UNIT 1

EE 387: CLASSICAL CONTROL SYSTEMS

INTRODUCTION TO UNIT ONE

A system is a collection of matter, parts or components which are included inside a specified boundary. The boundary serves as a separation between the system and its surroundings. The effects which originate outside the system and act directly on it and are not affected by changes in the system are known as input of the system. In this unit, we shall find out what a control system is. We shall examine some practical control systems and finally consider some mathematical/computational tools used in the design and analysis of control systems.

Learning Objectives:

1. To get a basic understanding of what control systems are, to know the basic types of control systems, their components and characteristics.
2. To recognize and appreciate the importance of control systems in real life processes.
3. To get a working knowledge of the Laplace transform- an essential tool in control system modeling and analysis.
4. To understand block diagrams, transfer functions and signal flow graph descriptions of control systems.

SECTION 1-1: DESCRIPITON OF A CONTROL SYSTEM

1-1.1: What is a Control System?

A control system is basically an interconnection of various components in a particular configuration to provide a desired system response or output. The system controls the variable output to the desired value by applying proper input or controlling signal to the system input terminals. The input-output relationship of the system represents the cause and effect relationship of the system and mathematically represents the process in the system by which the input signal through system parameters controls the output signal to have desired output. The basic illustration of a control system is shown below:

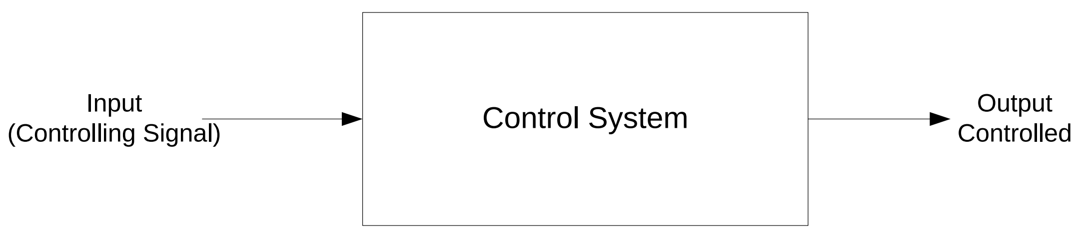


Figure 1.1: Basic Control System

1-1.2 Practical Applications of Control Systems

- In the process industries control is used to regulate level, pressure and temperature of chemical refinery vessels.
- In a steel rolling mill, the position of the rolls is controlled according to the measured thickness of the steel coming off the finishing line.

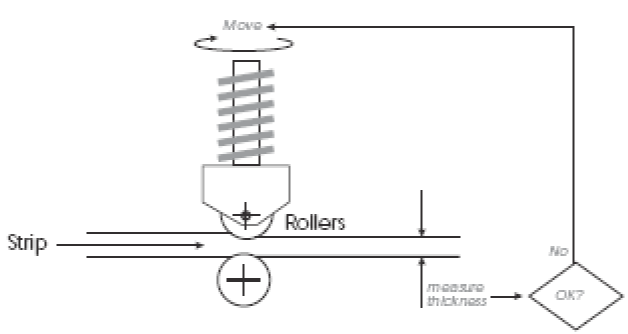


Figure 1.2: Thickness Control in Steel Rolling Mill

Control systems also find application in many devices used in the home. The following are some examples.

- **CD Players:** the position of the laser spot in relation to the microscopic pits in a Compact Disc is controlled
- **Video Recorders:** the tracking of the record and play-back heads is controlled by controlling the velocity of the tape
- **Central heating systems** use thermostats to measure and control the temperature in the room
- **Washing machines** use sequencing controls to provide a variety of wash cycles and temperature controls to avoid damage to delicate fabrics

Control systems are also employed in controlled furnace heating systems, antenna servo systems, temperature control systems etc.

1-1.3 Describing Control Systems

Input and Output and of a Control System

Using a lift-car as an example,the fourth-floor button is pressed on the ground floor. The lift-car rises to the fourth floor with a speed and floor levelling accuracy designed for passenger comfort.

The fourth floor button is the input shown by a step command. The lift does not mimic the input — this would be undesirable for passenger comfort as well as impossible with finite power supplied by motor. Instead, the input represents the position we would like the lift to be in when the lift has stopped moving. The lift itself follows the *lift response curve*. Two factors make the output different from the input.

First consider the instantaneous change in the input against the gradual change in the output. Physical entities cannot change their position or velocity instantaneously. The state changes through a path dictated by the physical devices and the way it aquires and dissipates energy. The lift undergoes a gradual change as it moves from the ground to fourth floor — called the *transient response*.

After the transient response is complete, the physical system approaches its steady-state response which is an approximation to the commanded or *desired response*. This occurs when the lift reaches the fourth floor. The accuracy of the lift's final level is the second factor that makes the output different from the input. The difference is called the *steady-state error*.

Steady-state error may also be a feature of the system being controlled and it is one of the features that the control engineer considers when specifying the desired behaviour. For example, when tracking a satellite, some error may be tolerated provided that the satellite stays close to the centre of the tracking radar beam. However, if a robot is inserting a chip into a PCB the steady-state error must be zero.

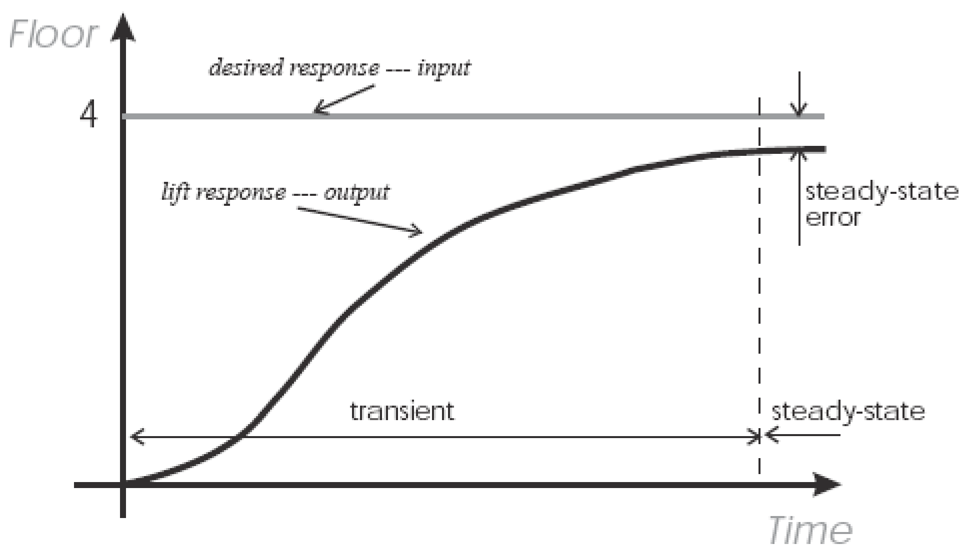


Fig 1.3: Lift Response Curve

1-1.4 Advantages of Control Systems

We tolerate the differences between desired response and actual response because of the many advantages of control systems.

- Power Amplification: Satellite dish can be positioned by a low power knob at the input but requires large power to rotate. Power gain is one good reason for building control systems.

- Dangerous Applications — remote control of a robot arm for handling nuclear material.

- Compensation for Human Deficiencies — e.g. to help handicapped people or the exo-skeleton used by Ripley in *Aliens*.
- Convenience by change of form of Input — Temperature control is by the position of a dial on a thermostat, output is heat.
- Compensation of Disturbances — Typical control variables are temperature, position and velocity, voltage, current or frequency. The control system must yield the correct output even in the presence of disturbances.

1-1.5 Types of Control Systems

There are two types of control systems namely:

- Open Loop Control Systems
- Closed-loop Control System

1-1.5.1 Open Loop Control System

An open loop system is that system in which the control action is independent of the system output. Thus an open loop system utilizes a controller to control the system process in such a way as to obtain the desired output without considering the actual system output as shown below.

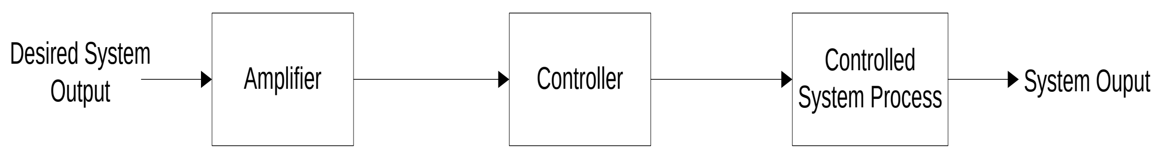


Fig 1.4: Open Loop Control System

This type of system is sensitive to surrounding conditions like vibrations, voltages, aging, etc:

ADVANTAGES OF OPEN-LOOP SYSTEM

- Relatively simple, resulting in cost, reliability and maintainability advantages.
- Inherently stable.

DISADVANTAGES OF OPEN-LOOP SYSTEM

- Relatively slow in response to demanded changes.
- Inaccurate, due to lack of corrective action for error (departure of actual value from desired value)

Examples are the control system in a microwave oven, automatic toaster system, traffic control system etc.

1-1.5.2 Closed-loop Control System

A closed-loop control system is that system in which the control action is somehow dependent upon the system output. Thus a closed-loop control system measures the actual system output, compares it with the input and determines the error which is then used for controlling the system output to have the desired value. The block diagram of a closed-loop control system is shown below:

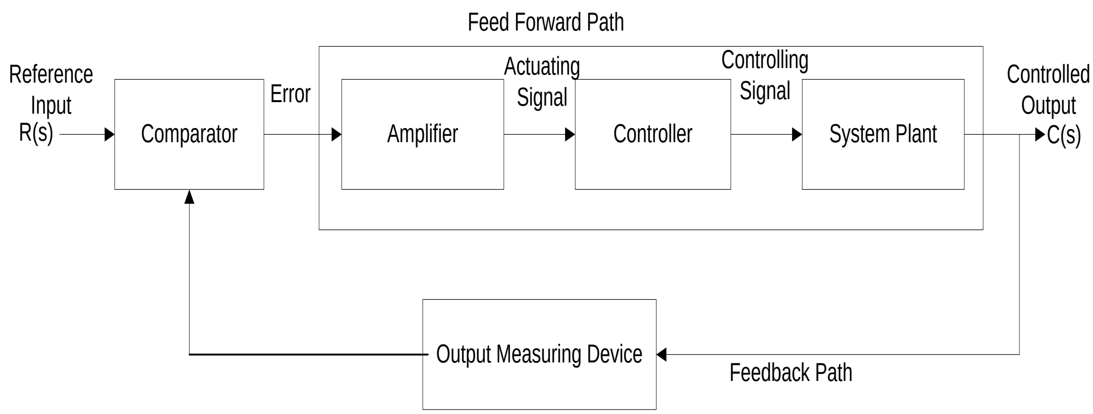


Fig 1.5: Closed-loop Control System

ADVANTAGES OF CLOSED-LOOP SYSTEM

- Relatively fast in response to demanded changes.
- Relatively accurate in matching actual to desired value.
- Relatively independent of operating conditions
- Transient performance and steady-state errors can be controlled more conveniently and with a greater degree of flexibility than with open loop systems- often by simple adjustment gains in the loop and by redesign of the controller (called the compensation).

DISADVANTAGES OF CLOSED-LOOP SYSTEM

- Relatively complex, and therefore more expensive than open-loop systems
- Potentially unstable under fault condition.

A familiar example of a closed-loop control system is thermostatic control of room temperature. The desired temperature as set on the thermostat is the reference; the actual temperature is recorded on a thermometer. When the room temperature falls below the reference temperature the thermostat is activated, a valve in the furnace is opened and heat is sent into the room until the desired temperature is reached.

SECTION1-2: ANALYTICAL TOOLS USED IN CONTROL SYSTEMS

1-2.1 The Laplace Transform

The Laplace transform of a general signal *x(t)* is defined as

$$X(s) = \int_{-\infty}^{+\infty} x(t)e^{-st}dt$$

and we note in particular that it is a function of the independent variable *s* corresponding to the complex variable in the exponent of e^{-st}. The complex variable *s* can be written as $s = (\sigma + j\omega)$, with σ and ω the real and imaginary parts, respectively.

The region of the complex variable s for which the integral converges is called the existence region or *the region of convergence* and is designated *ROC*

The Transform defined above is called *Bilateral Laplace Transform* to distinguish it from the *Unilateral Laplace Transform.* The bilateral transform involves an integration from - ∞ to + ∞, while the unilateral transform has limits of integration from 0 to + ∞.

Laplace transform is used in the modeling of linear time-invariant analog system as a transfer function. It may also be used to solve for the response in this type of system.

$x(t)$ is Laplace transformable if it satisfies the following two conditions:

i. $x(t)$ exists and is single valued every where for $0 \le t$ with t as a real variable.

ii. The value of integral $\int x(t)e^{-\sigma t}dt$ for function $x(t)$ is less than infinity for any real value of σ.

Example

Let the signal $x(t) = e^{-at}u(t)$. The Laplace transform is

$$X(s) = \int_{-\infty}^{\infty} e^{-at}u(t)e^{-st}dt = \int_{0}^{\infty} e^{-(s+a)t}dt,$$

Or, with $s = (\sigma + j\omega)$,

$$X(\sigma + j\omega) = \int_{0}^{\infty} e^{-(\sigma+a)t}e^{-j\omega t}dt$$

Thus

$$X(\sigma + j\omega) = \frac{1}{(\sigma + a) + j\omega}, \qquad \sigma + a > 0,$$

Or equivalently, since $s = (\sigma + j\omega)$ and $\sigma = \mathrm{Re}\{s\}$,

$$X(s) = \frac{1}{s+a}, \quad \mathrm{Re}\{s\} > -a.$$

That is,

$$e^{-at}u(t) \overset{L}{\leftrightarrow} \frac{1}{s+a}, \qquad \mathrm{Re}\{s\} > -a$$

For example, for a = 0, $x(t)$ is the unit step with Laplace transform $X(s) = \frac{1}{s}$, $\mathrm{Re}\{s\} > 0$.

LAPLACE TRANSFORM PAIRS - $X(s) = \int_{-\infty}^{+\infty} x(t)e^{-st}dt$

Signal	**Transform**	**ROC**
1. $A\delta(t)$	A	All s
2. $Ae^{at}u(t)$	$\frac{A}{s-a}$	Re[s]>a
3. Au(t)	$\frac{A}{s}$	Re[s]>0
4. $Ae^{(a+j\omega)t}u(t)$	$\frac{A}{s-a-j\omega}$	Re[s]>a
5. $Ae^{at}\cos(\omega t)u(t)$	$\frac{A(s-a)}{(s-a)^2+\omega^2}$	Re[s]>a
6. $Ae^{at}\sin(\omega t)u(t)$	$\frac{A}{\omega}\frac{1}{(s-a)^2+\omega^2}$	Re[s]>a
7. $A\cos(\omega t)u(t)$	$\frac{As}{s^2+\omega^2}$	Re[s]>0
8. $A\sin(\omega t)u(t)$	$\frac{A\omega}{s^2+\omega^2}$	Re[s]>0

9. $Ae^{at}\cos(\omega t+\alpha)u(t)$	$\frac{A[(s-a)\cos\alpha-\omega\sin\alpha]}{(s-a)^2+\omega^2}$ $=\frac{0.5Ae^{j\alpha}}{s-(a+j\omega)}+\frac{0.5Ae^{-j\alpha}}{s-(a-j\omega)}$	Re[s]>0
10. $Atu(t)$	$\frac{A}{s^2}$	Re[s]>0
11. $At^n u(t)$	$\frac{An!}{s^{n+1}}$	Re[s]>0
12. $Ate^{at}u(t)$	$\frac{A}{(s-a)^2}$	Re[s]>a
13. $Ae^{at}u(-t)$	$\frac{-A}{s-a}$	Re[s]<0
14. $Au(-t)$	$\frac{-A}{s}$	Re[s]<0
15. $Ae^{at}\cos(\omega t)u(-t)$	$\frac{-A(s-a)}{(s-a)^2+\omega^2}$	Re[s]<a

Table 1.1

UNILATERAL LAPLACE TRANSFORM PROPERTIES AND RELATIONS

$$X(s)=\int_0^{+\infty} x(t)e^{-st}dt$$

Property	**Signal**	**Transform**
1. Linearity	ax(t)+bg(t)	aX(s)+bG(s)
2. Scale Change	x(at)	$\frac{1}{a}X(\frac{s}{a})$
3. Time Derivative	$\frac{dx(t)}{dt}$	sX(s)-x(0)

4. Nth derivative	$\frac{d^N x(t)}{dt^N}$	$s^N X(s) - s^{N-1}x(0)$ $- s^{N-2}\frac{dx(0)}{dt}$ $- \ldots - \frac{d^{N-1}x(0)}{dt^{N-1}}$
5.Time Integral	$\int_0^t x(\tau)d\tau$	$\frac{X(s)}{s}$
6. Complex Differentiation	$tx(t)$	$-\frac{d}{ds}X(s)$
7. Time Shift (a>0)	$x(t-a)u(t-a)$	$e^{-as}X(s)$
8. Frequency shift (a>0)	$e^{at}x(t)$	X(s-a)
9. Convolution	$x_3(t) = \int_0^t x_1(\tau)x_2(t-\tau)d\tau$ $= \int_0^t x_2(\tau)x1(t-\tau)d\tau$	$X_3=X_1(s).X_2(s)$ $X_3=X_2(s).X_1(s)$
10. Final Value Theorem(FVT)	$\lim_{t \mapsto \infty} x(t)$	$\lim_{s \mapsto 0} sX(s)$, *provided all poles of* $sX(s)$ *have negative real parts*
ie	$x(\infty) = \lim_{s \mapsto 0} sX(s)$	
11.Initial Value Theorem(IVT)	$\lim_{t \mapsto 0} x(t)$	$\lim_{t \mapsto \infty} sX(s)$
ie	$x(0^+) = \lim_{s \mapsto \infty} sX(s)$	

Table 1.2

Laplace Transform of Periodic Functions

Let $X(s)$ be the complete Laplace transform of periodic function $x(t)$ of period T and $X_1(s)$ be the Laplace transform of one cycle of periodic function located at origin. Thus as the function is periodic, each cycle of function will have Laplace transform $X_1(s)$. By using the shifting property,

$$X(s) = \frac{X_1(s)}{1-e^{-Ts}},$$ where T is the period of the function.

1-2.1.1 Inverse Laplace Transform

$$x(t) = L^{-1}[X(s)] = \frac{1}{2\pi j}\int_{c-j\omega}^{c+j\omega} X(s)e^{st}dt$$

In general $X(s) = \dfrac{P(s)}{Q(s)} = \dfrac{P(s)}{(s+s_1)(s+s_2).......(s+s_n)}$

The poles may be first order, multiple order, real or complex.

(i) Evaluation of simple poles by partial fractions

$$X(s) = \frac{s+2}{s(s+1)(s+3)} = \frac{A_0}{s} + \frac{A_1}{s+1} + \frac{A_2}{s+3}$$

By residue Theorem

$$A_k = \left[(s+s_q)\frac{P(s)}{Q(s)}\right]_{S=S_q} = \left[\frac{P(s)}{Q^{'}(s)}\right]_{s=sq}$$

Hence $A_0=2/3$, $A_1=-\frac{1}{2}$, $A_2=-1/6$

And $x(t) = [2/3 - e^{-t}/2 - e^{-3t}/6]u(t)$

(ii) For a multiple root, $X(s) = \dfrac{P(s)}{Q(s)} = \dfrac{P(s)}{(s+s_q)^r(s+s_1)}$

$$X(s) = \frac{A_{qr}}{(s+s_q)^r} + \frac{A_{q(r-1)}}{(s+s_q)^{r-1}} + ... + \frac{A_{q1}}{s+s_q} + \frac{A_1}{s+s_1}$$

Generally,

$$A_{q(r-k)} = \frac{1}{k!}\frac{d^k}{ds^k}\left[(s+s_q)^r \frac{P(s)}{Q(s)}\right]_{S=-Sq}$$

eg:

$$X(s) = \frac{1}{(s+2)^3(s+3)} = \frac{A_{13}}{(s+2)^3} + \frac{A_{12}}{(s+2)^2} + \frac{A_{11}}{s+2} + \frac{A_2}{s+3}$$

$$A_{13} = \left[(s+2)^3 X(s)\right]_{S=-2} = 1 \qquad A_{12} = \frac{d}{ds}\left[(s+2)^3 X(s)\right]_{S=-2} = -1$$

$$A_{11} = \frac{1}{2}\frac{d^2}{ds^2}\left[(s+2)^3 X(s)\right]_{S=-2} = 1 \qquad A_2 = \left[(s+2)^3 X(s)\right]_{S=-3} = -1$$

Hence

$$x(t) = \left[\frac{t^2 e-^{2t}}{2} - te^{-2t} + e^{-2t} - e^{-3t}\right]u(t)$$

(iii) Complex

$$X(s) = \frac{20}{(s^2+6s+25)(s+1)} = \frac{As+B}{s^2+6s+25} + \frac{C}{s+1}$$

Multiplying equation by s+1 and substituting s=-1

We get C=1 and $X(s) - \dfrac{1}{s+1} = \dfrac{-(s+5)}{(s+3)^2+4^2}$

And $X(s) = \dfrac{1}{s+1} - \dfrac{s+3}{(s+3)^2+4^2} - \dfrac{2}{(s+3)^2+4^2}$

$$x(t) = \left[e^{-t} - e^{-3t}\cos 4t - \frac{e^{-3t}}{2}\sin 4t\right]u(t)$$

1-2.2 Transfer Functions

The transfer function is the ratio of the Laplace transform of the output variable to the Laplace transform of the input variable with initial conditions assumed to be zero.

1-2.3 Block Diagrams

Block diagrams are unidirectional blocks which represent the transfer function. They are generally connected in cascade, parallel, canonically or combination of these to represent the complete block diagram of the system.
In cascade, the blocks are connected in series as shown:

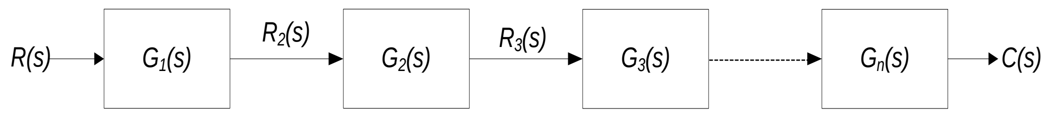


Figure 1.6: Cascade Connection of *n* Blocks

The overall transfer function of the system above is given by:

$$\begin{aligned}
\frac{C(s)}{R(s)} &= \frac{R_2(s)}{R(s)} \cdot \frac{R_3(s)}{R_2(s)} \cdots\cdots \frac{C(s)}{R_n(s)} \\
&= G_1(s) \cdot G_2(s) \cdot G_3(s) \cdots\cdots G_n(s) \\
&= \prod_{i=1}^{n} G_i(s)
\end{aligned}$$

A parallel connection is shown below:

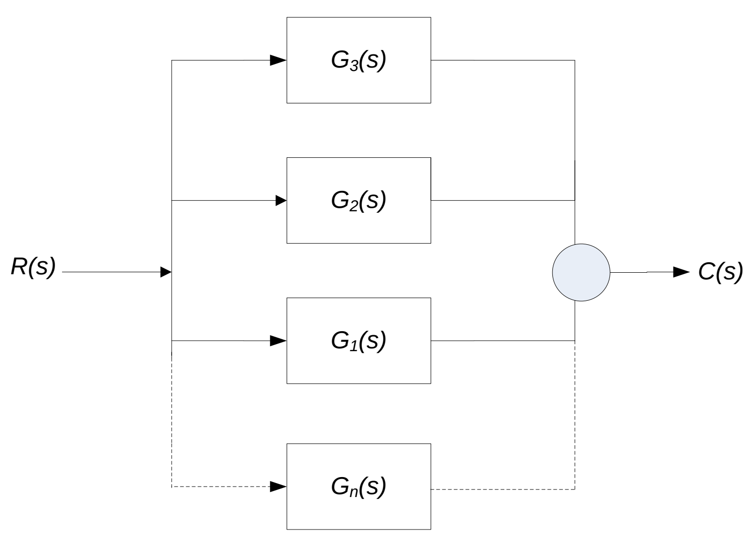


Figure 1.7: Parallel Connection of *n* Blocks

The overall transfer function is

$$\frac{C(s)}{R(s)} = G_1(s) + G_2(s) + \cdots + G_n(s)$$
$$= \coprod_{i=1}^{n} G_i(s)$$

A canonical connection is shown below:

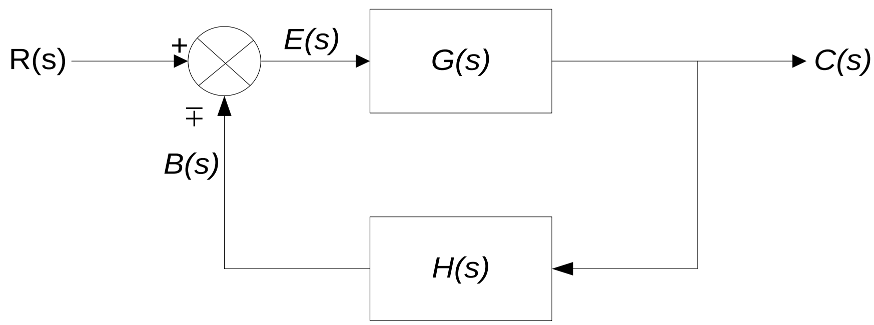


Figure 1.8: A Canonical Connection with a Single Loop

Depending on the sign, the overall transfer function is given by:

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

1-2.3.1 TABLE OF IMPORTANT BLOCK DIAGRAM TRANSFORMATIONS

The following transformations are used to transform complex block diagrams into the simplest equivalent. This consists of a single block which has a transfer function that converts the transformed input variable into the same transformed output variable as the original block diagram.

SOME IMPORTANT BLOCK-DIAGRAM TRANSFORMATIONS

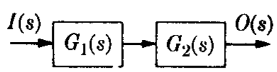
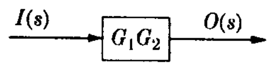
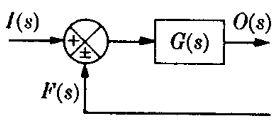
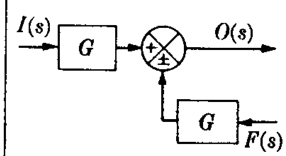
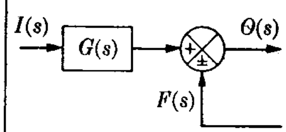
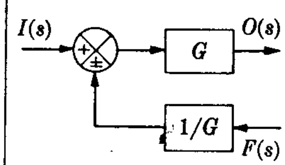
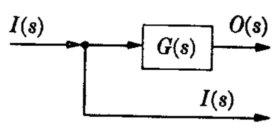
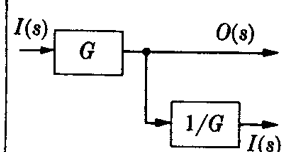
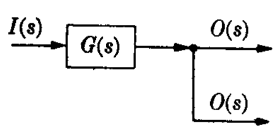
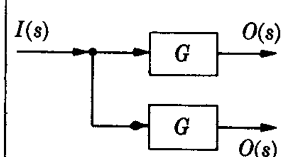
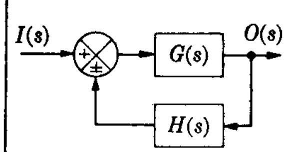
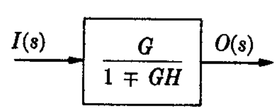
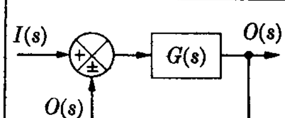
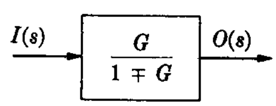
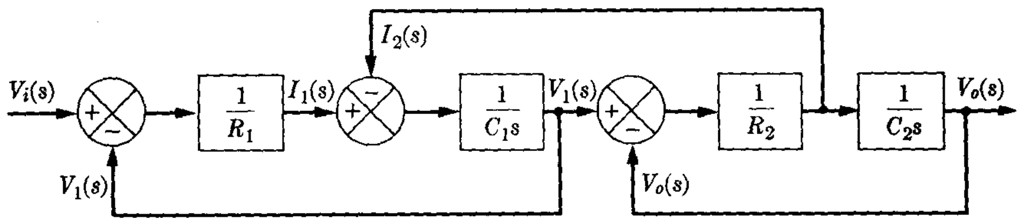
Original diagram	Equivalent diagram	Remarks
		1. Combining two blocks in cascade, $I \cdot G_1 \cdot G_2 = I(G_1 \cdot G_2)$
		2. Moving a summing point behind a block, $(I \pm F)G = (I \cdot G) \pm (F \cdot G)$
		3. Moving a summing point ahead of a block, $(I \cdot G) \pm F = \left(I \pm \frac{1}{G} \cdot F\right) G$
		4. Moving a pickoff point behind a block.
		5. Moving a pickoff point ahead of a block.
		6. Eliminating a feedback loop (see p. 263 for proof).
		7. Special case for Rule 6 above ($H = 1$).

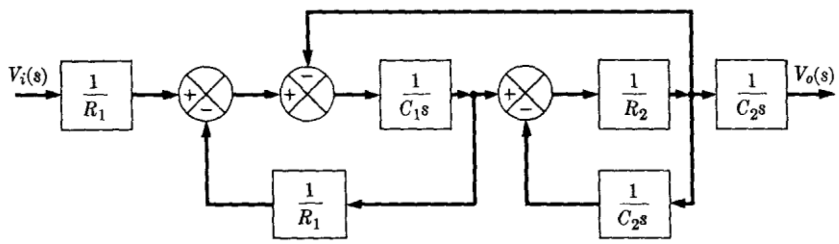
Table 1.3

EXAMPLE 3.1

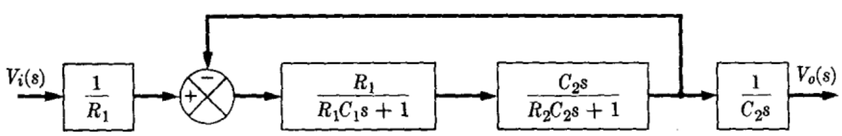
Reduce the block diagram below to open loop form.



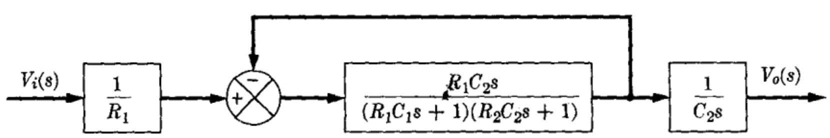
SOLUTION



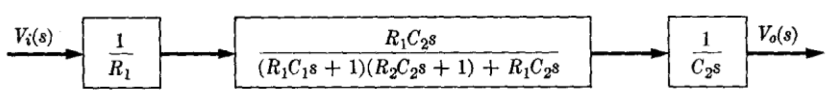
(a)



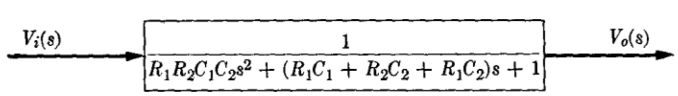
(b)



(c)



(d)



(e)

(a) Moving first summing point behind block $1/R_1$ (Rule No. 2) and last pickoff point ahead of block $1/C_2s$ (Rule No. 5). (b) Eliminating two feedback loops (Rule No. 6). The order of two consecutive summing points is interchangeable. (c) Combining two cascaded blocks (Rule No. 1). (d) Eliminating the last feedback loop (Rule No. 6). (e) Combining the three-cascaded blocks in (d) (Rule No. 1).

1-2.4 Signal Flow Graph

A signal flow graph is also used to denote graphically the transfer function of a control system. A signal flow graph consists of *nodes* which are connected by *directed branches*. Each node j has associated with it a system variable (node signal) X_j, and each directed branch *jk* from node *j* to node *k* has an associated *branch transmittance* T_{jk}. The nodes sum up all the signals that enter them, and transmit the sum signals (node signals) to all *outgoing branches*. In this sense, the nodes are like repeater stations. If the node signals are transformed system variables, then the branch transmittance T_{jk} is essentially the transfer function which specifies the manner in which the signal node k depends upon that at node *j*. The contribution of the signal at node *j*, X_i, to the signal at node *k* is given by:

$$X_k = \sum_j T_{jk} X_j$$

where the summation is taken over all branches entering the k^{th} node.

A typical signal flow graph is shown below:

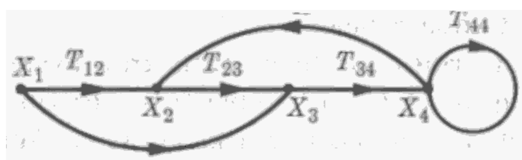


Figure 1.9: A Typical Signal Flow Graph

The equations connecting the (transformed) system variables at the four nodes form a set of linear algebraic equations:

$$X_1 = X_1$$
$$X_2 = T_{12}X_1 + T_{42}X_4$$
$$X_3 = T_{13}X_1 + T_{23}X_2$$
$$X_4 = T_{34}X_3 + T_{44}X_4$$

The above equations can be solved simultaneously for X_2, X_3, and X_4 in terms of X_1. However, the signal flow graph can be reduced by means of a few simple rules of transformation to one with a single directed branch going from the *source (excitation) node* X_1 to the *dependent (response) node* X_2, X_3 or X_4. A *source node* is a node which has only outgoing branches and a *dependent node* is one which has one or more incoming branches. The *overall transmittance* between a source node and a specified dependent node is called a *graph transmittance*.

Nodes with one outgoing branch and more than one incoming branch are called *summing points* while nodes with one incoming branch and more than one outgoing branch are called *pick-off point.*
Signal flow graphs can be represented by block diagram and vice versa. For example, the system represented above in block diagram representation is shown below:

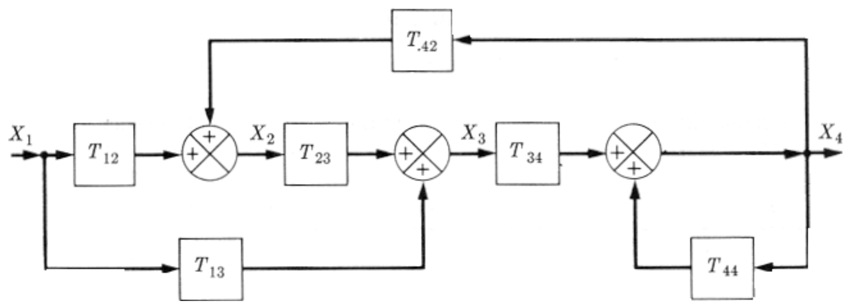


Figure 1.10: Block Diagram Representation of Figure 1.9

1-2.4.1 IMPORTANT SIGNAL FLOW GRAPH TRANSFORMATIONS

SOME IMPORTANT FLOW-GRAPH TRANSFORMATIONS

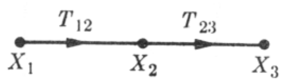
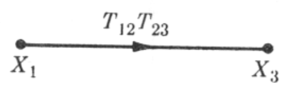
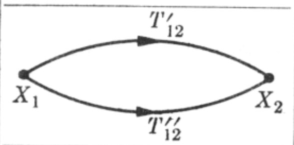
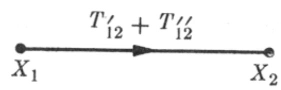
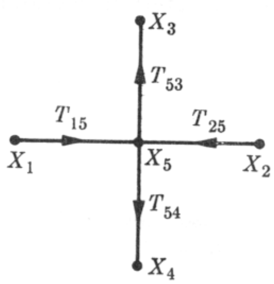
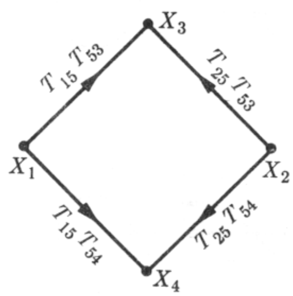
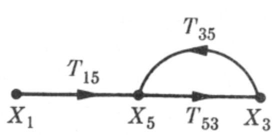
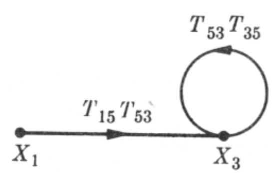
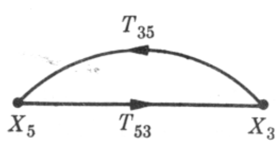
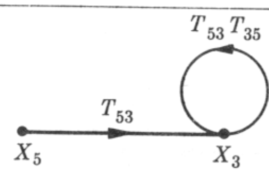
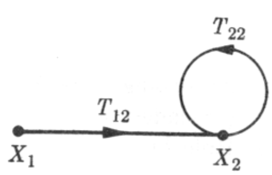
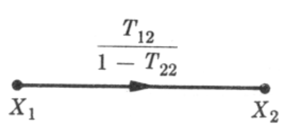
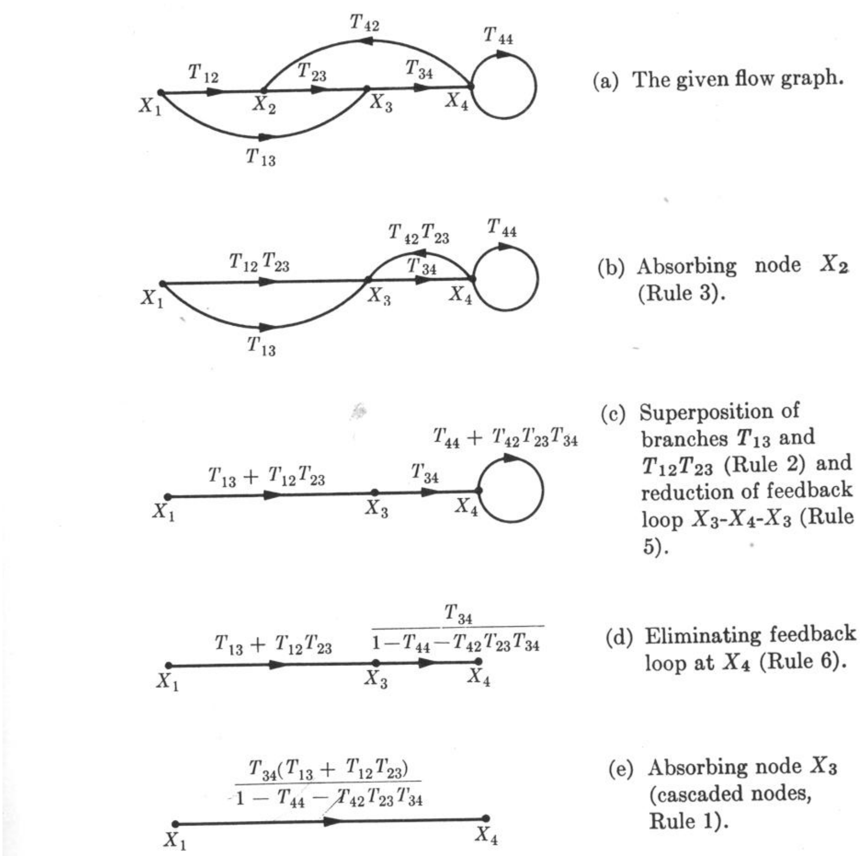
Original graph	Equivalent graph	Remarks
		1. Cascaded nodes, $X_3 = T_{23}X_2 = T_{23}(T_{12}X_1) = T_{12}T_{23}X_1$
		2. Superposition, $X_2 = T'_{12}X_1 + T''_{12}X_1 = (T'_{12} + T''_{12})X_1$
		3. Absorbing a node, $X_5 = T_{15}X_1 + T_{25}X_2$ $X_3 = T_{53}X_5,$ $X_4 = T_{54}X_5,$ or $X_3 = T_{15}T_{53}X_1 + T_{25}T_{53}X_2$ $X_4 = T_{15}T_{54}X_1 + T_{25}T_{54}X_2$
		4. Absorbing a node (special case of Rule 3 above), $T_{54} = 0,$ $X_2 = X_3$
		5. Reducing a feedback loop (special case of Rule 4 above), $T_{15} = 1$
		6. Eliminating a feedback loop, $X_2 = T_{12}X_1 + T_{22}X_2,$ or $X_2 = \frac{T_{12}}{1 - T_{22}} X_1$

Table 1.4

Example 3.2



1-2.4.2 Mason's Gain Formula

Mason's rule allows one to find the transfer function, by inspection, of either a block diagram or a signal flow graph.

Mason's Gain Formula states that the overall transmittance between a source node and a sink node in a signal flow graph may be written as:

$$T = \frac{1}{\Delta}\sum_k T_k \Delta_k$$

Where

$\Delta = 1 -$ (sum of all individual loop gain)

$+$ (sum of all products of loop gains for two non-touching feedback loops)

-(sum of all products of loop gains for three non-touching feedback loops) $+ \ldots,$

Δ_k = re-evaluation of Δ with all loops overlapping forward k^{th} open path set to zero.

T_k = path transmittance of the k^{th} open path

Δ is called the graph determinant, and Δ_k is called the path factor for the k^{th} open path.
The summation is taken over all open paths between the source node and the sink node under consideration. The loop transmittance of a feedback loop is defined as the product of the transmittances of the branches forming the loop.

By Mason's Rule, the transmittance of fig 3.4 can be found as follows:

Two open paths:	(a) X_1-X_3-X_4	$T_a = T_{13}T_{35}$
	(b) X_1- X_2- X_3- X_4	$T_b = T_{12}T_{23}T_{34}$
Two feedback loops:	(a) X_4- X_4	$T_A = T_{44}$
	(b) X_4- X_2- X_4	$T_A = T_{42}T_{23}T_{34}$

No nontouching loops touch open paths (a) and (b).

Both feedback loops touch open paths (a) and (b).
$\Delta = 1 - (T_A + T_B) = 1 - T_{44} - T_{42}T_{23}T_{34}$
$\Delta_a = \Delta_b = 1$

Hence

$$T = \frac{(T_a\Delta_a + T_b\Delta_b)}{\Delta} = \frac{T_{34}(T_{13} + T_{12}T_{23})}{1 - T_{44} - T_{42}T_{23}T_{34}}$$

UNIT 1 ASSIGNMENT

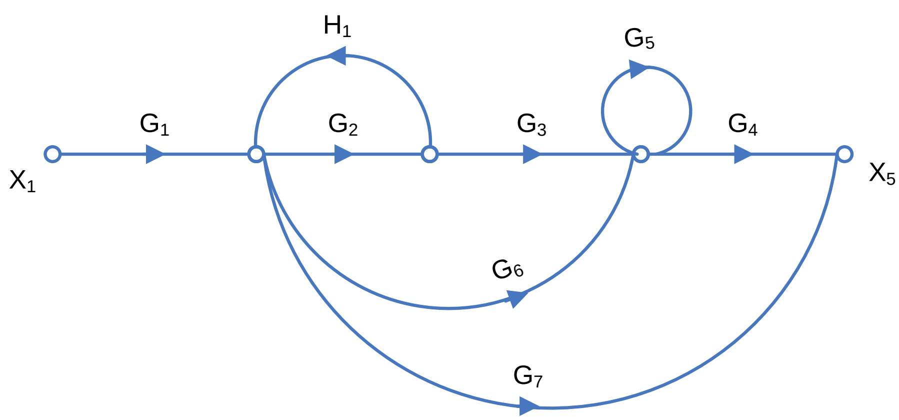
1. Find the inverse Laplace transform of the functions

   i. $F(s) = \dfrac{20}{s(s+6)(s+8)}$

   ii. $F(s) = \dfrac{10(s+6)}{s(s+6)(s^2+3s+2)(s^2+7s+12)}$

2. Find the Laplace transform of $f(t) = 5u(t) + 4e^{-2t}$ and $f(t) = t - 2e^{-t}$

3. Find the transmittance of the graph shown below using Mason's formula.



UNIT 2

SYSTEM ANALYSIS, DESIGN AND PERFORMANCE

INTRODUCTION TO UNIT TWO

Control systems are dynamic in nature. They respond to an input by undergoing a transient response prior to reaching a steady state response the generally resembles the input. In this unit the importance of transient and steady-state response is examined in detail. The new concept of stability is also introduced.

Further in the unit, the steps involved in the analysis and design of control systems introduced. The unit ends with a detailed discussion of control system performance parameters and time domain specification of control systems of the first and second order.

Learning objectives:

1. To get an understanding of transient response, steady-state accuracy, and stability as pertains in the analysis and design of control systems.
2. To gain knowledge of the five major stages involved in the analysis and design of control systems.
3. To gain an understanding of the system parameters such as stability, disturbance rejection, sensitivity to parameter variation etc. used in the evaluation of control system performance.
4. To understand the time domain specifications (peak overshoot, rise time, settling time etc.) of a system

SECTION 2-1: INTRODUCTION TO THE ANALYSIS AND DESIGN PROCESS

In the analysis and design of control systems, the following issues are of the greatest concern:

- Transient response
- Steady-state accuracy
- Stability

2-1.1 Transient Response

In the example of a hard-disk drive for a computer the transient is related to the read-write time. Read/write cannot take place until the head is in place over the correct track of the disk. So speed of the read/write head over the surface of the disk from one track to another will be important for the control of the hard-disk drive.

2-1.2 Steady State

Steady-state is concerned with the state of a system after it arrives at the desired output

- *lift system*: when the lift car reaches the fourth floor;
- *hard disk controller*: when the read-write head is over the correct track on the hard disk.

We are concerned with the **accuracy** of the steady-state.

- The floor of the lift must be sufficiently level with the floor of the corridor to allow passengers to safely enter or leave the car;
- the read-write head would yield disk errors if it was not positioned correctly over a track on the disk surface;
- a satellite tracking system must keep the satellite within its beam width.

2-1.3 Stability

Stability is the most important system specification. If a system is unstable, transient performance and steady-state errors are moot points. An unstable system cannot be designed for a specific transient response or a steady-state error requirement.

2-1.3.1 What is Stability?

There are many definitions for stability depending upon the kind of system or ones point of view. In this section, we limit ourselves to a consideration of the stability of *linear time- invariant systems* (LTIs).

Recall that the response of an LTI system is given by

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

- An LTI is *stable* if the natural response approaches zero as time approaches infinity:

  $$c_{\text{natural}}(t)|_{t\to\infty} = 0$$

  Only the forced response remains as $t \to \infty$.

  $$c(t)|_{t\to\infty} = c_{\text{forced}}(t)$$

- An *unstable* system has a natural response that grows without bound, so that:

  $$c_{\text{natural}}(t)|_{t\to\infty} = \infty$$

  and therefore

  $$c(t)|_{t\to\infty} = \infty$$

- A *marginally stable* system has a natural response that neither grows nor decays as $t \to \infty$ but either oscillates or remains at a constant value.

Physically, an unstable system whose natural response grows without bound can cause damage to the system, adjacent property or human life. In practice many systems are designed with limit stops to prevent runaway. From the time-response point of view, instability is indicated by transients that get bigger and consequently by a total response that does not reach a steady state.

The response of a control system is given as:

$$\text{Total Response} = \text{Forced Response} + \text{Natural Response}$$

When related to the solution of the differential equation used in representing a given control system, the natural response is obtained from the homogeneous solution of the differential equation, and the forced response from the particular solution.

The natural response (or homogeneous solution) describes the way a system acquires or dissipates energy. The form and nature of the natural response depends only on the system, not its inputs.

The form or nature of the forced response (or particular solution) depends on the input.

In some systems, the natural response grows without bound rather than diminishing or oscillating. Eventually the natural response is so much bigger than the forced response that the system becomes "out of control."

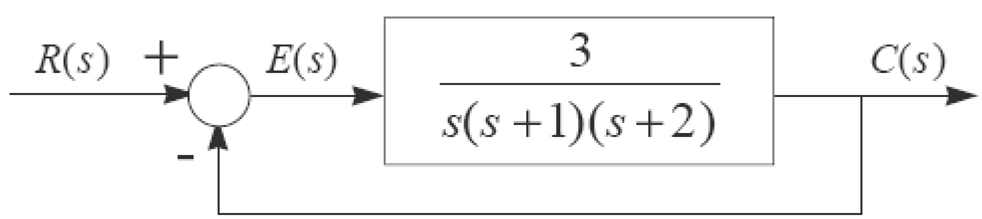
- This condition is called *instability*.
- It could lead to self destruction.

2-1.3.2 How Do We Determine if a System is Stable?

Recall from our study of system poles and zeros that poles to the left of the imaginary axis in the s-plane—a region called the *left-half plane* (LHP)—yield responses that are either decaying exponentials or damped sinusoids. These natural responses decay to zero as time approaches infinity. Thus:

> **Closed-loop stability:** a closed loop control system is stable if all the closed-loop poles are located in the left half plane. ■

Example *Determine the stability of the closed-loop control system shown below.*



Solution: The closed-loop transfer function is

$$G_c(s) = \frac{3}{s^3 + 3s^2 + 2s + 3}.$$

The poles are the zeros (roots) of the closed-loop characteristic equation (CLCE)

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

That is:

$$s = -2.672, -0.164 \pm j1.047.$$

The pole-locations and the resulting response are illustrated below.

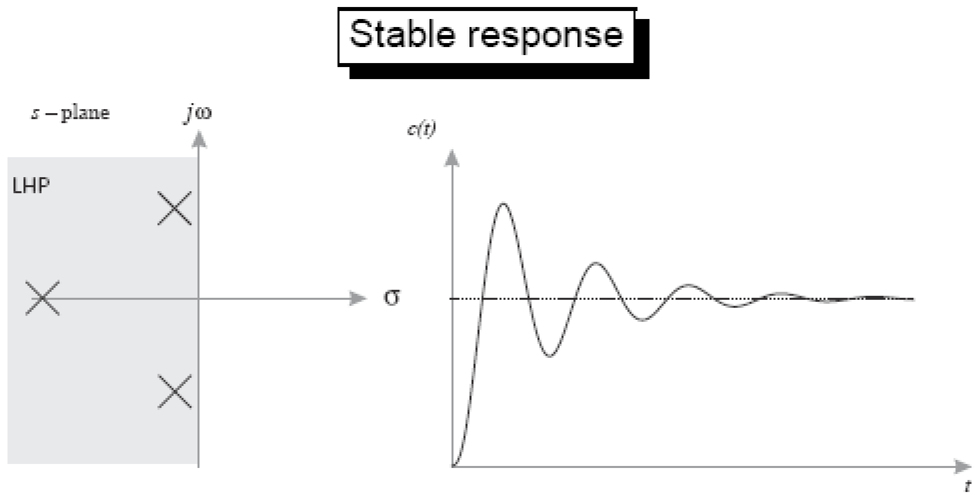
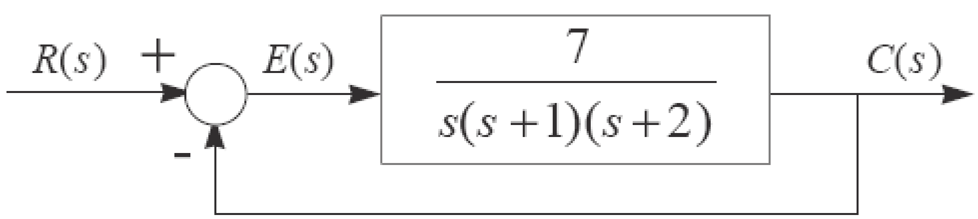


Figure 2.1: Graphical Representation of Stable Response

2-1.4 Instability

A system which has all its poles in the LHP is stable: all the poles will be negative real or complex with negative real parts. On the other hand, poles in the right-half plane are positive or complex with positive real parts. They produce responses which are increasing exponentials or increasing sinusoids. These grow without bound and hence yield unstable responses. Thus if there are poles in the right-half plane the system is unstable.

Example *Determine the stability of the closed-loop control system shown in the figure below.*



Solution: This time the closed-loop transfer function is

$$G_c(s) = \frac{7}{s^3 + 3s^2 + 2s + 7}$$

but now the poles are:

$$s = -3.087, +0.0434 \pm j1.505.$$

Thus, two poles are in the right-half plane (RHP) and the resulting response is unstable .

If there are double or triple poles on the imaginary axis the response will be of the form $At^n \cos(\omega t + \phi)$ $n = 1, 2, \ldots$. Such responses also grow without bound since clearly $t^n \to \infty$ when $t \to \infty$. Thus:

> **Instability of a closed-loop control system:** A system is unstable if its closed-loop transfer function has at least one pole in the right-half plane and/or poles of *multiplicity* > 1 on the imaginary axis. ■

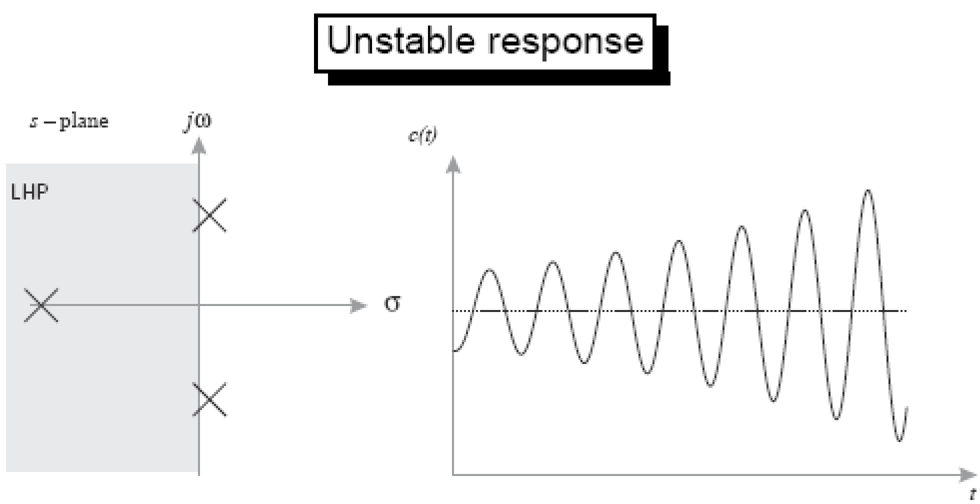


Figure 2.2: Graphical Representation of an Unstable System

2-1.5 Marginal Stability

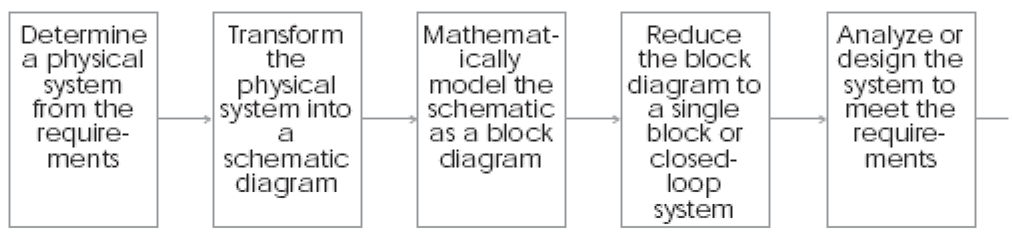
Finally, if a system has a single pair of poles on the imaginary axis, or a single pole at the origin, then we say that the system is *marginally stable*. It will have a natural response containing an undamped oscillation or a constant value as $t \to \infty$.

The topic of stability is discussed again in later sections where the concepts of boundedness and unboundedness are introduced.

2-1.6 The Design and Analysis Sequence

It consists of five basic stages:

1. Determine a physical system from the requirements.
2. Transform the physical system into a schematic.
3. Construct a mathematical model.
4. Perform block-diagram reduction.
5. Analysis and design.



The Control Systems Design and Analysis Sequence

Although the sequence is shown to be linear, it need not be and in practice there will be iterations between stages. Also, it is often the case that the control engineer does not have influence over the first stage and may have to design controllers for existing plant. Another point to be aware of is that the early stages are often quite difficult! In the next sections we review each of the stages in the sequence.

Example of analysis and design process: Antenna Azimuth Position Control

The output is the azimuth angle $\theta_o(t)$ follows the input $\theta_i(t)$ of potentiometer. A model of this system is used in the lab.

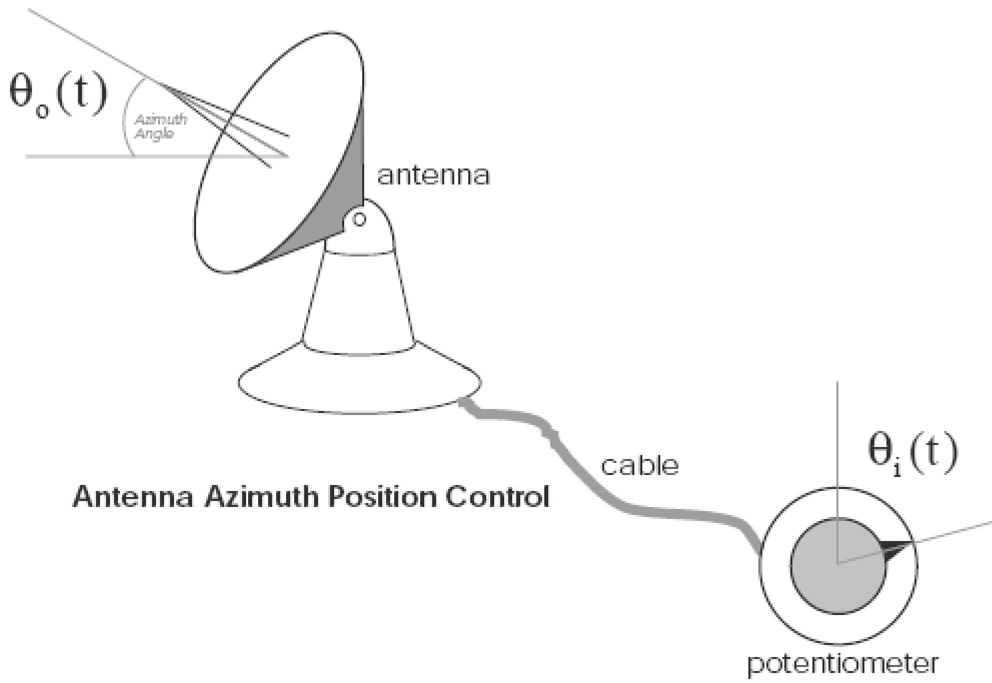


Figure 2.3: Antenna Azimuth Control System

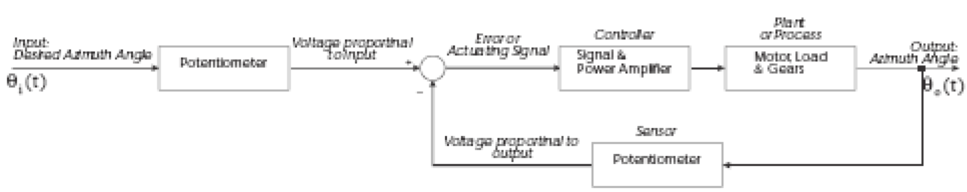


Figure 2.4: Block diagram of Control System

Block Schematic Diagram

A block schematic diagram of the system is shown in Slide 45.

The system normally operates to drive the error to zero.

When the input = the output there is no actuating signal, the motor is not driven. The motor is only driven when the input $\neq$ the output.

The bigger the error the faster the motor.

Transient Performance

What happens if the gain of the signal amplifier is increased?

The motor is driven harder, but the actuating signal is still zero when input = output. The difference will be in transients — motor driven harder so will move faster. The increased speed leads to increased momentum so the system may overshoot the final value and be forced by the system to reverse its direction. The result may be a diminishing oscillation.

Steady State Error

In some systems there is, and the increase in gain will tend to reduce its value. This leads to a trade-off between transient performance and steady-state error. To combat this extra components may need to be added to the system to allow both the gain and the transients to be adjusted. This is called *compensation*.

System error: for a feedback control system is defined as the difference between the demanded output ($r(t)$) and the actual output ($c(t)$)

$$e(t) = r(t) - c(t).$$

Steady-state error: is defined as the difference between the demanded output and the actual output as $t \to \infty$.

$$e(\infty) = \left.(r(t) - c(t))\right|_{t\to\infty}.$$

2-1.7 Test Waveforms Used In Control Systems

To determine the actual response of a system, some test waveforms are used to excite the system, the response or output obtained is used in the analysis and design process. Below is a table of some of the commonly used test waveforms and their uses.

Test Waveforms Used in Control Systems

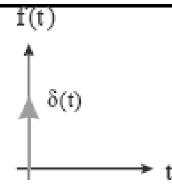
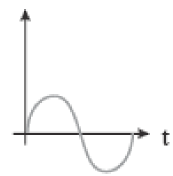
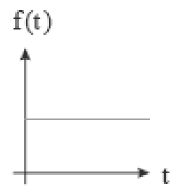
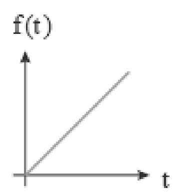
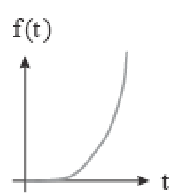
Input	*Function*	*Description*	*Sketch*	*Use*
Impulse	$\delta(t)$	$\delta(t) = \infty$ for $0^- < t < 0^+$; $= 0$ elsewhere; $\int_{0^-}^{0^+} \delta(t)dt = 1$		Transient response modelling
Sinusoid	$\sin \omega t$			Transient response modelling; Steady-state error
Step	$u(t)$	$u(t) = 1$ for $t \geq 0$; $= 0$ elsewhere		Transient response; Steady-state error
Ramp	$tu(t)$	$tu(t) = t$ for $t \geq 0$; $= 0$ elsewhere		Steady-state error
Parabola	$\frac{1}{2}t^2u(t)$	$\frac{1}{2}t^2u(t) = \frac{1}{2}t^2$ for $t \geq 0$; $= 0$ elsewhere		Steady-state error

Table 2.1

2-1.7.1 Use of Simple Test Inputs

Impulse, sinusoid, step, ramp and parabolic inputs are used to excite the system in order to determine the actual response of the system.

- *Impulse* is used to inject energy into the system so that its natural response may be obtained. This can be used to determine what the transfer function of an unknown system is.
- *Sinusoid* is used to determine the steady-state and transient behaviour from frequency response measurements. It can also be used to determine the transfer function of an unknown system.
- *Step* is used to analyse steady-state and transient performance
- *Ramp* and *Parabolic* inputs are used to determine steady-state accuracy..

SECTION 2-2: CONTROL SYSTEM PERFORMANCE

The performance of a control system may be evaluated in terms of its:

1. Stability
2. Transient Response
3. Steady-state Accuracy
4. Sensitivity to parameter variation
5. Disturbance Rejection

2-2.1 Stability as a Performance Parameter

A stable system responds in some reasonable manner to an applied input. For an unstable system there is little apparent relation between the system input and the system output. For linear time-invariant systems, the bounded-input, bounded-output (BIBO) definition of stability is used.

A system is bounded-input, bounded-output stable, if, for every input, the output remains bounded for all time.

By letting $G(s) = G_c(s)G_p(s)$, the transfer function is

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

Let $G(s) = \frac{N_G(s)}{D_G(s)}$ and $H(s) = \frac{N_H(s)}{D_H(s)}$ where $N_G(s)$, $D_G(s)$, $N_H(s)$ and $D_H(s)$ are all polynomials in s. Substituting these into the transfer function,

$$\begin{aligned} T(s) &= \frac{N_G(s)/D_G(s)}{1 + N_G(s)N_H(s)/D_G(s)D_H(s)} \\ &= \frac{N_G(s)D_H(s)}{D_G(s)D_H(s) + N_G(s)N_H(s)} = \frac{P(s)}{Q(s)} \end{aligned}$$

where $P(s)$ and $Q(s)$ are polynomials in s.

The system characteristic equation is given by,

$$Q(s) = D_G(s)D_H(s) + N_G(s)N_H(s) = 1 + G(s)H(s) = 1 + \frac{N_G(s)N_H(s)}{D_G(s)D_H(s)} = 0$$

Let the characteristic equation be represented in factored form as

$$Q(s) = a_n \prod_{i=1}^{n} (s - p_i) = a_n (s - p_1)(s - p_2)\cdots(s - p_n) = 0$$

where a_n is a constant. Note that the roots of the characteristic equation are the same as the poles of the closed-loop transfer function.

From (1)

$$C(s) = T(s)R(s) = \frac{P(s)}{a_n \prod_{i=1}^{n} (s - p_i)} R(s)$$

$$= \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} + \cdots + \frac{k_n}{s - p_n} + C_r(s)$$

where $C_r(s)$ is the sum of the terms, in the partial-fraction expansion, that originate in the poles of $R(s)$. Thus $C_r(s)$ is the forced response. The inverse Laplace transform of C(s) yields

$$c(t) = k_1 e^{p1t} + k_2 e^{p2t} + \cdots + k_n e^{pnt} + c_r(t) = c_n(t) + c_r(t)$$

The $c_n(t)$ terms are the natural response terms, since these terms originate in the poles of the transfer function and the functional forms are independent of the input. If $r(t)$ is bounded, all terms in $c_r(t)$ will remain bounded, since $c_r(t)$ is of the functional form of r(t). Thus if the output becomes unbounded, it is because at least one of the natural-response terms, $k_i e^{pit}$, has become unbounded. This unboundedness cannot occur if the real part of each of the roots p_i of the characteristic equation is negative.

Thus for a linear time-invariant system to be stable, all the roots of the characteristic equations (poles of the closed-loop transfer function) must lie in the left half of the s-plane.

If the characteristic equation has roots on the imaginary axis ($j\omega$-axis) with all other roots in the left half-plane, the steady-state output will sustain oscillations for a bounded input, unless the input is a sinusoid (which is bounded) whose frequency is equal to the magnitude of the $j\omega$ axis. For this case, the output becomes unbounded. Such a system is called a marginally stable, since only certain bounded inputs (sinusoids of the frequency of the poles) will cause the output to become unbounded.

For an unstable system, the characteristic equation has at least one root in the right half of the s-plane; for this case the output will become unbounded for any input.

NOTE:
For the case of an m^th-order repeated root, the partial-fraction expansion of C(s) yields terms of the form

$$L^{-1}\left(\frac{k}{(s-p_i)^m}\right)=\frac{k}{(m-1)!}t^{m-1}e^{pit}$$

This term is bounded if the real part of p_i is negative, which the same condition is developed above.

In summary, a linear time-invariant system is bounded-input, bounded-output stable provided all roots of the system characteristic equation (poles of the closed-loop transfer function) lie in the left half of the s-plane.

Example:
Comment on the stability of a system with the closed-loop transfer function given by each of the following.

1. $T(s)=\dfrac{2}{(s+1)(s+2)}$

2. $T(s)=\dfrac{10(s+3)}{(s+1)(s-3)(s+4)}$

3. $T(s)=\dfrac{1}{s^2+1}$ What happens when an input *r(t)* = sin (*t*) is applied?

2-2.2 Sensitivity

The sensitivity of a system is defined as the ratio of the percentage change in the system transfer function to the percentage change in the system parameter. This ratio is denoted by *S* and expressed as:

$$S=\frac{\Delta T(s)/T(s)}{\Delta b/b}=\frac{\Delta T(s)}{\Delta b}\frac{b}{T(s)}$$

In this equation $\Delta T(s)$ is the change in the transfer function *T(s)* due to the change of the amount Δb in the parameter *b*. By definition, the sensitivity function is given by

$$S_b^T=\lim_{\Delta b\to 0}\frac{\Delta T(s)}{\Delta b}\frac{b}{T(s)}=\frac{\partial T(s)}{\partial b}\frac{b}{T(s)}$$

Sensitivity analysis may need to be performed in order to determine how changes in system parameters will affect the performance of the system. Systems must be built to withstand changes in parameters due to causes such as temperature, pressure, etc.

2-2.3 Disturbance Rejection

In a control system, we have the input to the plant that is used to control the plant, and this input is called the *manipulated variable*. However, any control system will have other inputs that influence the plant output and that, in general, we do not control. We call these inputs *disturbances* and usually attempt to design the control system such that these disturbances have a minimum effect on the system.

Feedback in systems, controls and partially eliminates the disturbance signal in the system. These disturbance signals cause the system to generate inaccurate output. The examples of disturbance signals are thermal noise in amplifiers, wind gusts on antenna radar systems, synchro noise in servo systems, distortion produced due to non-linear system etc. Disturbance in control system can be introduced in the following parts of the system.

i. Forward path of the system

ii. Feedback path of the system

iii. Output of the system.

I. Disturbance in the forward path of the system: These disturbances may be produced either due to properties of elements present in the forward path or due to the surrounding condition changes of forward path elements. Consider the signal flow graph shown below with disturbance signal $T_d(s)$:

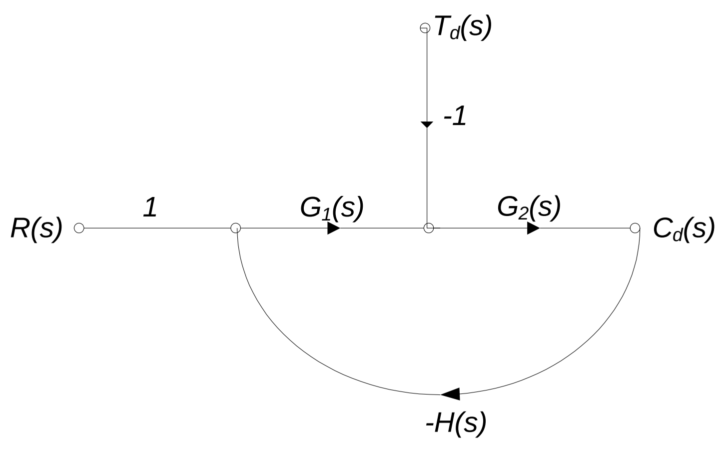


Figure 2.5: Disturbance Signal in the Forward Path

The ratio of output signal $C(s)$ to the disturbance signal $T_d(s)$ is

$$M_d(s) = \frac{C_d(s)}{T_d(s)} = \frac{-G_2(s)}{1 + G_1(s)G_2(s)H(s)}$$

If at the working range of s in $G_1(s)$, $G_2(s)H(s)$ is much larger than one, then

$$M_d(s) = \frac{C_d(s)}{T_d(s)} \cong \frac{-G_2(s)}{G_1(s)G_2(s)H(s)}$$

$$= \frac{-1}{G_1(s)H(s)}$$

$$C_d(s) = \frac{-T_d(s)}{G_1(s)H(s)}$$

If $G_1(s)$ is made very large the effect of disturbance $T_d(s)$ on the output will be very small. Thus it is always desired to have large $G_1(s)$, so that the output may not change due to disturbance in the forward path of the system. The sensitivity of the system for changes in the forward path transfer function can be found.

II. Disturbance in the Feedback Path: These disturbances are produced due to changes in the feedback path elements like the sensing device. Consider the signal flow graph of a closed-loop system with the disturbance $T_d(s)$ in the feedback path.

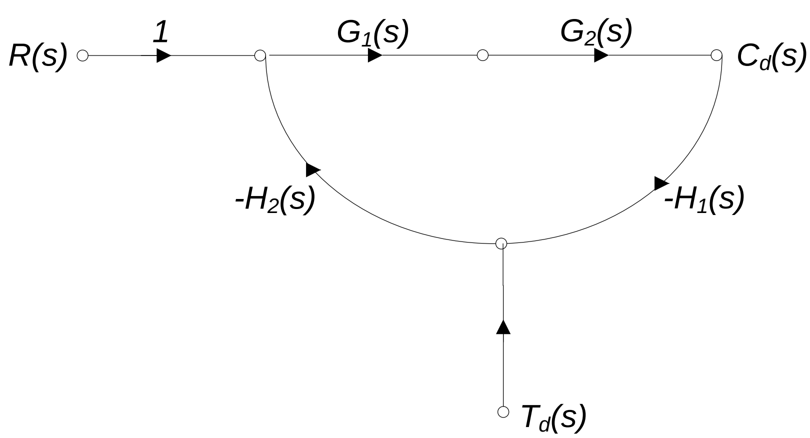


Figure 2.6: Disturbance Signal in the Feedback Path

$$M_d(s) = \frac{C_d(s)}{T_d(s)} = \frac{-G_1(s)G_2(s)H_2(s)}{1 + G_1(s)G_2(s)H_1(s)H_2(s)}$$

For large values of $G_1(s)$, $G_2(s)$ and $H_2(s)$, the above equation reduces

$$M_d(s) = \frac{C_d(s)}{T_d(s)} \cong \frac{-1}{H_1(s)}$$

$$C_d(s) = -\frac{T_d(s)}{H_1(s)}$$

Thus by choosing proper value of $H_1(s)$, the effect of disturbance in the feedback path can be reduced at the output of the system. The sensitivity of the system for disturbances due to feedback path variation can be found.

III. Disturbance at the Output of the System: The disturbances at the output of the system are introduced at the output terminals of the system as illustrated in the diagram below:

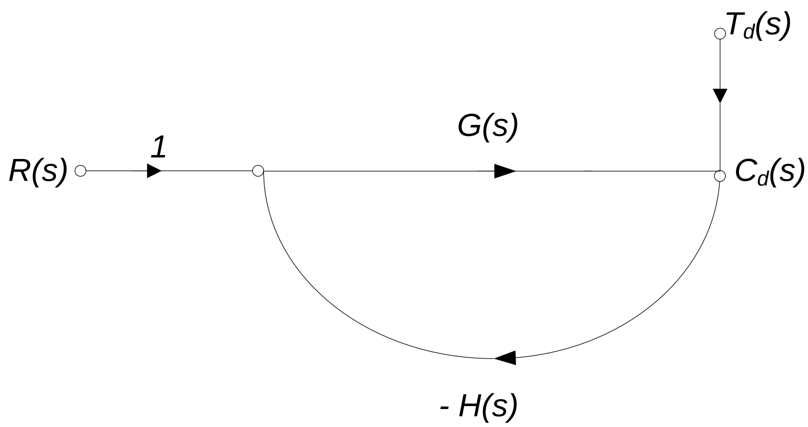


Figure 2.7: Disturbance Signal at the Output of the System

$$M_d(s) = \frac{C_d(s)}{T_d(s)} = \frac{1}{1 + G(s)H(s)}$$

For larger values of $G(s)H(s)$, the equation reduces to

$$C(s) \cong \frac{1}{G(s)H(s)} \cdot T_d(s)$$

Hence, the effect of disturbance on the output can be controlled by changing the values of $G(s)$ and $H(s)$. The sensitivity of the system to disturbance at the output terminals for variations in $G(s)$ and $H(s)$ can be found.

2-2.4 STEADY-STATE ACCURACY

2-2.4.1Characterization of System Errors

For the unit feedback system, the system error is

$$e(t) = r(t) - c(t)$$

Thus

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

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

The error transfer function $\frac{1}{1+G(s)}$ may be expanded in a power series as

$$\frac{1}{1+G(s)} = C_0 + C_1 s + C_2 s^2 + C_3 s^3 + \ldots$$

so that

$$E(s) = C_0 R(s) + C_1 sR(s) + C_2 s^2 R(s) + \ldots$$

or in time domain

$$e(t) = C_0 r(t) + C_1 \frac{dr(t)}{dt} + C_2 \frac{d^2 r(t)}{dt^2} + \ldots$$

The coefficients C_i, are known as the (dynamic) error coefficients of the system.

C_0 is the position error coefficient
C_1 is the velocity error coefficient
C_2 is the acceleration error coefficient

2-2.4.2 Determination of Error Coefficients

Consider a function

$$f(s) = C_0 + C_1 s + C_2 s^2 + \ldots = \frac{1}{1+G(s)}$$

Then

$$C_0 = f(0) \qquad C_1 = f'(0) \qquad C_2 = \frac{1}{2.1} f''(0)$$

In general $\quad C_i = \frac{1}{n!}\left[\frac{d^n f}{ds^n}\right]_{s=0}$, n = 0, 1, 2, …

In Bode Form

$$G(s) = K \frac{\prod_{j=1}^{m}(1+sT_Z j)}{s^K \prod_{j=1}^{n}(1+sT_P j)}$$

$$\frac{1}{1+G(s)} = \frac{s^K \prod_{j=1}^{n}(1+sT_P j)}{s^K \prod_{j=1}^{n}(1+sT_P j) + K\prod_{j=1}^{m}(1+sT_Z j)}$$

2-2.5 System Type or Class Number

Feedback control systems in the unit feedback configuration are classified by means of a class or type number. This is simply the number of integrators, $1/s$, in the open loop transfer function, $G(s)$. Type n $\rightarrow \dfrac{1}{s^n}$.

For n = 0, i.e. type 0 system $\quad C_0 = \dfrac{1}{1+K_0}$

For n = 1, i.e. type 1 system $\quad C_0 = 0, \qquad C_1 = \dfrac{1}{K_1}$

For n = 2, i.e. type 2 system $\quad C_0 = 0, \qquad C_1 = 0, \qquad C_2 = \dfrac{1}{K_2}$

Example:

A feedback control system has the open loop transfer function

$$G(s) = \frac{10(s+10)}{s(s+5)}$$

Determine the error in the response of the system to an input, $r(t) = t^2$.

Solution:

It is necessary to evaluate only C_0, C_1 and C_2.

Now $\quad G(s) = \dfrac{20(1+0.1s)}{s(1+0.2s)} \qquad$ Bode gain $K_1 = 20$

System is type 1 $\qquad \therefore C_0 = 0 \qquad C_1 = \dfrac{1}{K_1} = \dfrac{1}{20}$

$$C_2 = \frac{1}{2}\left[\frac{d^2}{ds^2}\left(\frac{1}{1+\omega_0(s)}\right)\right]_{s=0}$$

$$= \frac{1}{2}\left[\frac{d}{ds}\frac{d}{ds}\frac{s(s+5)}{s(s+5)+10(s+10)}\right]\Bigg|_{s=0}$$

$$= 5\left[\frac{d}{ds}\frac{s^2+20s+50}{\left(s(s+5)+10(s+10)\right)^2}\right]_{s=0} = \frac{1}{400}$$

Thus

$$e(t) = C_0 t^2 + C_1.2t + C_2.2$$

$$e(t) = 0 + \frac{2}{20}t + \frac{2}{400}$$

$$e(t) = \frac{t}{10} + \frac{1}{200}$$

2-2.6 Steady State Errors

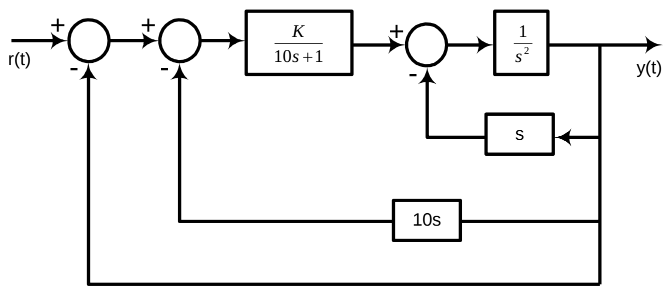
Steady state error in a system is

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

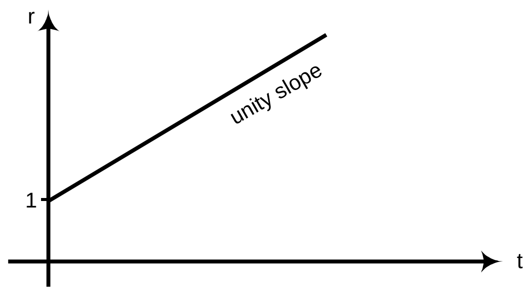
$$e(\infty) = \lim_{s \to 0} sE(s)$$

Ex

1.



If the input is



at rest.

2-2.7 EVALUATION OF SYSTEM RESPONSE

2-2.7.1 Poles and Zeros and System Response

The output response of a linear system is the sum of the forced response and the natural response

$$c(t) = c_f(t) + c_n(t).$$

Solving differential equations or using inverse Laplace transforms allow us to evaluate the output response. But this is laborious and time consuming. By having an analysis technique that is quick and easy to apply we will increase productivity. Knowledge of the effects of poles and zeros gives us a means of qualitatively evaluating the response of a system by inspection.

> **Poles of a Transfer Function:** The poles of a transfer function are those values of the transfer function variables that cause the transfer function to become infinite. ■

> **Zeros of a Transfer Function:** The zeros of a transfer function are those values of the transfer function variables that cause the transfer function to become zero. ■

Let $G(s)$ be a transfer function

$$G(s) = \frac{b(s)}{a(s)} = \frac{s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}$$

The poles of $G(s)$ are solutions of

$$a(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0$$

that is, they are *zeros* of the *denominator polynomial* $a(s)$. If the denominator of $G(s)$ is zero then $G(s) = \infty$.

The zeros of $G(s)$ are solutions of

$$b(s) = s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0 = 0$$

that is, they are *zeros* of the *numerator polynomial* $b(s)$. If the numerator of $G(s)$ is zero then $G(s)$ is also 0.

Example: *Determine the step response of a transfer function.*

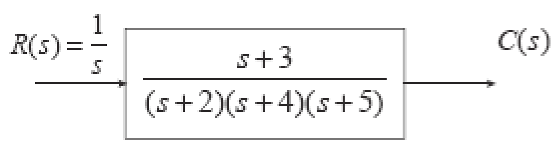
Solution: The step response is

$$\begin{aligned} C(s) &= \frac{s+2}{s(s+5)} \\ &= \frac{A}{s} + \frac{B}{s+5} \\ &= \frac{2/5}{s} + \frac{3/5}{s+5} \end{aligned}$$

thus

$$c(t) = \frac{2}{5} + \frac{3}{5}e^{-5t}.$$

Example: *Use poles to evaluate the system response of the system shown in the figure below by inspection.*



Solution:

$$C(s) = \underbrace{\frac{K_1}{s}}_{\text{forced response}} + \underbrace{\frac{K_2}{s+2} + \frac{K_3}{s+4} + \frac{K_4}{s+5}}_{\text{natural response}}$$

$$c(t) = \underbrace{K_1}_{c_f(t)} + \underbrace{K_2e^{-2t} + K_3e^{-4t} + K_4e^{-5t}}_{c_n(t)}.$$

2-2.8 Transient Response Analysis

In general, the overall transfer function of a closed-loop negative feedback control system can be written as:

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

Factorizing

$$\frac{C(s)}{R(s)} = \frac{K\prod_{j=1}^{m}(s+z_j)}{\prod_{e=1}^{n+2r}(s+p_e)}$$

$$= \frac{K\prod_{j=1}^{m}(s+z_j)}{\prod_{i=1}^{n}(s+\sigma_i)\prod_{k=1}^{r}[s+(\alpha_k + j\omega_k)][s+(\alpha_k - j\omega_k)}$$

Where m are the zeros at z_j, *n* are the real pole at σ_i, and *r* are the pairs of complex conjugate poles. For input *r(t)*, the response of the system is

$$c(t) = L^{-1}[C(s)]$$

$$= L^{-1}\left[R(s)\frac{K\prod_{j=1}^{m}(s+z_j)}{\prod_{i=1}^{n}(s+\sigma_i)\prod_{k=1}^{r}[s+(\alpha_k + j\omega_k)][s+(\alpha_k - j\omega_k)}\right]$$

For unit input r(t)=1,

$$c(t) = L^{-1}\left[\frac{K\prod_{j=1}^{m}(s+z_j)}{s\prod_{i=1}^{n}(s+\sigma_i)\prod_{k=1}^{r}[s+(\alpha_k + j\omega_k)][s+(\alpha_k - j\omega_k)}\right]$$

Let A_i and B_k be the partial fractions of the above equation, then

$$c(t) = L^{-1}\left[\frac{1}{s}+\sum_{i=1}^{n}\frac{A_i}{(s+\sigma_i)}+\sum_{k=1}^{r}\frac{B_k}{[s^2+2\alpha_k s+(\alpha_k^2+\omega_k^2)}\right]$$

Assuming no repeated roots,

$$c(t) = 1 + \sum A_t e^{-\sigma_i t} + \sum \frac{B_k}{\omega_k} e^{-\alpha_k t} \sin(\omega_k t)$$

Steady state Transient response terms Response term

The transient response terms can be of exponential form, damped sinusoid or a combination of both depending on the properties of the system characteristic equation. Since the real root σ_i and the real part of the complex α_k , appears as exponents in the time response, the damping of time response is controlled by σ_i and α_k . The imaginary part of the complex root $j\omega_k$ appears as the frequency of the sinusoidal term and thus controls the oscillation frequencies of the time response.

2-2.9 Time Domain Specifications

These are defined in terms of unit step, ramp and parabolic functions. Each response has a steady state and a transient component.

Steady-state performance is a measure of system accuracy when a specific input is applied. Figures of merit are the error constants, $K_{n.}$ while transient response is normally described in terms of the unit step function response.

Typical specifications of transient response are peak overshoot and measures of speed response.

2-2.10 First-Order System Response and Specifications

A first-order system has a transfer function of the form

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

which is also called a "*first-order lag*". The system has one pole at $s = -a$ as shown in the *pole-zero diagram*. For the step response, the input signal transform is

$$R(s) = \frac{1}{s}$$

so the step response is

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

In the time domain the step response is:

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

The forced response $c_f(t) = 1$ is generated by the pole of $1/s$. The natural response $c_n(t) = e^{-at}$ is generated by the pole of $a/(s+a)$.

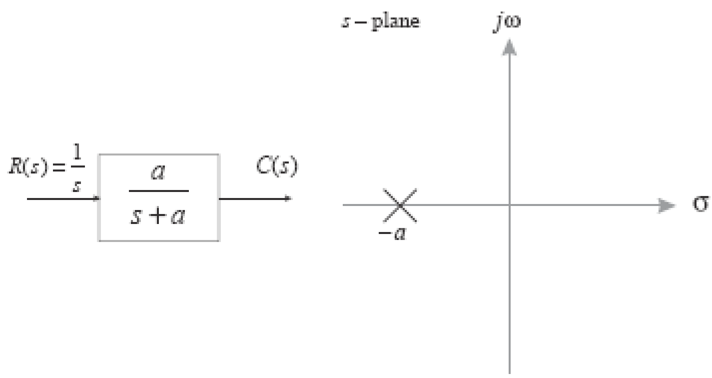


Figure 2.8: A First-Order System

Significance of a

The parameter a is very important for specifying the performance of a first-order system. It is significant because

$$e^{-at}\big|_{t=1/a} = e^{-1} = 0.37.$$

Or, alternatively, the step response

$$c(t) = 1 - e^{-at}\big|_{t=1/a} = 0.63.$$

That is, we can relate the shape of the time response to the parameter a.

Time Constant

The parameter $\tau = 1/a$ is called the "*time constant*" of the first-order response. It is the time taken for e^{-at} to decay to 37% of its initial value or the time taken for the step response to reach 63% of its final value. τ has units s. a has units s^{-1} or frequency. a is called the *exponential frequency*. The derivative of e^{-at} at $t = 0$ is $-a$, so a is the initial rate of change of the exponential at $t = 0$. Thus the time constant can be considered a transient response specification for a first-order system since it is related to the speed of response.

The time constant can be evaluated directly from the pole-zero plot. The pole is $s = -a$. The further to the left of s-plane, the larger a hence the smaller the time constant $\tau = 1/a$ and the faster the response. A plot of a typicalfirst-order response is shown in the figure that follows.

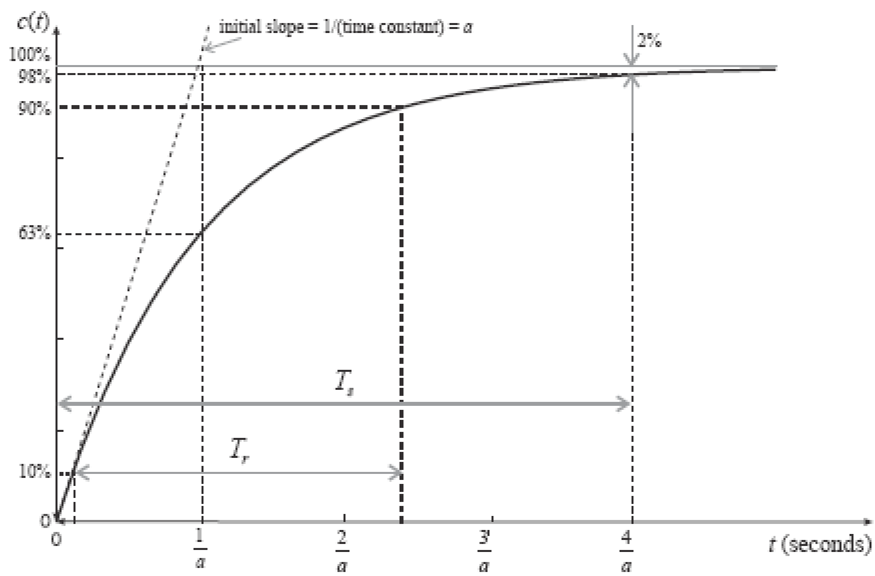


Figure 2.9: Graphical Representation of a First Order Response

Rise Time

The rise-time (symbol T_r units s) is defined as the time taken for the step response to go from 10% to 90% of the final value. For a first-order system it is rather easily derived by solving $c(t) = e^{-at}$ for $c(t_{0.1}) = 0.1c_{\text{final}}$ and $c(t_{0.9}) = 0.9c_{\text{final}}$ from which

$$T_r = \frac{2.31}{a} - \frac{0.11}{a} = \frac{2.2}{a}.$$

Settling Time

We shall define the settling-time (symbol T_s units s) to be the time taken for the step response to come to within 2% of the final value of the step response.[1] From the definition,

$$c(T_s) = 0.98c_{\text{final}}$$

which gives

$$T_s = \frac{4}{a}.$$

In other words the 2% settling-time for a first-order system is $4\times$ the time constant.

2-2.11 The General Second-Order Response

Definitions of Some Important Terms

Natural frequency ω_n is defined as the frequency of oscillation of a second-order system without damping. E.g. the frequency of oscillation of an RLC circuit with R shorted. It has units of rad.s^{-1}.

Damping ratio ζ is defined by

$$\begin{aligned} \zeta &= \frac{\text{Exponential decay frequency}}{\text{Natural frequency}} \\ &= \frac{1}{2\pi}\frac{\text{Natural period}}{\text{Exponential time constant}} \end{aligned}$$

The damping ratio has no units. That is the damping ratio of a system that decays to zero after 3 oscillations in 1 ms is the same as that of a system that decays to zero in three oscillations in 1 hour. It is independent of the speed of response or the rate of the oscillation.

Derivation of Formulae

Let us derive formulae for these quantities from their definitions, given that

$$G(s) = \frac{b}{s^2 + as + b}.$$

Without damping, the term $as = 0$ and we have

$$G(s) = \frac{b}{s^2 + b}.$$

The poles are imaginary, and the frequency of oscillation of this system is ω_n by definition. Thus

$$\omega_n = \sqrt{b},$$
$$b = \omega_n^2.$$

Assuming an underdamped system. The complex poles have a real part given by

$$s = -\frac{a}{2}.$$

The magnitude of the real part is the exponential decay "frequency" σ_d. Thus $\sigma_d = a/2$ and from the definition

$$\zeta = \frac{\sigma_d}{\omega_n} = \frac{a/2}{\omega_n},$$

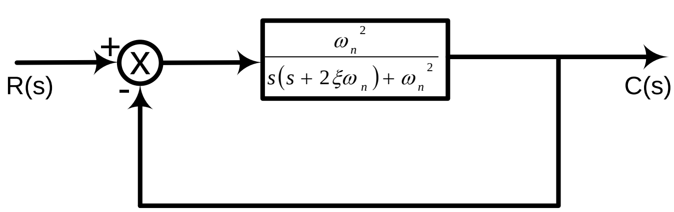
hence

$$a = 2\zeta\omega_n.$$

In general then, the second order system has the **canonical form**:

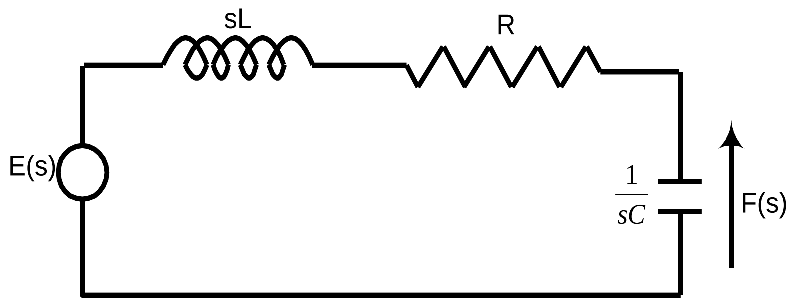
$$\boxed{G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}}$$

Knowledge of the parameters ω_n and ζ may be used to determine the type of motion of any particular second-order system.



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

Figure 2.10: Block Diagram of Second Order Control System



$$G(s)=\frac{F(s)}{E(s)}=\frac{1/sL}{sL+R+\frac{1}{sC}} \qquad =\frac{1/LC}{s^2+\frac{R}{L}s+\frac{1}{LC}} \qquad =\frac{\omega_n{}^2}{s^2+2\xi\omega_n s+\omega_n{}^2}$$

Figure 2.11: An RLC Circuit That Models a Second Order System.

Some problems in second order systems involve finding such parameters as the damping ration and the natural frequency for the system represented by a given transfer function.

Example *Find the natural frequency ω_n and damping ratio ζ for the system with transfer function:*

$$G(s)=\frac{36}{s^2+4.2s+36}.$$

Solution: Comparing to the standard form

$$G(s)=\frac{36}{s^2+4.2s+36}=\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$$

$$\omega_n^2=36\rightarrow\omega_n=\sqrt{36}=6.$$

$$\begin{aligned}2\zeta\omega_n &= 4.2\\ \zeta &= \frac{4.2}{2\omega_n}=\frac{4.2}{2\times 6}=0.35.\end{aligned}$$

2-2.11.1 Pole-Zero Locations

Having defined ω_n and ζ, let us relate these quantities to the pole locations in the s-plane.

Solving the *characteristic equation*

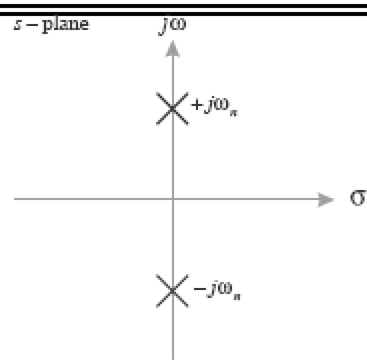
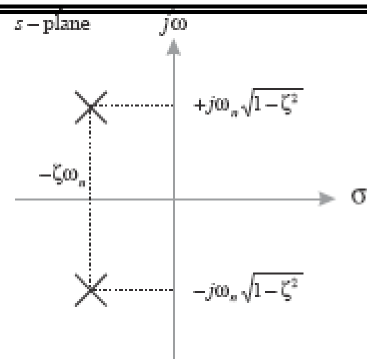
$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0$$

gives

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

The various types of response for a given value of natural frequency ω_n are a function of ζ and may be summarised as shown in the table below.

The relationship between damping ratio, pole location and transient reponsefor second-order systems with transfer function $G(s) = \omega_n^2/(s^2 + 2\zeta\omega_n$ $\omega_n^2)$.

ζ	Pole Locations	Type of Response
$\zeta = 0$		Oscillatory
$\zeta < 1$		Underdamped

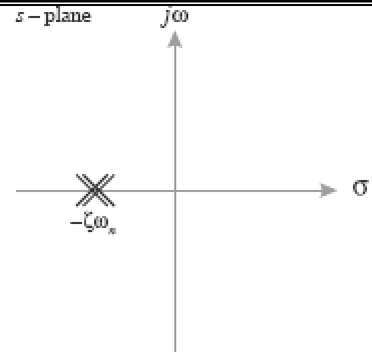
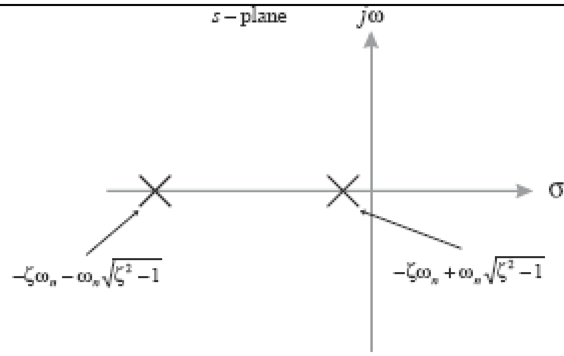
ζ	Pole Locations	Type of Response
$\zeta = 1$		Critically Damped
$\zeta > 1$		Overdamped

Table 2.2

Example *Describe the nature of the second-order system response via the value of the damping ratio for the systems with transfer function:*

$$
\begin{aligned}
G(s) &= \frac{12}{s^2 + 8s + 12} \\
G(s) &= \frac{16}{s^2 + 8s + 16} \\
G(s) &= \frac{20}{s^2 + 8s + 20}.
\end{aligned}
$$

Solution: In all cases the transfer function is of the form

$$G(s) = \frac{b}{s^2 + as + b}$$

so $a = 2\zeta\omega_n$ and $\omega_n = \sqrt{b}$ hence $\zeta = a/(2\sqrt{b})$. $a = 8$ in all cases.

For

$$G(s) = \frac{12}{s^2 + 8s + 12}$$

$b = 12$ hence $\zeta = 8/(2\sqrt{12}) = 2/\sqrt{3} > 1$: system response is **overdamped**.

For

$$G(s) = \frac{16}{s^2 + 8s + 16}$$

$b = 16$ hence $\zeta = 8/(2\sqrt{16}) = 1$: system response is **critically damped**.

For

$$G(s) = \frac{20}{s^2 + 8s + 20}$$

$b = 20$ hence $\zeta = 8/(2\sqrt{20}) = 2/\sqrt{5} < 1$: system response is **underdamped**.

The dynamic behaviour of second order systems can be described in terms of two parameters ξ and ω_n.

Four cases will be considered

1. **Underdamped case** ($0 < \xi < 1$)

$$\frac{C(s)}{R(s)} = \frac{\omega_n^{\,2}}{(s + \xi\omega_n + j\omega_d)(s + \xi\omega_n - j\omega_d)}$$

where $\omega_d = \omega_n\sqrt{1 - \xi^2}$

= damped natural frequency

For a unit step input

$$R(s) = \frac{1}{s}$$

$$C(s) = \frac{1}{s} - \frac{s + \xi\omega_n}{(s + \xi\omega_n)^2 + \omega_d} - \frac{\xi\omega_n}{(s + \xi\omega_n)^2 + \omega_d^{\ 2}}$$

$$c(t) = 1 - \frac{e^{-\xi\omega_n t}}{\sqrt{1 - \xi^2}} \sin\left(\omega_d t + \tan^{-1}\frac{\sqrt{1 - \xi^2}}{\xi}\right) \qquad t>0$$

If $\xi = 0$, $c(t) = 1 - \cos\omega_n t$
ω_n = undamped natural frequency of the system.

Consider the second order system

$$G(s) = \frac{b}{s^2 + as + b}$$

with $b = 9$ and a selected to illustrate each type of response.

IF $a = 3$ we have the system shown in the figure below.
The step response will be

$$C(s) = \frac{9}{s(s^2 + 3s + 9)}.$$

The pole at $s = 0$ comes from the input (forced response) and there are two complex poles at $s = -1.5 \pm j2.598$. The step response of this system will be of the form

$$c(t) = K_1 + e^{-1.5t}(K_2 \cos 2.598t + K_3 \sin 2.598t).$$

Comparing the time response with the pole locations, we see that the exponential decay term is related to the real-part of the complex pole pair. Real part: $\sigma_d = 1.5$; exponential term $e^{-1.5t} = e^{-\sigma_d t}$. The frequency of the oscillatory term is related to the imaginary part of the complex pole pair. Imaginary part: $\omega_d = 2.598$; sinusoidal term: $K_2 \cos 2.598t + K_3 \sin 2.598t$

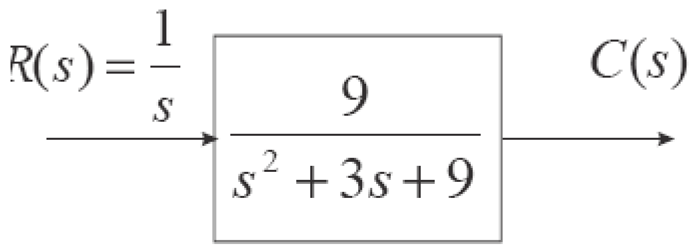


Figure 2.12: Block Diagram for Underdamped System

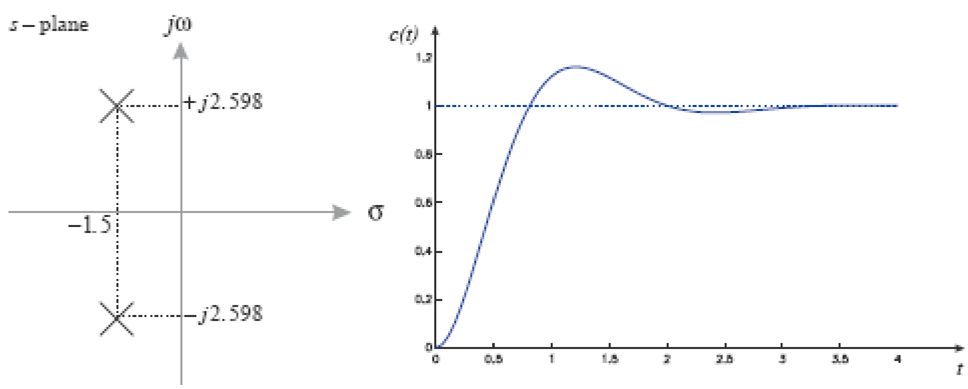


Figure 2.13: Graphical Representation of the Response of an Underdamped System

2. **Critically damped case** $(\xi = 1)$

For a step input

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

$$c(t) = 1 - e^{-\omega_n t}(1 + \omega_n t) \qquad\qquad t \geq 0$$

Consider the second-order system

$$G(s) = \frac{b}{s^2 + as + b}$$

with $b = 9$ and a selected to illustrate each type of response.

IF $a = 6$ we have the system shown.

The step response will be

$$C(s) = \frac{9}{s(s^2 + 6s + 9)} = \frac{9}{s(s+3)^2}.$$

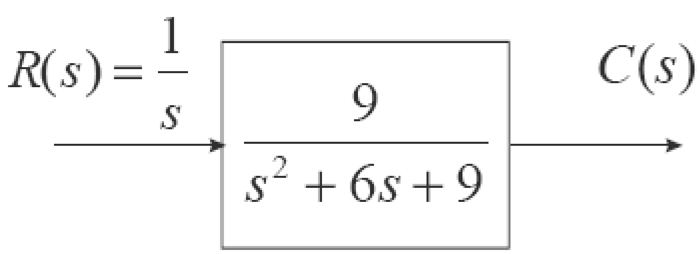


Figure 2.14: Block Diagram of a Critically Damped System

The pole at $s = 0$ comes from the input (forced response) and there are two real and equal poles at $s = -3$. The step response of this system will be of the form

$$c(t) = K_1 + K_2 e^{-3t} + K_3 t e^{-3t}.$$

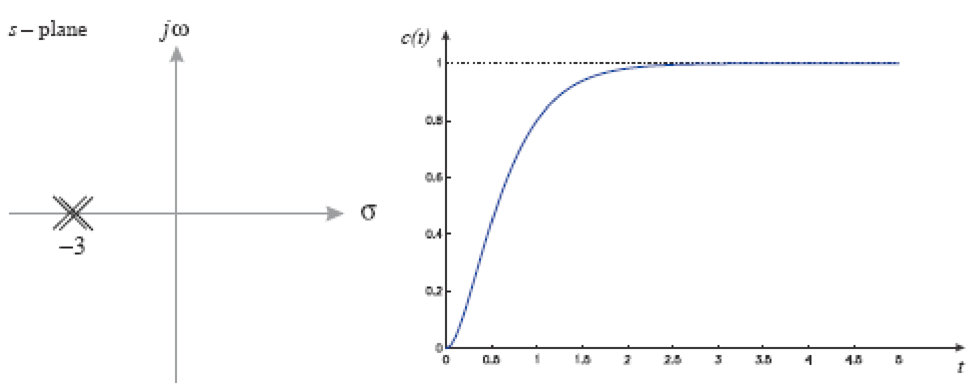


Figure 2.15: Graphical Representation of the Response of a Critically Damped System

3. **Overdamped case** $(\xi > 1)$

$$C(s) = \frac{\omega_n^{\ 2}}{s(s+s_1)(s+s_2)}$$

where $s_{1,2} = \left[\xi \pm \sqrt{\xi^2 - 1}\right]\omega_n$

$$c(t) = 1 + \frac{\omega_n}{2\sqrt{\xi^2 - 1}}\left(\frac{e^{-s_1 t}}{s_1} - \frac{e^{-s_2 t}}{s_2}\right) \qquad t \geq 0$$

If $\xi >> 1$, then $|s_2| << |s_1|$ and we can neglect $(-s_1)$ and

$$c(t) = 1 - e^{-\left(\xi - \sqrt{\xi^2 - 1}\right)\omega_n t} \qquad t \geq 0$$

Once the faster decaying exponential term has disappeared, the response is similar to that of a first order system.

Consider the second-order system

$$G(s) = \frac{b}{s^2 + as + b}$$

with $b = 9$ and a selected to illustrate each type of response.

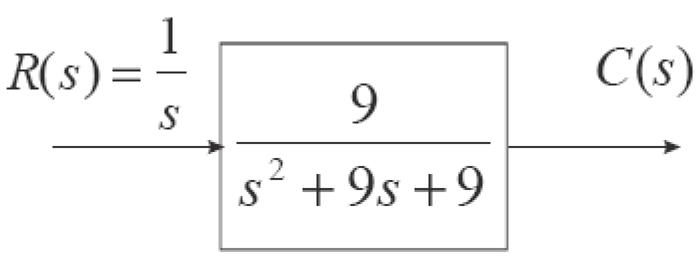


Figure 2.16: Block Diagram of an Overdamped System

IF $a = 9$ we have the system shown

The step response will be

$$C(s) = \frac{9}{s(s^2+9s+9)} = \frac{9}{s(s+7.854)(s+1.146)}.$$

The pole at $s = 0$ comes from the input (forced response) and there are two real poles at $s = -7.854$ and $s = -1.146$ respectively. Thus the step response will be of the form

$$c(t) = K_1 + K_2 e^{-7.854t} + K_3 e^{-1.146t}.$$

The response consists of the some of two first-order responses. The response due to the largest pole at $s = -7.854$ has a time constant of $\tau_{\text{fast}} = 0.127$ s. Left to itself, this pole would have a settling time of $T_{s_{\text{fast}}} = 0.51$ s. The response due to the smallest pole at $s = -1.146$ has a time constant of $\tau_{\text{slow}} = 0.873$ s. This pole has a settling time of $T_{s_{\text{slow}}} = 3.49$ s. Settling time $T_{s_{\text{slow}}}$ is about 4 times slower than $T_{s_{\text{fast}}}$. Thus the "slow pole" dominates the later stages of the response while the "fast pole" dominates the early part of the response. In addition, the initial slope is zero.

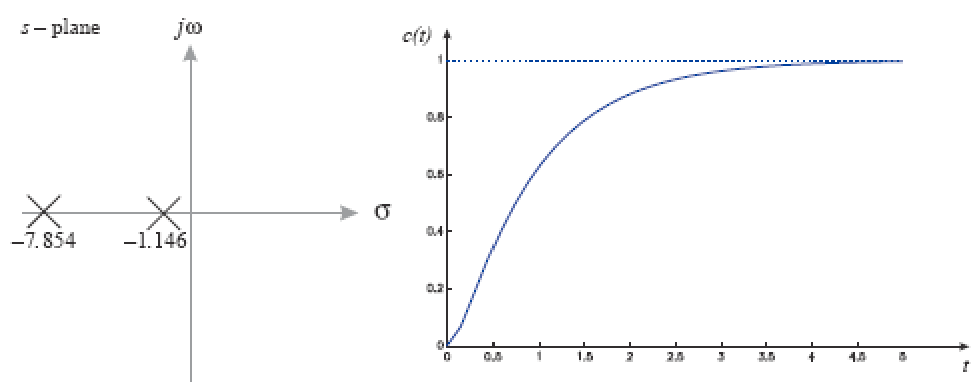


Figure 2.17: Graphical Diagram of the Response of an Over damped System

4. **Undamped Case**
Again, consider the second-order system

$$G(s) = \frac{b}{s^2 + as + b}$$

with $b = 9$ and a selected to illustrate each type of response.

IF $a = 0$ we have the system shown.

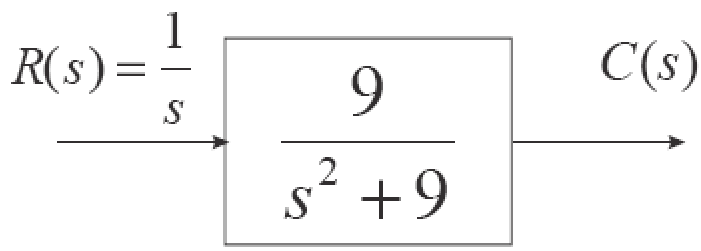


Figure 2.18: Block Diagram of an Undamped System

IF $a = 0$ we have the system shown.
The step response will be

$$C(s) = \frac{9}{s(s^2 + 9)}.$$

The pole at $s = 0$ comes from the input (forced response) and there are two imaginary poles at $s = \pm j3$. The step response of this system will be of the form

$$c(t) = K_1 + K_2 \cos 3t.$$

The frequency of oscillation of the undamped response is called the *natural frequency* ω_n. In this case the natural frequency $\omega_n = 3$ rad.s^{-1}.

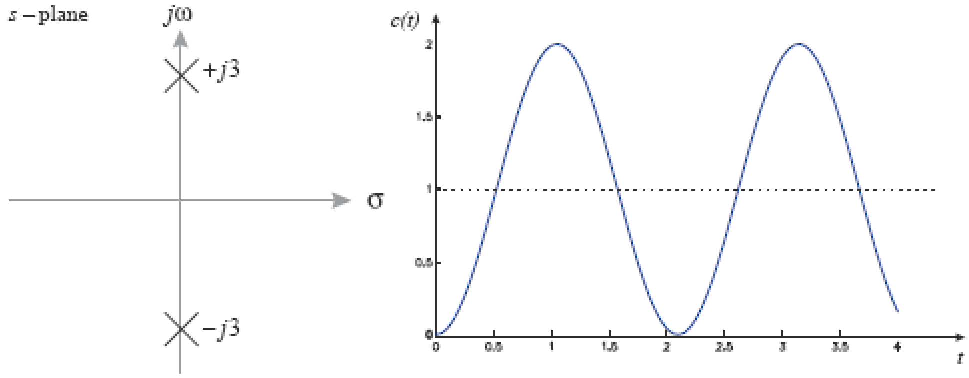


Figure 2.19: Graphical Representation of an Undamped Response

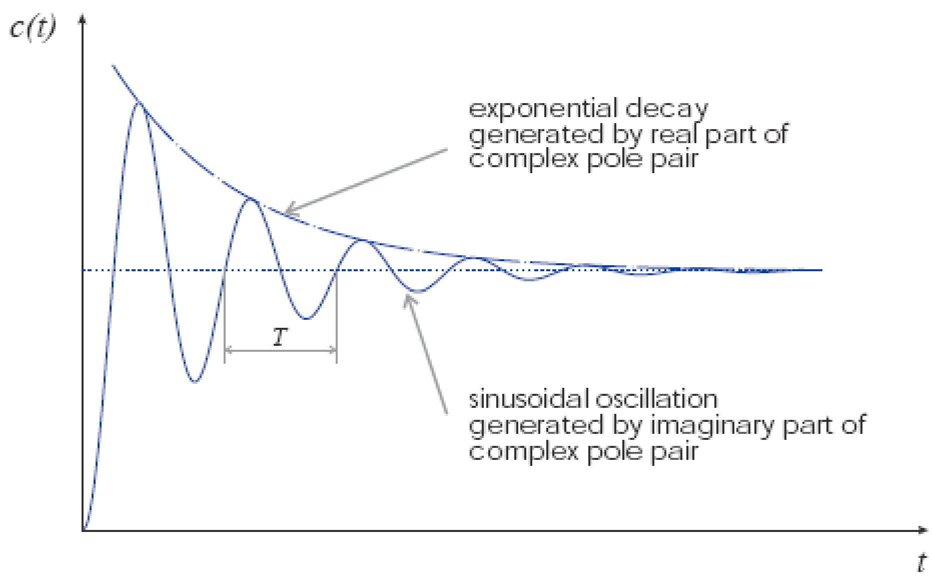


Figure 2.20: Graphical Representation of a Decaying Oscillation

2-2.11.2 The Specification of Second-Order Response

In order to specify a second-order response we need to define some performance measures based on a typical underdamped response curve. The specifications are:

Rise time T_r: the amount of time required to go from 10% to 90% of the final value.

Peak Time T_p: the time taken to reach the first, or maximum, peak.

$$\text{Peak time } T_p = \frac{\pi}{\omega_n} = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}}$$

Peak Overshoot, *M_p*

It is the amplitude of the first overshoot and steady-state for a unit step function input. It's a measure of relative stability. It is given by

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

Percent overshoot $\%OS$: the the amount that the waveform overshoots the steady state, or final value at the peak time, expressed as a percentage of the steady-state value. The percentage overshoot is thus thepeak overshoot expressed as a percentage of the steady-state value.

Settling time T_s: the amount of time required for the transient's damped oscillations to stay within $\pm 2\%$ of the final value.

$$T_s = \frac{4}{\zeta\omega_n} \text{ for 2\% error.}$$

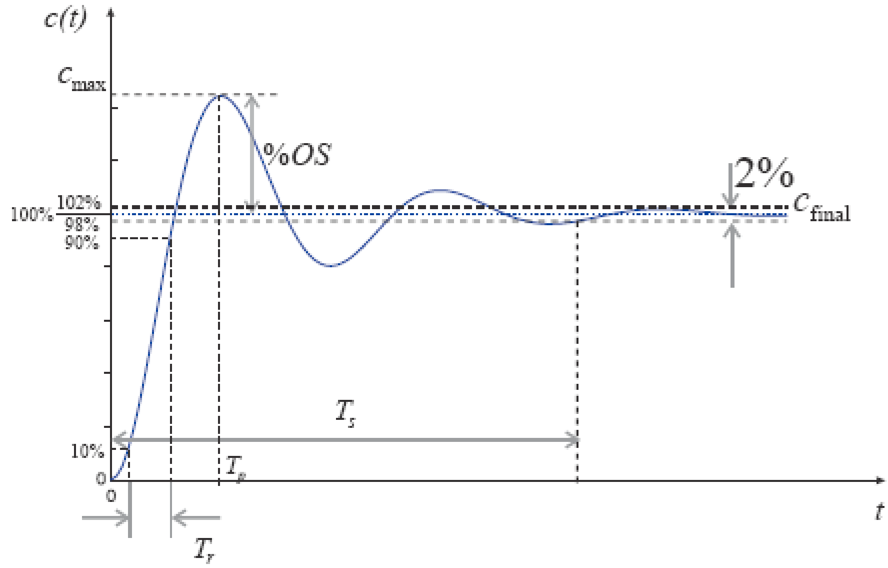


Figure 2.21: Graphical Representation of Second-Order Response Specifications

Rise time, settling time and peak time yield information about the speed and "quality" of the transient response. They can help the designer achieve a given speed of response without excessive overshoot or oscillations.

It should be noted that the last two specifications (T_r and T_s) are the same as those used for first-order systems and they may also be used for overdamped and critically damped second-order systems too. In fact, these specifications may also used for systems with order higher than two, provided that the response is of the same approximate shape. However, analytical formulae relating the time-response specification parameters to pole-and-zero locations can only be developed for second order systems.

UNIT TWO ASSIGNMENT

1. Describe the nature of the second-order system response via the value of the damping ration for the systems with transfer function:

$$
\begin{aligned}
G(s) &= \frac{12}{s^2+8s+12} \\
G(s) &= \frac{16}{s^2+8s+16} \\
G(s) &= \frac{20}{s^2+8s+20}.
\end{aligned}
$$

2.

For each of the second-order systems below, find ζ, ω_n, T_s, T_p, T_r, and $\%OS$.

i. $G(s) = 120/(s^2 + 12s + 120)$

ii. $G(s) = 0.01/(s^2 + 0.002s + 0.01)$

iii. $G(s) = 10^9/(s^2 + 6280s + 10^9)$

3.

Find the step response for each of the sytems with transfer function:

i. $G(s) = 5/(s + 5)$

ii. $G(s) = 20/(s + 20)$

Also find the time constant, rise time and settling time in each case.

4.

A unity feedback system has an open-loop transfer function

$$G(s) = \frac{1}{s(s+1)(s+2)}$$

Find the system error at time t= 4 sec, when at time t=0 an input $r(t) = 1 + 2t + 3t^2$ is applied to the system at rest.

UNIT 3

FEEDBACK CONTROL SYSTEMS AND STABILITY ANALYSIS

INTRODUCTION TO UNIT THREE

Automatic Control uses feedback to improve the performance of a large range of technological systems, from the steam engine to the space station.

Feedback control is used in radios (amplifiers), CD players (laser tracking), automobiles (cruise control, engines and suspensions), flight control (autopilots and stability augmentation), spacecraft (attitude control and guidance), machine tools, robots, power plants, materials processing, etc. In this unit, you will be introduced to the broad picture of control systems as well as some control system analysis tools and techniques.

In this unit, feed back control systems are introduced. This is followed by the stability analysis of systems using such techniques as the Routh-Hurwitz criterion, Root Locus method, Nyquist Criterion etc.

Learning objectives

1. To gain an understanding of feedback control systems and how they differ from open loop control systems.
2. To understand the use of the Routh-Hurwitz criterion and the Root Locus technique in determining the stability of a control systems,
3. To understand the use of frequency response techniques such as the Nyquist and Bode diagram to evaluate the stability of a control system.

SECTION 3-1: FEEDBACK CONTROL SYSTEMS

3-1.1 What is a feedback control system?

Feedback (closed) loop control system is shown below:

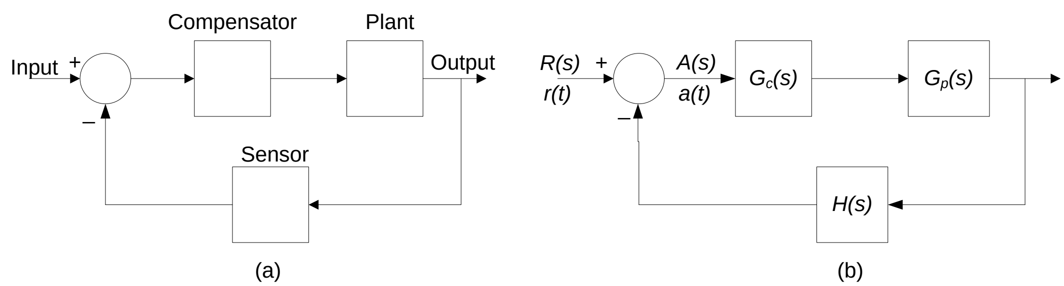


Figure 3.1: A Closed-loop Control System

The **plant** is the physical system, or process, to be controlled, and for the diagram above, it includes all the power amplifiers, actuators, gears, and so on. The **sensor** is the instrumentation that measures the output signal and converts this measurement to a usable signal at the summing junction. The **compensator** is a dynamic system purposely added to the loop to enhance the closed-loop system characteristics.

The transfer function of the system is given by:

$$T(s) = \frac{C(s)}{R(s)} = \frac{G_c(s)G_p(s)}{1+G_c(s)G_p(s)H(s)} \qquad (1)$$

For any single-input, single-output feedback control system as shown in fig 4.1, the following definitions hold:

1. The system will have a variable, $c(t)$, that we wish to control. This is known as the **controlled variable** or the **system output**.
2. The system will have an **input variable** or **system input**, $r(t)$, the value of which is a measure or indication of (but not necessarily equal to in magnitude or units) the **desired value of the system output**, $c_d(t)$. Note that $c_d(t)$ is not a variable that appears in the system.
3. The difference between the desired value of the output and the actual output is the **system error**, $e(t)$:

System error = e(t) = $c_d(t) - c(t)$

Again note that the system error is not a signal within the system.

3-1.2 Analysis and Design of Feedback Systems

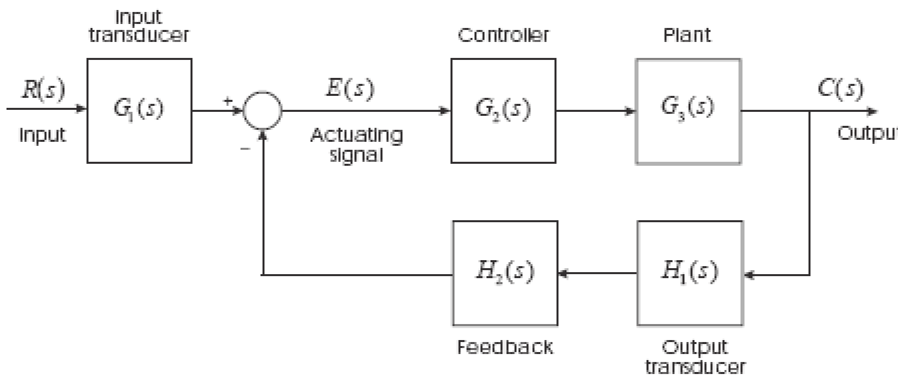


Figure 3.2: Feedback from a Control System Topology

Now, for the simplified system

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

The block diagram of this reduced "closed-loop" control system is shown below.

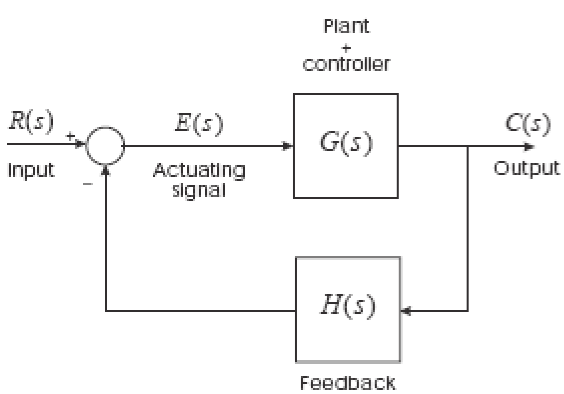


Figure 3.3: Simplified Feedback Control System Topology

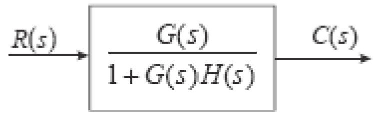


Figure 3.4: Reduced Feedback Control System

Interpretation of the generalized closed-loop transfer function

- The transfer function $G(s)H(s)$ is called the "*loop transfer function*".
- $1+G(s)H(s) = 0$ is called the "*closed-loop characteristic equation*" (CLCE). As we shall see, the CLCE is a very important equation in feedback control systems analysis and design.
- $G_c(s)$ is called the "*closed-loop transfer function*".

3-1.3 Unity-Gain Feedback

The "*unity-gain feedback*" canonical form is shown in the figure below

- $G_o(s)$ is called the "*open-loop transfer fuction*".

In comparison with the previous model, $H(s) = 1$, hence, we have

$$G_c(s) = \frac{G_o(s)}{1 + G_o(s)}$$

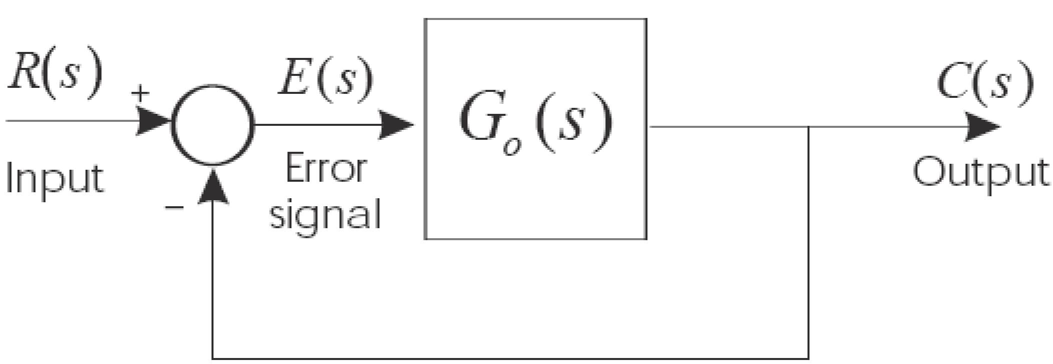


Figure 3.5: Unity-gain Feedback in Canonical Form

SECTION 3-2: STABILITY ANALYSIS AND DESIGN OF CONTROL SYSTEMS

3-2.1 Routh-Hurwitz Criterion

This criterion is an analytical procedure for determining if all roots of a polynomial have negative real parts. The limitation of this method is that it does not find the exact location of the roots.

The Routh-Hurwitz Criterion applies to a polynomial of the form

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

The coefficients of Q(s) are arranged to form the Routhian array, where the first two rows are the coefficients of the polynomial.

$$\begin{array}{l|lllll}
s^n & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \ldots \\
s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \ldots \\
s^{n-2} & b_1 & b_2 & b_3 & b_4 & \\
s^{n-3} & c_1 & c_2 & c_3 & c_4 & \ldots \\
. & . & . & & & \\
. & . & . & & & \\
. & . & . & & & \\
s^2 & k_1 & k_2 & & & \\
s^1 & l_1 & & & & \\
s^0 & m_1 & & & &
\end{array}$$

The equations for the coefficients of the array as follows:

$$b_1 = -\frac{1}{a_{n-1}} \begin{vmatrix} a_n & a_{n-2} \\ a_{n-1} & a_{n-3} \end{vmatrix}, \qquad b_2 = -\frac{1}{a_{n-1}} \begin{vmatrix} a_n & a_{n-4} \\ a_{n-1} & a_{n-5} \end{vmatrix}, \cdots$$

$$c_1 = -\frac{1}{b_1} \begin{vmatrix} a_{n-1} & a_{n-3} \\ b_1 & b_2 \end{vmatrix}, \qquad c_2 = -\frac{1}{b_1} \begin{vmatrix} a_{n-1} & a_{n-5} \\ b_1 & b_3 \end{vmatrix}, \ \ldots$$

The number of polynomial roots in the right half-plane is equal to the number of sign changes in the first column of the array. Again, if any coefficient other than a_0 is zero, there are imaginary roots or roots with positive real parts.

Example1.
The Routhian array of the polynomial

$$Q(s) = (s+2)(s^2 - s + 4) = s^3 + s^2 + 2s + 8$$

is given by:

s^3	1	2
s^2	1	8
s^1	-6	
s^0	8	

Since there are two sign changes on the first column (from 1 to -6 and -6 to 8), there are two roots of the polynomial in the right half-plane.

3-2.1.1 Special Cases

1. For this case, the first element in a row is zero, with at least one nonzero element in the same row. Either substitute a small positive number δ for the zero and then proceed to evaluate the rest of the array or substitute in the original equation $s = \frac{1}{x}$ and proceed.

Example 1:
Consider the polynomial

$$Q(s) = s^5 + 2s^4 + 2s^3 + 4s^2 + 11s + 10$$

The Routhian array is given as

s^5	1	2	11
s^4	2	4	10
s^3	$\cancel{0}\,\delta$	6	
s^2	$\frac{-12}{\delta}$	10	
s^1	6		
s^0	10		

From the array, there are two sign changes in the first column whether δ is assumed positive or negative. The number of sign changes in the first column is always independent of the assumed sign of δ, which leads to the conclusion that the system that falls under this case is always unstable.

2. In this case all the elements in a row of the Routhian array are zero. A zero row indicates there are roots which are negatives of each other. Here an auxiliary equation can be formed from the preceding row. An array can be completed by replacing the all-zero row by the coefficients obtained by differentiating the auxiliary equation. The roots of the auxiliary equation are also the roots of the original equation.

Example 2

$Q(s) = s^4 + 2s^3 + 11s^2 + 18s + 18 = 0$

The Routhian Array is

$$
\begin{array}{c|ccc}
s^4 & 1 & 11 & 18 \\
s^3 & \cancel{2}\ ^1 & \cancel{18}\ ^9 & \\
s^2 & \cancel{2}\ ^1 & \cancel{18}\ ^9 & \\
s^1 & 0 & &
\end{array}
$$

Form auxiliary equation from the s^2 row

$$s^2 + 9 = 0$$

$$s = \pm j3$$

Differentiating the equation gives

$$2s + 0 = 0$$

The array can then be completed as

$$
\begin{array}{c|c}
s^1 & 2 \\
s^0 & 9
\end{array}
$$

Example 3

$Q(s) = s^5 + s^4 + 3s^3 + 4s^2 + s + 2 = 0$

The Routhian Array is

$$
\begin{array}{c|ccc}
s^5 & 1 & 3 & 1 \\
s^4 & 1 & 4 & 2 \\
s^3 & -1 & -1 & \\
s^2 & 3 & 2 & \\
s^1 & -\frac{1}{3} & & \\
s^0 & 2 & &
\end{array}
$$

There are four sign changes.

Example 4

In a feedback system with

$$F(s)=\frac{P(s)}{Q(s)}=\frac{K(s+2)}{s(s+5)(s^2+2s+5)+K(s+2)}$$

The value of K is an adjustable parameter (+ve or –ve) and its value determines the stability of the system.

Here $$Q(s)=s^4+7s^3+15s^2+(25+K)s+2K=0$$

s^4	1	15	$2K$
s^3	7	$25+K$	
s^2	$80-K$	$14K$	
s^1	$\frac{(80-K)(25+K)-98K}{80-K}$		
s^0	$14K$		

From s^2 row $K<80$

From s^0 row $K>0$

From s^1 row $K=28.1,\qquad K=-71.1$

Condition for stability $0<K<28.1$

Routh's stability does not give the degree of stability, i.e. the amount of overshoot and the settling time of the controlled variable.

3-2.2 Root Locus Analysis And Design

Root locus is a powerful method of analyst and transient response. It is a graphical method for obtaining the roots of the system characteristic equation in *s* plane and thereby investigating the system response performance.

The root locus can be used to qualitatively describe the performance of a system as various parameters are changed. For example, the effect of gain on settling time and peak time can be vividly displayed. The root locus can also be used to assess stability, Ranges of stability, ranges of instability and the conditions under which a system begins to oscillate can all be investigated.

The root locus if a system is a plot of the roots of the characteristic equation of the closed-loop system as a function of the gain.

3-2.2.1 The Control System Problem.

Assume the general closed-loop transfer function shown in the block diagram below

Further, we assume that the loop-transfer function is $KG(s)H(s)$ and the poles and zeros of $G(s)$ and $H(s)$ are known. Note also that they do not change as a function of K.

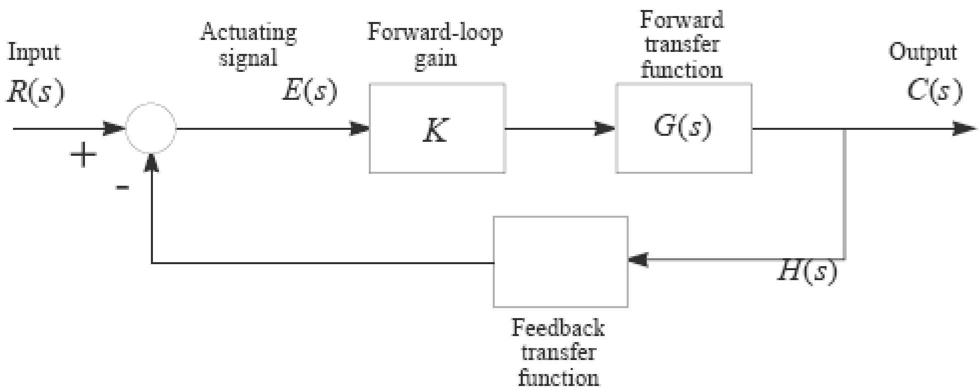


Figure 3.6 Closed-loop Control System

3-2.2.2 Closed-Loop Transfer Function

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

Where

$$G(s) = \frac{N_G(s)}{D_G(s)} \quad H(s) = \frac{N_H(s)}{D_H(s)}$$

hence

$$G_c(s) = \frac{KN_G(s)D_H(s)}{D_G(s)D_H(s) + KN_G(s)N_H(s)}$$

where $N_G(s)$, $D_H(s)$, $N_H(s)$ and $D_H(s)$ are factored numerator and denominator polynomials of plant and feedback respectively

Observations

- We know the zeros and poles of $G(s)$ and $H(s)$
- zeros of $G_c(s)$ consist of the zeros of $G(s)$ and the poles of $H(s)$
- The poles $G_c(s)$ are the roots of the closed-loop characteristic equation

$$D_G(s)D_H(s) + KN_G(s)N_H(s) = 0$$

and are not immediately known. In fact they vary as a function of the open-loop gain K.

Example Poles and Zeros of a Closed-Loop System *If*

$$G(s) = (s+1)/[(s(s+2)]$$

and

$$H(s) = (s+3)/(s+4)$$

what are the zeros and poles of the closed loop transfer function $G_c(s)$?

Solution: The poles of $G(s)$ are $s = 0, -2$, $H(s)$ has a single pole at $s = -4$. $G(s)$ has a single zero at $s = -1$ and $H(s)$ has a single zero at $s = -3$. The loop-transfer function, $KG(s)H(s)$ therefore has poles at $s = 0, -2, \text{ and } -4$ and zeros at $s = -1 \text{ and } -3$. Now

$$G_c(s) = \frac{K(s+1)(s+4)}{s^3 + (6+K)s^2 + (8+4K)s + 3K}$$

Thus the zeros of $G_c(s)$ consist of the zeros of $G(s)$ and the poles of $H(s)$. The poles of $G_c(s)$ are not immediately known without factoring the denominator, and they are a function of K Since a system's transient response and stability are dependent on the poles of $G_c(s)$ we have no knowledge of the system's performance until we factor the characteristic polynomial for specific values of K. The *root locus* will be used to give us a vivid picture of the poles of $G_c(s)$ as K varies.

3-2.2.3 Complex numbers and their vector representation

Any *complex number* $\sigma + j\omega$ can be graphically represented by a vector as shown in Slide 133. The complex number can also be described in polar form with magnitude M and angle θ as $M\angle\theta$ or $[Me^{j\theta}]$.

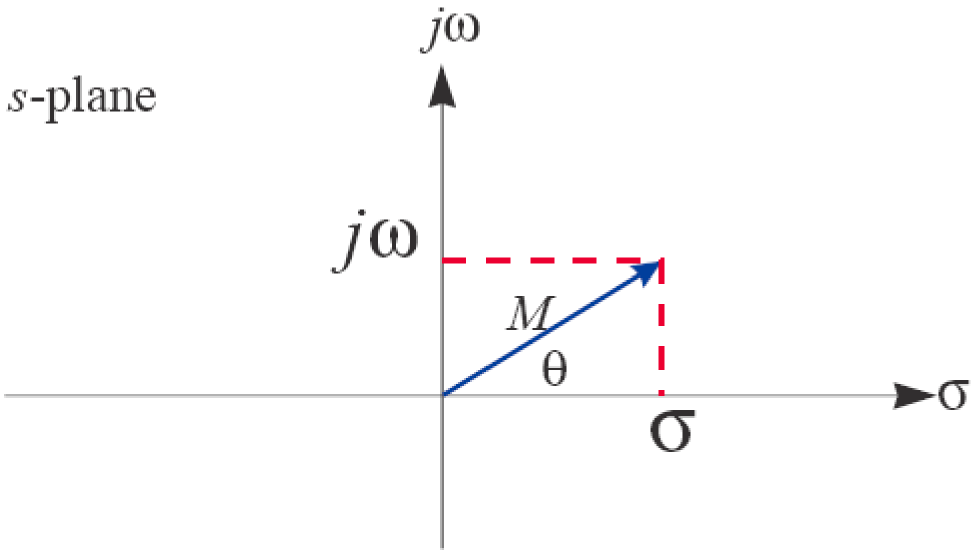
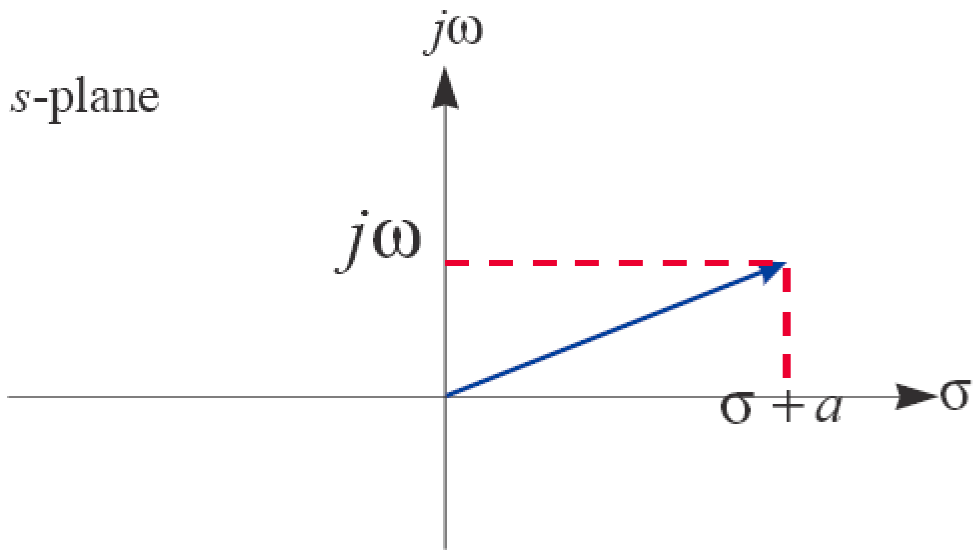


Figure 3.7:Vector Representation of s = σ+jω

If the complex number is substituted into a complex function $F(s)$, another complex number will result. For example, if $F(s) = (s + a)$, then substituting the complex number $s = \sigma + j\omega$ yields $F(s) = (\sigma + a) + j\omega$, another complex number. This number is shown in the figure below.

Notice that $F(s)$ has a zero at $s = -a$. If we translate the vector a units to the left, as in firgure, we have the alternative representation of a complex number that originates at the zero of $F(s)$ and terminates on the point $s = \sigma + j\omega$. Both vectors are equivalent since they both have the same lengths and direction.



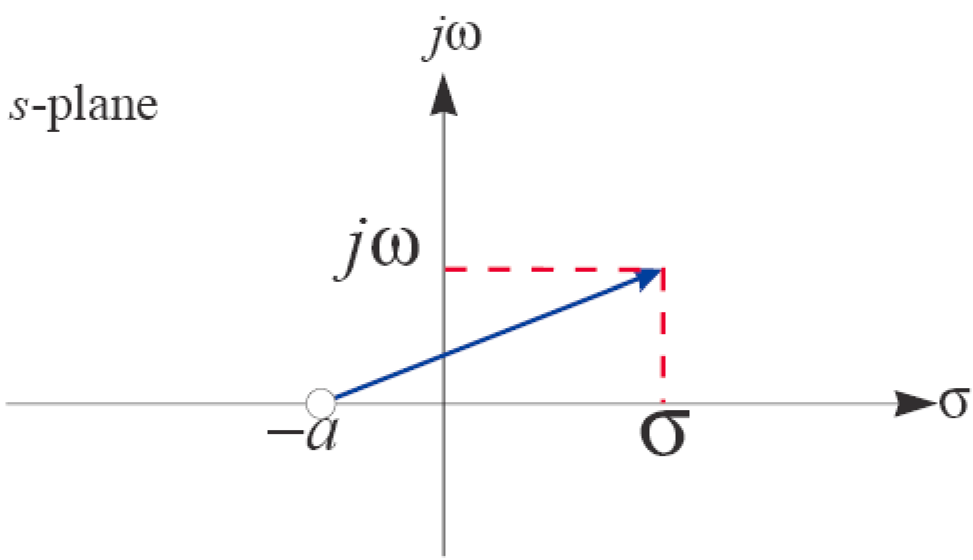


Figure 3.8: Alternate Representation of F(s)=(s+a)

Consider a complex function

$$
\begin{aligned}
F(s) &= \frac{\prod_{i=1}^{m}(s+z_i)}{\prod_{j=1}^{n}(s+p_j)} \\
&= \frac{\prod \text{ numerator's complex factors}}{\prod \text{ denominator's complex factors}}
\end{aligned}
$$

Then, the magnitude of F(s) is given as

$$\begin{aligned} M &= \frac{\prod \text{ zero lengths}}{\prod \text{ pole lengths}} \\ &= \frac{\prod_{i=1}^{m} |(s+z_i)|}{\prod_{j=1}^{n} |(s+p_j)|} \end{aligned}$$

And the angle of F(s) as

$$\begin{aligned} \theta &= \sum \text{ zero angles} - \sum \text{ pole angles} \\ &= \sum_{i=1}^{m} \angle(s+z_i) - \sum_{j=1}^{n} \angle(s+p_j) \end{aligned}$$

The symbol $\prod$ means "product," m is the number of zeros, and n is the number of poles. Each factor of the numerator $(s+z_i)$ and denominator $(s+p_j)$ is a complex number representable as a vector. The function defines the complex arithmetic to be performed to define $F(s)$ at any point s. Since every complex factor can be thought of as a vector, the magnitude, M, of $F(s)$ at any point $F(s)$ is equal to the product of the "zero lengths" divided by the product of the "pole lengths". A zero length $|(s+z_i)|$, is the magnitude of the vector drawn from the zero of $F(s)$ at $-z_i$ to the point s, and a pole length $|(s+p_j)|$, is the magnitude of the vector drawn from the pole of $F(s)$ at $-p_j$ to the point s. The zero angle is the angle, measured from the positive real axis, of a vector drawn from a zero of $F(s)$ at $-z_i$ to the point s, a pole angle is the angle, measured from the positive real axis, of the vector drawn from a pole of $F(s)$ at $-p_j$ to the point s.

Example Evaluation of a complex function via vectors *Given*

$$F(s) = \frac{(s+1)}{s(s+2)}$$

find $F(s)$ at the point $s = -3 + j4$

Solution: The problem is graphically depicted in Slide. 140, where each vector $(s + \alpha)$, of the function is shown terminating on the selected point $s = -3 + j4$. The vector originating at the zero at -1 is

$$\sqrt{20}\angle 116.57^\circ$$

The vector originating at the pole at the origin is

$$5\angle 126.87^\circ.$$

The vector originating at the pole at -2 is

$$\sqrt{17}\angle 104.04^\circ$$

Substituting these values into the equations for the magnitude M and the angle θ gives

$$M\angle\theta = \frac{\sqrt{20}}{5\sqrt{17}}\angle 116.57^\circ - 126.87^\circ - 104.04^\circ = 0.217\angle - 114.34^\circ$$

as a result of evaluating $F(s)$ at the point $s = -3 + j4$.

We are now ready to define the root locus

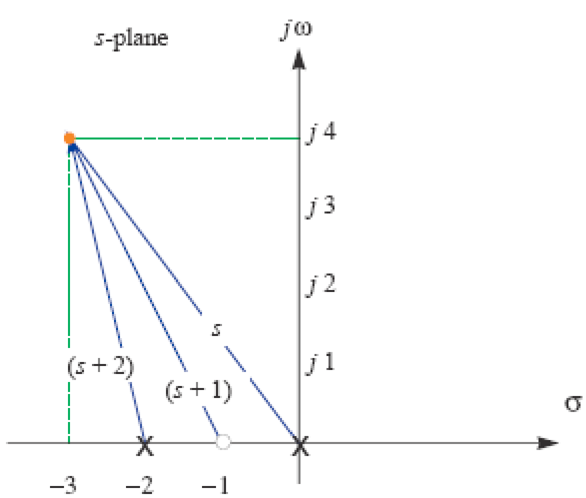


Figure 3.9: Vector Representation of $F(s) = (s+1)/[s(s+2)]\big|_{s=-3+j4}$

3-2.2.4 Plotting the Root Locus

Consider the position control system shown below

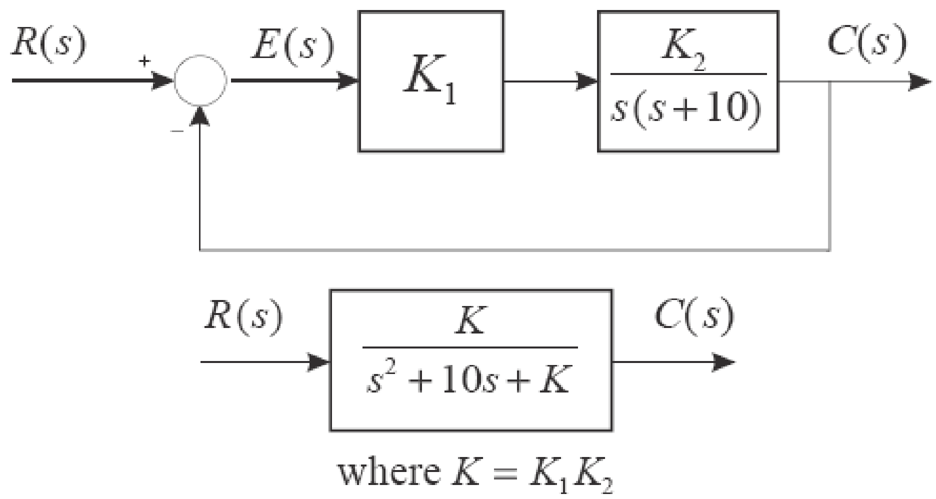


Figure 3.10: Block Diagram for a Position Control System

The closed-loop poles of the system change as the open-loop gain $K = K_1K_2$ is varied. The table below was formed by applying the quadratic formula to the closed-loop characteristic equation

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

Gain: K	**Pole 1**	**Pole 2**
0	0	−10.0000
5.0000	−9.4721	−0.5279
10.0000	−8.8730	−1.1270
15.0000	−8.1623	−1.8377
20.0000	−7.2361	−2.7639
25.0000	−5.0000	−5.0000
30.0000	$-5.0000 - j2.2361$	$-5.0000 + j2.2361$
35.0000	$-5.0000 - j3.1623$	$-5.0000 + j3.1623$
40.0000	$-5.0000 - j3.8730$	$-5.0000 + j3.8730$
45.0000	$-5.0000 - j4.4721$	$-5.0000 + j4.4721$
50.0000	$-5.0000 - j5.0000$	$-5.0000 + j5.0000$

Pole location as a function of K

Table 3.1

In the following figure this data is shown graphically: each pole and its gain is shown on the s-plane.

As the gain increases, the pole at $s = -10$ when $K = 0$, moves to the right along the real-axis, while the pole at $s = 0$ when $K = 0$, moves to the left. The poles meet at $s = -5$, when $K = 25$ and become complex. One pole moves upward along the line $s = -5 + j\omega$ and the other moves downward. In the next figure, we show the path of the poles for a continuous variation in K as solid directed lines[1]. It is this *representation of the path of the closed-loop poles as the gain is varied* that we call the *root locus*. For the rest of this discussion we limit ourselves to *positive* open-loop gain, $K \geq 0$.[2]

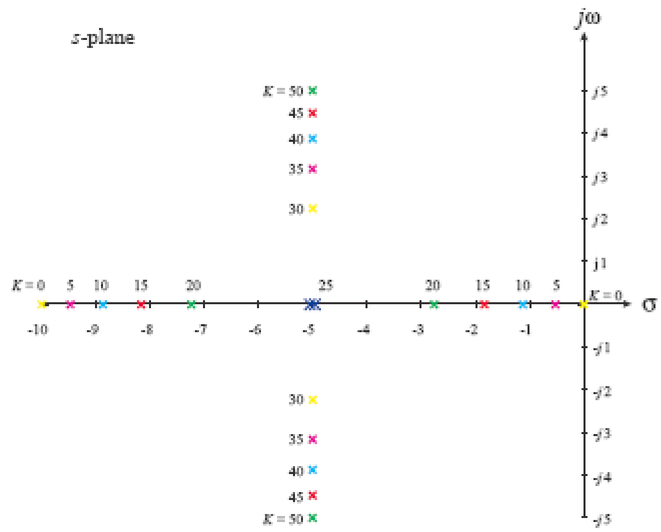


Figure 3.11: Plot of Values in Table 3.1

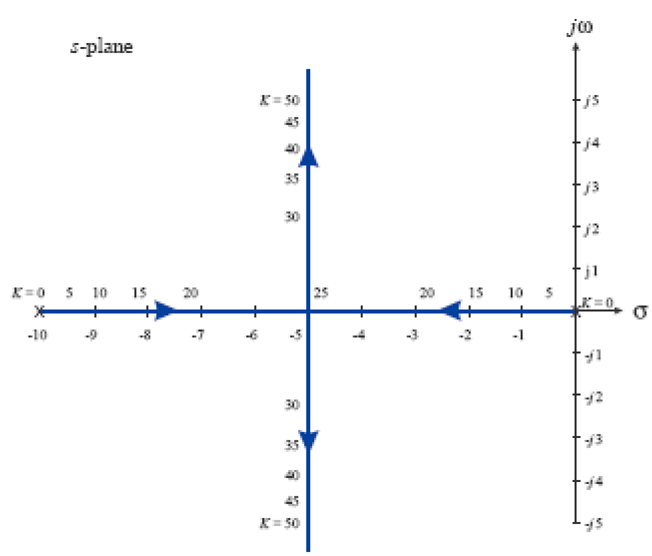


Figure 3.12: Plot for Continuous Variations in K

Commenting on the performance of the example system that the root-locus reveals we note the following

- The system has real-but-distinct poles for $0 < K < 25$ so the response is *over-damped*.
- There are two real-but-equal poles when $K = 25$ and the response is *critically damped*.
- The system has complex poles for $K > 25$ and the response is *under-damped* but we also note the following:
  - The real-part is constant ($\sigma_d = 5$) so the settling-time, which is inversely proportional to the σ_d is constant for all $K > 25$.
  - The damped natural frequency ω_d increases with K which means that the peak-time reduces as K increases (system becomes faster), but
  - at the same time the damping ratio reduces, so the peak-overshoot increases.
- Finally, because the root locus never crosses the imaginary axis, the system is never in a state or marginal stability (*undamped response*) and the closed-loop system is stable for all gain.

3-2.2.5 Rules for sketching the root locus

The root locus of a system could be *plotted* by sweeping through the s-plane and finding all points for which the angles added up to an odd multiple of $180°$. However this is tedious without computer-aid. Nonetheless the *angle criterion* can be used to *sketch* the root locus without the effort required to *plot* the root locus. Once a sketch is obtained it is only necessary to plot points of interest.

Rule 1: Number of Branches

The number of branches equals the number of closed-loop poles which in turn equals the number of open-loop poles. A branch is the path taken by a single closed-loop pole.

Rule 2: Symmetry

The closed-loop characteristic polynomial of any realisable system will have real coefficients. Any polynomial with real coefficients can only have roots that are themselves real, or complex. If the poles are complex, they will always be present as complex conjugate pairs. Thus, the root locus will be symmetrical about the real axis.

Rule 3: Real-axis Segments

The angle criterion can be used to determine where, if anywhere, the closed-loop poles will be located on the real-axis. To illustrate, consider the figure below which shows the poles and zeros of a general open-loop system.

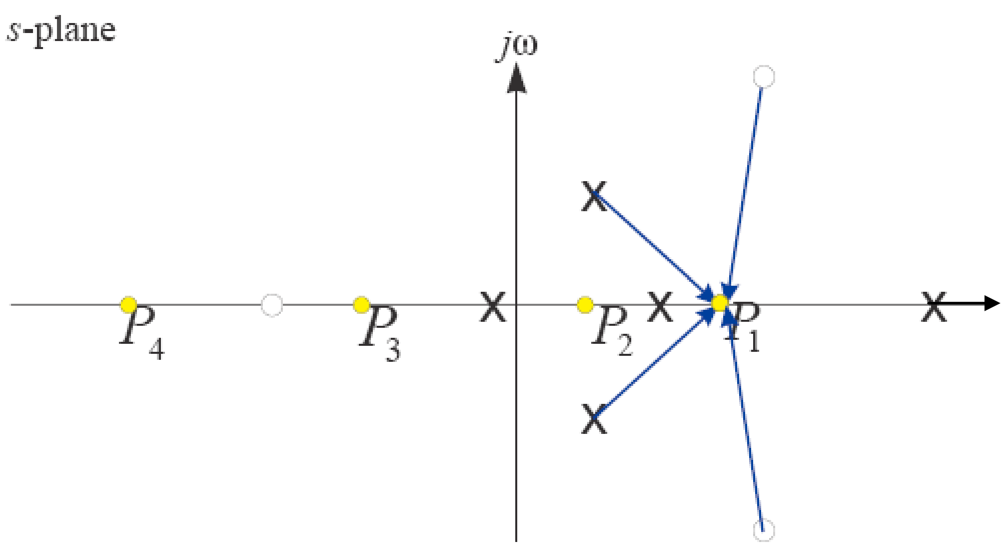


Figure 3.13: Poles and Zeros of a General Open-Loop System

If we calculate the angular contribution of each point P_1, P_2, P_3 and P_4, along the real axis the following can be observed.

- At each point, the angular contribution of each open-loop complex pole or zero pair is zero (the positive angle above the real axis is cancelled by an equal negative angle below), and
- the contribution of the open-loop real poles or zeros to the left of the respective point is zero.

Thus only open-loop poles or zeros to the right of the respective point contribute to the angle criterion. If we then calculate the contribution at each point using only the open-loop poles and zeros to the right we notice:

- The angle alternates between $0°$ and $180°$, and
- The angle is $180°$ for regions of the real axis that exist to the right of an odd-number of poles and/or zeros.

Summarizing these findings in a rule, we have

> The root locus on the real axis lies to the left of an odd-number of open-loop real-axis poles and/or finite open-loop zeros.

Rule 4: Start and End- Points

The root locus begins at the open-loop poles of $G(s)H(s)$ and ends at the finite and infinite zeros of $G(s)H(s)$.

Rule 5: Asymptotic Behaviour

Consider a system with transfer function

$$KG(s)H(s) = \frac{K}{s(s+1)(s+2)}$$

For this system, there are three finite poles $s = 0, -1, -2$ and no finite zeros. How could we apply rule 4 to this case? We get round the problem in typical mathematical style by inventing the concept of *infinite* poles and zeros. If a function $\rightarrow \infty$ as $s \rightarrow \infty$ we say that it *as a pole at ininity*. Similarly, if a function $\rightarrow 0$ as $s \rightarrow \infty$ we say that the function *as a zero at infinity*. For example, the transfer function $G(s) = s$ has a pole at infinity as $G(s)|_{s\rightarrow\infty} \rightarrow \infty$. Similarly the function $G(s) = 1/s$ has a zero at infinity because $G(s)|_{s\rightarrow\infty} \rightarrow 0$.

Every transfer function has the same number of poles as zeros, if we include the infinite zeros. Thus, the transfer function in the equation above contains three finite poles and finite zeros.

To illustrate, let $s \rightarrow \infty$ then Equation (x) becomes

$$KG(s)H(s) \approx \frac{K}{s^3} = \frac{K}{s \cdot s \cdot s} \qquad \text{(y)}$$

Each s in the denominator causes the loop transfer-function, $KG(s)H(s)$ to become zero as s approaches infinity. Hence, Equation (y) has three zeros at infinity.

Thus, for Equation (x), the root locus begins at the finite poles of $KG\ \ (s)H(s)$ and ends up at the infinite zeros. But the question remains: where are the infinite zeros? We need to know in order to show the root locus moving from the poles to the infinite zeros. Rule 5 helps to locate these zeros at infinity.

The root locus approaches straight lines as asymptotes at the locus approaches infinity. Further, the equation of the aymptotes is given by the real-axis intercept, σ_0, and angle, θ, as follows:

$$\sigma_0 = \frac{\sum \text{finite poles} - \sum \text{finite zeros}}{n-m}$$

$$\theta = \frac{(2r+1)180°}{n-m}$$

where n is the number of finite poles, m is the number of finite zeros (and, therefore $n-m$ is the number of infinite zeros), $r = 0, 1, 2, \ldots$, and the angle is given in degrees relative to the direction of the positive real axis.

Summary of the Rules

- **Rule 1**: *Number of branches* is equal to the number of closed-loop poles.
- **Rule 2**: *Symmetry*—the root locus is symmetrical about the real axis.
- **Rule 3**: *Real-axis segments* a point on the real axis is on the root locus if there are an odd number of open-loop poles and/or finite zeros to the right
- **Rule 4**: *Start and end-points* the root locus starts at the n open-loop poles and ends at the m open-loop zeros and $m-n$ infinite zeros of the system.
- **Rule 5**: *Asymptotic behaviour* The root locus approaches the infinite open-loop zeros along asymptotes. The asymptotes are straight lines centred on the real axis at $\sigma_0 = (\sum \text{open} - \text{loop poles} - \sum \text{open} - \text{loop zeros})/(n-m)$ and the angles are given by $\theta = (2r+1)180°/(n-m), \ r = 0, 1, 2, \ldots$

3-2.2.6 Refining the Root Locus Diagram

Real-axis break-away and break-in points

Consider a system with transfer function

$$KG(s)H(s) = \frac{K(s-3)(s-5)}{(s+1)(s+2)}.$$

The root locus breaks-away from the real axis somewhere in the region $-2 < \sigma < -1$ and breaks-in to the real axis somewhere between $3 < \sigma < 5$. There are several ways to determine exactly where these break-in and break-out points occur.

The most commonly described method is to use the fact that the gain K reaches a maximum at the point of a break-away, and is at a minimum at the point of a break-in. The argument for this is based on the fact that the gain increases as the closed-loop poles move way from the open-loop poles and towards the open-loop zeros. Thus, the gain at a break-away point is a maximum for poles on the real axis (greater gains causes the poles to become complex). At a break-in point, the gain is greater as the poles move away from the break-in point towards the zeros, so the gain is a minimum at the point of break-in.

Because the gain is a minimum or a maximum, then we can use differentiation to determine the location of the break-in and break-away points.

Calculation of $j\omega$-Axis crossing

If the root locus crosses the imaginary axis into the right-half plane of the s-plane, the system will become unstable for the value of K for which the poles are on the imaginary axis. Therefore, it is important to be able to evaluate the point of the $j\omega$ crossing. This can be done by use of the Routh-Hurwitz criterion.

Example ***Frequency and gain at imaginary axis crossing.*** *For the previous example which had*

$$KG(s)H(s) = \frac{K(s+3)}{s(s+1)(s+2)(s+4)}$$

determine the value of K and the corresponding frequency ω of the imaginary axis crossing point.

Solution: The closed-loop characteristic equation is

$$1+KG(s)H(s) = 1+\frac{K(s+3)}{s(s+1)(s+2)(s+4)} = s^4+7s^3+14s^2+(8+K)s+3K = 0$$

The Routh table for this system is shown below.

$$
\begin{aligned}
KG(s)H(s) &= \frac{K(s+3)}{s(s+1)(s+2)(s+4)} \\
G_c(s) &= \frac{K(s+3)}{s^4+7s^3+14s^2+(8+K)s+3K}
\end{aligned}
$$

s^4	1	14	$3K$
s^3	7	$8+K$	0
s^2	$90-K$	$21K$	0
s^1	$\frac{-K^2-65K+720}{90-K}$	0	0
s^0	$21K$	0	0

Row s^1 is the most important from the point of view of stability. This is a zero row when $K = 9.65$. The corresponding value of ω can be determined by use of the rule for dealing with a row of zeros in the Routh array[1]. However, since this rule is not discussed in detail during lectures, we shall find the imaginary axis crossing as follows. Substitute $K = 9.65$ into the closed-loop characteristic equation giving

$$s^4 + 7s^3 + 14s^2 + 17.65s + 28.95.$$

Now substitute $s = j\omega$ into this equation to give

$$\omega^4 - 7j\omega^3 - 14\omega^2 + 17.65j\omega + 28.95 = 0.$$

Equating real and imaginary parts gives

$$
\begin{aligned}
\omega^4 - 14\omega^2 + 28.95 &= 0 \\
-7\omega^3 - 17.65\omega &= 0
\end{aligned}
$$

giving

$$\omega = \sqrt{17.65/7} = 1.59.$$

Thus the imaginary axis crossing occurs when $\omega = 1.59$ rad/s.

Note that as a result of this analysis, we can also state that the system is stable for $0 \le K < 9.65$.

3-2.2.7 Angles of Arrival and Departure

If the open-loop system has complex poles or zeros, then the sketch can be further refined by calculating the angles of departure from the poles and arrival at the zeros. The method of doing this is to take a small excursion from one of the complex poles (or zeros) and use the angle criterion with the remaining poles and zeros to determinine the angle of departure (or arrival) that balances the angle criterion. The departure (arrival) from the conjugate pole (zero) will be the negative of the angle calculated above, because of symmetry.

Example:

Now let us consider the following example.

Root Locus plot for

$$GH = \frac{K}{s(s+1)(s^2+4s+5)}$$

$$GH = \frac{K}{s(s+1)(s+2+j)(s+2-j)}$$

The poles are at 0, -1, -2 ± j

1. <u>Number of Loci</u>
   Number of branches of root locus = number of poles of *GH(s)* = 4.

2. <u>Starting and Ending Points</u>
   Branches start (K = 0) on open loop poles of *GH(s)* and end ($K = \infty$) on open loop zeros or infinity.

3. <u>Loci on Real Axis</u>
   For $K > 0$ locus exists on the real axis which lies to the left of an odd number of poles and zeros.

   For $K < 0$ the locus exists on the real axis which lie to the left of an even number of poles and zeros (ignore complex roots).

4. <u>Asymptotes to the root Loci</u>
   As K → ∞, the branches approach asymptotes at angles of

$$\gamma = \frac{(1+2m)180}{n_P - n_Z} \qquad \text{For } K > 0$$

$$\gamma = \frac{360m}{n_P - n_Z} \qquad \text{For } K < 0$$

Example

$$\gamma = \frac{(1+2m)180}{4} = \pm 45, \pm 135$$

5. Intersection of Asymptotes

Asymptotes intercept the real axis at the "centre of gravity" of the poles and zeros given by

$$\sigma_c = \frac{\sum_{i-1}^{n_P} \mathrm{Re}(p_i) - \sum_{i=1}^{n_Z} \mathrm{Re}(z_i)}{n_P - n_Z}$$

Here

$$\sum \mathrm{Re}(p_i) = 0 - 1 - 2 - 2 = -5$$

$$\sum \mathrm{Re}(z_i) = 0$$

$$\sigma_c = -5/4 = -1.25$$

6. Angles of Departure and of Arrival

Angle of departure from a complex pole

For $K > 0$,

$$\phi_{d+} = (2m+1)180 + \sum \text{ angles from zeros - } \sum \text{ angles from other poles}$$

For $K < 0$,

$$\phi_{d-} = \phi_{d+} - 180$$

Angle of approach to a complex zero is

For $K > 0$,

$$\psi_{a+} = \sum \text{ angles from poles - } \sum \text{ angles from other zeros - } (2m+1)180$$

For $K < 0$,

$$\psi_{a-} = \psi_{a+} - 180$$

Here

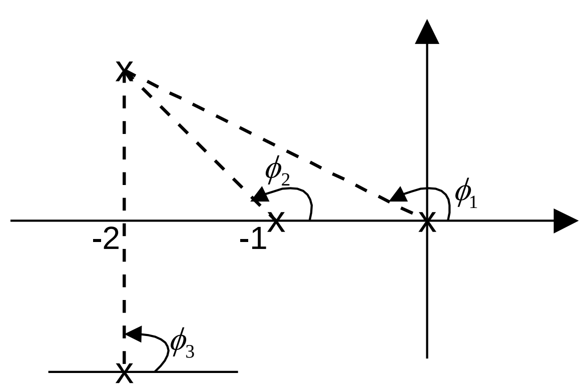
$$\phi_{d+} = (2m+1)180 - \phi_1 - \phi_2 - \phi_3$$

$$\phi_1 = 180 - \tan^{-1} \tfrac{1}{2} = 180 - 26.6^\circ = 153.4^\circ$$

$$\phi_2 = 180 - \tan^{-1} 1 = 135^\circ$$

$$\phi_3 = 90^\circ$$

$$\phi_{d+} = 540 - 378.4 = 161.6$$



7. Branches leave (or strike) the real axis at right angles

8. Intersection with the imaginary axis

Imaginary axis crossing point are obtained from

$$F(s) = \frac{P(s)}{Q(s)} = \frac{K}{s(s+1)(s^2+4s+5)+K}$$

$$\therefore Q(s) = s^4 + 5s^3 + 9s^2 + 5s + K$$

s^4	1	9	K
s^3	$\cancel{5}\ 1$	$\cancel{5}\ 1$	
s^2	8	K	
s^1	$\frac{8-K}{K}$		
s^0	K		

For stability $0 < K < 8$

Imaginary roots occur when

$$8 - K = 0$$
$$\therefore \quad K = 8$$

Auxiliary equation is:

$$8s^2 + K = 0$$

$$\therefore s = \pm j\sqrt{\frac{K}{8}} = \pm j$$

9. Departure from real axis

If we define $G(s)H(s) = \dfrac{K}{\omega(s)} = -1$, then the break-away and break-in points on the real axis can easily be calculated for a system where $\dfrac{d\omega(s)}{ds}$ is of the second order.

Break-away from real axis is at maximum or minimum of K.

$$\sum_{i=1}^{n_P} \frac{1}{\delta_b + P_i} = \sum_{i=1}^{n_Z} \frac{1}{\delta_b + Z_i}$$

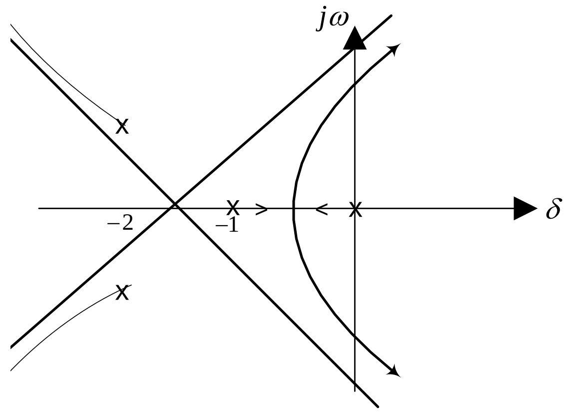


Figure 3.14: Sketch of the Root Locus for $GH = \dfrac{K}{s(s+1)(s^2+4s+5)}$

3-2.3 Frequency Response Techniques

The root locus method, and methods of analysis and design which require knowledge of the open-loop pole of the system are known as time response methods. An alternative view point, which is equally relevant to linear systems defined by transfer functions are the group of frequency response methods.

These consider the behaviour of systems as revealed by, usually sinusoidal (or harmonic) response measurements. Stability, steady-state performance and transient response can all be determined from frequency response measurements of the plant and actuators, and knowledge of the open-loop frequency response can guide the design of the closed-loop system.

Advantages of Frequency Response Methods

- Frequency response data is easier to obtain experimentally.
- Frequency response methods can be used if a model of the plant and actuators are difficult to obtain.
- Frequency response methods can be used for systems with time-delays.
- Compensators can be simpler to design and can be designed if only experimental data about the system is available.
- Frequency response methods can be used to determine properties, such as the existence of limit cycles and stability, of nonlinear systems.

Frequency Response is Magnitude versus Frequency for the Control Ratio

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

3-2.4 Bode Diagrams

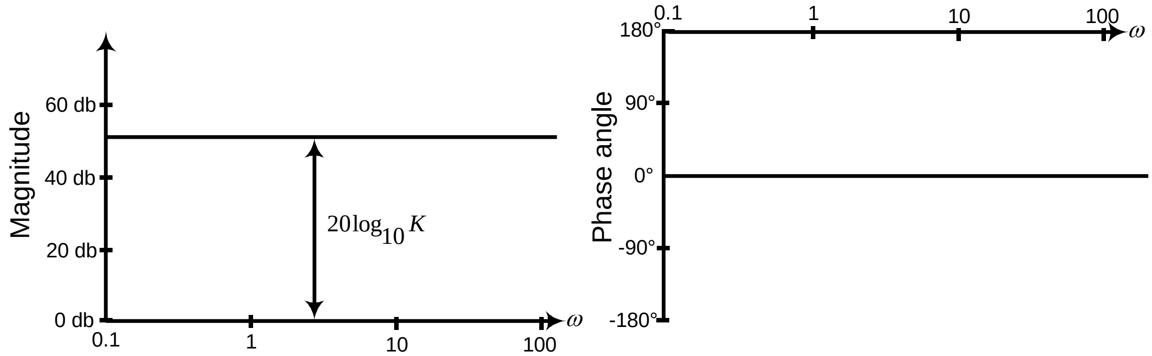
The bode plot consists of two graphs in rectangular co-ordinates. These are:

i. magnitude of $G(j\omega)H(j\omega)$ in db versus $\log_{10}\omega$.

ii. phase angle of $G(j\omega)H(j\omega)$ versus $\log_{10}\omega$.

1. Bode plot for gain term: K

$$K|_{db} = 20\log_{10} K$$
$$\angle K = 0°$$



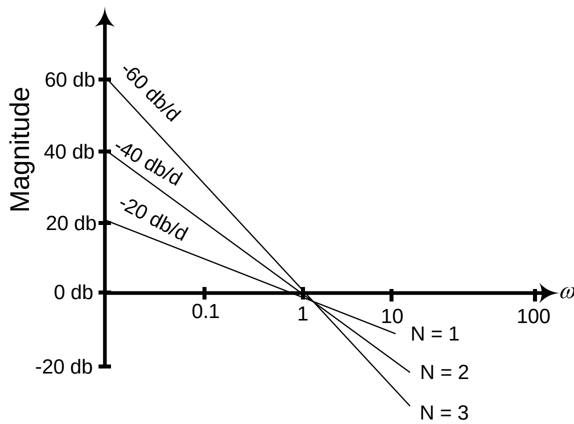
2. Bode plot for term: $\dfrac{1}{(j\omega)^N}$

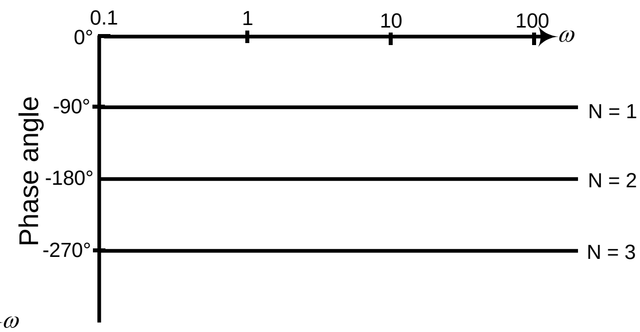
The magnitude in db is

$$20\log_{10}\left|\frac{1}{(j\omega)^N}\right| = -20\log N_{10}\omega$$

And

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





3. Bode plot for term: $(1 + j\omega T)$

The magnitude in db is

$$20\log_{10}|1 + j\omega T| = 20\log_{10}\sqrt{1+\omega^2T^2}$$

The following approximations are made

i. $\omega T << 1$ $\therefore 20\log_{10}|1 + j\omega T| = 20\log_{10}1 = 0db$

ii. $\omega T >> 1$ $\therefore 20\log_{10}|1 + j\omega T| = 20\log_{10}\omega T$

The bode plot consists of two portions

1. A line on 0 db axis up to $\omega = \frac{1}{T}$
2. A line having a slope of +20 db/decade and intersects 0 db axis at $20\log_{10}\omega T = 0$

The intersection point is $\omega = \frac{1}{T}$.

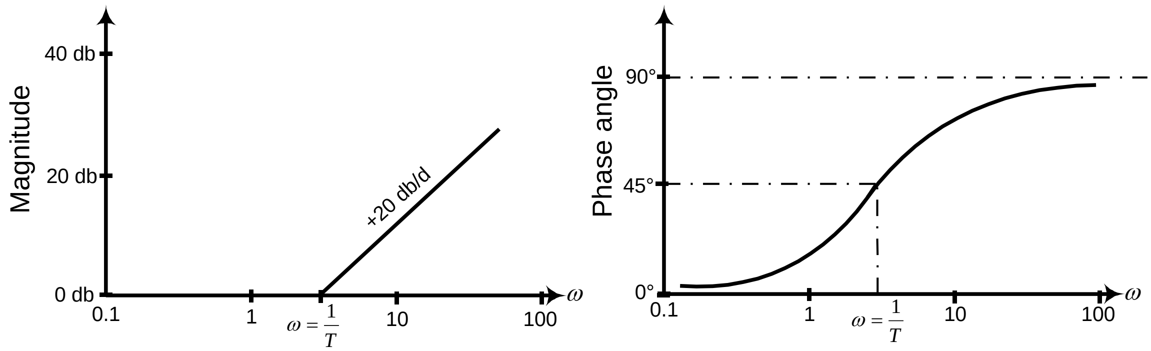
The phase angle is given by

$$\phi = \tan^{-1}\left(\frac{\omega T}{1}\right)$$

✓ At very low values of ω, $\phi \to 0$

✓ At $\omega = \frac{1}{T}$, $\phi = \tan^{-1}1 = 45°$

✓ At very low values of ω, $\phi \to 90°$



4. Bode plot for term: $\dfrac{1}{(1 + j\omega T)}$

The magnitude in db is

$$20\log_{10}\left|\frac{1}{1+j\omega T}\right| = -20\log_{10}\sqrt{1+\omega^2 T^2}$$

The following approximations are made

i. $\omega T << 1$ $\quad \therefore 20\log_{10}\left|\dfrac{1}{1+j\omega T}\right| = -20\log_{10} 1 = 0db$

ii. $\omega T >> 1$ $\quad \therefore 20\log_{10}\left|\dfrac{1}{1+j\omega T}\right| = -20\log_{10}\omega T$

The bode plot consists of two portions

1. A line on 0 db axis up to $\omega = \frac{1}{T}$
2. A line having a slope of -20 db/decade and intersects 0 db axis at $-20\log_{10}\omega T = 0$

The intersection point is $\omega = \frac{1}{T}$.

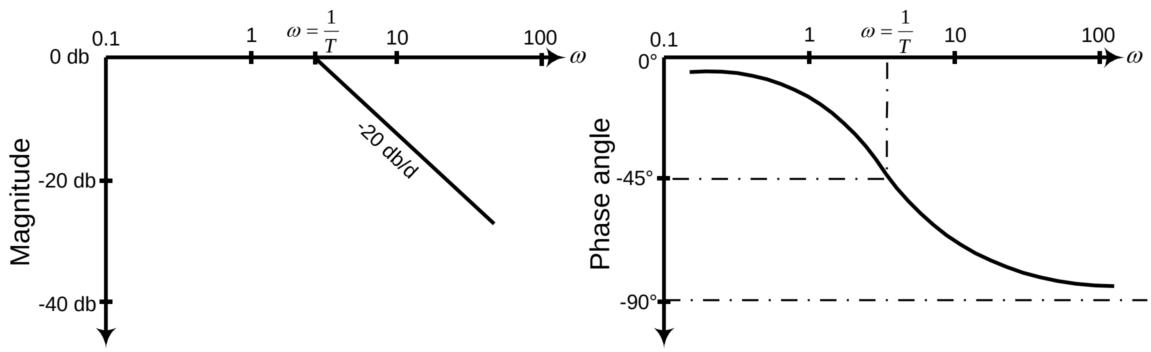
The phase angle is given by

$$\phi = -\tan^{-1}\left(\frac{\omega T}{1}\right)$$

✓ At very low values of ω, $\phi \to 0$

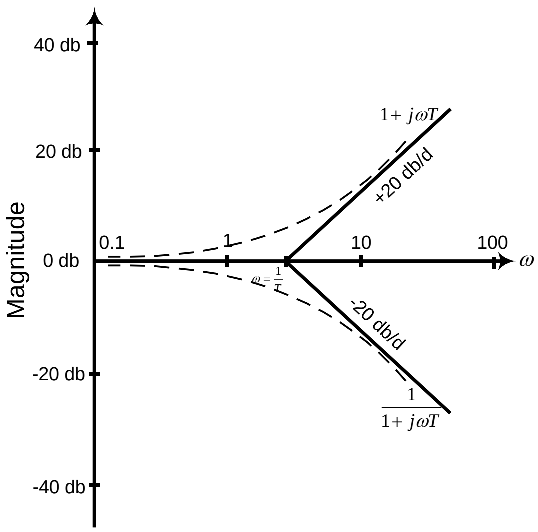
✓ At $\omega = \frac{1}{T}$, $\phi = -\tan^{-1} 1 = -45°$

✓ At very low values of ω, $\phi \to -90°$



Error due to approximation:

The magnitude Bode plot for $(1+j\omega T)$ and $\dfrac{1}{(1+j\omega T)}$ are asymptotic plots. The two plots match each other at lower and higher values of frequencies w.r.t. corner frequency $\omega = \frac{1}{T}$. At $\omega = \frac{1}{T}$ maximum error is noted.



At $\omega = \frac{1}{T}$ magnitude is approximated as 0 b exact magnitude is:

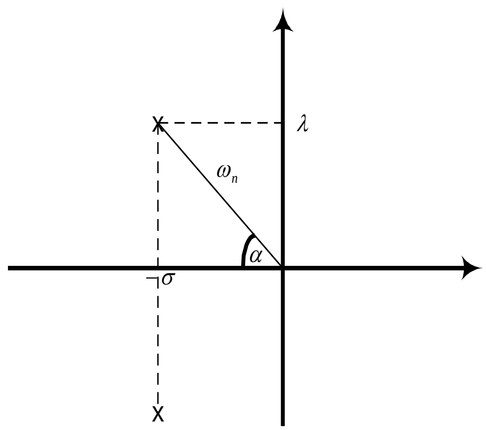
$$20\log_{10}\left|\left(1+j\frac{1}{T}T\right)\right| = 20\log_{10}\sqrt{\left(1+\frac{1}{T^2}T^2\right)}$$
$$= 20\log_{10}\sqrt{2}$$
$$= 3db$$

Therefore, the maximum error due to approximation is 3 db.

Similarly error due to term $\dfrac{1}{(1+j\omega T)}$ at $\omega = \frac{1}{T}$ is -3 db.

5. Bode plot for the quadratic term

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



For K=1, the general form of a complex pole function is

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

We define poles as

$$p=-\sigma+j\lambda \qquad \bar{p}=-\sigma-j\lambda$$

$$(s-p)\left(s+\bar{p}\right)=s^2+2\sigma s+\left(\sigma^2+\lambda^2\right)$$

comparing with the characteristic polynomial

$$\omega_n^{\ 2}=\sigma^2+\lambda^2$$

$$\sigma=\xi\omega_n$$

where $\xi=\dfrac{\sigma}{\omega_n}$ = damping ratio

$$=\frac{\sigma}{\sqrt{\sigma^2+\lambda^2}}$$

$$=\cos\alpha$$

ω_n = undamped natural frequency

ω_d =damped natural frequency

$$=\omega_n\sqrt{1-\xi^2}$$

If s = jω then the general complex pole is given by

$$G(j\omega)=\frac{1}{1-\left(\dfrac{\omega}{\omega_n}\right)^2+j2\dfrac{\xi\omega}{\omega_n}}$$

The magnitude in db is

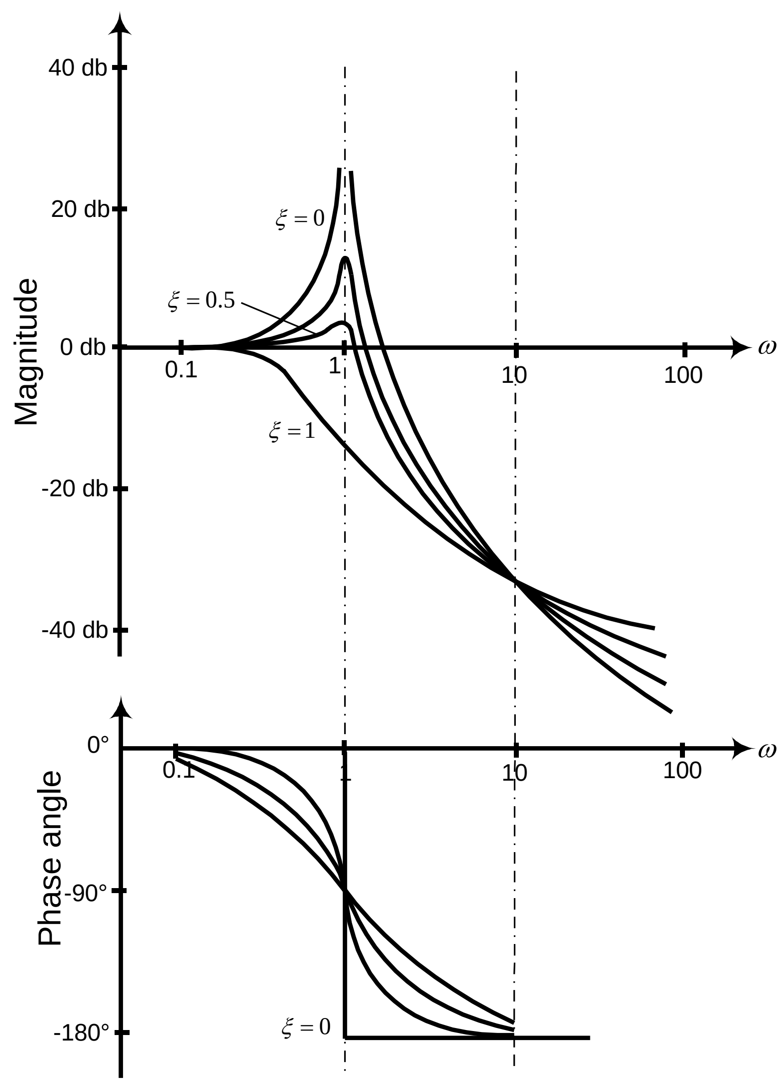
$$=-20\log_{10}\left[\left\{1-\left(\frac{\omega}{\omega_n}\right)^2\right\}^2+\left(\frac{2\xi\omega}{\omega_n}\right)^2\right]^{1/2}$$

For ω << 1 $\qquad L_m\ G=0$

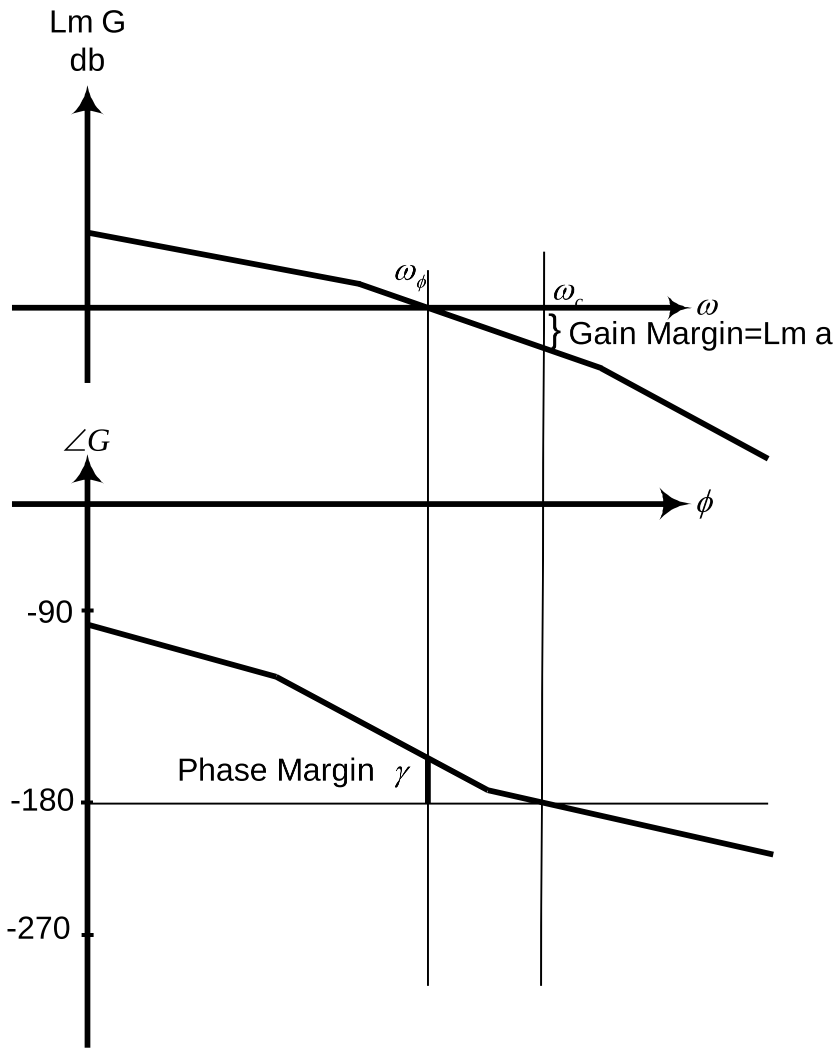
ω >> 1 $\qquad L_m\ G=-20\log_{10}\left(\dfrac{\omega}{\omega_n}\right)^2=-40\log_{10}\dfrac{\omega}{\omega_n}$

The phase angle is given by

$$\phi = -\tan^{-1}\frac{2\xi\dfrac{\omega}{\omega_n}}{1-\dfrac{\omega^2}{\omega_n{}^2}}$$



3-2.4.1 Frequency Domain Specifications



- **Gain Crossover**: Frequency ω_ϕ at which $|G(s)H(s)| = 1$
- **Phase Margin**: Amount of phase shift at the frequency ω_ϕ that would just produce instability. It is designated by $\gamma = 180 + \phi$ where ϕ is negative.
- **Phase Crossover**: Frequency ω_c at which the phase angle is -180°.
- **Gain Margin**: Additional gain that just makes the system unstable at the frequency ω_c. i.e. $|G(j\omega_c)H(j\omega_c)| = \frac{1}{a}$, where a is the factor by which the gain must be changed in order to produce instability.
- **Bandwidth (BW)**: Range of frequencies over which the magnitude ratio does not differ by more than -3 db from its magnitude at a specified frequency.
- **Resonance Peak** M_m is a measure of relative stability and is the maximum value of the magnitude of the closed-loop frequency response.
- **Resonance Frequency** ω_m is the frequency at which M_m occurs. $M_m = \max \omega_m \left|\frac{C}{R}(j\omega_m)\right|$

3-2.5 Correlation between Frequency and Time Responses

For a simple 2nd order system

$$M_p = 1 + e^{-\zeta\pi/\sqrt{1-\zeta^2}} \leftrightarrow\leftrightarrow M_m = \frac{1}{2\zeta\sqrt{1-\zeta^2}}$$

$$\omega_d = \omega_n\sqrt{1-\zeta^2} \leftrightarrow\leftrightarrow \omega_m = \omega_m\sqrt{1-2\zeta^2}$$

Hence

1. As ω_m increases we have a faster time of response, T_r and T_s
2. $0.4 < \zeta < 0.707$

   $1 < M_m < 1.4$

   $1.05 < M_p < 1.25$
3. The closer the $\omega_0(j\omega)$ curve to the (-1+j0) point, the larger is the value of M_m.
4. As K_n increases we have a greater accuracy (i.e. e_{ss} decreases).
5. Phase margin γ is directly related to

   for $0.4 < \zeta < 0.707$ and $450 < \gamma < 650$

Comments:

$\zeta < 0.4$ yields excessive overshoot

$\zeta > 0.7$ responds sluggishly

Also

$$e_{ss} = \frac{1}{K}$$

Therefore for large K, e_{ss} is small. Also a large K suppresses undesirable effects caused by dead zone, backlash, coulomb friction, etc.

Note:

The frequency response is given by

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

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

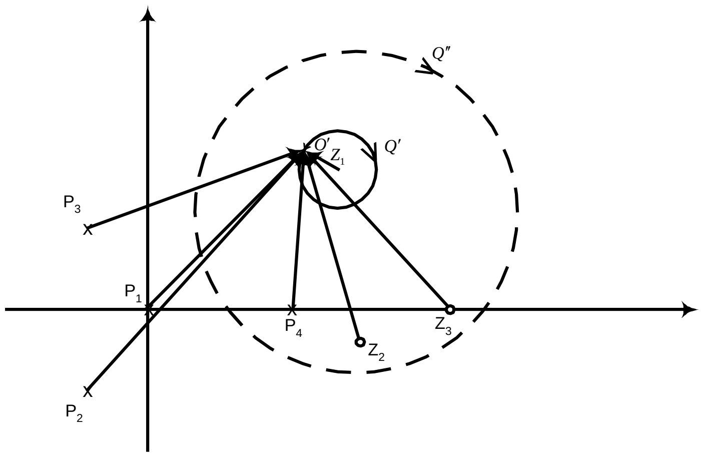
Differentiating the above equation with respect to frequency and setting it to zero

M_m occurs at $\omega_m\sqrt{1-2\zeta^2}$ and $M_m = \dfrac{1}{2\zeta\sqrt{1-\zeta^2}}$

3-2.6 Nyquist Stability Criterion

For a stable system, none of the roots of $Q(s)=1+G(s)H(s)=0$ should lie in the right half of the s-plane or on the $j\omega$ axis.

$$Q(s)=\frac{(s-Z_1)(s-Z_2)\ldots(s-Z_m)}{(s-P_1)(s-P_2)\ldots(s-P_n)}$$



As O' is rotated clockwise once around the closed curve Q', the length $(s-Z_1)$ rotates through a net angle of 360º. All the other directed segments rotate through a net angle of 0º.

For the larger contour Q'' which includes zeros Z_1, Z_2, Z_3 and pole P_4 each of the directed line segments from the enclosed poles and zeros rotates through a net angle of 360º.

Hence the total net angular rotation experienced by

Q(s) = Number of poles enclosed – Number of zeros enclosed
= 1 – 3 = - 2

(- sign indicates clockwise rotation).

If we consider a closed contour such that the whole right half plane is encircled, thus encircling all zeros and poles of *Q(s)* that have positive real parts, then the net rotations N of *Q(s)* about the origin is equal to the total number of poles P_R minus its total number of zeros Z_R in the right half s-plane.

$$N = P_R - Z_R$$

For a stable system *Q(s)* can have no zeros in the right half s plane.

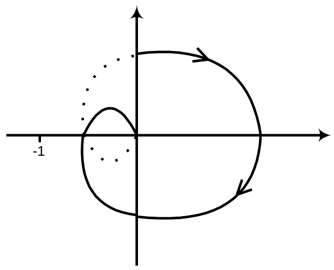
Nyquist stability criterion states that for a stable system, the net number of rotations of *G(s)H(s)* about the (-1 + j0) point must be zero when there are no poles of G(s)H(s) in the right half s-plane.

3-2.6.1 General Stability Criterion

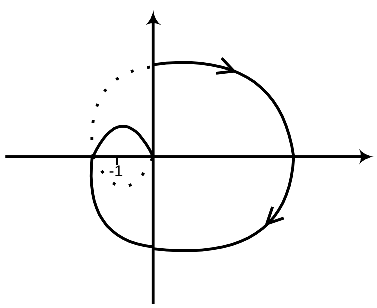
If an open loop transfer function $G_0(s)$ has P_0 poles in the right half of the s-plane and if the Nyquist diagram, $G_0(j\omega)$ encircles the critical point (-1, 0) N times in the clockwise direction, then the closed-loop transfer function has P_c poles in the right half s-plane where

$$P_c = P_0 + N$$

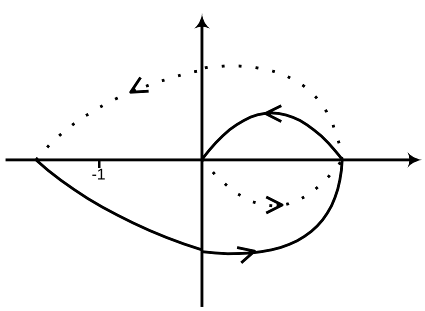
The closed-loop system is stable if and only if $P_c = 0$.



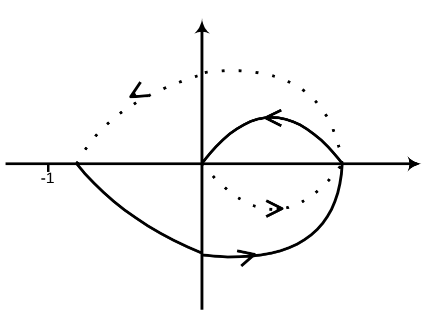
$N = 0$, $P_0 = 0$, $P_c = 0$
Closed-loop system is stable.



$N = 2$, $P_c = 2$
Closed-loop system is unstable

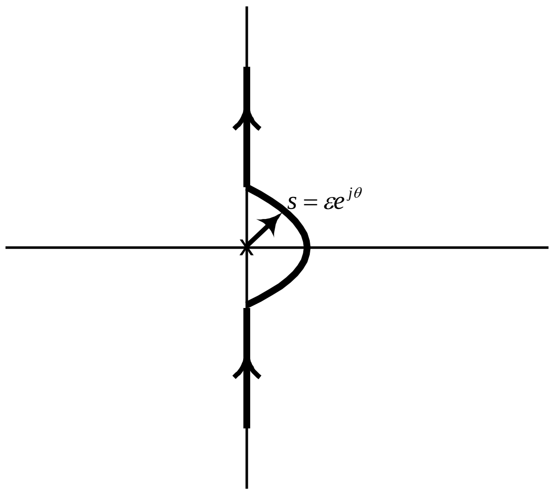


$N = -1$, $P_0 = 1$, $P_c = 0$
Closed-loop system is stable



$N = 0$, $P_0 = 1$, $P_c = 1$
Closed-loop system is unstable

3-2.6.2 Effect of Poles at the Origin



$$GH(s) = \frac{K_1}{s(1+s)}$$

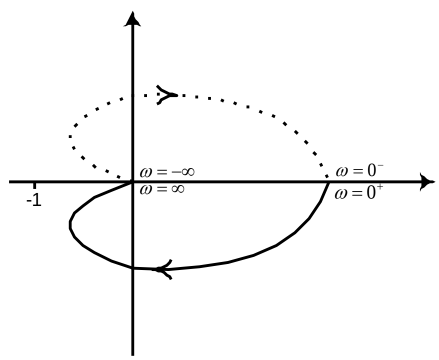
Let $\quad s = \varepsilon e^{j\theta} \qquad\qquad \varepsilon << 1$

$$GH(s) = \frac{K_1}{\varepsilon e^{j\theta}} = \frac{K_1}{\varepsilon} e^{-j\theta}$$

As s goes from $\varepsilon \angle {-\pi/2}$ to $\varepsilon \angle {\pi/2}$

GH(s) goes from $\frac{K_1}{\varepsilon} \angle {\pi/2}$ to $\frac{K_1}{\varepsilon} \angle {-\pi/2}$

Also $\frac{K_n}{s^n} \Rightarrow$ net rotation of $n\pi$

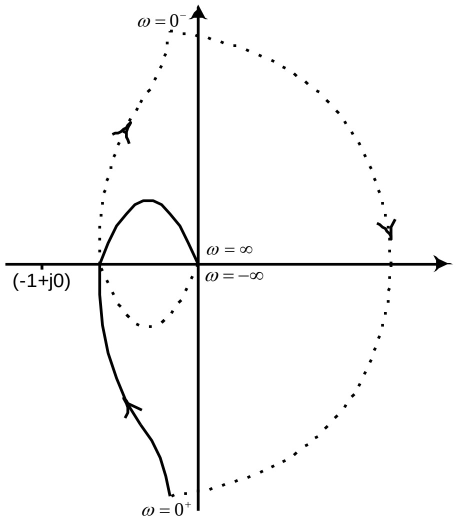


Nyquist Plot
Type 0 system

$$G(s)H(s) = \frac{K_0}{(1+T_1 s)(1+T_2 s)}$$

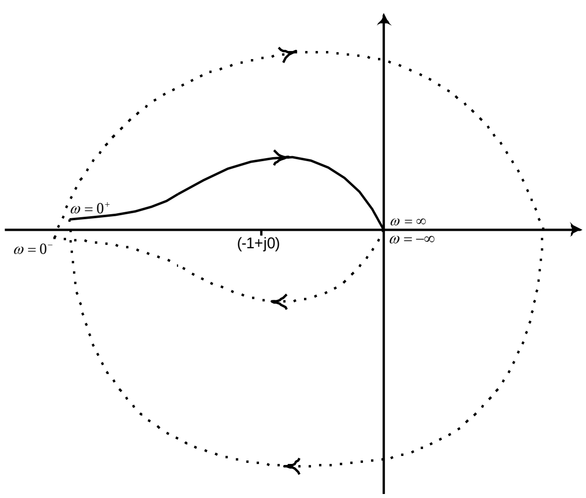
As $s \to 0$
$GH \to K_0$
As $s \to \infty$
$\phi \to -180$
Here $N = 0$, $P_R = 0$, $Z_R = 0$
System is always stable

Type 1 system



$$G(s)H(s) = \frac{K_1}{s(1+T_1 s)(1+T_2 s)}$$

Type 2 System



$$G(s)H(s) = \frac{K_2}{s^2(1+T_1 s)}$$

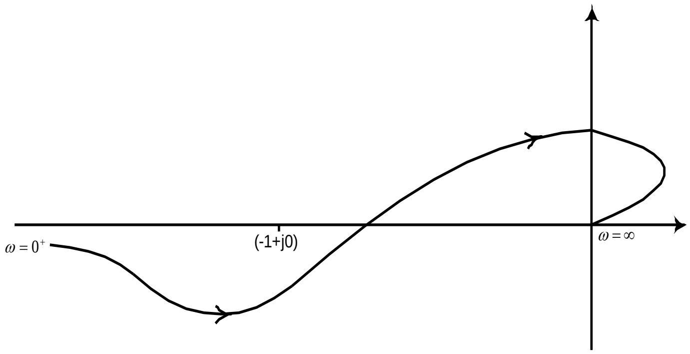
Here N = -2

$\therefore -2 = P_R - Z_R$

$\therefore Z_R = 2$

System is unstable

s^2 in the denominator will result in a net rotation of 360° in the vicinity of $\omega = 0$



$$G(s)H(s) = \frac{K_2{'}(1+sT_1)}{s^2(1+sT_f)(1+sT_m)(1+sT_z)}$$

As $\omega \to 0^+$, the polar plot for a Type 2 system is below the real axis if

$$\sum(T_{num}) - \sum(T_{den}) > 0$$

i.e. $T_1 > (T_f + T_m + T_z)$

3-2.6.3 Effects of Gain

The Nyquist diagram will have the same phase for a given frequency ω no matter what the gain K. Increasing the gain magnifies the magnitude of the Nyquist diagram at each radial value equally as shown below.

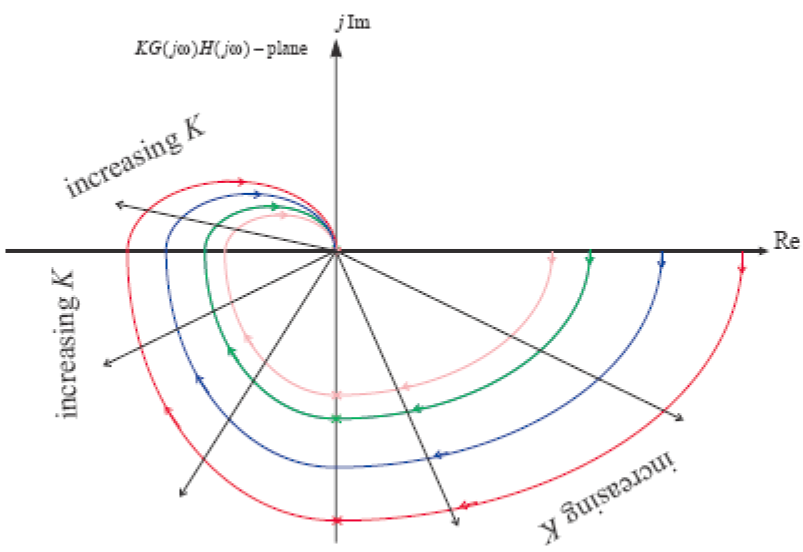


Figure 3.15: Effects of Gain on the Nyquist Diagram

3-2.6.4 Simplified Nyquist Criterion

If a system has no open-loop poles or zeros in the right-half plane, the system is closed-loop stable if and only if the point $-1 + j0$ lies to the left of the open-loop Nyquist diagram relative to an observer traveling along the plot in the direction of increasing frequency.

3-2.6.5 Gain and Phase Margins

In the interests of safety and good design, we would wish to be sure that the system would not be merely stable, but rather has a guaranteed margin of stability. The quantities *gain* and *phase margin* are measures of this margin of stability. They can be used to assess the closeness of the system to instability and also, in design, to ensure a good margin of stability.

3-2.6.6 Gain Margin

If the frequency at which the Nyquist diagram crosses the real axis is called ω_π (the frequency at which the phase is π radians or 180°)[1] then $A = |KG(j\omega_\pi)H(j\omega_\pi)|$ is the magnitude of the frequency response at this frequency. We define a measure of closeness to the critical point called the "gain margin" (g_m) as $g_m = 1/A$. The gain margin must be greater than zero for a stable system and usually we specify the gain margin to be $g_m \geq 6$ dB ($g_m \geq 2$). However, you should not attempt to make the gain margin too large since the corresponding system gain will be small with the result that the system tends to be slow and steady-state errors large. The definition and measurement of the gain margin is illustrated in the figure below.

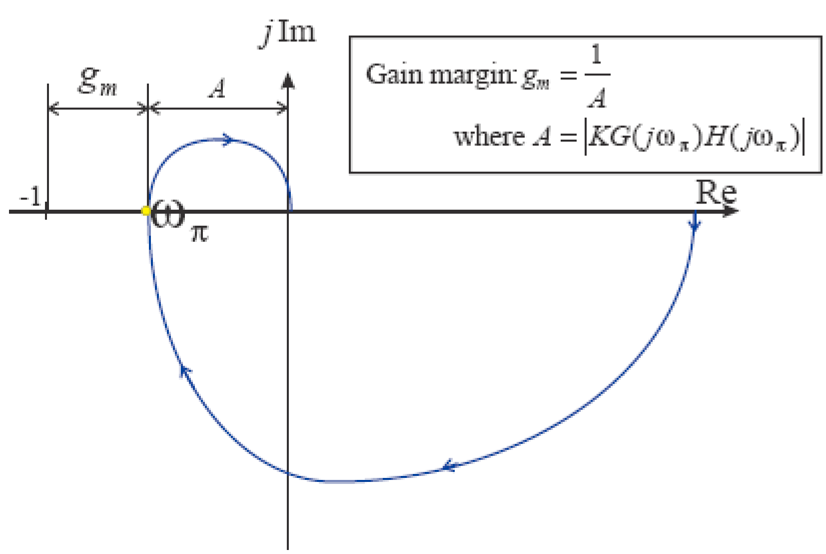


Figure 3.16: Calculation of Gain Margin using the Nyquist Diagram

3-2.6.7 Phase margin

The phase margin is a measure, in terms of phase, of closeness to the critical point. Since the magnitude of the open-loop frequency response at ω_π is unity when the system is marginally stable, if we consider the phase when the magnitude is unity, the system will be guaranteed to be stable if the phase is as illustrated in the figure that follows.

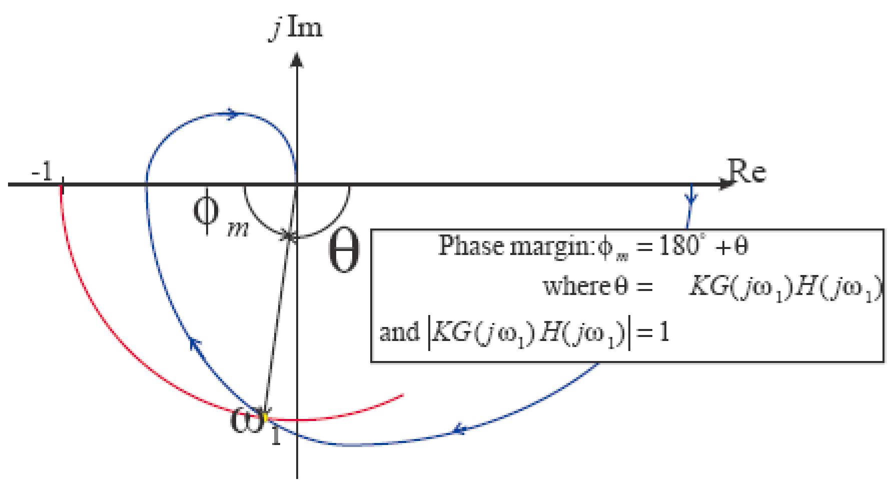


Figure 3.17: Calculation of Phase Margin using the Nyquist Diagram

The frequency at which the magnitude is unity is called the *gain cross-over frequency* ω_1 and is determined by solving the equation

$$|KG(j\omega_1)H(j\omega_1)| = 1$$

for ω_1. The corresponding phase is

$$\theta = \angle KG(j\omega_1)H(j\omega_1)$$

and the phase margin is defined to be

$$\phi_m = 180^\circ + \theta.$$

Since the phase is negative for a proper transfer function, the phase margin will be less than 180° if the system is stable and negative if the system is unstable.

For design purposes, the phase margin is usually kept in the range $30^\circ \leq \phi_m \leq 60^\circ$. Once again, if the phase margin is made too high there will be consequences on the performance of the resulting system.

UNIT THREE ASSIGNMENT

1. Determine whether or not the systems with the transfer functions given below are stable:

i.

$$\frac{10s^2 + 2s + 3}{s^5 + 2s^4 + 4s^2 + 2s + 7}$$

ii.

$$\frac{4s^2 + 6s + 3}{s^4 + 8s^3 + 24s^2 + 32s + 16}$$

2. A feedback control system has the following transfer function:

$$G_o(s) = \frac{K(s+2)}{s(s+1)(s^2+2s+2)}.$$

Determine the range of values for K for which the feedback control system is stable.

3.

Use the root locus technique to determine the range of values of K for which the system with the open-loop transfer function

$$G_o(s) = \frac{K}{(s-1)(s+2)(s+8)}$$

is closed-loop stable.

4.

A feedback control system has open-loop transfer function

$$G_o(s) = \frac{K(s+4)}{s(s+1)^2}$$

Sketch the Nyquist diagram for this system, and determine the gratest value of K for which the closed-loop is stable.

5.

The open-loop transfer function of a unity feedback control system is given by

$$G(s) = \frac{K_1}{s(1+0.1s)(1+s)}$$

(a) Plot the asymptotic bode diagrams for $K_1 = 1$.
(b) Determine the value of K_1 so that the gain margin of the system is 10db.
(c) Determine the value of K_1 so that the phase margin if the system is 60^o.

UNIT 4

COMPENSATION AND CONTROLLERS

INTRODUCTION TO UNIT FOUR

In the design of real life control systems, there is the need to take into consideration, the effects of extraneous inputs which are technically referred to as disturbances.
Such inputs are cannot be eliminated, and their presence affects the overall output of the system, resulting in deviations of the output form the desired values.
To ensure that the system outputs stay within acceptable bounds and hence meet performance specifications, compensation networks are introduced into feedback control systems. This unit deals with the various types of system compensations. Consideration is also given to the various types of compensation systems based on the root locus technique and the frequency response methods. Latter sections of the unit introduce three term PID (proportional, integral and differential) controllers. The unit ends with a section on feedback compensation.

Learning objectives

1. To gain an understanding of the some single-degree-of-freedom (SDF) compensation systems.
2. To gain knowledge of cascade compensation based on the root locus and frequency response techniques.
3. To understand three term PID controllers
4. To gain a fair knowledge of feedback compensation.

SECTION 4-1COMPENSATION

4-1.1 Types of System Compensations

To meet performance specifications for feedback control systems, appropriate compensation networks must usually be introduced into the systems. A system with one compensator is called a *single-degree-of-freedom (SDF)* system. A *two-degree-of freedom system (TDF)* has two compensators.

Methods of SDF system compensation are:

1. Cascade
2. Feedback and
3. Minor Feedback

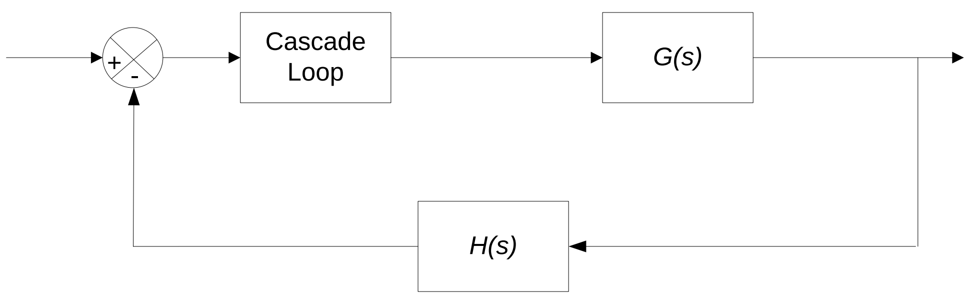


Figure 4.1(a): Cascade Compensation

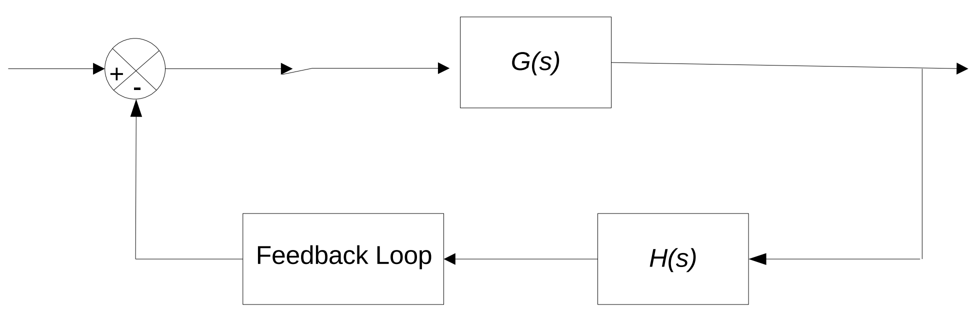


Figure 4.1(b): Feedback Compensation

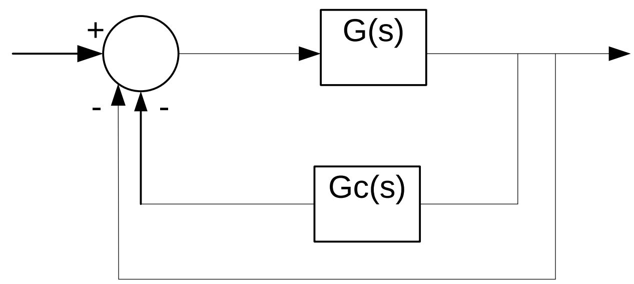


Figure 4.1(c): Minor Feedback Compensation

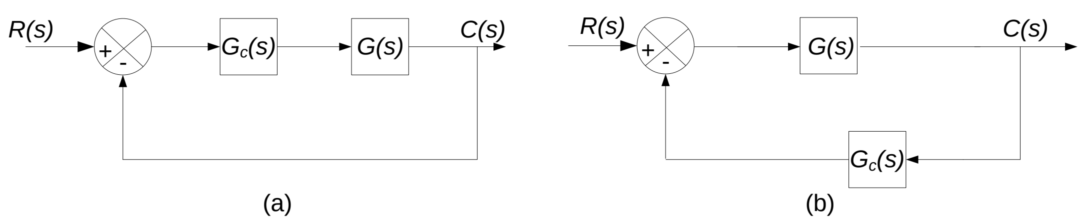


Figure 4.2: Single-degree-of-freedom Systems

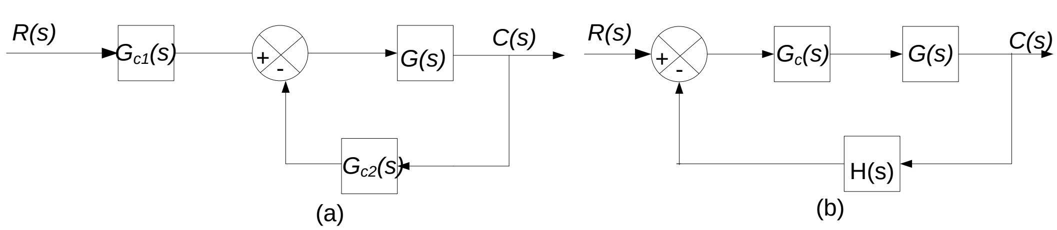


Figure 4.3: Two-degree-of-freedom Systems

4-1.1.1 Compensation is needed for the following cases.

1. Stable system with satisfactory transient response but large steady-state error (e_{ss})
   SOLUTION
   Gain must be increased to reduce e_{ss} without appreciably reducing the system stability. [Increasing K_n increases transient response].
   [Lag compensator is used for this purpose]

2. System is stable but has a poor transient response.
   SOLUTION
   Root locus must be moved to the left of the imaginary axis

3. Stable system but both transient and steady-state response are unsatisfactory.
   SOLUTION
   Locus must be moved to the left and gain increased.

4. System is unstable for all values of gain.
   SOLUTION
   Locus must be reshaped so that part of each branch falls in the left half *s* plane, thereby stabilizing the system

4-1.2 CASCADE COMPENSATION – ROOT LOCUS

4-1.2.1 Lag Compensation

If the steady –state error is too large the error can be reduced without appreciably changing the dominant roots of the characteristic equation by means of integral plus proportional control $\left[1+\frac{a}{s}\right]$.

A circuit which approximates this is given by

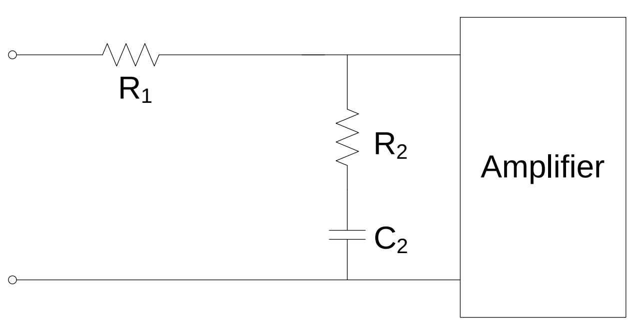


Figure 4.4: Circuit Approximation of a Proportional plus Integral Control

This gives the approximate proportional + integral output.

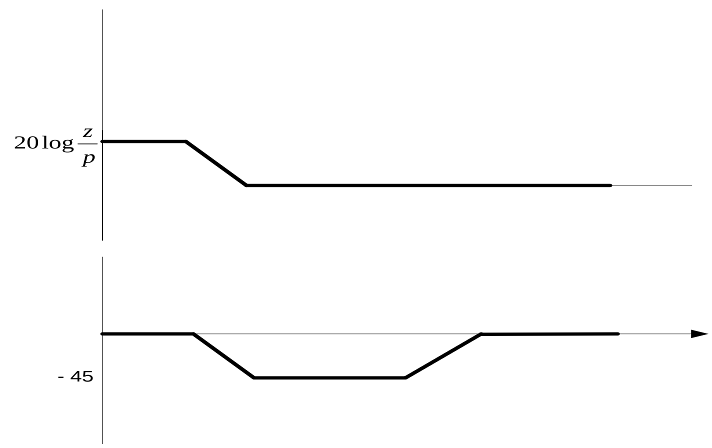
$$G_c = \frac{s+z}{s+p} \qquad |z| > |p|$$

For most of the original locus to remain practically unchanged the pole $s=-p$ and the zero $s=-z$ of the compensator are placed close together and near the origin. The angle contributed by the compensator at the original closed-loop dominant root should be less than about 5 degrees.

The required gain of the compensator is then given approximately by

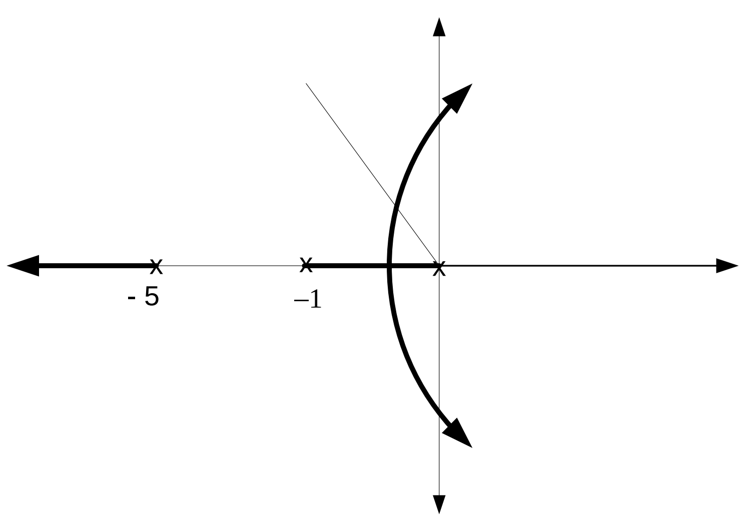
$$\alpha = \left|\frac{z}{p}\right|$$

The size of α limited by the physical parameters required in the network $\alpha = 10$ is often used.



Example 1:

$$G(s) = \frac{K}{s(s+1)(s+5)}$$



For a damping ratio of $\zeta = 0.5$

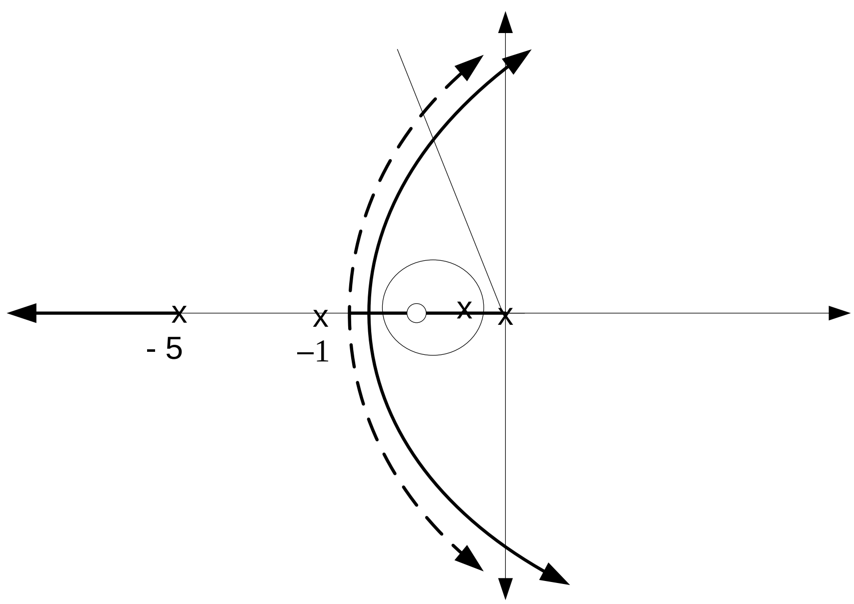
Dominant Roots are $s_{1,2} = -0.4 \pm j0.81$

Static Loop Sensitivity $K = |s||s+1||s+5| = 4.17$

$$K_1 = \lim_{s \to 0} sG(s) = \frac{K}{5} = 0.83 \sec^{-1}$$

$\omega_n = 0.9$ rad/sec

Let $G_c(s) = \dfrac{s+0.1}{s+0.01}$



For $\zeta = 0.45$

Dominant roots are $s_{1,2} = -0.36 \pm j0.72$

$$K' = \frac{|s||s+1||s+5||s+0.1|}{|s+0.01|} = 3.75$$

$$K_1' = \frac{K' \times 0.1}{0.01 \times 5} = K' \times 2 = 3.75 \times 2 = 7.5 \sec^{-1}$$

Increase in Gain $\dfrac{7.5}{0.83} = 9.03$ and $\omega_n = 0.8$ rad/sec

4-1.2.2 Lead Compensation

When the transient response is to be improved it is necessary to reshape the locus so that it is moved further to the left of the imaginary axis. This can be achieved using an ideal derivative compensation.

In practice we have

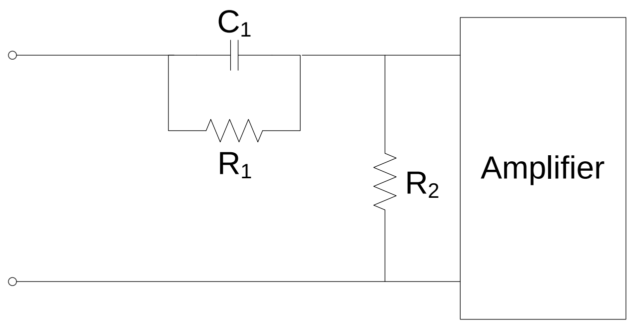


Figure 4.5: Circuit Approximation of a Derivative Controller

$$G_c = \frac{s+z}{s+p} \qquad\qquad |z| < |p|$$

Choose $\alpha = \left|\frac{z}{p}\right|$ small so that the location of the pole is far to the left and has a small effect on the important part of the root locus. In practice $\alpha = 0.1$.

Example 2:

$$G(s) = \frac{K}{s(s+1)(s+5)}$$

With $\alpha = 0.1,\ G_c(s) = \dfrac{s+z}{s+p} = \dfrac{s+1}{s+10}$

Here for a damping ratio of $\zeta = 0.45$

Dominant roots are $s_{1,2} = -1.6 \pm j3.2$

$K_1{}' = 2.96$ Increase in gain $= \dfrac{K_1{}'}{K_1} = \dfrac{2.96}{} = 3.5$

$\omega_n{}' = 0.5 \qquad$ Increase in gain $= \dfrac{\omega_n{}'}{\omega_n} = \dfrac{3.51}{0.9} = 3.9$

4-1.2.3 Lag-Lead Compensation

We now have

$$G_c(s) = \left(\frac{s+0.1}{s+0.01}\right)\left(\frac{s+1}{s+10}\right)$$

Thus $G(s) = G_\alpha(s)G_c(s) = \dfrac{K'(s+0.1)}{s(s+0.01)(s+5)(s+10)}$

For $\alpha = 0.45$

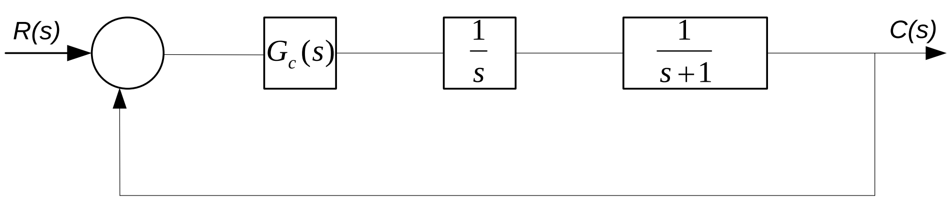
$s_{1,2} = --1.55 \pm j3.08$

$K' = 145$

$K_1{}' = \dfrac{K'(0.1)}{(0.01)(5)(10)} = 29\,\text{sec}^{-1}$

$\omega_n = 3.45$

Example 3:



(a) A block diagram of an altitude rate control system is shown above.

Use the root locus method to design a compensator $G_c = \dfrac{s+z}{s+p}$ so that the dominant complex poles have damping factor $\alpha = 0.5$ and undamped frequency $\omega_n = 4$. The ratio $\alpha = \left|\dfrac{z}{p}\right|$ should be maximum.

(b) Sketch the root locus of the compensated system.

SOLUTION

$$G_\alpha(s) = \frac{1}{s(s+1)}$$

Locate $s_1 = -2 + j2\sqrt{3}$

$$\cos^{-1}\zeta = 60^\circ$$

$$\varphi_1 = 180 - \tan^{-1}(2\sqrt{3}) = 180 - 73.9$$

Contribution of the compensator pole and zero at s_1 is given by

$$(\phi_p - \phi_o) + \varphi_1 + 120 = 180$$

$$\therefore (\phi_p - \phi_o) = 180 - (180 - 73.9) - 120 = -46.1^\circ$$

From graph

$$G_c = \frac{s+2}{s+p} = \frac{s+2.5}{s+6.8}$$

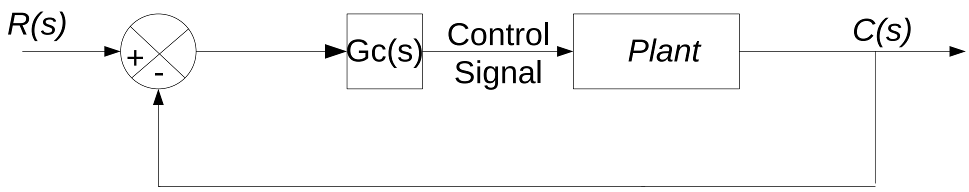
$$\alpha = \frac{2.5}{6.8} = 0.37$$

Root locus of the compensated system is

$$G(s) = G_\alpha(s)G_c(s) = \frac{s+2.5}{s(s+1)(s+6.87)}$$

4-1.3 Cascade Compensation – Frequency Response

Here a network is inserted in cascade with the plant.

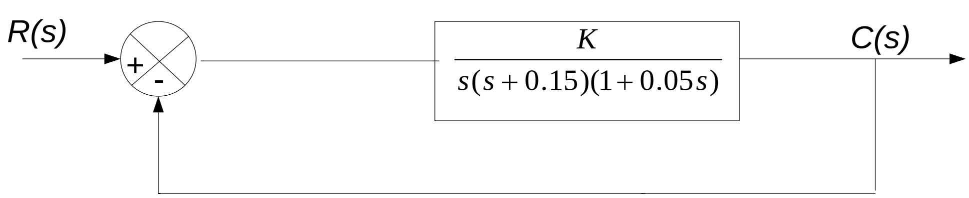


The simplest case of cascade compensation is where $G_C(s) = K$, known as the proportional control.

Gain control

Example

Determine the gain constant *K*, in the open-loop transfer function of the system below to give a closed-loop system having a phase margin of 40 ^{o}. what is the corresponding gain margin?



Solution

Set K = 1 and draw Bode diagrams. A phase margin is obtained for a cross-over frequency of $\omega_1 = 6.3 rad/\sec$. The required Bode gain is 18 dB i.e. K = 7.9. The new gain margin is 12.5 dB.

4-1.3.1Choice of Poles and Zeros

The choice of poles and zeros of a compensator by the root locus method determines the poles of the closed –loop system, which in turn permits evaluation of the closed-loop time response.

Using the polar or log plot, we can determine the new values of M_m, ω_m and K_n by the frequency- response method. However the closed-loop poles are not explicitly determined. Furthermore the correlation between M_m, ω_m and the time response is only qualitative and the presence of real roots near the complex dominant roots alters the correlation more.

The design procedure is usually based on selecting a value for M_m and then finding the corresponding values for ω_m and then finding the gain K_n required. If the desired performance specifications are not met, then compensating devices must be used to reshape the frequency response plot of the basic system.

4-1.3.2 Lag Compensator

When M_m and ω_m are satisfactory but steady state error is too large, it is necessary to increase K_n without reducing ω_m too much.

Now

$$G_c(s) = \frac{s+z}{s+p}$$

$$= \frac{z}{p}\frac{1+\frac{s}{z}}{1+\frac{s}{p}}$$

$$= \alpha\frac{1+Ts}{s+\alpha Ts} \qquad \alpha > 1$$

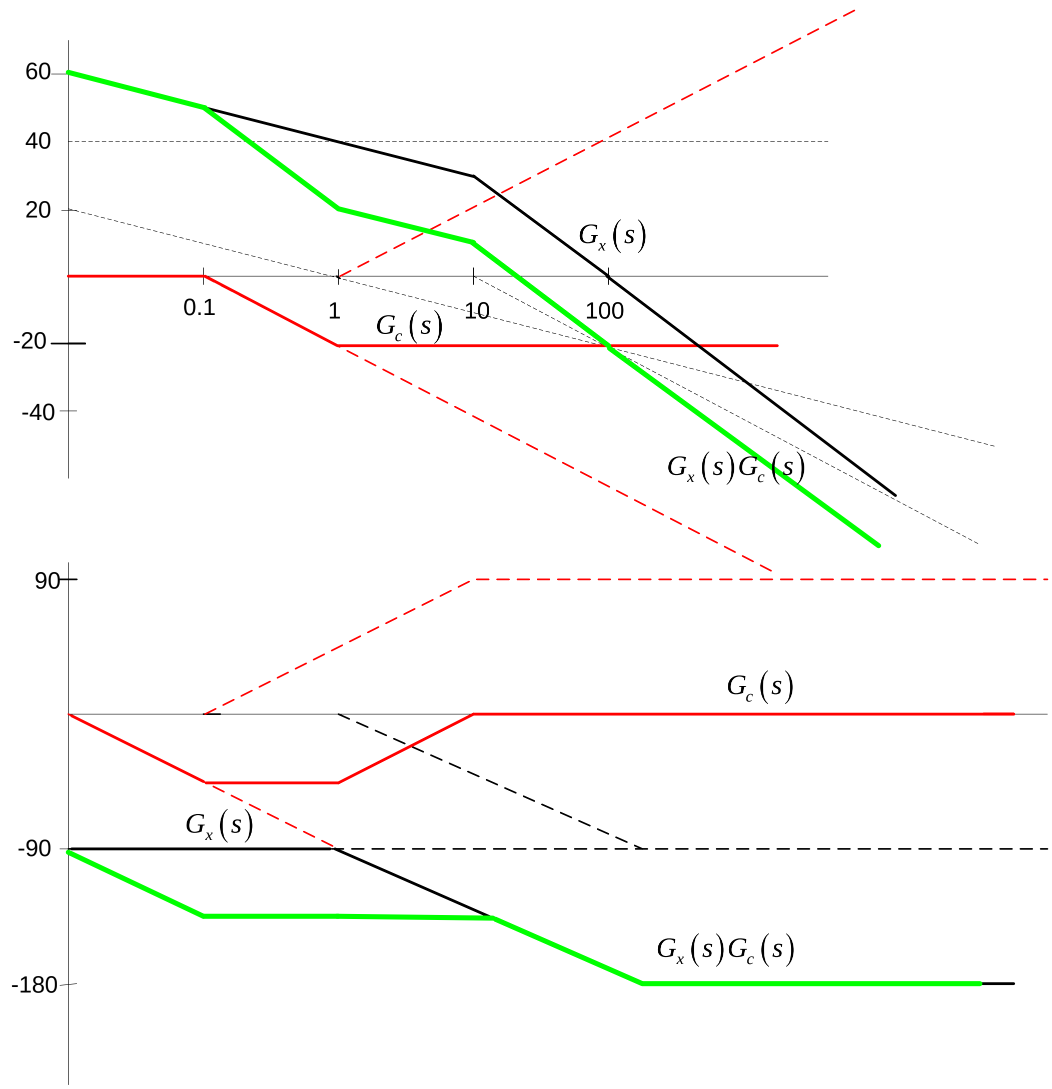
The negative angle of the lag compensator tends to reduce ω_m. To limit this decrease in ω_m, $G_c(s)$ is designed so that it introduces a small angle (-5^o) at the original resonant frequency.

Another method is to select the lag compensator so that the magnitude of its angle is 5^o or less at the original phase margin frequency, ω_ϕ.

The value of T of the compensator is selected so that the attenuation $L_m\alpha$ occurs at the original phase –margin frequency $\omega_{\phi 1}$.

The gain of $G_x G_c(s)'$ is then increased by A so that the L_m curve has a value of 0 dB at the frequency $\omega_{\phi 2}$.

$$G_x(s) = \frac{100}{s\left(1+\dfrac{s}{10}\right)} \qquad G_c(s) = \frac{1+s}{1+\dfrac{s}{0.1}}$$



4-1.3.3 Lead Compensation

$$G_c(s) = \frac{s+z}{s+p}$$

$$= \alpha \frac{1+Ts}{s+\alpha Ts} \qquad\qquad \alpha < 1$$

Application of the lead compensation can be based on adjusting the phase–margin and the phase–margin frequency.

The lead compensator introduces a positive angle over a relatively narrow bandwidth. By properly selecting the value of *T*, the phase-margin frequency can be increased from $\omega_{\phi 1}$ to $\omega_{\phi 2}$. The location of the compensator angle curve must be such as to produce the specified phase-margin at the highest possible frequency.
This location determines the value of the time constant T. The gain of $G_x G_c(s)'$ must be increased for the L_m curve to have the value of 0db at the frequency $\omega_{\phi 2}$.

Analysis shows that the farther to the right the compensator curves are placed (ie the smaller *T* is made), the larger is the new gain of the system.

You may also select the time constant *T* in the numerator of $G_c(s)$ to equal to or slightly less than the largest time constant in the denominator of the original system.

Now

$$G_c(s) = |G_c(j\omega)| \angle \phi_c = \alpha \frac{1 + j\omega T}{1 + j\omega\alpha T}$$

Phase angle $\phi_c = \tan^{-1}\omega T - \tan^{-1}\omega\alpha T$

$$T^2 + \frac{\alpha - 1}{\omega\alpha \tan\phi_c} + \frac{1}{\omega^2\alpha} = 0$$

The frequency $\omega_{\max}$ at which max phase shift occurs for a given *T* and α can be determined by setting the derivative of ϕ_c with respect to frequency ω equal to zero.

$$\omega_{\max} = \frac{1}{T\sqrt{\alpha}}$$

$$\phi_c(\max) = \sin^{-1}\frac{1-\alpha}{1+\alpha}$$

SECTION 4-2: CONTROLLERS

4-2.1 The Three-Term or PID Controller

The controller generates a control signal which is a combination of signals that are proportional to, the integral of, and the derivative of the error signal. Its input is the error signal and its output is the control signal.

$$u(t) = K\left[e + \frac{1}{T_R}\int edt + T_D\frac{de}{dt}\right]$$

where K = Proportional gain

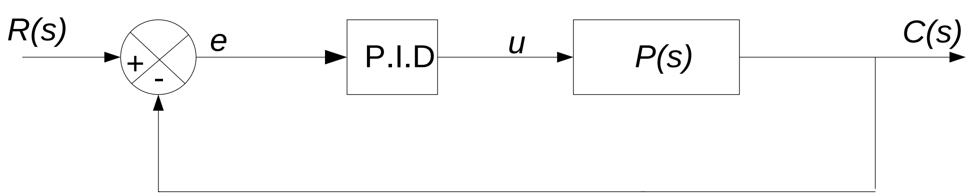
T_R = Reset time

T_D = Rate time (or derivative time)

In industry, the controller may typically be expected to ensure that the overall system :

1. is stable
2. is fast in response to command changes
3. has reasonable damping
4. has zero steady-state error to step change in command and
5. rejects the effects of disturbances.

We now consider individual control actions.
Consider the system



with a second order plant

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

4-2.2 Proportional Action

Here P.I.D $=G_c(s) = K$

$$G_c(s)P(s) = \frac{KG}{s^2 + 2\zeta\omega_n s + \omega_n{}^2}$$

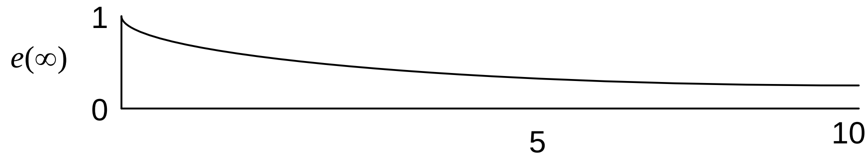
The error transfer function

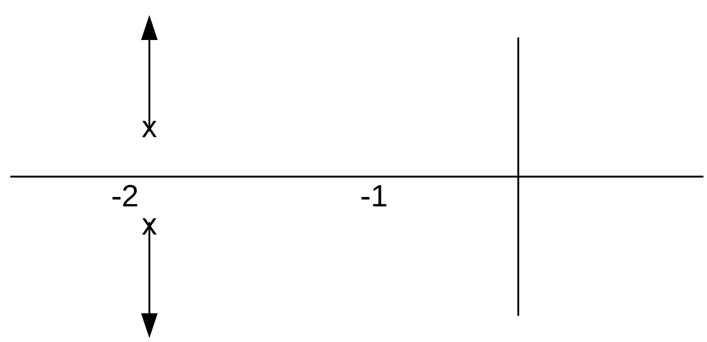
$$\omega_e(s) = \frac{E(s)}{R(s)} = \frac{1}{1+G_c(s)P(s)} = \frac{s^2 + 2\zeta\omega_n s + \omega_n{}^2}{s^2 + 2\zeta\omega_n s + KG + \omega_n{}^2}$$

For a unit-step change in set point, the steady state error is

$$e(\infty) = \lim_{s\to 0} s.\omega_e(s).\frac{1}{s}$$

$$= \frac{\omega_n{}^2}{KG + \omega_n{}^2}$$





As K increases, system gets more and more oscillatory. Thus for proportional control only, an increase in gain reduces steady-state error and increase in gain reduces steady-state error, but makes the system more oscillatory.

4.2.3 Integral Action

Here $G_C(s) = \frac{1}{T_R s}$

The open-loop transfer function is

$$G_C(s)P(s) = \frac{G/T_R}{s(s^2 + 2\zeta\omega_n s + \omega_n{}^2)}$$

For a unit-step change in set point, the steady state error is

$$e(\infty) = l\,\lim_{s\to 0} s.\omega_e(s).\frac{1}{s} = 0$$

Thus the steady state error = 0

Integral control action, by introducing phase lag into the system, has a destabilizing effect on the system. But it has the potential of reducing the steady-state error due to a step input to zero.

4-2.4 Derivative Action

Derivative controller transfer function is $G_C(s) = T_D s$

Open-loop transfer function

$$G_C(s)P(s) = \frac{GT_D s}{s^2 + 2\zeta\omega_n s + \omega_n{}^2}$$

Thus derivative control increases system damping as seen also from the closed-loop transfer function

$$\omega_e(s) = \frac{GT_D s}{s^2 + (2\zeta\omega_n + T_D G)s + \omega_n{}^2}$$

SUMMARY

In practice, integral and derivative control are invariably accompanied by proportional control as proportional-plus-integral (PI) or proportional-plus-derivative (PD) control, or full- blown P.I.D control.

n an industrial application, the problem is to set the values of the controller constant (K, T_R, T_D) so as to give the closed-loop system the desired response characteristics. This is the problem of PID controller tuning which is considered next.

4-2.5 PID Controller Tuning

There is no need for a mathematical model of the plant and so is popular in industries. However an intuitive knowledge of plant behaviour is useful for proper tuning of the controller.

Two examples of tuning PID controllers are due to Ziegler and Nichols.

4-2.5.1 Ziegler and Nichols Continuous Cycling (Ultimate Method.)

1. Turn off the integral and derivative controls of the system ie set $T_R = \infty$ and $T_D = 0$
2. Adjust K until the system just bursts into continuous oscillation. Note the value $K = K_c$ for this condition.
   Record the oscillation and measure the period of oscillation, P_u, known as the ultimate period.
3. The following controller settings are then recommended by Ziegler and Nichols.

Proportional control only

Set $K = 0.5K_c$

PI control

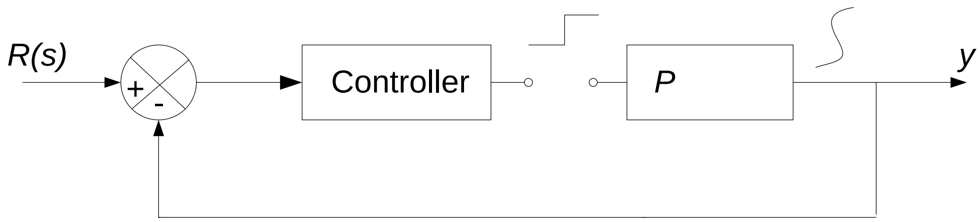
$K = 0.45K_c$ and $T_R = \dfrac{P_u}{1.2}$

PID control

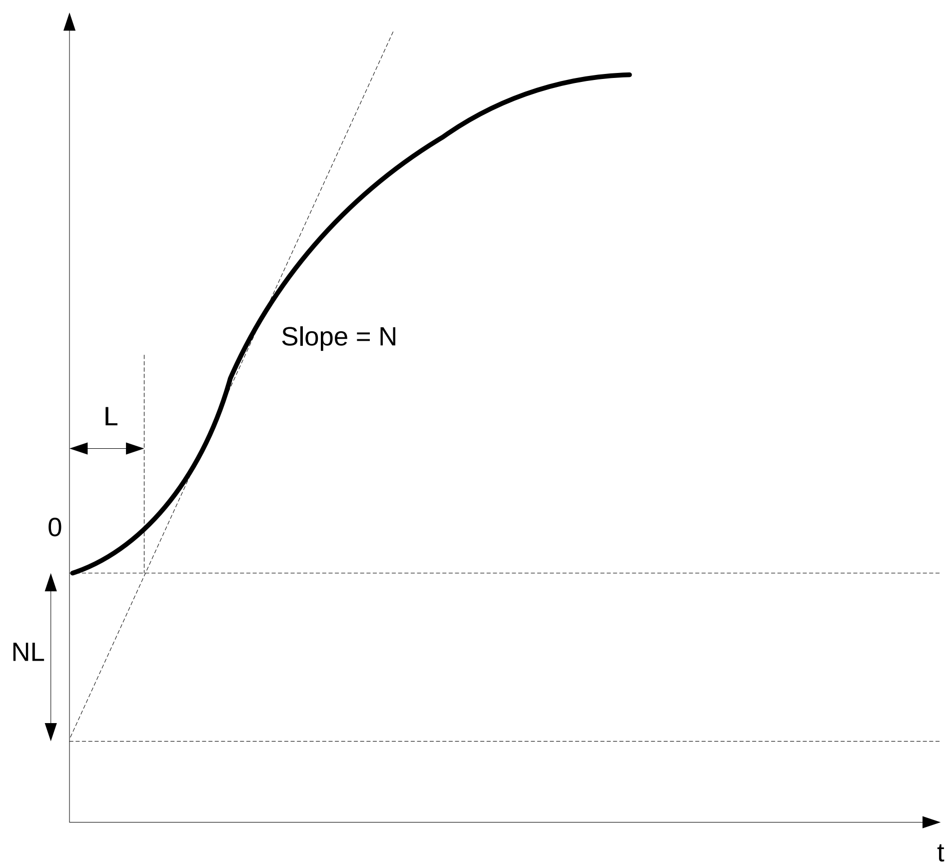
$K = 0.6K_c$, $T_R = \dfrac{P_u}{2}$ and $T_D = \dfrac{P_u}{8}$

4-2.5.2 Ziegler-Nichols reaction Curve Method

1. Break the loop as shown below



2. Apply a step input size ΔU to the plant and record the step response as below; this is the step response of the plant, also known as the reaction curve of the plant.



3. Measure the delay, *L*, the slope, *N*, and the length *NL*.

Ziegler and Nichols recommend the following controller settings.

Proportional control only

Set $K_c = \dfrac{\Delta U}{NL}$

PI control

$K_c = 0.9\dfrac{\Delta U}{NL}$ and $T_R = \dfrac{L}{0.3}$

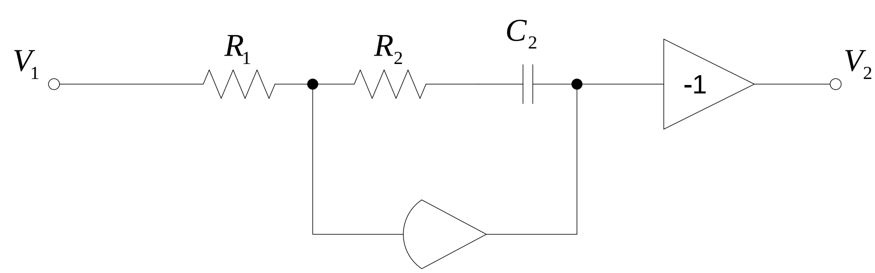
PID control

$K_c = 1.2\dfrac{\Delta U}{NL}$, $T_R = 2L$ and $T_D = 0.52$

4-2.6 ACTIVE PI, PD CONTROLLERS

Some simple active filters often used for realising PI and PD controllers are as follows:

1. **Active PI controller**

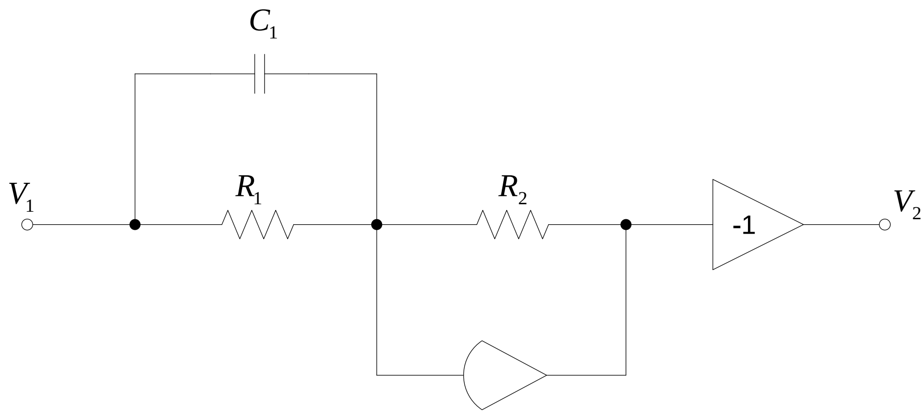


The transfer function is

$$\frac{V_2(s)}{V_1(s)} = \frac{R_2}{R_1}\left(1 + \frac{1}{sC_2R_2}\right)$$

giving the following identifications $K = \dfrac{R_2}{R_1}$ and $T_R = C_2R_1$

4-2.6.2 Active PD controller



$$\frac{V_2(s)}{V_1(s)} = \frac{R_2}{R_1}\left(1 + sC_1R_1\right)$$

so that $K = \dfrac{R_2}{R_1}$ and $T_D = C_1R_1$

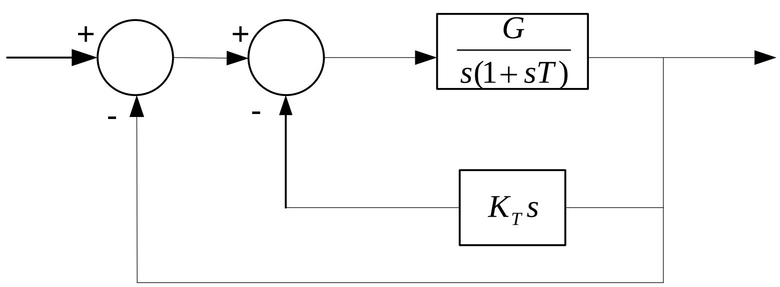
4-2.7 Feedback Compensation

Tacho or velocity feedback

An example is the velocity feedback in a position servo using a tacho-generator.

$G_C(s) = K_T(s)$

Where K_T is the tacho constant



If steady state is too large, the error can be reduced without appreciably changing the dominant roots of the characteristic equation by means of integral + proportional control (lag compensator).

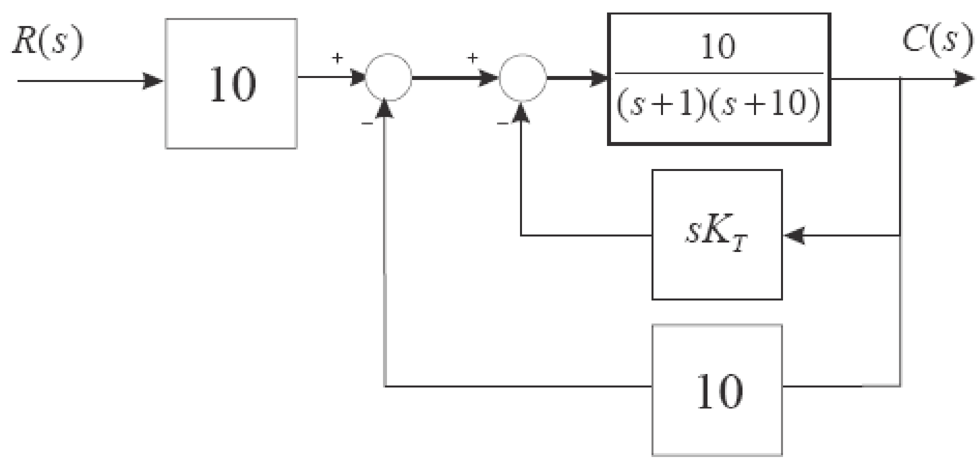
For root locus plot the pole and zero of compensator must be close together and near the origin $\alpha = /z/p/ \cong$ desired increase in gain the size of α is limited by the physical parameters required in the network $\alpha = 10$ is often used.

UNIT FOUR ASSIGNMENT

1. A control system with open-loop transfer function $G_o(s) = K/(s(s+2))$ is to have ideal damping with steady-state velocity error 10%. Show that these criteria cannot simultaneously be achieved with gain compensation alone. If dynamic compensation was used to satisfy the criteria, where would the closed-loop poles need to be located?

2.

For the feedback control system with block diagram shown in the fig. below explain how velocity feedback with coefficient K_T is used to improve the transient performance. Determine the value for K_T for which the transient response is ideal, and the steady-state error in this case.



3.

A feedback control system with unity feedback has a transfer function:

$$G(s) = \frac{K_1}{s(1+0.02s)(1+0.01s)}$$

It is desired that the closed-loop roots are located at s= -10+j35

(a) Sketch the root locus of the original system.
(b) Design a compensator which will achieve the above.
(c) How has the compensated system improved the original system?