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

By the end of 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 $j314\,\mathrm{rad/s}$, then this responds to a pure sinusoid of frequency $314\,\mathrm{rad/s}$ (i.e. $50\,\mathrm{Hz}$).

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:

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

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

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

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

Chapter 3

The Block Diagram Representation & Characteristics of Feedback Systems

Learning Outcomes

Chapter 4

Time Domain Analysis of Linear Systems: Time Domain Specifications

Learning Outcomes

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

5.1.1 Definitions of Stability, Instability and Marginal Stability

5.2 Stability of LTI system from Transfer Function

5.3 Necessary Conditions of Stability

5.4 Routh's Stability Criterion

5.4.1 Formation of Routh's Tables

5.5 Routh-Hurwitz Stability Criterion for Linear Feedback Control Systems

5.5.1 Examples

5.6 Limitations of Routh-Hurwitz Criterion

5.7 Relative Stability Analysis

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

$$e_{ss} = e(\infty) = \lim_{t \to \infty} e(t) \tag{1}$$

$$= \lim_{s \to 0} sE(s) = \lim_{s \to 0} \frac{sR(s)}{1 + G(s)} \tag{2}$$

Thus the steady state error is expressed as

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

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:

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

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

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

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

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

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

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

$$K_v = \lim_{s \to 0} sG(s) = \lim_{s \to 0} \frac{K(s + z_1)(s + z_2)\dots(s + z_m)}{s(s + p_1)(s + p_2)\dots(s + p_n)} = \frac{Kz_1 z_2 \dots z_m}{p_1 p_2 \dots p_n} = K_1 \text{(say)}$$

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

$$e(\infty) = \lim_{s \to 0} \frac{sR(s)}{1 + G(s)} = \lim_{s \to 0} \frac{s}{1 + G(s)} \frac{1}{s^3} \tag{7}$$

$$= \lim_{s \to 0} \frac{1}{s^2 + s^2 G(s)} = \frac{1}{\lim_{s \to 0} s^2 G(s)} \tag{8}$$

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

At the end of 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, $a_1, ..., a_n$ and $b_1, ..., b_m$ 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
20	-7.24	-2.76
25	-5	-5

K	Pole 1	Pole 2
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:

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

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

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