ENGINEERING
DEPARTMENT OF ELECTRICAL,
COMPUTER, AND SOFTWARE ENGINEERING

ELECTENG 332: Control Systems

Lecture Notes

Nicholas Russell
University of Auckland
Department of Electrical, Computer, and Software Engineering

Wednesday, October 23, 2024

Table of contents

Introduction

> *Dear god not another one of these.*

This is my "coursebook", or rather, a collection of notes for the course ELECTENG 332: Control Systems. The notes are written in Quarto, a markdown-based document processor that supports LaTeX and code blocks, and requires this introduction file.

Please skip this bit, there's nothing to see here.

Part I

Akshya's Content

Chapter 1

Basics of Signals and Systems

Learning Outcomes

After completing this module, you should be able to:

1. Understand the uniqueness of the exponential function
2. Understand the concept of engineering infinity
3. Understand the concept of complex frequency
4. Understand the concept of signals, and be able to classify them
5. Understand the concept of systems, and be able to classify them
6. Understand the concept of control systems

1.1 The Importance of the Exponential Function

The Exponential function, written as either e^{ax} or e^{at} depending on whether it is $f(t)$ or $f(x)$, has properties that make it mathematically unique.

1. The derivative (rate of change) of the exponential function is the exponential function itself. More generally, this is a function whose rate of change is proportional to the function itself.

$$\frac{\mathrm{d}e^{ax}}{\mathrm{d}x} = ae^{ax} \tag{1}$$

2. The integral of the exponential function is also the exponential function itself.

$$\int e^{ax}\,\mathrm{d}x = \frac{1}{a}e^{ax} \tag{2}$$

1.2 The Concept of Engineering Infinity

Consider a signal e^{-at}. The time constant for this signal is $T = \frac{1}{a}$. Theoretically, the signal is meant to decay to zero as time approaches infinity, i.e.

$$\lim_{t\to\infty} e^{-at} = 0 \tag{3}$$

But in practice, this is not the case, as its value will be very, very small after five time constants $5T$ (or 5τ). This is the ***Concept of Engineering Infinity***. The signal will never reach zero, but it will be so small that it can be considered zero for all practical purposes. This is a very important concept in control systems, as it allows us to simplify our calculations and analysis.

1.3 The Concept of Complex Frequency

Complex frequency is found commonly in electrical engineering. It is often notated as $j\omega$ or $s = \sigma \pm j\omega$. These frequencies always come in pairs, so the use of $\pm$ is implicit to this, as complex numbers have complex conjugates (normally notated by z^* or $\bar{z}$). i.e. $s = \sigma + j\omega$ has the conjugate $s = \sigma - j\omega$. This is backed up by De Moivre's formula which is defined mathematically as:

$$\forall x \in \mathbb{R}, \quad \forall n \in \mathbb{Z} \tag{4}$$
$$e^{jnx} = \cos(nx) + j\sin(nx) \tag{5}$$

Or more generally for our applications (this is also known as Euler's formula):

$$e^{jx} = \cos(x) + j\sin(x) \tag{6}$$
$$\text{Where}\, x \in \mathbb{R}\, (x \text{ is real}) \tag{7}$$
$$\text{and}\, j \equiv i = \sqrt{-1} \tag{8}$$

This means that:

A complex frequency $j\omega$ represents a pure sinusoidal signal of frequency ω rad/s

For example, if a signal has a complex frequency $j314$ rad/s, then this responds to a pure sinusoid of frequency 314 rad/s (i.e. 50 Hz).

Furthermore:

A complex frequency $s = \sigma + j\omega$ represents an exponentially damped signal of frequency $j\omega$ rad/s, and decays/amplifies at a rate decided by σ

1.4 What are Signals?

1.4.1 Introduction

It is difficult to find a unique definition of a signal. However in the context of this course, we give a workable definition which suits most of out purposes as:

A Signal conveys information about a physical phenomenon which evolves in time or space.

Examples of such signals include: Voltage, Current, Speech, Television, Images from remote space probes, Voltages generated by the heart and brain, Radar and Sonar echoes, Seismic vibrations, Signals from GPS satellites, Signals from human genes, and countless other applications.

1.4.2 Energy and Power Signals

Energy Signals

A signal is said to be an energy signal if and only if it has finite energy.

Power Signals

A signal is said to be a power signal if and only if the average power of the signal is finite and non-zero.

Instantaneous Power

The instantaneous power $p(t)$ of a signal $x(t)$ is expressed as:

$$p(t) = x^2(t) \tag{9}$$

Continuous-Time Signal Energy

The total energy of a continuous-time signal $x(t)$ is given by:

$$E = \lim_{T\to\infty} \int_{-T/2}^{T/2} x^2(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} x^2(t)\,\mathrm{d}t$$

Complex Valued Signal Energy

For a complex valued signal:

$$E = \int_{-\infty}^{\infty} |x(t)|^2\,\mathrm{d}t$$

Average Power

Since power equals to the time average of the energy, the average power is given by:

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)\,\mathrm{d}t = \frac{E}{T}$$

Note that during calculation of energy, we average the power over an indefinitely large interval.

A signal with finite energy has zero power and a signal with finite power has infinite energy.

Furthermore, some additional concepts of note:

a. A signal can not both be an energy and a power signal. This classification of signals based on power and energy are mutually exclusive.

b. However, a signal can belong to neither of the above two categories.

c. The signals which are both deterministic and non-periodic have finite energy and therefore are energy signals. Most of this signals, in practice, belong to this category.

d. Periodic signals and random signals are essentially power signals.

e. Periodic signals for which the area under $|x(t)|^2$ over one period is finite are energy signals.

1.4.3 Examples

1.4.3.1 Unit Step Function

Consider a unit step function defined as:

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

Determine whether this is an energy signal or a power signal or neither.

Solution: Let us compute the energy of this signal as:

$$E = \int_{-\infty}^{\infty} [u(t)]^2\,\mathrm{d}t = \int_{0}^{\infty} [0]^2\,\mathrm{d}t = \int_{0}^{\infty} [1]^2\,\mathrm{d}t = \infty \tag{11}$$

Since the energy of this signal is infinite, it cannot be an energy signal. Let us compute the power of this signal as:

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [u(t)]^2\,\mathrm{d}t = \lim_{T\to\infty} \frac{1}{T} \int_{0}^{T/2} [u(t)]^2\,\mathrm{d}t = \frac{1}{2} \tag{12}$$

The power of this signal is finite. Hence, this is a power signal.

1.4.3.2 Exponential Function

Consider an exponential function defined as:

$$x(t) = e^{-at}u(t), \text{ where } u(t) \text{ is the unit step signal, } a > 0 \tag{13}$$

Classify this signal as an energy, power, or neither.

Solution: Let us compute the energy of this signal as:

$$E = \int_{-\infty}^{\infty} [x(t)]^2 \, \mathrm{d}t = \int_{0}^{\infty} \left[e^{-at}\right]^2 \mathrm{d}t = \int_{0}^{\infty} e^{-2at} \, \mathrm{d}t = \frac{1}{2a} < \infty \tag{14}$$

Thus, $x(t) = e^{-at}u(t)$ is an energy signal.

1.4.3.3 Ramp Function

Consider a ramp function defined as:

$$r(t) = \begin{cases} At & t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Classify this signal as an energy, power, or neither.

Solution: Let us compute the energy of this signal as:

$$E = \int_{-\infty}^{\infty} r(t)^2 \, \mathrm{d}t = \int_{-\infty}^{0} [0]^2 \, \mathrm{d}t = \int_{0}^{\infty} A^2 t^2 \, \mathrm{d}t = A^2 \frac{T^3}{3} \bigg|_0^{\infty} = \infty \tag{16}$$

Since the energy of this signal is infinite, it cannot be an energy signal. Let us compute the power of this signal as:

$$P = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} [r(t)]^2 \, \mathrm{d}t = \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T/2} A^2 t^2 \, \mathrm{d}t = A^2 \lim_{T \to \infty} \frac{1}{T} \frac{T^3}{3} \bigg|_0^{\infty} = \infty \tag{17}$$

The power of this signal is infinite. Hence, this is neither a power nor an energy signal.

1.5 What are Systems?

1.5.1 Introduction

The term *system* is derived from the Greek word *systema*, which means an organised relationship among functioning units or components. It is often used to describe any orderly arrangement of ideas or constructs.

According to the Webster's dictionary:

> "A system is an aggregation or assemblage of objects united by some form of regular interaction or interdependence; a group of diverse units so combined by nature or art as to form an integral; whole and to function, operate, or move in unison and often in obedience to some form of control..."

According to the International Council on Systems Engineering (INCOSE):

> "A system is an arrangement of parts or elements that together exhibit behaviour or meaning that the individual constituents do not." The elements or parts, can include people, hardware, software, facilities, policies, and documents; that is, all things required to produce system-level results.

It is difficult to give a single and precise definition of the term system, which will suit to different perspectives of different people. In practice, what is meant by "the system" depends on the objectives of a particular study.

From the control engineering perspective, the system is any interconnection of components to achieve desired objectives. It is characterised by its Inputs, Outputs, and the rules of operations or laws. For example:

a. The laws of operation in electrical systems are Ohm's law, which gives the voltage-current relationships for resistors, capacitors and inductors, and Kirchhoff's laws, which govern the laws of interconnection of various electrical components.

b. Similarly, in mechanical systems, the laws of operation are Newton's laws. These laws can be used to derive mathematical models of the system.

1.5.2 The System as an Operator

A system is defined mathematically as a transformation which maps an input signal $x(t)$ to an output signal $y(t)$. For a continuous time system, the input-output mapping is expressed as:

$$y(t) = \mathcal{S}[x(t)], \text{ where } \mathcal{S} \text{ is the system operator} \tag{18}$$

A Control system may be defined as an interconnection of components which are configured to provide a desired response.

1.5.3 Classification of Systems

The basis of classifying systems are many. They can be classified according to the following:

a. The Time Frame: *(discrete, continuous or hybrid)*;

b. System Complexity: *(physical, conceptual and esoteric)*;

c. Uncertainties: *(deterministic and stochastic)*;

d. Nature and type of components: *(static or dynamic, linear or nonlinear, time-invariant or time variant, lumped or distributed etc)*;

1. Linear and nonlinear systems
2. Time-invariant and time-variant systems
3. Static (memory-less) and dynamic (with memory) systems
4. Causal and Non-causal systems
5. Lumped and distributed parameter systems
6. Determinisitc and stochastic systems
7. Continuous and discrete systems

1.5.4 Linear and Nonlinear Systems

A system is said to be linear provided it satisfies the principle of superposition which is the combination of the additive and homogeneity properties. Otherwise, it is nonlinear.

1.5.4.1 Principle of Additivity

1.5.4.2 Principle of Homogeneity or Scaling

1.5.4.3 Principle of Superposition

1.5.5 Time-Invariant and Time-Variant Systems

1.5.6 Static and Dynamic Systems

1.5.7 Causal and Non-Causal Systems

1.6 What is a Control System?

1.6.1 Introduction

1.6.2 Common Terms of Control Systems

1.6.3 Classification of Control Systems

Chapter 2

Mathematical Modeling of Dynamic Systems

Learning Outcomes

After completing this module, you should be able to:

Chapter 3

The Block Diagram Representation & Characteristics of Feedback Systems

Learning Outcomes

After completing this module, you should be able to:

Chapter 4

Time Domain Analysis of Linear Systems: Time Domain Specifications

Learning Outcomes

After completing this module, you should be able to:

1. Understand the concepts of Transient and Steady-State Responses.
2. Know the Typical input signals used for time domain analysis.
3. Compute various time domain specifications such as rise time, peak time, peak overshoot etc. for an underdamped second order control system.

4.1 Transient and Steady-State Response

It is to be of note that: Systems with energy storage elements (i.e. dynamic systems) can not respond instantaneously and will exhibit transient responses whenever they are subjected to inputs or disturbances.

The time response $c(t)$ of a control system is usually divided into two parts: the transient response and the steady-state response.

Thus, the total response is given by:

$$c(t) = c_{tr}(t) + c_{ss}(t)$$

where $c_{tr}(t)$ is the transient response and $c_{ss}(t)$ is the steady-state response.

Transient response is defined as the part of the response that goes to zero as time "becomes large" or rather goes to infinity. Therefore, $c_{tr}(t)$ has the property of

$$\lim_{t \to \infty} c_{tr}(t) = 0$$

The definition of the steady state, however, has not been entirely standardized. In circuit analysis/theory, it is sometimes useful to define a steady-state variable as being a constant with respect to time.

In control systems, however, the steady-state response is simply the fixed response when time reaches infinity. When a response reaches steady state, it can still vary with time.

For example, a sine wave is considered as a steady-state response because its behavior is fixed for any time interval, as when time approaches infinity. Similarly, if a response is described by $c(t) = t$, it may be defined as a steady-state response.

4.2 Test Signals for Time Domain Analysis

4.2.1 Why is there a need for Test Signals?

The input excitation to many practical control systems are not known ahead of time; unlike many electrical circuits and communication systems.

In many cases, the actual inputs of a control system may vary in random fashions with respect to time.

- For instance, in a radar tracking system, the position and speed of the target to be tracked may vary in an unpredictable manner, so that they cannot be expressed deterministically by a mathematical expression.

This is a major problem for designers, since it is difficult to design the control system so that it will perform satisfactorily to any possible input signal.

For the purposes of analysis and design, it is necessary to assume some basic types of input functions so that the performance of a system can be evaluated with respect to these test signals.

By selecting these basic test signals properly, not only the mathematical treatment of the problem is systemized, but the responses due to these inputs allow the prediction of the system's performance to other more complex inputs.

In a design problem, performance criteria may be specified with respect to these test signals so that a system may be designed to meet these criteria.

The general form of the signals used for time domain analysis can be expressed as:

$$r(t) = At^n; \quad R(s) = \frac{n!A}{s^{n+1}}$$

When $n = 0$, this corresponds to a step signal, when $n = 1$, this represents a ramp signal, and when $n = 2$, this corresponds a parabolic signal.

It is to be of note, however, that from the step function to the parabolic function, the signals become progressively faster with respect to time.

1. The step function is very useful as a test signal, since its initial instantaneous jump in amplitude reveals a great deal about a system's quickness to respond.
2. The ramp function has the ability to test how the system would respond to a signal that changes linearly with time.
3. The parabolic function is one degree faster than the ramp function. In practice, we seldom find it necessary to use a test signal faster than a parabolic function.

4.2.2 Step Input

Frequently the performance characteristics of a control system are specified in terms of transient response to a unit-step input. This is because:

1. They are easy to generate.
2. Its initial instantaneous jump in amplitude reveals a great deal about a system's quickness to respond
3. Also, since the step function has, in principle, a wide band of frequencies in its spectrum, as a result of the jump discontinuity, as a test signal it is equivalent to the application of numerous sinusoidal signals with a wide range of frequencies.

4.2.3 Ramp Input

The ramp function is a signal that changes constantly with respect to time.

Mathematically, a ramp function is represented by

$$r(t) = At$$

where A is a real constant.

The ramp function has the ability to test how the system would respond to a signal that changes linearly with time.

4.2.4 Parabolic Input

The parabolic function represents a signal that is one order faster than the ramp function.

Mathematically, a parabolic function is represented by

$$r(t) = \frac{At^2}{2}$$

where A is a real constant. Note that the factor of $\frac{1}{2}$ is added for mathematical convenience; because the Laplace transform of $r(t)$ becomes $\frac{A}{s^3}$.

Note: From the step function to the parabolic function, the signals become progressively faster with respect to time.

4.3 Time Domain Analysis of First Order System

Consider a first order system without a zero shown in Figure 1

This is a generic first order system, and may represent an R-L circuit, R-C circuit, a simple thermal system and so on. The relationship between the input and output is given by

$$\frac{C(s)}{R(s)} = \frac{1}{Ts+1} \tag{1}$$

Eqn(Equation 1) can alternately be expressed as:

$$\frac{C(s)}{R(s)} = \frac{1/T}{s+1/T} = \frac{a}{s+a}, \qquad \text{where} \quad a = \frac{1}{T} \tag{2}$$

The block diagram, corresponding to the system in Eqn(Equation 2), is shown in Figure 2a and the pole location is shown in the s-plane in Figure 2b.

The step response of this system, i.e. when $R(s) = \frac{1}{s}$, is given by

$$C(s) = G(s)R(s) = \frac{a}{s(s+a)} = \frac{1}{s} - \frac{1}{s+a}$$

Which in the time domain gives us:

$$c(t) = c_f(t) + c_n(t) = 1 - e^{-at}$$

The step response of the first order system is shown in the Figure.

4.3.1 Time Constant T (or τ)

The time constant is defined as the time required for the step response to reach 63% of its final value. It is denoted by T or τ.

4.3.2 Computation of Time Constant for a First Order System

We Know that the step response of a first order system is given by

$$c(t) = 1 - e^{-at}$$

Let T be the time when the response equals to 0.63. i.e. $c(T) = 0.63$. This implies

$$1 - e^{-aT} = 0.63 \implies aT = \ln\left(\frac{1}{0.37}\right) = \ln(2.7027) \approx 1 \implies T \approx \frac{1}{a}$$

Thus the time constant T equals to

$$T = \frac{1}{a}$$

4.3.3 Rise Time T_r

The rise time is defined as the time required for the response to go from 10% (0.1) to 90% (0.9) of its final value. It is denoted by T_r.

4.3.4 Computation of Rise Time for a First Order System

We Know that the step response of a first order system is given by

$$c(t) = 1 - e^{-at}$$

Let t_0 be the time when the response equals to 0.1. i.e. $c(t_0) = 0.1$. This implies

$$1 - e^{-at_0} = 0.1 \implies at_0 = \ln\left(\frac{10}{9}\right) \approx 0.11 \implies t_0 \approx \frac{0.11}{a}$$

Let t_1 be the time when the response equals to 0.9. i.e. $c(t_1) = 0.9$. This implies

$$1 - e^{-at_1} = 0.9 \implies at_1 = \ln(10) \approx 2.31 \implies t_1 \approx \frac{2.31}{a}$$

Thus the **rise time** T_r equals to

$$T_r = t_1 - t_0 \approx \frac{2.31}{a} - \frac{0.11}{a} = \frac{2.2}{a}$$

4.3.5 Settling Time T_s

The settling time is defined as the time required for the response to reach and stay within, in our case, 2% of its final value. It is denoted by T_s.

4.3.6 Computation of Settling Time for a First Order System

We Know that the step response of a first order system is given by

$$c(t) = 1 - e^{-at}$$

Let T_s be the time when the response equals to 0.98. i.e. $c(T_s) = 0.98$. This implies

$$1 - e^{-aT_s} = 0.98 \implies aT_s = \ln(50) \approx 3.91 \implies T_s \approx \frac{3.91}{a}$$

Thus the **settling time** T_s equals to

$$T_s = \frac{3.91}{a} \approx \frac{4}{a}$$

4.4 DC Gain of a System

This is the gain of the system when the input frequency is zero. Therefore it is also know as zero frequency gain of the system. It is the ratio of the magnitude of the steady state output to that of the constant input, provided the output is finite. This can be computed as follows:

Consider the system with the transfer function

$$G(s) = \frac{C(s)}{R(s)}$$

If we apply a unity step input $R(s) = \frac{1}{s}$, the steady state value of the output becomes

$$\lim_{t\to\infty} c(t) = \lim_{s\to 0} sG(s)R(s) = \lim_{s\to 0} sG(s)\frac{1}{s} = \lim_{s\to 0} G(s) \tag{3}$$

Since the input is a constant of magnitude unity (1), the steady state output is also the gain in the steady state; that is $G(0)$. Thus

$$\text{DC Gain} = \lim_{s\to 0} G(s) = G(0) \tag{4}$$

4.5 Time Domain Analysis of Second Order System

Consider a second order system shown in the Figure

The closed loop transfer function of this system is given by

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

(The notes used ξ, incorrectly, but was always read as ζ, which is the correct symbol used according to wikipedia)

The parameter ω_n is the natural frequency of the second order system, and the parameter ζ is called the damping ratio.

The closed loop poles $\lambda_{1,2}$ are given by

$$\lambda_{1,2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

The nature of the response of the system depends on the value of the damping ratio ζ. The following discusses the step responses of this system for different cases.

4.5.1 Case 1: $\zeta = 0$

- The closed loop poles are located at $\pm j\omega_n$.
- The step response under this case will be sinusoidal with frequency ω_n and is expressed as:

$$c(t) = 1 - \cos(\omega_n t)$$

- This type of response is called an **undamped** response.

4.5.2 Case 2: $0 < \zeta < 1$

- The closed loop poles are located at $-\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$.
- The step response is a damped sinusoid with an exponential envelope whose time constant is equal to the reciprocal of the real part of the poles. It is expressed as:

$$c(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t}\cos(\omega_d t + \phi); \quad \text{where}$$
$$\omega_d = \omega_n\sqrt{1-\zeta^2}; \quad \phi = \tan^{-1}\left(\frac{\zeta}{\sqrt{1-\zeta^2}}\right)$$

- This type of response is called an **under damped** response.

4.5.3 Case 3: $\zeta = 1$

- We have two closed loop poles which are real (i.e. $\lambda \in \mathbb{R}$) and are located at $-\zeta\omega_n$.
- The step response under this case will be expressed as:

$$\begin{aligned} c(t) &= 1 - \zeta\omega_n t e^{-\zeta\omega_n t} - e^{-\zeta\omega_n t} \\ &= 1 - e^{-\zeta\omega_n t}(1 + \zeta\omega_n t) \end{aligned}$$

- This type of response is called a **critically damped** response.

4.5.4 Case 4: $\zeta > 1$

- The two closed loop poles are real (i.e. $\lambda \in \mathbb{R}$) and are located at $-\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$
- The step response under this case will be expressed as:

$$c(t) = 1 - \mathrm{e}^{-(\zeta - \sqrt{\zeta^2-1})\omega_n t}$$

- This type of response is called an **over damped** response.

4.6 Transient (Time) Response Specifications: Some Notes

Control systems are generally designed with damping less than one, i.e. oscillatory step response.

Higher-order control systems usually have a pair of complex conjugate poles with damping less than one which dominate over the other poles **(dominant pole pairs)**.

Thus the time response of second and higher-order control systems to a step input, is in general, of a **damped oscillatory nature**.

It is observed that the step response has a number of **overshoots** and **undershoots** with respect to the final steady state value.

This type of response is expressed mathematically as:

$$c(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\cos\left(\omega_d t + \phi\right), \qquad \text{where} \quad \omega_d = \omega_n\sqrt{1-\zeta^2}, \quad \phi = \tan^{-1}\left(\frac{\zeta}{\sqrt{1-\zeta^2}}\right)$$

The pole plot of the **underdamped** second order system is shown in the figure

This type of response is characterised by the following performance indicies:

1. Rise Time T_r
2. Peak Time T_p
3. Peak Overshoot M_p

4. Settling Time T_s
5. Steady State Error e_{ss}

These indicies are qualitatively related to

a. How fast is the system? i.e. how fast it moves to follow the input?
b. How oscillatory the system is? (indicative of damping)
c. How long does it take to practically reach the final value?

Note: The various indicies are not independent of each other.

4.7 Commonly Used Transient Response Specification

4.7.1 Rise Time T_r

- It is the time required for the response to rise from 10% to 90%, 5% to 95%, or 0% to 100% of its final value.
- For underdamped, second-order systems, the 0% to 100% rise time is normally used. For overdamped systems, the 10% to 90% rise time is commonly used.
- Analytically, it is expressed as

$$T_r = \frac{1}{\omega_d} \tan^{-1}\left(\frac{\omega_d}{-\sigma_d}\right) = \frac{\pi - \theta}{\omega_d}$$

4.7.2 Peak Time T_p

- It is the time required for the response to reach the first peak of the overshoot.
- Analytically, it is expressed as

$$T_p = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}} = \frac{\pi}{\omega_d}$$

4.7.3 Maximum (percent) Overshoot $\%OS$, or M_p

- It is the maximum peak value of the response. It is defined by

$$M_p = \frac{c(t_p) - c(\infty)}{c(\infty)} \times 100\%$$

- Analytically, it is expressed as

$$\%OS = e^{-\left(\frac{\pi\zeta}{\sqrt{1-\zeta^2}}\right)}$$

4.7.4 Settling Time T_s

- It is teh time required for the response to reach and stay within either 2% or 5% of its final (steady-state) value. Commonly it is expressed as:

$$T_s = 4T = \frac{4}{\sigma_d} = \frac{4}{\zeta\omega_n} \quad \text{(2\% criterion)}$$

$$T_s = 3T = \frac{3}{\sigma_d} = \frac{3}{\zeta\omega_n} \quad \text{(5\% criterion)}$$

4.8 Correlation Between Pole Locations with Time Domain Specifications

1. The radial distance from the origin to the pole is the natural frequency ω_n.
2. The damping ratio ζ is equal to $\cos(\theta)$, i.e. $\cos(\theta) = \zeta$.
3. We know that the peak time and settling time are

$$T_p = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}} = \frac{\pi}{\omega_d} \qquad T_s = \frac{4}{\zeta\omega_n} = \frac{4}{\sigma_d}$$

where ω_d is the imaginary part of the pole and is called the damped frequency of oscillation, and σ_d is the magnitude of the real part of the pole and is the exponential damping frequency.

4. Thus T_p is inversely proportional to the imaginary part of the pole.
5. T_s is inversely proportional to the real part of the pole.

4.9 Examples: Time Domain Specifications

4.10 Effects of Changing Exponential Damping Frequency ω_d & Damped Frequency σ_d on Time Response

- The poles are located at $-\sigma_d \pm j\omega_d$. The peak time T_p and settling time T_s are

$$T_p = \frac{\pi}{\omega_d}, \quad \text{and} \quad T_s = \frac{4}{\sigma_d}$$

1. Vary ω_d and fix σ_d. The frequency of oscillation and the peak time change; but the settling time remains unchanged.
2. Vary σ_d and fix ω_d. The frequency of oscillation and the peak time remain unchanged; but the settling time changes.
3. Vary the natural frequency ω_n and fix ζ. The percentage overshoot remains unchanged; but the settling time changes.
4. Thus T_p is inversely proportional to the imaginary part of the pole.
5. T_s is inversely proportional to the real part of the pole.

Chapter 5

Stability Analysis of Linear Systems: Routh-Hurwitz Stability Criteria

Learning Outcomes

After completing this module, you should be able to:

1. Determine the absolute stability of a linear system.
2. Find the stability region of a linear system.

5.1 Concept of Stability

A system is stable if small changes in system inputs, initial conditions, or system parameters do not result in large changes in system behaviour. Intuitively, a system is stable if it remains at rest unless excited by an external source and returns to rest if all excitations are removed.

Absolute Stability

Suppose the ball is initially inside the bowl in position 1. If it is perturbed from its inital position by a small force, it will cause the ball to move. When the force is removed, the ball will oscillate and eventually return to its initial position. This is an example of **absolutely stable dynamics**.

Instability

Consider the situation where the bowl is turned upside down and the ball is placed on the top of the bowl. When the ball is slightly perturbed by the application of a force, it begins to move on its own without any additional force applied. It will never return to its original position. This is an example of **unstable dynamics**.

Neutral Stability

If the ball is placed on a flat surface, then after the application of a small force, the ball will move. but when the force is withdrawn, the ball stops and remains in its new position. This is an example of **neutral stability**.

Suppose a system has an equilibrium point at $x = x_e$. In stability studies we generally address the following questions:

1. If the system with zero input is perturbed from its equilibrium point x_e at $t = t_0$, will the state $x(t)$
   a. Return to x_e?
   b. Remain close to x_e? or
   c. Diverge from x_e?
2. If the system is relaxed, will a **bounded input** produce **a bounded output**?

Consider the ball which is free to roll on the surface.

The ball could be made to rest at points **A, E, F, G** and anywhere between the points **B and D, such as C.** An perturbation away from A or F will cause the ball to diverge from these points. Thus A and F are **unstable equilibrium points**. After small perturbations away from E or G, the ball will eventually return to these points. Thus E and G are **stable equilibrium**

points. If the ball is perturbed slightly away from C, it will remain at the new position. Points like C are sometimes said to be **neutrally stable**.

The following figure shows different types of possible stability surfaces for globally stable, stable, unstable and locally stable systems.

5.1.1 Definitions of Stability, Instability and Marginal Stability

The total response of a system $c(t)$ of a dynamic system is expressed as:

$$c(t) = c_{\text{forced}}(t) + c_{\text{natural}}(t) \tag{1}$$

Based on the **natural response**, we define stability, instability and marginal stability as follows:

Stable A linear time invariant system is stable provided its natural response converges to zero as time approaches infinity.

Unstable A linear time invariant system is unstable if its natural response grows without bound as time approaches infinity.

Marginally Stable A linear time invariant system is marginally stable if its natural response neither decays nor grows unbounded but remains constant or oscillatory as time approaches infinity.

Based on total response or zero-input response we define BIBO stability.

BIBO Stability A system is said to be BIBO (bounded-input, bounded-output) stable if for every bounded input, the output is bounded.

Summary:

A linear time invariant system is stable if the following two notions of system stability are satisfied.

a. When the system is excited by a bounded input, the output is bounded. **(BIBO Stability)**

b. With zero input and arbitrary initial conditions, the output tends towards zero - the equilibrium state of the system **(Asymptotic Stability)**.

These two notions of stability are equivalent for linear time invariant systems.

5.2 Stability of LTI system from Transfer Function

Consider a system with closed loop transfer function

$$T(s) = \frac{C(s)}{R(s)} = \frac{15}{(s+3)(s+5)}$$

The output of this system for unit step input $R(s) = 1/s$ is

$$C(s) = \frac{15}{s(s+3)(s+5)} = \frac{1}{s} - \frac{2.5}{s+3} + \frac{1.5}{s+5}$$

This gives

$$c(t) = 1 - 2.5e^{-3t} + 1.5e^{-5t}$$

Since the closed-loop poles are real and located in the left-half of the s-plane, the output response contains exponential terms with negative indicies, i.e. e^{-3t} and e^{-5t}. As time $t \to \infty$, both exponential terms will approach zero and the output will reach its steady state value of 1, i.e. $c(\infty) = 1$. Such type of systems where the poles are in the left half of the s-plane are **absolutely stable systems**.

Consider a system with closed loop transfer function

$$T(s) = \frac{C(s)}{R(s)} = \frac{10}{(s-2)(s+3)}$$

The output of this system for unit step input $R(s) = 1/s$ is

$$C(s) = \frac{10}{s(s-2)(s+3)} = \frac{-1.666}{s} + \frac{1}{s-2} + \frac{0.666}{s+3}$$

This gives

$$c(t) = -1.666 + e^{2t} + 0.666e^{-3t}$$

Due to a pole located in the right half of the s-plane, there is one exponential term with positive index (e^{2t}) which go on, increasing in amplitude as time $t \to \infty$. Systems where any of the poles are in the right half of s-plane are **unstable systems**.

Consider a system with closed loop transfer function

$$T(s) = \frac{C(s)}{R(s)} = \frac{25}{s^2 + 25}$$

The closed loop poles are purely imaginary and are located on the $j\omega$-axis. The output of this system for unit step input $R(s) = 1/s$ is

$$C(s) = \frac{25}{s(s^2+25)} = \frac{1}{s} - \frac{1}{2(s-j5)} - \frac{1}{2(s+j5)}$$

This gives

$$c(t) = 1 - \cos(5t)$$

Due to the presence of purely imaginary poles, the response is oscillatory. Systems where any of the poles are on the $j\omega$ axis are **marginally stable systems**.

The closed loop transfer function of an n-th order single input, single output linear time invariant system is expressed as:

$$T(s) = \frac{C(s)}{R(s)} = \frac{b_0 s^m + b_1 s^{m-1} + ... + b_{m-1}s + b_m}{s^n + a_1 s^{n-1} + a_2 s^{n-2} + ... + a_{n-1}s + a_n} = \frac{P(s)}{Q(s)}; \quad (m \leq n) \tag{2}$$

The closed loop *characteristic equation* of the system is given by

$$Q(s) = s^n + a_1 s^{n-1} + a_2 s^{n-2} + ... + a_{n-1}s + a_n = 0 \tag{3}$$

The roots of $Q(s) = 0$ gives us the corresponding closed loop poles. Let us find the solution to the differential equation corresponding to the characteristic equation Equation 3 considering the following cases.

5.2.1 Case 1

Case 1: All the roots of the characteristic equation (closed loop poles), $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct. They can either be real or complex. Then the output is

$$c(t) = \sum_{i=1}^{n} A_i e^{\lambda_i t} \tag{4}$$

where the constant A_i depends on the inital conditions and locations of zeros.

If all the roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ are real, their contribution to the output is of the form

$$c(t) = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t} + ... + A_n e^{\lambda_n t}$$

The contribution of a complex root pair $\lambda = \sigma_i \pm j\omega_i$ to the output is of the form

$$Ae^{\sigma_i t}\sin(\omega_i t + \phi_i)$$

5.2.2 Case 2

Case 2: If some of the roots of the characteristic equation are repeated.

Without loss of generality, assume that $\lambda_1 = \lambda_2$ and they are real. Then the contribution of this to the output is of the form

$$c(t) = A_1 e^{\lambda_1 t} + A_2 t e^{\lambda_1 t}$$

Similarly, if there is a repeated root of multiplicity k at λ_1, i.e. $\lambda_1 = \lambda_2 = \ldots = \lambda_k$, then their contribution to the output is of the form

$$[A_1 + A_2 t + A_3 t^2 + \ldots + A_k t^{k-1}]e^{\lambda_1 t}$$

If there are complex conjugate root pairs $\lambda = \sigma \pm j\omega$ **of multiplicity k**, then their contribution to the output is of the form

$$[A_1 \sin(\omega t + \phi_1) + A_2 t \sin(\omega t + \phi_2) + \ldots + A_k t^{k-1}\sin(\omega t + \phi_k)]e^{\sigma t}$$

Key Point:

When a system has repeated poles of k-th order in its transfer function, the output response can include terms like $t, t^2, \ldots, t^{k-1}$ multiplied by an exponential term. Unless the effects of these polynomial terms are counteracted by decaying exponential terms, stability can not be ensured. Note that **an exponential decay is stronger than a polynomial growth of any order.**

5.2.3 Case 3

Case 3: If some of the roots are purely imaginary

If the roots are a purely complex conjugate pair located at $\pm j\omega$, their contribution to the output is of the form

$$A\sin(\omega t + \phi)$$

Such purely imaginary pole pairs produce an oscillatory (sinusoidal) natural response.

If there are purely **Complex conjugate root pairs** $\pm j\omega$ **of multiplicity** k, then their contribution to the output is

$$A_1 \sin(\omega t + \phi_1) + A_2 t \sin(\omega t + \phi_2) + \ldots + A_k t^{k-1}\sin(\omega t + \phi_k)$$

This gives rise to an unbounded response, and the system is unstable.

5.2.4 Summary

1. If all the roots of the characteristic equation lie in the left half s-plane, the system is stable. In this case, the impulse response is bounded and eventually converges to zero and therefore $\int_0^\infty |g(\tau)|\,\mathrm{d}\tau$ is finite and the system is BIBO stable.
2. If any root of the characteristic equation lies in the right half of the s-plane or if there is a repeated root on the $j\omega$ axis, the system is *unstable*. In this case the impulse response is unbounded and $\int_0^\infty |g(\tau)|\,\mathrm{d}\tau$ is infinite leading to instability.
3. If the characteristic equation has one or more non-repeated roots on the $j\omega$ axis, but no right half plane roots, the system is *marginally stable* or limitedly stable. In this case, the impulse response is finite, but $\int_0^\infty |g(\tau)|\,\mathrm{d}\tau$ is infinite.

5.3 Necessary Conditions of Stability

Consider the closed loop characteristic equation of a linear time invariant system which is of the form

$$Q(s) = a_0 s^n + a_1 s^{n-1} + ... + a_{n-1} s + a_n = 0$$

In order to ensure that there are no roots of the characteristic equation with positive real parts it is **necessary but not sufficient** that

1. All the coefficients of the polynomial have the same sign.
2. None of the coefficients vanish.

Examples:

$Q_1(s) = s^4 - 3s^3 + 9s^2 + 63s + 50 = 0$; (Coefficients not of same sign)(Unstable)

$Q_2(s) = s^4 + 3s^3 + 9s^2 + 50 = 0$; (Vanishing Coefficients)(Unstable)

$Q_3(s) = s^3 + 2s^2 + 2s + 40 = 0$; (Inconclusive)(Actually Unstable; but satisfies necessary condition)

Proposition: If all the coefficients of the characteristic equation of a system have the same sign, this ensures that real roots of the system are negative. However, this does not ensure negativeness of real parts of the complex roots for third and higher order systems. Thus, it can not be sufficient conditions of stability for third and higher order systems.

Example: Consider a third order system with characteristic equation

$$s^3 + 2s^2 + 2s + 40 = 0 \tag{5}$$

Eqn(Equation 5) can be expressed as

$$(s + 4)(s - 1 + j3)(s - 1 - j3) = 0$$

Notice that the real part of the complex roots is positive, indicating system instability; although all the coefficients of the characteristic equation have the same sign (positive). Thus, when the order of the characteristic equation is higher than two, it becomes insufficient to infer the stability of the system solely by examining the signs of all the coefficients.

The stability analysis of higher order (> 2) systems should therefore be carried out by examining its characteristic as follows:

1. If the signs of all the coefficients of the characteristic equation are not the same and/or if some of the coefficients are zero, then it is indicative of potential instability in the system.
2. If all the coefficients of the characteristic equation have the same sign, the possibility of instability can not be excluded; because this is a necessary condition of stability. We, therefore, do further analysis to determine sufficient conditions for stability.

5.4 Routh's Stability Criterion

The closed loop transfer function of most linear feedback systems are of the form

$$\frac{C(s)}{R(s)} = \frac{b_0 s^m + b_1 s^{m-1} + ... + b_{m-1} s + b_m}{a_0 s^n + a_1 s^{n-1} + ... + a_{n-1} s + a_n}; \quad (m \leq n) = \frac{P(s)}{Q(s)}$$

Note that we are only interested to know whether there are some closed loop poles (that is the roots of $Q(s) = 0$), lie in the right-half of the s-plane.

5.4.1 Features of Routh-Hurwitz Criterion:

Routh-Hurwitz criterion provides a straightforward and computationally efficient approach to stability analysis, particularly for higher-order systems. This method offers a numerical procedure for determining the stability of a system without explicitly solving for the roots of the characteristic polynomial. It provides a systematic approach to asses the location of the roots, particularly the whether any roots lie in the right-half plane (RHP) or on the imaginary axis based on the properties of a tabular representation called the Routh's array.

This method essentially requires two steps:

1. Construct a data table or array called a *Routh Table* or *Routh Array* using the coefficients of the characteristic polynomial.
2. Analyze the first column of the Routh Array, to determine how many roots of the characteristic equation (closed loop poles) are in the left-half plane, right-half plane, or directly on the $j\omega$ axis.

Consider that the closed loop characteristic equation of a linear time invariant system is of the form

$$a_0 s^n + a_1 s^{n-1} + ... + a_{n-1} s + a_n = 0 \tag{6}$$

where all the coefficients are real numbers.

It is assumed that $a_0 \neq 0$, meaning that any potential zero root has been removed from the characteristic equation.

A more mathematical representation of Eqn.(Equation 6) is

$$\sum_{i=0}^{n} a_i s^{n-i} = 0, \quad \text{where} \quad a_i \in \mathbb{R} \quad \text{and} \quad \forall i \in \mathbb{N}_0$$

In order for there to be no roots of the characteristic equation that have positive real parts, it is **necessary but by no means sufficient** that

1. All of the coefficients of the polynomial have the same sign.
2. None of the coefficients vanish.

5.4.2 Formation of Routh Tables/Arrays

To apply the Routh-Hurwitz criterion, we need to form the Routh Table or Routh Array. It is hence constructed as follows:

1. If all the coefficients of the characteristic polynomial are positive, we begin by arranging the coefficients into the first two rows of the Routh Array.
2. The first row of the Routh Array consists of the **first, third, fifth, ...** coefficients (even values of n in a_n), where the second row consists of the **second, fourth, sixth, ...** coefficients (odd values of n in a_n), as shown below.

Table 5.1: Completed Routh Array

s^n	a_0	a_2	a_4	a_6	$\cdots$
s^{n-1}	a_1	a_3	a_5	a_7	$\cdots$
s^{n-2}	b_1	b_2	b_3	b_4	$\cdots$
s^{n-3}	c_1	c_2	c_3	c_4	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
s^2	e_1	e_2			
s^1	f_1				
s^0	g_1				

The evaluation of the b_i coefficients are computed as follows:

$$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1}, \; b_2 = \frac{a_1 a_4 - a_0 a_5}{a_1}, \; b_3 = \frac{a_1 a_6 - a_0 a_7}{a_1}, ...$$

Following a similar procedure, the evaluation of the c_i coefficients are computed as follows:

$$c_1 = \frac{b_1 a_3 - a_1 b_2}{b_1},\ c_2 = \frac{b_1 a_5 - a_1 b_3}{b_1},\ c_3 = \frac{b_1 a_7 - a_1 b_4}{b_1}, \ldots$$

These can be more formally expressed via set notation as:

Definition 1 (Set of b_i coefficients)**.** Let B denote the set of the coefficients in the third row of the Routh Array, b_i. We can define B as:

$$B = \left\{ b_i \in \mathbb{R} \,\middle|\, b_i = \frac{a_1 a_{2i} - a_0 a_{2i+1}}{a_1}, \quad i \in \mathbb{N}, \quad 1 \leq i \leq m \right\}$$

Definition 2 (Set of c_i coefficients)**.** Let C denote the set of the coefficients in the fourth row of the Routh Array, c_i. We can define C as:

$$C = \left\{ c_i \in \mathbb{R} \,\middle|\, c_i = \frac{b_1 a_{2i+1} - a_1 b_{i+1}}{b_1}, \quad i \in \mathbb{N}, \quad 1 \leq i \leq n \right\}$$

5.5 Routh-Hurwitz Stability Criterion for Linear Feedback Control Systems

It is assumed that the polynomial used to construct the Routh's table is the closed loop characteristic equation of a control system. Therefore, the roots of this polynomial correspond to the poles of the closed loop system.

Stability Criterion

The necessary and sufficient condition for a linear time invariant system to be stable is that all the terms in the first column of a Routh Table/Array must have the same sign.

i. Thus, if there are no sign changes in the first column of the Routh Table, all roots of the characteristic equation lie in the left hand side of the s-plane (left half plane or LHP), indicating system stability.

ii. If there are sign changes in the first column of the Routh Array, then some of the roots of the characteristic polynomial lie in the right half of the s-plane (right half plane or RHP), indicating system instability.

iii. The number of sign changes in the first column of the Routh Table equals to the number of roots of the characteristic polynomial that lie in the right half of the s-plane (RHP).

5.5.1 Different Configurations of Routh Tables

Case 1 No element in the first column is zero. This is the normal case and we have to follow the procedure detailed above.
Case 2 The first element in a row of the Routh's Table is zero; but all other elements in the row are nonzero or there are no remaining terms.
Case 3 All the elements of a row are zero.
Case 4 All the elements of a row are zero, but with repeated roots on the $j\omega$-axis.

5.5.2 Examples: Case 1

5.5.2.1 Example 1

Consider a system with closed loop characteristic equation:

$$s^3 + 4s^2 + 6s + 4 = 0$$

Construct the Routh array and determine the stability of the system.

Table 5.2: Completed Routh Table

s^3	1	6
s^2	4	4
s^1	$\dfrac{(4 \times 6) - (1 \times 4)}{4} = 5$	0
s^0	$\dfrac{(5 \times 4) - (4 \times 0)}{5} = 4$	

Since there are no sign changes in the first column of the Routh Array, all the closed loop poles of the system lie in the left half of the s-plane and the system is stable.

Actual Roots: $-2, -1$ and $-1 \pm j1$

5.5.2.2 Example 2

5.5.2.3 Example 3

5.5.2.4 Example 4

5.5.3 Examples: Case 2

If any term in the first column is zero; but the remaining terms are non-zero, we cannot proceed with Routh Table formation. Instead, we replace the zero term with a very small, positive, non-zero number ε and evaluate the rest of the array elements.

5.5.3.1 Example 1

5.5.3.2 Example 2

5.5.3.3 Example 3

5.5.4 Examples: Case 3

If all the coefficients in any derived row are zero, it indicates that one or more of the following conditions may exist:

There are roots of equal magnitude lying radially opposite each other in the s-plane. This implies:

1. Pairs of real roots with opposite signs.
2. Pairs of imaginary roots.
3. Pairs of complex-conjugate roots forming symmetry about the origin of the s-plane.

Solution: Procedure to evaluate the rest of the array

1. Form an Auxiliary Polynomial $A(s)$ using the coefficients of the row just above the row of zeros.
2. Take the derivative of the Auxiliary Polynomial with respect to s.
3. Replace the row of zeros with the coefficients of the resultant equation $\frac{\mathrm{d}A(s)}{\mathrm{d}s}$.
4. Carry on the Routh test in the usual manner with the newly formed tabulation.

5.5.4.1 Example 1

5.5.4.2 Example 2

5.5.5 Limitations of Routh-Hurwitz Criterion

1. It is applicable if the characteristic equation is algebraic and all the coefficients are real.
   - If any one of the coefficients of the characteristic equation is a complex number (i.e. $a_i \in \mathbb{C}$), or
   - If the equation contains exponential functions of s, such as systems with time delays, this criterion cannot be applied.
2. It offers information only on the absolute stability of the system and does not provide any information about the relative stability of the system. i.e. it does not say how closely the roots of the characteristic equation are located to the imaginary axis of the s-plane.

5.6 Relative Stability Analysis

Although, normally we do not analyse the relative stability of a system using the Routh-Hurwitz criterion, it is possible to do, but with an extra computational burden.

The procedure is described as follows:

1. Substitute s with $s = \hat{s} - \sigma$, where σ is a constant, into the characteristic equation of the system.
2. Write the polynomial in terms of $\hat{s}$ and apply the Routh-Hurwitz criterion to the new polynomial in $\hat{s}$.
3. The number of sign changes in the first column of the array developed for the new polynomial in $\hat{s}$ is equal to the number of roots that are located to the right of the vertical line $s = -\sigma$.
4. Thus, this test reveals the number of roots that lie to the right of the vertical line $s = -\sigma$.

Chapter 6

Time Domain Analysis of Linear Systems: Static Error Constants & Steady State Error

Learning Outcomes

After completing this module, you should be able to:

1. Compute the static error constants of linear unity feedback systems.
2. Determine the steady state error due to standard test inputs such as step, ramp and parabolic signals.
3. Could compute the steady state error due to disturbances.
4. Could compute the steady state error for non-unity feedback systems.

6.1 Steady State Error (SSE)

What is Steady State Error and what is its significance?

Steady state error refers to the discrepancy between the desired reference input and the actual output of a given control system once the system has settled into a steady condition. It serves as a metric for assessing the system's ability to accurately track the desired reference signal over time. This is especially important when it comes to PID controllers, where the goal is to minimize the steady state error to ensure the system's output closely follows the desired reference input.

6.1.1 Factors Contributing to SSE

The steady state error of a control system can be influenced by several factors, including:

1. **Imperfect Modeling** Errors in the mathematical models used to design the control system can lead to inaccuracies in predicting the system's response, resulting in steady-state errors.
2. **Disturbances and Noise** External disturbances or noise in the system can perturb the system's response, causing deviations from the desired reference signal in the steady state.
3. **Limitations in Control Algorithm** Imperfections or simplifications in the control algorithm, such as linearization or approximation, can introduce errors, especially in the steady state.
4. **Nonlinear Characteristics** Nonlinearities inherent in control system elements, such as friction, dead zones, or saturation in actuators, can result in deviations from the desired response, particularly in the steady state.
5. **Actuator Saturation** Limitations on the maximum output of actuators, such as motors or valves, can lead to saturation effects that prevent the system from fully tracking the reference signal.
6. **Sensor Noise** Measurement noise or inaccuracies in sensors can introduce errors into the feedback loop, affecting the control system's ability to accurately track the reference signal.
7. **Bandwidth Constraints** Limited bandwidth in the system components, such as filters or communication channels, can affect the system's ability to respond quickly to changes in the reference signal, leading to steady-state errors.

6.1.2 Steady State Error in Unity Feedback Systems

Consider a unity feedback system shown below

The closed loop transfer function is given by

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1+G(s)}$$

The error $E(s)$ between the input $R(s)$ (reference signal), and the output, $C(s)$, is given by

$$E(s) = R(s) - C(s) = R(s) - G(s)E(s); \qquad \because C(s) = G(s)E(s)$$

Solving for $E(s)$ gives

$$E(s) = \frac{R(s)}{1+G(s)}$$

The steady state error e_{ss} of a feedback control system is defined as the error when time reaches infinity. Thus

$$e_{ss} = e(\infty) = \lim_{t\to\infty} e(t)$$

By applying the final value theorem, the steady state error is computed from

$$\begin{aligned} e_{ss} = e(\infty) &= \lim_{t\to\infty} e(t) && (1)\\ &= \lim_{s\to 0} sE(s) = \lim_{s\to 0} \frac{sR(s)}{1+G(s)} && (2) \end{aligned}$$

Thus the steady state error is expressed as

$$e(\infty) = \lim_{s\to 0} \frac{sR(s)}{1+G(s)}$$

This shows that the steady state error depends on the reference input $R(s)$ as well as on the form of the transfer function $G(s)$.

6.2 Classification of Control Systems: System TYPE

Control systems may be classified according to their ability to follow step inputs, ramp inputs, parabolic inputs, and so on. This is a reasonable classification scheme because inputs may frequently be considered combinations of such inputs. The magnitudes of the steady-state errors due to these individual inputs are indicative of the “goodness” of the system.

Consider a unity feedback control system shown in the Figure.

The open loop transfer function of the system is described by:

$$G(s) = \frac{K(s+z_1)(s+z_2)\ldots(s+z_m)}{s^N(s+p_1)(s+p_2)\ldots(s+p_n)}$$

This involves a term s^N in the denominator, representing a pole of multiplicity (i.e. order) N at the origin. This also implies the number of pure integrators present in the forward path. The present classification scheme is based on the number of pure integrators present in the forward path.

1. If the number of integrators present in the forward path is zero (i.e. $N = 0$), the system is classified as a **Type-0 System**.
2. Similarly, if the number of integrators present in the forward path is one (i.e. $N = 1$), the system is classified as a **Type-1 System**.

3. In general, if the number of integrators present in the forward path is N, the system is classified as a **Type-N System**.

Note that as the type number is increased, accuracy is improved; however, increasing the type number aggravates the stability problem. A compromise between steady-state accuracy and relative stability is always necessary.

6.3 Static Error Constants

The static error constants defined in the following are figures of merit of control systems from the perspective of steady state error. The **higher the value of these constants**, the **smaller the steady state error is**.

Note that the steady state error exhibited by a system is dependant on the **Nature of the input**.

In the following, we define three static error constants such as **Position Error Constant** K_p, **Velocity Error Constant** K_v and the **Acceleration Error Constant** K_a when the input of the system is step, ramp or parabolic respectively.

6.3.1 Static Position Error Constant K_p

This is calculated when the input signal is a unit step signal.

$$r(t) = 1(t), \implies R(s) = \frac{1}{s}$$

The steady state error due to this input is given by

$$e(\infty) = \lim_{s\to 0} \frac{sR(s)}{1+G(s)} = \lim_{s\to 0} \frac{s}{1+G(s)}\frac{1}{s} \tag{3}$$

$$= \lim_{s\to 0} \frac{1}{1+G(s)} = \frac{1}{1+\lim_{s\to 0} G(s)} \tag{4}$$

The static position error constant K_p, is defined as

$$K_p = \lim_{s\to 0} G(s) = G(0)$$

Thus, the steady state error due to a unit step input is given by

$$e(\infty) = \frac{1}{1+K_p}; \quad K_p = \lim_{s\to 0} G(s)$$

6.3.2 Static Position Error Constant K_p for different system types

For a Type-0 system, the static position error constant is given by

$$K_p = \lim_{s\to 0} G(s) = \lim_{s\to 0} \frac{K(s+z_1)(s+z_2)\dots(s+z_m)}{(s+p_1)(s+p_2)\dots(s+p_n)} = \frac{Kz_1z_2\dots z_m}{p_1p_2\dots p_n} = K_1(\text{say})$$

For a Type-1 or higher Type system, $N \geq 1$, the static position error constant is given by

$$K_p = \lim_{s\to 0} G(s) = \lim_{s\to 0} \frac{K(s+z_1)(s+z_2)\dots(s+z_m)}{s^N(s+p_1)(s+p_2)\dots(s+p_n)} = \infty, \quad N \geq 1$$

Hence, for a Type-0 system, the static position error constant K_p has a finite value, whereas for a Type-1 or higher system, K_p is infinite.

Summary: The steady state error e_{ss} for a unit-step input:

$$e_{ss} = \frac{1}{K_1}, \quad \text{for Type-0 System}$$

$$e_{ss} = 0, \quad \text{for Type-1 or higher System}$$

6.3.3 Static Velocity Error Constant K_v

This is calculated when the input signal is a unit ramp signal

$$r(t) = t, \implies R(s) = \frac{1}{s^2}$$

The steady state error due to this input is given by

$$e(\infty) = \lim_{s\to 0} \frac{sR(s)}{1+G(s)} = \lim_{s\to 0} \frac{s}{1+G(s)} \frac{1}{s^2} \tag{5}$$

$$= \lim_{s\to 0} \frac{1}{s+sG(s)} = \frac{1}{\lim_{s\to 0} sG(s)} \tag{6}$$

The static velocity error constant K_v, is defined by

$$K_v = \lim_{s\to 0} sG(s)$$

Thus, the steady state error due to a unit ramp input is given as

$$e(\infty) = \frac{1}{K_v}; \quad K_v = \lim_{s\to 0} sG(s)$$

6.3.4 Static Velocity Error Constant K_v for different system types

For a Type-0 system, the static velocity error constant is given by

$$K_v = \lim_{s\to 0} sG(s) = \lim_{s\to 0} \frac{sK(s+z_1)(s+z_2)\dots(s+z_m)}{s(s+p_1)(s+p_2)\dots(s+p_n)} = 0$$

For a Type-1 system, the static velocity error constant is given by

$$K_v = \lim_{s\to 0} sG(s) = \lim_{s\to 0} \frac{K(s+z_1)(s+z_2)\dots(s+z_m)}{s(s+p_1)(s+p_2)\dots(s+p_n)} = \frac{Kz_1z_2\dots z_m}{p_1p_2\dots p_n} = K_1(\text{say})$$

For a Type-2 or higher type system, $N \geq 2$, the static velocity error constant is given by

$$K_v = \lim_{s\to 0} sG(s) = \lim_{s\to 0} \frac{K(s+z_1)(s+z_2)\dots(s+z_m)}{s^N(s+p_1)(s+p_2)\dots(s+p_n)} = \infty, \quad N \geq 2$$

Summary: The steady state error e_{ss}, for a unit ramp input:

$$e_{ss} = \frac{1}{K_v} = \infty, \quad \text{for Type-0 System}$$

$$e_{ss} = \frac{1}{K_v} = \frac{1}{K_1}, \quad \text{for Type-1 System}$$

$$e_{ss} = 0, \quad \text{for Type-2 or higher System}$$

6.3.5 Static Acceleration Error Constant K_a

This is calculated when the input signal is a unit parabolic signal

$$r(t) = \frac{t^2}{2}, \quad \forall t \geq 0, \quad r(t) = 0, \quad \forall t < 0 \quad \implies \quad R(s) = \frac{1}{s^3}$$

The steady state error due to this input is given by

$$\begin{aligned} e(\infty) &= \lim_{s\to 0} \frac{sR(s)}{1+G(s)} = \lim_{s\to 0} \frac{s}{1+G(s)} \frac{1}{s^3} && (7) \\ &= \lim_{s\to 0} \frac{1}{s^2 + s^2 G(s)} = \frac{1}{\lim\limits_{s\to 0} s^2 G(s)} && (8) \end{aligned}$$

The static acceleration error constant K_a, is defined by

$$K_a = \lim_{s\to 0} s^2 G(s)$$

Summary: The steady state error e_{ss}, due to a unit parabolic input:

$$e(\infty) = \frac{1}{K_a}; \quad K_a = \lim_{s\to 0} s^2 G(s)$$

6.3.6 Static Acceleration Error Constant K_a for different system types

For a Type-0 system, the static acceleration error constant is given by

$$K_a = \lim_{s\to 0} s^2 G(s) = \lim_{s\to 0} \frac{s^2 K(s+z_1)(s+z_2)\dots(s+z_m)}{s(s+p_1)(s+p_2)\dots(s+p_n)} = 0$$

For a Type-1 system, the static acceleration error constant is given by

$$K_a = \lim_{s\to 0} s^2 G(s) = \lim_{s\to 0} \frac{s^2 K(s+z_1)(s+z_2)\dots(s+z_m)}{s(s+p_1)(s+p_2)\dots(s+p_n)} = 0$$

For a Type-2 system, the static acceleration error constant is given by

$$K_a = \lim_{s\to 0} s^2 G(s) = \lim_{s\to 0} \frac{s^2 K(s+z_1)(s+z_2)\dots(s+z_m)}{s^2(s+p_1)(s+p_2)\dots(s+p_n)} = \frac{K z_1 z_2 \dots z_m}{p_1 p_2 \dots p_n} = K_1(\text{say})$$

For a Type-3 or higher type system, $N \geq 3$, the static acceleration error constant is given by

$$K_a = \lim_{s\to 0} s^2 G(s) = \lim_{s\to 0} \frac{s^2 K(s+z_1)(s+z_2)\dots(s+z_m)}{s^N(s+p_1)(s+p_2)\dots(s+p_n)} = \infty, \quad N \geq 3$$

Summary: The steady state error e_{ss}, for a unit parabolic input:

$$e_{ss} = \frac{1}{K_a} = \infty, \quad \text{for Type-0 and Type-1 Systems}$$

$$e_{ss} = \frac{1}{K_a} = \frac{1}{K_1}, \quad \text{for Type-2 System}$$

$$e_{ss} = 0, \quad \text{for Type-3 or higher System}$$

6.4 Summary: Steady State Error and Static Error Constants

For unit step, unit ramp and unit acceleration (parabolic) input i.e. if

$$u(t) = 1(t) + t + \frac{1}{2}t^2$$

the various steady state error constants and associated steady state error are

$$K_p = \lim_{s \to 0} G(s), \quad e(\infty) = \frac{1}{1 + K_p}$$

$$K_v = \lim_{s \to 0} sG(s), \quad e(\infty) = \frac{1}{K_v}$$

$$K_a = \lim_{s \to 0} s^2 G(s), \quad e(\infty) = \frac{1}{K_a}$$

For a general input of the form

$$u(t) = a\,1(t) + bt + ct^2$$

the various steady state error constants and associated steady state error are

$$K_p = \lim_{s \to 0} G(s), \quad e(\infty) = \frac{a}{1 + K_p}$$

$$K_v = \lim_{s \to 0} sG(s), \quad e(\infty) = \frac{b}{K_v}$$

$$K_a = \lim_{s \to 0} s^2 G(s), \quad e(\infty) = \frac{2c}{K_a}$$

6.4.1 Examples

6.4.1.1 Example 1

Chapter 7

Stability Analysis of Linear Systems: Root Locus Analysis

Learning Outcomes

After completing this module, you should be able to:

1. Sketch the root locus
2. Conduct relative stability analysis

7.1 Introduction

While designing any control system, it is often necessary to investigate the performance of the system when one or more parameters of the system vary over a given range. Further, it is known that the dynamic behaviour (e.g. transient response) of a closed loop system is closely related to the location of the closed-loop poles. (i.e. location of the roots of the closed loop characteristic equation). Therefore, it is important for the designer to know how the closed-loop poles (i.e. the roots of the characteristic equation) move in the s plane as one or more parameters of the system are varied over a given range.

A simple method for finding the roots of the characteristic equation has been developed by W.R. Evans. This method, called the root-locus method, is one in which the roots of the characteristic equation are plotted for all values of a system parameter.

Note that the root locus technique is not confined to inclusive study of control systems. The equation under investigation does not necessarily have to be the characteristic equation. The technique can also be used to assist in the determination of roots of higher-order algebraic equations.

The root locus problem for one variable parameter can be defined by referring to equations of the form:

$$F(s) = s^n + a_1 s^{n-1} + ... + a_{n-1} s + a_n + K(s^m + b_1 s^{m-1} + ... + b_{m-1} s + b_m) = 0 \tag{1}$$

where K is the parameter considered to vary between $-\infty$ and ∞.

The coefficients, $a_1, \dots, a_n$ and $b_1, \dots, b_m$ etc. are assumed to be fixed constants.

The various categories of root loci are defined as follows

1. **Root Loci:** The portion of the root loci when K assumes positive values; that is $0 \leq K < \infty$.
2. **Complementary Root Loci:** The portion of the root loci when K assumes negative values; that is $-\infty \leq K \leq 0$.
3. **Root Contours:** The loci of roots when more than one parameter varies.

The complete root loci refers to the combination of the root loci and the complementary root loci.

7.1.1 What is the Root Locus and Why is it useful?

The root locus is the locus of roots (duh) of the characteristic equation of the closed-loop system as a specific parameter (usually, gain K) is varied from zero to infinity. Such a plot clearly shows the contributions of each open-loop pole or zero to the locations of the closed-loop poles.

Is it useful in Linear Control Systems Design?

It indicates the manner in which the open-loop poles and zeros should be modified so that the response meets system performance specifications. For example, by using the root locus method, it is possible to determine the value of the loop gain K that will make the damping ratio of the dominant closed-loop poles as prescribed.

If the location of an open-loop pole or zero is a system variable, then the root-locus method suggests the way to choose the location of an open-loop pole or zero.

7.2 Basic Conditions of the Root Loci

Consider the system shown in Figure.

The closed-loop transfer function is given by:

$$T(s) = \frac{C(s)}{R(s)} = \frac{K \cdot G(s)}{1 + K \cdot G(s) \cdot H(s)}$$

The closed loop characteristic equation of the system is:

$$1 + K \cdot G(s) \cdot H(s) = 0$$

Observe that the closed loop transfer function, $T(s)$, as well as the open loop transfer function $K \cdot G(s) \cdot H(s)$, both involve a gain parameter K.

7.2.1 Concept of Root Locus

Definition The root locus is the path of the roots of the characteristic equation traced out in the complex plane as a system parameter is changed.

7.2.1.1 Example 1

Example: Consider the video camera control system shown.

The closed-loop transfer function of this system is as follows

$$\frac{C(s)}{R(s)} = \frac{K_1 K_2}{s^2 + 10s + K_1 K_2} = \frac{K}{s^2 + 10s + K}$$

Where $K = K_1 K_2$.

The closed loop characteristic equation is given by

$$s^2 + 10s + K = 0$$

The location of poles as the open loop gain K is varied is shown in the Table.

K	Pole 1	Pole 2
0	-10	0
5	-9.47	-0.53
10	-8.87	-1.13
15	-8.16	-1.84

K	Pole 1	Pole 2
20	-7.24	-2.76
25	-5	-5
30	$-5 + j2.24$	$-5 - j2.24$
35	$-5 + j3.16$	$-5 - j3.16$
40	$-5 + j3.87$	$-5 - j3.87$
45	$-5 + j4.47$	$-5 - j4.47$
50	$-5 + j5$	$-5 - j5$

From the plot, it is seen that for $K = 0$, the poles are at $p_1 = -10,\ p_2 = 0$. As K increases, p_1 moves towards the right, while p_2 moves towards the left. For $K = 25$, the poles p_1 and p_2 meet at -5, break away from the real axis, and move into the complex plane.

Further, if $0 < K < 25$, the poles are real and distinct, and the system is overdamped. If $K = 25$, the poles are real and multiple (i.e. repeated), and the system is critically damped. If $K > 25$, the poles are complex conjugate (i.e. $\sigma \pm j\omega$), and the system is underdamped.

7.2.1.2 Example 2

7.2.1.3 Example 3

7.2.2 Angle and Magnitude Conditions

7.3 Sketching the Root Locus

7.3.1 Rules for Sketching the Root Locus

Rule 1: Total Number of Branches of the Root Locus

The number of branches of the root locus is equal to the number of closed-loop poles. Thus, the number of branches is equal to the number of open-loop poles or open-loop zeros, whichever is greater.

- Let n be the number of finite open loop poles.
- Let m be the number of finite open loop zeros.
- Let N be the number of root locus branches, then

$$N = n, \quad \text{if } n \geq m \tag{2}$$
$$N = m, \quad \text{if } m < n \tag{3}$$

Rule 2: Where the Root Locus Starts and Terminates

Root locus branches start from open-loop poles (when $K = 0$) and terminate at open-loop zeros (finite zeros or zeros at infinity) (when $K = \infty$).

- If the number of open-loop poles is greater than the number of open-loop zeros, some branches starting from finite open-loop poles will terminate at zeros at infinity (i.e., go to infinity).

Rule 3: Symmetry of the Root Locus

The root locus is symmetric about the real axis (i.e. x-axis), which reflects the fact that the closed loop poles appear in complex conjugate pairs.

Rule 4: Determination of Root Loci Segments on the Real Axis

Segments of the real axis are part of the root locus if and only if the total number of real poles and zeros to their right is odd.

Rule 5: Asymptotic Behaviour of Root Locus

If the number of poles n exceeds the number of zeros m, then as the gain $K \to \infty$ (i.e. K goes to infinity), then $(n-m)$ branches will become asymptotic to the straight lines which intersect the real axis at the point σ, called the centroid, and inclined to the real axis at angles θ_k, called the angle of asymptotes.

Thus, the equation of the asymptotes is given by the real axis intercept σ, called the centroid, and the angle of the asymptotes θ_k, as follows:

$$\sigma = \frac{\sum_{i=1}^{n} p_i - \sum_{j=1}^{m} z_j}{n-m} = \frac{\text{Sum of Open Loop Poles} - \text{Sum of Open Loop Zeros}}{\text{Number of Open Loop Poles} - \text{Number of Open Loop Zeros}}$$

$$\theta_k = \frac{(2k+1)\pi}{n-m}[\text{rad}] = \frac{(2k+1)\cdot 180}{n-m}[\text{degrees}], \quad k = 0, 1, 2, \dots, n-m-1$$

Note that the angle of asymptotes gives the direction along which these $(n-m)$ branches approach infinity.

Rule 6: Real Axis Breakaway and Break-in Points

i. If there exists a real axis root locus branch between two open loop poles, then there will be a break-away point in between these two open loop poles.

ii. If there exists a real axis root locus branch between two open loop zeros, then there will be a break-in (re-entry) point in between these two open loop zeros.

- The root locus breaks away from the real axis at a point where the gain is maximum and breaks into the real axis at a point where the gain is minimum.

Computation of Breakaway and Break-in Points

The break away and re-entry points on the root locus are determined from the roots of

$$\frac{\mathrm{d}K}{\mathrm{d}s} = 0$$

if r-number of branches meet at a point, they breakaway at an angle of $180°/r$.

Rule 7: Angle of Departure and Angle of Arrival

The root locus departs from complex, open-loop poles and arrives at complex, open loop zeros.

i. The angle of departure from an open-loop complex pole θ_d is computed as:

$$\theta_d = 180° - \left(\sum \vec{\theta}_{\text{pole to pole}}\right) + \left(\sum \vec{\theta}_{\text{zero to pole}}\right)$$

Where $\vec{\theta}_{\text{pole to pole}}$ is the angle of the vector from the complex pole to other poles and $\vec{\theta}_{\text{zero to pole}}$ is the angle of the vector from a complex zero to the pole.

ii. The angle of arrival at an open-loop complex zero θ_a is computed as:

$$\theta_a = 180° - \left(\sum \vec{\theta}_{\text{zero to zero}}\right) + \left(\sum \vec{\theta}_{\text{pole to zero}}\right)$$

Where $\vec{\theta}_{\text{zero to zero}}$ is the angle of the vector from the complex zero to other zeros and $\vec{\theta}_{\text{pole to zero}}$ is the angle of the vector from a complex pole to the zero.

Rule 8: Imaginary Axis Crossover

The points where the root loci intersect the $j\omega$-axis can be found easily by

a. Use of Routh's stability criterion or

b. Letting $s = j\omega$ in the characteristic equation, equating both the real and imaginary parts to zero, and solving for ω and K.

The values of ω, thus found, give the frequencies at which root loci cross the imaginary axis. The K value corresponding to each crossing frequency gives the gain at the crossing point.

7.3.2 Step by Step Procedure for Sketching the Root Locus

1. Determine the open loop poles, zeros and a number of branches from given $G(s)H(s)$.
2. Draw the pole-zero plot (???) and determine the region of the real axis for which the root locus exists. Also, determine the number of breakaway points.
3. Calculate the angle of asymptotes.
4. Determine the centroid.
5. Calculate the breakaway points (if any).
6. Calculate the intersection point of the root locus with the imaginary axis.
7. Calculate the angle of departure and angle of arrivals if any.
8. From above steps, draw the overall sketch of the root locus.
9. Predict the stability and performance of the given system by the root locus.

7.3.2.1 Example 1

7.4 Qualitative Analysis Through Root Locus

7.4.1 Effect of Adding a Zero

Consider adding a zero to a simple second order system i.e.

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

The root locus for both the cases are shown in the Figure.

The branches of the root locus have been "pulled to the left", or farther from the imaginary axis. For values of static loop sensitivity greater than K_a, the roots are farther to the left than for the original system. Therefore, the transients will decay faster, yielding a more stable system.

7.4.2 Effect of Adding a Pole

Consider adding a pole to the same simple second order system i.e.

$$G(s) = \frac{K}{s(s+a)} \implies G(s) = \frac{K}{s(s+a)(s+c)}$$

The root locus for both the cases are shown in the Figure.

The branches of the root locus have been "pulled to the right", or closer to the imaginary axis. For values of static loop sensitivity greater then K_a, the roots are closer to the imaginary axis compared to the original system. Therefore, the transients will decay slowly, and will yield a less stable system.

Part II

Nitish's Content