In fact the following approximation is used:

$$M_r = \left|G(j\omega)\right|_{@\omega=\omega_n} = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \text{ for } \zeta < \frac{\sqrt{2}}{2} \tag{16}$$

1.1.4.1 Poles location and transient response

Consider a generic second order system pair of poles such as those provided by the expression (10). Geometrically, in the s plane, each of the root equation coefficients refers to the characteristics indicated in the figure below.

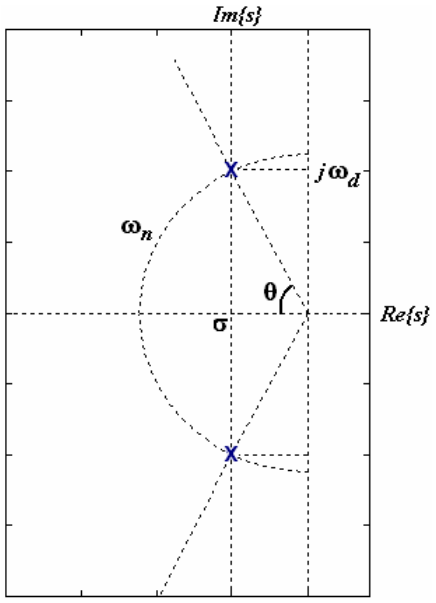


Fig 5. Location of poles as function of the parameters $\{\sigma, \omega_d, \omega_n, \zeta\}$

Changing the location of the two poles implies a change in system response. The effect, on the system response to a unit step, given the variation of each parameter $\{\sigma, \omega_d, \omega_n, \zeta\}$ can be summarized as follows:

- The settling time is inversely proportional to σ;
- The rise-time is inversely proportional to the vector pole module. More specifically is approximately equal to $T_R \approx 1.8/\omega_n$;
- The overshoot is directly proportional to θ where $\theta = \cos^{-1}(\zeta)$;

- The peak-time is inversely proportional to ω_d;
- The bandwidth is proportional to ω_n.

1.1.5 Reducing the system order

Often the system mathematical models are high-order differential equations. However, in many situations, these models can be approximated by differential equations of lower order with little information loss. The simplification is usually carried out by neglecting system modes with low influence on the overall transient response. The influence of a particular pole (or pair of complex poles) on the response is mainly determined by two factors:

- The real part of the pole;
- Relative magnitude of the residue at the pole.

The pole real part determines the rate, by which, the transient term due to this pole decays. The transient component decay is proportional to the magnitude of the pole's real part.

On the other hand, the relative magnitude of the residual, i.e. the coefficient associated with the *decreasing* in time exponential, determines the percentage of the total response due to this pole in particular. The relative magnitude of the residue, associated to a particular pole, may be drastically reduced due to the presence of a geometrically close zero.

Normally a pole, or pair of poles, is non-dominant if they are located far to the left, on the s plane, of those considered dominant (e.g. a decade or more).

[note **]** A decade refers to a ratio of frequencies equal to 10 (ten times higher or lower). An alternative specification is to express the relationship in octaves (two times higher or lower).

After the removal of one or more poles / zeros, the transfer function DC gain should be rescaled so that both transfer functions (primary and reduced) exhibit the same gain.

In order to illustrate what has been said let us consider the following system transfer function:

$$G(s) = \frac{120}{(s+0.5)(s+5)}$$

It has two poles, one in $s = -0.5$ and the other in $s = -5$, and the DC gain is 48. Since the transient response on the pole at -5 decays 10 times faster than the one at -0.5, we can try to approximate the 2nd order transfer function to a first order one with the following aspect.

$$G(s) = \frac{24}{s+0.5}$$ (Note the gain scaling ot the zero frequency!)

In order to compare the transient behaviour of both transfer functions, the step response is presented in the following figure.

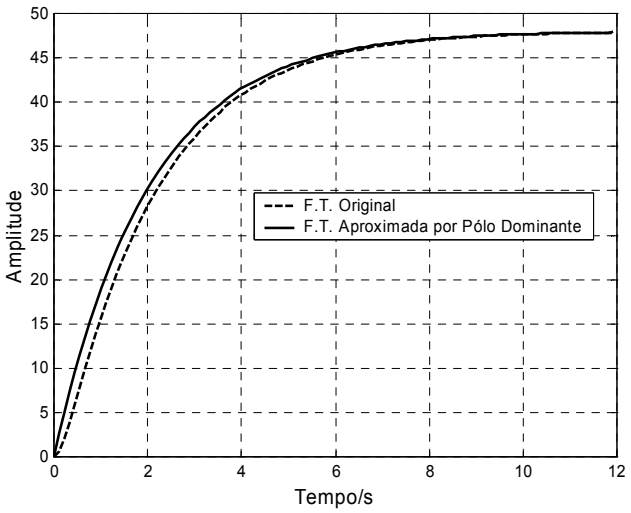


Fig 6. Step response of both original and reduced system.

As one can see, the dynamic behaviour of both systems is quite similar having a slight discrepancy only in the initial transient.

1.1.6 Noise Immunity vs. Bandwidth

The system bandwidth is directly proportional to the distance between the dominant poles and the origin of the s-plane. In other words, the system response time decreases and the output signal became more similar to the input one.

In order to illustrate this statement, the next figure presents the step response of three systems with different bandwidth. It's possible to verify that the response speed of the system with the pole at -10 is higher than the one from the system

with pole at -1. Moreover the response becomes to appear more “step shape like” for systems with high magnitude poles.

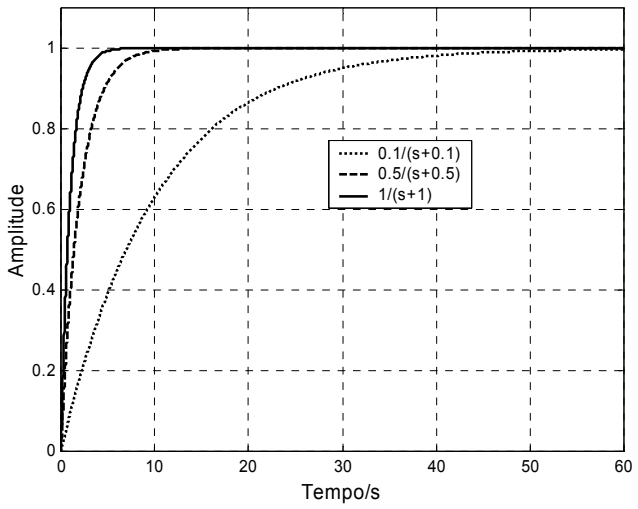


Fig 7. Step response of three different first order systems

So it is intended that the system has the highest possible bandwidth right?

Now imagine the response of two first order systems, one with a pole at 0.1 and another with a pole at 1, to a step contaminated with white noise (random signal with flat spectral density). Figure 8 present the simulation results for this set-up.

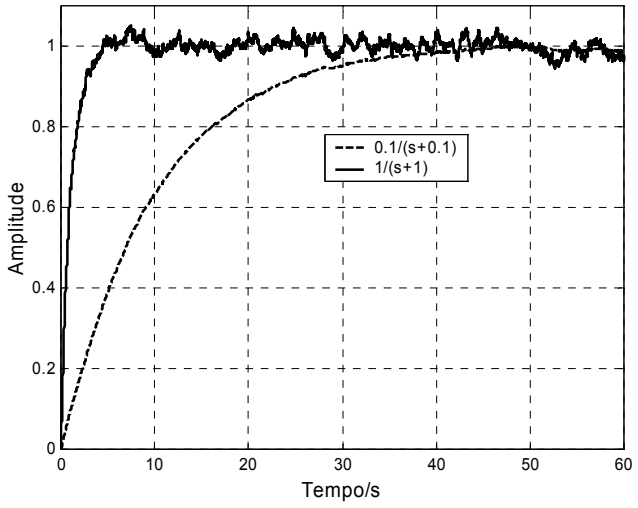


Fig 8. Response of two first order systems to a step contaminated with white noise (signal to noise ratio of approximately 6dB)

An analysis to the above figure reveals that a faster system has lower noise immunity than the slower one. Thus, there is a clear compromise between speed of response and noise immunity.

1.1.7 Systems linearization

All analysis and design techniques proposed in this curricular unit begin with the assumption that the system is linear. Large part of the classic tools, for both control systems analysis and synthesis, are based on the manipulation of linear differential equations (in time or in the frequency domain). This is due to the easier and faster mathematical manipulation of linear differential equations when compared with the numerical treatment usually required by nonlinear models. However, in reality, there are not linear systems. At least, a physical system is always conditioned by the nonlinear phenomena of saturation.

However, often, a physical system operates only around a given operating point and, within that dynamic range of operation, the system behaviour is approximately linear. Since the objective of a control system is to keep the process variables as close as possible to an equilibrium point, often a compensator can be designed considering the system linear if the operating zone is linearized. Both the linearized model and linear analysis method are valid within the operating point.

[note **]** Once the control system is synthesized, it is advisable to carry out a numerical simulation of the system with all its nonlinearities.

The linearization can be seen as the process of finding a linear model that approximates a non-linear one. This can be done in various ways depending on whether or not a mathematical model of the system is available. If so, the linearization can be carried out by expanding the nonlinear terms in Taylor series and neglecting the terms with order higher than unity. Alternatively this can be done from the data obtained experimentally. From the data collected, and given that the closed-loop controlled system remains near a given operating point, by system identification procedures or sometimes even graphically, one develops a linear model valid around that operating point [6].

1.1.8 Feedback system

An open-loop control system only has a proper behaviour if:

- The system model is accurate;

- There are no external disturbances;
- If the system parameters vary in a deterministic manner.

Since these conditions rarely occur, most systems rely on control by feedback. A simple feedback system can be modelled by the following block diagram:

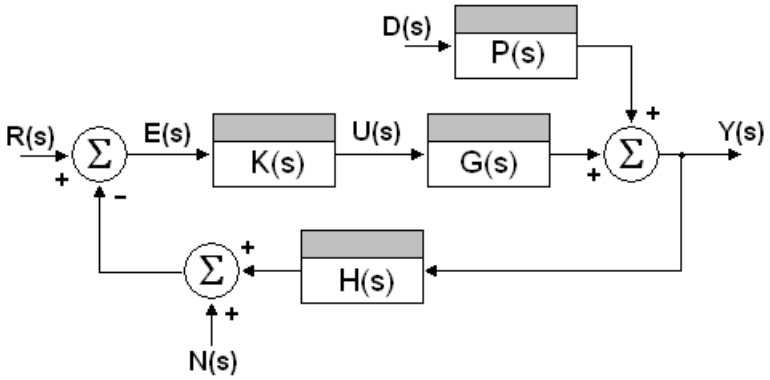


Fig 9. Block diagram of a feedback control system

Where $R(s)$ refers to the reference signal that the system should follow $E(s)$ the error signal, $U(s)$ is the controller output signal, $D(s)$ is the disturbance, $Y(s)$ is the output signal and $N(s)$ refers to measurement noise[3] introduced by the sensors. The transfer functions $K(s)$, $G(s)$, $H(s)$ and $P(s)$ refers to the controller, the plant, the sensor and the disturbance. In some circumstances, there is a pre-filter located after the signal $R(s)$ whose purpose is to eliminate the effect of some closed-loop zeros.

The design of some of the Figure 9 blocks (in particular $K(s)$) should make the overall system behaviour to act as imposed by the project constraints. More specifically the system must be able to:

- Follow the reference signal with the least possible error.
- Reject disturbances and error signals.

Just before moving on, a recap of some of the nomenclature, associated with the block diagrams of control systems, is in order. Taking into consideration the diagram shown in Figure 9, we present the following definitions:

- $K(s)G(s)$ - Direct transfer function
- $K(s)G(s)H(s)$ - Open-loop transfer function

[3] Usually of high-frequency.

- $Y(s)/R(s) = \dfrac{G(s)K(s)}{1+G(s)K(s)H(s)}$ - Closed-loop transfer function

[note **]** Although it seems that the open-loop transfer function should be $K(s)G(s)$ in fact, and considering $H(s)$ as the sensing element, the way you measure the system output is considered inherent to the system itself. The sensor dynamics cannot be separated from the system dynamics itself.

[note **]** Still regarding the stability margins, consider the closed-loop transfer function. One observes that, for a given frequency, the magnitude of the transfer function is infinite if the open loop gain is equal to -1. This corresponds, in terms of Bode plots, to a 0dB magnitude and a -180^{o} phase. It is from this relation that one can infer on the closed-loop stability by using open-loop information.

1.1.8.1 Sensitivity of closed-loop system

As already said, a closed-loop system has greater immunity to variations in the system dynamics so it has an increased ability to cope with variations in system parameters. In order to validate what has been said let us consider a control system with unity feedback as shown in the figure below.

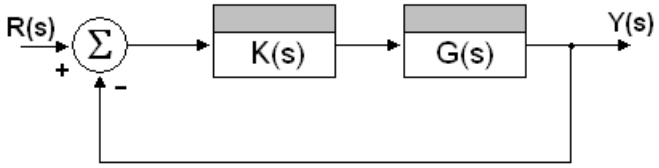


Fig 10. Closed-loop system with unity feedback.

The closed-loop transfer function has the following expression:

$$\frac{Y(s)}{R(s)} = T(s) = \frac{K(s)G(s)}{1+K(s)G(s)} \tag{17}$$

Now we evaluate the sensitivity of the closed-loop transfer function to some system variations. In order to do this one computes the sensibility of the closed-loop regarding the system transfer function, i.e. $\partial T(s)/\partial G(s)$

$$\frac{\partial T(s)}{\partial G(s)} = \frac{K(s)\left(1+K(s)G(s)\right)-K^{2}(s)G(s)}{\left(1+K(s)G(s)\right)^{2}} = \frac{K(s)}{\left(1+K(s)G(s)\right)^{2}} \tag{18}$$

multiplying and dividing by $G(s)$ become:

$$\frac{\partial T(s)}{\partial G(s)} = \frac{K(s)G(s)}{G(s)\left(1+K(s)G(s)\right)^2} = T(s)\frac{1}{G(s)\left(1+K(s)G(s)\right)} \tag{19}$$

This leads to the closed-loop transfer function relative variation:

$$\frac{\partial T(s)}{T(s)} = \frac{1}{\left(1+K(s)G(s)\right)} \cdot \frac{\partial G(s)}{G(s)} \tag{20}$$

It follows that the closed loop transfer function is insensitive to variations in the process transfer function for the frequencies to which the open-loop transfer function is high. That is, if

$$K(s)G(s)\big|_{s=s_0} \to \infty \tag{21}$$

then

$$\left.\frac{\partial T(s)}{T(s)}\right|_{s=s_0} \to 0 \tag{22}$$

Thus, for the design of a robust controller (insensitive system dynamics variations), it is necessary to find $K(s)$ so that the magnitude of the transfer function of open loop, is high for the frequencies at which there are significant variations in the transfer function of the system.

Another particularity of a closed-loop control system has to do with his ability to overcome the effect of disturbances on the controlled variable. In fact, analyzing the effect of $D(s)$ on the system output of the system of figure 9 (considering unity feedback and both $R(s)=0$ and $N(s)=0$) one gets,

$$\frac{Y(s)}{D(s)} = S(s) = \frac{P(s)}{1+G(s)K(s)} \tag{23}$$

This expression refers to the so-called sensitivity function. One concludes that, in order to reduce the influence of disturbances, the sensitivity function must provide low values for the frequencies components present in the disturbance. The same is to say that, considering $P(s)$ constant and equal to one, the open-loop transfer function must have a gain as high as possible in the disturbance frequency range.

The same reasoning can be carried out by considering now the measurement error. Still based on the image of Figure 9, considering unity feedback and the signals $R(s)$ and $D(s)$ equal to zero, the influence of measurement error in the

output signal is modelled by the following transfer function:

$$\frac{Y(s)}{N(s)} = -\frac{G(s)K(s)}{1+G(s)K(s)} \tag{24}$$

Therefore, to reduce the influence of measurement error, the closed-loop transfer function must provide low values for the frequencies range present in the noise.

Finally note that, to minimize the set-point tracking error, the closed-loop transfer function should be constant, and close to unity, for the range of frequencies present in the reference signal [8]. Thus, taking into account the system closed-loop sensitivity to both disturbance and noise, it appears that there is compatibility between the set-point tracking criterion and the disturbance rejection. However there is incompatibility between this objective and the measurement error reduction.

1.1.8.2 Steady-state error

In many control systems design, one of the imposed criteria has to do with the system steady-state response. For a closed loop stable system the level of the system output signal, $y(t)$, tends to be, in steady state, as close as possible to the magnitude of the command signal $r(t)$. The difference between these two values is called steady state error and can be computed by:

$$e_{ss} = \lim_{t\to\infty} e(t) = \lim_{t\to\infty} \left[r(t) - y(t) \right] \tag{25}$$

or, alternatively, in the Laplace domain,

$$e_{ss} = \lim_{s\to 0} sE(s) = \lim_{s\to 0} s\left[R(s) - Y(s) \right] \tag{26}$$

1.1.8.2.1 System with unity feedback

In the case of a unity feedback system, like that presented in figure 10, the steady state error can be determined from the open loop transfer function as,

$$E(s) = R(s) - Y(S) = R(s) - K(s)G(s)E(s) = R(s) - G_{OL}(s)E(S) \tag{27}$$

Solving in order to $E(s)$ become,

$$E(s) = \frac{1}{1+G_{OL}(s)} R(s) \tag{28}$$

Applying the final-value theorem,

$$e_{ss} = \lim_{s\to 0} \frac{sR(s)}{1+G_{OL}(s)} \tag{29}$$

For a polynomial excitation signal of degree k :

$$R(s) = \frac{1}{s^{k+1}}, \forall k \in \mathbb{N}_0^+ \tag{30}$$

The steady state error expression takes the following aspect:

$$e_{ss} = \lim_{s\to 0} \frac{1}{\left(1+G_{OL}(s)\right)s^k} \tag{31}$$

By admitting that the system is type l and has a transfer function with form;

$$G_{OL}(s) = \frac{N(s)}{s^l \cdot D(s)} \tag{32}$$

One concludes, by replacing (31) in (30), that:

$$e_{ss} = \lim_{s\to 0} \frac{s^l D(s)}{\left(s^l D(s)+N(s)\right)s^k} \tag{33}$$

From an analysis of the previous expression one can formulate the following conclusions:

- If $l > k$ the steady-state error is zero.
- If $l < k$ the steady-state error is infinite.
- If $l = k$

$$e_{ss} = \lim_{s\to 0} \frac{1}{\left(1+\frac{N(s)}{s^l D(s)}\right)s^l} = \lim_{s\to 0} \frac{1}{\left(1+G_{OL}(s)\right)s^l} \tag{34}$$

Systems of type 0, I and II are the most common as well as order 0, I and II excitation signals (steps, ramps and parabolas). The following table summarizes the e_{ss} values for all combinations between these three pairs of cases.

	Order 0	**Order 1**	**Order 2**
Type 0	$1/(1+K_P)$	∞	∞
Type I	0	$1/K_V$	∞
Type II	0	0	$1/K_A$

Tabela 1. Steady-state errors as function of system type and signal order (for unity feedback)

In the above table, the parameters K_P, K_V and K_A are designated by position, velocity and acceleration constants respectively. These constants are computed by the following relationships (derived from equation (32)):

$$K_P = \lim_{s\to 0} G_{OL}(s),\ K_V = \lim_{s\to 0} sG_{OL}(s) \text{ and } K_A = \lim_{s\to 0} s^2 G_{OL}(s)$$

[note **]** Often, the system steady state error for a step, ramp or parabola input is called position, velocity and acceleration error respectively.

1.1.8.2.2 Non-unity feedback system

For the generic closed-loop system structure with transfer function in the feedback loop equals to $H(s)$ (like the one shown in Figure 9), the steady state error can be determined from the following expression,

$$E(s) = R(s) - Y(S) = R(s) - \frac{K(s)G(s)}{1 + K(s)G(s)H(s)} R(s) = R(s) - G_{CL}(s)R(S) \qquad (35)$$

The same is to say,

$$E(s) = \left(1 - G_{CL}(s)\right) R(s) \qquad (36)$$

Applying the final-value theorem, and for an k - order excitation signal, the steady state error expression takes the following aspect:

$$e_{ss} = \lim_{s\to 0} \frac{1 - G_{CL}(s)}{s^{k+1}} \qquad (37)$$

1.1.9 First-order closed-loop systems

Consider a first order, causal and stable, system of the form:

$$\kappa G(s) = \frac{\kappa}{s + a} \qquad (38)$$

As one has seen earlier, the system bandwidth is equal to the pole magnitude. Additionally, and for $\kappa/a > 1$, the open-loop gain crossover frequency is $\omega_{gc} = \kappa$.

In closed-loop, with unity feedback, the transfer function becomes:

$$\frac{\kappa G(s)}{1+\kappa G(s)} = \frac{\kappa}{s+a+\kappa} \tag{39}$$

and changing the parameterization one obtains,

$$\frac{\kappa G(j\omega)}{1+\kappa G(j\omega)} = \left(\frac{\kappa}{\kappa+a}\right) \cdot \frac{1}{\frac{j\omega}{\kappa+a}+1} \tag{40}$$

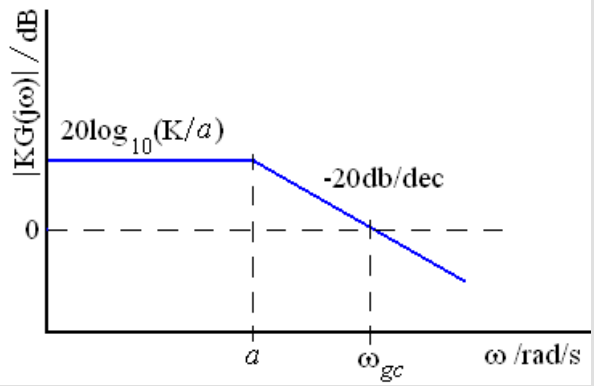
From the previous equation one can conclude that, since $\kappa/(\kappa+a)$ is less than unity, the magnitude of frequency response never crosses the line of 0dB. So there are not resonance peaks for first-order systems.

[proof **]** that the open-loop gain crossover frequency is $\omega_{gc} = \kappa$.

Using the Bode plots for,

$$\kappa G(j\omega) = \left(\frac{\kappa}{a}\right) \cdot \frac{1}{j(\omega/a)+1}$$

One gets the following asymptotic outline:



Since between a and ω_{gc} there are $\log_{10}(\omega_{gc}/a)$ decades, the attenuation over the frequency a is therefore $-20 \cdot \log_{10}(\omega_{gc}/a)$. It is known that for $\omega = \omega_{gc}$ the magnitude is 0dB hence,

$$20 \cdot \log_{10}(\kappa/a) - 20 \cdot \log_{10}(\omega_{gc}/a) = 0$$

which implies that $\omega_{gc} = \kappa$.

Additionally, and relatively to the open-loop system, there is an increase in bandwidth. For high gain values, the bandwidth of the closed loop is approximately equal to the gain crossover frequency. This statement is validated by the following figure.

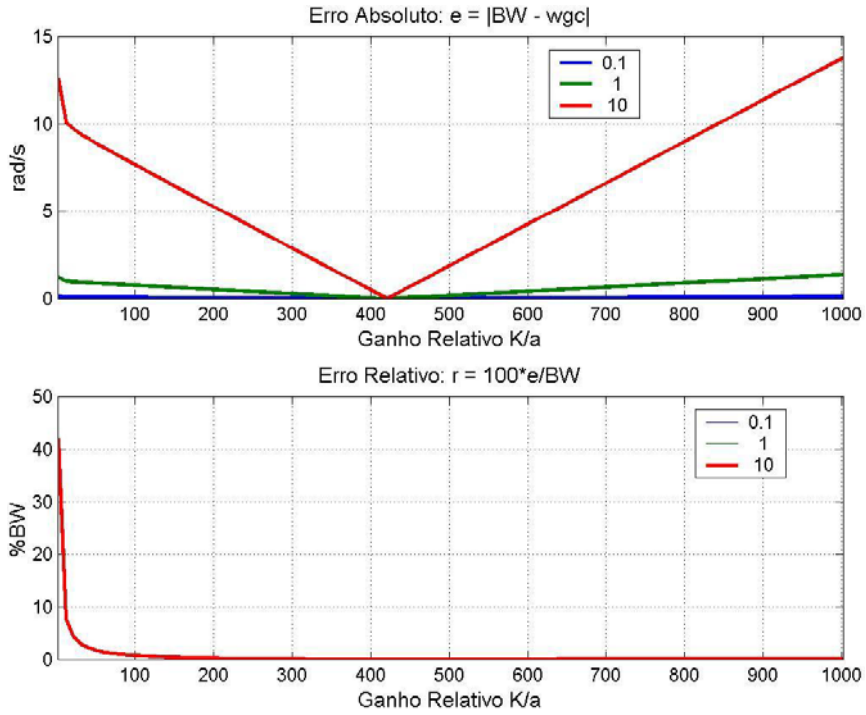


Fig 11. Absolute and relative error of approximation $BW_{cl} \approx \omega_{gc}$

It thus appears that the approach $BW_{cl} \approx \omega_{gc}$ is valid, within a tolerance of ± 1%, for DC open loop gains higher than 40dB. That is, if the gain is 100 times the pole module, the relationship in question remains within the defined limit. Note, however, that for lower values, the error of this approach can be quite high. For the simulated cases a relative error around 45% was obtained for a relative gain of 2, i.e. $\kappa/a = 2$.

1.1.10 Closed-loop second order systems

Consider the 2nd order system (causal and stable) in standard form:

$$\kappa G(s) = \frac{\kappa\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{41}$$

The transfer function under closed-loop unity feedback is:

$$\frac{\kappa G(s)}{1+\kappa G(s)} = \frac{\kappa\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2 + \kappa\omega_n^2} = \frac{\kappa\omega_n^2}{s^2 + 2\zeta\omega_n s + \left(\omega_n\sqrt{1+\kappa}\right)^2} \tag{42}$$

which can be rewritten as

$$\frac{\kappa G(s)}{1+\kappa G(s)} = \frac{\kappa\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_{ncl}{}^2} = \frac{\kappa_{cl}\omega_{ncl}^2}{s^2 + 2\zeta_{cl}\omega_{ncl} s + \omega_{ncl}{}^2} \tag{43}$$

where

$$\omega_{ncl} = \omega_n \sqrt{\kappa + 1} \tag{44}$$

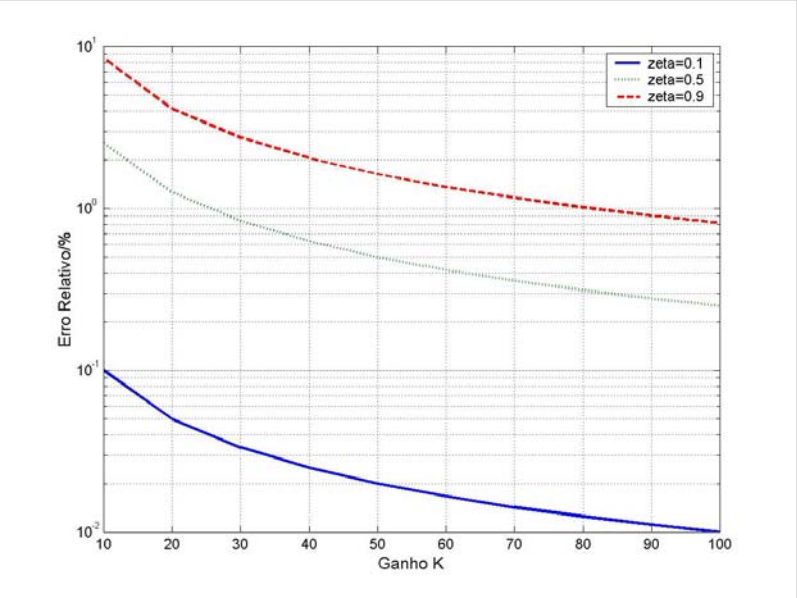
$$\zeta_{cl} = \frac{\zeta}{\sqrt{\kappa + 1}} \tag{45}$$

$$\kappa_{cl} = \frac{\kappa}{\kappa + 1} \tag{46}$$

- From expression (43) it follows that the system closed-loop bandwidth is higher than the open-loop bandwidth ($BW_{cl} \approx \omega_{ncl}$).
- By (44) one concludes that the closed-loop damping factor is lower than the open-loop damping one. So the overshoot will be higher.
- And from (45) the closed loop gain is less than the open-loop gain and less than one.

[note **]**

For high values of κ , the gain crossover frequency (ω_{gc}) is approximately equal to the undamped natural frequency of the closed loop, i.e. $\omega_{gc} \approx \omega_{ncl}$. More specifically the approach is valid, with an error below 10%, for values of $\kappa > 10$ and $\zeta < 1$ as illustrated in the following figure.



[note **]**

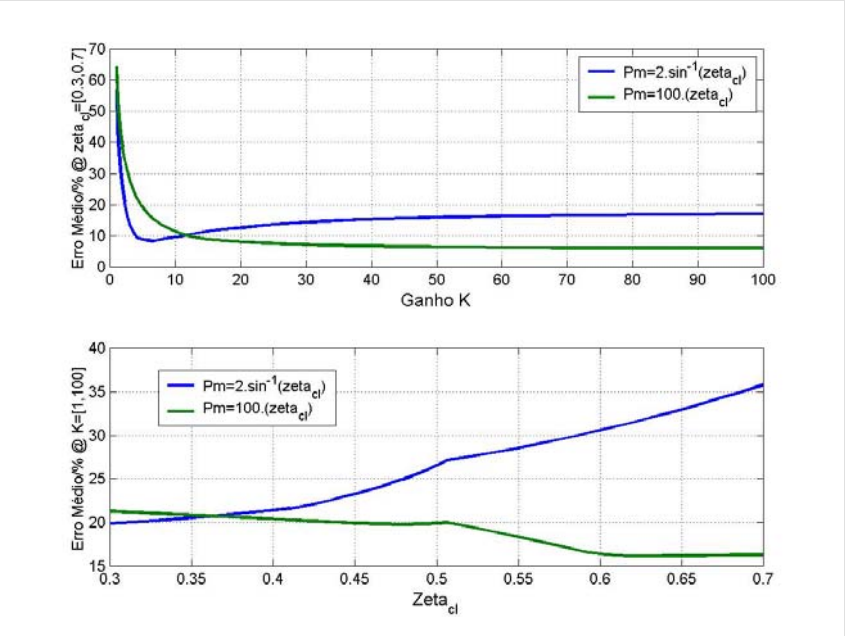
As noted earlier, there is a close relationship between phase margin and the closed loop damping ratio. An approximate relationship is given by the following formula [14]:

$$P_m \approx 2 \cdot \sin^{-1}(\zeta_{cl}) \qquad (1)$$

Another approach for phase margins less than 70 °, consists of [6]

$$\zeta_{cl} \approx \frac{P_m(\text{in degrees})}{100} \qquad (2)$$

The following figure illustrates the quality of each approximation.



It appears that for gain values below 3, the mean relative error increases exponentially. Moreover, it is observed that the approach by (1) only produces satisfactory results for gains between 3 and 10. From this point forward it is advisable to use the relation (2).

1.1.11 Open-loop vs. closed-loop response

The classical design and analysis techniques use the open-loop system response and try to predict his closed loop behaviour. Even if only a rude approximation of the closed-loop behaviour can be obtained from the open-loop transfer function, the following rules can be taken into account as an aid in control systems design.

In this framework, consider a system with direct transfer function $G_D(s)$. By evaluating $G_D(s)$ along the axis $j\omega$ one obtain the direct frequency response $G_D(j\omega)$. In closed loop, and taking into consideration the feedback transfer function $H(s)$, the frequency response will be:

$$G_{CL}(j\omega) = \frac{G_D(j\omega)}{1 + G_D(j\omega)H(j\omega)} = \frac{G_D(j\omega)}{1 + G_{OL}(j\omega)} \tag{47}$$

For values of $\left|G_{OL}(j\omega)\right| \gg 1$,

$$G_{CL}(j\omega) \approx \frac{G_D(j\omega)}{G_D(j\omega)H(j\omega)} = \frac{1}{H(j\omega)} = \frac{1}{\left|H(j\omega)\right|} e^{-j\angle H(j\omega)} \tag{48}$$

It appears then that for high magnitude values of the open loop transfer function, the closed loop frequency response has its magnitude approximately equal to the inverse feedback transfer function and phase with opposite sign. In the particular case of unity feedback ($H(j\omega) = 1$) the closed loop frequency response magnitude is approximately constant and equal to 0dB. The phase is also constant and equal to 0 º.

On the other hand, for values $\left|G_{OL}(j\omega)\right| \ll 1$,

$$G_{CL}(j\omega) \approx G_D(j\omega) \tag{49}$$

In this case the closed loop frequency response is approximately equal to the direct frequency response (both magnitude and phase).

In the vicinity of the gain crossover frequency, (for $\left|G_{OL}(j\omega)\right| \approx 1$) the closed loop frequency response magnitude strongly depends on the phase margin. Due to this fact, the relationship between the gain crossover frequency and closed-loop bandwidth mismatches. This discrepancy increases when the closed-loop zeta decreases. Thus, as a rule-of-thumb, one can say that the bandwidth of the closed loop is usually within one to two times the gain crossover frequency,

$$\omega_{gc} \le BW_{CL} \le 2 \cdot \omega_{gc} \tag{50}$$

An useful heuristic for control systems design is to consider the bandwidth equal to the gain crossover frequency in case of a phase margin of 90º or a bandwidth twice the gain crossover frequency if the system has an open loop phase margin of 45º [6]. That is,

$$BW_{CL} = \omega_{gc}\Big|_{Pm=90^o} \text{ or } BW_{CL} = 2\omega_{gc}\Big|_{Pm=45^o} \quad (51)$$

1.2 Control Systems Design

In this section some classical controller design techniques are reviewed. Please note that we are concerned in analog controller design for linear, time-invariant, causal and minimum phase systems. The techniques addressed are based into two distinct graphical approaches: the root locus and the Bode plots. Both techniques have substantial differences when compared. Besides the fact that the former is a time-domain analysis and design technique and the later a frequency domain approach, there are one major difference among them: the root-locus design procedure requires the knowledge of the system pole-zero location (i.e. a satisfactory process model must be known). On the other hand a Bode plot can be obtained experimentally and then be used for analysis and synthesis.

[**note**] Please note once again that the controller classical design techniques are based on the open-loop transfer function in order to **predict** the closed-loop system response.

1.2.1 The root-locus

The root locus shows the location of the closed-loop poles as a function of a given transfer function parameter variation (usually, but not exclusively, the gain). Besides the possibility of determining the stability and relative stability in closed loop, the root locus is also a common controller design tool [4] [10].

1.2.2 Bode diagrams

In control systems, Bode plots can be used for various purposes including the determination the values for of some figures-of-merit and for controller design. In the control system design framework there are two types of Bode diagrams: open-loop and closed-loop. Open-loop diagrams can be use to:

- Determine relative stability margins;
- Determine the system type (noting the slope of the frequency response);
- Controller design: Due to the diagrams addictive nature, the association

effect of a given compensator to the system can be easily determined.

On the other hand, closed-loop diagrams can be used to:

- Determine the bandwidth (a measure of system response and noise immunity);
- To determine the relative stability (the resonance peak in the Bode plot of closed loop is a reliable indicator of relative stability).

1.2.3 Controllers Types

At present time, there are a myriad of different type of controllers. Some of them are used in the industry and in our everyday life machines. Most of them are still in the theoretical domain and live in books and conference papers. Due to the time constraint imposed by the present curricular unit, only a brief treatment on a reduced number of controller strategies will be taken. Two classical control strategies, the PID and the Lead/Lag controller, will be addressed.

PID stands for proportional-integral-derivative and are the most common controller type in the process industry. With three degrees of freedom, this controller is able to meet most of the closed loop specifications (e.g., gain and phase margins or steady-state error). The PID has the following standard transfer function:

$$K(s) = K_P + K_d s + \frac{K_i}{s}$$

where K_P, K_d and K_i are the proportional, derivative and integral coefficients respectively. As one can see, the PID transfer function contributes with two zeros and a pole at the origin.

On the other hand, a lead/lag controller only has one zero and one pole. The canonical transfer function of such controller has the following structure:

$$K(s) = \kappa \frac{aTs + 1}{Ts + 1}$$

where the controller is of lag type if $|a| > 1$ and of lead type if $|a| < 1$.

Besides these two controller types, an alternative algebraic method is also presented. This method, denoted by unity feedback controller (UFC), consists

on an open-loop transfer function pole placement strategy. In order to use this method one must know where to put the closed-loop transfer function poles.

1.2.4 Controller design by pole-placement

Once again, the main control system objective is to reshape the natural system behaviour in order to obtain a new one, developed around the original, capable of meeting the desired design constraints. As can be imagined, there are many ways to do this task. The usual and basic strategy involves information feedback.

So, consider the following unity feedback system:

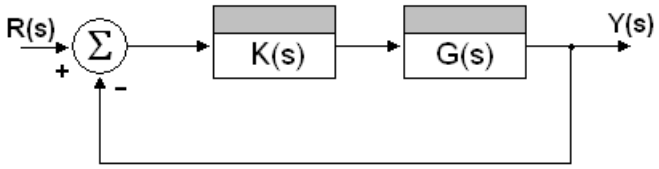


Fig 12. Closed-loop control system with unity feedback.

Taking into consideration the system transfer function $G(s) = N(s)/D(s)$ and the compensator transfer function $K(s) = B(s)/A(s)$, the aim of the control system is to make the dynamic behaviour of the closed-loop transfer function close to the desired one. Let $G_{CL}(s) = P(s)/Q(s)$ be the desired closed-loop transfer function.

Since the plant transfer function is considered fixed (otherwise any change may involve the physical alteration of the plant) the closed-loop dynamic is tuned by selecting a proper controller transfer function.

The closed loop transfer function of the system shown in figure 12 is:

$$G_{CL}(s) = \frac{G(s)K(s)}{1 + G(s)K(s)} = \frac{B(s)N(s)}{A(s)D(s) + B(s)N(s)}$$

and, since we want the system to display the behaviour dictated by the transfer function $G_{CL}(s)$, then the following relationship must be verified:

$$\frac{B(s)N(s)}{A(s)D(s) + B(s)N(s)} = \frac{P(s)}{Q(s)}$$

which leads to the following pair of project equations:

$$N(s)A(s) = P(s) \tag{52}$$

and

$$A(s)D(s) + B(s)N(s) = Q(s) \tag{53}$$

Please note that the closed-loop transfer function numerator, $N(s)B(s)$ cannot be changed. This is because $N(s)$ is intrinsic to the plant and $B(s)$ depends on equation {52} solution. Thus, using this project strategy, only the pole's position can be controlled. Thus, since the location of zeros also contributes to the system dynamic behaviour (e.g. the error in steady state) in general one cannot validate all the constraints imposed initially.

Since, as already mentioned before, the original system dynamics is considered unchangeable, equation {52} is of the type,

$$a \cdot X + b \cdot Y = c \tag{54}$$

known, in number theory, as the Diophantine equation. Thus, the objective of this technique is then proposed for solving a polynomial equation.

Note that often the Diophantine equation solution is not unique. Moreover, sometimes, the final design results in improper and physically impossible controller. However, it is possible to guarantee the existence of a proper controller if the following condition is verified:

If the system is of order n, strictly proper and irreducible, then there exists a controller of order $n-1$ for a order $2n-1$ characteristic polynomial $Q(s)$.

The general solution of Diophantine equation is presented by below taking into consideration the following transfer function expressions for both process and controller,

$$G(s) = \frac{N_n s^n + N_{n-1} s^{n-1} + \cdots + N_0}{D_n s^n + D_{n-1} s^{n-1} + \cdots + D_0} \tag{55}$$

$$K(s) = \frac{B_{n-1} s^{n-1} + B_{n-2} s^{n-2} + \cdots + B_0}{A_{n-1} s^{n-1} + A_{n-2} s^{n-2} + \cdots + A_0} \tag{56}$$

and assuming a closed-loop characteristic polynomial of the form:

$$Q(s) = R_{2n-1} s^{2n-1} + \cdots + R_1 s + R_0 \tag{57}$$

Multiplying the appropriate terms and equating the coefficients of identical

powers leads to the following set of matrix equations,

$$S(N,D)\cdot X = R \tag{58}$$

where

$$S(N,D)=\begin{bmatrix} D_0 & N_0 & 0 & 0 & \cdots & 0 & 0 \\ D_1 & N_1 & D_0 & N_0 & \cdots & \vdots & \vdots \\ \vdots & \vdots & D_1 & N_1 & \cdots & 0 & 0 \\ D_n & N_n & \vdots & \vdots & \cdots & D_0 & N_0 \\ 0 & 0 & D_n & N_n & \cdots & D_1 & N_1 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & D_n & N_n \end{bmatrix} \tag{59}$$

is called the Sylvester matrix and has order $2n$. On the other hand X and R are vectors with the following format:

$$X = \begin{bmatrix} A_0 & B_0 & A_1 & B_1 & \cdots & A_{n-1} & B_{n-1} \end{bmatrix}^T \tag{60}$$

$$R = \begin{bmatrix} R_0 & R_1 & R_2 & R_3 & \cdots & R_{2n-2} & R_{2n-1} \end{bmatrix}^T \tag{61}$$

Finally, the controller transfer function coefficients are obtained by solving the matrix equation,

$$X = S(N,D)^{-1} R \tag{62}$$

1.2.5 Tuning PID Controllers

In this section some control PID tuning techniques are presented. The first is an empirical method based on the free or forced system response. In addition, an analytical tuning method similar to the pole-placement, is reviewed in section §1.2.4. Frequency response based methods are also possible and will be addressed in some proposed exercises.

1.2.5.1 Ziegler and Nichols method

One method for tuning PID controller uses a set of empirical rules proposed in 1942 by Ziegler and Nichols. From the open-loop system step response, or evaluating the system closed-loop response at the edge of instability, was possible to derive a set of heuristics for easy tuning of a three degrees-of-freedom controller. However it should be noted that, although simple, a controller tuned by this method cannot achieve system closed-loop behaviour able to meet specific requirements (e.g. overshooting, settling time and so on).

However, its main advantage lays in the fact that a system mathematical model it is not necessary for the design process (as opposed to most techniques). Taking into consideration Ziegler and Nichols work one presents two different tuning controllers rules for stable systems.

1.2.5.1.1 Reaction Curve Method

The first technique follows from the open-loop system step response. In case the step response can be approximated to a first order system, with pure delay in the time, we get the following relevant variables:

- Slope of the tangent at the inflection point of the response;
- Interception of this tangent with the axis of time.

As can be seen in figure 13, the slope of the tangent is obtained from $m = K/\tau$ and the time delay t_d from the point where the line crosses the time axis.

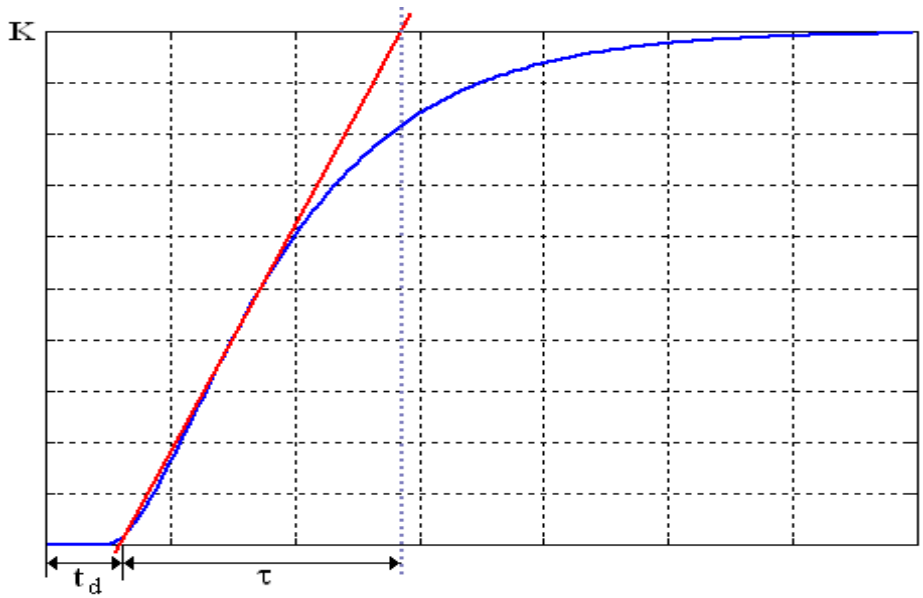


Fig 13. Reaction-curve tuning procedure.

Using this values, and based on the table below, the controller parameters are tuned.

	Kp	Ki	Kd
P	$1/(m \cdot t_d)$		
PI	$0.9/(m \cdot t_d)$	$0.3K_P/t_d$	
PID	$1.2/(m \cdot t_d)$	$K_P/(2 \cdot t_d)$	$0.5 \cdot K_P \cdot t_d$

Tabela 2. Ziegler-Nichols tuning rules for the reaction curve method.

It should be noted that the criteria defined in the tuning table normally lead to a

¼ decay. The same is to say that the closed-loop transient response magnitude decays to 25% after a period of oscillation. This performance criterion implies a closed-loop zeta approximately equal to 0.22 which can be considered a good compromise between speed of response and adequate stability margins.

[note **]**

The poles of a canonical second order system are located at $s = -\zeta\omega_n \pm j\omega_d$. Thus the system impulse response has the following appearance:

$$h(t) = \frac{\omega_n^2}{\omega_d} e^{-\zeta\omega_n t} \cdot \sin(\omega_d t)$$

As one can see the transient component decays exponentially. For the sinusoidal component attenuation to be 25% after a period it is necessary that

$$e^{-\zeta\omega_n t}\Big|_{t=T} = \frac{1}{4} \Rightarrow e^{-\zeta\omega_n T} = 0.25 \Rightarrow e^{-\zeta\omega_n \frac{2\pi}{\omega_d}} = 0.25 \Rightarrow e^{-\frac{\zeta 2\pi}{\sqrt{1-\zeta^2}}} = 0.25$$

solving in order to zeta one gets the solution $\zeta \approx 0.2155$.

1.2.5.1.2 Sensitivity Limit Method

In this second method, the parameters adjustment criterion is based on the system evaluation at the stability limit. More specifically, for the particular case of an asymptotically stable system in the region $0 \le \kappa \le K_C$, the following tuning rules are specified:

	Kp	Ki	Kd
P	$0.5 \cdot K_C$		
PI	$0.45 \cdot K_C$	$0.6K_P\omega_C/\pi$	
PID	$0.6 \cdot K_C$	$K_P\omega_C/\pi$	$K_P\pi/(4 \cdot \omega_C)$

Tabela 3. Ziegler-Nichols PID tuning rules

Where K_C refers to the critical gain and ω_C the oscillation frequency (imaginary part of the closed-loop poles for $\kappa = K_C$). If a model is available, the values for K_C and ω_C can be algebraically determined by the Routh stability criterion.

Other approaches can be used. For example the root-locus or the Bode plots are two tools that enable us to find both the critical gain and critical frequency. In the root-locus one searches for the gain that places the closed-loop poles

over the imaginary axis. This value refers to the critical gains. The pole vector magnitude at this point defines the oscillation frequency.

[note **] Routh stability criterion**

The Routh stability criterion is a method for finding the existence of poles in the right half-plane and can be applied to systems such as:

$$G(s)=\frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_0}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0}$$

The stability is analyzed from the characteristic equation:

$$a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0 = 0$$

The criterion is applied by constructing a table or matrix of the form:

s^n	a_n	a_{n-2}	a_{n-4}	$\cdots$
s^{n-1}	a_{n-1}	a_{n-3}	a_{n-5}	$\cdots$
$\vdots$	α_1	α_2	α_3	$\cdots$
$\vdots$	β_1	β_1	β_1	$\cdots$

where $\alpha_1=\frac{a_{n-1}a_{n-2}-a_n a_{n-3}}{a_{n-1}}$, $\alpha_2=\frac{a_{n-1}a_{n-4}-a_n a_{n-5}}{a_{n-1}}$, $\cdots$

and $\beta_1=\frac{\alpha_1 a_{n-3}-a_{n-1}\alpha_2}{\alpha_1}$, $\beta_2=\frac{\alpha_1 a_{n-5}-a_{n-1}\alpha_3}{\alpha_1}$, $\cdots$

All the characteristic equation roots have negative values if, and only if, the Routh's first column elements have all the same sign. Otherwise, the number of roots with positive real parts is equals to the number of sign changes.

Obs1: A row of zeros associated to the line s indicates that the polynomial has a pair of roots that satisfy the auxiliary equation $A\cdot s^2 + B = 0$ where A and B are the first and second elements of the row s^2.

Obs2: If any of the elements of the first column is zero (except the last) the zero is replaced by an infinitesimal amount ε. This amount is used for computing subsequent factors.

On the other hand, using the Bode diagrams, one can find the critical gain value as the one that makes the system gain margin equal to zero. The critical frequency is the crossover phase frequency.

1.2.5.2 Bode diagrams controller design

An alternative design way of PID controllers is based on the open loop system frequency response outline. Due to the addictive behaviour of Bode plots, and knowing both the open-loop frequency response and desired close-loop frequency response, it is often possible to find out, in an expeditious way, the controller coefficients.

For Bode plot controller design, an alternative controller transfer function parameterization is used. The transfer function has now the following structure:

$$K(s) = \frac{K_d s^2 + K_p s + K_i}{s} = \frac{K_i \left(\frac{s}{\omega_1} + 1\right)\left(\frac{s}{\omega_2} + 1\right)}{s} \tag{63}$$

where

$$K_p = \frac{K_i}{\omega_1} + \frac{K_i}{\omega_2} \tag{64}$$

and

$$K_d = \frac{K_i}{\omega_1 \omega_2} \tag{65}$$

[proof **]**

$$K(s) = \frac{K_d s^2 + K_p s + K_i}{s} = K_d \frac{s^2 + K_p / K_d \, s + K_i / K_d}{s}$$

The roots of the numerator are:

$$\begin{cases} s = -\frac{1}{2K_d}\left(K_p + \sqrt{K_p^2 - 4K_i K_d}\right) = -\omega_1 \\ s = -\frac{1}{2K_d}\left(K_p - \sqrt{K_p^2 - 4K_i K_d}\right) = -\omega_2 \end{cases}$$

so $K(s)$ can be rewritten as follows:

$K(s) = K_d \frac{(s+\omega_1)(s+\omega_2)}{s}$, I.e.

$$K(s) = \omega_1 \omega_2 K_d \frac{\left(\frac{s}{\omega_1} + 1\right)\left(\frac{s}{\omega_2} + 1\right)}{s}$$

The product $\omega_1 \omega_2$ is:

$$\omega_1 \cdot \omega_2 = \left[\frac{1}{2K_d}\left(K_p + \sqrt{K_p^2 - 4K_i K_d}\right)\right] \cdot \left[\frac{1}{2K_d}\left(K_p - \sqrt{K_p^2 - 4K_i K_d}\right)\right] =$$

$$= \frac{1}{4K_d^2}\left(K_p^2 - \left(K_p^2 - 4K_iK_d\right)\right) = \frac{1}{4K_d^2}\left(4K_iK_d\right) = \frac{K_i}{K_d}$$

hence,

$$K(s) = K_i \frac{\left(\frac{s}{\omega_1}+1\right)\left(\frac{s}{\omega_2}+1\right)}{s}$$

It is easy to see that the constant K_p and K_d can be taken from the values of the variables ω_1, ω_2 and K_i as follows:

$$K_d = \frac{K_i}{\omega_1\omega_2} \text{ and } K_p = K_i\frac{\omega_1+\omega_2}{\omega_1\omega_2}$$

1.2.5.3 Analytical design strategy

By knowing the following performance criteria:

- Steady state error;
- Bandwidth;
- Phase margin.

an analytical PID controller design technique can be derived.

The phase margin can be derived, for example, by knowing the maximum allowable overshoot and the bandwidth can be indirectly inferred by the settling time.

To illustrate the procedure consider the open loop transfer function of a system controlled by a PID (for a unity feedback loop):

$$G_{OL}(s) = \left(K_p + K_d s + \frac{K_i}{s}\right)G(s) \tag{66}$$

If case of a type p system, the compensated system is of type $p+1$ (due to the additional pole at the origin added by the controller). From section § 1.1.8.2 one already knows that the error constant is equal to the steady-state error inverse and is given by:

$$K_{p+1} = \lim_{s\to 0} s^{p+1}\left(\frac{K_p s + K_d s^2 + K_i}{s}\right)G(s) = \lim_{s\to 0} s^p K_i G(s) = \frac{1}{e_{ss}} \tag{67}$$

Thus, for a given steady state error, one of the controller parameter is immediately obtained: the integral constant K_i

It has been said earlier that the closed-loop natural frequency corresponds to the open-loop gain crossover frequency. It is also known that the phase margin can be obtained from the closed loop damping coefficient. Thus, for the frequency $\omega = \omega_{gc}$, the compensator must have unity gain and phase equal to $\theta_{\omega=\omega_{gc}} = P_m - 180^{\circ}$.

Resulting from these facts, and since the integral constant is known, the following equality can be written [14]:

$$\left(K_p + j\omega_{gc}K_d + \frac{K_i}{j\omega_{gc}} \right) G(j\omega_{gc}) = 1e^{j\theta_{\omega=\omega_{gc}}} \tag{68}$$

i.e.,

$$K_p + j\omega_{gc}K_d = \frac{1e^{j\theta_{\omega=\omega_{gc}}}}{G(j\omega_{gc})} + j\frac{K_i}{\omega_{gc}} \tag{69}$$

which leads to that

$$K_p = \mathrm{Re}\left\{ \frac{1e^{j\theta_{\omega=\omega_{gc}}}}{G(j\omega_{gc})} + j\frac{K_i}{\omega_{gc}} \right\} \text{ and } K_d = \mathrm{Im}\left\{ \frac{1e^{j\theta_{\omega=\omega_{gc}}}}{G(j\omega_{gc})} + j\frac{K_i}{\omega_{gc}} \right\} \cdot \frac{1}{\omega_{gc}} \tag{70}$$

1.2.6 Lead/Lag controller design strategies

One of the simplest forms of a compensator is just a filter with one pole and one zero. In this context two basic controller types will be reviewed: the phase lead and the phase lag controller.

A lead controller, as its name implies, add positive phase to the system. On the other hand a lag controller add negative phase to the system. The controller type choice is application specific. However, usually, a phase advance controller is used in situations where an increase in bandwidth and phase margin is needed. On the other hand, a lag controller has an opposite effect: tends to decrease the bandwidth but increasing the steady state performance.

1.2.6.1 Phase-lead controllers

A phase lead controller has the following effects on the control system behaviour:

- Increase of relative stability by increasing the phase margin;

- Increasing the bandwidth;
- Increase the steady-state error;
- Increased time response;
- Reduction of overshoot (higher zeta)
- Poor noise immunity.

1.2.6.1.1 Design strategy: Bode diagrams

Consider the transfer function of a lead compensator parameterized as follows:

$$K(s) = \kappa \frac{aTs+1}{Ts+1}\ ,\ a>1 \tag{71}$$

for $s = j\omega$ one obtain,

$$K(j\omega) = \kappa \frac{j\omega aT+1}{j\omega T+1} = \left|K(j\omega)\right| \angle K(j\omega)$$

The controller provides a phase advance that can be computed by:

$$\angle K(j\omega) = \Phi = \tan^{-1}\left(aT\omega\right) - \tan^{-1}\left(T\omega\right) \tag{72}$$

The frequency, at which the phase advance is maximal, can be calculated by solving the following equation:

$$\Phi_{MAX} = \frac{d}{d\omega}\left(\tan^{-1}\left(aT\omega\right) - \tan^{-1}\left(T\omega\right)\right) = 0$$

leading to

$$\frac{d}{d\omega}\left(\tan^{-1}\left(aT\omega\right) - \tan^{-1}\left(T\omega\right)\right) = \frac{aT}{1+\left(aT\omega\right)^2} - \frac{T}{1+\left(T\omega\right)^2} = 0$$

For the above equation to be zero it is necessary that

$$aT\cdot\left[1+\left(T\omega\right)^2\right] - T\cdot\left[1+\left(aT\omega\right)^2\right] = 0$$

so,

$$aT + aT^3\omega^2 - T - a^2T^3\omega^2 = 0 \Rightarrow$$
$$aT^3\omega^2\left(1-a\right) = T\left(1-a\right) \Rightarrow$$
$$\omega^2 = \frac{1}{aT^2}$$

and thus

$$\omega = \frac{1}{T\sqrt{a}} \tag{73}$$

Thus the frequency at which the phase advance is maximal, occurs at:

$$\omega = \frac{1}{T\sqrt{a}} = \omega_{\Phi}$$

the phase advance, at ω_{Φ}, is maximal and its value is equal to:

$$\Phi_{MAX} = \tan^{-1}\left(\sqrt{a}\right) - \tan^{-1}\left(\frac{1}{\sqrt{a}}\right) \tag{74}$$

leading to

$$\sin\left(\Phi_{MAX}\right) = \frac{a-1}{a+1}$$

and ultimately, to the value of constant a regarding the maximum phase advance:

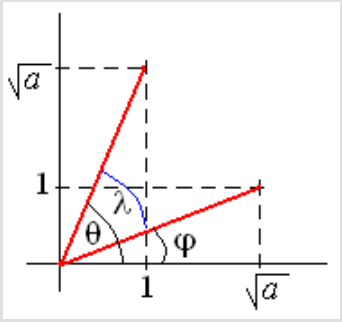
$$a = \frac{1+\sin(\Phi_{MAX})}{1-\sin(\Phi_{MAX})} \tag{75}$$

[proof **]**

Consider the following equality:

$$\Phi_{MAX} = \tan^{-1}\left(\sqrt{a}\right) - \tan^{-1}\left(\frac{1}{\sqrt{a}}\right) = \theta - \varphi = \lambda\,.$$

Geometrically, the expression has the following aspect:



Note: $\tan(\tau) = \frac{y}{x}$

Form the vector calculus theory one knows that the inner product between two vectors $\vec{a} =< a_1, a_2, \cdots, a_n >$ and $\vec{b} =< b_1, b_2, \cdots b_n >$ is given by:

$$\vec{a}\cdot\vec{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \left|\vec{a}\right|\cdot\left|\vec{b}\right|\cdot\cos(\varepsilon)$$

where ε refers to the angle formed by the two vectors and $\left|\vec{a}\right| = \sqrt{a_1^2 + a_2^2 + \cdots a_n^2}$

and $\left|\vec{b}\right| = \sqrt{b_1^2 + b_2^2 + \cdots b_n^2}$ refers to the absolute value of the vectors $\vec{a}$ and $\vec{b}$ respectively.

So it is easy to verify that

$$\cos(\lambda) = \frac{< \sqrt{a}, 1 > \cdot < 1, \sqrt{a} >}{\left| < \sqrt{a}, 1 > \right| \cdot \left| < 1, \sqrt{a} > \right|} = \frac{\sqrt{a} + \sqrt{a}}{\sqrt{1+a} \cdot \sqrt{1+a}} = \frac{2\sqrt{a}}{1+a}$$

By the fundamental theorem of trigonometry,

$$\cos(\lambda) = \sqrt{1 - \sin^2(\lambda)} = \frac{2\sqrt{a}}{1+a}$$

so,

$$\sin(\lambda) = \sqrt{1 - \frac{4a}{(1+a)^2}} = \sqrt{\frac{1 - 2a + a^2}{(1+a)^2}} = \sqrt{\frac{(a-1)^2}{(1+a)^2}} = \frac{a-1}{a+1}$$

$$a - 1 = (a+1) \cdot \sin(\lambda) \Rightarrow a = \frac{1 + \sin(\lambda)}{1 - \sin(\lambda)} \underset{\lambda = \Phi_{MAX}}{\rightarrow} a = \frac{1 + \sin(\Phi_{MAX})}{1 - \sin(\Phi_{MAX})}$$

[alternative proof **]**

Taking into consideration that

$\tan^{-1}\left(\frac{1}{a}\right) = \frac{\pi}{2} - \tan^{-1}(a)$ one get the following equality:

$$\Phi_{MAX} = \tan^{-1}\left(\sqrt{a}\right) - \tan^{-1}\left(\frac{1}{\sqrt{a}}\right), \Phi_{MAX} = \tan^{-1}\left(\sqrt{a}\right) - \frac{\pi}{2} + \tan^{-1}\left(\sqrt{a}\right)$$

$$\left(\frac{1}{2}\right)\left(\Phi_{MAX} + \frac{\pi}{2}\right) = \tan^{-1}\left(\sqrt{a}\right)$$

$\tan\left(\left(\frac{1}{2}\right)\left(\Phi_{MAX} + \frac{\pi}{2}\right)\right) = \sqrt{a}$ which implies that $a = \tan^2\left(\left(\frac{1}{2}\right)\left(\Phi_{MAX} + \frac{\pi}{2}\right)\right)$. Thus,

$$a = \left(\frac{\sin\left(0.5 \cdot \left(\Phi_{MAX} + 0.5\pi\right)\right)}{\cos\left(0.5 \cdot \left(\Phi_{MAX} + 0.5\pi\right)\right)}\right)^2 = \left(\frac{\left(\sin(0.5 \cdot \Phi_{MAX}) + \cos(0.5 \cdot \Phi_{MAX})\right)}{\left(\cos(0.5 \cdot \Phi_{MAX}) - \sin(0.5 \cdot \Phi_{MAX})\right)}\right)^2$$

$$a = \left(\frac{\sin^2(0.5 \cdot \Phi_{MAX}) + \cos^2(0.5 \cdot \Phi_{MAX}) + 2 \cdot \sin(0.5 \cdot \Phi_{MAX}) \cdot \cos(0.5 \cdot \Phi_{MAX})}{\cos^2(0.5 \cdot \Phi_{MAX}) - 2 \cdot \sin(0.5 \cdot \Phi_{MAX}) \cdot \cos(0.5 \cdot \Phi_{MAX}) + \sin^2(0.5 \cdot \Phi_{MAX})}\right)$$

$$a = \left(\frac{1 + 2 \cdot \sin(0.5 \cdot \Phi_{MAX}) \cdot \cos(0.5 \cdot \Phi_{MAX})}{1 - 2 \cdot \sin(0.5 \cdot \Phi_{MAX}) \cdot \cos(0.5 \cdot \Phi_{MAX}))}\right) = \frac{1 + \sin\left(\Phi_{MAX}\right)}{1 - \sin\left(\Phi_{MAX}\right)}$$

The additional gain contribution, from the transient component, at frequency $\omega = \omega_\Phi$ is:

$$M = \left| \frac{j\omega_\Phi aT + 1}{j\omega_\Phi T + 1} \right| \Leftrightarrow M_{dB} = 20\log_{10}\left(\left| \frac{j\omega_\Phi aT + 1}{j\omega_\Phi T + 1} \right| \right)_{dB}$$

$$M_{dB} = 20\log_{10}\left(\left| \frac{j\sqrt{a} + 1}{ja^{-0.5} + 1} \right| \right)_{dB} = 20\log_{10}\left(\sqrt{1+a} \right) - 20\log_{10}\left(\sqrt{\frac{1+a}{a}} \right)$$

in other words

$$M_{dB} = 10\log_{10}(1+a) - 10\log_{10}\left(\frac{1+a}{a} \right) = 10\log_{10}(a) \qquad (76)$$

Using the above derived equations a design procedure is systematized by the following seven steps [14].

[note **]** With this algorithm is not possible to specify the gain crossover frequency.

Algorithm:

Step 1 of 7: Calculate the gain κ so the error constant has the desired value

Step 2 of 7: Draw the Bode plot of $\kappa G(j\omega)$ and find the phase margin.

Step 3 of 7: Compute the amount of phase lead necessary Φ and add to it five or ten degrees (security margin).

Step 4 of 7: Find the value of a from the expression $a = \frac{1+\sin(\Phi)}{1-\sin(\Phi)}$

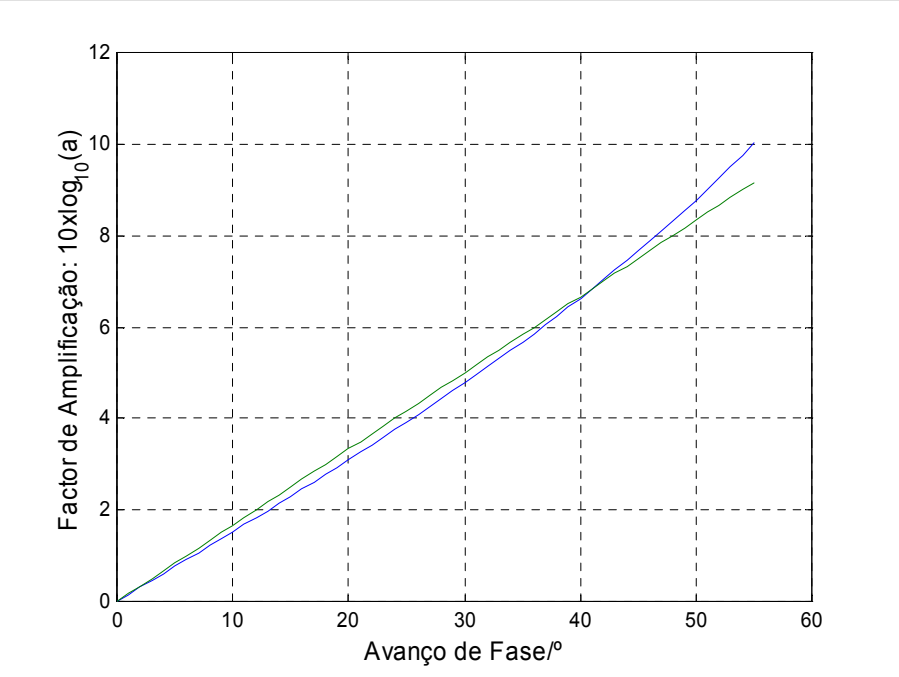
[note **]** The practical limit to a is 10 which is equivalent to a maximum phase increase of 55º.

Step 5 of 7: Search for the frequency at which the gain is $\left|\kappa G(j\omega)\right|_{dB} = -10 \cdot \log_{10}(a)$. The value of this frequency will be the gain crossover frequency.

[note **]** The phase advance, for each zero, should not be higher than 45 ° so that its position does not significantly interfere with the frequency response magnitude by changing the gain crossover frequency ω_{gc}.

[note **]** **Relationship between attenuation and phase advance**.

For a phase advance between 0º and 55º, the ratio between the signal amplification (due to the controller) and the phase advance has, in dB, the following aspect



If we compare the result profile with a line through the origin it can be said that, for an error lower than 0.85dB (0.25dB between 0º and 45º), the attenuation to consider in the design algorithm is approximately 1/6 of the phase lead (in degrees) needed.

Step 6 of 7: Calculate T from $T = \dfrac{1}{\omega_{gc}\sqrt{a}}$

Step 7 of 7: Draw the Bode plots for $K(j\omega)G(j\omega)$ to confirm the design constraints. One should also simulate the response of the closed-loop system

1.2.6.1.2 Design Strategy: By analytic manipulation

Just like was done for the PID controller, another lead controller tuning strategy can be obtained analytically. Consider a system, with transfer function $G(s)$, in series with a phase lead controller with transfer function:

$$K(s) = \kappa \frac{aTs+1}{Ts+1}$$

Consider also that the system must have, in closed loop, a steady-state error

less than or equal to δ and a bandwidth ω_c. In addition, the phase margin must be equal to P_m degrees.

From the first specification, and depending on the order of the polynomial type input signal (step, ramp or parabola), the gain κ is obtained from the final-value theorem. For example if the input is a step and the system is type 0 then

$$\varepsilon = \frac{1}{1+K_P} \leq \delta$$

where

$$K_P = \lim_{s\to 0} K(s)G(s) = \kappa \cdot \lim_{s\to 0} G(s)$$

so

$$\kappa \geq \frac{1-\delta}{\delta \cdot \lim_{s\to 0} G(s)}$$

Parameters a and T are obtained from the other performance criteria. Considering that the bandwidth is approximately equal to the gain crossover frequency then, at frequency $\omega = \omega_{gc} = \omega_c$, the system should exhibit a phase equal to $\phi = Pm - 180$. Therefore,

$$K(j\omega)G(j\omega)\big|_{@\,\omega=\omega_c} = e^{j\phi}$$

Considering that, at the frequency $\omega = \omega_c$ the system magnitude and phase are $|G(j\omega)| = M$ and $\measuredangle G(j\omega) = \theta$ then

$$K(j\omega)G(j\omega)\big|_{@\,\omega=\omega_c} = K(j\omega)\big|_{@\,\omega=\omega_c} M \cdot e^{j\theta} = e^{j\phi}$$

i.e.,

$$K(j\omega)\big|_{@\,\omega=\omega_c} = \frac{1}{M} e^{j(\phi-\theta)}$$

Because,

$$K(j\omega) = \kappa \frac{jaT\omega + 1}{jT\omega + 1}$$

then

$$K(j\omega)\big|_{@\,\omega=\omega_c} = \kappa \frac{jaT\omega_c + 1}{jT\omega_c + 1} = \frac{1}{M} e^{j(\phi-\theta)}$$

or

$$\frac{jaT\omega_c+1}{jT\omega_c+1}=\frac{1}{\kappa\cdot M}e^{j(\phi-\theta)}=\frac{1}{\kappa\cdot M}e^{j(Pm-180-\theta)}$$

Solving the above equation one gets:

$$a=-\left(\frac{(\kappa\cdot M)\cos(Pm-\theta)+1}{\kappa\cdot M\left[\kappa\cdot M+\cos(Pm-\theta)\right]}\right) \text{ and } T=\frac{\kappa\cdot M+\cos(Pm-\theta)}{\omega_c\sin(Pm-\theta)}$$

or,

$$aT=-\left(\frac{(\kappa\cdot M)\cos(Pm-\theta)+1}{\kappa\cdot M\cdot\omega_c\sin(Pm-\theta)}\right) \text{ and } T=\frac{\kappa\cdot M+\cos(Pm-\theta)}{\omega_c\sin(Pm-\theta)} \quad (77)$$

[proof]

$\frac{jaT\omega_c+1}{jT\omega_c+1}=\frac{1}{\kappa\cdot M}e^{j(Pm-180-\theta)}$ multiplying the left term by the conjugate and applying Euler's identity to the right term one obtain:

$$\frac{(jaT\omega_c+1)(-jT\omega_c+1)}{(T\omega_c)^2+1}=\frac{1}{\kappa\cdot M}\left(\cos(Pm-180-\theta)+j\sin(Pm-180-\theta)\right)$$

$$\frac{1+a(T\omega_c)^2+jT\omega_c(a-1)}{(T\omega_c)^2+1}=\frac{1}{\kappa\cdot M}\left(\cos(Pm-180-\theta)+j\sin(Pm-180-\theta)\right)$$

Separating the real from the imaginary part we have that:

$$\frac{1+a(T\omega_c)^2}{(T\omega_c)^2+1}=\frac{1}{\kappa\cdot M}\cos(Pm-180-\theta)=-\frac{1}{\kappa\cdot M}\cos(Pm-\theta) \quad (1)$$

$$\frac{T\omega_c(a-1)}{(T\omega_c)^2+1}=\frac{1}{\kappa\cdot M}\sin(Pm-180-\theta)=-\frac{1}{\kappa\cdot M}\sin(Pm-\theta) \quad (2)$$

Solving (1) in order to a

$$a=-\frac{1}{\kappa\cdot M(T\omega_c)^2}\cos(Pm-\theta)\left[(T\omega_c)^2+1\right]-\frac{1}{(T\omega_c)^2} \quad (3)$$

Substituting (3) in (2) and simplifying one gets:

$$-\frac{\cos(Pm-\theta)}{\kappa\cdot M(T\omega_c)^2}\left[(T\omega_c)^2+1\right]-\frac{1}{(T\omega_c)^2}=-\frac{\sin(Pm-\theta)}{\kappa\cdot M\cdot T\omega_c}\left[(T\omega_c)^2+1\right]+1$$

$$\frac{\cos(Pm-\theta)}{\kappa\cdot M(T\omega_c)^2}\left[(T\omega_c)^2+1\right]-\frac{T\omega_c\cdot\sin(Pm-\theta)}{\kappa\cdot M(T\omega_c)^2}\left[(T\omega_c)^2+1\right]=-\left(\frac{(T\omega_c)^2+1}{(T\omega_c)^2}\right)$$

$$\frac{(T\omega_c)^2+1}{\kappa\cdot M(T\omega_c)^2}\left[\cos(Pm-\theta)-T\omega_c\sin(Pm-\theta)\right]=-\left(\frac{(T\omega_c)^2+1}{(T\omega_c)^2}\right)$$

$$T = \frac{\kappa \cdot M + \cos(Pm - \theta)}{\omega_c \sin(Pm - \theta)}$$

Substituting this last result in (3),

$$a = -\frac{1}{\kappa \cdot M (T\omega_c)^2}\cos(Pm - \theta)\left[(T\omega_c)^2 + 1\right] - \frac{1}{(T\omega_c)^2}$$

$$a = -\frac{\cos(Pm - \theta)}{\kappa \cdot M} - \left(\frac{\cos(Pm - \theta) + \kappa \cdot M}{\kappa \cdot M}\right)\left(\frac{1}{(T\omega_c)^2}\right)$$

$$a = -\frac{\cos(Pm - \theta)}{\kappa \cdot M} - \left(\frac{\cos(Pm - \theta) + \kappa \cdot M}{\kappa \cdot M}\right)\left(\frac{\omega_c^2 \sin^2(Pm - \theta)}{\omega_c^2\left[\kappa \cdot M + \cos(Pm - \theta)\right]^2}\right)$$

$$a = -\frac{\cos(Pm - \theta)}{\kappa \cdot M} - \left(\frac{\cos(Pm - \theta) + \kappa \cdot M}{\kappa \cdot M}\right)\left(\frac{\sin^2(Pm - \theta)}{\left[\kappa \cdot M + \cos(Pm - \theta)\right]^2}\right)$$

$$a = -\frac{\cos(Pm - \theta)}{\kappa \cdot M} - \left(\frac{\sin^2(Pm - \theta)}{\kappa \cdot M\left[\kappa \cdot M + \cos(Pm - \theta)\right]}\right)$$

$$a = -\left(\frac{(\kappa \cdot M)\cos(Pm - \theta) + \cos^2(Pm - \theta) + \sin^2(Pm - \theta)}{\kappa \cdot M\left[\kappa \cdot M + \cos(Pm - \theta)\right]}\right)$$

$$a = -\left(\frac{(\kappa \cdot M)\cos(Pm - \theta) + 1}{\kappa \cdot M\left[\kappa \cdot M + \cos(Pm - \theta)\right]}\right)$$

[note **]** For the controller to be stable it is necessary that T is positive. Additionally, in order to ensure that the system is minimum phase, the value of aT should also be positive. After some tests it was found that, using this strategy, the gain crossover frequency cannot be arbitrary chosen. In fact, the crossover frequency is restricted to the values that make the controller minimum phase.

1.2.6.2 Phase lag Controllers

A phase lag controller usually contributes for the following behavioural system changes:

- Increase of relative stability by increasing the phase margin
- Decrease the bandwidth
- Decreased the steady-state error
- Reduction of overshoot (higher zeta)

1.2.6.2.1 Design strategy: Bode plots

Consider a phase lag controller transfer function parameterized as follows:

$$K(s) = \kappa \frac{aTs+1}{Ts+1},\ a<1 \tag{78}$$

Just like what was done to the phase lead compensator, the gain constant is estimated in order to satisfy the steady-state error requirements. The parameters a and T are designed so that the required phase margin is met. For this type of controllers, the gain decreases with increasing frequency and the maximum gain reduction is $20\cdot\log_{10}(a)$.

[proof **]**

$$\lim_{s\to\infty}\frac{aTs+1}{Ts+1} = a \Rightarrow 20\cdot\log_{10}(a)\big|_{dB}$$

Usually one defines, for design sake, that the minimum lag controller phase contribution occurs a decade ahead of the zero location, i.e.

$$\omega_{\min\phi} = 10\cdot\omega_z \tag{79}$$

where

$$\omega_z = \frac{1}{aT} \tag{80}$$

leading to,

$$\omega_{\min\phi} = \frac{10}{aT} \tag{81}$$

Thus, once a is selected, the variable T is chosen so that the zero crossover frequency is away (toward the Bode diagram's left) from the system critical frequency (otherwise the additional phase lag can destabilize the system).

Just like for the phase lead controller, below is presented a set of steps that can be followed in order to design a lag controller [1]:

Algorithm:

Step 1 of 6: Calculate the gain κ so that the error constant has the desired value

Step 2 of 6: Sketch the Bode plot of $\kappa G(j\omega)$

Step 3 of 6: If the phase margin is insufficient, one must estimate the

frequency value at which the phase margin is satisfied (add 5º for safety). This frequency will be the desired gain crossover frequency (ω_{gc})

Step 4 of 6: Find the gain $\mathrm{P} = \left|\kappa G(j\omega)\right|_{dB}$ at the frequency $\omega = \omega_{gc}$. Compute a from $a = 10^{-\mathrm{P}/20}$.

[note **]** The practical limit to a is 0.1. To add more phase lag is necessary to cascade compensators.

Step 5 of 6: To minimize the controller phase contribution one must estimate T by $T = \dfrac{10}{a\omega_{gc}}$.

Step 6 of 6: Confirm the final design by drawing the Bode plot of $K(j\omega)G(j\omega)$. Also one must simulate the system closed-loop response.

1.2.6.2.2 Design strategy: Analytically

Another tuning strategy can be obtained by using some closed-form expressions just like it was done for the PID controller and for the phase lead controller. In fact, the strategy behind this method has much in common with the analytical technique used in the lead compensator design. Thus, considering a system with transfer function $G(s)$ in series with a phase lag controller with transfer function:

$$K(s) = \kappa \frac{aTs+1}{Ts+1} \text{ with } a<1 \tag{82}$$

and assuming that the system must have, in closed loop, a steady-state error less than or equal to δ, a bandwidth ω_c and a phase margin of Pm degrees than:

$$a = -\left(\frac{(\kappa \cdot M)\cos(Pm-\theta)+1}{\kappa \cdot M\left[\kappa \cdot M + \cos(Pm-\theta)\right]} \right) \tag{83}$$

and

$$T = \frac{\kappa \cdot M + \cos(Pm-\theta)}{\omega_c \sin(Pm-\theta)} \tag{84}$$

where M and θ refer to the gain and phase (in degrees) that the system displays at frequency ω_c. The gain κ is obtained by the maximum allowable steady state error.

[note **]** Since the analytical technique the poles and zeros time constants are obtained through the division by $\sin(Pm-\theta)$, this method does not work if the sine argument approaches 180º. Hence, it is possible that a given set of performance criteria are not attainable with this method. In my point-of-view the sine argument must not be greater than, or equal to, 180º since, in this case, the sine function returns a negative number or zero. With negative values the controller are unstable or non-minimum phase.

[◄ CHAPTER 1]

Chapter

2

Discrete-Time Control

2.1 Sampling and Reconstruction

DIGITAL control has to do with the replacement of the previously reviewed analog controllers by algorithms running on digital processors as computers, microcontrollers, DSP's or ASIC's. However, since the information signals typically presented in a control loop are analog (continuous in time), the addition of a digital controller requires an intermediate stage of signal discretization (A/D conversion). As we shall see later, in most cases there is also the need for reverse conversion. That is transforming a signal from the digital domain back to analog domain (an operation performed by D/A converters). The figures that follow are intended to illustrate what one has just said. The “switches” in figure 15 represent the sampling process basic devices: the samplers

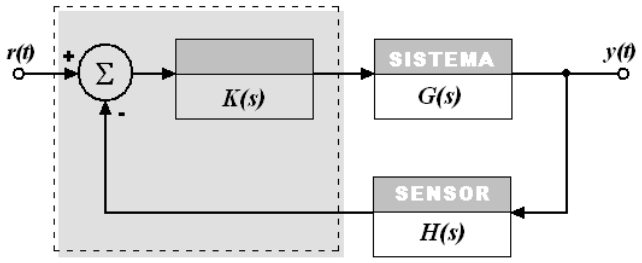


Fig 14. Continuous controller.

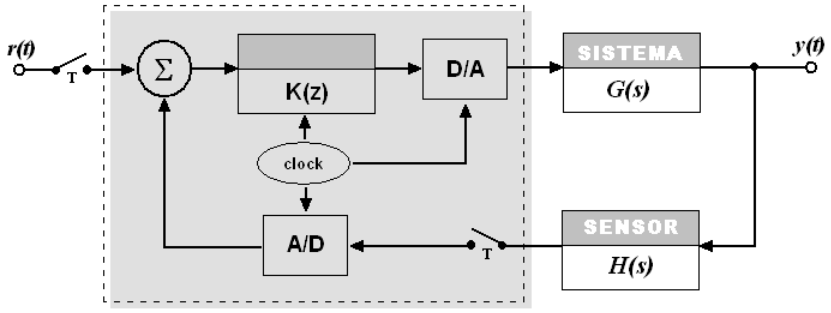


Fig 15. Digital controller.

The signal digitalization step requires:

- Signal sampling at time intervals T usually regular (it is possible to have variable sampling rates). After this process, one gets a discrete-time signal but continuous in amplitude. However, digital processors can't perform operations with infinite precision. Thus, the amplitude must also be discretized by a quantization operation.
- The quantization of an analog input, to its digital counterpart, depends on the number of binary digits used (bits). For example, using a 10 bits quantization one can represent 1024 different levels. The approximation resolution is equal to one part in 1024 times de conversion reference amplitude. For example a 8 bit A/D converter with minimum and maximum reference conversion values equal to -1 and +1, has a resolution of 2/256 or 1/128. If the signal amplitude does not match an integer multiple of the resolution, the quantization process drive, as output, the binary equivalent of the closest value to quantify. So the quantization process adds additional measurement error. The theoretical minimum error added is equal to half the least significant bit i.e. $\pm\frac{1}{2}LSB$. Figure 16 show the 2 bit quantization error effect for a ramp signal.

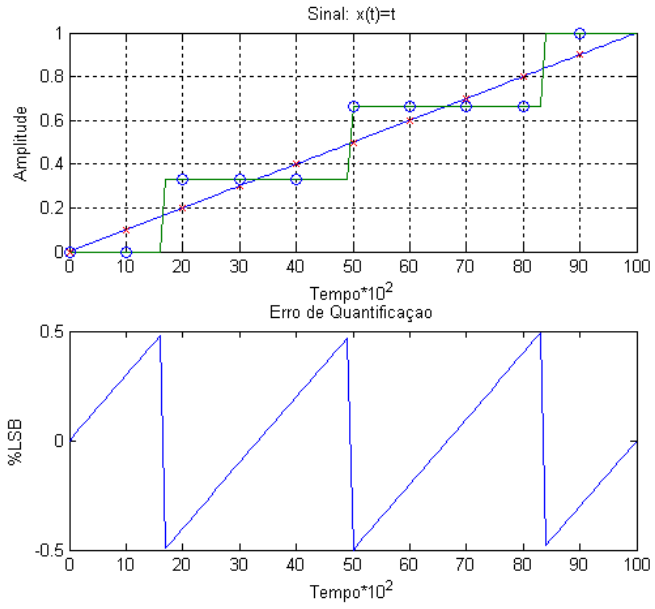


Fig 16. Quantization error due to 2 bit A / D converter.

The quantization phenomenon, as well as its control system effect, will be addressed further on, in section § 2.1.3.

Since a digital control system usually requires sampling operations, its understanding is of fundamental importance. For this reason the following section lays down the sampling process mathematical foundations. This knowledge is essential for digital control systems analysis and design.

2.1.1 Process Sampling[4]

Whenever a digital processor is involved, for measurement, signal processing or control, the data and systems involved are, in their nature, discrete-time. In this section we are particularly interested in discrete-time signals obtained by sampling, in time, an analog[5] signal. This section purpose is to establish a mathematical model for the sampling process. This model will be useful in order to take into consideration possible closed-loop dynamics changes (compared to the analog one). In this framework consider the following figure:

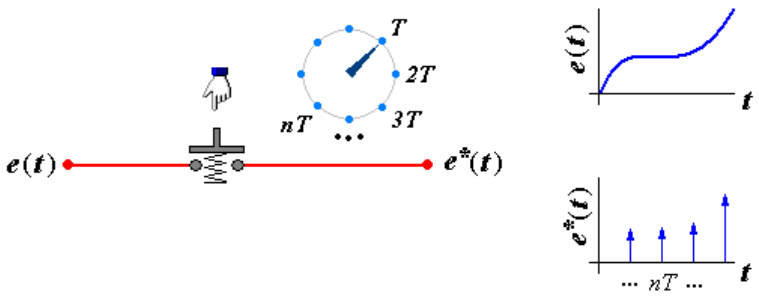


Fig 17. Figurative model of the ideal sampling process.

Let's imagine an analog electrical signal $e(t)$ (for example a voltage) applied upstream to the previous figure switch. Consider also that the switch is pressed in regular time intervals $0,T,\cdots,nT,\cdots,\forall n\in\mathbb{N}_0$ and during an infinitely small instant. At the switch downstream one predict the appearance of the theoretical signal with the appearance presented in the figure 18.

[Note]	This is an ideal model of sampling since the output signal is composed by a sequence of (non-physical) impulses (symbolized by arrows). In the real world there are no impulses but short duration pulses [12].

One can see that the sampled signal is a weighted impulse sequence (impulse *train*). The weighting factor is not more than the signal amplitude at each sampling time nT. Thus, the sampled signal can be written as a weighted sum

[4] A previous study to Annex A2 is advised.
5 An analog signal is a continuous signal both in time and amplitude.

of shifted in time impulses:

$$e^*(t) = e(0)\delta(t) + e(1)\delta(t-T) + \cdots + e(nT)\delta(t-nT) + \cdots = e(t)\cdot\sum_{n=0}^{+\infty}\delta(t-nT) \qquad (85)$$

[note **]** The function $\delta(t)$, designated by impulse or Dirac delta *r*epresents a theoretical signal without physical existence. Conceptually this signal describes a pulse with infinitely small duration and infinitely high amplitude.

$$\delta(t) = \begin{cases} 0 \text{ , if } t \neq 0 \\ \infty \text{ , if } t = 0 \end{cases}$$

This signal also admits representation in the discrete-time domain. In this case the signal are physically realisable and has the following formulation:

$$\delta[n] = \begin{cases} 0 \text{ , if } n \neq 0 \\ 1 \text{ , if } n = 0 \end{cases}$$

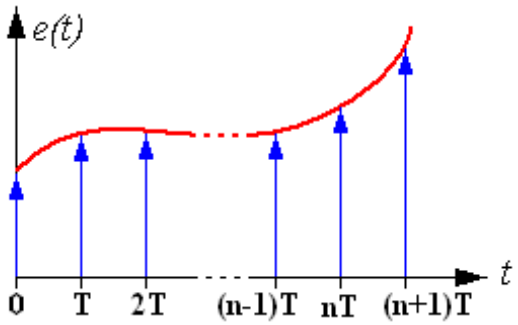


Fig 18. Relationship between the upstream and downstream signals of the ideal sampler.

From the previous expression the sampling process can be viewed as the product of a periodic impulse sequence with period T by the sampled signal $e(t)$. In other words we are witnessing an amplitude modulation strategy where the carrier is the impulse train and the modulating signal is $e(t)$. This concept is illustrated in the following figure [12].

Let us now consider the effect of signal sampling in the frequency domain. In order to do this let's apply the Fourier transform to the sampled signal $e^*(t)$:

$$e^*(t) = e(t) \cdot \sum_{n=0}^{+\infty} \delta(t - nT) \overset{F}{\rightarrow} E^*(j\omega) = \frac{1}{2\pi} E(j\omega) * \Delta(j\omega) \tag{86}$$

where $E(j\omega)$ refers to the Fourier transform of $e(t)$ and $\Delta(j\omega)$ is the Fourier transform of the impulse *train*.

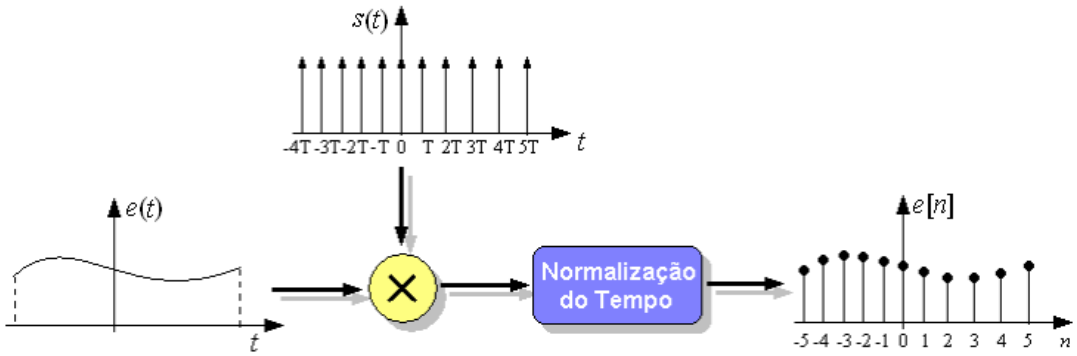


Fig 19. Sampling seen as an amplitude modulation.

[note]

$$r(t) = s(t) \cdot p(t) \overset{F}{\rightarrow} R(j\omega) = \frac{1}{2\pi}\left[S(j\omega) * P(j\omega)\right]$$

$$R(j\omega) = S(j\omega) * P(j\omega) \overset{F^{-1}}{\rightarrow} r(t) = s(t) \cdot p(t)$$

Since the impulse sequence is periodic with period T, its Fourier transform is:

$$\Delta(j\omega) = \sum_{n=-\infty}^{+\infty} 2\pi \cdot C_k \cdot \delta(\omega - k\omega_o) \tag{87}$$

where

$$C_k = \frac{1}{T} \int_{-T/2}^{T/2} \delta(t) \cdot e^{-jk\omega_o t} dt = \frac{1}{T} \tag{88}$$

[note] Sifting property [4]

$$\int_{-\infty}^{+\infty} f(t)\delta(t - t_o)dt = f(t_o)$$

$$\int_{a}^{b} f(t)\delta(t - t_o)dt = \begin{cases} f(t_o) & \text{se } a \le t_o \le b \\ 0 & \text{remain cases} \end{cases}$$

Then,

$$\Delta(j\omega) = \frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta(\omega - k\omega_o) \tag{89}$$

which leads to the conclusion that, in the Fourier domain, an impulse *train* in the

time domain is also an impulse *train* but in the frequency domain. In the frequency domain the impulses appear spaced by $\omega_o = 2\pi/T$ and weighted by a constant factor equal to $2\pi/T$.

As already said, the sampling signal frequency spectrum can be obtained by the following operation:

$$E^*(j\omega) = \frac{1}{2\pi} E(j\omega) * \Delta(j\omega) \tag{90}$$

The convolution between the two spectra, $E(j\omega)$ and $\Delta(j\omega)$, is calculated using the convolution integral as follows:

$$E^*(j\omega) = \frac{1}{2\pi}\left[\int_{-\infty}^{+\infty} E(j\Omega)\cdot\Delta(j(\omega-\Omega))\,d\Omega\right] \tag{91}$$

Substituting $\Delta(j\omega)$ by $\frac{2\pi}{T}\sum_{k=-\infty}^{+\infty}\delta(\omega-k\omega_o)$ we have:

$$E^*(j\omega) = \frac{1}{2\pi}\left[\int_{-\infty}^{+\infty} E(j\Omega)\cdot\frac{2\pi}{T}\sum_{k=-\infty}^{+\infty}\delta(\omega-k\omega_o-\Omega)\,d\Omega\right] \tag{92}$$

i.e.,

$$E^*(j\omega) = \frac{1}{T}\left[\int_{-\infty}^{+\infty} E(j\Omega)\cdot\sum_{k=-\infty}^{+\infty}\delta(\omega-k\omega_o-\Omega)\,d\Omega\right]$$

$$E^*(j\omega) = \frac{1}{T}\left[\int_{-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} E(j\Omega)\cdot\delta(\omega-k\omega_o-\Omega)\,d\Omega\right]$$

Since the integral of the sum is equal to the sum of integrals one get,

$$E^*(j\omega) = \frac{1}{T}\left[\sum_{k=-\infty}^{+\infty}\int_{-\infty}^{+\infty} E(j\Omega)\cdot\delta(\omega-k\omega_o-\Omega)\,d\Omega\right]$$

By the sifting property, and because $\delta(\omega-k\omega_o-\Omega)$ is only nonzero for $\omega-k\omega_o-\Omega=0$ or. $\Omega=\omega-k\omega_o$, one get,

$$E^*(j\omega) = \frac{1}{T}\sum_{k=-\infty}^{+\infty} E(j(\omega-k\omega_o)) \tag{93}$$

This last expression means that the sampled signal spectrum is periodic in frequency with fundamental period ω_o. Effectively the sampled signal spectrum is equal to the continuous signal spectrum repeated indefinitely with a period that depends on T. Additionally the spectrum energy of the sampled signal is T

times lower than the continuous signal spectrum.

In order to illustrate the sampling effects in the frequency domain consider an arbitrary signal $e(t)$, band limited, whose spectrum has the following generic profile:

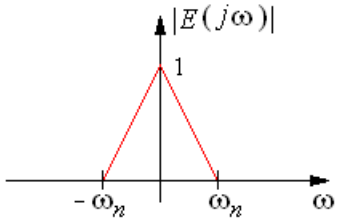


Fig 20. Frequency spectrum (magnitude) of a generic signal.

In graphical terms, the expression (92) represents the overlap of $|E(j\omega)|$ replicas shifted in frequency by $k\omega_o$ and amplitude scaled by 1/T. Figure 21 show the magnitude of $|E^*(j\omega)|$.

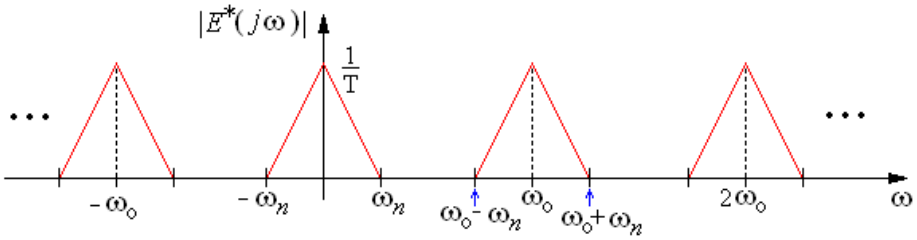


Fig 21. Frequency spectrum after sampling the signal.

From the previous figure, one can presume that it's possible to reconstruct the continuous-time signal from its sampled version. To do this we just have to *eliminate* the sampled signal spectral components above and below $\pm\omega_n$. This operation can be performed by a low-pass filter.

It thus appears that, theoretically, it's possible to obtain the continuous-time signal from the sampled one by filtering. However, in order to be possible, two conditions must be fulfilled. The first refers to how the signal is filtered and the second to how the spectrum is distributed. Regarding the former, this will be the study subject in section § 2.1.4.

Concerning the second condition, and observing figure 21, one concludes that, to be possible the original signal recovery, overlap of adjacent bands between the replicas is not allowed. Since the relative position between adjacent spectra

depends on the sampling frequency ω_o then a necessary condition for the invertibility of the sampling operation that $\omega_n < \omega_o - \omega_n$, or $\omega_o > 2\omega_n$. This condition is known as the Nyquist theorem (sometimes Shannon's theorem). It establishes that the sampling rate must be greater than twice the maximum signal frequency component (with significant amplitude) to be sampled.

[note **]** The frequency equal to half the sampling frequency is called the Nyquist frequency. This convention will be followed in the course of this document.

If this condition is not fulfilled, a time phenomenon known as *aliasing*, occurs.. In this case one witness a distortion where the signal frequency components, greater than half the sampling frequency, are translated to the limited interval $[-\omega_o/2, \omega_o/2]$. The effect of continuous signals under-sampling will be subject to further analysis in the following section.

Just to conclude one must warn that, in the discrete domain, the signal spectrum is often represented using a frequency axis normalization by a factor equal to the sampling period. Thus the axis ω becomes the axis ω_d and the relationship between them can be expressed by the following equality:

$$\omega_d = \omega T \tag{94}$$

The frequency ω_d is usually called the “digital frequency” and, as can be inferred from the above expression, this frequency does not have the explicit notion of time. Then it’s measured in radians per sample [12]. By using this frequency normalization figure 21 is replaced by the following alternative representation:

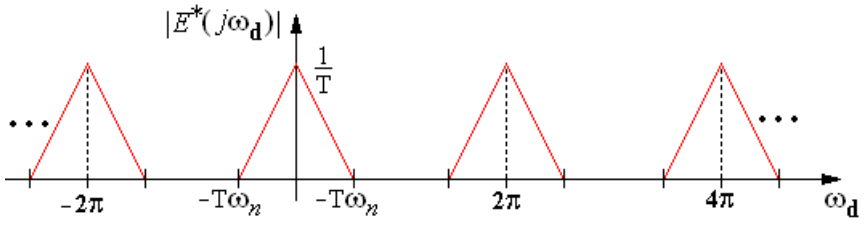


Fig 22. Frequency spectrum as a function of digital frequency.

So, in the digital domain, the sampling frequency is equal to 2π. In fact, replacing in expression (93) ω by $\omega_o = 2\pi/T$ we obtain $\omega_d = 2\pi$. Additionally, taking into account the previous figure, the relationship between the Fourier transform of continuous signal and its sampled version is:

$$E(j\omega) = T \cdot E^*(j\omega_d) \text{ to } -\pi < \omega_d < \pi \tag{95}$$

where π represents half the sampling frequency (the Nyquist frequency).

2.1.2 Sampling distortion aspects

As mentioned earlier, a strange phenomenon can arise when a continuous signal is sampled in time: high frequency components of the analog signal may appear as low frequency components (but with unchanged amplitude) in the discrete-time signal. This phenomenon is called *aliasing* and occurs whenever the sampling frequency is less than twice the maximum frequency component of the sampled signal.

In order to illustrate what was just said, consider the example of a simple monochrome signal, $x(t) = \sin(4\pi t)$, sampled at two different rates 2.5 Hz and 10 Hz. The results can be evaluated by visual inspection to figure 23 (the markers represent acquired signal samples).

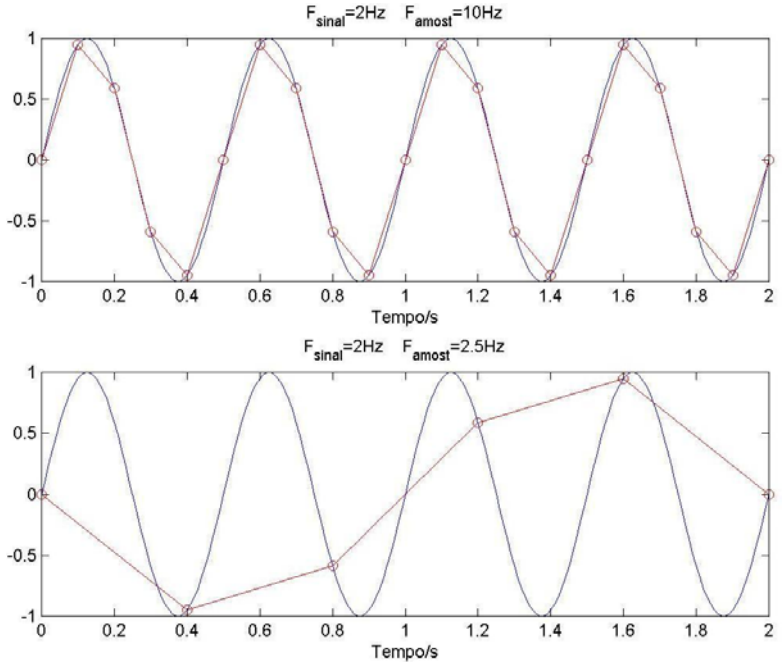


Fig 23. Sampling frequency effect: aliasing example.

Note that, contrary to what may seem at first glance, only the samples are part of the discrete signal. However, for better visual perception of the *aliasing* phenomenon, the markers are connected by line segments. In reality what one are doing, when joining the dots with line segments, is a sampled signal reconstruction using a 1st order linear interpolation:

$$x(t) = \frac{x[k+1] - x[k]}{T} t + x[k] \quad kT \le t \le (k+1)T$$

From the previous figure it is also clear that the discrete signal, that has been undersampled, seems to have a frequency lower than the frequency of the analog signal. Furthermore it is found that the digital signal frequency (after reconstruction) is equal to 0.5Hz!

In general, the value of a given frequency component, of a undersampled signal, can be obtained by [1]:

$$\omega_{alias} = \left| \left(\left(\omega + \frac{\omega_o}{2} \right) \% \omega_o \right) - \frac{\omega_o}{2} \right| \tag{96}$$

where the % operation refers to the remain after division.

Another alternative is to recurrently subtract the sampling frequency component signal until the result is smaller, in magnitude, than the Nyquist frequency. At this point the resulting frequency is the signal apparent frequency. This procedure can be summarized by the following equation,

$$\omega_{alias} = \left| \omega - \omega_o \left\lfloor \frac{2\omega + \omega_o}{2\omega_o} \right\rfloor \right| \tag{97}$$

where the operator $\lfloor \cdot \rfloor$ rounds its argument toward zero.

For example, if $\omega_o = 2.5$ and $\omega = 26$ this means that, using equation (96),

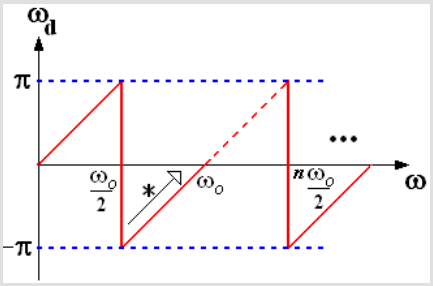
$$\omega_{alias} = \left| (27.25 \% 2.5) - 1.25 \right| = \left| 2.25 - 1.25 \right| = 1 \text{ rad/s}$$

On the other hand, by subsequent subtractions one gets,

$$\omega_{alias} = 26 - 2.5 - 2.5 - \cdots - 2.5 = 26 - 10 \cdot 2.5 = 1 \text{ rad/s}$$

[note] In control systems, aliasing causes another more subtle problem. In general, in continuous-time control system, signals contaminated with high-frequency noise outside the control system bandwidth, don't usually affect the system response. The same cannot be said for sampled systems since frequencies above π will be folded back to the frequency range of interest.

[note] Besides the transformation from high to low frequencies, under-sampling also has a reverse effect on the spectrum. For some frequency ranges one witnesses a decrease (increase) in frequency when the digital analog frequency increases (decreases). This phenomenon can be evidenced by observing the following figure.



For analog frequencies between $k\,\omega_0/2$ and $k\omega_0, \forall k \in \mathbb{Z}$, a frequency increased implies a decrease in module, of the digital frequency (remember that the negative sign refers to phase information only).

Apparently, the aliasing problem seems simple to fix: we sample an analog signal ensuring that the sampling frequency obeys the Nyquist theorem.

However things are not so simple. This is because frequency spectrum of a real analog signal never ends abruptly at a given frequency. It's not possible to acquire an analog physical signal which is band limited in bandwidth (like the one shown in Figure 20). In theory, due to random noise, the spectrum of such signals extends from minus infinity to plus infinity. This implies that, independently of the chosen sampling frequency, there will always be sidebands overlapping. However, almost all of the signal energy is contained in a narrow frequencies range. Thus the signal high frequency components, conveying no information, must be eliminated or severely attenuated. A low-pass filter is usually in charge of this task. The location of this filter, within a

control system loop, is shown in figure 24.

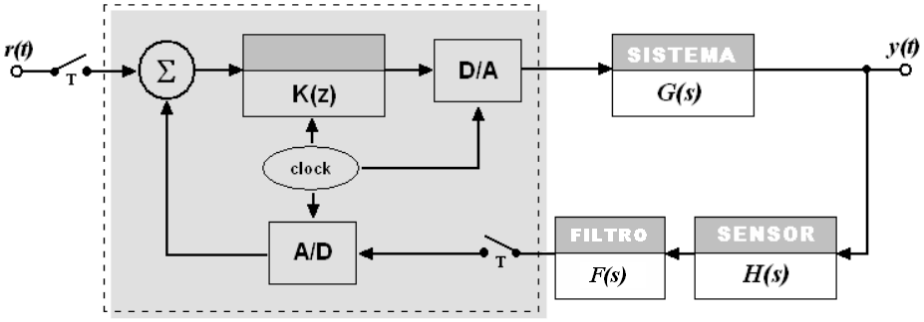


Fig 24. Reducing the effect of aliasing by introducing a pre-filter $F(s)$

Obviously this filter must be analog. Frequently it's implemented using electronic active and passive components such as operational amplifiers, capacitors and resistors.

Typically these pre-filters, inserted upstream of the sampler, are called *anti-aliasing* filters. The choice often falls to filters with a first-order type transfer function:

$$F(s) = \frac{\omega_o}{2} \cdot \frac{1}{s - (\omega_o / 2)} \tag{98}$$

However it is also common to find higher order filters as the case of Butterworth and Bessel filters. The latter have the advantage of having an almost linear phase (within the frequency range of interest) which implies low distortion of the signal profile.

Note that the bandwidth of the anti-aliasing filter is usually much higher than the bandwidth of the system. This implies that the additional dynamics introduced by the filter can be neglected in the design procedure. However the influence of the filter should be taken into account in the global simulation of the control system. This theme will be deeper analysed in section § 2.6.2.

2.1.3 Quantization

In the digital control context there are three facts usually unavoidable:

- A digital controller is based on a digital processor (computer, µC, DSP, ASIC, etc.).. This component is responsible for establishing the relationship between the control signal and the system information.

- A digital processor deals with binary coded information. Due to this fact, the arithmetic operations performed in a digital processor, always have finite precision. The operation accuracy depends on the device wordlenght.
- The control processes are usually "analog". Thus one needs a control signal decoder to translate the processor's bit string into, for example, an analog voltage.

These considerations are objectively illustrated in figures 15 or 24. Ignoring the digital processor type, there are two system blocks: an A/D converter and D/A converter. Each component performs two separate functions. The A/D converter is responsible for:

- Sampling the signal (sample & hold);
- Encode the signal. i.e. by comparing the input signal magnitude to a reference minimum and maximum threshold, it converts a given value to a *n* bit binary word.

[note **]** In practice the A/D and D/A conversion are performed, on electrical signals, using integrated circuits.

On the other hand, the D/A converter take a binary string and, from a pair of fixed limits, turn it into an analog value. Besides decoding it also performs a reconstruction operation that will be a study subject on the subsequent section.

Returning to the A/D conversion, signal encoding involves the loss of information. This is because, viewed from another perspective, a signal with an infinite number of levels is transformed into a finite number of levels signal (level *quantized*). In an A/D converter the number of quantization levels depends on the number of resolution bits and is approximately equal to $2^n - 1$. In the case of an A/D converter, with reference imposed by $\left[X_{MIN}, X_{MAX}\right]$, the quantization effect can be modelled by the following expression:

$$x_{quant}(t) = q \cdot round\left\{x(t)/q\right\} \tag{99}$$

where

$$q = \frac{X_{MAX} - X_{MIN}}{2^n - 1}$$

and n refers to the number of converter bits. Figure 25 shows the effect of an 8

bits quantization of a error signal injected to a controller. Note that the quantization performance depends on the dynamic range in which the conversion is done (the relationship between the signal amplitude limits and the A/D conversion limits).

Additionally, due to finite precision of the digital processor, a rounding or truncation error resulting from arithmetic operations should also be considered. Although the latter problem is not easily noticeable when using floating-point format (for example computations using MatLab®), but is evident in fixed-point format. For example performing multiplication operations on a 8 bit microcontroller using Q7 format.

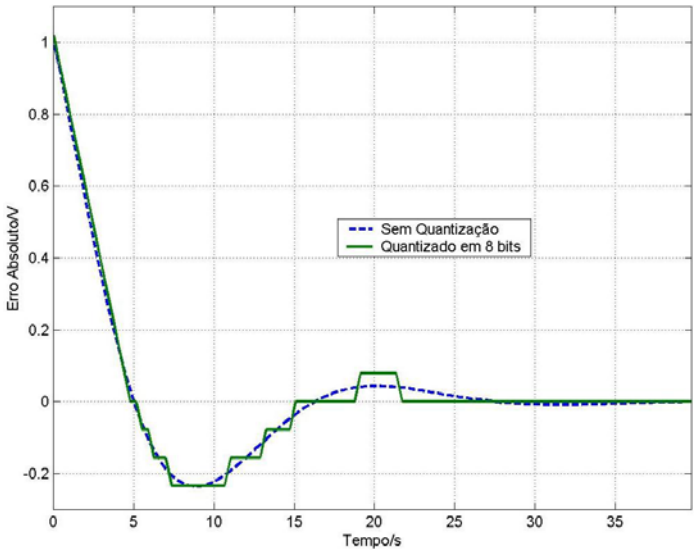


Fig 25. Effect of amplitude quantization. (Dynamic range of conversion between -10 and 10 V and 8-bit coding)

[note **]** The quantization errors introduced by 16 or 32 bits processor are usually negligible in the digital control context.

In general, in controller design procedures, the quantization effect is neglected and only examined, at the end, with computer simulation.

2.1.4 Reconstruction

This section talks about the reconstruction problem of a discrete-time signal. Strictly speaking, in practice, one are more interested in "construction" than "reconstruction." This is because the control signal is digitally and not obtained by time discretization. The term “rebuilding” is usually employed bearing in mind the recovery of a sampled signal that has been originally continuous.

At first the mathematical concepts, underlying the reconstruction of a sampled signal, are presented. Additionally we present the conditions, under which, this recovery can be attained. The reconstruction problem, in the context of digital control, is then addressed. Prosecuting this subject, the dynamic influence of the D / A converter in the control loop, will be analysed.

2.1.4.1 Ideal Reconstruction

Theoretically, if properly sampled, a continuous signal in time can be reconstructed using an ideal low-pass filter. For a convenient reconstruction is necessary that the filter has cut-off frequency $\omega_n < |\omega_c| < \omega_o/2$ and magnitude, in pass-band, equal to T as shown in the figure below.

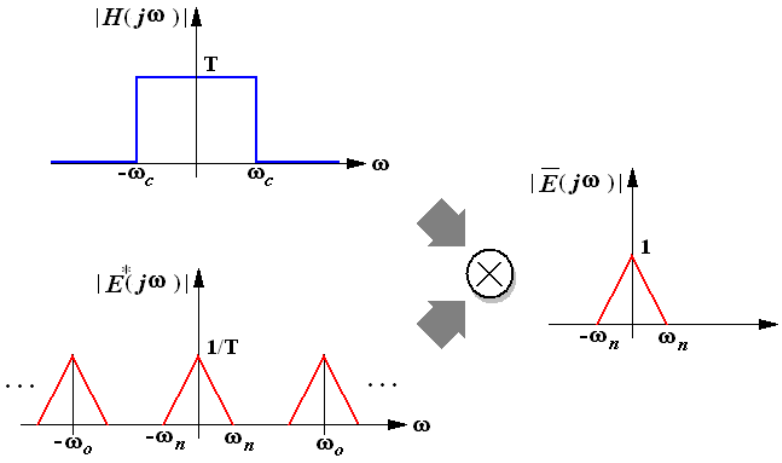


Fig 26. Reconstruction of a sampled signal (magnitude normalized)

The reconstruction sequence of a continuous time signal, from its sampled version, is the product of the sequence Fourier transforms by the ideal low-pass filter Fourier transform. In mathematical terms one write:

$$\bar{E}(j\omega) = E^*(j\omega) \cdot H(j\omega) \tag{100}$$

were $\bar{E}(j\omega)$ is the Fourier transform of the reconstructed signal $\bar{e}(t)$. In the time domain the previous relation is expressed as:

$$\bar{e}(t) = e^*(t) * h(t) \tag{101}$$

On the other hand, the impulse response of the ideal low-pass filter can be easily derived using the inverse Fourier transform definition. Thus,

$$h(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} T e^{j\omega t} d\omega$$

leading to,

$$h(t)=\frac{T}{2\pi}\cdot\left[\frac{e^{j\omega t}}{jt}\right]_{-\omega_c}^{\omega_c}=\frac{T}{2\pi jt}\cdot\left[e^{j\omega_c t}-e^{-j\omega_c t}\right]=\frac{T}{\pi t}\cdot\sin(\omega_c t)=\frac{\omega_c T}{\pi}\operatorname{sinc}(\omega_c t)$$

Considering the equation (101) and the convolution definition one get:

$$\overline{e}(t)=\sum_{k=-\infty}^{+\infty}e(kT)h(t-kT)$$

and finally,

$$\overline{e}(t)=\frac{\omega_c T}{\pi}\sum_{k=-\infty}^{+\infty}e(kT)\cdot\operatorname{sinc}\left(\omega_c(t-kT)\right)$$

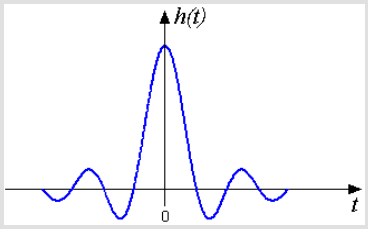
For $\omega_c=\frac{\omega_o}{2}$ and given that $T=\frac{2\pi}{\omega_o}$ the previous expression becomes:

$$\overline{e}(t)=\sum_{k=-\infty}^{+\infty}e(kT)\cdot\operatorname{sinc}\left(\frac{\omega_o}{2}(t-kT)\right) \qquad (102)$$

Although the ideal reconstruction operation may be achievable mathematically, in practice this approach start of an invalid assumption: the existence of ideal filters. In fact, and analyzing the impulse response of the filter, it turns out that it's not causal hence physically unrealisable.

[note **]**

A linear time-invariant system, with impulse response $h(t)$, is causal if, and only if, the impulse response is zero for values of $t<0$. This consideration is quite simply to understand. Consider, for example, the following impulse response:



The impulse response is the response of a system to an impulse. An impulse response, as the one showed in the previous figure, suggests that, even before a pulse is applied to the system, he begin to responds. On other words, an impulse $\delta(t)$ is applied to the filter at $t=0$ and the response system anticipates the cause that gave rise to it. Of course, in physically realizable systems, the system cannot sense something that has not yet happen.

2.1.4.2 Real Reconstruction

Usually, even in a digital control system, the control plant is continuous in time. Thus, even if a discrete signal can sometimes be used to drive system directly, this procedure is rarely used due to high frequency components of the signal injected into the actuators. Thus the decoded control signal is usually converted into a continuous-time signal.

Let us now consider how this conversion can be done. In the previous section, the conversion procedure suggested the use of a ideal low-pass analog filter. Now let's imagine the closed-loop structure illustrated by figure 27.

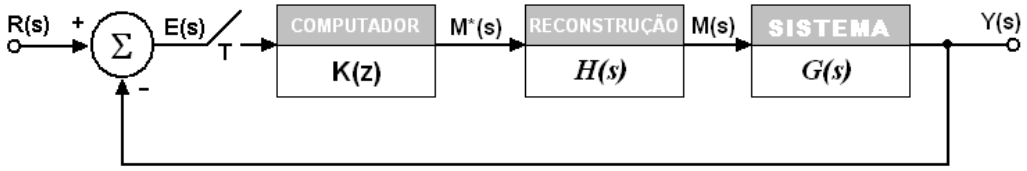


Fig 27. Microprocessor control of an analog process

At a given time instant kT the microprocessor outputs a control signal $m^*(kT)$ using some control law. This value is applied to the process and the next control signal value is applied only in the next instant $(k+1)T$.

Between discrete time instants, kT and $(k+1)T$, which control values should be applied to the system?

[note **]** In real-time control there are causality constraints. Thus it's not possible to access the control signal future values.

At the present instant, only $m^*(kT)$ and its past values are known. For this reason it is necessary, only from the known data, to **forecast** the control signal values between sampling periods. It's necessary to infer the most likely values between the present sample and the next one.

[note **]** The predictive nature of the reconstruction system is closely related to the non-causality of the ideal analog filter introduced in section § 2.1.4.1

One way to accomplish this prediction is by polynomial extrapolation [13]. From

the present and past knowledge, and using a polynomial function, one predict the more likely values of the control signal between the sampling instants kT and $(k+1)T$.

Let's see how a possible polynomial can be derived. First it's necessary to expand the signal $m(t)$, using a Taylor series, around the point $t = kT$:

$$m(t) = m(kT) + (t-kT)\cdot\left.\frac{dm(t)}{dt}\right|_{t=kT} + \cdots + \frac{(t-kT)^p}{p!}\cdot\left.\frac{d^p m(t)}{dt^p}\right|_{t=kT} \tag{103}$$

[note **]** Taylor series expansion of a function $f(x)$ around $x = a$:

$$f(x) = \sum_{k=0}^{+\infty} \frac{(x-a)^k}{k!}\cdot\left.\frac{d^k f(x)}{dx^k}\right|_{x=a}$$

The analysis of the previous expression provides some clues regarding the impossibility of knowing, with certainty, the control signal values between samples. In first place the polynomial can be infinite and, in second place, the coefficients calculation requires the knowledge of the signal derivative at point $t = kT$. The formal derivative definition is given by the following equality :

$$\left.\frac{dm(t)}{dt}\right|_{t=kT} = \lim_{h\to 0}\left.\frac{m(t+h)-m(t)}{h}\right|_{t=kT} \tag{104}$$

It's clear that the derivative operator is not causal and requires knowledge of signal values for posterior moments regarding $t = kT$. Thus, because one has only knowledge of past and present values of the command signal, in the discrete-time domain the derivative is approximated by:

$$\left.\frac{dm(t)}{dt}\right|_{t=kT} \approx \left.\frac{m(t)-m(t-T)}{T}\right|_{t=kT} \text{ i.e. } \frac{dm(kT)}{dt} \approx \frac{m(kT)-m((k-1)T)}{T} \tag{105}$$

Because it's impossible to compute an infinite number of derivatives, expression (103) must be truncated at a certain point. With this strategy higher order derivatives are ignored.

Since one can select the truncation point, one alternative is to disregard all derivatives of order higher than zero. This strategy leads to a zero order polynomial extrapolator commonly referred to as zero-order (*zero-order hold* - ***ZOH***). In this context, and since the approximation must be valid only between

sampling periods, the reconstructed signal is:

$$m(t) = m(kT) \text{ for } kT \le t < (k+1)T \tag{106}$$

With this signal model the prediction made is of *naïve* type: the model assumes that the command value does not change between samples. In the time domain the appearance of zero-order holder output signal is:

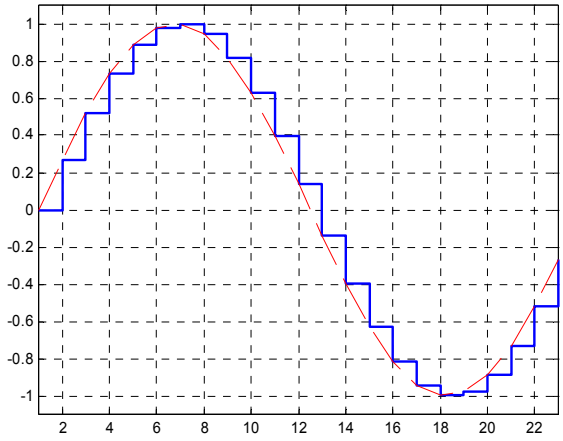


Fig 28. A discrete-time signal reconstructed by a zero-order holder

An alternative way to write the previous expression is:

$$m(t) = m(kT) \cdot \left[u(t-kT) - u(t-(k+1)T)\right] \tag{107}$$

Where $u(t)$ concerns the Heaviside function (discrete-time step function). The zero-order hold work as one input one output system. The present sample is fed to the input and the ZOH deliver, at the output and during one sample period, a control signal prediction based on this sample. If one applies an impulse to the ZOH input then he reacts with his impulse response. In mathematical terms, the ZOH impulse response is obtained by:

$$h_{zoh}(t) = \delta(kT) \cdot \left[u(t-kT) - u(t-(k+1)T)\right]$$

where $\delta(t)$ refers to the impulse function. Thus, by definition, $\delta(kT)$ is only nonzero for $k = 0$ and then the previous expression takes the following form:

$$h_{zoh}(t) = \left[u(t) - u(t-T)\right] \tag{108}$$

Were $u(t) - u(t-T)$ represents the "window function" with length *T*. Then one concludes that the zero-order hold creates a pulse for each input impulse.

The frequency response of this device can therefore be calculated by applying the Fourier transform to the previous expression or, alternatively, by assessing its Laplace transform along the $j\omega$ axis.

Under the latter strategy one gets,

$$H(s) = \frac{1}{s} - \frac{e^{-sT}}{s} = \frac{1 - e^{-sT}}{s} \tag{109}$$

Since $H(j\omega) = H(s)_{s=j\omega}$, the previous expression becomes,

$$H(j\omega) = \frac{1 - e^{-j\omega T}}{j\omega} \tag{110}$$

Or else,

$$H(j\omega) = e^{-j\omega\frac{T}{2}} \left\{ \frac{e^{j\omega\frac{T}{2}} - e^{-j\omega\frac{T}{2}}}{j\omega} \right\}$$

Taking into consideration the Euler identities the Fourier transform becomes,

$$H(j\omega) = 2\frac{e^{-j\omega\frac{T}{2}}}{\omega} \sin\left(\omega\frac{T}{2}\right)$$

in other words

$$H(j\omega) = T \cdot \text{sinc}\left(\omega\frac{T}{2}\right) \cdot e^{-j\omega\frac{T}{2}}$$

and finally, if we note that $\omega_o = \frac{2\pi}{T}$

$$H(j\omega) = T \cdot \text{sinc}\left(\pi\frac{\omega}{\omega_o}\right) \cdot e^{-j\pi\frac{\omega}{\omega_o}} \tag{111}$$

[note **]** Euler's Identities

$$\cos(\theta) = \frac{e^{j\theta} + e^{-j\theta}}{2} \text{ and } \sin(\theta) = \frac{e^{j\theta} - e^{-j\theta}}{2j}$$

A draft of the zero-order holder frequency response is illustrated in figure 29. As one can analyze by equation (111), the amplitude frequency response is a damped sinusoidal and the zero-crossings occur at integer multiples of ω_o. The spectrum amplitude is sampling period dependent and the system phase

changes linearly with the frequency. However, as can be seen in the figure below, there are multiple unwanted peaks in odd multiples of the Nyquist frequency, i.e. $(2p+1)\omega_o/2$, $p=1,2,\cdots$.

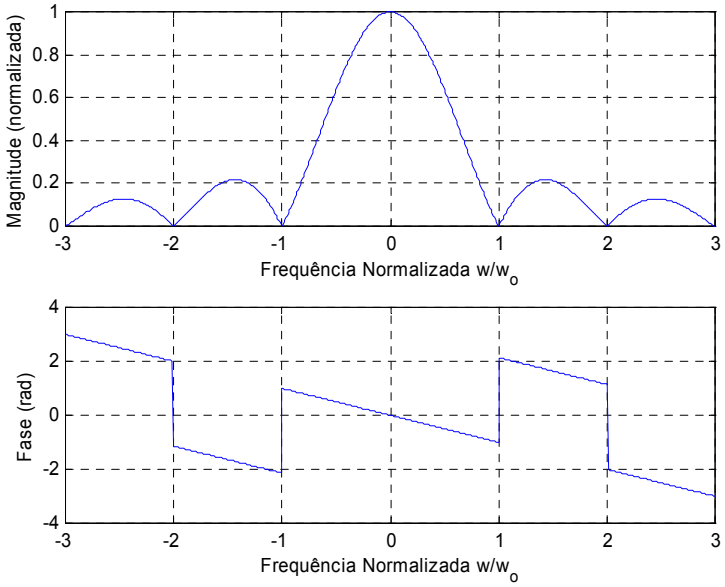


Fig 29. Frequency response of a zero-order (magnitude normalized)

Note also that, at the Nyquist frequency, the attenuation is approximately 4 dB and the gain in pass-band is not constant. Due the latter consideration, a spectrum distortion of the applied signal is observed. Additionally, from expression (111), it can be concluded that the *ZOH* performance as extrapolator strongly depends on the sampling frequency. In fact if $\omega_0 \to \infty$ then $H(j\omega) \to T$. This means that the output signal can be made arbitrarily close to the input provided that the sampling period can be made arbitrarily small.

Although other reconstruction strategies can be derived, such as first-order holders (see problem E21) the zero-order holder is, by far, the most used (a common D/A converter performs exactly this function). Therefore this subject will not be prosecuted.

2.1.4.3 Effect of the ZOH dynamics

The transfer function of a zero-order holder has the following appearance:

$$G_{zoh}(s) = \frac{1-e^{-sT}}{s} \tag{112}$$

Using this expression we draw two important conclusions. First, and given the

final-value theorem, we see that:

$$\lim_{s\to 0} G_{zoh}(s) = \lim_{s\to 0} \frac{1-e^{-sT}}{s} = \left(\frac{0}{0}\right) \tag{113}$$

Using the l'Hopital rule, the limit as s approaches zero , become

$$\lim_{s\to 0} \frac{1-e^{-sT}}{s} = \lim_{s\to 0} T\frac{e^{-sT}}{1} = T \tag{114}$$

Please recall that, at the beginning of this chapter, it was seen that the sampling process has, as a side effect, the spectrum amplitude scaling by a factor inversely proportional to the sampling period,

$$E^*(j\omega) = \frac{1}{T}E(j\omega) \text{ to } -\frac{\omega_o}{2} < \omega < \frac{\omega_o}{2} \tag{115}$$

Thus, if a zero-order is used upstream the system, then it's not necessary to adjust the controller gain (this concept will be reviewed later on when talking about the z transform).

In order to detect another details about the ZOH transfer function, one begin to expand $G_{zoh}(s)$ using a Taylor series around $T = 0$:

$$e^{-sT} = 1 - sT + \frac{1}{2}(sT)^2 - \frac{1}{6}(sT)^3 + \frac{1}{24}(sT)^4 - \frac{1}{120}(sT)^5 + \cdots$$

Substituting in equation (110) one gets,

$$G_{zoh}(s) = \frac{1-\left(1 - sT + \frac{1}{2}(sT)^2 - \frac{1}{6}(sT)^3 + \frac{1}{24}(sT)^4 - \frac{1}{120}(sT)^5 + \cdots\right)}{s}$$

and after simplification:

$$G_{zoh}(s) = T\left(1 - \frac{1}{2}(sT) + \frac{1}{6}(sT)^2 - \frac{1}{24}(sT)^3 + \frac{1}{120}(sT)^4 - \cdots\right)$$

On the other hand it is easy to show that:

$$T\cdot e^{-s\frac{T}{2}} = T\left(1 - \frac{1}{2}sT + \frac{1}{8}(sT)^2 - \frac{1}{48}(sT)^3 + \frac{1}{384}(sT)^4 - \cdots\right)$$

Now neglecting the terms of order greater than one:

$$G_{zoh}(s) \approx T\cdot e^{-s\frac{T}{2}} \tag{116}$$

In fact, as shown in the figure 30, the accuracy of approximation (116) decreases with the sampling period increase.

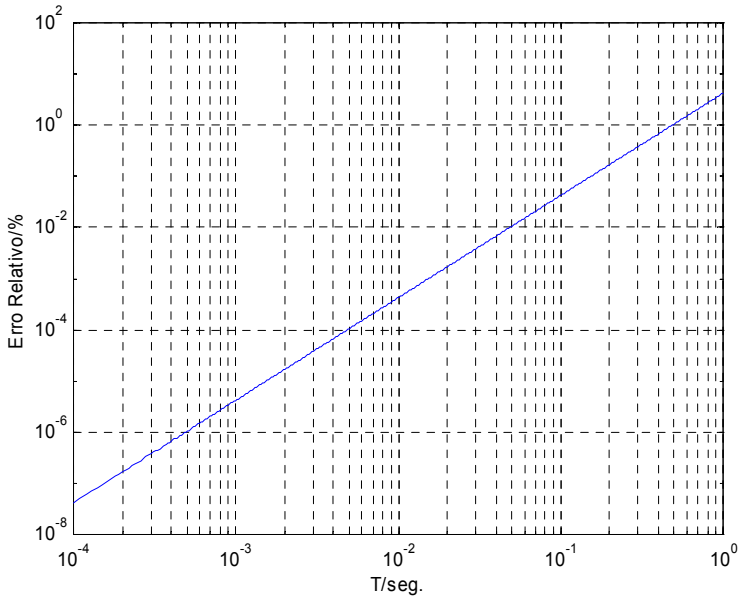


Fig 30. Relative error of approximation $G_{zoh}(s) \approx T \cdot e^{-sT/2}$

This approximation is very useful for controller design using frequency based design techniques like Bode plots. In terms of Bode diagrams, a zero-order holder contributes with a gain equal to $20 \cdot \log_{10}(T)$ and a phase lag that decreases linearly with frequency with a slope equal to half the sampling period (pure time delay). If the sampling period is large then the phase slope in acute and the ZOH has a large impact on the overall system frequency response. This is because a delay in a control loop is always cause of instability.

[note **]** As one will see further ahead, the zero-order contribute to system destabilization by reducing the system phase margin. The phase margin decrease is directly proportional to the sampling period.

When the design method is based on time-domain techniques such as the root locus, often a pure time delay of $T/2$ seconds can be crudely modelled by a first order system with the following appearance:

$$G_{zoh}(s) \approx T \frac{2/T}{s + 2/T} \tag{117}$$

An alternative way is to express the pure time delay using the Padé approximation [6]. This strategy finds a set of parameters in order to minimize the error between the exponential McLauren series expansion and an arbitrary order k transfer function. The pure delay Padé approximation is obtained by

solving the following minimization problem subject to proper transfer functions:

$$\min(\varepsilon) = \min\left\{ \Upsilon\left(e^{-sT_d}\right) - \Upsilon\left(\frac{b_1 s + b_0}{a_1 s + 1}\right) \right\} \tag{118}$$

In the above expression $\Upsilon(\cdot)$ refers to the McLauren series expansion of a function. It should be noted that the complex exponential is analytic for any finite value of s (it. has n order derivative). Thus, given that the series expansion of both components is the sum of an infinite number of terms, the minimization of (118) involves the solution of an infinite number of equations with a finite number of unknowns. Thus, for example, in case of first order approximation, the solution of the minimization problem has only three degrees of freedom thus the McLauren series expansion is made only up to order three.

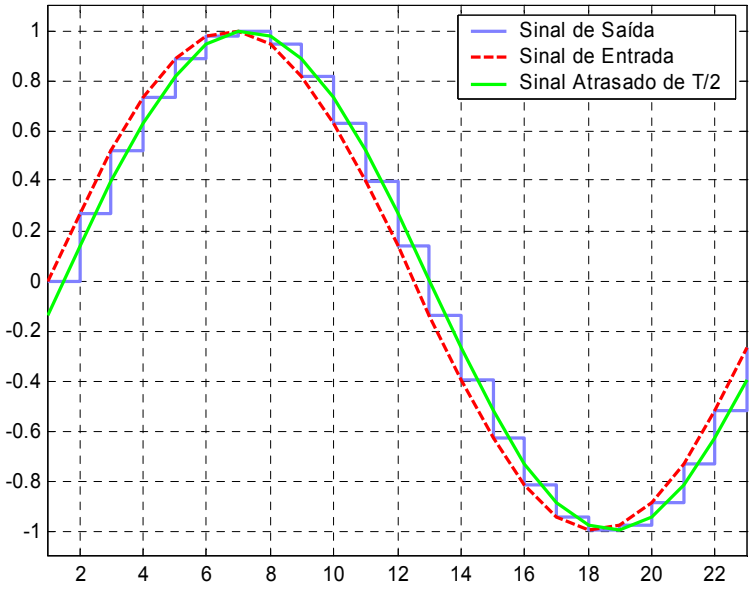


Fig 31. Response time in a zero-order and its approximate model for the analysis of digital control systems in the frequency domain

For the case of the zero-order hold, the first order Padé approximation is:

$$G_{zoh}(s) \approx T \frac{1-(T/4)s}{1+(T/4)s} \tag{119}$$

Any of the previous ZOH approximations can be used to estimate the negative impact on system stability, due to sampling. The approximation (116) is especially suitable for frequency-domain design techniques and equations (117) or (119) for time-domain techniques. Figure 31 illustrates the quality of approximation by equation (116).

2.2 The starred transform and the Z transform

The sampler-and-hold operation, illustrated in the figure below, can be analyzed by taking into consideration the waveform of Figure 33.

Fig 32. Block diagram of a sampler / zero-order

Fig 33. Generic example of a signal applied upstream of the system of Figure 32 and its output signal.

One observes that $m(t) = e(t)\big|_{t=kT}$ and, in between sampling instants, the reconstructed signal can be written as:

$$m(t) = \sum_{k=0}^{+\infty} e(kT) \cdot \{u(t-kT) - u(t-(k+1)T)\} \tag{120}$$

Now applying the Laplace transform one obtains,

$$M(s) = \sum_{k=0}^{+\infty} e(kT) \left\{ \frac{e^{-kTs}}{s} - \frac{e^{-(k+1)Ts}}{s} \right\}$$

$$M(s) = \sum_{k=0}^{+\infty} e(kT) \cdot e^{-kTs} \cdot \left\{ \frac{1-e^{-Ts}}{s} \right\}$$

By putting $\frac{1-e^{-Ts}}{s}$ in evidence becomes:

$$M(s) = \frac{1-e^{-Ts}}{s} \cdot \sum_{k=0}^{+\infty} e(kT) \cdot e^{-kTs} \tag{121}$$

As one can see, the factored term is independent of the input signal $e(t)$ and

corresponds to an already familiar element: the transfer function of the zero-order holder.

On the other hand, the factor

$$\sum_{k=0}^{+\infty} e(kT)\cdot e^{-kTs}$$

depends on both the sampling period and the input signal. Represents in the time domain, a weighted sum of pulses shifted in time, i.e. the product of the input signal $e(t)$ by a periodic pulse train with period T. Thus, this second factor, represents the action of the ideal sampler and is defined, as already introduced in section § 2.1.1, as $E^*(s)$.

Due to what was above said, the output signal of an ideal sampler is defined as the signal whose Laplace transform is:

$$E^*(s) = \sum_{k=0}^{+\infty} e(kT)\cdot e^{-kTs} \tag{122}$$

The $E^*(s)$ is usually called the starred transform of $E(s)$.

[note **]** If $e(t)$ is discontinuous in $t = kT$ then $e(kT)$ is taken into $e(kT^+)$. In other words, the value that $e(t)$ takes when t approaches kT by right hand values.

On the other hand, even without explicit intention, the stared transform was already defined in section § 2.1.1 on a different perspective. It was then demonstrated that an alternative definition for a starred transform would be:

$$E^*(j\omega) = \frac{1}{T}\sum_{k=-\infty}^{+\infty} E(j(\omega - k\omega_o))$$

which, in the Laplace domain, has the following aspect:

$$E^*(s) = \frac{1}{T}\sum_{k=-\infty}^{+\infty} E(s - jk\omega_o) \tag{123}$$

This expression allows us to conclude one of the first starred transform properties:

The starred transformed $E^*(s)$ is periodic in s with period $j\omega_o$ where ω_o refers to the sampling period.

This conclusion can be strengthened if one considers that in section § 2.2.1 it was shown that $E^*(j\omega)$ is no more than (unless a scale factor) the signal spectrum $E(j\omega)$ repeated from ω_o to ω_o.

Another fundamental starred transform property states:

If $E(s)$ has a pole at $s=a$ then his star has transformed into an infinite number of poles located, according to the plane s In $s=a-jk\omega_o$ for any $k\in\mathbb{Z}$.

Note that the same cannot be said about the zeros. Thus, even if they check the first property, normally the zeros are not mapped into the s plane in the same way as the poles as we shall see later with an example.

Suppose a second order system with a pair of complex conjugate poles with the following format:

$$E(s)=\frac{1}{(s+\sigma-j\omega)(s+\sigma+j\omega)},\ \forall\sigma\in\mathbb{R}^+,\forall\omega\in\mathbb{R}$$

If $\omega<\omega_o/2$ the poles of $E^*(s)$ are $s=-\sigma\pm j(\omega-k\omega_o),\forall k\in\mathbb{Z}$ and the poles map has the aspect illustrated in figure 34.

[note **]** Usually the frequency range between $[-\omega_o/2,\omega_o/2]$ is designated by primary strip and the remains by complementary strips.

The question that now arises is:

A different pole-zero map for $E^*(s)$ is obtained for every different poles constellation of $E(s)$?

No. If we consider, for example, $\omega=\omega_o/4$ then the $E^*(s)$ map of poles, regarding this modes, is exactly identical to the map that one would obtain if $\omega=3\omega_o/4$. In general one can say that any $E(s)$ pole located in $s=-\sigma\pm j(\omega-k\omega_o)$ result in a identical map of poles for $E^*(s)$. This statement can be seen, bellow, in figure 35.

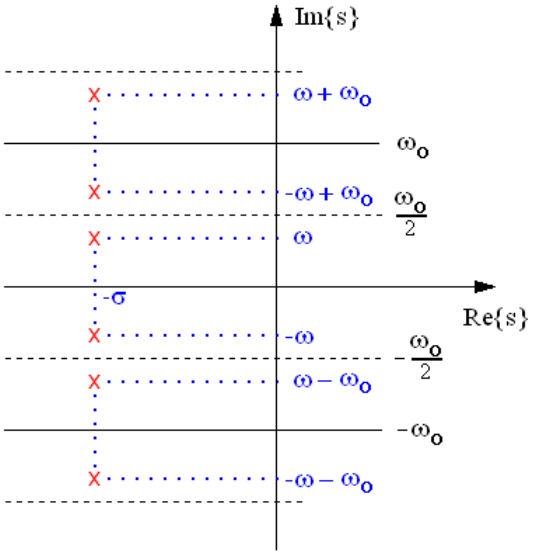


Fig 34. Map of poles of the star transformed into a system of second order under-damped.

Now taking into consideration only the primary strip, and if one apply the inverse Laplace transform, we observe that, while the signal for the leftmost s plane maintain its frequency, the rightmost one will show a lower frequency. What has been said lead us to the problem of frequency aliasing distortion (see § 2.1.2). Effectively, in the second case, the signal modes are above the sampling frequency. Thus, eliminating the secondary strips (by the filtering process), one observe the appearance of a lower frequency signal.

2.2.1 Evaluation of E * (s) in closed form

The starred transform format presented earlier (equations (122) and (123)) has an algebraic limited applicability (usually restricted only to time series). An alternative way to calculate a system starred transform, if the Lapace transform is known, results from applying the following equation [13].

$$E^*(s) = \sum_{\substack{\text{at the poles} \\ \text{of E}(\lambda)}} \text{Res}\left\{ E(\lambda) \cdot \frac{1}{1 - e^{-T(s-\lambda)}} \right\} \tag{124}$$

where the operator $\text{Res}\{\cdot\}$ refers to the residues of the argument expression. The calculation of the residues associated with each of the poles follows one of the two possibilities:

- The system has a simple pole at $s = a$,

$$\text{Res}\left\{\frac{E(\lambda)}{1-e^{-T(s-\lambda)}}\right\}\Bigg|_{\lambda=a} = (\lambda-a)\frac{E(\lambda)}{1-e^{-T(s-\lambda)}}\Bigg|_{\lambda=a}$$

- The system has multiple poles at $s=a$ with multiplicity m,

$$\text{Res}\left\{\frac{E(\lambda)}{1-e^{-T(s-\lambda)}}\right\}\Bigg|_{\lambda=a} = \frac{1}{(m-1)!}\left(\frac{d^{m-1}}{d\lambda^{m-1}}\left[(\lambda-a)^m\frac{E(\lambda)}{1-e^{-T(s-\lambda)}}\right]\right)_{\lambda=a}$$

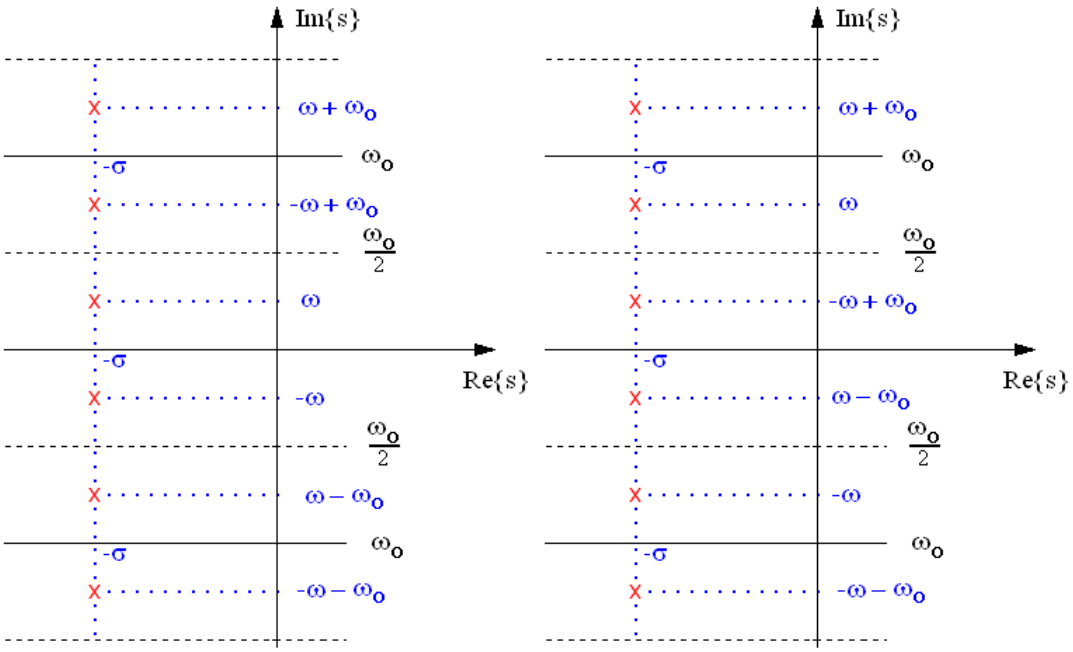


Fig 35. Map of poles for the same continuous system sampled at different rates.

In cases where the system transfer function includes a pure time-delay (an integer number of the sampling period),

$$E(s) = e^{\pm kTs}\cdot E'(s),\ \forall k\in\mathbb{Z}$$

then

$$E^*(s) = e^{\pm kTs}\sum_{\substack{\text{nos pólos}\\ \text{de } E'(\lambda)}}\text{Res}\left\{E'(\lambda)\cdot\left(1-e^{-T(s-\lambda)}\right)^{-1}\right\}$$

A few paragraphs ago, when talk about the properties of the starred transform, one said that the continuous transfer function zeros were not mapped in the same way as the poles. This statement can be validated through the following minimum phase system:

$$E(s) = \frac{s+a}{s+b}$$

By the residue theorem one obtain,

$$E^*(s) = \lambda + b\frac{\lambda + a}{\lambda + b}\frac{1}{1 - e^{-T(s-\lambda)}}\bigg|_{\lambda=-b}$$

which leads to:

$$E^*(s) = \frac{(a-b)}{1 - e^{-Ts}\cdot e^{-Tb}}$$

The $E^*(s)$ pole is located at:

$$1 - e^{-Ts}\cdot e^{-Tb} = 0 \Rightarrow e^{-Ts} = e^{Tb} \Rightarrow s = -b$$

However, the complex exponential function is periodic with period $2k\pi$ for all $k \in \mathbb{Z}$ (just look at Euler's formula!). Thus,

$$1 - e^{-Ts}\cdot e^{-Tb} = 0 \Leftrightarrow 1 - e^{-T(s + j2k\pi T^{-1})}\cdot e^{Tb} = 0$$

Since $\omega_o = 2\pi/T$

$$1 - e^{-T(s + jk\omega_o)}\cdot e^{Tb} = 0$$

which leads to $s = -b - jk\omega_o$. Thus, as predicted, the poles within the primary strip have the same location as the $E(s)$ poles. On the other hand, while $E(s)$ has a finite zero $E^*(s)$ has no zeros!

2.2.2 The Z transform

As we shall see later, the starred transformed is a useful tool for discrete time systems analysis. However, the transfer function of a sampled system, unlike the continuous systems counterpart, does not appear as a ratio of polynomials (note the example of the complex exponential in the previous section). Moreover, and recalling the aspect of a sampled system poles-zeros map, we find that these are infinite in number which does not help when using the singularities location for system analysis. Thus, an alternative strategy is presented.

This new strategy is nothing more than a variable swap: e^{sT} in $E^*(s)$ is replaced by the variable z. With this procedure the sampled system transfer function can be written as a z polynomial ratio. This transformation is appropriately designated by z transform and, regarding the starred transform, can be described mathematically as:

$$E(z) = E^*(s)\big|_{z=e^{sT}} \tag{125}$$

Equations (122) and (124) can take a different form by expressing them as z functions:

$$E(z)=\sum_{k=-\infty}^{+\infty} e(kT)\cdot z^{-k} \quad \text{(Bilateral transform)} \tag{126}$$

$$E(z)=\sum_{k=0}^{+\infty} e(kT)\cdot z^{-k} \quad \text{(Unilateral transform)}$$

$$E(z)=\sum_{\substack{nos\ pólos\\ de\,E(\lambda)}} \operatorname{Res}\left\{E(\lambda)\cdot\left(1-z^{-1}e^{T\lambda}\right)^{-1}\right\} \tag{127}$$

Since the z transformed was derived from the Laplace transform, it inherits many of its features. One of them is the concept of convergence region. In this case the convergence is assured if $|E(z)|<\infty$, that is.

$$|E(z)|=\left|\sum_{k=-\infty}^{\infty} e(kT)z^{-k}\right|\le\sum_{k=-\infty}^{\infty}|e(kT)|\left|z^{-1}\right|^{k}<\infty \tag{128}$$

The set of values for which the z transform converges is called the convergence region. As we shall see, graphically the convergence region consists on a ring in the plane z cantered at the origin who's upper and lower limits can be a circle or extend to infinity.

[note **]** Remember that the Laplace transform is for continuous-time the same way as the z transform is for discrete-time systems

A summary of z transform properties is presented in the following text box. This subject is further explored in detail in [12] and [14].

2.2.3 Modified Z Transform

In section § 2.2.1 when talked about systems with pure time delays, integer multiples of the sampling frequency, it was said that they admit starred transform representation as:

$$E^{*}(s)=e^{\pm skT}\sum_{\substack{nos\ pólos\\ de\,E'(\lambda)}} \operatorname{Res}\left\{E'(\lambda)\cdot\left(1-e^{-T(s-\lambda)}\right)^{-1}\right\} \tag{129}$$

where $E'(s)$ refers only to the polynomial transfer function component. Likewise, and taking into consideration what was said in section § 2.2.2, the

previous equation is replaced by the following z shaped one:

$$E(z) = z^{\pm k} \sum_{\substack{\text{nos pólos} \\ \text{de } E'(\lambda)}} \text{Res}\left\{ E'(\lambda) \cdot \left(1 - z^{-1} e^{T\lambda}\right)^{-1} \right\} \tag{130}$$

But how to find a system z transform if the pure time delay is not an integer multiple of the sampling period? For example the system

$$E(s) = \frac{e^{-1.2s}}{s+1}$$

sampled at a rate $T = 0.013$?

[properties **]**

Linearity

$$\mathcal{Z}\left\{ a \cdot e(kT) + b \cdot f(kT) \right\} = a \cdot E(z) + b \cdot F(z),\ \forall a, b$$

Time Shift

$$\mathcal{Z}\left\{ e(kT - nT) \cdot u(kT - nT) \right\} = z^{-n} \cdot E(z)$$

Final Value Theorem

$$\lim_{k \to \infty} e(kT) = \lim_{z \to 1} (z-1) E(z)$$

In order to analyze such systems it's necessary to have the z transform of the time delay function. As we have seen earlier, the starred transform of a signal $e(t)$, is the Laplace transform of the product of this signal by a periodic sequence of pulses of period T ,

$$E^*(s) = \mathcal{L}\left(e(t) \cdot \sum_{k=0}^{+\infty} \delta(t - kT) \right)$$

The signal pure time delay refers to a displacement in the growing sense of the time axis. Thus, shifting the signal $e(t)$ of a fractional amount of the sampling period, say ΔT with

$$\Delta T = (1-m)T \text{ with } 0 \le m \le 1$$

the previous expression takes the following form:

$$E^*(s) = \mathcal{L}\left(e(t - (1-m)T) \cdot \sum_{k=0}^{+\infty} \delta(t - kT) \right) \tag{131}$$

Note that sampling is not delayed, only the signal. Now, since the function is dependent on an additional parameter (m), the previous equation is rewritten as:

$$E^*(s,m) = \mathcal{L}\left(e(t-(1-m)T) \cdot \sum_{k=0}^{+\infty} \delta(t-kT) \right)$$

Since the starred transform is identical to the z transform with $z = e^{sT}$ then

$$E(z,m) = E^*(s,m)\Big|_{z=e^{sT}} = \mathcal{L}\left(e(t-(1-m)T) \cdot \sum_{k=0}^{+\infty} \delta(t-kT) \right)_{z=e^{sT}}$$

or,

$$E(z,m) = \mathcal{L}\left(e(t-T+mT) \cdot \sum_{k=0}^{+\infty} \delta(t-kT) \right)_{z=e^{sT}}$$

Pure-time delays, integer multiples of the sampling period, can be factored back, the previous expression is replaced by the following one,

$$E(z,m) = e^{-sT}\mathcal{L}\left(e(t+mT) \cdot \sum_{k=0}^{+\infty} \delta(t-kT) \right)_{z=e^{sT}}$$

that is,

$$E(z,m) = z^{-1}\mathcal{L}\left(e(t+mT) \cdot \sum_{k=0}^{+\infty} \delta(t-kT) \right)_{z=e^{sT}} \qquad (132)$$

It is known that the Laplace transform of the product of two variables in time is given by the complex convolution integral.

[note **]**

If $x(t) \underset{\mathcal{L}^{-1}}{\overset{\mathcal{L}}{\rightleftarrows}} X(s)$ and $y(t) \underset{\mathcal{L}^{-1}}{\overset{\mathcal{L}}{\rightleftarrows}} Y(s)$ then $x(t)\cdot y(t) \overset{\mathcal{L}}{\rightarrow} \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} X(\lambda)\cdot Y(s-\lambda)d\lambda$

This integral can be solved using a theorem derived from complex analysis: the residual method (the same method used to derive equations (124) and (127)).

[note **]**

If $x(t) \underset{\mathcal{L}^{-1}}{\overset{\mathcal{L}}{\rightleftarrows}} X(s)$ and $y(t) \underset{\mathcal{L}^{-1}}{\overset{\mathcal{L}}{\rightleftarrows}} Y(s)$ then

$$\mathcal{L}\left[x(t)\cdot y(t)\right] = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} X(\lambda)\cdot Y(s-\lambda)d\lambda = \sum_{\substack{\text{At the poles} \\ of\ X(\lambda)}} \text{Res}\left\{X(\lambda)Y(s-\lambda)\right\}$$

Thus, considering that $e(t)$ has Laplace transform $E(s)$, and the pulse sequence has Laplace transform equal to:

$$\Delta(s) = \mathcal{L}\left[\sum_{k=0}^{+\infty}\delta(t-kT)\right] = \int_0^{\infty}\sum_{k=0}^{+\infty}\delta(t-kT)\cdot e^{-st}dt = \sum_{k=0}^{+\infty}e^{-skT} \quad \text{(Unilateral)} \tag{133}$$

Since $\Delta(s)$ is a geometric series with ratio e^{-sT} than,

$$\Delta(s) = \frac{1}{1-e^{-sT}} \tag{134}$$

[note **]** Summing the terms in a geometric progression:

$$\sum_{k=a}^{b} r^k = \frac{r^a - r^{b+1}}{1-r}$$

Thus, and given the residue theorem, expression (132) is replaced by the following one:

$$E(z,m) = z^{-1}\left[\sum_{\substack{nos\ pólos\\ de\ E(\lambda)}} \text{Res}\left\{e^{mT\lambda}E(\lambda)\cdot\Delta(s-\lambda)\right\}\right]_{z=e^{sT}} \tag{135}$$

Substituting $\Delta(s-\lambda)$ by expression (134), evaluated $s = s-\lambda$, the above equality is replaced by,

$$E(z,m) = z^{-1}\left[\sum_{\substack{nos\ pólos\\ de\ E(\lambda)}} \text{Res}\left\{e^{mT\lambda}E(\lambda)\cdot\frac{1}{1-e^{-T(s-\lambda)}}\right\}\right]_{z=e^{sT}} \tag{136}$$

i.e.

$$E(z,m) = z^{-1}\left[\sum_{\substack{nos\ pólos\\ de\ E(\lambda)}} \text{Res}\left\{e^{mT\lambda}E(\lambda)\cdot\frac{1}{1-z^{-1}e^{T\lambda}}\right\}\right] \tag{137}$$

Which is the more effective form to compute the modified transformed z from the Laplace transform of a signal or system.

An alternative way of establishing the modified z transform is verifying, from expression (130), that

$$E^*(s,m)=\mathcal{L}\left(e(t-(1-m)T)\cdot\sum_{k=0}^{+\infty}\delta(t-kT)\right)=\mathcal{L}\left(\sum_{k=0}^{+\infty}e(t-(1-m)T)\cdot\delta(t-kT)\right)$$

Given the definition of discrete-time impulse, the previous equation takes the following form:

$$E^*(s,m)=\mathcal{L}\left(\sum_{k=0}^{+\infty}e(kT-(1-m)T)\cdot\delta(t-kT)\right)=\sum_{k=0}^{+\infty}e(kT-(1-m)T)\cdot e^{-kTs}$$

For $k=0$ then $E^*(s,m)=\mathcal{L}\left\{e\left(-(1-m)T\right)\delta(t)\right\}$. As $(1-m)T$ is always positive $\forall m\in[0,1]$ then $-(1-m)T$ concerns a negative value. Additionally, since $e(t)=0$ for $t<0$ the previous relationship is replaced by the following one,

$$E^*(s,m)=\sum_{k=1}^{+\infty}e(kT-(1-m)T)\cdot e^{-kTs}$$

which leads to the alternative modified z transform formulation:

$$E(z,m)=\sum_{k=1}^{+\infty}e(kT-(1-m)T)\cdot z^{-k} \tag{138}$$

This parameterization provides an alternative to expression (136) and may be useful in cases where the signal comes in time-series format.

Finally, in order to illustrate the modified z transform conversion, consider the following example where you one want to obtain the modified z transform for the following system,

$$E(s)=\frac{e^{-1.2s}}{s+1}$$

sampled at a rate $T=0.013$. This system can be rewritten as

$$E(s)=\frac{e^{-T(k+\Delta)s}}{s+1}=\frac{e^{-T(k+1-m)s}}{s+1},\ \forall k\in\mathbb{Z}^+\ \forall m\in[0,1]$$

Due to this fact we obtain that

$$k=\left\lfloor\frac{1.2}{T}\right\rfloor=92$$

Note, once again, that $\lfloor\cdot\rfloor$ refers to the floor rounding operator. Using the computed k one find the value of Δ as follows:

$$\Delta = \frac{1.2 - kT}{T} \approx 0.308$$

which leads to that

$$m = 1 - \Delta \approx 0.692$$

Then, the $E(s)$ modified z transform is:

$$E(z,m) = z^{-1}\mathcal{Z}_m\left[\frac{e^{-(k+1)Ts}\cdot e^{mTs}}{s+1}\right] = z^{-94}\mathrm{Res}\left\{\frac{e^{mT\lambda}}{\lambda+1}\cdot\frac{1}{1-z^{-1}e^{T\lambda}}\right\}_{\lambda=-1} \approx \frac{0.991}{z-0.987}z^{-93}$$

[suggestion **]**	Perform the same procedure, but this time with $T = 0.2$

Comparing the expressions (127) and (137), one can say that, for the case of systems with pure delays integer multiples of the sampling period, both transforms are related by the following equality:

$$E(z) = \lim_{m\to 0} zE(z,m) \tag{139}$$

Additionally, and just like in the ordinary z transform, pure time-delays integer multiples of the sampling period can be factored,

$$\mathcal{Z}_m\left(e^{\pm skT}E(s)\right) = z^{\pm k}\mathcal{Z}_m\left(E(s)\right) \tag{140}$$

Please note that the z transform tables cannot be applied directly to his modified version. Therefore new tables must be derived (usually by the expression (137)). Annex A4 gives some transform pairs for the most common signals or systems.

2.2.4 Inverse Z transform and difference equations

Typically, a digital controller is conceptually, a set of equations that operate in a time-domain sampled signal (usually the error). In this curricular unit, one must think that those equations are implemented and solved by digital microprocessors.

Just like some analog controller design techniques, the design of digital controllers is made using the frequency domain representation of the open-loop system. One start from the system Laplace transform and try to obtain the controller transfer function (in the z domain) in order the closed-loop system to cope with the defined performance criteria.

As already mentioned, typically the implementation of the controller is in the time domain then, in order to physically implement the control law in the microprocessor based system, there's a need to convert the frequency domain controller transfer function to the digital time domain. This procedure is performed through the inverse z transform operation.

Formally, the inverse z transform is described mathematically by:

$$e[k] = \frac{1}{2\pi j} \oint_c E(z) z^{k-1} dz \tag{141}$$

where the contour integral is taken counter-clockwise in the convergence region containing the origin of $E(z)$.

There are several techniques for obtaining the inverse z transform from a discrete-time transfer function. Most of them bypasses the computation of the contour integral solution by using tables and, if necessary, a pre-arrangement of the original function by splitting up or taking rational expansions [10] [13]. An alternative strategy is based on the evaluation of the contour integral using the Cauchy residue theorem. Thus, the inverse transform of $E(z)$ can be obtained given the following relation [13]:

$$e[k] = \sum_{\substack{\text{at poles of} \\ E(z)z^{k-1}}} \text{Res}\left\{ E(z) \cdot z^{k-1} \right\} \tag{142}$$

Note that, unlike the equation {126}, residues are taken to the poles of $E(z)z^{k-1}$. Thus, if $k < 0$ it's necessary to evaluate, besides the poles of $E(z)$, the residues of $k+1$ poles at the origin. However, since this course interest lies only in causal systems, the equation (142) is evaluated only for $k \geq 0$. Note that for $k = 0$ it is necessary to calculate the residue at $z = 0$.

Just like in the starred transform, the residue assessment depends on the poles multiplicity.

- If the system has a simple pole at $z = a$,

$$\text{Res}\left\{ E(z)z^{k-1} \right\}_{z=a} = (z-a)E(z)z^{k-1}\Big|_{z=a}$$

- If the system has a multiple pole at $z = a$ multiplicity m,

$$\text{Res}\left\{ E(z)z^{k-1} \right\}_{z=a} = \frac{1}{(m-1)!}\left(\frac{d^{m-1}}{dz^{m-1}}\left[(z-a)^m E(z)z^{k-1} \right] \right)\Bigg|_{z=a}$$

In the normal context of control engineering, the discrete transfer function is usually a ratio of polynomials in z with the generic form:

$$G(z)=\frac{Y(z)}{U(z)}=\frac{b_m z^m+b_{m-1}z^{m-1}+\cdots+b_0}{z^n+a_{n-1}z^{n-1}+\cdots+a_0} \tag{143}$$

As seen above, to ensure causality it is necessary that the denominator polynomial degree is equal or greater than the numerator polynomial degree. Re-arranging the above expression one obtain:

$$\left(1+a_{n-1}z^{-1}+\cdots+a_0 z^{-n}\right)Y(z)=\left(b_m z^{m-n}+b_{m-1}z^{m-n-1}+\cdots+b_0 z^{-n}\right)U(z) \tag{144}$$

Applying the inverse transform (see [properties] pg. 88) we may write:

$$\begin{aligned} y[k]+a_{n-1}y[k-1]+\cdots+a_0 y[k-n]= \\ b_m u[k-m+n]+b_{m-1}u[k-m+n+1]+\cdots+b_0 u[k-n] \end{aligned} \tag{145}$$

Such equation types are called difference equations and they are what actually define the control rules implemented in digital microprocessors.

[note **]** In the difference equation $y[k]$ refers, in fact, to $y(kT)$. The notation presented is based on the precedent introduced by Oppenheim and Schaffer (1998) and aims to be a more compact, and less ambiguous, way to represent discrete sequences. So one can say that:

$$y[k]=y(kT)=y(t)\big|_{t=kT}$$

[note **]**

In many control systems publications it's possible to detect an alternative difference equations presentation. This alternative is based on the definition of a time shift operator. This shift may be in order to advance time (*forward shift operator*) or to time delay (*backward shift operator*). The shift operator is denoted by the letter q and, when applied to a function in time, performs the following operation:

- $q\cdot e(kT)=e(kT+T)=e[k+1]$ (time advance)
- $q^{-1}\cdot e(kT)=e(kT-T)=e[k-1]$ (time delay)

In a general way,

- $q^p\cdot e(kT)=e(kT+pT)=e[k+p],\quad \forall p\in\mathbb{Z}$

For example, consider the following transfer function:

$$\frac{Y(z)}{U(z)} = \frac{b_m z^m + b_{m-1} z^{m-1} + \cdots + b_0}{z^n + a_{n-1} z^{n-1} + \cdots + a_0} = \frac{B(z)}{A(z)}, \quad \forall n, m \in \mathbb{Z},\ n \geq m,$$

Dividing both terms by z^n and considering that the excess of poles over zeros is $d = n - m$ one gets:

$$\frac{Y(z)}{U(z)} = \frac{z^{m-n} + b_{m-1} z^{m-n-1} + \cdots + b_0 z^{-n}}{1 + a_{n-1} z^{-1} + \cdots + a_0 z^{-n}} = \frac{z^{-d} + b_{m-1} z^{-d-1} + \cdots + b_0 z^{-d-m}}{1 + a_{n-1} z^{-1} + \cdots + a_0 z^{-n}} = \frac{B^*(z)}{A^*(z)}$$

where $A^*(z)$ and $B^*(z)$ refer to the reciprocal polynomials. Factoring z^{-d}, and taking the inverse z transform, the differences equation can be written as:

$$A^*(q^{-1}) y(kT) = B^*(q^{-1}) u(kT - dT)$$

Note that $z \neq q$ since the former is a complex variable and the second is an operator. However, to some extent, one can say that

$$q^p \cdot e(kT) = \mathcal{Z}^{-1}\left(z^p E(z)\right) \text{ if } e(kT) \underset{z^{-1}}{\overset{z}{\rightleftarrows}} E(z)$$

2.3 Mapping the s into the z plane

As we have just seen, there is a close relation between the complex variable z and the complex variable s. The link between both variables was defined in section §2.2.2 as:

$$z = e^{sT} \tag{146}$$

In a conventional linear and time invariant analog system, the transfer function becomes a ratio of polynomials in s. The location, in the s plane, of the poles and zeros defines the system dynamic behaviour. In the same fashion, discrete-time systems can also be described by a ratio of polynomials in z. Also the values of z that make the function equal to zero are called the transfer function zeros. In the same way, the values of z that make the transmission infinite are called the transfer function poles. Like its continuous-time counterpart, the dynamic behaviour of a discrete system is also closely linked to the poles and zeros location on a map called the z plane.

Being z a complex variable makes sense that the z plane, like the s plane, is the Argand plane for complex numbers. Moreover, given the existence of a relationship between the complex variables s and z seems obvious that there

also a relationship between both plans. Indeed it is. During this section will observe how the plane s is transformed into the plane z from the relationship (146).

As already said, the variable s being complex, has both real and imaginary parts described generically by:

$$s = \sigma + j\omega \tag{147}$$

Given the equation (146) it's possible to state that

$$z = e^{(\sigma + j\omega)T} = e^{\sigma T} \cdot e^{j\omega T} \tag{148}$$

The z module and phase are $e^{\sigma T}$ and $e^{j\omega T}$ respectively. Additionally, since the complex exponential is a periodic function with period $2k\pi$, $\forall k \in \mathbb{Z}$, the following relationship holds true,

$$z = e^{\sigma T} \cdot e^{j(\omega T + 2\pi k)} = e^{\sigma T} \cdot e^{jT\left(\omega + \frac{2\pi}{T}k\right)} = e^{\sigma T} \cdot e^{jT(\omega + \omega_o k)}, \forall k \in \mathbb{Z} \tag{149}$$

From this expression one concludes that the s plane singularities, for frequencies integer multiples of the sampling frequency ω_o, are mapped to the same location in the z plane.

Now let's analyse what happens, in the z domain, for different values of s. If $s = 0$ then,

$$z = e^{0T} = 1$$

We conclude that, in the discrete plane, a singularity at $s = 0$ is translated to $z = 1$. In general, taking into consideration expression (147), any singularity located at:

$$s = 0 + j\frac{2k\pi}{T} = 0 + jk\omega_o, \ \forall k \in \mathbb{Z}$$

will have its position exactly at $z = 1$.

Now we relax the s complex variable imaginary value and force the real part to be zero. This situation is equivalent to evaluate the s map along the $j\omega$ axis. In this case, equation (148) is reduced to,

$$z = e^{j\omega T} \tag{150}$$

By varying the analog frequency ω from $-\omega_o/2$ to $\omega_o/2$ one observe that z describes, in the Argand plane, a **circumference with unity radius**. This result

allows us to conclude that the Laplace's $j\omega$ axis is mapped, on the z plane, into a unit radius circle.

[suggestion] Run the following code in MATLAB® and try to see what is going on for alpha values outside the specified range. (Note that if $\omega = \alpha\omega_o/2$ with $\alpha \in [-1,1]$ then $z = e^{j\alpha\pi}$)

```
»alpha =- 1:0.01:1;
»Z = exp (j * pi * alpha);
»plot (real (z), imag (z))
»axis square
```

Now one assume the real part of the variable s is positive. That is $\mathrm{Re}\{s\} > 0$. This constraint defines, for the s plane, only the right hand half-plane limited, at left, by the $j\omega$ axis. This situation leads to $|z| > 1$ and hence $z = |z| \cdot e^{j\omega T}$ regards all the space outwards the unity radius circle. Thus, poles or zeros located in the right hand half-plane on the s plane are mapped, in the z plane, into the space region outside the unity circle.

The last situation concerns the location of continuous-time singularities located at the left-hand half-plane. All s points, forced to obey at $\mathrm{Re}\{s\} < 0$, are mapped, in the z domain, into the unity circle interior region. For this reason, to be stable, all the causal discrete-time system poles must lye inside the unity circle. The following figure illustrates graphically all the cases considered in the previous paragraphs.

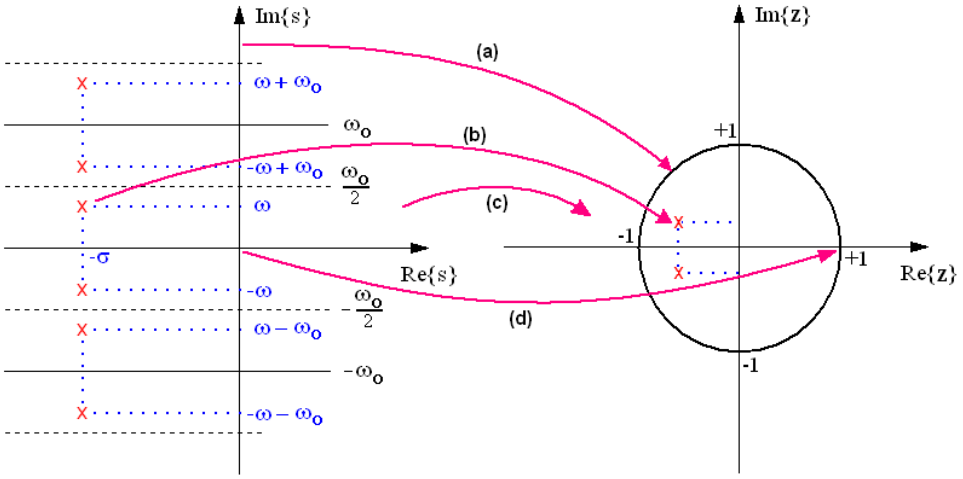


Fig 36. Mapping the s into the z plane: (a) the imaginary axis is transformed into a unit radius circle (b) the left half-plane is converted inside the unit circle (c) the right half-plane becomes the plane bounded below by the circle (d) singularities at the origin are now at the point z = 1.

A final remark concerns the fact that, in sampled system, the Laplace transform lead to an infinite number of poles. This would not be representable in the plane s however, in the plane z, all the secondary stripes are mapped into the same spatial location as the primary stripe. For this reason only a finite number of poles are represented in z.

2.3.1 Discrete-time system frequency response

The frequency response, in an analog system, is obtained through the system behaviour to pure frequency excitation signals within a given frequency range. If a mathematical model of the system is available, this evaluation can be made from the Laplace transform evaluated at $s = j\omega$.

The same concept can be extrapolated to systems expressed in z. Indeed, since,

$$z = e^{sT}$$

evaluating the frequency response for a discrete system is equivalent to evaluate the complex variable z for the cases in which $s = j\omega, \forall\omega$. That is,

$$z = e^{j\omega T}$$

Thus, a discrete-time system with transfer function $G(z)$ has frequency response $G\left(e^{j\omega T}\right)$.

As mentioned in section § 2.1.1, and can also be observed by equation (93), the product of frequency ω by the sampling period T result in a quantity designated by digital frequency ω_d. Thus, the frequency response is,

$$G\left(e^{j\omega_d}\right) = G\left(z\right)\Big|_{|z|=1 \text{ and } \omega_d = \omega T}$$

for a frequency range within the interval $[0, \pi]$. Additionally, since for sigma equal to zero the module of z is equal to unity, the evaluation of frequency response is made over the unit radius circumference.

[note **]** In the case discrete system has Fourier transform, then the z transform is equivalent to the Fourier transform when $|z| = 1$,. That is when $z = e^{j\omega_d}$ if $|z| = 1$.

[note] For a system, or discrete sequence, to admit Fourier transform it is necessary that the z transform convergence region includes the unit circle. The convergence region is the set of all z values that makes the transform convergent. The convergence region describes regions bounded by origin concentric circles. For cases where the systems is expressed by ratio of polynomials in z, the convergence region never includes the poles. Additionally, for causal systems, the convergence region is always bounded below by a circle.

2.3.1.1 Frequency response geometric evaluation

Consider the figure 37 where, in the z plane, a generic point $z = e^{j\omega_a}$ is represented. Graphically one observes that, at frequency $\omega_a = 0$, z refers to the point $(1, j0)$. Increasing the frequency the point moves counterclockwise around the unit circle. At the frequency $\omega_a = \pi$ (i.e. for ω equal to half the sampling frequency) z refers to the point $(-1, j0)$. Now for $\omega_a = 2\pi$ the point z located again at $(1, j0)$. This situation is repeated at integer multiples of 2π. This phenomenon is, of course, a sampling consequence.

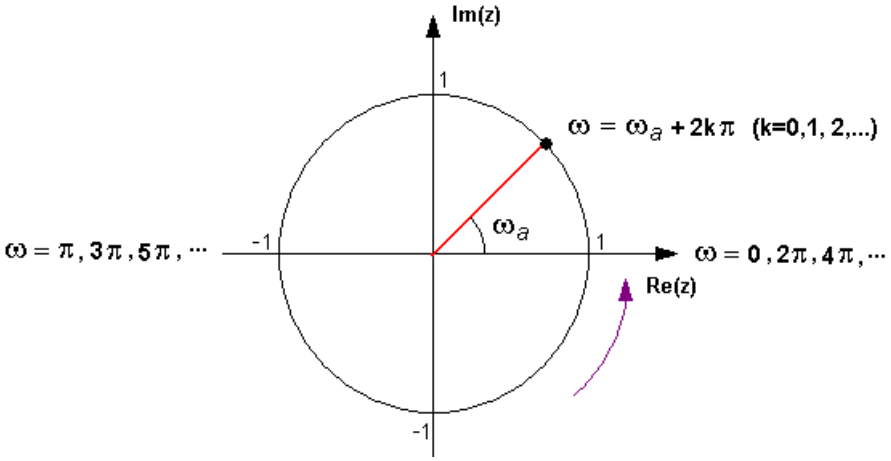


Fig 37. Evaluation of the geometric location of a point $z = e^{j\omega_d}$ according to the digital frequency.

Due to the relationship between both the z and Fourier transforms, from the pole-zero map is possible to evaluate the magnitude and phase of the Fourier transform. So, consider a linear, time-invariant discrete-time system represented by a transfer function parameterized as follows,

$$G(z)=\frac{\prod_{i=1}^{m}(z-w_i)}{\prod_{i=1}^{n}(z-p_i)},\ \forall n,m\in\mathbb{Z}\ \text{e}\ n>m$$

Since $(z-p_i)$ and $(z-w_i)$ are vectors in the z plane, the Fourier transform can be evaluated as follows:

- The magnitude of the Fourier transform is equal to the product of the magnitude of all the zero vectors divided by the product of the magnitude of all pole vectors:

$$\left|G(e^{j\omega_d})\right|=\frac{\prod_{i=1}^{m}\left|e^{j\omega_d}-w_i\right|}{\prod_{i=1}^{n}\left|e^{j\omega_d}-p_i\right|}$$

- The phase is equal to the sum of the phases of all zero vectors minus the sum of the phases of all pole vectors pole (note: the angles are taken in reference to the positive real axis).

$$\measuredangle G(e^{j\omega_d})=\sum_{i=1}^{m}\measuredangle\left(e^{j\omega_d}-w_i\right)-\sum_{i=1}^{n}\measuredangle\left(e^{j\omega_d}-p_i\right)$$

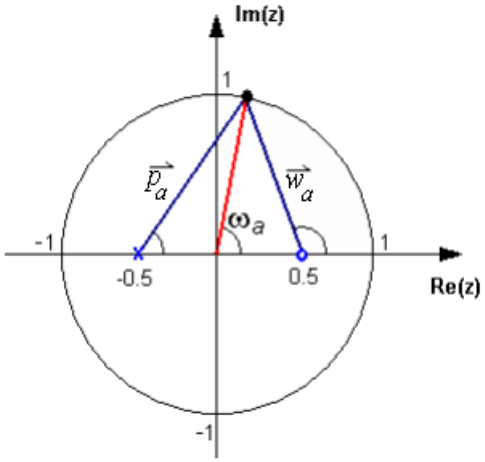


Fig 38. Vectors pole and zero for a generic frequency ω_a

Consider, as an example, the following causal system defined in z by;

$$G(z)=\frac{z-0.5}{z+0.5}$$

This transfer function has a zero at $z=0.5$ and a pole at $z=-0.5$. For a general frequency ω_a we can plot the pole and zero vectors illustrated in Figure 38:

For concrete values of ω_d (measured in rad / sample) one have obtained the results presented in table 4:

ω_d	$\lvert\vec{p}\rvert$	$\lvert\vec{w}\rvert$	$\measuredangle\vec{p}$	$\measuredangle\vec{w}$	$\lvert G(e^{j\omega_d})\rvert$	$\measuredangle G(e^{j\omega_d})$
0	1.5	0.5	0	0	1 / 3	0
$\pi/2$	$\sqrt{5}/2$	$\sqrt{5}/2$	1.1071	2.0345	1	0.9274
π	0.5	1.5	π	π	3	0

Tabela 4. Evaluation of the system response G (z) for some frequency values

Assessing for a wider set of values, one obtains the Fourier transform module and phase profile illustrated in figure 39.

Before to end this subject the following considerations are presented:

- The poles, when placed near the unit circle, produce well-defined peaks in the response at the corresponding angular frequency.
- Zeros on the unit circle have the effect of producing a null response for the corresponding angular frequency.

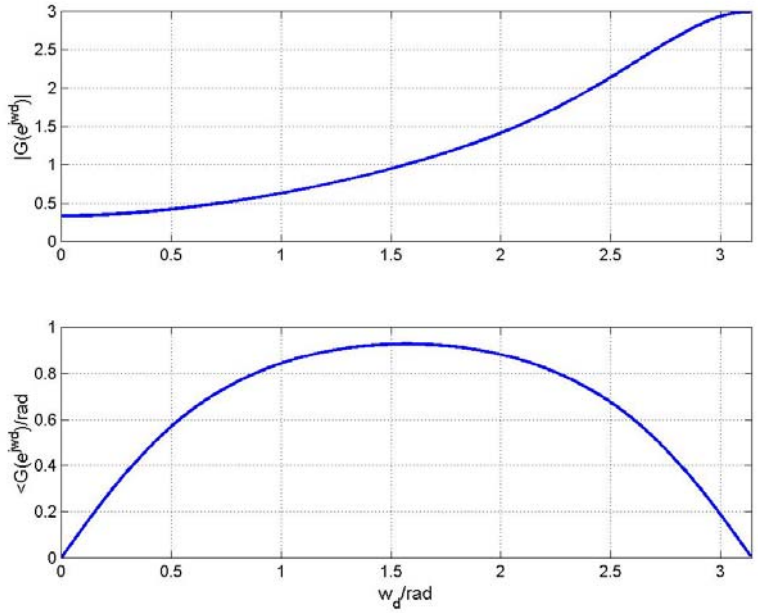


Fig 39. Frequency response for the system G (z).

2.3.1.2 Discrete-time system stability

In the first chapter it was seen that for a linear, time invariant and causal system to be stable it was necessary that the all poles possess negative real part. In the case of discrete systems is easy to see that the stability condition requires all

the poles to be inside the unit circle.

This can be illustrated by considering the following example: Consider two causal first-order discrete-time systems with poles at $z = 0.9$ and $z = 1.02$:

$$G_1(z) = \frac{1}{z - 0.9}$$

$$G_2(z) = \frac{1}{z - 1.02}$$

It is easy to see that, geometrically, the first pole is inside the unit circle and the second is outside. Applying the inverse z transform one obtains the following impulse responses:

$$h_1[k] = (0.9)^k u[k]$$

$$h_2[k] = (1.02)^k u[k]$$

Figure 40 presents the graphical aspect, of both expressions, for the first thirty samples.

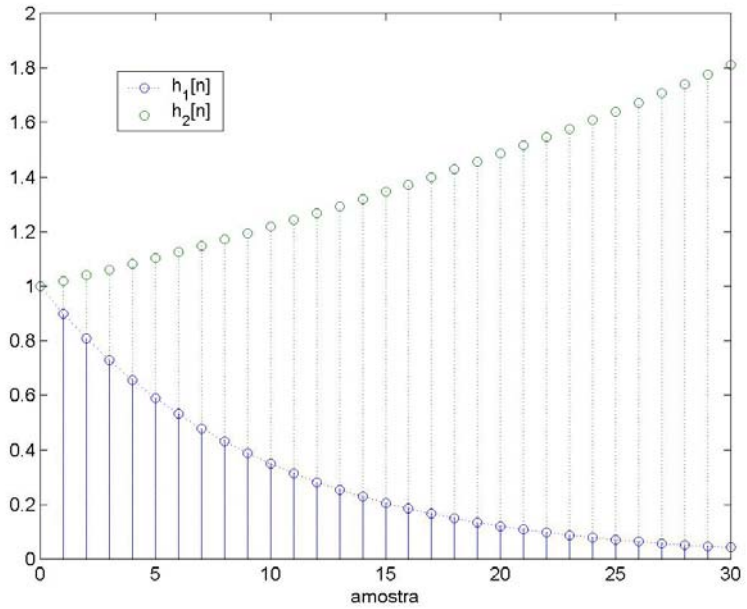


Fig 40. Impulse response system $G_1(z)$ and $G_1(z)$

As one might suspect, the impulse response $h_2[k]$ is not absolutely summable. The n terms sum of $h_2[k]$ lead to a geometric series with ratio greater than one and then divergent. Thus, poles whose magnitude is less than one, contribute to the transient response with terms that decay to zero over time. In other way,

poles with higher than unity modules, lead to transient terms that exponentially increase in time.

What if, in the above problem, the poles had negative values? In stability terms nothing changes. Just the time response shape is changed. So consider the impulse response of the two previous systems but now with poles at $z=-0.9$ and $z=-1.02$:

$$h_1[k]=\left(-0.9\right)^k u[k]$$

$$h_2[k]=(-1.02)^k u[k]$$

The impulse response behaviour, for the first thirty samples, concerning both systems is illustrated at figure 41.

Comparing the previous with figure 40 one concludes that both systems tend to the same values. However, the way they do it is different. In this second case the system appears to show oscillation like an under-damped system. In fact, this phenomenon could be anticipated since the impulse responses of both systems have a term of type $(-1)^k$ which is alternately positive or negative depending on the exponent parity. So, unlike analog systems, discrete systems with only one pole can oscillate. Due to this fact, poles with negative real part are given the name "*ringing poles*".

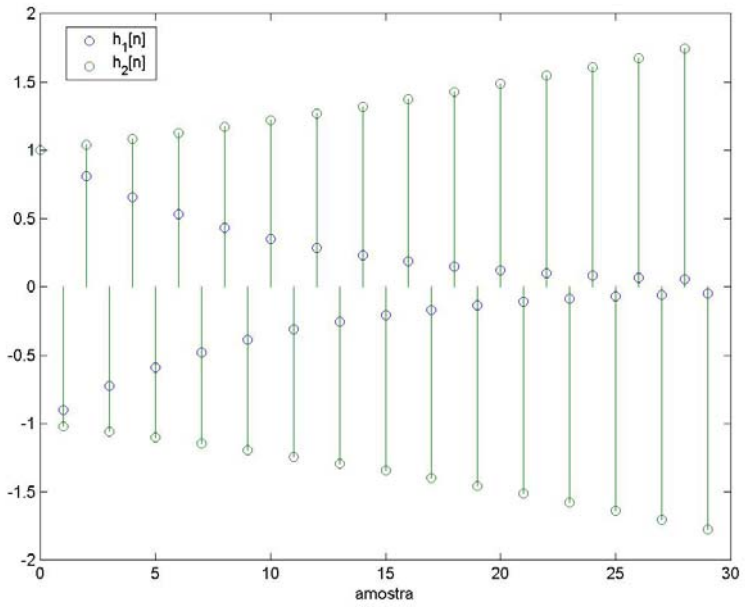


Fig 41. Impulse response for discrete-time systems with negative poles.

2.3.2 Continuous Transfer Functions Discretization

The whole theory derived so far, to discrete-time signals, was based on the concept shown in Figure 17. That is, a discrete-time signal is a sequence of amplitude values of a continuous-time signal taken regularly at specified time intervals. Thus, the discrete sequence can be viewed as a weighted sum of discrete impulses displaced in time. The Laplace transform of this sequence, and subsequent change of variable, led to the concept of the z transform:

$$\mathcal{L}\left(e(kT)\right) = E^*(s) = E(z)\big|_{z=e^{sT}} = \sum_{k=0}^{+\infty} e(kT)\cdot z^{-k} = \mathcal{Z}\left(E(s)\right) \tag{151}$$

What would happen if the sequence $e(kT)$ was obtained, by sampling the response of an analog system, to a given excitation signal? What is the relationship between the analog system transfer function and the sequence's z transform?

To answer these questions consider the following figure:

Fig 42. Transfer function of a system with sampler

The physical system transfer function that originates from a given excitation signal $u(t)$, the signal $e(t)$ is:

$$\frac{E(s)}{U(s)} = G(s) \Rightarrow E(s) = G(s)\cdot U(s)$$

On the other hand, the output signal starred transform is

$$E^*(s) = \left[E(s)\right]^* = \left[G(s)\cdot U(s)\right]^*$$

and for $z = e^{sT}$,

$$E(z) = \mathcal{Z}\left[E(s)\right] = \mathcal{Z}\left[G(s)\cdot U(s)\right] \tag{152}$$

However, the transfer function $G(z)$ is defined as,

$$G(z) = \frac{E(z)}{U(z)} \tag{153}$$

where

$$U(z) = \mathcal{Z}\left[U(s)\right]$$

Replacing (152) and (153) one obtains the relationship between the z transform and its Laplace transform:

$$G(z) = \frac{\mathcal{Z}\left[G(s)\cdot U(s)\right]}{U(z)} \tag{154}$$

Thus, contrary to what might be assumed, $G(z) \neq \mathcal{Z}\left[G(s)\right]$. In fact, what the previous expression tells us is that the analog system z transform depends on the excitation signal. However the transfer function should be independent of the input signal profile. Indeed it is so, but with this strategy what one is trying to map into z is not the system dynamics but its time response. It's intended to find a z function for the system from the input/output continuous signals observations in discrete instants (similar to system identification procedures). The $G(z)$ transfer function serves the specific objective to preserve the input/output relationship of the analog system (at least in the sampling instants).

[note **]**

Note that $\mathcal{Z}\left[G(s)\cdot U(s)\right]$ is, in general, different from $G(z)\cdot U(z)$. In fact, considering, for example,

$G(s) = U(s) = \frac{1}{s}$ we obtain

$G(z) = U(z) = \frac{z}{z-1}$ then,

$$G(z)\cdot U(z) = \frac{z^2}{z^2 - 2z + 1}$$

However,

$$\mathcal{Z}\left[G(s)\cdot U(s)\right] = \mathcal{Z}\left[\frac{1}{s^2}\right] = T\frac{z}{z^2 - 2z + 1}$$

Thus, for this particular example, the z transform of the product is only equal to the product of the z transform if the sampling period is unity.

For example, for the particular case of an impulse input, the equation (154) takes the following form:

$$G(z) = \mathcal{Z}\left[G(s)\right] \tag{155}$$

This mean that the z transform of the Laplace transform of an analog system only preserves the impulse response. In fact observe figure 43 which shows the response of,

$$G(s) = \frac{1}{s+1} \text{ and } G(z) = \frac{z}{z - e^{-T}},\ T = 0.2 \tag{156}$$

to two separate signals: an impulse and a unit step.

If the objective was to ensure accuracy of the step response samples the z transform should be:

$$G(z) = \frac{\mathcal{Z}\left[G(s)\cdot s^{-1}\right]}{\mathcal{Z}\left[s^{-1}\right]} = (1 - z^{-1})\mathcal{Z}\left[G(s)\cdot s^{-1}\right] = \frac{1 - e^{-T}}{z - e^{-T}} \tag{157}$$

The result is further illustrated in figure 44.

Looking closely at the equation (157) one observes that the z transform of the system $G(s)$ is taken as if the system was cascaded to a zero-order hold.

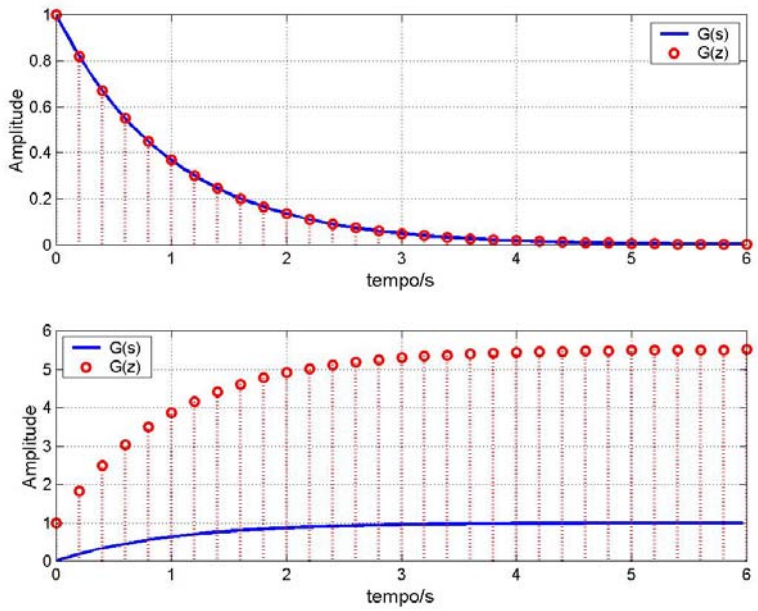


Fig 43. Impulse response and step response of an analog system and its z model taken from $\mathcal{Z}\left[G(s)\right]$. Note the impulse response samples adjustment accuracy at the sampling instants. Compare now with the step response (figure below).

The effect of continuous-time transfer function discretization can also be seen in the frequency domain. Thus consider the Bode plot illustrated in figure 45.

As you would suspect, the discrete-time system frequency response (measured

to the Nyquist frequency) does not exactly match with the continuous-time system frequency. Depending on the excitation signal to preserve, one obtains different transfer functions and then, several frequency response profiles.

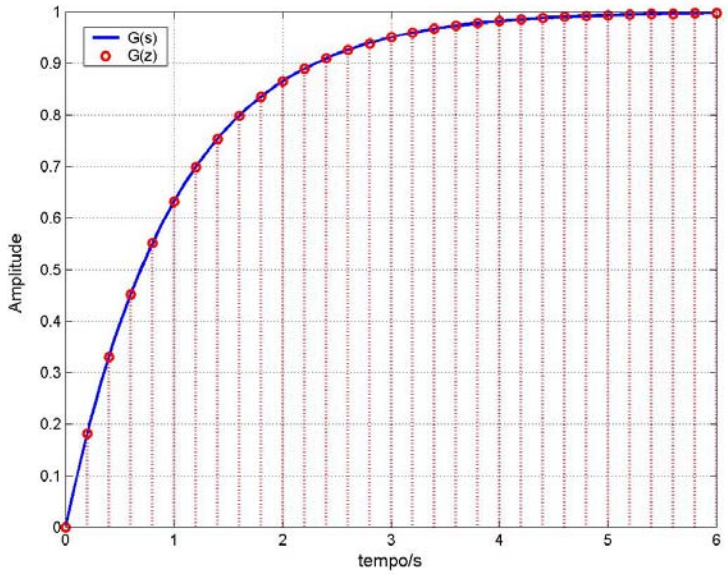


Fig 44. Step response of the system of equation {157}

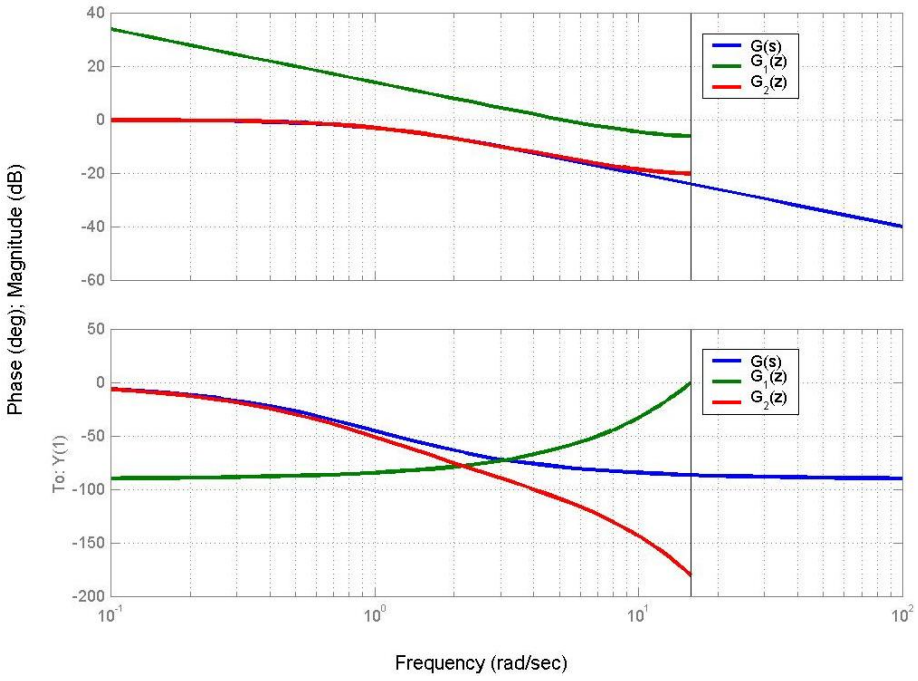


Fig 45. Frequency response for three different cases: the analog system, transformed into their discrete system z obtained from the z transformed into the cascade system and zero-order.

An alternative way to discretize the continuous-time transfer function, besides the one from the input/output signals observation, result from a direct

transformation using some $s \to z$ mapping law. Please note that the relationship between the two variables has been already established as:

$$z = e^{sT} \tag{158}$$

that is.

$$s = \frac{1}{T}\ln(z) \tag{159}$$

However this transformation law is not desirable since it converts a ratio of polynomials in s into a sum ratio of nonlinear functions in z. Thus, several methods have been proposed in order to circumvent this problem [7].

The diversity of existing methods is an indicator of the ineptitude of one of them meets all the requirements for a mapping strategy. As we will see, each method has advantages and disadvantages that should be considered for the controller discretization.

2.3.2.1 Euler forward and backward

As already noted, the direct application of the transformation (159) is not feasible as it transforms a linear rational fraction s in a fraction of non-linear in z. In order to avoid this problem, one option is to approximate (159) into a polynomial in z. The simplest method of doing this is to expand it in Taylor series around $z = 1$ (because we want a good match of both functions at low frequencies). In this context, the expression (159) is replaced by the following equality,

$$s = \frac{1}{T}\ln(z)\bigg|_{z=1} + \frac{1}{Tz}\bigg|_{z=1}(z-1) - \frac{1}{Tz^2}\bigg|_{z=1}\frac{(z-1)^2}{2} + \cdots \tag{160}$$

The logarithmic function admits an infinite number of derivatives, so the previous expression should be truncated at some point. One strategy is to neglect all terms of order greater than one. Therefore, the above expression is reduced to,

$$s = \frac{1}{T}\ln(z)\bigg|_{z=1} + \frac{1}{Tz}\bigg|_{z=1}(z-1) \tag{161}$$

which leads to the following relationship between s and z,

$$s = \frac{z-1}{T} \tag{162}$$

What this approach means in reality and how far it is valid?

To answer the first part of the question considers a continuous-time system governed by the following differential equation:

$$\frac{dx(x)}{dt} = y(t) \tag{163}$$

with zero initial conditions. Applying the Laplace transform one obtains:

$$sX(s) = Y(s) \tag{164}$$

where $X(s) = \mathcal{L}\{x(t)\}$ and $Y(s) = \mathcal{L}\{y(t)\}$. Discretizing the system using the relationship (162) one gets,

$$\frac{(z-1)}{T} X(z) = Y(z) \tag{165}$$

Now, applying the inverse transform yields the following difference equation,

$$\frac{x[k+1] - x[k]}{T} = y[k] \tag{166}$$

Comparing the expression {163} with the last one, we conclude that the approximation {162} t in the frequency domain is equivalent to approximate the first derivative to a first difference. In other words, the derivative is taken as the difference between the signal samples of $x(t)$ at $t = (k+1)T$ and $t = kT$ divided by the sampling period (note the non-causal expression of (166)). Because the calculation of this derivative approach requires a signal sample value ahead of the present moment this method is often called "Euler *Forward*".

The second part of the previously raised question concerns the approximation quality of (162). Remember that this mapping strategy was obtained by polynomial expansion around the lower frequencies.

Let us first examine how, using the Euler forward discretization technique, the s plane is mapped into the z plane. Thus, solving the expression (162) in order to z we obtain

$$z = sT + 1 \tag{167}$$

Since $s = \sigma + j\omega$ one gets,

$$z = (\sigma T + 1) + j\omega T \tag{168}$$

For $\sigma = 0$ the pure imaginary axis is mapped, at z domain, into a vertical line passing through point $z = 1$. On the other hand, the Laplace left hand half-plane, is transformed into a half-plane to the left of $z = 1$. For analog poles located on the right half-plane one found that, in the digital domain, these poles are transformed into poles at the right hand of the vertical line that passes through the point $z = 1$. The figure 46 seeks to illustrate these considerations.

We then concludes that:

- The left half-plane in s is not mapped within a circle of unit radius in the plane z (although this includes it);
- Stable analog systems can provide unstable digital ones. In fact, depending on the sampling period, poles in the left half-plane in s can be transformed in poles outside the unit circle in z.
- The frequency outline of the z plane does not follow the circumference of unit radius. Instead follow a vertical line that passes the point $z = 1$ (Note, however, that around this point, the frequency response is very close as forced by the performed Taylor expansion).

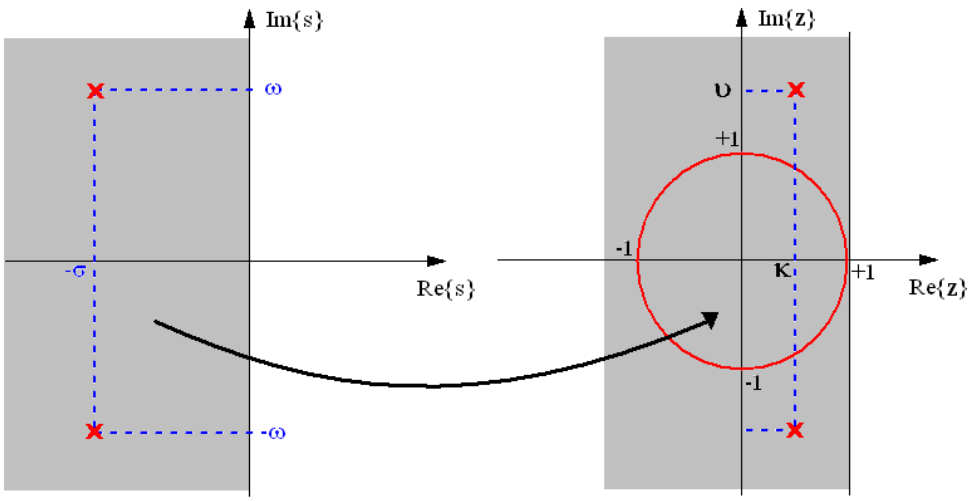


Fig 46. Planar mapping s to plane z transformation using the "Forward Euler".

Resulting from these conclusions we can say that, unless high sampling frequencies are used, this mapping is undesirable. An alternative has to do with how the derivative is calculated numerically. Instead of using the present and next sample value one can use the present and the previous sample. The following relation reflects this strategy:

$$\frac{x[k]-x[k-1]}{T}=y[k] \tag{169}$$

This derivative numerical method approximation is designated by "Euler Backward."

The mapping carried out by this technique follows the law:

$$s=\frac{1-z^{-1}}{T}=\frac{z-1}{zT}\ \text{(Note the pole at the origin to delay the signal)} \tag{170}$$

To analyze how the s plane is mapped into the z plane the above expression is solved in order to z resulting in,

$$z=\frac{1}{1-sT} \tag{171}$$

Since $s=\sigma+j\omega$ one get,

$$z=\frac{1}{(1-\sigma T)-j\omega T} \tag{172}$$

For $\sigma=0$,

$$z=\frac{1}{1-j\omega T}=\frac{1}{2}\frac{(1-j\omega T)+(1+j\omega T)}{1-j\omega T}=\frac{1}{2}\left(1+\frac{1+j\omega T}{1-j\omega T}\right)=\frac{1}{2}+\frac{1}{2}e^{j\varphi} \tag{173}$$

where

$$\varphi=\tan^{-1}(\omega T)+\tan^{-1}(\omega T)=2\cdot\tan^{-1}(\omega T) \tag{174}$$

In expression (173), the complex exponential impress, in the z plane, a circle of radius $1/2$ and the "*offset*" shifts the circle centre, along the positive real axis, by an amount equal to $1/2$. It is easy to verify that the stability region of the Laplace plane is transformed within this circle, and therefore, the right half-plane of the s map becomes the entire plane outside the same circle.

As done previously for the "Euler *Forward*" method, figure 47 presents a picture that geometrically illustrates the relationship between plans s and z for the "Euler Backward" method. By observation of the this figure one concludes that:

- As in the previous case, the left half-plane in s is not mapped exactly inside a unit radius circle in the z plane;
- It is possible to stabilize unstable analog systems after the discretization process.

- The frequency outline of the z plane also does not follow the unit radius circumference. Moreover, there is a degradation of the sampling frequency for points away from $z=1$.

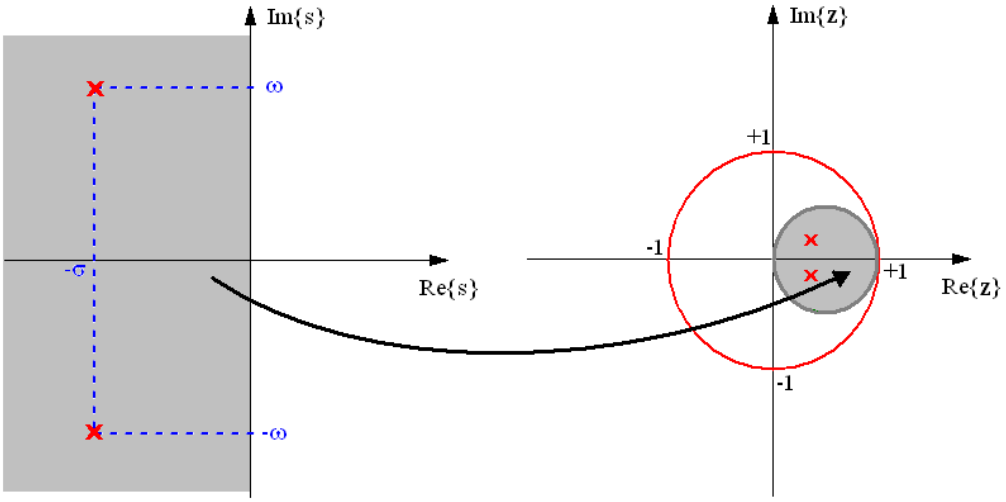


Fig 47. Mapping s plane to z plane using the " Euler Backward."

To conclude, let the say that often one want, after the analog transfer function discretization process, to maintain the original frequency response. Thus, although the two techniques presented are easily applied, they do not preserve the impulse response and severely distort the frequency response (at least for relatively low sampling frequencies or digital frequencies away from $\omega_d = 0$).

Thus, in the following section, an alternative technique is presented. This new mapping strategy is also derived from a polynomial approximation of expression (162). But now with the advantage of mapping, the s plane $j\omega$ axis, into the unit radius circle interior at the z plane.

2.3.2.2 Bilinear or "Tustin" transformation.

The most common form of analog transfer functions discretization is by direct replacement of complex variable s by a first order Padé approximation, around $z=1$, of expression (162). This technique is called bilinear or Tustin transformation.

With the results obtained in section § 2.1.4.3 (equation (119)) it is possible to extrapolate to the case in which one intends to approximate the equation (162) by a ratio of two polynomials. Therefore, solving in order to s, it's easy to verify that the relationship between the complex variables s and z is;

$$s = \frac{2}{T}\frac{z-1}{z+1} \tag{175}$$

[note **]** The above expression could also be obtained by applying the trapezoidal approximation's numerical integration method [4] [13].

Or,

$$z = \frac{2+sT}{2-sT} \tag{176}$$

given that $s = \sigma + j\omega$,

$$z = \frac{(2+\sigma T)+ j\omega T}{(2-\sigma T)- j\omega T} \tag{177}$$

Evaluating the previous expression along the $j\omega$ axis,

$$z = \frac{2+ j\omega T}{2- j\omega T} \tag{178}$$

That, in polar form, has the following aspect:

$$z = e^{j\omega_d} \text{ where } \omega_d = 2\cdot\tan^{-1}\left(\frac{\omega T}{2}\right) \tag{179}$$

Graphically the previous equation represents in the complex plane, a circle with unity radius. Moreover, if $\sigma < 0$, is easy to see that the numerator modulus of expression (177) is lower than the denominator modulus which leads to an approximation modulus less than one. Thus, the entire left half-plane of the Laplace domain is transformed into the interior of the unit radius circle. Similarly, for $\sigma > 0$, this mapping strategy results in the conversion of the right half-plane into the outer circumference of unity radius. Thus a stable analog system has, as its discrete equivalent, a stable filter. However this transformation leads to a problem of frequency slide, i.e. the relationship between analog frequency and digital frequency is not linear. More specifically the relationship is,

$$\omega = \frac{2}{T}\tan\left(\frac{\omega_d}{2}\right) \tag{180}$$

This last expression establishes the association between the s plane frequency and the z plane digital frequency.

[proof **]**

Starting from

$s = \frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}$ and evaluating z on the unit circle, i.e. $z = e^{j\omega T}$, one get

$j\omega = \frac{2}{T}\frac{1-e^{-j\omega T}}{1+e^{-j\omega T}}$ which, after factorization, becomes:

$$j\omega = \frac{2}{T}\frac{e^{-j\omega\frac{T}{2}}\left(e^{j\omega\frac{T}{2}} - e^{-j\omega\frac{T}{2}}\right)}{e^{-j\omega\frac{T}{2}}\left(e^{j\omega\frac{T}{2}} + e^{-j\omega\frac{T}{2}}\right)}$$

Using the Euler relations one obtain

$$j\omega = \frac{2}{T}\frac{e^{-j\omega\frac{T}{2}}2j\left(\frac{e^{j\omega\frac{T}{2}} - e^{-j\omega\frac{T}{2}}}{2j}\right)}{e^{-j\omega\frac{T}{2}}2\left(\frac{e^{j\omega\frac{T}{2}} + e^{-j\omega\frac{T}{2}}}{2}\right)} = \frac{2}{T}j\frac{\sin\left(\frac{\omega T}{2}\right)}{\cos\left(\frac{\omega T}{2}\right)} = \frac{2}{T}j\tan\left(\frac{\omega T}{2}\right)$$

and finally $\omega = \frac{2}{T}\tan\left(\frac{\omega_d}{2}\right)$

[alternative proof **]**

From expression (179) and solving in order to ω.

This frequency distortion tends to be negligible for higher sampling frequencies. In fact, observe figure 48. As one can see, by selecting a sufficiently high sampling frequency, the effect of distortion is minimized. Specifically, for sampling frequencies greater than twenty times the system bandwidth, the distortion is kept below 1%

2.3.2.3 Pole-Zero mapping

Another method for converting a transfer function, from the Laplace domain into the z domain, is based on the relationship between plans s and z. That is a pole or zero at $s = a$ is converted into a pole or zero at $z = e^{aT}$. Thus, knowing the analog transfer function singularities location, it's possible to establish a discrete transfer function whose poles and zeros are the poles and zeros of the

original transfer function transformed by the relationship (162). This technique is easy to apply and the algebraic manipulation can be summarized in the three following points:

- Begin by mapping out all the finite poles and zeros in accordance to the relationship $z = e^{sT}$.
- If the numerator order is less than the denominator (which happens often) add zeros in $z = -1$ until both terms have the same degree [6]. With the introduction of these "artificial" zeros at the Nyquist frequency the digital system frequency response at $\omega \to \omega_0 2^{-1}$ is similar to the analog at $\omega \to \infty$
- Finally set up the DC gain so that both transfer functions have the same value.

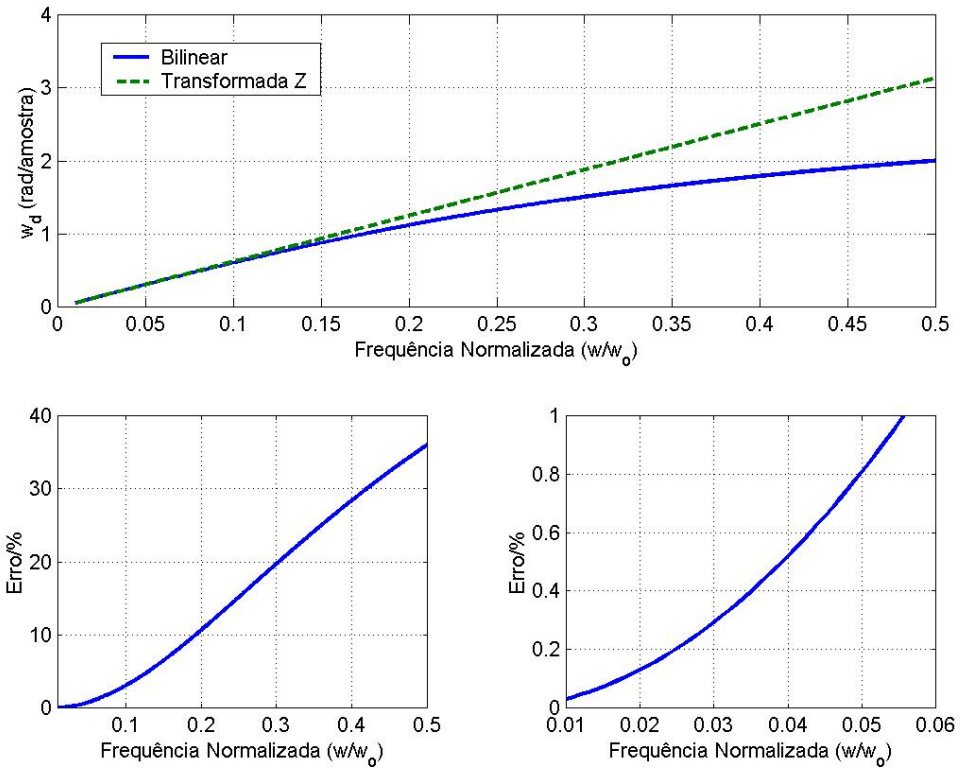


Fig 48. Frequency distortion due to bilinear transformation. Below right, the graphic detail of the relative error.

To illustrate the procedure consider the following case:

$$G(s) = \frac{s+1}{s^3 + 2s^2 + 2s}$$

Factoring the transfer function:

$$G(s)=\frac{s+1}{s\left(s^2+2s+2\right)}=\frac{s+1}{s\left(s+1+j\right)\left(s+1-j\right)}$$

Converting finite poles and zeros (assuming a unit sampling period) one obtains:

$$G(z)=\frac{z-e^{-1}}{\left(z-e^{0}\right)\left(z-e^{-1+j}\right)\left(z-e^{-1-j}\right)}=\frac{z-e^{-1}}{\left(z-1\right)\left(z^2-2e^{-1}\cos(1)\cdot z+e^{-2}\right)}$$

Introducing two more zeros in $z=-1$

$$G(z)=\frac{\left(z^2+2z+1\right)\left(z-e^{-1}\right)}{\left(z-1\right)\left(z^2-2e^{-1}\cos(1)\cdot z+e^{-2}\right)}$$

Finally, adjusting the gain so that $sG(s)\big|_{s=0}=(z-1)G(z)\big|_{z=1}=0.5$ becomes,

$$G(z)\doteq 0.146\frac{\left(z^2+2z+1\right)\left(z-e^{-1}\right)}{\left(z-1\right)\left(z^2-2e^{-1}\cos(1)\cdot z+e^{-2}\right)}=\frac{0.146z^3+0.238z^2+0.039z-0.054}{z^3-1.398z^2+0.533z-0.135}$$

If $G(z)$ was the transfer function of a digital controller, the difference equation that would be implemented in a digital processor is obtained by:

$$G(z)=\frac{U(z)}{E(z)}=\frac{0.146+0.238z^{-1}+0.039z^{-2}-0.054z^{-3}}{1-1.398z^{-1}+0.533z^{-2}-0.135z^{-3}}$$

and applying the inverse transform of z one get:

$$u[k]=1.398u[k-1]-0.533u[k-2]+0.135u[k-3]+$$
$$+0.146e[k]+0.238e[k-1]+0.039e[k-2]-0.054e[k-3]$$

where $u[k]$ refers to the control signal applied to the process at time $t=kT$ and $e[k]$ the error signal obtained at time $t=kT$. As you may suspect there is a slight problem here. The control signal applied at the instant k depends on the error signal also at the same instant. This would be irrelevant if the "hardware" process inputs, outputs and perform the calculations in zero time. However this is impossible. Thus, in order to account for the delay introduced by the system, one considers that, during a sampling instant, the machine should have enough time to perform all necessary operations. For example if a control variable is sampled every second, then the hardware have, theoretically, a second to carry out all the data transmission and processing operations.

Thus, in order to analyse the controller performance obtained by analog

controller discretization, the effect of computer time delay must be taken into consideration. Thus, a slight change to the previously presented discretization technique is presented. This modification refers specifically in adding fictitious zeros to the transfer function. In order to maintain one sample delay between input and output, the relative degree of the transfer function must be one. That is for strictly proper systems, zeros at $z=1$ will be added until the denominator degree exceed, by one, the numerator degree [6]. In the case of proper systems, such as lead and lag controllers, the numerator degree is identical to the denominator. In this case it's not possible to add zeros. Additionally it seems even to be an extra zero. This problem is solved by adding a delay to the transfer function by introducing a pole. The pole should be added in order to have, as little as possible, influence into the system dynamic response. On the s plane that location would be $s=-\infty$. The z plane equivalent location is, obviously, $z=0$.

2.4 Sample Period Choice

The sampling frequency choice is not a trivial task. However, when the sampling theorem was derived, things seemed simple: one choose a sampling frequency that is greater than twice the maximum frequency component in the signal. However, a closer look reveals that things are not so obvious. First one need to know the signal maximum frequency component. *How to do it?*

One possibility could be carried out using a microprocessor and running the FFT on the signal. However this alternative is flawed because, to be implemented, the signal must already be sampled. On the other hand, and unlike the examples shown, the spectrum of a real signal never ends abruptly at a given frequency. Typically, the actual spectra extend from minus infinity to plus infinity. Some of these components can be part of the observable phenomenon but, most is due to noise that overlaps the signal of interest. Noise can arise in various parts of the spectrum and can be due to several phenomena including thermal agitation, magnetic induction, etc. Thus, there is some impossibility of knowing exactly what the upper limit of the signal is. Moreover, there will always be sidebands overlap.

[note **]** The FFT can be used to verify the existence of aliasing. If, after computing the FFT, there are frequency components with significant energy very close to the Nyquist frequency then there is a strong possibility that aliasing has occurred.

As already said, the choice of the sampling period is critical in digital control. The design of digital controllers by emulation start from the analog controller transfer function and is discretized using one of the technique already reviewed. Since the digital controller quality of approximation, regarding the analog controller, increases with increasing sampling frequency is clear that the sampling period influences the performance of the controller such as:

- Set-point tracking;
- Load disturbance rejection and measurement noise;
- Sensitivity to non-modeled dynamics.

If, by one hand, it is desirable to have a high sampling frequency on the other this value should be limited to the minimum necessary to carry out the numerical calculations. In fact, the algorithm computational load along with processor performance sets up the upper limit of the sampling period. Taking into consideration that the "hardware" has the capacity to meet any demand imposed by the system, some guidelines have been proposed, more or less empirical, for the choice of the sampling period.

It is often recommended in literature, a sampling period between one tenth to one fourth of the system rise time. That is,

$$\frac{T_R}{10} < T < \frac{T_R}{4} \tag{181}$$

The same is to say that the sampling period should be chosen so that [1]:

$$\frac{0.2}{\omega_{ncl}} < T < \frac{0.6}{\omega_{ncl}} \tag{182}$$

where ω_{ncl} refers to the natural frequency of the dominant closed loop poles.

Another rule of thumb, used frequently in digital signal processing, states that the sampling frequency should be five times higher than the highest frequency component where one wants to have identical analog and digital filter features.

That is:

$$\omega_o \geq 5 \cdot \omega_H \tag{183}$$

For example, for a low pass filter, ω_H can be equal to five times the bandwidth. This implies that the digital filter will behave in a similar fashion to the analog one until around twenty-five times the bandwidth. Thus, a conservative rule states that the sampling frequency should have a minimum of 20 times the closed-loop bandwidth and a maximum of 40 times the same bandwidth,

$$20 \cdot BW_{cl} \leq \omega_o \leq 40 \cdot BW_{cl} \tag{184}$$

On the other hand, as we saw in the section concerning signal reconstruction, a D/A converter is often put between the controller and the continuous system. This retention implies, as has been seen, that the control signal is delayed an amount approximately equal to half the sampling period. As expected, this delay affects the phase margin and then the system stability. Thus, a rule for choosing the sampling period indicates that the deterioration of stability, by the zero-order, holder is small and tolerable if the time delay is less than a tenth of the rise time. That is,

$$\frac{T}{2} \leq \frac{T_R}{10} \Rightarrow T \leq \frac{T_R}{5} \qquad \text{(Compare to (181))} \tag{185}$$

The relationship between the zero-order hold phase margin degradation and the sampling period can be analyzed by an alternative approach. As already be seen, the dynamics of a ZOH can be approximated by:

$$G_{zoh}(s) \approx T \cdot e^{-s\frac{T}{2}} \underset{s=j\omega}{\Rightarrow} G_{zoh}(j\omega) \approx T \cdot e^{-j\omega\frac{T}{2}} \tag{186}$$

Assuming that the phase margin degradation imposed by this element must be contained between 5º and 10º that is,

$$5^o \times \frac{\pi}{180} < \omega_{gc} \frac{T}{2} < 10^o \times \frac{\pi}{180} \tag{187}$$

which implies that

$$5^o \times \frac{2\pi}{180 \cdot \omega_{gc}} < T < 10^o \times \frac{2\pi}{180 \cdot \omega_{gc}} \tag{188}$$

By considering the gain crossover frequency (in open loop) equal to the system bandwidth one gets that the phase margin decrease, due to the zero-order hold, is limited by:

$$\frac{\pi}{18 \cdot BW_{cl}} < T < \frac{2\pi}{18 \cdot BW_{gc}} \tag{189}$$

leading to:

$$18 \cdot BW_{cl} < \omega_o < 36 \cdot BW_{cl} \qquad \text{(Compare to (184))} \tag{190}$$

2.5 Digital Control Systems Analysis

Effectively this section deals with hybrid systems:, systems composed of both discrete and continuous-time components. One will show the closed loop transfer function for some of the most common feedback control topologies. Finally section § 2.5.3 presents two algebraic techniques to analyze the system stability in the z domain.

2.5.1 Open-loop sampled systems

Usually, in digital control systems co-exist, simultaneously, continuous and discrete transfer functions. The way to handle this situation is based on the introduction of "dummy" samplers for the variables of interest. That is, despite continuous in time, one considers their values only at discrete time instants [10].

In this section we present four different cases of open-loop hybrid systems.

CASE I: Open-loop sampled system.

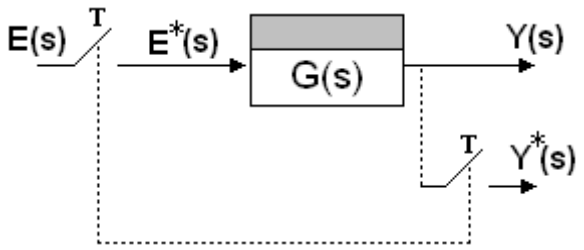


Fig 49. Simple system sampled in open loop. Note that the sampler output is fictitious and is in phase with the sampler with physical existence.

Observation of the previous figure leads to the following variables relationship:

$$Y(s) = G(s) \cdot E^*(s) \tag{191}$$

This system z analysis starts from the fictitious sampling of $Y(s)$ (Shown dashed in above figure). Thus, applying the starred transform to equation (191) one gets:

$$Y^*(s) = \left[G(s) \cdot E^*(s) \right]^* \tag{192}$$

In this context one could of * as a kind of sampling operator. After this "operator" all the already sampled variables can be factored out of the operation. This is because the synchronized sampling of an sampled signal is the same sampled signal. Thus, the previous expression takes the following form:

$$Y^*(s) = G^*(s) \cdot E^*(s) \tag{193}$$

Due to the relationship between the starred and z transform, the previous equation can be rewritten as:

$$Y(z) = G(z) \cdot E(z)\big|_{z=e^{sT}} \tag{194}$$

CASE II: Elements separated by ideal samplers.

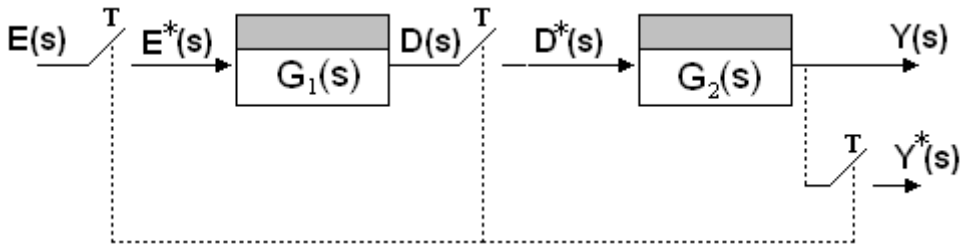


Fig 50. Cascade of two systems separated by ideal samplers. It is considered that all samplers are in phase.

In this case,

$$Y(s) = G_2(s) \cdot D^*(s) \tag{195}$$

and

$$D(s) = G_1(s) \cdot E^*(s) \tag{196}$$

Applying the star operator to both terms becomes,

$$D^*(s) = G_1^*(s) \cdot E^*(s) \tag{197}$$

Replacing the previous result on expression (195) and applying the star one

gets:

$$Y^*(s) = G_2^*(s) \cdot G_1^*(s) \cdot E^*(s) \tag{198}$$

In other words

$$Y(z) = G_1(z) \cdot G_2(z) \cdot E(z) \tag{199}$$

CASE III: Cascaded elements not separated by samplers.

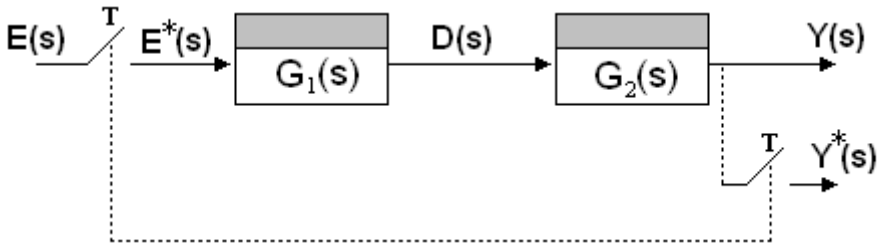


Fig 51. Cascade elements not separated by ideal samplers

This third case are similar to the previous one but without the sampler between $G_1(s)$ and $G_2(s)$. The analysis to the above system leads to:

$$Y(s) = G_2(s) \cdot D(s) \tag{200}$$

and

$$D(s) = G_1(s) \cdot E^*(s) \tag{201}$$

Replacing (201) into (200) one obtains:

$$Y(s) = G_2(s) \cdot G_1(s) \cdot E^*(s) \tag{202}$$

Applying the star is transformed into

$$Y^*(s) = \left[G_2(s) \cdot G_1(s) \cdot E^*(s) \right]^* \tag{203}$$

which leads to:

$$Y^*(s) = E^*(s) \cdot \left[G_2(s) \cdot G_1(s) \right]^* = E^*(s) \cdot \overline{G_2 G_1(s)}^* \tag{204}$$

where $\overline{G_2 G_1(s)}^*$ regards the starred transform of the product between $G_1(s)$ and $G_2(s)$.

Often common sense misleads us. From the expression (203) one may be tempted to say that

$$Y^*(s) = E^*(s) \cdot \left[G_2(s) \cdot G_1(s) \right]^* = E^*(s) \cdot G_2^*(s) \cdot G_1^*(s) \tag{205}$$

But that's wrong. This is because, as already seen, in general

$$\mathcal{Z}\left\{ G_1(s) \right\} \cdot \mathcal{Z}\left\{ G_2(s) \right\} \neq \mathcal{Z}\left\{ G_1(s) \cdot G_2(s) \right\} \tag{206}$$

That is,

$$G_1^*(s) \cdot G_2^*(s) \neq \overline{G_1 G_2(s)}^* \tag{207}$$

Case IV: Cascaded elements separated by samplers and excited by continuous signals.

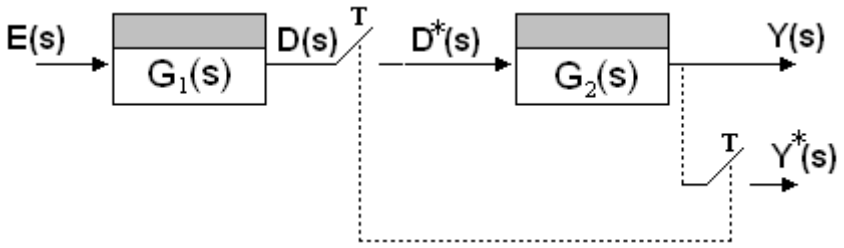


Fig 52. Cascade elements excited by continuous signal in time.

From the previous figure the following expression is derived:

$$Y(s) = G_2(s) \cdot D^*(s) \tag{208}$$

and

$$D(s) = G_1(s) \cdot E(s) \tag{209}$$

Applying the starred transform to the previous equation one gets,

$$D^*(s) = \overline{G_1 E(s)}^* \tag{210}$$

Substituting the last result in (208) lead to,

$$Y(s) = G_2(s) \cdot \overline{G_1 E(s)}^* \tag{211}$$

Finally, applying the starred transform to (211):

$$Y^*(s) = G_2^*(s) \cdot \overline{G_1 E(s)}^* \tag{212}$$

Since, usually, $\overline{G_1 E(s)}^*$ cannot be factored in $G_1^*(s) \cdot E^*(s)$ the system of figure 52 does not admit representation in transfer function format.

From the analysis of these four cases there are three important concepts to remember:

- The sampling of a sampled signal results in the sampled signal:

$$\left[E^*(s)\right]^* = E^*(s) \tag{213}$$

- Normally the sampling of the product of two signals is different from the product of the sampled signals,

$$\left[E_1(s)\cdot E_2(s)\right]^* = E_1E_2^*(s) \neq E_1^*(s)\cdot E_2^*(s) \tag{214}$$

- It is not possible to derive a transfer function if the signals applied upstream to continuous systems are not previously sampled [13].

2.5.2 Closed-loop sampled systems

Finding the transfer function for sampled systems is not a trivial task since there is no transfer function for the ideal sampler [13]. This statement is even more pronounced when it comes to closed loop sampled systems. Depending on the operations sequence, it is possible to reach a point where input/output variables factorization is not possible implying the impossibility of obtaining a system transfer function. To illustrate this situation consider the following closed-loop control system:

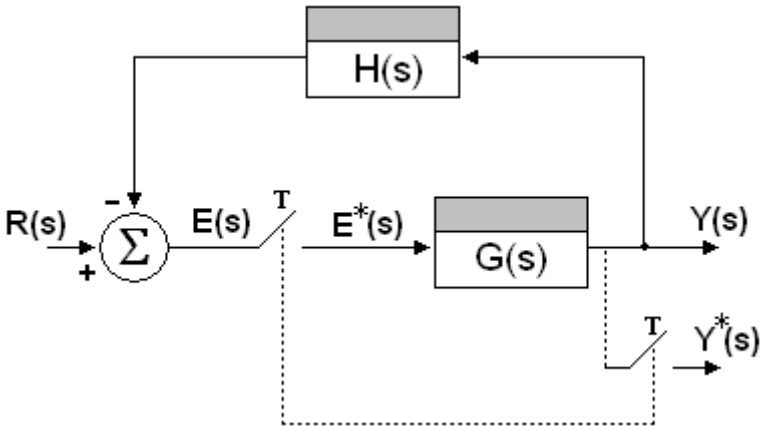


Fig 53. Feedback system with an ideal sampler in the loop

From the block diagram one can write:

$$Y(s) = G(s)\cdot E^*(s) \tag{215}$$

and

$$E(s) = R(s) - H(s)\cdot Y(s) \tag{216}$$

Replacing (216) in (215) by first applying the starred transform to $E(s)$ one

obtain:

$$Y(s) = G(s)R^*(s) - G(s)\cdot\overline{HY(s)}^* \tag{217}$$

By applying the starred transform to $Y(s)$ leads to,

$$Y^*(s) = G^*(s)R^*(s) - G^*(s)\cdot\overline{HY(s)}^* \tag{218}$$

Since $\overline{HY(s)}^*$ can't be factored, the previous expression cannot be solved for $Y^*(s)$ and, therefore, it's not possible to compute the transfer function.

On the other hand, rewriting the equation (216) as,

$$E(s) = R(s) - H(s)\cdot G(s)\cdot E^*(s) \tag{219}$$

and taking the starred transform one obtain,

$$E^*(s) = R^*(s) - \overline{HG(s)}^*\cdot E^*(s) \tag{220}$$

This leads to,

$$E^*(s) = \frac{R^*(s)}{1+\overline{HG(s)}^*} \tag{221}$$

replacing $E^*(s)$ is in (215),

$$Y(s) = G(s)\cdot\frac{R^*(s)}{1+\overline{HG(s)}^*} \tag{222}$$

Now applying the starred transformed to (222)

$$Y^*(s) = G^*(s)\cdot\frac{R^*(s)}{1+\overline{HG(s)}^*} \tag{223}$$

i.e.

$$Y(z) = G(z)\cdot\frac{R(z)}{1+\overline{HG(z)}} \tag{224}$$

We conclude by saying that there is a need for special care in handling variable and the starred transform application sequence.

Thus, in order to make easy the analysis of such systems, the following algorithm is presented [13]:

Step 1 of 3: Represent all the samplers input by a variable name;

Step 2 of 3: Write these variables as a function of each sampler output;

Step 3 of 3: Apply the starred transform.

2.5.3 Algebraic techniques for stability analysis

As referred in section § 2.3.1.2, a discrete system, causal, linear and time-invariant is stable if all the characteristic equation roots have a module less than unity. That is the discrete system poles should be located inside the unit radius circumference. However, such an analysis involves the calculation of all system modes which may be irrelevant or difficult to deal. Of course, this statement considers the treatment of a problem in algebraic rather than numerical form.

This section presents two algebraic techniques to analyze the stability of discrete-time systems. Both techniques have in common that they do not require explicit calculation of the system poles.

2.5.3.1 Routh-Hurwitz criterion for discrete-time systems

It is known that in a continuous-time LTI system, the stability limit is the imaginary axis while in a discrete system the stability bound is a geometric unit radius circle. Thus, stability analysis techniques used for continuous systems cannot be applied directly to discrete systems. However this problem can be circumvented by transforming the discrete system into a continuous one using, for example, the (inverse) bilinear transformation.

$$z \to \frac{2+Ts}{2-Ts} \tag{225}$$

With this strategy the unit circle of the plane z is transformed into the $j\omega$ axis. Hence the stability bound becomes the same as the analog system. For this reason it's now possible to use continuous-time stability analysis techniques. Among the universe of existing techniques, in this curricular unit, one focus on the Routh-Hurwitz criterion. The application of the Routh criterion for a discrete system is done by following the steps:

Step 1 of 3: Determine the closed-loop system transfer function;

Step 2 of 3: Apply the transformation expressed in equation (225);

Step 3 of 3: Apply the Routh criterion following the same procedure than for continuous systems (see text box in section § 1.2.5.1).

As for the continuous case, the Routh criterion can be used to determine the discrete system critical gain. That is the gain for which the roots cross the

imaginary axis. This value of gain is the gain to which the system is marginally stable and hence can be used to determine the critical frequency. However, for discrete systems, the critical frequency found using the previous algorithm has to be transformed in order to find the exact discrete-time system critical frequency. This transformation is performed using the relationship,

$$\omega_d = 2 \cdot \tan^{-1}\left(T \frac{\omega}{2} \right) \tag{226}$$

2.5.3.2 Jury's Criterion

As seen in the previous section, stability analysis of discrete systems can be performed using an adaptation of the classical Routh criterion. However the application of this technique requires the transformation of z to s which may lead to a lot of algebra manipulation. An alternative stability analysis technique, which can be used directly in discrete-time systems, is the Jury stability test.

Thus, consider a discrete system characteristic equation with the form:

$$Q(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0 = 0 \ , \ a_n > 0 \tag{227}$$

The Jury table is formed by using the polynomial $Q(z)$ as follows:

z^0	z^1	z^2	$\cdots$	z^{n-1}	z^n
a_0	a_1	a_2	$\cdots$	a_{n-1}	a_n
a_n	a_{n-1}	a_{n-2}	$\cdots$	a_1	a_0
b_0	b_1	b_2	$\cdots$	b_{n-1}	
b_{n-1}	b_{n-2}	b_{n-3}	$\cdots$	b_0	
c_0	c_1	$\cdots$	c_{n-2}		

(228)

where

$$b_k = \begin{vmatrix} a_0 & a_{n-k} \\ a_n & a_k \end{vmatrix}, c_k = \begin{vmatrix} b_0 & b_{n-k-1} \\ b_{n-1} & b_k \end{vmatrix} \text{ and } d_k = \begin{vmatrix} c_0 & c_{n-k-2} \\ c_{n-2} & c_k \end{vmatrix} \tag{229}$$

Note that the number of Jury's table rows is equal to $2(n-1)-1$. The necessary and sufficient conditions for discrete system stability are:

$$\begin{aligned} & Q(1) > 0 \\ & (-1)^n Q(-1) > 0 \\ & |a_0| < a_n \\ & |b_0| > |b_{n-1}| \\ & |c_0| > |c_{n-2}| \\ & \ \ \vdots \end{aligned} \tag{230}$$

The Jury criterion can be applied as follows:

Step 1 of 2: Check the first three conditions of (230). If any is violated stop the process and conclude that the system is unstable. Otherwise go to the next item

Step 2 of 2: Construct the Jury table checking, as each row is being calculated, the remaining conditions. If any of them is violated conclude that the system is unstable.

2.6 Digital Control Design

While it is possible to design a digital control system in discrete-time domain, one of the most common techniques begins by design an analog controller in the continuous-time domain: a strategy called "design by emulation". In this technique, one starts to design an analog controller transfer function in order to meet the proposed performance criteria. Then, from the obtained continuous-time transfer function, and using a discretization techniques, one obtain the discrete-time equivalent controller and, at the end, the difference equation governing the filter behaviour.

Although, for a good performance of this technique, the system should be over-sampled, this design technique is very appealing since one can use the knowledge of continuous-time controller design. Note however that the digital control systems design has additional considerations such as sampling effects, quantization and reconstruction. Due to this phenomena influence the system closed-loop dynamic behaviour of the emulated discrete controller should not coincide exactly with the one anticipated for the analog control system. However, for fast sampling and low quantization errors, the behaviour is very similar.

As already said, this section presents a digital controller design technique based on the digitization of continuous controllers. In a first phase we analyze the effect, of the sampled system elements, on the control loop. More specifically we are talking about zero-order holders, anti-aliasing filters and quantizers. Later a concrete example will be used to present the basic controller “design by emulation” steps.

2.6.1 Zero-order hold system impact

In section § 2.1.4.3 it was noted that a zero-order had, as a side effect, the system phase margin deterioration due to the time-delay introduced. As has been demonstrated, the time-delay value was in the vicinity of half the sampling period. This section steps forward in order to analyze the effect, on a control loop, of the zero-order holder. For this consider the analog system of the figure below:

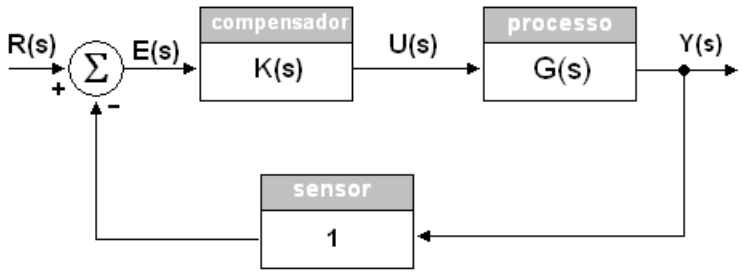


Fig 54. Analog control system in closed loop.

where

$$G_{cl}(s) = \frac{Y(s)}{R(s)} = \frac{K(s)G(s)}{1 + K(s)G(s)} \tag{231}$$

refers to its closed-loop transfer function.

Imagine now that the analog compensator, for example a phase advance controller, for various reasons should be replaced by an identical control strategy but embedded in a microcontroller. Thus, the "equivalent" scheme, from the digital system point-of-view, has the following aspect:

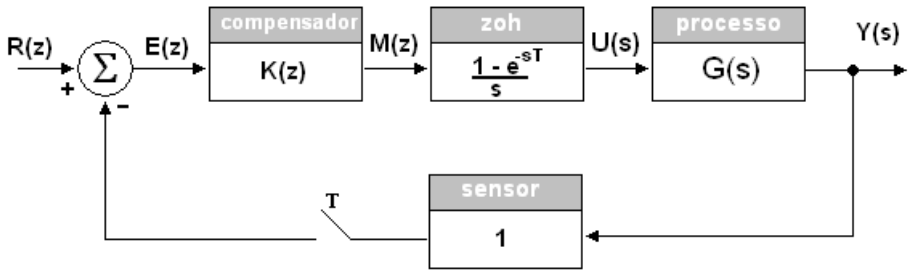


Fig 55. "Equivalent" digital control system of the previous figure.

Its closed-loop transfer function is:

$$G_{cl}(z) = \frac{Y(z)}{R(z)} = \frac{K(z)\cdot\overline{GG_{zoh}}^{*}(s)}{1 + K(z)\cdot\overline{GG_{zoh}}^{*}(s)} \tag{232}$$

where $G_{zoh}(s)$ refers to the zero-order hold transfer function.

Consider that the analog controller transfer function $K(s)$ was discretized using the appropriate technique so that the discretization influence in the closed-loop dynamics is negligible. Having said that, one begins by analyzing the system unit-step response represented in figure 56.

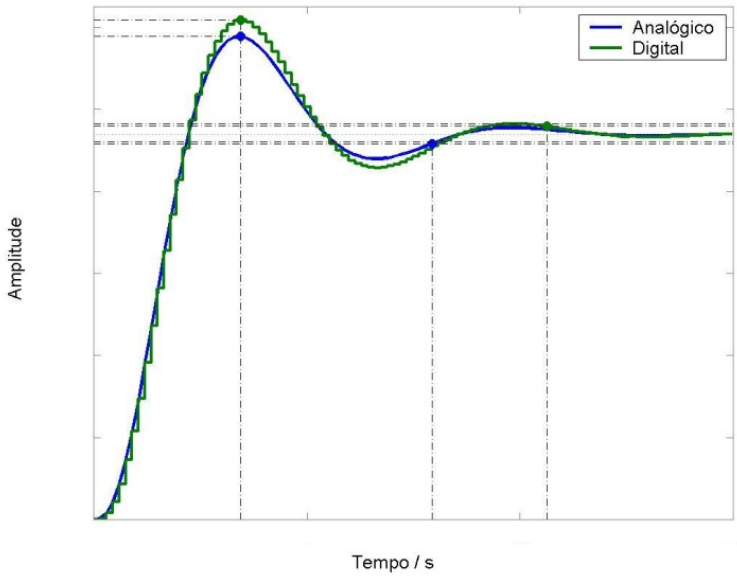


Fig 56. Step responses of both an analog controller and his digital equivalent.

As can be seen, not only an overshoot increase is observed, but also the settling time increases (remember that both performance criteria are related to zeta). Since one ensures that the controller digitalization effect is negligible, the difference between the analog and digital transient response can only be due to the zero-order hold influence. In fact, the increased instability observed is due to a phase margin reduction as shown in Figure 57.

In numerical terms, the phase decay felt was approximately equal to:

$$\phi = \omega \frac{T}{2}\bigg|_{\omega=\omega_{gc}} \text{ (Radians)}$$

Resulting from this phase reduction one would expect an overshoot increase which, in fact, was felt (look again at figure 56). In addition, and taking into consideration all the study carried out on the zero-order hold dynamics, one knows that the phase margin deterioration will decrease with increasing

sampling frequency. For that, and other reasons already discussed, the sampling period is a parameter with great impact on the digital control systems performance.

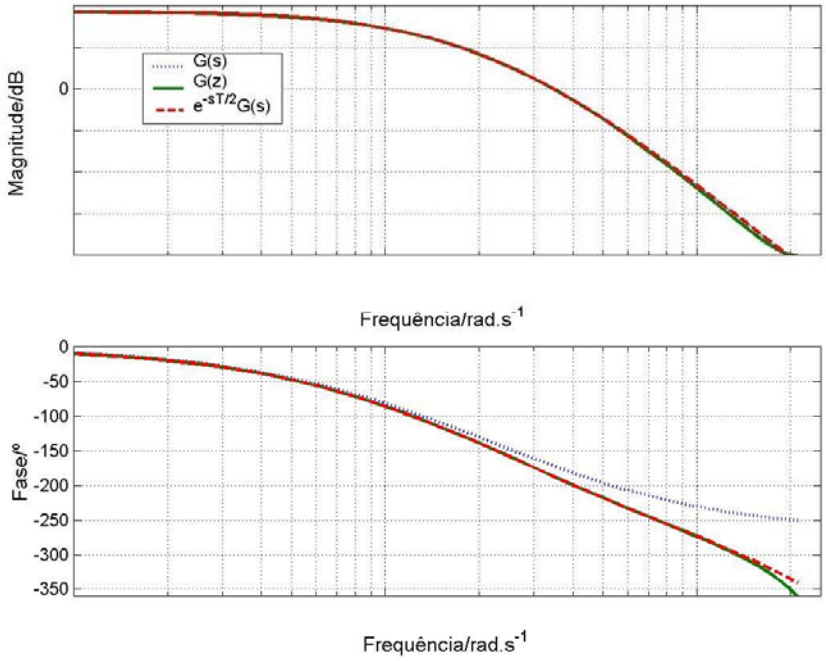


Fig 57. Frequency response comparison between an analog system and its digitized version (by the method of the step response invariance).

Concluding, the zero-order hold has a slightly destabilizing effect that can be overlooked or, alternatively, included in the design process. As a rule, and due to sampling frequencies generally involved, one expects phase margin decreases of less than 10º.

2.6.2 Effect of Anti-Aliasing Filter

We have already mentioned that a real-world signal is not "well-behaved" having frequency components that, theoretically, would extend to infinity. Thus, results from the sampling process, one would always expect some aliasing signal distortion.

One way to minimize this phenomenon requires the use of a pre-filter at sampler upstream. This filter will attenuate the energy of frequency components outside the interest band. Within a wide filter range type usually, and in order to minimize the system dynamics disturbance, the choice is a single pole filter with transfer function:

$$G_f(s) = \frac{1}{2\frac{s}{\omega_o} + 1} \tag{233}$$

In order to analyze the filter introduction effect on the control loop consider again the control system of the previous section but now with a first-order pre-filter placed before the sampler. Figure 58 show the required configuration.

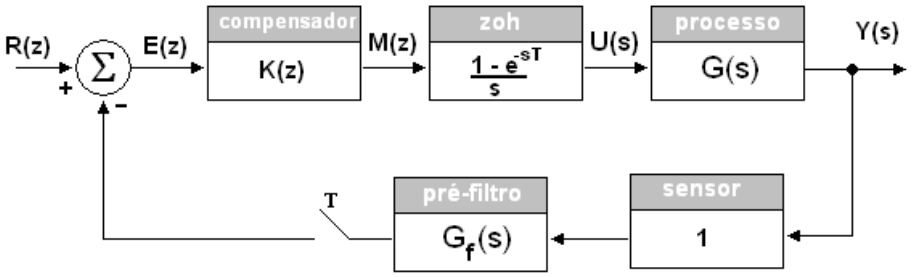


Fig 58. Introduction of an anti-aliasing filter in the loop control.

The transfer function of this new closed-loop system has the following form,

$$G_{cl}(z) = \frac{Y(z)}{R(z)} = \frac{K(z)\cdot\overline{GG_{zoh}}^{*}(s)}{1 + K(z)\cdot\overline{GG_f G_{zoh}}^{*}(s)} \tag{234}$$

Consider now the system open-loop frequency of the above configuration and compare it to the ones illustrated at figures 54 and 55.

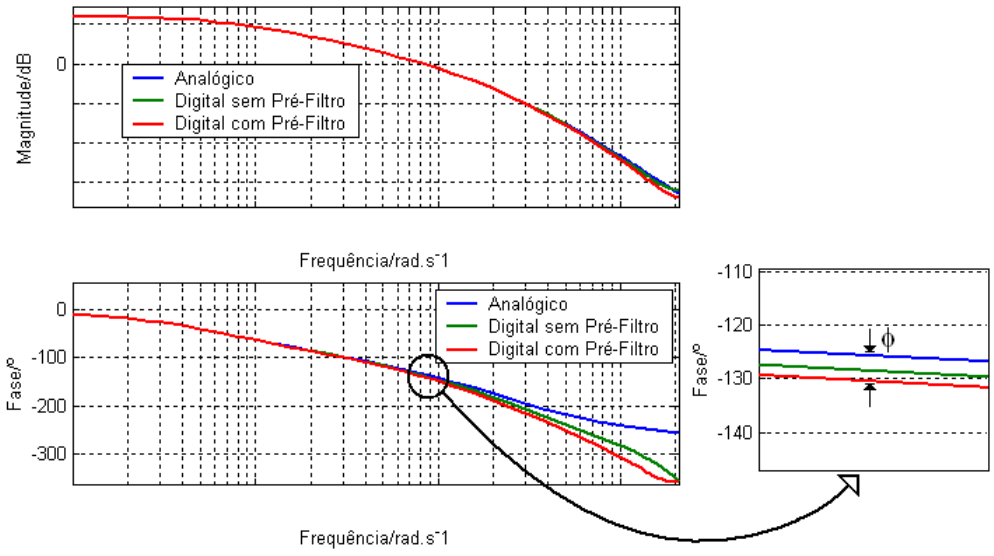


Fig 59. Frequency response of open loop, analog and digital systems (with and without pre-filter)

In terms of magnitude, only a small discrepancy is observed very close to the Nyquist frequency. In terms of phase, there is a phase margin decrease due to the additional delay introduced by the pole. Regarding the analog system, the

phase margin deterioration occurs by an amount ϕ which can be determined by:

$$\phi = \phi_{zoh} + \phi_{filtro} = \omega \frac{T}{2}\bigg|_{\omega=\omega_{gc}} + \tan^{-1}\left(2\frac{\omega}{\omega_o}\right)\bigg|_{\omega=\omega_{gc}} \quad \text{(Radians)}$$

The phase margin deterioration due to the filter can also be attested by an increase of step response overshoot (compared to the system without pre-filter). The figure below illustrates this fact.

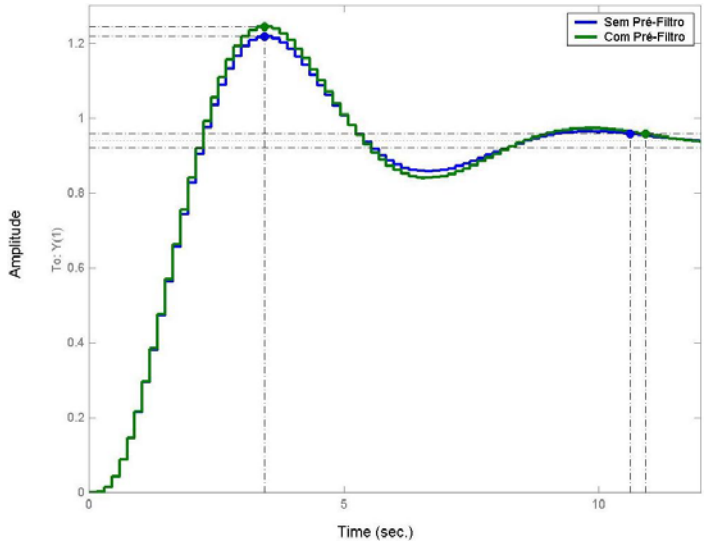


Fig 60. Step response of digital control systems with and without pre-filter.

The analysis shows that, concerning the behaviour with and without filter, the difference in dynamics is almost negligible. To be conservative one can take the pre-filter effect in the controller design by increasing the required phase margin. Due to the sampling period usually involved, the phase margin is deteriorated by an amount less than 6º.

Additionally it is worth reiterating that, in the digital control context, the anti-aliasing filter serves a very important purpose: to prevent the introduction of low-frequency disturbances in the control signal. To illustrate this phenomenon consider again the same control system with and without pre-filter. Also consider, for each of the two cases, the introduction of monochrome measurement error with signal-to-noise ratio of about 6dB and frequency slightly greater than twice the sampling frequency. This simulation strategy is characterized by figures 61 and 62.

For each case, the transfer functions $Y(z)$ over $N(z)$, is established. For the system represented in Figure 61,

$$Y(z) = -\frac{\overline{GG_{zoh}}^{*}(s)K(z)N(z)}{1+\overline{GG_{zoh}}^{*}(s)K(z)} \tag{235}$$

And for the system represented in figure 62 the transfer function is:

$$Y(z) = -\frac{\overline{GG_{zoh}}^{*}(s)\cdot K(z)\cdot\overline{G_f N}^{*}(s)}{1+\overline{GG_f G_{zoh}}^{*}(s)K(z)} \tag{236}$$

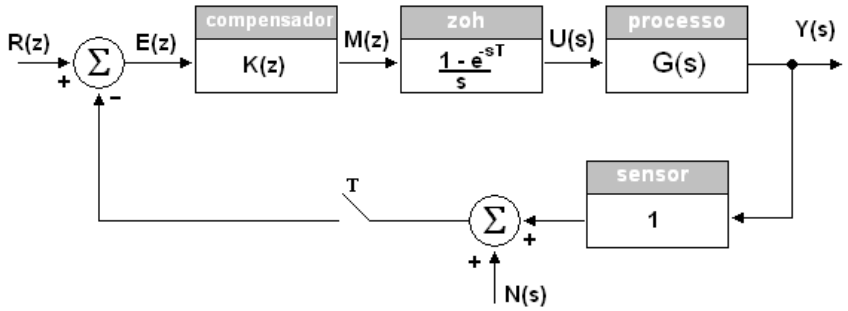


Fig 61. System measurement error contamination without pre-filter.

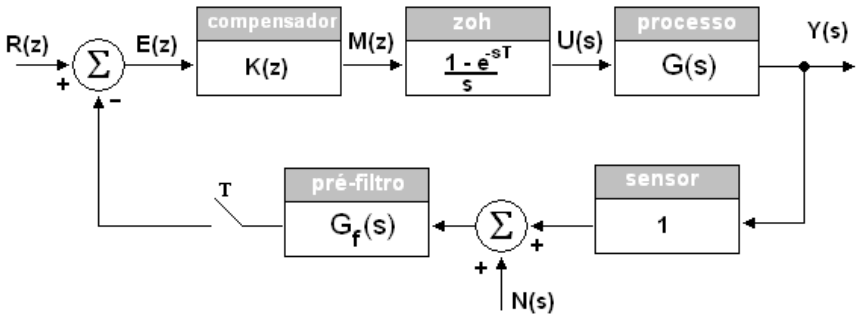


Fig 62. System measurement error contamination with pre-filter.

Just because $N(z)$ cannot be factored in this last expression, does not mean that the simulation cannot be performed [we recommend an analysis to the *script* associated with Figure 63 available on-line in a zip file]. The simulation results can be summarized by the step response illustrated in figure 63.

As expected the low-pass system behaviour was able to minimize the effect, on the output, of the measurement error. However, due to the aliasing phenomenon, this high-frequency noise becomes bandpass noise. Therefore the system fails to eliminate the measurement noise effect. For the digital system without pre-filter, the effect of measurement error at the output is very

clear. On the other hand, the introduction of a system pre-filter resulted in attenuation (by a factor of 5) of the measurement error.

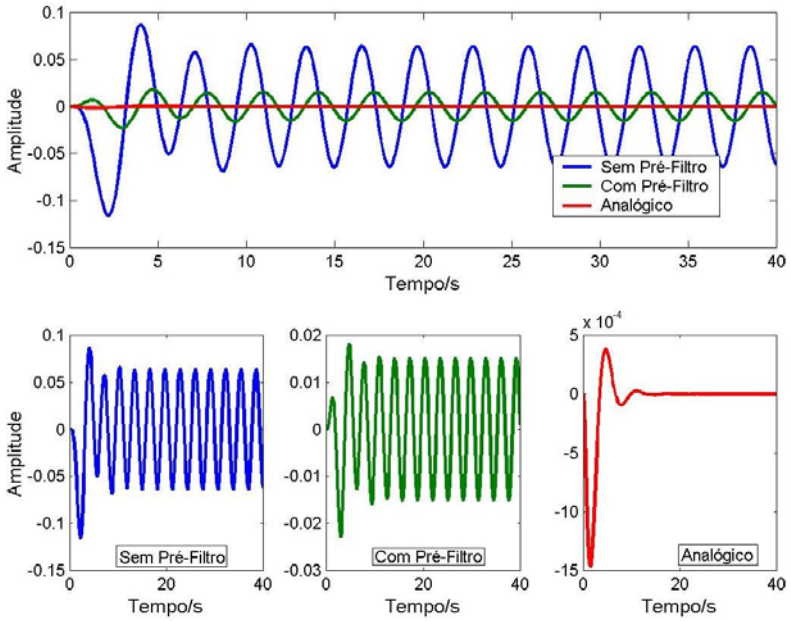


Fig 63. Response of analog and digital systems, with and without pre-filter, to out-of-band measurement noise.

2.6.3 Design by Emulation

This last section presents a full design-by-emulation strategy for digital controllers. The used technique requires the knowledge described in the first chapter. This is so because the first step of this procedure begins by designing an analog controller. This controller will force, some continuous-time process, to meet, in closed-loop, some proposed performance criteria.

The design-by-emulation procedure can be summarized by the following four steps:

Step 1 of 4: Derive the analog controller

Step 2 of 4: Choosing the sample period and add the elements associated with digital control systems.

Step 3 of 4: Discretize the control law

Step 4 of 4: Performance evaluation by simulation.

From the first step one can foreseen two different situations. One where the analog control system already exists and the aim is to convert it into digital. The second alternative admits that there is no controller and one must be designed from scratch.

The second step involves the addition, to the analogue system, of the dynamics associated with the elements that surround a digital control strategy. More specifically we talk about the A/D converter (modelled by an ideal sampler), the D/A converter (usually a ZOH) and the anti-aliasing filter. Please note that the additional dynamics effect introduced by the holder and filter can be considered in the first step. That is the phase margin deterioration by these elements can be taken into consideration during the analog controller design phase. Additionally, an appropriate sampling frequency must be selected. Usually this frequency selection is based on the closed-loop bandwidth or the step response.

Finally, after controller discretization, the sampled system performance should be evaluated. This assessment must be carried out given the control system relative stability as well as transient and steady-state responses. After ending the design iterative process, the controller transfer function must be converted into a difference equation for embed in a digital system processor. This last step may require the controller parameters round-off effect due to processor finite precision. There are some difference equation implementation strategies that try to minimize this effect [9] [13].

In order to illustrate the above discussed design procedure, the project-by-emulation of a digital controller, is presented. So, consider a open-loop system with the following transfer function:

$$G(s) = \frac{76}{(s+1)(s+3)^2} \tag{237}$$

The performance criteria to be met are:

- $e_{ss} < 10\%$;
- $BW \in [1,2]$ rad/s ;
- $P_m \approx 45^o$.

We begin the design procedure by first showing, in figure 64, the open-loop system Bode plot. The gain crossover frequency is approximately equal to 3.5 rad / s and phase margin is around 7º. As The phase margin is very low and the closed-loop bandwidth is too high (empirically $2\omega_{gc}$). These considerations suggest the use of phase lag compensator.

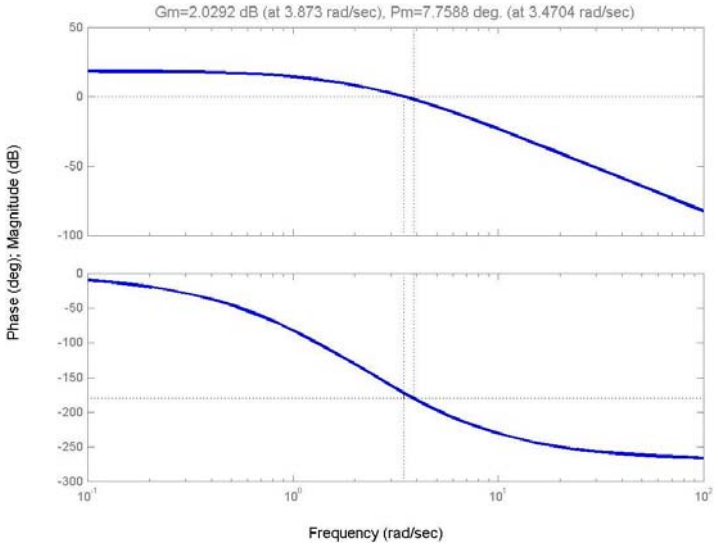


Fig 64. Open-loop system frequency response.

From the steady-state error one verify that the controller gain should obey the following restriction:

$$e_{ss} = \frac{1}{1 + K\left(76/9\right)} < 0.1 \Rightarrow K > 1.07 \tag{238}$$

Let's consider $K = 2$. In addition, and because we want a bandwidth between 1 and 2 rad / s, say 1.5 rad / s, and given the already stated rule of thumb (see equation (50)), the controller should make the gain crossover frequency to be $\omega_{gc} = 0.75$ rad/s. The figure below shows the frequency response of the system in series with the gain.

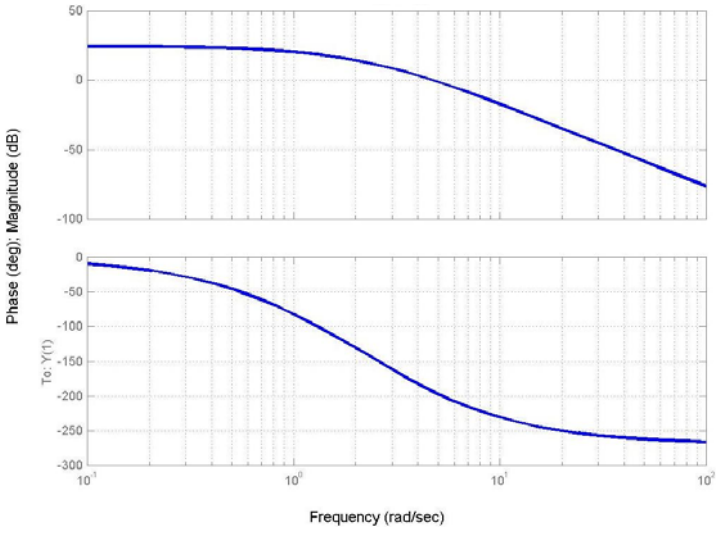


Fig 65. Frequency response of $KG(s)$

At frequency $\omega = \omega_{gc}$ one verify that the gain is equal to 22dB and the phase

equal to 65 degrees. Thus it is necessary to reduce the gain, at 0.75 rad/s frequency, of 22dB. Since you cannot change the DC gain, the desired frequency attenuation will be handled by a pole. It is known that an attenuation imposed by a pole is approximately equal to 20 dB per decade starting from the pole's frequency. So in this case, the pole should be placed a decade before of the frequency of interest. More specifically at,

$$\omega_p = \omega_{gc} \cdot \left(\sqrt{\left(10^{22/20}\right)^2 - 1} \right)^{-1} \approx 0.06 \tag{239}$$

For this case, the new open-loop frequency response is presented below in figure 66.

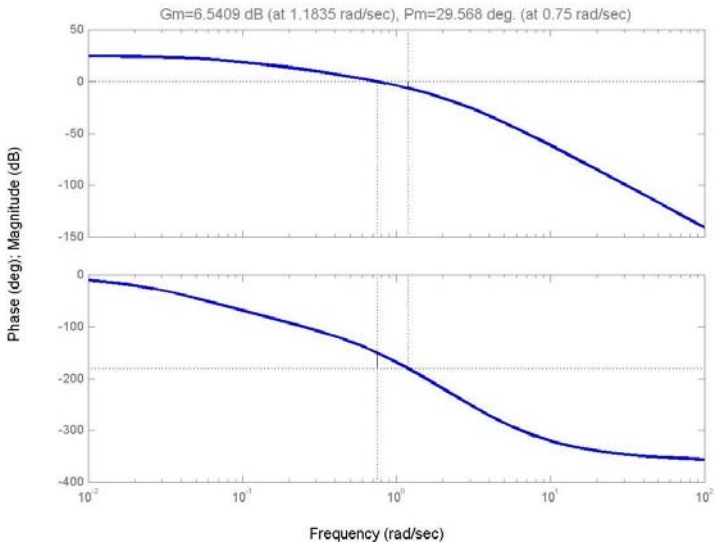


Fig 66. Frequency response of $KG(s)/\left(\omega_p^{-1}s+1\right)$

Now remains to increase the phase margin from 29.5 degrees to 45 degrees. However, since the controller is digital, to this value the effect of the ZOH and anti-aliasing filter will be added. Considering a sampling frequency of thirty times the closed-loop system bandwidth, closed-loop system, that is $\omega_0 = 30 \times 1.5 = 45$ rad/s then,

$$\phi = \phi_{zoh} + \phi_{filtro} = 0.75\frac{180^o}{45} + \tan^{-1}\left(2\frac{0.75}{45}\right) = 3^o + 2^o = 5^o$$

Thus, the phase lead is no longer 15.4° but rises to 20.4º. The phase lead is obtained by adding a zero to the system. If, at crossover frequency, the phase must increases 20.4 ° is necessary to set a zero at:

$$\omega_z = \frac{\omega_{gc}}{\tan\left(20.4^o\right)} \approx 2.02 \tag{240}$$

It is expected, due to the zero insertion, a gain crossover frequency increase. However, given that the phase lead required is less than 45º, the magnitude drift will not be very significant (certainly less than 3dB!). More specifically it is expected an increase in the magnitude of the frequency response of a factor equal to

$$20\log_{10}\left(\sqrt{\left(\frac{\omega_{gc}}{2.02}\right)^2+1}\right)\approx 0.6\ \text{dB}\Rightarrow \Delta=1.06 \tag{241}$$

Since the DC gain criterion was slightly over-sized, you can reduce it to 94% of its value, ie $K=2/1.06=1.87$ leading to the final controller transfer function:

$$K(s)=1.874\frac{0.497s+1}{16.9s+1} \tag{242}$$

The new open-loop frequency response has now the following profile:

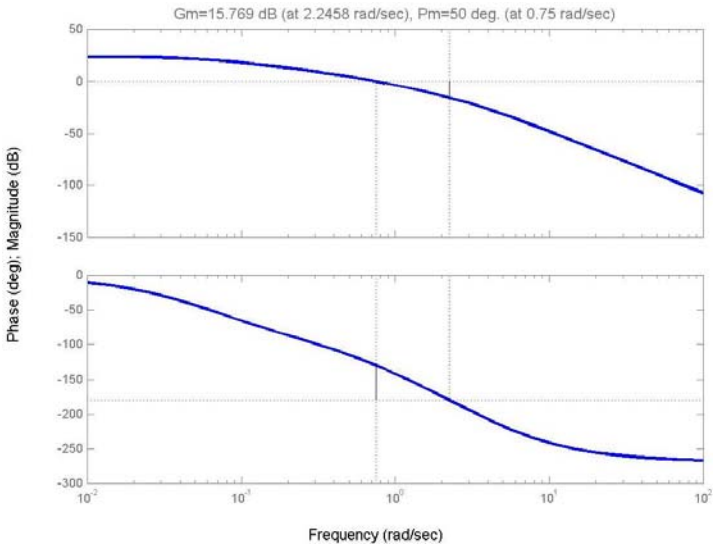


Fig 67. Frequency response of open loop end

Below, in figure 68, the closed loop step response is represented and, at figure 69, the closed loop frequency response (for unity feedback) is drawn.

As one can see, both steady-state error and phase margin criteria have been met. Additionally, from the figure 69, one can also conclude that the closed-loop system bandwidth is between the desired limits.

The next step is to discretize the controller transfer function followed by a full closed-loop system simulation. The controller discretization is typically performed by using the bilinear transform or backward Euler's method. In this

case one choose the Tustin method and neglecting the “warping” phenomenon (we have seen that for the chosen sampling frequency, this phenomenon is limited to 1%). From the discretization process follows that:

$$K(z) = \frac{0.0626 - 0.047 \cdot z^{-1}}{1 - 0.992 \cdot z^{-1}} \tag{243}$$

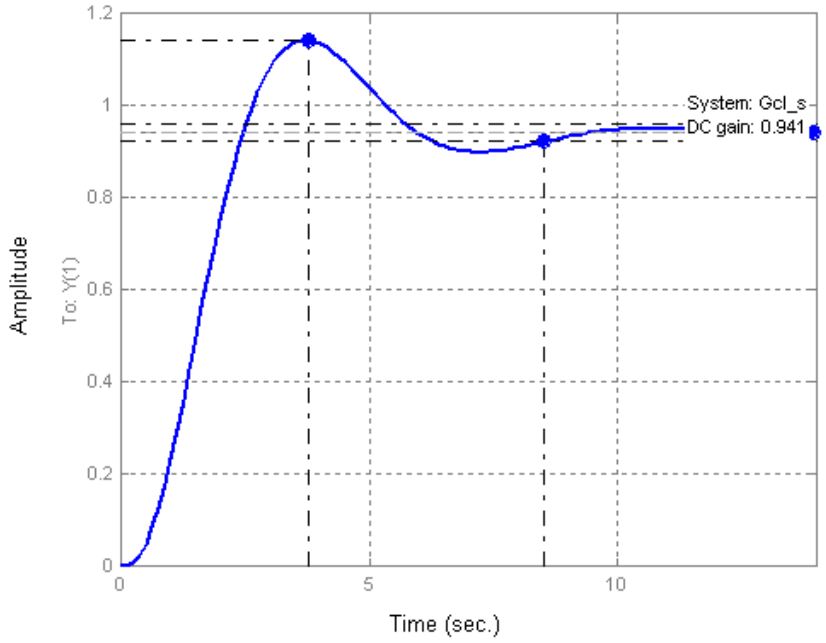


Fig 1. Closed-loop step response.

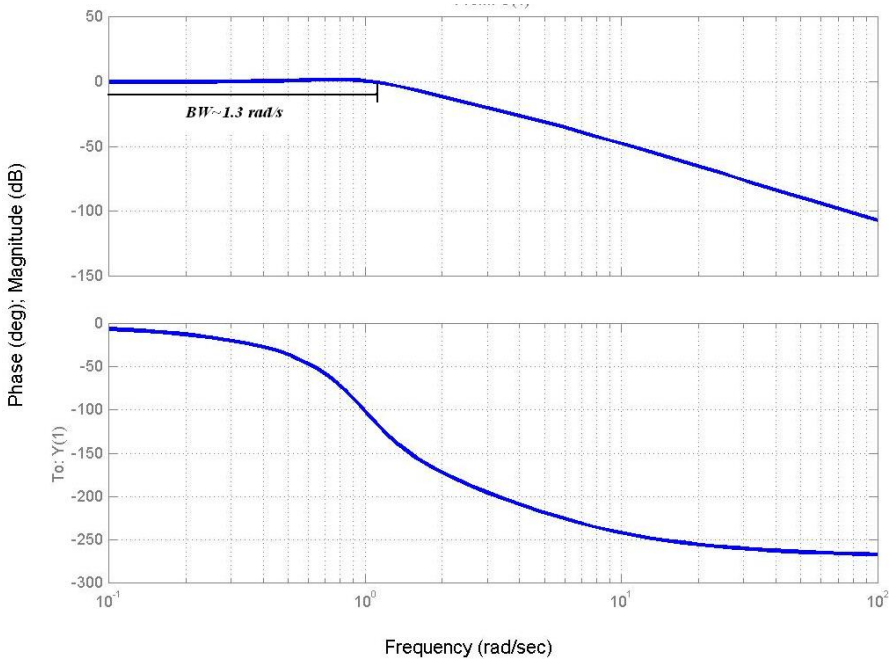


Fig 2. Closed-loop frequency response.

Considering the effect of ZOH and anti-aliasing filter (the process in series with the zero-order is discretized using the z transform) we obtain the following results illustrated by figures 70, 71 and 72.

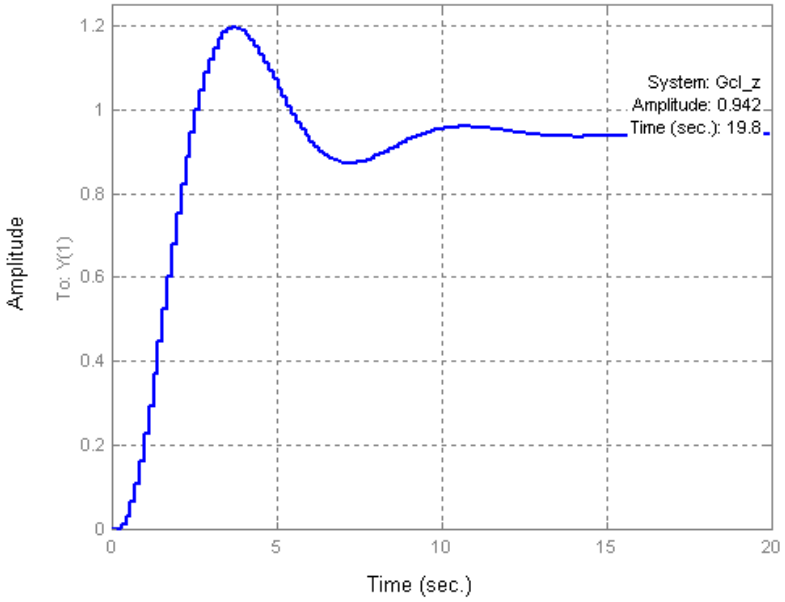


Fig 3. Digital control system closed-loop step response.

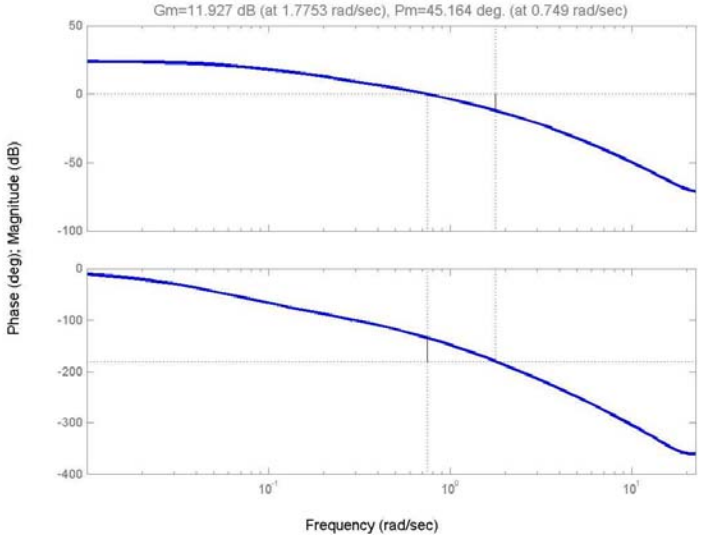


Fig 4. Digital control system open-loop frequency response.

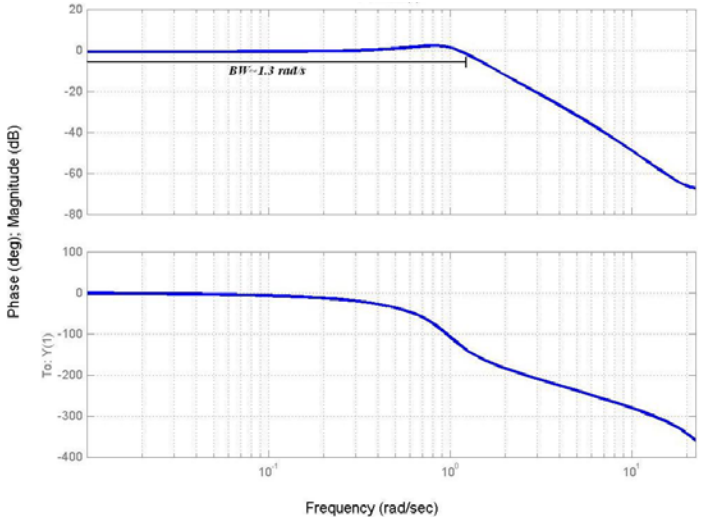


Fig 5. Digital control closed-loop frequency response.

We conclude therefore that, at least in simulation, all the design constraints were met. The next step refers to difference equation implementation and subsequent analysis of the controller performance "hardware-in-the-loop".

2.6.3.1 Digital processor effect

In the example shown the delay introduced by the processor was not considered. That is, it was assumed that the processor takes zero time to perform the I/O operations and algebraic calculation. However, in reality, this is not the case. Thus, depending on the delay effect, this new variable may, or may not, be taken into account in the design phase.

The following equation illustrates exactly what was just said. If the transfer function (243) was implemented in a digital processor, the difference equation that should be embedded possess the following structure:

$$u[k] = 0.992 \cdot u[k-1] + 0.0626 \cdot e[k] - 0.047 \cdot e[k-1] \tag{244}$$

As can be seen, the calculation of the present output control signal requires the present value of the error. In terms of actual implementation this will mean that the present value of the error signal would approximate the value of the error signal obtained after a sampling instant. This effect would be more evident as the sampling period gets smaller. Obviously this approach could have devastating effects on the behaviour of the closed-loop system.

In order to circumvent this problem, the analog controller is designed by taking into account the time delay due to information processing. In the case of the example reviewed above, this delay is reflected in a deterioration of the phase margin equal to:

$$-\omega_{gc} T \approx 6^{\circ} \tag{245}$$

In this context, a new controller is designed resulting in the following transfer function:

$$K(z) = \frac{0.07731 \cdot z^{-1} - 0.06257 \cdot z^{-2}}{1 - 0.9918 \cdot z^{-1}} \tag{246}$$

Comparing this last expression with equation (243) one observe the appearing of a pole at the origin. This pole is responsible for the delay of a sample of the

input signal.

In order to analyze the performance of this new control system, subsequently a set of images are presented in order to show the relative stability of both (243) and (246) controllers regarding the processing delay. Additionally the new controller open-loop and closed-loop Bode diagrams are presented.

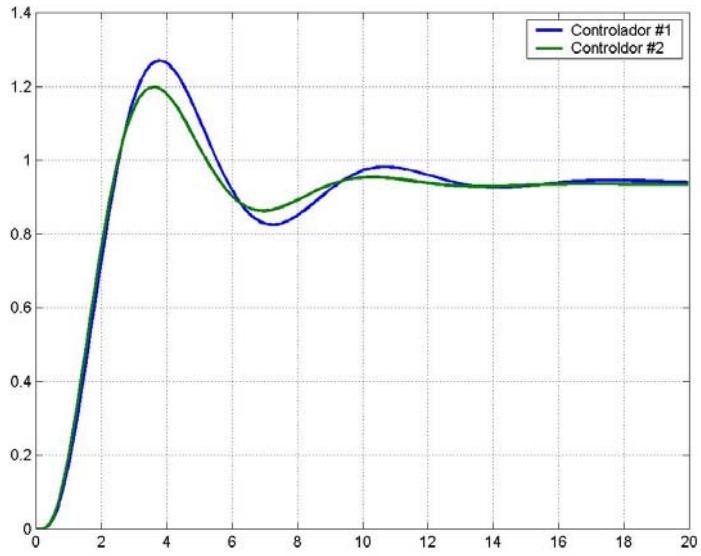


Fig 6. Unit step response of the system (including the processing delay) using the controller defined in equation {243} (Controller # 1) and the controller using {246} (Controller # 2)

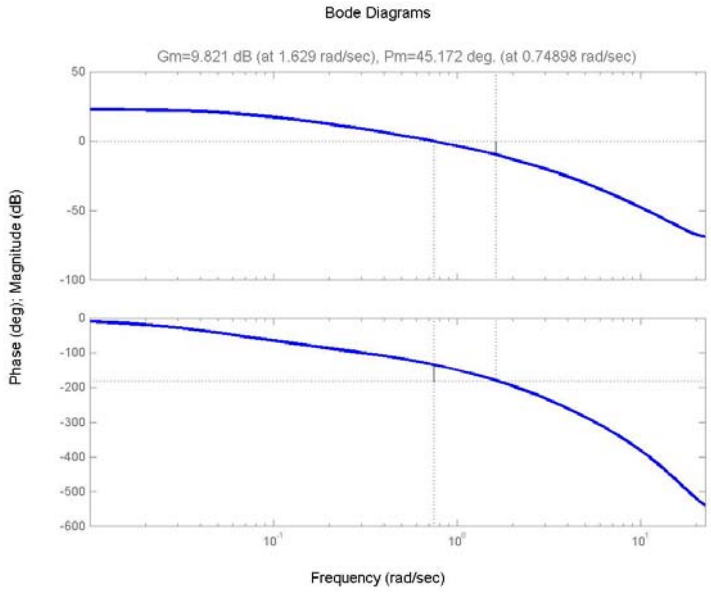


Fig 7. Frequency response for open-loop digital control system. Note the maintenance of the gain crossover frequency and phase margin.

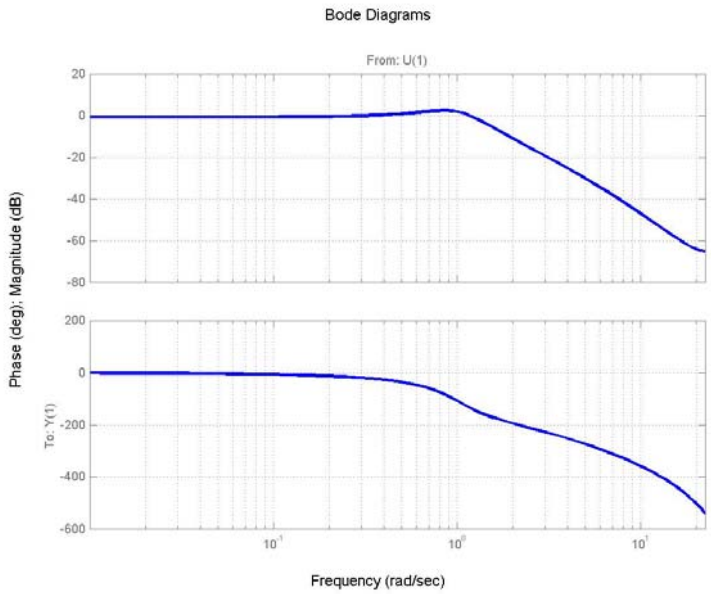


Fig 8. Frequency response in closed loop. There is a bandwidth within the limits imposed by design criteria.

Figure 73 highlights the relative stability decrease due to the processing effect delay. Thus, the act of neglecting this factor contributed to an increase in the overshoot equal to 8%. Hence the inclusion, during the design process, of the effect of all components that influence the overall dynamics, typically contributes to a better designed controller.

[◄ CHAPTER 2]

E

Exercises

PART I: Analysis and Design of Analog Control Systems

E1: *Relationship between set-point tracking and first order system pole location.*

Using MATLAB® analyze the step response of a first order system as a function of pole location. For this you should consider the system:

$$G(s) = \frac{a}{s+a}$$

for $a = \{0.1,1,10,100\}$. What conclusions can you draw?

E2: *Relationship between noise immunity and first order system pole location.*

Using MATLAB® analyze the noise immunity of a first order system as a function of pole location. For this you should consider the system:

$$G(s) = \frac{a}{s+a}$$

for $a = \{0.1,1,10,100\}$ and a unity step contaminated with white noise. The signal/noise ratio should be 6dB. What conclusions can you draw?

E3: *Frequency response of a first order system*.

Use MATLAB® to obtain the Bode plot for the system:

$$G(s) = \frac{a}{s+a}$$

with $a = \{0.1,1,10,100\}$. What conclusions can you draw? What are the values for the gain and phase margins?

E4: *Effect of σ to the step response of a second-order system*.

Use MATLAB® to simulate the unity step response of a second order

system:

$$G(s)=\frac{\omega_d^2+\sigma^2}{(s-\sigma+j\omega_d)(s-\sigma-j\omega_d)}$$

for $\sigma=\{-0.5;-1;-5\}$ and considering ω_d constant and equal to 1. Observe what happens to the following performance criteria:

- *Settling Time*
- *Overshoot*
- *Rise Time*
- *Peak Time*

E5: *Effect of ω_d in the step response of a second order system.*

Use MATLAB® to simulate the unity step response of a second order system:

$$G(s)=\frac{\omega_d^2+\sigma^2}{(s-\sigma+j\omega_d)(s-\sigma-j\omega_d)}$$

for $\omega_d=\{0.5,1,5\}$ and considering σ constant and equal to 1. Observe what happens to the following performance criteria:

- *Settling Time*
- *Overshoot*
- *Rise Time*
- *Peak Time*

E6: *Effect of ω_n in the step response of a second order system.*

Use MATLAB® to simulate the unity step response of a second order system:

$$G(s)=\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$$

for $\omega_n=\left\{\sqrt{2}/2,\sqrt{2},5\sqrt{2}\right\}$ and considering $\zeta=\sqrt{2}/2$. Observe what happens to the following performance criteria:

- *Settling Time*
- *Overshoot*
- *Rise Time*
- *Peak Time*

E7: Effect of ζ in *the step response of a second-order system.*

Use MATLAB® to simulate the unity step response of a second order

system:

$$G(s)=\frac{\omega_n^2}{\left(s+\omega_n\cdot e^{j\theta}\right)\left(s+\omega_n\cdot e^{-j\theta}\right)}$$

for $\theta=\{30,45,60\}(^{\circ})$ and considering $\omega_n=\sqrt{2}$. Observe what happens to the following performance criteria:

- *Settling Time*
- *Overshoot*
- *Rise Time*
- *Peak Time*

E8: For each of the subsequent systems, and using the MATLAB®, compare the step response of the original system with the one obtained from the alternative system approximated by dominant pole(s).

a) $G(s)=\dfrac{1}{(s+0.1)(s+1)}$

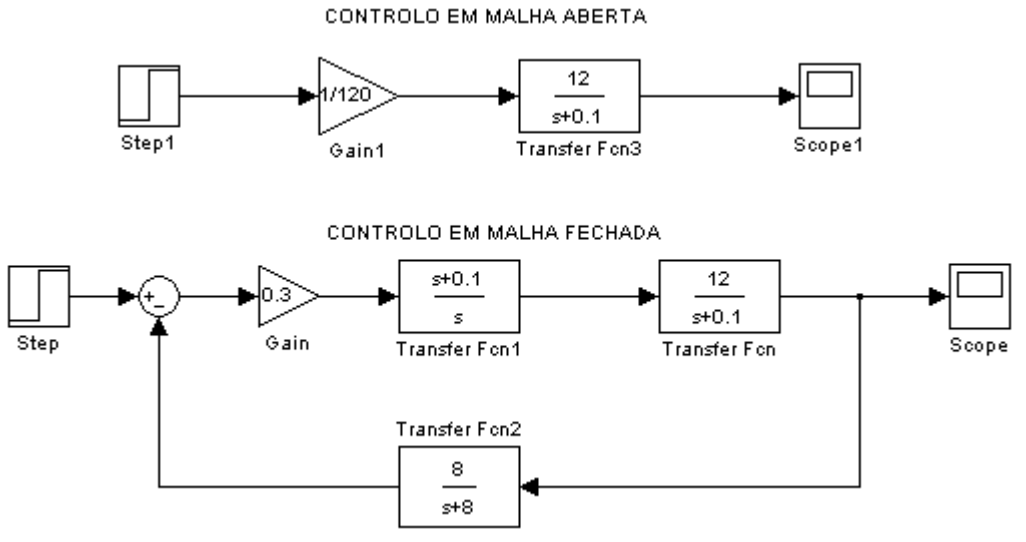
b) $G(s)=\dfrac{(s+0.2)}{(s+0.1)(s+1)}$

c) $G(s)=\dfrac{2810.1\cdot(s+4)}{(s+3.8)(s+6)(s^2+2s+17)(s^2+10s+29)}$

E9: *open loop vs. closed loop*

Consider the following transfer function of a plant whose temperature must be regulated:

$$G(s)=\frac{12}{s+0.1}$$



a) Use SIMULINK® to compute the open and closed-loop system response to a sudden change of the reference (0ºC - 100°C).
b) Simulate the response of both control systems to a step shape disturbance load with amplitude equal to -5 º C.
c) In the block diagram above, `Transfer Fcn2` represents the transfer function of a thermistor and its signal conditioning system. Analyze the effect, on the system response, to a step input with amplitude 100ºC, if the information provided by the sensor is contaminated with unity variance white noise.

E10: *Root-locus analog controller design.*

Consider a system with the following open-loop transfer function:

$$G(s) = \frac{0.8}{s+1}$$

Design a controller so that the system display, in closed loop, the following characteristics:

a) Bandwidth of about 2 rad / s and maximum steady-state error of around 5%.
b) Bandwidth of about 0.5 rad / s and maximum steady-state error of around 5%.

E11: *Root-locus analog controller design.*

Consider the following open-loop system transfer function:

$$G(s) = \frac{5}{(s+2)(s+3)}$$

Using the root locus design a controller so that the system display the following characteristics.

- Unit step error less than 0.05
- Phase Margin = 45 °
- Closed-loop bandwidth approximately equal to 6 rad / s.

E12: *Tuning a PID controller using the Ziegler-Nichols rules.*

Consider the following open-loop system:

$$G(s) = \frac{1}{s(s+1)(s+3)}$$

Project a PID controller using the Ziegler and Nichols tuning rules. Additionally, and using MATLAB®, analyze the response of the closed-loop system (unity feedback) with and without compensation.

E13: *Tuning a PID controller using the Ziegler-Nichols rules.*

Consider the following open-loop system:

$$G(s) = \frac{e^{-0.1s}}{(s+1)}$$

Project a PID controller using the Ziegler and Nichols tuning rules. Additionally, and using MATLAB®, analyze the response of the closed-loop system (unity feedback) with and without compensation.

E14: *Analytic PID design.*

Consider the following open-loop system:

$$G(s) = \frac{400}{s^2 + 30s + 200}$$

Design a PID controller (analytically) in order the system to display, in closed loop, the following characteristics:

- Error = 0.1 the unit ramp
- Overshoot = 10% and settling time = 2s.

E15: *Bode plot phase lead controller design.*

Consider the system:

$$G(s) = \frac{1}{s^2 + 0.2s + 0.1}$$

Design a controller so that the system display a steady-state error less than or equal to 1% and a phase margin of around 45 degrees.

E16: *Phase lead controller design.*

Design a lead controller for the system:

$$G(s) = \frac{72}{(s+1)(s+3)^2}$$

so that it displays the following characteristics.

- Error in steady state (the step) lower than or equal to 0.1
- Phase margin of 45 ° and bandwidth approximately equal to 1rad / s.

E17: Bode plot phase lag controller design.

Design a lag compensator for the system:

$$G(s)=\frac{10(s+5)}{(s+15)(s^2+8s+20)}$$

So the closed-loop system exhibits the following characteristics.

- Steady-state error less than or equal to 10%
- Overshoot exceeding 5%

E18: Design a compensator for the following system:

$$G(s)=\frac{10}{s(s+5)}$$

So that it displays the following characteristics.

- Error in steady state (the unit ramp) less than or equal to 5%
- Phase margin of around 40 º and bandwidth near 2 rad / s

PART II: Sampling and Reconstruction

E19: Determine $E^*(s)$ for the following signals:

a) $e(t)=u(t)$

b) $e(t)=e^{-t}$

c) $e(t)=t$

E20: Determine $E^*(s)$ for the following transfer functions:

a) $E(s)=\frac{1}{(s+1)(s+2)}$

b) $E(s)=\frac{s+2}{(s+1)s^2}$

c) $E(s)=\frac{s^2+5s+6}{s(s+4)(s+5)}$

E21: Determine the transfer function and frequency response of an ideal *first-order-hold*.

E22: Determine the Z transform $E(z)$ for the following signals:

a) $e(nT) = u(nT)$

b) $e[n] = e^{-n}$

c) Time-series obtained by sampling $e(t) = t$ every second.

E23: Determine the Z transform for the following transfer functions:

a) $E(s) = \dfrac{1}{(s+1)(s+2)}$

b) $E(s) = \dfrac{s+2}{(s+1)s^2}$

c) $E(s) = \dfrac{s^2+5s+6}{s(s+4)(s+5)}$

E24: Determine the modified Z transform $E(z,m)$ for the systems presented in the previous exercise.

E25: Find the modified Z transform for the following transfer functions:

a) $E(s) = \dfrac{20e^{-0.3Ts}}{(s+2)(s+5)}$, $T = 1s$

b) $E(s) = \dfrac{(s+2)e^{-0.2Ts}}{(s+1)s^2}$

c) $E(s) = \dfrac{2e^{-0.75s}}{s^2+2s+5}$, $T = 0.2s$

E26: Determine the discrete sequences $e(kT)$ associated with the following Z transforms:

a) $E(z) = \dfrac{z}{z^2-3z+2}$

b) $E(z) = \dfrac{-3.894z}{z^2+0.6065}$

c) $E(z) = \dfrac{z}{(z-1)^2}$

E27: Solve the following difference equations using Z transform

a) $m[k] = e[k] - e[k-1] - m[k-1]$

b) $x[k] - 3x[k-1] + 2x[k-2] = e[k]$

c) $y[k+2] - 6 \cdot y[k+1] + 8y[k] = e[k]$ to $y[0] = 1$ and $y[1] = 2$

E28: Consider the discrete system characterized by the difference equation:

$$y[k+1] = a \cdot y[k] + b \cdot x[k] \text{ where } 0 < a < 1 \text{ and } y[0] = 0$$

a) Determine the impulse response and make a sketch of the result.

b) Calculate the unit step response and sketch the result.

c) Determine the static gain of the system.

E29: Consider the discrete system characterized by the following differences equation:

$$y[k] = y[k-1] - 0.25 \cdot y[k-2] + x[k-1] + 0.5 \cdot x[k-2]$$

Determine the transfer function $Y(z)/X(z)$, identify the poles and zeros and represent them in the Z plane. What can we say about system stability?

PART IV: Open-loop discrete-time system response

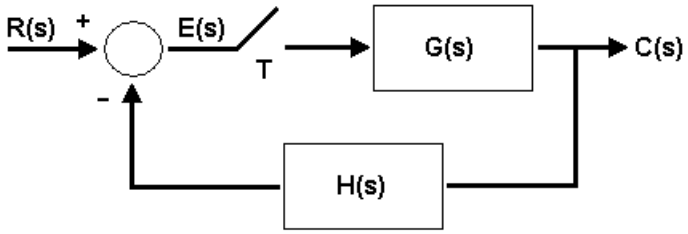
E30: Prove that, for a system consisting of an ideal sampler/zero order hold with input $E(s)$ and output $C(s)$, one have $C(z) = G(z)E(z)$ where $G(s) = C(s)/E^*(s)$.

E31: Consider a system consisting of an ideal sampler/zero order holder in series with a process with transfer function:

$$G(s) = \frac{1}{s+1}$$

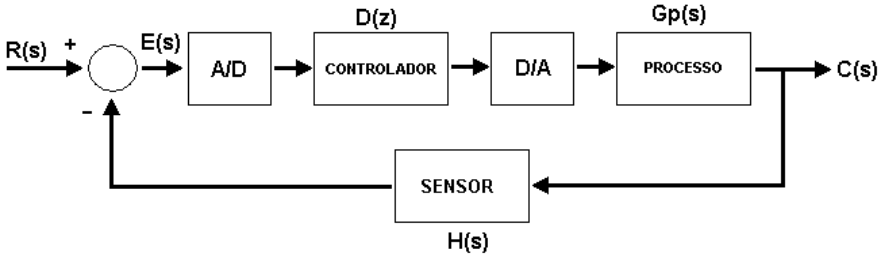
Assuming a unit step input determines $C(z)$ and $c(kT)$.

E32: Determine the closed-loop transfer function of the following system:

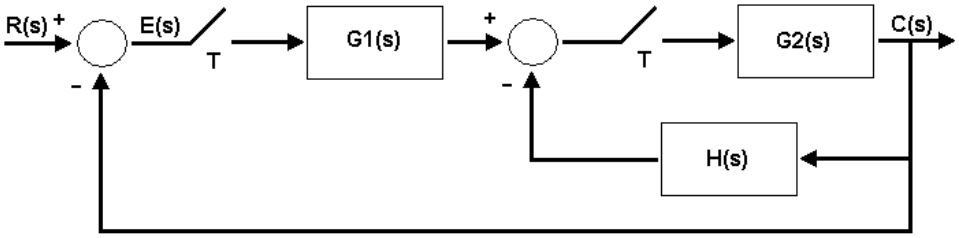


E33: Consider the following closed-loop control systems. Obtain the transfer function $C(z)/R(z)$.

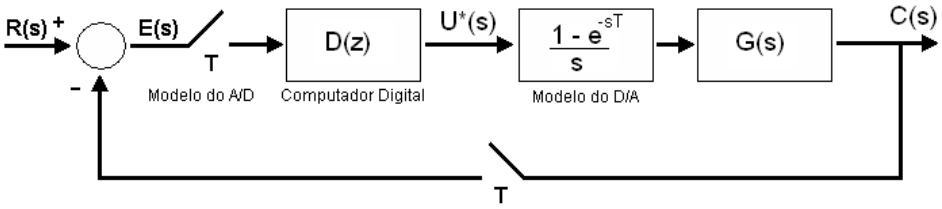
a)



b)



E34: For the following figure compute:

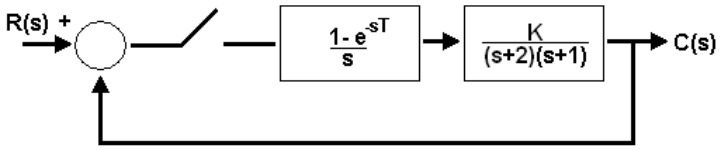


a) $C(z)/R(z)$.

b) The response $C(z)$ for the case where $G(s) = a/s + a$, $e^{-aT} = 0.5$ and the computer algorithm $u[k] = u[k-1] + k \cdot e[k]$ and $r(t) = u(t)$.

PART VI: Discrete-time systems stability analysis

E35: Check for what values K is the following system is stable.



Note: Consider $e^{-T} = 0.5$

a) Using the Routh-Hurwitz criterion for discrete systems.

b) Using the Jury stability criterion.

E36: Using the Jury's criterion characterized the stability of the following discrete system.

$$G(z) = \frac{1}{z^3 - 1.1z^2 - 0.1z + 0.2}$$

PART VII: Digital Control Design

E37: Consider a process with transfer function:

$$G(s) = \frac{1}{s(s+1)}$$

design a digital controller capable of implementing the transfer function:

$$K(s) = \frac{70}{(s+2)(s+1)}$$

for:

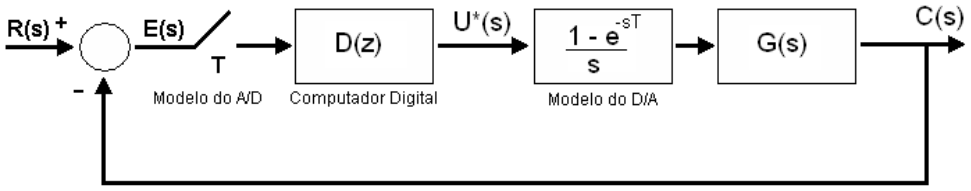
a) A sampling frequency of 20Hz.

b) A sampling frequency of 40Hz.

Note: Use in both cases an approximation of Euler (forward)

c) Determine, with the MATLAB®, the system response $K(s)G(s)$ to a unit step and compare it with the system response when the controller is digital.

E38: Consider the following feedback system



where

$$G(s) = \frac{1}{s^2 + 2s}$$

It is intended that the unity step overshoot in less than 10%, the rise time less than 5 seconds and the unit ramp error to lower than 2%.

a) Under these conditions determine an analog controller transfer function in order to satisfy these requirements.

b) Determine the digital controller transfer function obtained by emulation of the continuous controller using the bilinear transformation. What sampling frequency should be used?

[◄ EXERCISES]

A Appendices

A1. Laplace Transform

The Laplace transform is used to convert time domain signals and systems into a set of equations expressed in terms of a complex variable commonly designated by 's'. The Laplace transform can be unilateral or bilateral. The asymmetry of the unilateral version imply system causality since, in its specification, it's assumed that $x(t) = 0$ for $t < 0$.

The Laplace transform of a time signal $x(t)$ are described mathematically by:

$$X(s) = \mathcal{L}\{x(t)\} = \int_{-\infty}^{+\infty} x(t) \cdot e^{-st} dt \quad \text{Version (bilateral)}$$

or

$$X(s) = \mathcal{L}\{x(t)\} = \int_{0}^{+\infty} x(t) \cdot e^{-st} dt \quad \text{(One-sided version)}$$

where $\mathcal{L}\{\cdot\}$ refers to the transformation of Laplace. The complex variable s can be decomposed into $s = \sigma + j\omega$ where σ is the real part of s and ω is the imaginary part. The set of values for s which make the integral convergent is called the Laplace transform convergence region.

In the control systems framework, the application of the Laplace transform is closely related to the fact that, in most cases, the physical systems dynamics are expressed by constant coefficients ordinary differential equations. The Laplace transform application to this equations type turns them into simple polynomial equations. For example consider the following case:

$$y(t) = \frac{dx(t)}{dt}$$

The Laplace transform of this differential equation leads to:

$$Y(s) = \mathcal{L}\{y(t)\} = \int_0^{+\infty} \frac{dx(t)}{dt} e^{-st} dt$$

The antiderivative of the integral's argument can be computed as ,

$$P\left[\frac{dx(t)}{dt}e^{-st}\right] = P\left[\frac{dx(t)}{dt}e^{-st}\right]e^{-st} - P\left[P\left[\frac{dx(t)}{dt}\right]\cdot\frac{d}{dt}e^{-st}\right]$$
$$= x(t)e^{-st} + s\cdot P\left[x(t)e^{-st}\right]$$

and so,

$$Y(s) = x(t)e^{-st}\Big|_{-\infty}^{+\infty} + s\cdot\int_0^{+\infty} x(t)e^{-st}$$

Since,

$$X(s) = \int_0^{+\infty} x(t)e^{-st}dt$$

then

$$Y(s) = sX(s) - x(0)$$

where $x(0)$ is the initial value of $x(t)$ at time instant $t = 0$. If all the initial conditions are zero then the derivative operation in the time-domain is equivalent to multiply by s in the complex frequency domain.

One can return back to the time domain, from the Laplace domain, by using the inverse Laplace transform. This transform is presented formally as a contour integral over $s = \sigma + j\omega$ with the following form:

$$x(t) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} X(s)\cdot e^{st} ds$$

And often can be solved using the Cauchy's residue theorem.

To conclude this appendix, below are presented some of the fundamental properties of the Laplace transform.

Linearity

If $x_1(t) \rightleftarrows X_1(s)$ and $x_2(t) \rightleftarrows X_2(s)$ then

$$\mathcal{L}\{x_1(t)+x_2(t)\} = \mathcal{L}\{x_1(t)\} + \mathcal{L}\{x_2(t)\} = X_1(s) + X_2(s)$$

Homogeneity

If $x(t) \rightleftarrows X(s)$ then

$$\mathcal{L}\{\alpha x(t)\} = \alpha \mathcal{L}\{x(t)\} = \alpha X(s) \quad , \forall \alpha$$

Final Value Theorem

If $x(t) \rightleftarrows X(s)$ then

$$\lim_{t\to\infty} x(t) = \lim_{s\to 0} sX(s)$$

Initial Value Theorem

If $x(t) \rightleftarrows X(s)$ then

$$\lim_{t\to 0} x(t) = \lim_{s\to\infty} sX(s)$$

Differentiation

If $x(t) \rightleftarrows X(s)$ then

$$\mathcal{L}\left\{\frac{d^n x(t)}{dt^n}\right\} = s^n X(s) - \sum_{k=1}^{n} s^{n-k} \frac{d^{k-1}x(0)}{dt^{k-1}}$$

Integration

If $x(t) \rightleftarrows X(s)$ then

$$\mathcal{L}\left\{P^n\left[x(t)\right]\right\} = \frac{X(s)}{s^n} + \sum_{k=1}^{n} \frac{P^k\left[x(0)\right]}{s^{n-k+1}}$$

A2. Fourier Theory

The aim of the Fourier transforms is to convert a usually "complex" signal into a set of simplest treatment signals. For this the Fourier transform uses trigonometric functions as basis functions: the signal is decomposed in a linear combination of sinus and co-sinus. In particular, the signal is decomposed as a weighted sum of complex exponentials. The importance of this strategy rests on the fact that the linear time-invariant system response to a complex exponential signal is still a complex exponential signal with the same frequency and probably with different amplitude and phase[6]. Depending on the involved signal type (periodic, aperiodic, etc.) the representation, in terms of complex exponentials, can take the following aspects:

- **Fourier series**

Any periodic signal can be written as a weighted sum of harmonically related complex exponentials. That is a infinite periodic signal $x(t)$ with fundamental period T_o can be written as:

$$x(t) = \sum_{k=-\infty}^{+\infty} C_k \cdot e^{jk\omega_o t}$$

where $\omega_o = \frac{2\pi}{T_o}$ refers to the fundamental angular frequency and $C(k)$ represents a weighting function computed by:

$$C_k = \frac{1}{T_o} \int_{T_o} x(t) \cdot e^{-jk\omega_o t} dt$$

In other words, the Fourier series coefficients are calculated from the integral over one period of the signal.

- **Fourier transform of aperiodic signals**

Aperiodic signals can also be represented as a linear combination of complex exponentials. However, in this case, the exponentials are not harmonically related but infinitely close in frequency ω . Thus, if $x(t)$ is an aperiodic signal that admits representation in the Fourier domain, the synthesis and analysis

[6] This property is often given the name of sinusoidal fidelity.

equations are:

$$x(t) \underset{F^{-1}}{\overset{F}{\rightleftarrows}} X(j\omega)$$

$$X(j\omega) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j\omega t} dt$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(j\omega) \cdot e^{j\omega t} d\omega$$

- **Fourier transform of periodic signals**

If $x(t)$ is periodic with period T_0 then,

$$X(j\omega) = \sum_{k=-\infty}^{+\infty} 2\pi C_k \delta\left(\omega - k\omega_o\right)$$

where

$$C_k = \frac{1}{T_o} \int_{T_o} x(t) \cdot e^{-jk\omega_o t} dt$$

From the previous equations one verifies the existence of a relationship between the periodic signal Fourier transform and the series Fourier coefficients. That is, in terms of spectral representation, the Fourier transform of a periodic signal is always a set of impulses located at multiple harmonics of the fundamental frequency and weighted by the factor $2\pi C_k$.

Note: For a continuous-time signal to admit Fourier representation it must cope with the following three conditions (known as Dirichlet conditions):

- During a period or finite time interval $x(t)$ must have a finite number of maxima and minima.
- During a period or finite time interval $x(t)$ must have a finite number of discontinuities.
- The signal must be absolutely integrable, i.e. $\int_{-\infty}^{+\infty} \left|x(t)\right| dt < \infty$

- **Fourier transform of discrete signals**

A discrete sequence $x[n]$ has Fourier transform $X(e^{j\omega})$ given by:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] \cdot e^{-j\omega n}$$

if the series exists. The inverse Fourier transform of a discrete signal is:

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega$$

- **Discrete Fourier Transform (DFT)**

The discrete Fourier transform of a discrete-time signal is, itself, a discrete sequence and consists of Fourier transform samples taken at N equally spaced points in the frequency:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi kn}{N}} \quad , 0 \le k \le N-1$$

where

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot e^{j\frac{2\pi kn}{N}} \quad , 0 \le n \le N-1$$

Note that in the case of discrete-time signals, the frequency ω actually refers to the digital frequency ω_d. The relationship between analog and digital frequency is:

$$\omega_d = \omega \cdot T$$

where T refers to the sampling period. For an analog frequency equal to the sampling frequency (inverse of the sampling period), the digital frequency is 2π radians per sample. Moreover, it appears that the Fourier transform for discrete signals is periodic with period 2π. In order to validate look to the following proof:

$$X(e^{j(\omega+2k\pi)}) = \sum_{n=-\infty}^{+\infty} x[n] \cdot e^{-j(\omega+2k\pi)n} = \sum_{n=-\infty}^{+\infty} x[n] \cdot e^{-j\omega n} \cdot e^{-j2k\pi n} \quad , \forall k \in \mathbb{Z}$$

as n only takes integer values then $e^{-j2\pi kn} = \cos(2k\pi n) - j\sin(2k\pi n) = 1$ and so,

$$X(e^{j(\omega+2k\pi)}) = \sum_{n=-\infty}^{+\infty} x[n] \cdot e^{-j\omega n} = X(e^{j\omega})$$

A3. Some Laplace Transform Pairs

Function Time $e(t), t > 0$	*Laplace Transform* $E(s)$
$\delta(t)$	1
$\delta(t-t_0)$	e^{-st_0}
$u(t)$	$\frac{1}{s}$
t	$\frac{1}{s^2}$
$\frac{t^2}{2}$	$\frac{1}{s^3}$
t^{k-1}	$\frac{(k-1)!}{s^k}$
e^{-at}	$\frac{1}{s+a}$
$t \cdot e^{-at}$	$\frac{1}{(s+a)^2}$
$t^k \cdot e^{-at}$	$\frac{(k-1)!}{(s+a)^k}$
$1-e^{-at}$	$\frac{a}{s(s+a)}$
$t-\frac{1-e^{-at}}{a}$	$\frac{a}{s^2(s+a)}$
$1-(1+at)e^{-at}$	$\frac{a^2}{s(s+a)^2}$
$e^{-at}-e^{-bt}$	$\frac{b-a}{(s+a)(s+b)}$
$\sin(at)$	$\frac{a}{s^2+a^2}$
$\cos(at)$	$\frac{s}{s^2+a^2}$
$\frac{1}{b}e^{-at}\sin(bt)$	$\frac{1}{(s+a)^2+b^2}$
$e^{-at}\cos(bt)$	$\frac{a^2+b^2}{s\left((s+a)^2+b^2\right)}$
$\frac{1}{ab}+\frac{e^{-at}}{a(a-b)}+\frac{e^{-bt}}{b(b-a)}$	$\frac{1}{s(s+a)(s+b)}$

A4. Some transform pairs Z

Function Time $e(t), t>0$	*Z transform* $E(z)$	*Modified Z Transform* $E(z,m)$
$u(t)$	$\frac{z}{z-1}$	$\frac{1}{z-1}$
t	$\frac{Tz}{(z-1)^2}$	$\frac{mT}{z-1}+\frac{Tz}{(z-1)^2}$
$\frac{t^2}{2}$	$\frac{T^2 z(z+1)}{2(z-1)^3}$	$\frac{T^2}{2}\left(\frac{m^2}{z-1}+\frac{2m+1}{(z-1)^2}+\frac{2}{(z-1)^3}\right)$
t^{k-1}	$\lim_{a\to 0}(-1)^{k-1}\frac{\partial^{k-1}}{\partial a^{k-1}}\left(\frac{z}{z-e^{-aT}}\right)$	$\lim_{a\to 0}(-1)^{k-1}\frac{\partial^{k-1}}{\partial a^{k-1}}\left(\frac{e^{-amT}}{z-e^{-aT}}\right)$
e^{-at}	$\frac{z}{z-e^{-aT}}$	$\frac{e^{-amT}}{z-e^{-aT}}$
$t\cdot e^{-at}$	$\frac{Tze^{-aT}}{\left(z-e^{-aT}\right)^2}$	$\frac{Te^{-amT}\left(e^{-aT}+m\left(z-e^{-aT}\right)\right)}{\left(z-e^{-aT}\right)^2}$
$t^k\cdot e^{-at}$	$(-1)^k\frac{\partial^k}{\partial a^k}\left(\frac{z}{z-e^{-aT}}\right)$	$(-1)^k\frac{\partial^k}{\partial a^k}\left(\frac{e^{-amT}}{z-e^{-aT}}\right)$
$1-e^{-at}$	$\frac{z\left(1-e^{-aT}\right)}{(z-1)\left(z-e^{-aT}\right)}$	$\frac{1}{(z-1)}-\frac{e^{-amT}}{\left(z-e^{-aT}\right)}$
$t-\frac{1-e^{-at}}{a}$	$\frac{z\left(z\left(aT-1+e^{-aT}\right)+\left(1-e^{-aT}-aTe^{-aT}\right)\right)}{a(z-1)^2\left(z-e^{-aT}\right)}$	$\frac{T}{(z-1)^2}+\frac{amT-1}{a(z-1)}+\frac{e^{-amT}}{a\left(z-e^{-aT}\right)}$
$1-(1+at)e^{-at}$	$\frac{z}{z-1}-\frac{z}{z-e^{-aT}}-\frac{aTe^{-aT}z}{\left(z-e^{-aT}\right)^2}$	$\frac{1}{z-1}-\left(\frac{1+amT}{z-e^{-aT}}+\frac{aTe^{-aT}}{\left(z-e^{-aT}\right)^2}\right)e^{-amT}$
$e^{-at}-e^{-bt}$	$\frac{\left(e^{-aT}-e^{-bT}\right)z}{\left(z-e^{-aT}\right)\left(z-e^{-bT}\right)}$	$\frac{e^{-amT}}{\left(z-e^{-aT}\right)}-\frac{e^{-bmT}}{\left(z-e^{-bT}\right)}$
$\sin(at)$	$\frac{z\sin(aT)}{z^2-2z\cos(aT)+1}$	$\frac{z\sin(amT)+\sin\left((1-m)aT\right)}{z^2-2z\cos(aT)+1}$
$\cos(at)$	$\frac{z\left(z-\cos(aT)\right)}{z^2-2z\cos(aT)+1}$	$\frac{z\cos(amT)-\cos\left((1-m)aT\right)}{z^2-2z\cos(aT)+1}$
$\frac{1}{b}e^{-at}\sin(bt)$	$\frac{ze^{-aT}\sin(bT)}{b\left(z^2-2ze^{-aT}\cos(bT)+e^{-2aT}\right)}$	$\frac{e^{-amT}\left(z\sin(bmT)+e^{-aT}\sin\left((1-m)bT\right)\right)}{b\left(z^2-2ze^{-aT}\cos(bT)+e^{-2aT}\right)}$
$e^{-at}\cos(bt)$	$\frac{z^2-ze^{-aT}\cos(bT)}{z^2-2ze^{-aT}\cos(bT)+e^{-2aT}}$	$\frac{e^{-amT}\left(z\cos(bmT)+e^{-aT}\sin((1-m)bT)\right)}{z^2-2ze^{-aT}\cos(bT)+e^{-2aT}}$

[APPENDICES ◄]

R References

[1] Ástrom, K and Wittenmark, B., "Adaptive Control", Addison-Wesley 1985, ISBN 0-201-55866-1

[2] Coelho, JP, "Digital Signal Processing", 2001-2002 Bragança Polytechnic Institute - School of Technology and Management.

[3] Coelho, JP, "Sensors and Actuators", 2003-2005 Bragança Polytechnic Institute - School of Technology and Management.

[4] Distefano, JJ and Stubberud, and AR Williams, IJ "Feedback and Control Systems", McGraw-Hill 1972

[5] Dutton, K. Thompson S, Barraclough, B. "The Art of Control Enginnering" Prentice Hall 1997. ISBN 0-201-17545-2

[6] Franklin, G.F., Powell, and Emami-Naeini J.D., A. "Feedback Control of Dynamic Systems", Addison-Wesley 1994, ISBN 0-201-52747-2

[7] Kowalczuk, Z. *Discrete Approximation of Continuous-time systems: a survey*. IEE Proceedings-G, Vol 140, No. 4, pp 264-278, 1993.

[8] Moscinski, J. and Ogonowski, Z., "Advanced Control with Matlab and Simulink," Ellis Horwood Limited 1995, ISBN 0-13-309667-4

[9] Ogata, K. "Discrete-Time Control Systems", Prentice Hall 1995, ISBN 0-13-034281-5.

[10] Ogata, K., "Modern Control Engineering". Prentice Hall 1997. ISBN 0-13-043245-8

[11] Oppenheim, A.V, Willsky, and Nawab A.S., S.H.. "Signals and Systems". Prentice Hall 1997. ISBN 0-13-651175-9

[12] Oppenheim and Shaffer. "Discrete-Time Signal Processing", Prentice-Hall 1998, ISBN 0-13-754920-2

[13] Phillips, C.L. and Nagle, HT "Digital Control System Analysis and Design." Prentice Hall 1995. ISBN 0-13-317729-7

[14] Shahian, B. and Hassul, M. "Control System Design Using Matlab". Prentice Hall 1993. ISBN 0-13-014557-2.

[15] Svrcek, W. Mahoney, D. and Young, B.. "Real-Time Approach to Process Control", John Wiley & Sons 2000, ISBN 0-471-80363-4.

[◄ REFERENCES]

Index

[INDEX ◄]