UNIVERSITY OF AUCKLAND
Waipapa Taumata Rau
NEW ZEALAND

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

List of Figures

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:

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

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

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

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:

1. Understand the Basic advantages of closed-loop control system compared to open loop control system.
2. Understand the Sensitivity reduction due to parameter variations.
3. Understand the Reduction of effects of noise due to feedback.
4. Understand the Control of transient response (speed) by feedback.

3.1 Introduction to the Block Diagram Representation

A typical control systems is often complex and consists of interconnection of several components/elements which are practically nonlinear. Although, their input-output behaviour can be approximated by linear differential equations, it is difficult to derive the complete behaviour of a complex control system via a single equation or function. Therefore, it is always desirable to represent the dynamics of each of the components separately, and then connect them together to represent the whole system. This is the motivation of block diagram representation of control systems where each component of the control system is represented with a block.

Advantages:

1. Possible to represent a complex system by merely connecting the blocks of the components according to the signal flow.
2. The contribution of each component to the overall system performance can be evaluated.
3. It becomes easier to visualize the the functional operation of the system by examining the block diagram.
4. The overall transfer function can easily be computed.

Disadvantages:

1. No information about the physical construction of the system can be obtained. Therefore, a block diagram may represent many dissimilar and unrelated systems
2. The main source of energy is not explicitly shown.
3. The same system can be represented by different block diagrams.

3.2 Basic Elements of Block Diagrams

A block diagram essentially has three basic elements: the **block**, the **summing point** and the **junction or pick-off/take-off point**. In the following we will discuss each of these elements.

3.2.1 The Block

A block is used to represent the proportional relationship between two Laplace transformed signals which are the input and output of the block. The proportionally function is the transfer function or transmittance which relates the incoming and outgoing signals of the block. Thus, each of the individual blocks is the symbolic representation of the transfer function of a particular component of the control system (i.e. a piece of a larger puzzle). The complete control system is represented by the interconnection of required number of blocks. The steps of for block diagram representation are as follows:

1. Express the dynamics of each component by linear differential equations.
2. Take the Laplace transform ($\mathcal{L}$) of these differential equations. This converts these into the s-domain, meaning that they are now algebraic equations in terms of the complex frequency variable s.
3. Select the input and output variable of each component.
4. Compute the transfer function of each component. This is defined as the ratio of the Laplace transform of the output and the input under zero initial conditions.
5. Represent this transfer function in block form.

Note that for each block, the input may be regarded as the "cause", and the output as the "effect". Since the "effect" cannot produce the "cause", the block diagram is **"unidirectional"**. The transfer function $G(s)$ relates the Laplace transform $Y(s)$ of the output $y(t)$ to the laplace transform $X(s)$ of the input $x(s)$ through the relationship:

$$Y(s) = G(s)X(s)$$

Main Block Diagram Rule *The output signal of a block is the product of the transfer function of the given block and the input signal of the block.*

3.2.2 The Summing Point

The summing point is used to represent the addition or subtraction of signals. It can have any number of incoming signals. However, it can only have one outgoing signal. The algebraic sign of each incoming signal is shown next to the arrowhead of this signal.

3.2.3 The Junction or Pick-off/Take-off Point

The junction or pick-off/take-off point is the point in the block diagram where the same signal goes to more than one place. This is shown by a circle at the intersection where the signal splits.

3.3 Modelling of an Armature Controlled DC Motor

Consider an armature controlled D.C. Motor shown in the figure above, where:

R_a is the resistance of the armature (Ω). L_a is the inductance of the armature (H). i_a is the armature current (A). i_f is the field current (A). e_a is the applied armature voltage (V). e_b is the back emf (V). T_m is the mechanical torque developed by the motor (N m). T_d is the disturbance torque (N m). θ is the angular displacement of the motor shaft and load referred to the motor shaft (rad). J is the equivalent moment of inertia of the motor and load referred to the motor shaft (kgm^2). f_0 is the equivalent viscous friction coefficient of the motor and load referred to the motor shaft ($\left(\frac{\text{N m}}{\text{rad/s}}\right)$).

When DC motors are used for servo applications, their operation is restricted in the linear range of the magnetization curve. The air gap flux ϕ is therefore proportional to the field current i.e.

$$\phi = K_f i_f \tag{1}$$

where K_f is a constant.

The Torque T_m developed by the motor is proportional to the product of the armature current i_a and the air gap flux ϕ, i.e.

$$T_m \propto \phi i_a \implies T_m = K_1 K_f i_f i_a \tag{2}$$

Since in an armature controlled DC motor, the field current is kept constant, eqn(Equation 2) can be written as:

$$T_m = K_T i_a \tag{3}$$

where $K_T = K_1 K_f i_f$, and is called the motor torque constant.

The motor back emf is proportional to the speed and is given as

$$e_b = K_b \frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{4}$$

where K_b is the back emf constant.

Note that the DC motor is an electro-mechanical system. This means it has both electrical and mechanical dynamics.

The electrical dynamics of the DC motor associated with the armature circuit is given by

$$R_a i_a + L_a \frac{\mathrm{d}i_a}{\mathrm{d}t} + e_b = e_a \quad \text{where} \quad e_b = K_b \frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{5}$$

The mechanical dynamics of the DC motor is given by

$$J \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + f_0 \frac{\mathrm{d}\theta}{\mathrm{d}t} + T_d = T_m \quad \text{where} \quad T_m = K_T i_a \tag{6}$$

Taking the Laplace transform of eqn(Equation 3), eqn(Equation 4), eqn(Equation 5) and eqn(Equation 6) with the assumption of zero initial conditions gives

$$\begin{aligned}
\mathcal{L}\left[T_m\right] = K_T \mathcal{L}\left[i_a\right] &\implies T_m(s) = K_T I_a(s) \\
\mathcal{L}\left[e_b\right] = K_b \mathcal{L}\left[\frac{\mathrm{d}\theta}{\mathrm{d}t}\right] &\implies E_b(s) = K_b s \theta(s) \\
\mathcal{L}\left[R_a i_a\right] + \mathcal{L}\left[L_a \frac{\mathrm{d}i_a}{\mathrm{d}t}\right] + \mathcal{L}\left[e_b\right] = \mathcal{L}\left[e_a\right] &\implies (L_a s + R_a) I_a(s) = E_a(s) - E_b(s) \\
\mathcal{L}\left[J \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}\right] + \mathcal{L}\left[f_0 \frac{\mathrm{d}\theta}{\mathrm{d}t}\right] + \mathcal{L}\left[T_d\right] = \mathcal{L}\left[T_m\right] &\implies (J s^2 + f_0 s)\theta(s) = T_m(s) - T_d(s)
\end{aligned} \tag{7}$$

From the complete block diagram, if the torque disturbance $T_d(s) = 0$, then the transfer function between the angular position and the armature voltage is given by

$$G(s) = \frac{\theta(s)}{E_a(s)} = \frac{K_T}{s[(R_a + L_a s)(J s + f_0) + K_T K_b]} \tag{8}$$

If the armature circuit inductance L_a is neglected, then the transfer function in eqn(Equation 8) simplifies to

$$G(s) = \frac{\theta(s)}{E_a(s)} = \frac{\frac{K_T}{R_a}}{J s^2 + s\left(f_0 + \frac{K_T K_b}{R_a}\right)} \tag{9}$$

The term $f_0 + \frac{K_T K_b}{R_a}$ indicates that due to back emf, the effective damping (viscous friction) of the system increases. Let us us define the effective viscous friction as:

$$f = f_0 + \frac{K_T K_b}{R_a}$$

Then the transfer function in eqn(Equation 9) reduces to

$$\frac{\theta(s)}{E_a(s)} = \frac{\frac{K_T}{R_a}}{s(Js+f)} \tag{10}$$

The transfer function in eqn(Equation 10) can be expressed in the form

$$\frac{\theta(s)}{E_a(s)} = \frac{K_m}{s(s\tau_m+1)} \tag{11}$$

where $K_m = \dfrac{K_T}{fR_a}$ is the motor gain constant, and $\tau_m = \dfrac{J}{f}$ is the motor time constant.

3.4 Simple Rules of Block Diagram Reduction

The pictorial representation of a complex control system with several interconnected components by block diagram can provide better physical insight about the structure of the control system represented. If the various components of the system are non-interacting, i.e. there is no loading effect of one component on another, then it is possible to obtain the overall transfer function of the system by algebraically manipulating the block diagram using some simple rules of block diagram transformations.

3.4.1 Serially/Cascade Connected Blocks

When the blocks are connected in series/cascade, then the overall transfer function of all the blocks in this way, is the multiplication of the transfer functions of the individual blocks. i.e.

$$\text{If } G_1(s), G_2(s), \dots, G_n(s) \text{ are connected in series, then:}$$
$$\forall i \in \mathbb{N}, \quad G(s) = \prod_{i=1}^{n} G_i(s)$$

Where $\mathbb{N} \equiv \{1, 2, \dots, n\}$, i.e. the set of natural numbers, and $G(s)$ is the overall transfer function of the system connected in series.

3.4.2 Parallel Connected Blocks

When the blocks are connected in parallel, then the overall transfer function of all the blocks in this way, is the sum of the transfer functions of the individual blocks. i.e.

$$\text{If } G_1(s), G_2(s), \dots, G_n(s) \text{ are connected in parallel, then:}$$
$$\forall i \in \mathbb{N}, \quad G(s) = \sum_{i=1}^{n} \pm G_i(s)$$

Where $\mathbb{N} \equiv \{1, 2, \dots, n\}$, i.e. the set of natural numbers, and $G(s)$ is the overall transfer function of the system connected in parallel.

3.5 Desired Properties of a Control System

1. Accuracy
2. Sensitivity: Any control system should be insensitive to internal disturbances but sensitive to input signals only.
3. Noise: An undesired input signal is known as noise. A good control system should be able to reduce the effects of noise for better performance.
4. Stability: It is an important characteristic of the control system. For the bounded input signal, the output must be bounded (BIBO Stability) and if the input is zero, then the output must be zero. Then, such a control system is said to be a stable system.

5. Bandwidth: An operating frequency range decides the bandwidth of the control system. Bandwidth should be as large as possible for the frequency response of a good control system.
6. Speed (Speed of Response): It is the time taken by the control system to achieve its stable output. A good control system possesses high speed. The transient period for such a system is very small.
7. Oscillation: A small number of oscillations or constant oscillation of the output tend to indicate that the system is stable.

3.6 Characteristics of Feedback Systems

Some of the beneficial effects of feedback with high loop gain are:

1. The controlled variable accurately follows the desired value.
2. The effects of external disturbances on the controlled variable are significantly reduced.
3. The effects of variations of process and controller parameters is reduced.
   i. These variations occur due to wear, aging, environmental changes, etc.
4. The speed of response can be improved.

The cost of achieving these improvements includes:

a. Greater system complexity.
b. Need for much larger forward path gain
c. Possibility of instability. This may mean undesired/persistent oscillations of the output variable.

3.7 What is Feedback and What are its Effects?

Before we begin, let us first find the input-output relation in a feedback system. The block diagram of a feedback system is shown below.

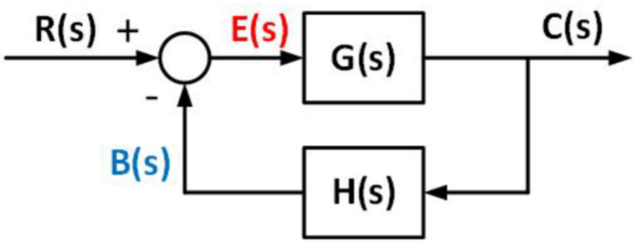

Figure 3.1: Feedback System Block Diagram

From the schematic, the relation between the output and input is computed as:

$$C(s) = G(s)E(s), \qquad \text{where} \quad E(s) = R(s) - B(s) = R(s) - C(s)H(s)$$

Simple algebraic manipulation gives:

$$\begin{aligned} C(s) &= G(s)E(s) \\ &= G(s)[R(s) - C(s)H(s)] \\ &= G(s)R(s) - G(s)C(s)H(s) \\ \text{or} \quad C(s)[1 + G(s)H(s)] &= G(s)R(s) \\ \text{Therefore,} \quad \frac{C(s)}{R(s)} &= \frac{G(s)}{1 + G(s)H(s)} \end{aligned}$$

Using this basic relationship of the feedback system structure, we can uncover some of the significant effects of feedback.

3.7.1 Effects of Feedback on System Gain

We can relate the input and output of a system *without* feedback as:

$$C(s) = G(s)R(s) \quad \text{or,} \quad \frac{C(s)}{R(s)} = G(s)$$

As computed previously, based on the schematic of a feedback system, the input-output relation is:

$$\begin{aligned} E(s) &= R(s) - B(s) = R(s) - C(s)H(s) \\ C(s) &= G(s)E(s) = G(s)[R(s) - C(s)H(s)] \\ \text{or,} \quad C(s) &= G(s)R(s) - G(s)C(s)H(s) \\ \text{or,} \quad C(s)[1 + G(s)H(s)] &= G(s)R(s) \\ \text{Therefore,} \quad \frac{C(s)}{R(s)} &= \frac{G(s)}{1 + G(s)H(s)} \end{aligned}$$

Thus, the feedback affects the gain of the open loop system by a factor $1 + G(s)H(s)$.

3.7.2 Effects of Feedback on Noise

To know the effect of feedback on noise, let us compare the transfer functions with and without feedback due to the noise signal alone. Consider an open loop system with noise signal as shown below.

The output is expressed as:

$$C(s) = [G_a(s)R(s) + N(s)]G_b(s) = G_a(s)G_b(s)R(s) + G_b(s)N(s)$$

The open loop transfer function due to noise alone can be obtained by making $R(s) = 0$ which gives:

$$\frac{C(s)}{N(s)} = G_b(s)$$

Consider the closed loop system with noise signal as shown below.

The output is expressed as:

$$C(s) = [(R(s) - H(s)C(s))G_a(s) + N(s)]G_b(s) = [G_a(s)R(s) - G_a(s)H(s)C(s) + N(s)]G_b(s)$$

Further simplification gives:

$$\begin{aligned} C(s) + G_a(s)G_b(s)H(s)C(s) &= G_a(s)G_b(s)R(s) + G_b(s)N(s) \\ \text{or,} \quad C(s)[1 + G_a(s)G_b(s)H(s)] &= G_a(s)G_b(s)R(s) + G_b(s)N(s) \end{aligned}$$

Thus

$$C(s) = \frac{G_a(s)G_b(s)}{1 + G_a(s)G_b(s)H(s)}R(s) + \frac{G_b(s)}{1 + G_a(s)G_b(s)H(s)}N(s)$$

The closed loop transfer function due to noise alone can be obtained by making $R(s) = 0$ which gives:

$$\frac{C(s)}{N(s)} = \frac{G_b(s)}{1 + G_a(s)G_b(s)H(s)}N(s)$$

This shows that in the closed loop control system, the gain due to the noise signal is decreased by a factor of $1+G_a(s)G_b(s)H(s)$. Note that, in most practical control systems, $(1 + G_a(s)G_b(s)H(s))$ is greater than 1.

3.7.3 Effects of Feedback on Speed of Response

3.7.3.1 Case 1: Open Loop System

Consider an open loop system with

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

The impulse response of this system is given by

$$c(t) = \mathcal{L}^{-1}\left[C(s)\right] = \mathcal{L}^{-1}\left[\frac{K}{s+a}\right] = Ke^{-at}$$

The time constant of the system τ or T associated with this mode of response equals to $\frac{1}{a}$.

3.7.3.2 Case 2: Closed Loop System

When the feedback loop is closed with unity feedback, then

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

The impulse response of this system is given by

$$c(t) = \mathcal{L}^{-1}\left[C(s)\right] = \mathcal{L}^{-1}\left[\frac{K}{s+a+K}\right] = Ke^{-(a+K)t}$$

The time constant of the system τ or T associated with this mode of response equals to $\frac{1}{a+K}$.

3.7.4 Effects of Feedback on Stability

Consider the basic feedback system.

The input output relation is given by:

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

From this it is obvious that when $G(s)H(s) = -1$, the output of the system will be infinite for any finite input, and the system is said to be unstable.

Thus, Feedback can cause a system that is originally stable to become unstable.

We will demonstrate that feedback can stabilize an unstable system.

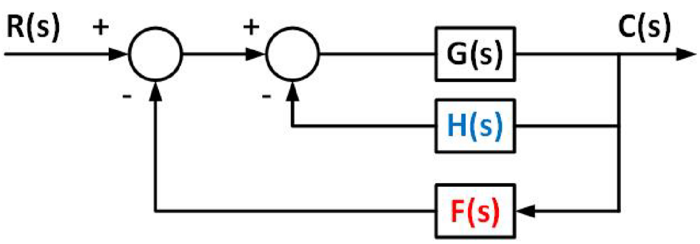

Figure 3.2: Two Loop Feedback System Block Diagram

Let us introduce another feedback loop through a negative feedback gain $F(s)$ as shown above. Then

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

The overall system can be made stable by properly selecting the outer feedback gain $F(s)$.

3.7.5 Reduction of the Effects of Parameter Variations

Let us define sensitivity on a quantitative basis. In the open loop case

$$C(s) = G(s)R(s)$$

Suppose, due to parameter variations, $G(s)$ changes to $[G(s) + \Delta G(s)]$.

The output of the open-loop system therefore changes to

$$C(s) + \Delta C(s) = [G(s) + \Delta G(s)]R(s), \qquad \text{Thus} \quad \Delta C(s) = \Delta G(s)R(s)$$

Similarly in the closed loop case, the output is given by

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

Due to variation $\Delta G(s)$ in the forward path transfer function, this changes to

$$C(s) + \Delta C(s) = \frac{G(s) + \Delta G(s)}{1 + [G(s) + \Delta G(s)]H(s)}R(s) = \frac{G(s) + \Delta G(s)}{1 + G(s)H(s) + \Delta G(s)H(s)}R(s)$$

Since $|G(s)| \gg |\Delta G(s)|$, the variation in the output can be expressed as:

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

1. **Open Loop System**

$$\Delta C(s) = \Delta G(s)R(s)$$

2. **Closed Loop System**

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

Thus, compared to the open loop system, the change in the output of the closed loop system due to variations in $G(s)$ is reduced by a factor of $[1 + G(s)H(s)]$ which often much greater than unity or 1.

3.7.6 Sensitivity Reduction Due to Feedback

The term *system sensitivity* is used to describe the relative variation in the overall transfer function $T(s) = \dfrac{C(s)}{R(s)}$ due to variation in $G(s)$ and is defined as:

$$\text{Sensitivity} = \frac{\text{Percentage change in } T(s)}{\text{Percentage change in } G(s)}$$

For a small incremental variation in $G(s)$, the sensitivity of T with respect to G is expressed quantitatively as:

$$S_{G(s)}^{T(s)} = \frac{\partial T(s)/T(s)}{\partial G(s)/G(s)} = \frac{\partial \ln T(s)}{\partial \ln G(s)}$$

The sensitivity of the closed loop system is

$$S_{G(s)}^{T(s)} = \frac{\partial T(s)}{\partial G(s)} \times \frac{G(s)}{T(s)} = \frac{(1+G(s)H(s)) - G(s)H(s)}{(1+G(s)H(s))^2} \times \frac{G(s)}{\frac{G(s)}{1+G(s)H(s)}} = \frac{1}{1+G(s)H(s)}$$

The sensitivity of the open loop system is

$$S_{G(s)}^{T(s)} = \frac{\partial T(s)}{\partial G(s)} \times \frac{G(s)}{T(s)} = 1, \quad (\because \text{Here } T(s) = G(s))$$

Thus due to variations in $G(s)$, the sensitivity of the closed loop system is reduced by a factor of $1 + G(s)H(s)$ compared to the open loop system.

The sensitivity of $T(s)$ with respect to the feedback sensor $H(s)$ is given as

$$S_{H(s)}^{T(s)} = \frac{\partial T(s)}{\partial H(s)} \times \frac{H(s)}{T(s)} = G(s)\left[\frac{-G(s)}{(1+G(s)H(s))^2}\right] \times \frac{H(s)}{\frac{G(s)}{1+G(s)H(s)}} = \frac{-G(s)H(s)}{1+G(s)H(s)}$$

This shows that for large values of $G(s)H(s)$ the sensitivity of the feedback system with respect to H is unity. This implies that changes in H directly affect the system output. Therefore, it is important to use feedback elements which remain substantially constant and do not vary with environmental changes.

Very often, we are interested to find the sensitivity of a system with respect to variation in a particular parameter or set of parameters.

Let the transfer function of the system be expressed as

$$T(s) = \frac{N(s,\alpha)}{D(s,\alpha)}; \quad \alpha = \text{parameter under consideration}$$

The sensitivity of $T(s)$ with respect to the parameter α is given by

$$S_{\alpha}^{T(s)} = \left.\frac{\partial \ln N(s)}{\partial \ln \alpha}\right|_{\alpha_0} - \left.\frac{\partial \ln D(s)}{\partial \ln \alpha}\right|_{\alpha_0} = S_{\alpha}^{N(s)} - S_{\alpha}^{D(s)}$$

where α_0 is the nominal value of the parameter around which variation occurs.

To have a highly accurate open-loop system, the components of $G(s)$ must be selected to rigidly meet the specifications. However, in a closed loop system, $G(s)$ may be less rigidly specified, since the effects of parameter variations can be mitigated by use of feedback. Furthermore, however, a closed loop system requires careful selection of components of the feedback sensor $H(s)$. Note that $H(s)$ is often made up of measuring elements which operate at lower power levels and are less costly.

3.7.7 Effect of High Gain in a Feedback System

It will be shown that using high gain in a feedback system can make the output track the input.

Consider a system with gain K.

The closed loop output $(C(s))$ and the error $(E(s))$ response can be expressed as:

$$C(s) = \frac{KG(s)}{1+KG(s)}R(s), \quad E(s) = \frac{1}{1+KG(s)}R(s)$$

From these, it is obvious that as $K \to \infty$, $C(s) \to R(s)$ and $E(s) \to 0$.

- Open Loop Gain: $|KG(s)| \gg 1$

- Closed Loop Gain: $\left|\frac{KG(s)}{1+KG(s)}\right| \approx 1$

Thus we can make the output track the input even if we do not know the exact value of the open loop gain.

3.8 Advantages and Disadvantages of Closed Loop Control

Advantages of a Closed Loop Control System:

1. More accurate even in the presence of non-linearity.
2. Highly accurate as any error is corrected due to the presence of the feedback signal.
3. Bandwidth range is large.
4. Facilitates automation.
5. The sensitivity of the system may be made small to make the system more stable.
6. The system is less affected by noise.

Disadvantages of a Closed Loop Control System:

1. Costly to implement
2. Complexity is increased.
3. More maintenance is required.
4. Feedback can lead to oscillations.
5. Feedback reduces overall gain.
6. Stability is the most major problem and more care is needed to design a stable closed loop system.

3.8.1 Comparison Between Open Loop and Closed Loop Control

Open Loop Control

1. The feedback element is absent.
2. An error detector is not present.
3. Easy to construct.
4. Have small bandwidth
5. Often Stable.
6. Less maintenance.
7. Often unreliable.

Closed Loop Control

1. The feedback element is present.
2. An error detector is always present.
3. Complicated construction.
4. Have large bandwidth.
5. May become unstable.
6. More maintenance.
7. Reliable.

3.9 Summary of Important Characteristics of a Feedback System

1. Decreased Sensitivity of the system to variations in the process parameters.
2. Improved rejection of disturbances.
3. Improved measurement noise attenuation.
4. Improved reduction in steady state error of the system.
5. Easy control and adjustment of the transient response of the system.

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:

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

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:

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

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

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

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

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

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. $\lambda_1 = \lambda_2 = ... = \lambda_k$, then their contribution to the output is of the form

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

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

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

Key Point:

When a system has repeated poles of k-th order in its transfer function, the output response can include terms like $t, t^2, \dots, t^{k-1}$ 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

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

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

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

$$\begin{aligned} e(\infty) &= \lim_{s\to 0} \frac{sR(s)}{1+G(s)} = \lim_{s\to 0} \frac{s}{1+G(s)} \frac{1}{s^2} && (5) \\ &= \lim_{s\to 0} \frac{1}{s+sG(s)} = \frac{1}{\lim\limits_{s\to 0} sG(s)} && (6) \end{aligned}$$

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

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

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

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

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:

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

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

Chapter 8

Frequency Domain Analysis of Linear Systems

Learning Outcomes

After completing this module, you should be able to:

- Analytically compute the frequency response of a second order system
- Determine various frequency domain specifications such as bandwidth, resonant frequency etc.
- Correlate the frequency domain specifications with time domain response.

8.1 What is the Frequency Response of a System?

The frequency Response of a system is the steady state response of the system to a sinusoidal input.

Define: Frequency Range, Amplitude

1. Apply one frequency
2. Study resulting response to make a valid amplitude and phase
3. Repeat with another frequency

For linear systems, the frequency of input and output signals remains the same. i.e. in linear systems, energy transfer from input to output occurs at the **same frequency**, while the ratio of magnitude of output signal to the input signal and the phase between the two signals may change.

8.2 Root Locus Method vs. Frequency Response Method

Advantages of the Root Locus Method:
- Good indicator of transient response
- Explicitly shows location of all closed-loop poles
- Trade-offs in the design are fairly clear

Disadvantages of the Root Locus Method:
- Requires a transfer function model (poles and zeros);
- Difficult to infer all performance metrics;
- Hard to determine response to steady-state (sinusoids)
- Hard to infer stability margins

Advantages of the Frequency Response Method:

- Frequency response methods are a good complement to the root locus techniques
- Can infer performance and stability from the same plot
- Can use measured data rather than a transfer function model
- Design process can be independent of the system order
- Time delays are handled correctly
- Graphical techniques (analysis and synthesis) are quite simple

8.3 Computing Steady State Response of a Linear System to a Sinusoidal Input

The steady-state output of a transfer function system can be obtained directly from the sinusoidal transfer function. That is, the transfer function in which the Laplace variable s is replaced by $j\omega$, where ω is frequency.

Consider the stable, linear time-invariant system with transfer function $G(s)$ shown above.
The input and output of the system are denoted by $x(t)$ and $y(t)$, respectively.
If the input $x(t)$ is a sinusoidal signal, the steady-state output will also be a sinusoidal signal of the same frequency, but with possibly different magnitude and phase angle.

Let us assume that the input signal is given by

$$x(t) = X\sin(\omega t)$$

Suppose that the transfer function $G(s)$ can be written as a ratio of two polynomials in s as:

$$G(s) = \frac{p(s)}{q(s)} = \frac{p(s)}{(s+p_1)(s+p_2)\cdots(s+p_n)}$$

Without the loss of generality, it is assumed that the transfer function has only distinct poles located at $-p_1, -p_2, \cdots, -p_n$. The Laplace Transform of the output $Y(s)$ is expressed as:

$$Y(s) = G(s)X(s) = \frac{p(s)}{q(s)}X(s)$$

It is worth noting that the steady-state response of a stable, linear, time-invariant system to a sinusoidal input does not depend on the initial conditions. (Thus, we can assume the zero initial condition.)

The Laplace Transform of the input signal $x(t) = X\sin(\omega t)$ is given by

$$X(s) = \frac{\omega X}{s^2+\omega^2}$$

Hence,

$$\begin{aligned} Y(s) &= G(s)X(s) = G(s)\frac{\omega X}{s^2+\omega^2} \\ &= \frac{a}{s+j\omega} + \frac{\bar{a}}{s-j\omega} + \frac{b_1}{(s+p_1)} + \frac{b_2}{(s+p_2)} + \cdots + \frac{b_n}{(s+p_n)} \end{aligned}$$

where a, and the b_i, ($i = 1, 2, \cdots, n$ aka $\forall i \in \mathbb{N}$) are constants and $\bar{a}$ is the complex conjugate of a.

Thus, the inverse Laplace Transform ($\mathcal{L}^{-1}$) of $Y(s)$ gives the output response to a sinusoidal input

$$y(t) = ae^{-j\omega t} + \bar{a}e^{j\omega t} + b_1e^{-p_1t} + b_2e^{-p_2t} + \cdots + b_ne^{-p_nt}$$

For a stable system, the poles lie in the left half of the s-plane. Therefore, as time approaches infinity, the terms $e^{-p_1t}, e^{-p_2t}, \cdots e^{-p_nt}$ approach zero. i.e.

$$\lim_{t\to\infty} e^{-p_it} = 0, \quad \text{where,} \quad i = 1, 2, \cdots, n \equiv \forall i \in \mathbb{N}$$

Thus, all the terms on the right hand side of the above equation, except the first two, drop out at steady state.

The steady-state response of the system is therefore given by:

$$y_{ss}(t) = ae^{-j\omega t} + \bar{a}e^{j\omega t}$$

Where the constants a and $\bar{a}$ can be evaluated as:

$$\begin{aligned} a = G(s) &= \frac{\omega X}{s^2+\omega^2}(s+j\omega)\Big|_{s=-j\omega} = \frac{XG(-j\omega)}{2j} \\ \bar{a} = G(s) &= \frac{\omega X}{s^2+\omega^2}(s+j\omega)\Big|_{s=+j\omega} = \frac{XG(+j\omega)}{2j} \end{aligned}$$

Since $G(j\omega)$ is a complex quantity, it can be written in the following form:

$$G(j\omega) = |G(j\omega)|\, e^{j\phi}$$

where $|G(j\omega)|$ represents the magnitude and ϕ represents the angle of $G(j\omega)$ i.e.,

$$\phi = \angle G(j\omega) = \tan^{-1}\left[\frac{\text{Imaginary part of } G(j\omega)}{\text{Real part of } G(j\omega)}\right] \equiv \tan^{-1}\left[\frac{\Im(G(j\omega))}{\Re(G(j\omega))}\right]$$

Also,

$$G(-j\omega) = |G(-j\omega)|\, e^{-j\phi} = |G(j\omega)|\, e^{-j\phi}$$

The expression for the steady-state output can now be simplified as:

$$\begin{aligned} y_{ss}(t) &= ae^{-j\omega t} + \bar{a}e^{j\omega t} \\ &= X|G(j\omega)|\,\frac{e^{j(\omega t+\phi)} - e^{-j(\omega t+\phi)}}{2j} \\ &= X|G(j\omega)|\,\frac{\exp\big(j(\omega t+\phi)\big) - \exp\big(-j(\omega t+\phi)\big)}{2j} \\ &= X|G(j\omega)|\sin(\omega t+\phi) \\ &= Y\sin(\omega t+\phi) \end{aligned}$$

where $Y = X|G(j\omega)|$.

Thus, the input $x(t) = X\sin(\omega t)$ results in $y_{ss}(t) = Y\sin(\omega t+\phi)$. i.e.

- If a stable, linear, time-invariant system is excited by a sinusoidal input, then at stead-state, the output will be sinusoidal.
- The frequency of the output will be equal to that of the input
- But the amplitude and phase of the output will, in general, be different from those of the input.

The amplitude of the output is given by the product of that of the input and $|G(j\omega)|$, while the phase angle differs from that of the input by the amount $\phi = \angle G(j\omega)$.

An example of input and output sinusoidal signals is shown below.

For sinusoidal inputs,

$$|G(j\omega)| = \left|\frac{Y(j\omega)}{X(j\omega)}\right|, \quad G(j\omega) = \frac{Y(j\omega)}{X(j\omega)} \quad \text{and} \quad \angle G(j\omega) = \angle\frac{Y(j\omega)}{X(j\omega)}$$

where $G(j\omega)$ is called the sinusoidal transfer function.

The sinusoidal transfer function of any linear system is obtained by substituting $j\omega$ for s in the transfer function of the system.

8.3.1 Example

Consider the system shown below

The transfer function $G(s)$ is given by

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

For sinusoidal input $x(t) = X\sin(\omega t)$, the steady-state output $y_{ss}(t)$ can be found as follows:

Substituting $j\omega$ for s in $G(s)$ yields

$$G(j\omega) = \frac{K}{1 + j\omega T}$$

The amplitude ratio of the output to the input $|G(j\omega)|$ and the phase angle ϕ is given by

$$|G(j\omega)| = \frac{K}{\sqrt{1 + \omega^2 T^2}} \quad \text{and} \quad \phi = \angle G(j\omega) = -\tan^{-1}(\omega T)$$

Thus, for the input $x(t) = X\sin(\omega t)$, the steady-state output $y_{ss}(t)$ is given by

$$y_{ss}(t) = \frac{XK}{\sqrt{1 + \omega^2 T^2}} \sin\big(\omega t - \tan^{-1}(\omega T)\big)$$

Note that for small values of ω, the amplitude of the steady-state output $y_{ss}(t)$ is almost equal to K times the amplitude of the input. The phase shift of the output is small for small values of ω. For large ω, the amplitude of the output is small and almost inversely proportional to ω. The phase shift approaches $-90°$ as ω approaches infinity. This is a phase-lag network.

8.4 Frequency Response of Closed Loop Systems

The closed loop transfer function of a single loop feedback control system is expressed as:

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

Under the sinusoidal steady state, $s = j\omega$ and the above equation becomes

$$M(j\omega) = \frac{C(j\omega)}{R(j\omega)} = \frac{G(j\omega)}{1 + G(j\omega)H(j\omega)}$$

The sinusoidal steady-state transfer function $M(j\omega)$ may be expressed in terms of its magnitude and phase as:

$$M(j\omega) = |M(j\omega)| \angle M(j\omega) = M(\omega)\angle\phi_m(\omega)$$

where

$$M(\omega) = \left|\frac{G(j\omega)}{1 + G(j\omega)H(j\omega)}\right| \frac{|G(j\omega)|}{|1 + G(j\omega)H(j\omega)|}$$

$$\phi_m(j\omega) = \angle G(j\omega) - \angle 1 + G(j\omega)H(j\omega)$$

8.4.1 Significance and Desired Characteristics of $M(\omega)$

$M(\omega)$ can be regarded as a magnification of the feedback control system. It is similar to the gain or amplification of an electronic amplifier. Ideally, an amplifier must have a flat gain for all frequencies; realistically, it should have a flat gain in the audio frequency range. In control systems, the ideal design criterion is similar. However, if it is desirable to keep the output $C(j\omega)$ identical to the input $R(j\omega)$ at all frequencies, then $M(j\omega)$ must be unity for all frequencies. However, $M(j\omega)$ can be unity only when $G(j\omega)$ is infinite, while $H(j\omega)$ is finite and nonzero.

An infinite magnitude for $G(j\omega)$ is impossible to achieve in practice, nor would it be desirable because

- Most control systems become unstable when its loop gain becomes very high.
- Furthermore, all control systems are subjected to noise. Thus, in addition to responding to the input signal, the system should be able to reject and suppress noise and unwanted signals.

This means that the frequency response of a control system should have a cutoff characteristic in general, and sometimes even a band-pass characteristic.

The figure below shows the gain and phase characteristics of an **ideal** low-pass filter which has a sharp cut-off frequency at ω_c.

Typical gain and phase characteristics of a feedback control system are shown below.

Note that, the great majority of control systems have the characteristics of a low-pass filter, so the gain decreases as the frequency increases.

8.5 Frequency Domain Characteristics & Specifications

In the design of linear control systems using frequency-domain methods, it is necessary to define a set of specifications so that the performance of the system can be identified. Specifications such as the maximum overshoot, damping ratio, and the like used in the time domain can no longer be used directly in the frequency domain. The following frequency domain specifications are often used in practice.

8.5.1 Resonant Peak, M_r and Resonant Frequency, ω_r

The resonant peak M_r is the maximum value of $|M(j\omega)|$. The magnitude of M_r gives indication on the relative stability of a stable closed-loop system. Normally, M_r corresponds to a large maximum overshoot of the step response. In practice the desirable value of M_r should be between 1.1 and 1.5.

The resonant frequency, ω_r, is the frequency at which the peak resonance M_r occurs.

8.5.2 Bandwidth, BW

The bandwidth BW is the frequency at which $|M(j\omega)|$ drops to 70.7% or 3 dB down from its zero-frequency value. The bandwidth of a control system gives indication on the transient response properties in the time domain.

- A large bandwidth corresponds to a faster rise time, since higher-frequency signals are more easily passed through the system.
- Conversely, if the bandwidth is small, only signals of relatively low frequencies are passed, and the time response will be slow and sluggish.
- Thus, the **bandwidth and rise time are inversely proportional** to each other.

Bandwidth also indicates the noise-filtering characteristics and the robustness of the system. Robustness represents a measure of the sensitivity of a system to parameter variations.

8.5.3 Cutoff Rate

Often, bandwidth alone is inadequate to indicate the ability of a system in distinguishing signals from noise. Sometimes it may be necessary to look at the slope of $|M(j\omega)|$, which is called the cutoff rate of the frequency response, at high frequencies. Two systems can have the same bandwidth, but the cutoff rates may be different.

8.6 A Second Order System: M_r, ω_r, and BW

Consider the closed loop transfer function of a second-order system

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

At sinusoidal steady-state, $s = j\omega$,

$$M(j\omega) = \frac{C(j\omega)}{R(j\omega)} = \frac{\omega_n^2}{(j\omega)^2 + 2\zeta\omega_n(j\omega) + \omega_n^2} = \frac{1}{1 + j2\left(\frac{\omega}{\omega_n}\right)\zeta - \left(\frac{\omega}{\omega_n}\right)^2}$$

Let $u = \frac{\omega}{\omega_n}$. Therefore,

$$\frac{1}{1 + j2\left(\frac{\omega}{\omega_n}\right)\zeta - \left(\frac{\omega}{\omega_n}\right)^2} = \frac{1}{1 + j2(u)\zeta - (u)^2} = M(ju)$$

Hence,

$$|M(ju)| = \frac{1}{\left[(1-u^2)^2 + (2\zeta u)^2\right]^{\frac{1}{2}}}$$

$$\angle M(ju) = \phi_m(ju) = -\tan^{-1}\left(\frac{2\zeta u}{1-u^2}\right)$$

The resonant frequency is determined by setting the derivative of $|M(ju)|$ to zero

$$\frac{\mathrm{d}|M(ju)|}{\mathrm{d}u} = -\frac{1}{2}\left[(1-u^2)^2 + (2\zeta u)^2\right]^{\frac{-3}{2}}\left(4u^3 - 4u + 8u\zeta^2\right) = 0$$

Thus, $4u^3 - 4u + 8u\zeta^2 = 4u(u^2 - 1 + 2\zeta^2) = 0$

The roots of this quadratic are $u_r = 0$ and $u_r = \sqrt{1 - 2\zeta^2}$.

Note:

- The solution $u_r = 0$ indicates that the slope of the $|M(ju)|$ vs. ω curve is zero at $\omega = 0$.
- If $\zeta \leq \frac{1}{\sqrt{2}}$, then $u_r = \sqrt{1 - 2\zeta^2}$ is real.
- If $\zeta > \frac{1}{\sqrt{2}}$, then u_r is complex.

Since $u = \frac{\omega}{\omega_n}$ hence $u_r = \frac{\omega_r}{\omega_n} = \sqrt{1 - 2\zeta^2}$, hence

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2} \quad \text{and} \quad M_r = \frac{1}{2\zeta\sqrt{1 - \zeta^2}}, \quad \zeta \leq \frac{1}{\sqrt{2}}$$

Note:

- Since the frequency, ω, is real, $\omega_r = \omega_n\sqrt{1 - 2\zeta^2}$ is meaningful only if $\zeta \leq \frac{1}{\sqrt{2}}$.
- If $\zeta \geq \frac{1}{\sqrt{2}}$, then $\omega_r = 0$ and $M_r = 1$.

It is important to note that, for the prototype second-order system, M_r is a function of the damping ratio ζ only, and $omega_r$ is a function of both ζ and ω_n. Furthermore, although taking the derivative of $|M(ju)|$ with respect to u is a valid method of determining M_r and ω_r, for higher-order systems, this analytical method is quite tedious and is not recommended.

By definition, the bandwidth is determined from $|M(ju)| = \frac{1}{\sqrt{2}}$. Thus

$$|M(ju)| = \frac{1}{\left[(1-u^2)^2 + (2\zeta u)^2\right]^{\frac{1}{2}}} = \frac{1}{\sqrt{2}}$$

Then,

$$\left[(1-u^2)^2 + (2\zeta u)^2\right]^{\frac{1}{2}} = \sqrt{2}$$

Hence,

$$u^2 = (1-\zeta^2) \pm \sqrt{4\zeta^4 - 4\zeta^2 + 2}$$

Since u must be real and positive, the bandwidth of the prototype second-order system is given by

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

The maximum overshoot of the unit step response in the time domain depends on ζ only.

$$M_p = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}$$

The resonance peak of the closed-loop frequency response M_r depends upon ζ only.

$$M_r = \frac{1}{2\zeta\sqrt{1-\zeta^2}}$$

The rise time increases with ζ, and the bandwidth decreases with the increase of ζ, for a fixed ω_n.

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

Therefore, bandwidth and rise time are inversely proportional to each other.

$$t_r \approx \frac{1 + 1.1\zeta + 1.4\zeta^2}{\omega_n}$$

Bandwidth is directly proportional to ω_n. Higher bandwidth corresponds to larger M_r.

Chapter 9

Frequency Domain Analysis of Linear Systems - Bode Plots

Learning Outcomes

After completing this module, you should be able to:

- Plot Bode Magnitude and Phase Plots
- Determine the stability information about the system from gain and phase margins
- Do Relative stability analysis.

9.1 Bode Plot (Corner Plot) of a Transfer Function

The Bode plot of the function $G(j\omega)$ is comprised of two plots

1. Amplitude Plot: Plot of the amplitude of $G(j\omega)$ in decibels (dB) versus $\log\omega$ or ω and
2. Phase Plot: Plot of the phase of $G(j\omega)$ in degrees as a function of $\log\omega$ or ω. Note: This is $\log(\cdots)$ and not $\ln(\cdots)$ i.e. $\log_e(\cdots)$.

A bode plot is also known as a corner plot or an asymptotic plot of $G(j\omega)$. These names stem from the fact that the Bode plot can be constructed by using straight-line approximations that are asymptotic to the actual plot.

Advantages of the Bode Plot

1. The magnitude of $G(j\omega)$ in the Bode plot is expressed in decibels, the product and division factors in $G(j\omega)$ become additions and subtractions, respectively. The phase relations are also added and subtracted from each other in a natural way.
2. The magnitude plot of the Bode plots of most $G(j\omega)$ functions encountered in control systems may be approximated by straight-line segments. This makes the construction of the Bode plot very simple.

9.1.1 Pole-Zero vs. Bode Notation

Consider the function

$$G(s) = \frac{K_1(s+z_1)(s+z_2)\cdots(s+z_m)}{s^N(s+p_1)(s+p_2)\cdots(s+p_n)}e^{-T_d s}$$

where K_1 and T_d are real constants, and the z's and the p's may be real or complex (in conjugate pairs) numbers.

Note: The equation above is the preferred form for root-locus construction, because the poles and zeros of $G(s)$ are easily identified.

However, for constructing the Bode plot manually, $G(s)$ is preferably written in the following form:

$$G(s) = \frac{K(1+T_1 s)(1+T_2 s)\cdots(1+T_m s)}{s^N(1+T_a s)(1+T_b s)\cdots(1+T_n s)}e^{-T_d s}$$

where K is a real constant, the T's may be real or complex (in conjugate pairs) numbers and T_d is the real time delay.

9.1.1.1 Example

The transfer function

$$G(s) = \frac{150(s+2)}{s^2(s+3)(s+5)(s+10)}$$

Can be represented as

$$\begin{aligned} G(s) &= \frac{150 \cdot 2(1+0.5s)}{s^2 \cdot 3 \cdot 5 \cdot 10(1+0.333s)(1+0.2s)(1+0.1s)} \\ &= \frac{2(1+0.5s)}{s^2(1+0.333s)(1+0.2s)(1+0.1s)} \end{aligned}$$

9.1.2 Magnitude & Phase Angle

The magnitude of $G(s)$ in dB and its phase angle $\phi(\omega)$ can be expressed as

$$\begin{aligned} |G(j\omega)|_{dB} &= 20\log_{10}|G(j\omega)| \\ &= 20\log_{10}|K| + 20\log_{10}|1+j\omega T_1| + \cdots + 20\log_{10}|1+j\omega T_m| \\ &\quad - 20\log_{10}|j\omega| - 20\log_{10}|1+j\omega T_a| - \cdots - 20\log_{10}|1+j\omega T_n| \end{aligned}$$

The phase of $G(j\omega)$ is

$$\begin{aligned} \phi(\omega) = \angle G(j\omega) &= \angle K + \angle(1+j\omega T_1) + \angle(1+j\omega T_2) + \cdots + \angle(1+j\omega T_m) \\ &\quad - 90°N - \angle(1+j\omega T_a) - \angle(1+j\omega T_b) - \cdots - \angle(1+j\omega T_n) \end{aligned}$$

Note: The magnitude and phase angle of the complex factor $1+j\omega T$ is given as:

$$|1+j\omega T| = \sqrt{1+\omega^2T^2} \quad \text{and} \quad \angle(1+j\omega T) = \tan^{-1}(\omega T)$$

9.1.3 Common Factor Types

$G(j\omega)$ can contain just five simple types of factors:

1. Constant factor: K
2. Poles or zeros at the origin of order N: $(j\omega)^{\pm N}$
3. Poles or zeros at $s = -\dfrac{1}{T}$ of order q: $(1+j\omega T)^{\pm q}$
4. Complex poles and zeros of order r: $\left(1 + 2\zeta\dfrac{\omega}{\omega_n} - \dfrac{\omega^2}{\omega_n^2}\right)^{\pm r}$
5. Pure time delay $e^{-j\omega T_d}$

where T_d, N, q, and r are positive integers.

Once we become familiar with the logarithmic plots of these basic factors, it is possible to utilise them in constructing a composite logarithmic plot for any general form of $G(j\omega)H(j\omega)$ by sketching the curves for each factor and adding or subtracting the individual curves graphically, because adding the logarithms of the gains corresponds to multiplying them together.

9.2 Bode Plot of a Pure Constant

Here, $G(s) = K$

The magnitude of $G(j\omega)$ in dB is given by

$$|G(j\omega)|_{dB} = 20\log_{10}|G(j\omega)| = 20\log_{10}|K|$$

The phase angle of $G(j\omega)$ is given by

$$\angle G(j\omega) = \angle K = \begin{cases} 0^\circ & K > 0 \\ -180^\circ & K < 0 \end{cases}$$

9.2.1 Examples

9.2.1.1 Example 1

Let $G(s) = 5$

$$\begin{aligned} |G(j\omega)| &= 5 \\ |G(j\omega)|_{dB} &= 20\log_{10}|5| = 13.9\,\text{dB} \\ \angle G(j\omega) &= \angle 5 = 0^\circ \end{aligned}$$

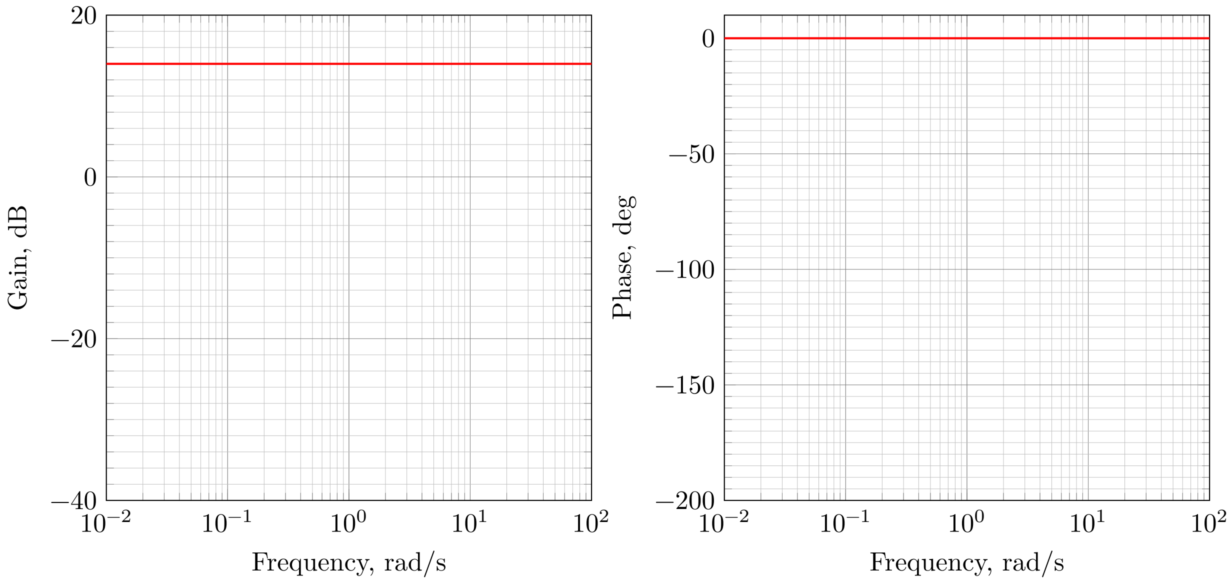

Figure 9.1: Bode Plot of $G(s) = 5$

9.2.1.2 Example 2

Let $G(s) = -5$

$$\begin{aligned}
|G(j\omega)| &= |-5| \\
|G(j\omega)|_{dB} &= 20\log_{10}|-5| = 13.9\,\mathrm{dB} \\
\angle G(j\omega) &= \angle -5 = -90°
\end{aligned}$$

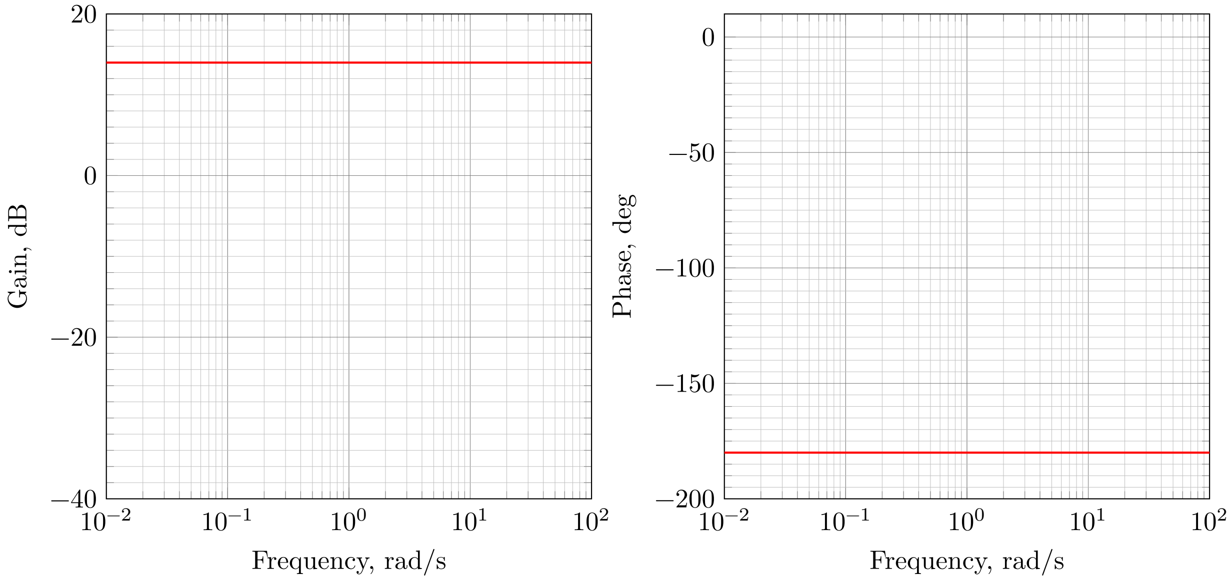

Figure 9.2: Bode Plot of $G(s) = -5$

9.3 Bode Plot of a Differentiator

Here, $G(s) = s$ The magnitude of $G(j\omega) = j\omega$ in dB is given by

$$\begin{aligned}|G(j\omega)|_{dB} &= 20\log_{10}|G(j\omega)| \\ &= 20\log_{10}|j\omega| = 20\log_{10}(\omega) \text{ dB}\end{aligned}$$

The log-magnitude curve is a straight line with a slope of 20 dB/decade.

The phase angle of $G(j\omega)$ is given by

$$\angle G(j\omega) = \angle j\omega = 90°$$

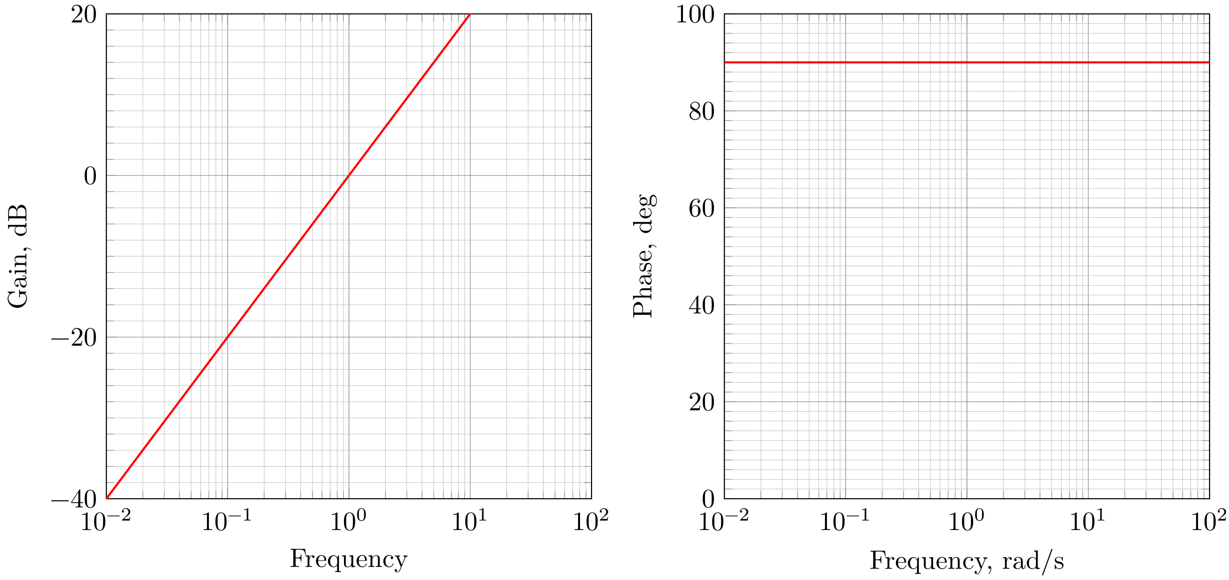

Figure 9.3: Bode Plot of $G(s) = s$

9.4 Bode Plot of an Integrator

Here, $G(s) = \dfrac{1}{s}$ The magnitude of $G(j\omega)$ in dB is given by

$$\begin{aligned} |G(j\omega)|_{dB} &= 20\log_{10}|G(j\omega)| \\ &= 20\log_{10}\left|\frac{1}{j\omega}\right| = 20\log_{10}\left(\frac{1}{\omega}\right) \\ &= -20\log_{10}(\omega)\ \text{dB} \end{aligned}$$

The log-magnitude curve is a straight line with a slope of $-20\,\text{dB/decade}$.

The phase angle of $G(j\omega) = \dfrac{1}{j\omega}$ is constant and equal to $-90°$.

$$\angle G(j\omega) = \angle\frac{1}{j\omega} = -90°$$

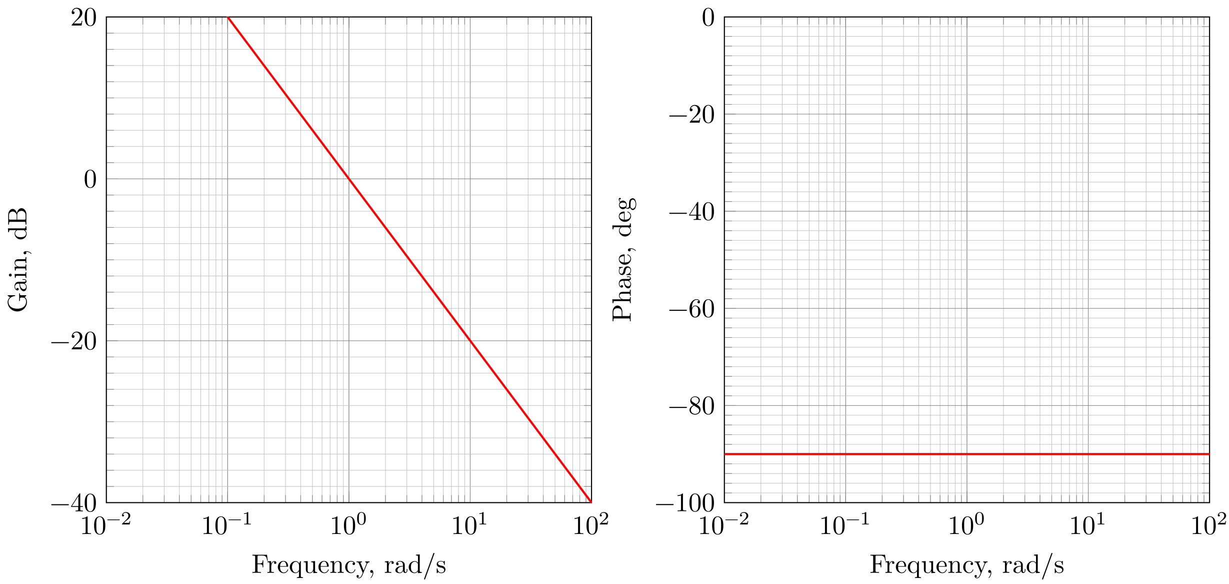

Figure 9.4: Bode Plot of $G(s) = \dfrac{1}{s}$

9.5 Bode Plot of a Simple Zero $1 + sT$

Here, $G(s) = 1 + sT$, where T is a positive real constant.

9.5.1 Magnitude

The magnitude of $G(j\omega)$ is given by

$$|G(j\omega)|_{\text{dB}} = 20\log_{10}|G(j\omega)| = 20\log_{10}\sqrt{1+\omega^2T^2}$$

To obtain asymptotic approximations of $|G(j\omega)|_{\text{dB}}$, we consider both very large and very small values of ω.

- At very low frequencies where $\omega T \ll 1$, ω^2T^2 is very very small compared to 1 and can be neglected. Hence
$$|G(j\omega)|_{\text{dB}} \cong 20\log_{10} 1 = 0\,\text{dB}$$
- At very high frequencies where $\omega T \gg 1$, we can approximate $1+\omega^2T^2$ by ω^2T^2. Thus,
$$|G(j\omega)|_{\text{dB}} \cong 20\log_{10}\left(\sqrt{\omega^2T^2}\right) \cong 20\log_{10}(\omega T)$$

Thus, the Bode magnitude plot at the low frequencies (i.e. $\omega T \ll 1$), is the constant 0 dB line.

The Bode magnitude plot at high frequencies (i.e. $\omega T \gg 1$), is a straight line with a slope of +20 dB/decade of frequency.

The frequency at which the two straight lines (called asymptotes) meet is called the *corner frequency* or *break frequency* and is given by

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

The corner frequency divides the frequency-response curve into two regions: a curve for the low-frequency region and a curve for the high-frequency region.

Note: The corner frequency is very important in sketching logarithmic frequency-response curves.

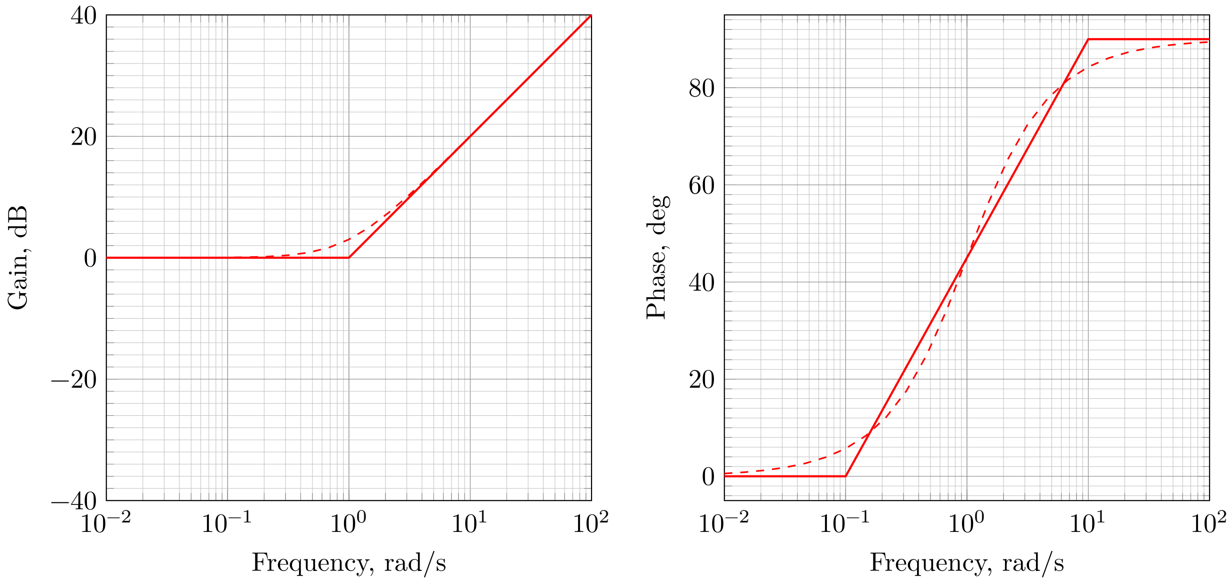

Figure 9.5: Bode Plot of $G(s) = 1 + sT$

9.5.2 Phase Angle

The exact phase angle ϕ of the factor $1 + j\omega T$ is: $\phi = \tan^{-1} \omega T$.

At zero frequency, the phase angle is 0°.

At the corner frequency, the phase angle is: $\phi = \tan^{-1} \frac{T}{T} = \tan^{-1} 1 = +45°$.

At infinity, the phase angle becomes 90°. Since the phase angle is given by an inverse tangent function, the phase angle is skew symmetric about the inflection point at $\phi = +45°$.

Similar to the magnitude curve, a straight-line approximation can be made for the phase curve.

Because the phase varies from 0° to 90°, we can draw a line from 0° at 1 decade below the corner frequency to 90° at 1 decade above the corner frequency.

The maximum deviation between the straight-line approximation and the actual curve is less than 6°.

9.6 Bode Plot of a Simple Pole $\dfrac{1}{1+sT}$

Here $G(s) = \dfrac{1}{1+sT}$, where T is a positive real constant.

9.6.1 Magnitude

The magnitude of $G(j\omega)$ is given by

$$|G(j\omega)|_{\text{dB}} = 20\log_{10}|G(j\omega)| = -20\log_{10}\sqrt{1+\omega^2T^2}$$

To obtain asymptotic approximations of $|G(j\omega)|_{\text{dB}}$, we consider both very large and very small values of ω.

- At very low frequencies where $\omega T \ll 1$, ω^2T^2 is very very small compared to 1 and can be neglected. Hence

$$|G(j\omega)|_{\text{dB}} \cong 20\log_{10} 1 = 0\,\text{dB}$$

- At very high frequencies where $\omega T \gg 1$, we can approximate $1+\omega^2T^2$ by ω^2T^2. Thus,

$$|G(j\omega)|_{\text{dB}} \cong -20\log_{10}\left(\sqrt{\omega^2T^2}\right) \cong -20\log_{10}(\omega T)$$

Thus, the Bode magnitude plot at the low frequencies (i.e. $\omega T \ll 1$), is the constant 0 dB line.

The Bode magnitude plot at high frequencies (i.e. $\omega T \gg 1$), is a straight line with a slope of -20 dB/decade of frequency.

The frequency at which the two straight lines (called asymptotes) meet is called the *corner frequency* or *break frequency* and is given by

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

The corner frequency divides the frequency-response curve into two regions: a curve for the low-frequency region and a curve for the high-frequency region.

Note: The corner frequency is very important in sketching logarithmic frequency-response curves.

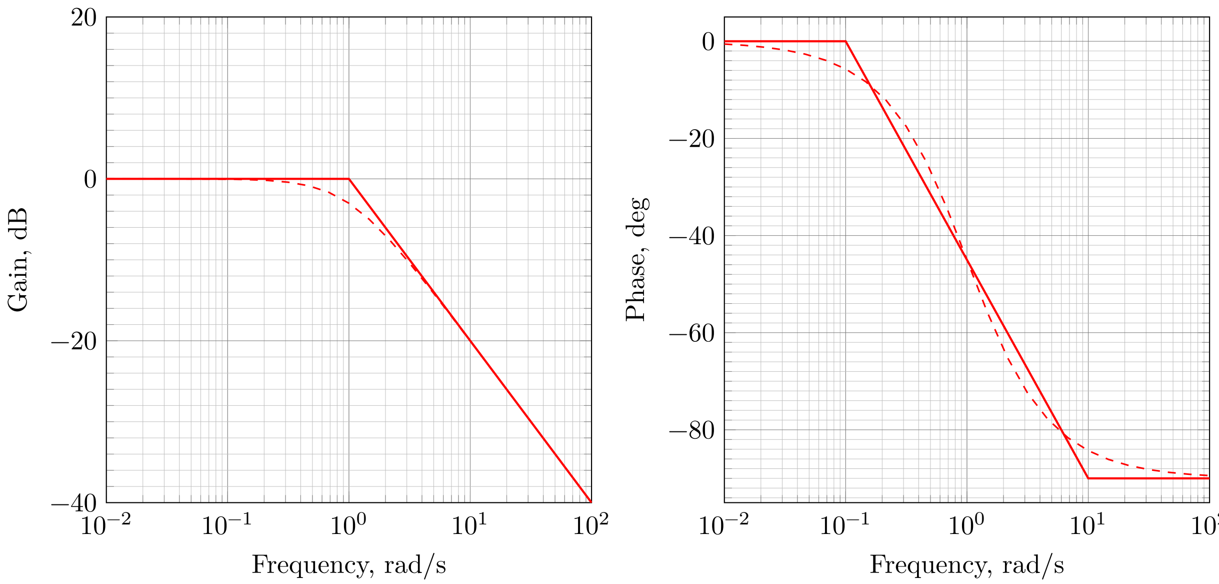

Figure 9.6: Bode Plot of $G(s) = \dfrac{1}{1+sT}$

9.6.2 Phase Angle

The exact phase angle ϕ of the factor $\dfrac{1}{1+j\omega T}$ is: $\phi = -\tan^{-1}\omega T$.

At zero frequency, the phase angle is 0°.

At the corner frequency, the phase angle is: $\phi = -\tan^{-1}\dfrac{T}{T} = -\tan^{-1}(1) = -45°$.

At infinity, the phase angle becomes $-90°$. Since the phase angle is given by an inverse tangent function, the phase angle is skew symmetric about the inflection point at $\phi = -45°$.

Similar to the magnitude curve, a straight-line approximation can be made for the phase curve.

Because the phase varies from 0° to $-90°$, we can draw a line from 0° at 1 decade below the corner frequency to $-90°$ at 1 decade above the corner frequency.

The maximum deviation between the straight-line approximation and the actual curve is less than 6°.

9.7 Advantages of Approximate Bode Plot using Asymptotes

The asymptotes are quite easy to draw and are sufficiently close to the exact curve.

The use of such approximations in drawing Bode diagrams is convenient in establishing the general nature of the frequency-response characteristics quickly with a minimum amount of calculation and may be used for most preliminary design work.

The actual $|G(j\omega)|$ dB plot is a smooth curve and deviates only slightly from the straight line approximation.

If accurate frequency-response curves are desired, corrections may easily be made.

The error between the actual magnitude curve and the straight-line asymptotes is symmetrical with respect to the corner frequency $\omega = \frac{1}{T}$.

In practice, an accurate frequency-response curve can be drawn by introducing a correction of 3 dB at the corner frequency and a correction of 1 dB at points one octave below and above the corner frequency and then connecting these points by a smooth curve.

9.8 Bode Plot of Quadratic Factors

Consider the function

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

We are interested only in the case when $\zeta < 1$, because otherwise $G(s)$ would have two unequal real poles, and the Bode plot can be obtained by considering $G(s)$ as the product of two transfer functions with simple poles. Now by letting $s = j\omega$, the above function can be written as

$$\begin{aligned} G(j\omega) &= \frac{1}{1 + \left(\frac{2\zeta}{\omega_n}\right) j\omega + \left(\frac{1}{\omega_n^2}\right) (j\omega)^2} = \frac{1}{1 + j\left(\frac{2\zeta}{\omega_n}\right) \omega - \left(\frac{1}{\omega_n^2}\right) \omega^2} \\ &= \frac{1}{\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right] + j2\zeta\left(\frac{\omega}{\omega_n}\right)} \end{aligned}$$

9.8.1 Magnitude

The magnitude of $G(j\omega)$ in dB is given by

$$\begin{aligned} |G(j\omega)|_{\text{dB}} &= 20\log_{10}|G(j\omega)| = 20\log_{10} \frac{1}{\sqrt{\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right]^2 + 4\zeta^2\left(\frac{\omega}{\omega_n}\right)^2}} \\ &= -20\log_{10}\sqrt{\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right]^2 + 4\zeta^2\left(\frac{\omega}{\omega_n}\right)^2} \end{aligned}$$

Asymptotic approximations to the frequency-response curves are not accurate for a factor with low values of ζ.

This is because the magnitude and phase of the quadratic factor depend on both the corner frequency and the damping ratio ζ.

To obtain the asymptotic frequency-response curve, we consider the two cases as follows:

1. At very low frequencies such that $\omega \ll \omega_n$, i.e. $\left(\frac{\omega}{\omega_n}\right) \ll 1$, then the log magnitude $|G(j\omega)|_{\text{dB}}$ becomes:
$$|G(j\omega)|_{\text{dB}} \cong -20\log_{10}(1) = 0\,\text{dB}$$
The low-frequency asymptote is thus a horizontal line at 0 dB.

2. At very high frequencies such that $\omega \gg \omega_n$, i.e. $\left(\frac{\omega}{\omega_n}\right) \gg 1$, then the log magnitude $|G(j\omega)|_{\text{dB}}$ becomes:
$$|G(j\omega)|_{\text{dB}} \cong -20\log_{10}\left(\frac{\omega^2}{\omega_n^2}\right) = -40\log_{10}\left(\frac{\omega}{\omega_n}\right)$$
The equation for the high-frequency asymptote is a straight line having the slope -40 dB/decade.

The high-frequency asymptote intersects the low-frequency asymptote at $\omega = \omega_n$.

This frequency, ω_n, is the corner frequency for the quadratic factor considered.

Note that the two asymptotes just derived are independent of the value of ζ.

Near the frequency $\omega = \omega_n$, a resonant peak occurs.

The damping ratio ζ determines the magnitude of this resonant peak

There exist errors in the approximation by straight-line asymptotes.

The actual magnitude curve of $G(j\omega)$ differs strikingly from the asymptotic curve. Because, the amplitude and phase curves of the second-order $G(j\omega)$ depend on not only the corner frequency, ω_n, but also on the damping ratio, ζ, which does not enter the asymptotic curve.

The magnitude of the error depends on the value of ζ. It is large for small values of ζ.

9.8.2 Phase Angle

The phase angle of the quadratic factor $\left[1 + 2\zeta\left(\frac{j\omega}{\omega_n}\right) + \left(\frac{j\omega}{\omega_n}\right)^2\right]^{-1}$ is

$$\phi = \angle \frac{1}{1 + 2\zeta\left(\frac{j\omega}{\omega_n}\right) + \left(\frac{j\omega}{\omega_n}\right)^2} = -\tan^{-1}\left[\frac{2\zeta\frac{\omega}{\omega_n}}{1 - \left(\frac{\omega}{\omega_n}\right)^2}\right]$$

The phase angle is a function of both ω and ζ.

At $\zeta = 0$, the phase angle equals $0°$.

At the corner frequency, $\omega = \omega_n$, the phase angle is $-90°$, regardless of ζ because

$$\phi = -\tan^{-1}\left(\frac{2\zeta}{0}\right) = -\tan^{-1}\infty = -90°$$

i.e. Division by zero leads to tangent of ∞.

The Bode plot of the phase angle of the quadratic factor is shown in the figure below.

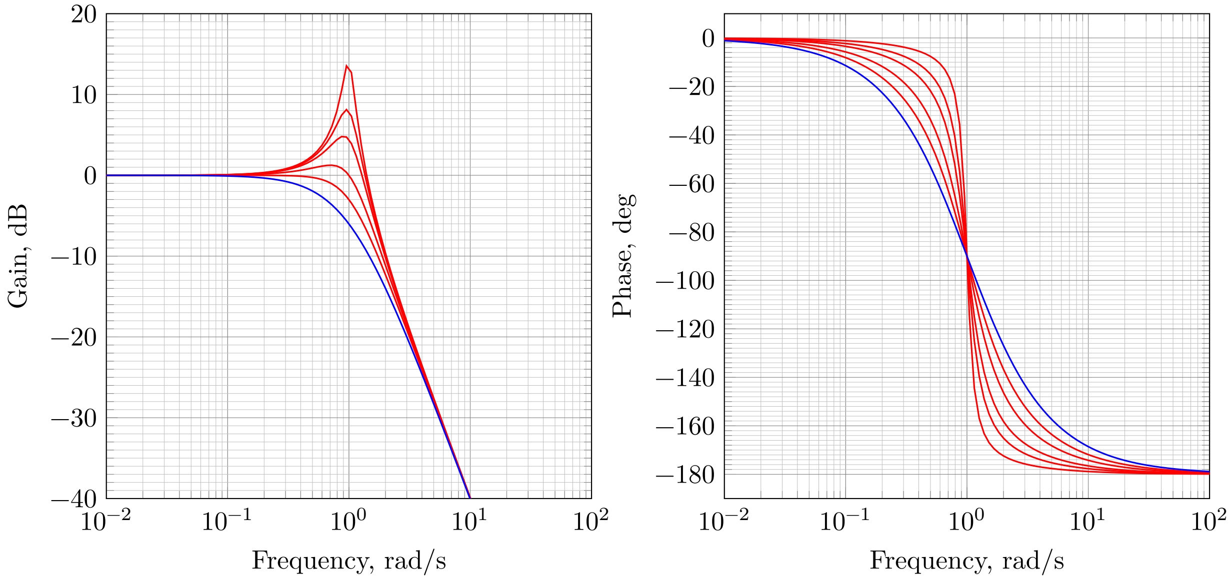

Figure 9.7: Bode Plot of $G(s) = \dfrac{\omega^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$

9.9 Rules for Drawing Bode Magnitude Plots with Simple Poles and Zeros

- First determine all the break points (pole and zero locations) and arrange in order of increasing frequency.
- Choose a frequency range for the plot that encompasses all these points, adding an extra decade of frequency above and below this range.

- Based on the poles and zeroes, make a quick sketch of the expected shape of the Bode plot. This will help you find the appropriate vertical scales.
- For a simple pole or zero of the form $(s + a)$, the slope of the uncorrected Bode plot changes at the break point $\omega = a$, increasing by 20 dB/decade for a zero and decreasing by 20 dB/decade for a pole.
- For a repeated pole or zero $(s + a)^r$, the slop changes by $r \times 20$ dB/decade, or 20 dB for each time the pole or zero is repeated.
- To find a reference level we first find the behaviour of the function for low frequencies $\omega \to 0$ or high frequencies $\omega \to \infty$. If the limiting behaviour approaches a constant value at these extremes, that is a good starting point. Otherwise we must evaluate the function numerically at some particular frequency, preferably in a region with a constant-value plateau.
- Once the uncorrected Bode plot is finished, a corrected version can be drawn. For simple/repeated roots the true response passes through a point that is $r \times 3$ dB below the uncorrected curve at the break point, or 3 dB for each time the pole or zero is repeated.

9.10 Relative Stability Analysis Using Bode Plots

Motivation: Stability is paramount.

- In practice it is not enough that a system is stable due to modelling uncertainties
- There must also be some margins of stability that describe how stable the system is
- We shall concentrate on minimum phase systems
- In time-domain, the closer the dominant closed-loop poles are to the imaginary axis, the poorer the system's relative stability.

9.11 Measures of Relative Stability

Gain Crossover Frequency

The frequency at which the magnitude of the open-loop transfer function gain ($|G(j\omega)|$) is unity or 0 dB is called the *Gain Cross Over Frequency.* It is denoted by ω_{gc}.

Phase Crossover Frequency

The frequency at which the phase angle of the open loop transfer function ($\angle G(j\omega)$) equals $-180°$ is called the *Phase Cross Over Frequency.* It is denoted by ω_{pc}.

Gain Margin

It is the amount of gain that can be allowed to increase – in the loop – at the phase crossover frequency, before the closed-loop system reaches instability.

In other words, the gain margin is the reciprocal of the magnitude $|G(j\omega)|$ at the frequency at which the phase angle is $-180°$, i.e. at the phase crossover frequency, ω_{pc}. If K_g denotes the gain margin of the system. Then

$$K_g = \frac{1}{\left|G(j\omega_{pc})\right|}, \quad K_g\text{dB} = 20\log_{10} K_g = -20\log_{10}\left|G(j\omega_{pc})\right|$$

The gain margin expressed in decibels, is positive if K_g is greater than unity, and negative if K_g is smaller than unity.

Thus, a positive gain margin (in decibels) means that the system is stable, and a negative gain margin (in decibels) means that the system is unstable.

For a stable minimum-phase system, the gain margin indicates how much the gain can be increased before the system becomes unstable. For an unstable system, the gain margin is indicative of how much the gain must be decreased to make the system stable.

Phase Margin

The phase margin is the amount of additional phase lag at the gain crossover frequency required to bring the system to the verge of instability.

The phase margin γ is 180° plus the phase angle ϕ_{gc} (in degrees) of the open-loop transfer function at the gain crossover frequency, or

$$\gamma = 180 + \phi_{gc}$$

9.12 Frequency Domain Stability: Summary

The stability of the control system based on the relation between the phase cross over frequency and the gain cross over frequency is listed below

1. If the phase cross over frequency ω_{pc} is greater than the gain cross over frequency ω_{gc}, (i.e. $\omega_{pc} > \omega_{gc}$), the control system is stable.
2. If the phase cross over frequency ω_{pc} is equal to the gain cross over frequency ω_{gc}, (i.e. $\omega_{pc} = \omega_{gc}$), the control system is marginally stable.
3. If the phase cross over frequency ω_{pc} is less than the gain cross over frequency ω_{gc}, (i.e. $\omega_{pc} < \omega_{gc}$), the control system is unstable.

The stability of the control system based on the relation between gain margin and phase margin is listed below.

1. If both the gain margin, GM, amd the phase margin, PM, are positive, then the control system is stable.
2. If both the gain margin, GM, and the phase margin, PM, are equal to zero, then the control system is marginally stable.
3. If the gain margin, GM, and/or the phase margin, PM, are/is negative, then the control system is unstable.