Department of Electrical Engineering
(ବୈଦ୍ୟୁତିକ ଯାନ୍ତ୍ରିକ ବିଭାଗ)
ଓଡ଼ିଶା ବୈଷୟିକ ଓ ଗବେଷଣା ବିଶ୍ୱବିଦ୍ୟାଳୟ
ODISHA UNIVERSITY OF TECHNOLOGY AND RESEARCH
(Formerly College of Engineering & Technology)
Ghatikia, Bhubaneswar-751029

Notes

Control System-I

Module-I

Industrial Control Examples

1.1 Basic definitions

Systems: A system is a combination of components that act together and perform a certain objective. A system need not be physical. The concept of the system can be applied to abstract, dynamic phenomena such as those encountered in economics. The word system should, therefore, be interpreted to imply physical, biological, economic, and the like, systems.

Plants. A plant may be a piece of equipment, perhaps just a set of machine parts functioning together, the purpose of which is to perform a particular operation.
such as a mechanical device, a heating furnace, a chemical reactor, or a spacecraft.

Controlled Variable: The *controlled* variable is the quantity or condition that is measured and controlled.
Control Signal or Manipulated Variable: The *control signal* or *manipulated* variable is the quantity or condition that is varied by the controller so as to affect the value of the controlled variable. Normally, the controlled variable is the output of the system.
Control means measuring the value of the controlled variable of the system and applying the control signal to the system to correct or limit deviation of the measured value from a desired value.

Disturbances. A disturbance is a signal that tends to adversely affect the value of the output of a system. If a disturbance is generated within the system, it is called *internal*, while an *external* disturbance is generated outside the system and is an input.

1.2 Mathematical models of physical systems

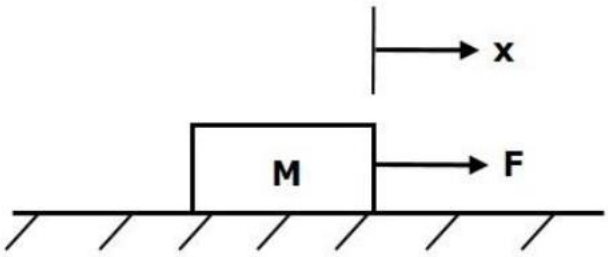
- **Differential equation modelling** of mechanical systems.
- There are two types of mechanical systems based on the type of motion.
    1. Translational mechanical systems
    2. Rotational mechanical system
- Translational mechanical systems move along a **straight line**.
- These systems mainly consist of three basic elements. Those are mass, spring and dashpot or damper.

If a force is applied to a translational mechanical system, then it is opposed by opposing forces due to mass, elasticity and friction of the system.

Since the applied force and the opposing forces are in opposite directions, the algebraic sum of the forces acting on the system is zero.

1.2.1 Mass

- Mass is the property of a body, which stores **kinetic energy**.
- If a force is applied on a body having mass **M**, then it is opposed by an opposing force due to mass.
- This opposing force is proportional to the acceleration of the body. Assume elasticity and friction are negligible.



$$F_m \propto a$$

$$=> F_m = Ma = M\frac{d^2x}{dt^2}$$

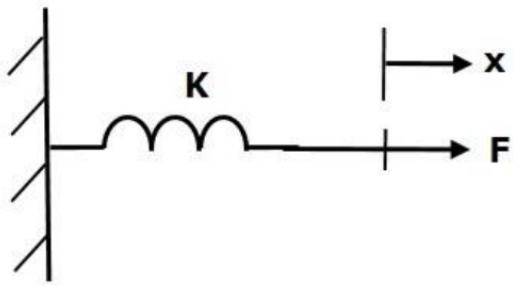
$$F = F_m = M\frac{d^2x}{dt^2}$$

Where,

- $\boldsymbol{F}$ is the applied force
- $\boldsymbol{F_m}$ is the opposing force due to mass
- M is the mass
- a is acceleration
- x is displacement

Spring

- Spring is an element, which stores **potential energy**.
- If a force is applied on spring **K**, then it is opposed by an opposing force due to elasticity of spring.
- This opposing force is proportional to the displacement of the spring. Assume mass and friction are negligible.



$$F_k \propto x$$

$$=> F_k = Kx$$

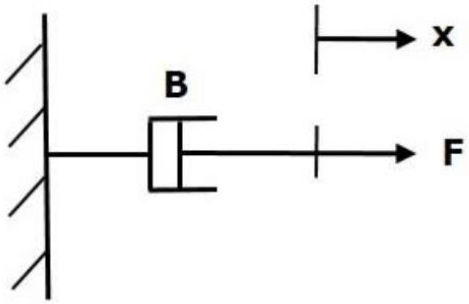
$$F = F_k = Kx$$

Where,

- $\boldsymbol{F}$ is the applied force
- $\boldsymbol{F_k}$ is the opposing force due to elasticity of spring
- $\boldsymbol{K}$ is spring constant
- $\boldsymbol{x}$ is displacement

Dashpot

- If a force is applied on dashpot **B**, then it is opposed by an opposing force due to **friction** of the dashpot.
- This opposing force is proportional to the velocity of the body. Assume mass and elasticity are negligible.



$$F_b \propto v$$

$$=> F_b = Bv = B\frac{dx}{dt}$$

$$F = F_b = B\frac{dx}{dt}$$

Where,

- $\boldsymbol{F_b}$ is the opposing force due to friction of dashpot
- $\boldsymbol{B}$ is the frictional coefficient
- $\boldsymbol{v}$ is velocity
- $\boldsymbol{x}$ is displacement

1.3 Modelling of Rotational Mechanical Systems

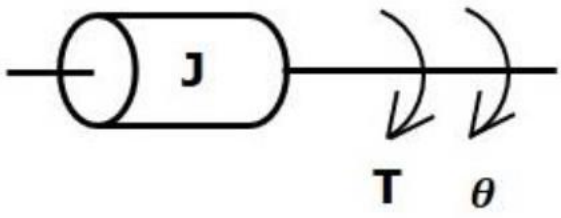
- Rotational mechanical systems move about a **fixed axis**.
- These systems mainly consist of three basic elements. Those are **moment of inertia**, **torsional spring** and **dashpot**.

If a torque is applied to a rotational mechanical system, then it is opposed by opposing torques due to moment of inertia, elasticity and friction of the system. Since the applied torque and the opposing torques are in opposite directions, the algebraic sum of torques acting on the system is zero.

Moment of Inertia

- In translational mechanical system, mass stores kinetic energy. Similarly, in rotational mechanical system, moment of inertia stores **kinetic energy**.

If a torque is applied on a body having moment of inertia **J**, then it is opposed by an opposing torque due to the moment of inertia. This opposing torque is proportional to angular acceleration of the body. Assume elasticity and friction are negligible.



$$T_j \propto \alpha$$

$$=> T_j = J\alpha = J\frac{d^2\theta}{dt^2}$$

$$T = T_j = J\frac{d^2\theta}{dt^2}$$

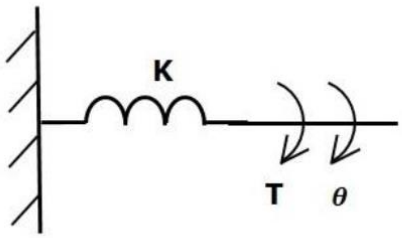
Where,

- ***T*** is the applied torque
- ***T_j*** is the opposing torque due to moment of inertia
- ***J*** is moment of inertia
- **α** is angular acceleration
- **θ** is angular displacement.

Torsional Spring

- In translational mechanical system, spring stores potential energy. Similarly, in rotational mechanical system, torsional spring stores **potential energy**.

If a torque is applied on torsional spring **K**, then it is opposed by an opposing torque due to the elasticity of torsional spring. This opposing torque is proportional to the angular displacement of the torsional spring. Assume that the moment of inertia and friction are negligible.



$$T_k \propto \theta$$

$$=> T_k = K\theta$$

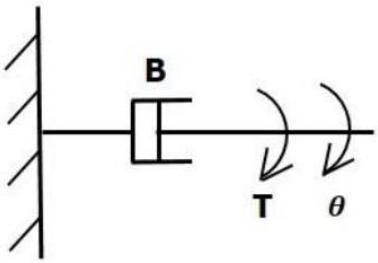
$$T = T_k = K\theta$$

Where,

- ***T*** is the applied torque
- T_k is the opposing torque due to elasticity of torsional spring
- ***K*** is the torsional spring constant
- **θ** is angular displacement

Dashpot

- If a torque is applied on dashpot **B**, then it is opposed by an opposing torque due to the **rotational friction** of the dashpot.
- This opposing torque is proportional to the angular velocity of the body. Assume the moment of inertia and elasticity are negligible



$$T_b \propto \omega$$

$$=> T_b = B\omega = B\frac{d\theta}{dt}$$

$$T = T_b = B\frac{d\theta}{dt}$$

Where,

- T_b is the opposing torque due to the rotational friction of the dashpot
- B is the rotational friction coefficient
- ω is the angular velocity
- θ is the angular displacement

1.4 Open Loop and Closed Loop Control Systems

- Control Systems can be classified as open loop control systems and closed loop control systems based on the **feedback path**.

Open loop control system

- Output is not fed-back to the input. So, the control action is independent of the desired output.

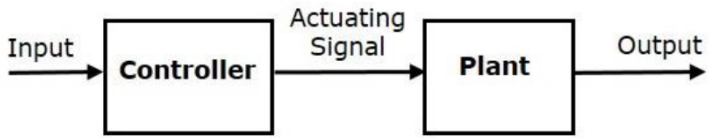


FIG: block diagram of the open loop control system

- An input is applied to a controller and it produces an actuating signal or controlling signal. This signal is given as an input to a plant or process which is to be controlled. So, the plant produces an output, which is controlled. Example: The traffic lights control system

Closed loop control systems

- Output is fed back to the input. So, the control action is dependent on the desired output.

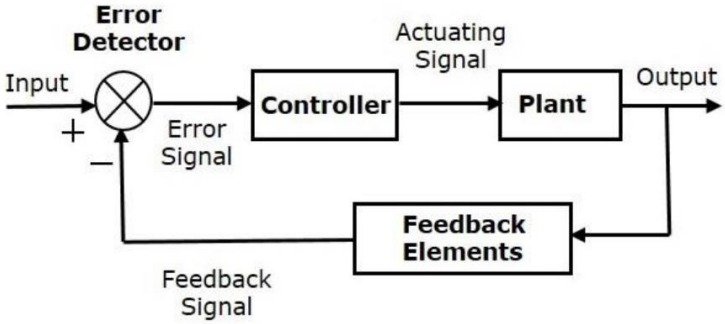


FIG: block diagram of negative feedback closed loop control system

- The error detector produces an error signal, which is the difference between the input and the feedback signal.
- Feedback signal is obtained from the block (feedback elements) by considering the output of the overall system as an input to this block. Instead of the direct input, the error signal is applied as an input to a controller.
- The controller produces an actuating signal which controls the plant. In this combination, the output of the control system is adjusted automatically till we get the desired response.
- Hence, the closed loop control systems are also called the automatic control systems. Example: Traffic lights control system having sensor at the input.

Difference between Open Loop and Closed Loop System

Open Loop Control Systems	Closed Loop Control Systems
Control action is independent of the desired output.	Control action is dependent on the desired output.
Feedback path is not present.	Feedback path is present.
These are also called as **non-feedback control systems**.	These are also called as **feedback control systems**.
Easy to design.	Difficult to design.
These are economical.	These are costlier.
Inaccurate.	Accurate.

Types of Feedback

There are two types of feedback:

- Positive feedback
- Negative feedback

Positive Feedback

- The positive feedback adds the reference input, $R(s)$ and feedback output.

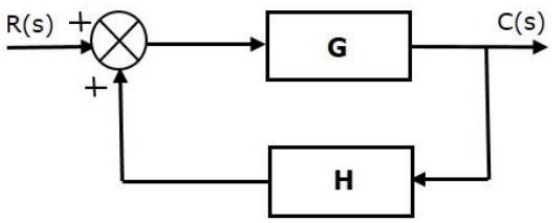


FIG: block diagram of positive feedback control system

The transfer function of positive feedback control system is,

$$T = \frac{G}{1-GH} \qquad \text{Eq. 1}$$

Where,

- T is the transfer function or overall gain of positive feedback control system.
- **G** is the open loop gain, which is function of frequency.
- **H** is the gain of feedback path, which is function of frequency.

Negative Feedback

- Negative feedback reduces the error between the reference input, $R(s)$ and system output.

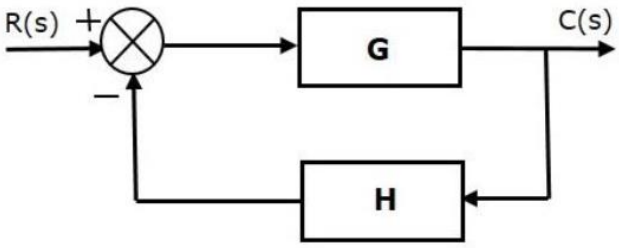


FIG: Block diagram of the negative feedback control system

Transfer function of negative feedback control system is,

$$T = \frac{G}{1+GH} \qquad \text{Eq. 2}$$

Where,

- T is the transfer function or overall gain of negative feedback control system
- **G** is the open loop gain, which is function of frequency
- **H** is the gain of feedback path, which is function of frequency

1.5 Benefits of Feedback

1.5.1 Effect of Feedback on Overall Gain

- From Eq. 2, the overall gain of negative feedback closed loop control system is the ratio of 'G' and (1+GH). So, the overall gain may increase or decrease depending on the value of (1+GH).
- If the value of (1+GH) is less than 1, then the overall gain increases. In this case, 'GH' value is negative because the gain of the feedback path is negative.
- If the value of (1+GH) is greater than 1, then the overall gain decreases. In this case, 'GH' value is positive because the gain of the feedback path is positive.

In general, 'G' and 'H' are functions of frequency. So, the feedback will increase the overall gain of the system in one frequency range and decrease in the other frequency range.

Effect of Feedback on Sensitivity

Sensitivity of the overall gain of negative feedback closed loop control system (**T**) to the variation in open loop gain (**G**) is defined as

$$\mathbf{S_G^T} = \frac{\partial T/T}{\partial G/G} = \frac{\textbf{Percentage change in T}}{\textbf{Percentage change in G}} \qquad \text{Eq. 3}$$

Where, **∂**T is the incremental change in T due to incremental change in G.

Eq. 3 can be written as

$$S_G^T = \frac{\partial T}{\partial G}\frac{G}{T} \qquad \text{Eq. 4}$$

Differentiating Eq. 2 with respect to G on both sides

$$\frac{\partial T}{\partial G} = \frac{\partial}{\partial G}\left(\frac{G}{1+GH}\right) = \frac{(1+GH).1-G(H)}{(1+GH)^2} = \frac{1}{(1+GH)^2} \qquad \text{Eq. 5}$$

From Eq. 2,

$$\frac{G}{T} = 1 + GH \qquad \text{Eq. 6}$$

Substitute Eq. 5 and Eq. 6 in Eq. 4

$$S_G^T = \frac{1}{(1+GH)^2}(1+GH) = \frac{1}{1+GH} \qquad \text{Eq.7}$$

So, we got the **sensitivity** of the overall gain of closed loop control system as the reciprocal of (1+GH). So, Sensitivity may increase or decrease depending on the value of (1+GH).

- If the value of (1+GH) is less than 1, then sensitivity increases. In this case, 'GH' value is negative because the gain of feedback path is negative.
- If the value of (1+GH) is greater than 1, then sensitivity decreases. In this case, 'GH' value is positive because the gain of feedback path is positive.

In general, 'G' and 'H' are functions of frequency. So, feedback will increase the sensitivity of the system gain in one frequency range and decrease in the other frequency range.

Effect of Feedback on Stability

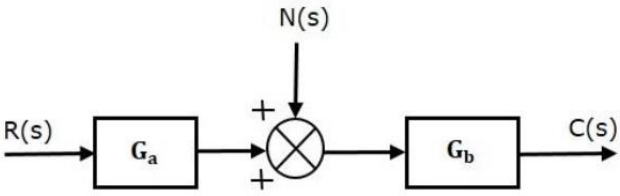
- A system is said to be stable, if its output is under control. Otherwise, it is said to be unstable.
- In Equation 2, if the denominator value is zero (i.e., GH = -1), then the output of the control system will be infinite. So, the control system becomes unstable.

Therefore, we have to properly choose the feedback in order to make the control system stable.

Effect of Feedback on Noise

To know the effect of feedback on noise, let us compare the transfer function relations with and without feedback due to noise signal alone.

Consider an **open loop control system** with noise signal as shown below.

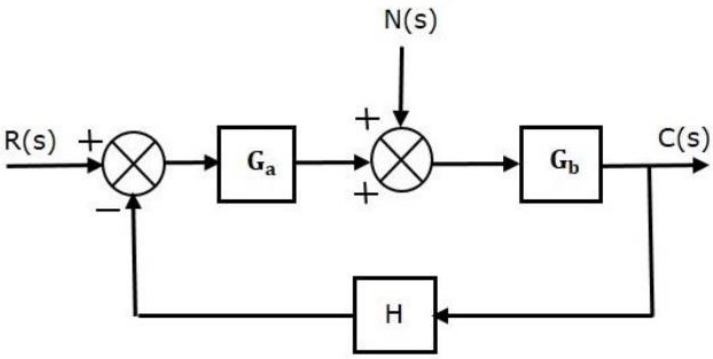


The **open loop transfer function** due to noise signal alone is

$$\frac{C(s)}{N(s)} = G_b \qquad \text{Eq. 8}$$

It is obtained by making the other input R(s) equal to zero.

Consider **a closed loop control system** with noise signal as shown below.



The **closed loop transfer function** due to noise signal alone is

$$\frac{C(s)}{N(s)} = \frac{G_b}{1+G_aG_bH} \qquad \text{Eq. 9}$$

It is obtained by making the other input R(s) equal to zero.

Comparing Eq. 8 and Eq. 9

- In the closed loop control system, the gain due to noise signal is decreased by a factor of $(1+G_aG_bH)$ provided that the term $(1+G_aG_bH)$ is greater than one.

Linear, Time-Invariant (LTI) Systems: Systems that satisfy both linear and time invariant criterion are considered LTI systems. The property of superposition makes LTI systems easier to analyze. By representing complex inputs as a superposition of basic signals, such as the impulse, we can then use superposition to determine the system output.

Transfer Function

- A simpler system or element maybe governed by first order or second order differential equation.
- When several elements are connected in sequence, say ―n elements, each one with first order, the total order of the system will be nth order.
- In general, a collection of components or system shall be represented by nth order differential equation
- In control systems, transfer function characterizes the input output relationship of components or systems that can be described by Liner Time Invariant Differential Equation

In a system having two or more components in sequence, it is very difficult to find graphical relation between the input of the first element and the output of the last element. This problem is solved by transfer function

Definition of Transfer Function

- Transfer function of a LTIV system is defined as the ratio of the Laplace Transform of the output variable to the Laplace Transform of the input variable assuming all the initial condition as zero.

Properties of Transfer Function

- The transfer function of a system is the mathematical model expressing the differential equation that relates the output to input of the system.
- The transfer function is the property of a system independent of magnitude and the nature of the input
- The transfer function includes the transfer functions of the individual elements. But at the same time, it does not provide any information regarding physical structure of the system.
- The transfer functions of many physically different systems shall be identical.
- If the transfer function of the system is known, the output response can be studied for various types of inputs to understand the nature of the system.

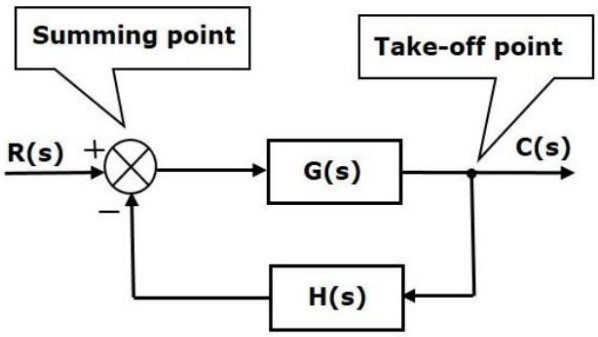
- If the transfer function is unknown, it may be found out experimentally by applying known inputs to the device and studying the output of the system

How you can obtain the transfer function (T. F.)

- Write the differential equation of the system
- Take the L. T. of the differential equation, assuming all initial condition to be zero.
- Take the ratio of the output to the input. This ratio is the T. F.

Block diagram algebra

Block diagrams consist of a single block or a combination of blocks. These are used to represent the control systems in pictorial form.



- Block diagram consists of two blocks having transfer functions G(s) and H(s).
- It is also having one summing point and one take-off point.
- Arrows indicate the direction of the flow of signals.

Block

- The transfer function of a component is represented by a block. Block has single input and single output.

The following figure shows a block having input X(s), output Y(s) and the transfer function G(s).

Transfer function,

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

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

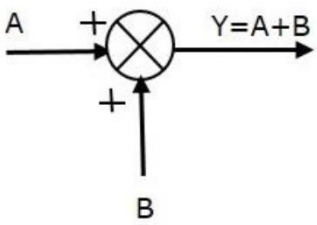
Output of the block is obtained by multiplying transfer function of the block with input.

Summing Point

- The summing point is represented with a circle having cross (X) inside it.
- It has two or more inputs and single output.
- It produces the algebraic sum of the inputs.
- It also performs the summation or subtraction or combination of summation and subtraction of the inputs based on the polarity of the inputs.

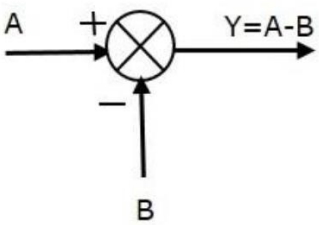
The following figure shows the summing point with two inputs (A, B) and one output (Y).

$Y = A + B$.



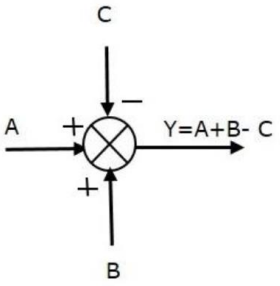
The inputs A and B are having opposite signs, i.e., A is having positive sign and B is having negative sign.

$Y = A + (-B) = A - B$



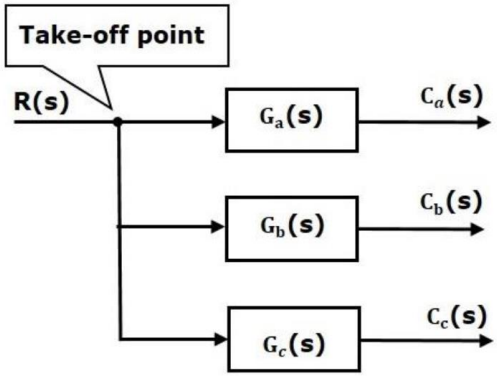
The inputs A and B are having positive signs and C is having a negative sign. So, the summing point produces the output **Y** as

$Y=A+B+(-C)=A+B-C$.

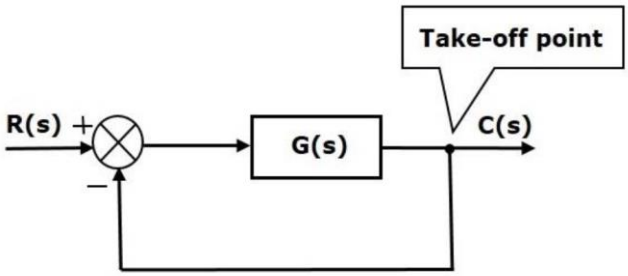


Take-off Point

- The take-off point is a point from which the same input signal can be passed through more than one branch.
- That means with the help of take-off point, we can apply the same input to one or more blocks, summing points.

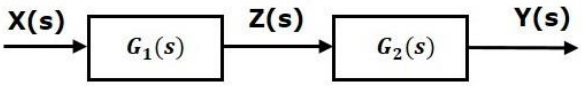


The take-off point is used to connect the output $C(s)$, as one of the inputs to the summing point.



Block Diagram Algebra

1. **Series Connection**



For this combination, we will get the output Y(s) as

$$Y(s)=G_2(s)Z(s)$$

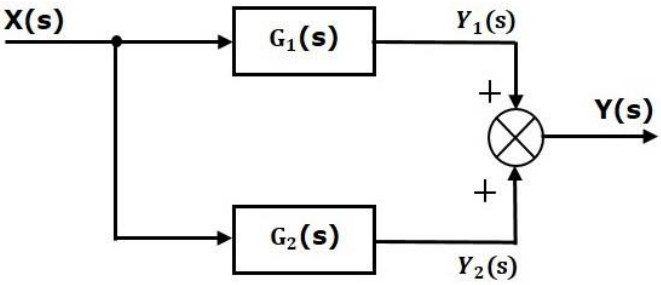
Where, $Z(s)=G_1(s)X(s)$

$$Y(s)=G_2(s)[G_1(s)X(s)]=G_1(s)G_2(s)X(s)$$

$$Y(s)=\{G_1(s)G_2(s)\}X(s)$$

Compare this equation with the standard form of the output equation, $Y(s)=G(s)X(s)$. Where, $G(s)=G_1(s)G_2(s)$

2. **Parallel connection**



For this combination, we will get the output Y(s) as

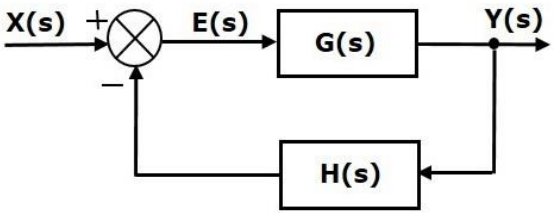
$$Y(s)=Y_1(s)+Y_2(s)$$

Where, $Y_1(s)=G_1(s)X(s)$ and $Y_2(s)=G_2(s)X(s)$

$$\Rightarrow Y(s)=G_1(s)X(s)+G_2(s)X(s)=\{G_1(s)+G_2(s)\}X(s)$$

Compare this equation with the standard form of the output equation, $Y(s)=G(s)X(s)$.
Where, $G(s)=G_1(s)+G_2(s)$

3. **Feedback Connection**



The output of the summing point is -

$$E(s)=X(s)-H(s)Y(s)$$

The output Y(s) is

$$Y(s)=E(s)G(s)$$

Substitute E(s) value in the above equation.

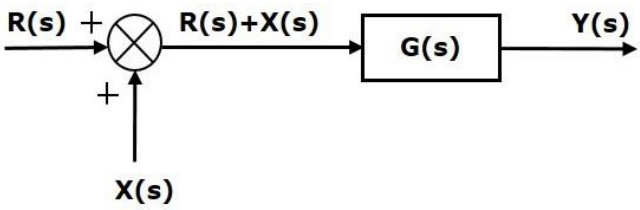
$$Y(s)=\{X(s)-H(s)Y(s)G(s)\}$$

$$Y(s)\{1+G(s)H(s)\}=X(s)G(s)\}$$

$$\Rightarrow \frac{Y(s)}{X(s)} = \frac{G(s)}{1+G(s)H(s)}$$

Therefore, the negative feedback closed loop transfer function is $\frac{G(s)}{1+G(s)H(s)}$

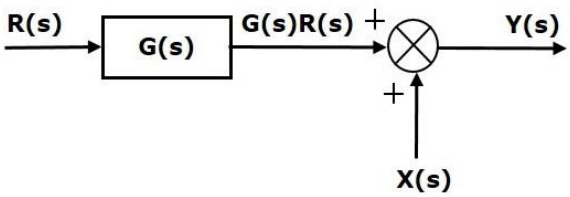
4. **Shifting summing point after the block**



Summing point has two inputs R(s) and X(s). The output of it is {R(s)+X(s)}. So, the input to the block G(s) is {R(s)+X(s)} and the output of it is

$$Y(s)=G(s)\{R(s)+X(s)\}$$

$$\Rightarrow Y(s)=G(s)R(s)+G(s)X(s)$$

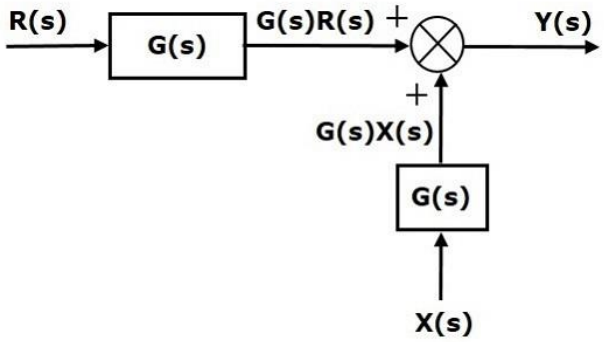


Output of the block G(s) is G(s)R(s).

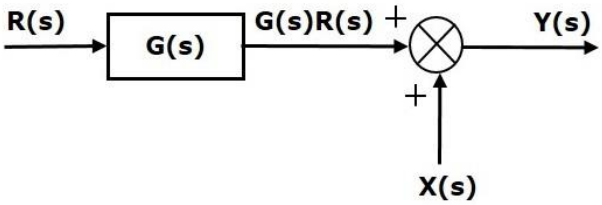
The output of the summing point is

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

The first term 'G(s)R(s)′ is same in both the equations. But there is difference in the second term. In order to get the second term also same, we require one more block G(s).



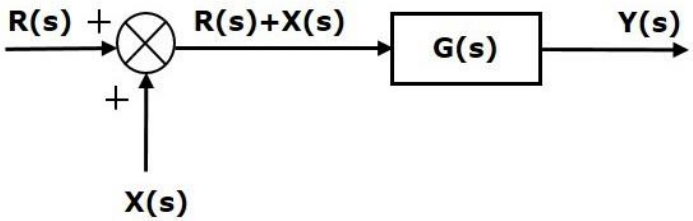
5. **Shifting summing point before the block**



Output of this block diagram is -

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

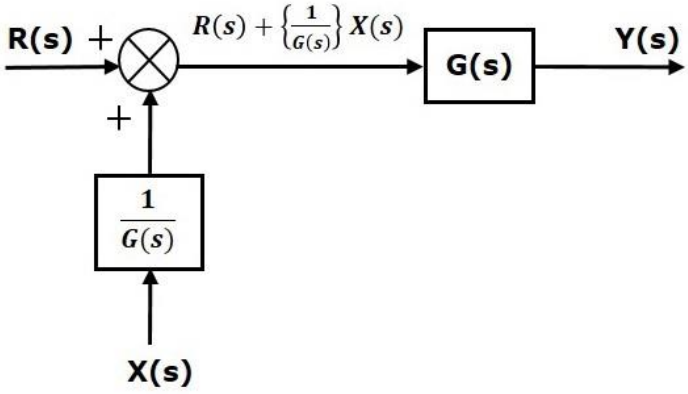
Now, shift the summing point before the block. This block diagram is shown in the following figure.



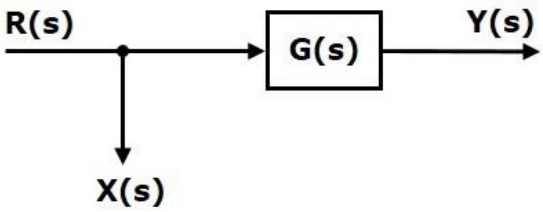
Output of this block diagram is -

$$Y(S)=G(s)R(s)+G(s)X(s)$$

The first term 'G(s)R(s)′ is same in both equations. But there is difference in the second term. In order to get the second term also same, we require one more block $\frac{1}{G(s)}$

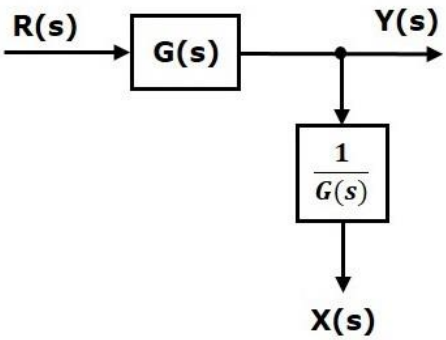


6. **Shifting take-off point after the block**

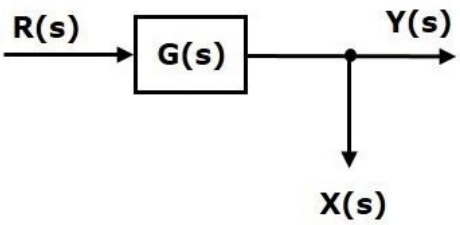


Here, X(s)=R(s) and Y(s)=G(s)R(s)

When you shift the take-off point after the block, the output Y(s) will be same. But there is difference in X(s) value. So, in order to get the same X(s) value, we require one more block 1/G(s).

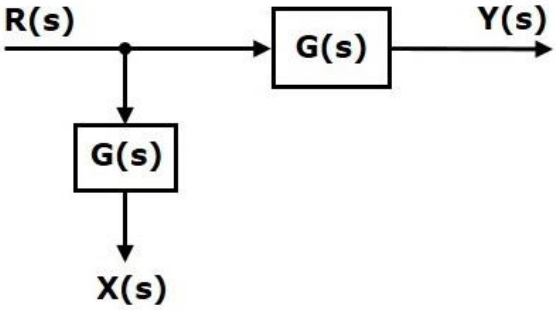


7. **Shifting take-off point after the block**



Here, X(s)=Y(s)=G(s)R(s)

When you shift the take-off point before the block, the output Y(s) will be same. But there is difference in X(s) value. So, in order to get same X(s) value, we require one more block G(s)



Signal flow graph

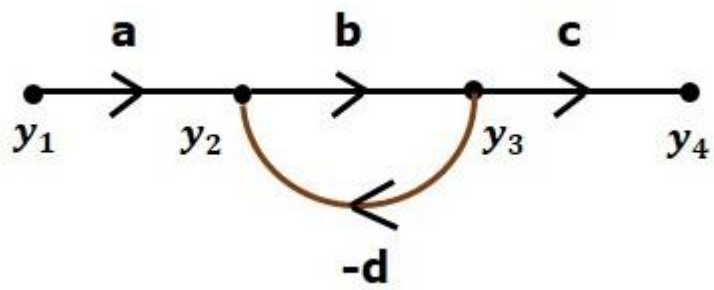
Signal flow graph is a graphical representation of algebraic equations

Nodes and branches are the basic elements of signal flow graph.

Node is a point which represents either a variable or a signal. There are three types of nodes — input node, output node and mixed node.

- **Input Node** − It is a node, which has only outgoing branches.
- **Output Node** − It is a node, which has only incoming branches.
- **Mixed Node** − It is a node, which has both incoming and outgoing branches.

Example



- The **nodes** present in this signal flow graph are **y_1, y_2, y_3** and **y_4**.

- **y_1** and **y_4** are the **input node** and **output node** respectively.
- **y_2** and **y_3** are **mixed nodes**.

Branch

Branch is a line segment which joins two nodes. It has both **gain** and **direction**. For example, there are four branches in the above signal flow graph. These branches have **gains** of **a, b, c** and **-d**.

Construction of Signal Flow Graph

Let us construct a signal flow graph by considering the following algebraic equations −

$$y_2=a_{12}y_1+a_{42}y_4$$

$$y_3=a_{23}y_2+a_{53}y_5$$

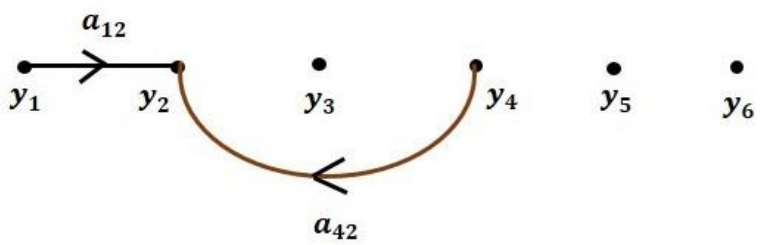
$$y_4=a_{34}y_3$$

$$y_5=a_{45}y_4+a_{35}y_3$$

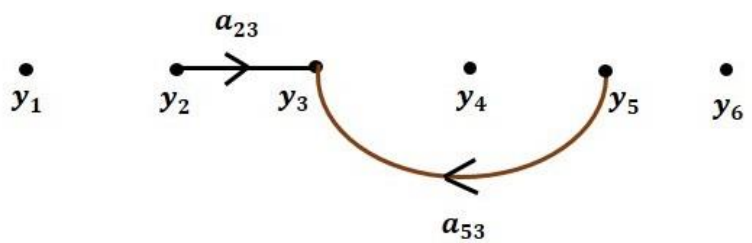
$$y_6=a_{56}y_5$$

There will be six **nodes** (y_1, y_2, y_3, y_4, y_5 and y_6) and eight **branches** in this signal flow graph. The gains of the branches are a_{12}, a_{23}, a_{34}, a_{45}, a_{56}, a_{42}, a_{53} and a_{35}.

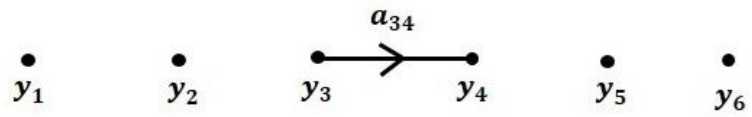
Step 1 − Signal flow graph for $y_2=a_{12}y_1+a_{42}y_4$



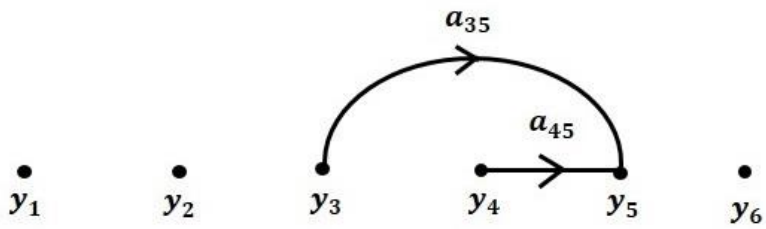
Step 2 − Signal flow graph for $y_3=a_{23}y_2+a_{53}y_5$



Step 3 − Signal flow graph for $y_4=a_{34}y_3$

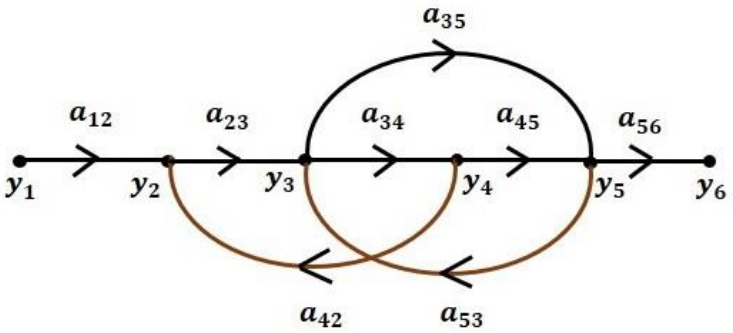


Step 4 − Signal flow graph for $y_5=a_{45}y_4+a_{35}y_3$



Step 5 − Signal flow graph for $y_6=a_{56}y_5$

Step 6 − Signal flow graph of overall system is



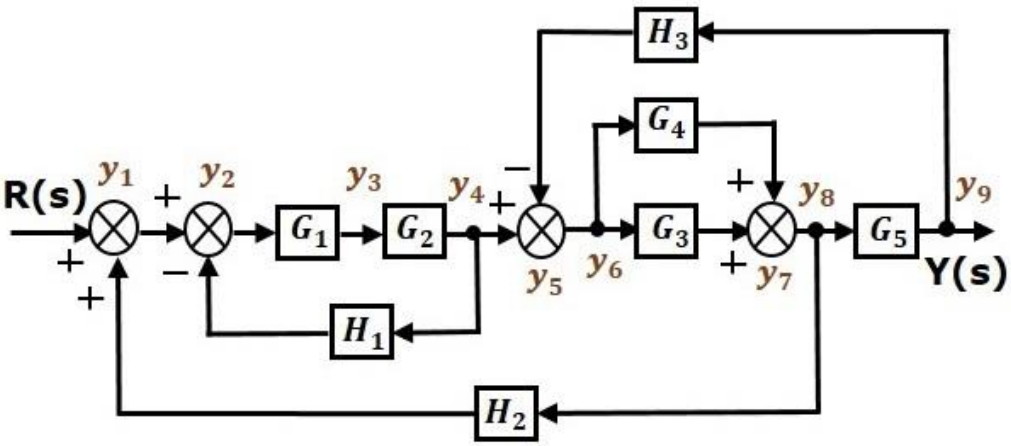
Conversion of Block Diagrams into Signal Flow Graphs

Follow these steps for converting a block diagram into its equivalent signal flow graph.

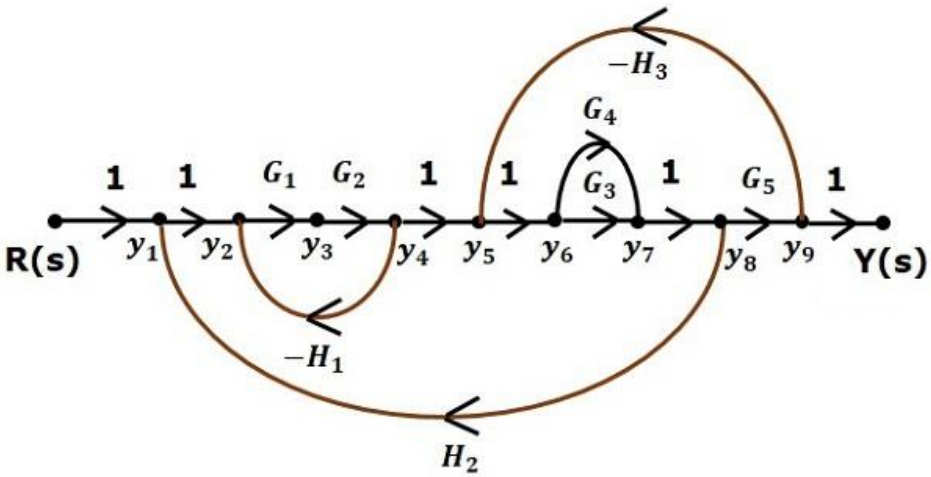
- Represent all the signals, variables, summing points and take-off points of block diagram as **nodes** in signal flow graph.
- Represent the blocks of block diagram as **branches** in signal flow graph.

- Represent the transfer functions inside the blocks of block diagram as **gains** of the branches in signal flow graph.
- Connect the nodes as per the block diagram. If there is connection between two nodes (but there is no block in between), then represent the gain of the branch as one. **For example**, between summing points, between summing point and takeoff point, between input and summing point, between take-off point and output.

Example



Represent the input signal R(s) and output signal C(s) of block diagram as input node R(s) and output node C(s) of signal flow graph.



Mason's Gain formula

Suppose there are 'N' forward paths in a signal flow graph. The gain between the input and the output nodes of a signal flow graph is nothing but the **transfer function** of the system. It can be calculated by using Mason's gain formula.

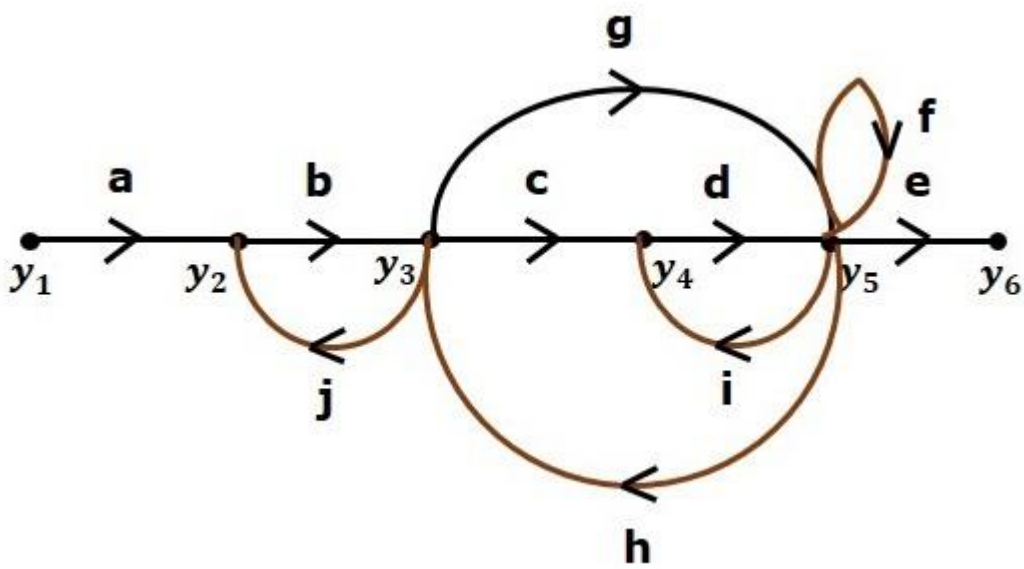
Mason's gain formula is

$$T = \frac{C(s)}{R(s)} = \frac{\Sigma_{i=1}^{N} P_i \Delta_i}{\Delta}$$

Where,

- **C(s)** is the output node
- **R(s)** is the input node
- **T** is the transfer function or gain between R(s)R(s) and C(s)C(s)
- **P_i** is the ith forward path gain
- Δ=1−(sum of all individual loop gains) +(sum of gain products of all possible two non-touching loops) −(sum of gain products o fall possible three non-touching loops)+...
- *Δ_i is obtained from Δ by removing the loops which are touching the ith forward path*

Example



Path

It is a traversal of branches from one node to any other node in the direction of branch arrows. It should not traverse any node more than once.

Examples − y2→y3→y4→y5 and y5→y3→y2

Forward Path

The path that exists from the input node to the output node is known as **forward path**.

Examples − y1→y2→y3→y4→y5→y6 and y1→y2→y3→y5→y6

Forward Path Gain

It is obtained by calculating the product of all branch gains of the forward path.

Examples − abcde is the forward path gain of y1→y2→y3→y4→y5→y6 and abge is the forward path gain of y1→y2→y3→y5→y6

Loop

The path that starts from one node and ends at the same node is known as **loop**. Hence, it is a closed path.

Examples − y2→y3→y2 and y3→y5→y3

Loop Gain

It is obtained by calculating the product of all branch gains of a loop.

Examples − bj is the loop gain of y2→y3→y2 and gh is the loop gain of y3→y5→y3

Non-touching Loops

These are the loops, which should not have any common node.

Examples − The loops, y2→y3→y2 and y4→y5→y4 are non-touching.

- Number of forward paths, N = 2.
- First forward path is - y1→y2→y3→y4→y5→y6.
- First forward path gain, p1=abcde.
- Second forward path is - y1→y2→y3→y5→y6
- Second forward path gain, p2=abge.
- Number of individual loops, L = 5.

- Loops are y2→y3→y2, y3→y5→y3, y3→y4→y5→y3, y4→y5→y4 and y5→y5.
- Loop gains are - l1=bj, l2=gh, l3=cdh, l4=di and l5=f.
- Number of two non-touching loops = 2.
- First non-touching loops pair is - y2→y3→y2, y4→y5→y4.
- Gain product of first non-touching loops pair, l1l4=bjdi
- Second non-touching loops pair is - y2→y3→y2, y5→y5.
- Gain product of second non-touching loops pair is - l1l5=bjf

Higher number of (more than two) non-touching loops are not present in this signal flow graph.

Δ=1− (sum of all individual loop gains) + (sum of gain products of all possible two non-touching loops) − (sum of gain products of all possible three non-touching loops)+…

Substitute the values in the above equation,

Δ=1−(bj+gh+cdh+di+f) +(bjdi+bjf) −(0)
⇒Δ=1−(bj+gh+cdh+di+f) +bjdi+bjf

There is no loop which is non-touching to the first forward path.

So, Δ1=1.
Similarly, Δ2=1. Since, no loop which is non-touching to the second forward path.

Substitute, N = 2 in Mason's gain formula

$$T = \frac{C(s)}{R(s)} = \frac{\Sigma_{i=1}^{2} P_i \Delta_i}{\Delta}$$

$$T = \frac{C(s)}{R(s)} = \frac{P_1 \Delta_1 + P_2 \Delta_2}{\Delta}$$

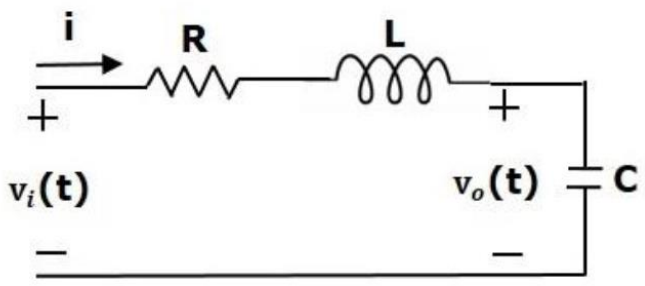
Substitute all the necessary values in the above equation

$$T = \frac{C(s)}{R(s)} = \frac{(abcde)1 + (abge)1}{1 - (bj + gh + cdh + di + f) + bjdi + bjf}$$

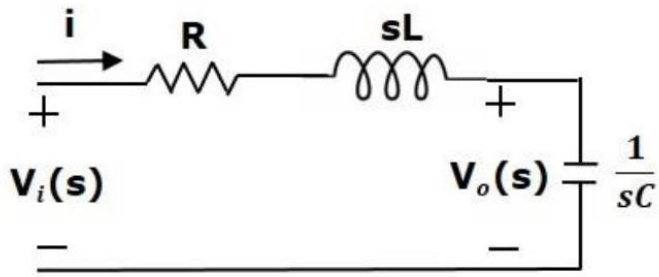
$$\Rightarrow T = \frac{C(s)}{R(s)} = \frac{(abcde) + (abge)}{1 - (bj + gh + cdh + di + f) + bjdi + bjf}$$

Block Diagram Representation of Electrical Systems

Consider a series of RLC circuit as shown in the following figure. Where, Vi(t) and Vo(t) are the input and output voltages. Let $i(t)$ be the current passing through the circuit. This circuit is in time domain.



By applying the Laplace transform to this circuit, will get the circuit in s-domain.



Chapter 2

Transient Response

In practice, the input signal to a control system is not known ahead of time but is random in nature.
Only in some special cases is the input signal known in advance and expressible analytically.

In analysing and designing control systems set up by specifying particular test input signals and by comparing the responses of various systems to these input signals.

Standard Test Signals

The commonly used test input signals are step functions, ramp functions, acceleration functions, impulse functions, sinusoidal functions.
We use test signals such as step, ramp, acceleration and impulse signals. With these test signals, mathematical and experimental analyses of control systems can be carried out easily, since the signals are very simple functions of time.

- If the inputs to a control system are gradually changing functions of time, then a **ramp function of time** may be a good test signal.
- If a system is subjected to sudden disturbances, a **step function of time** may be a good test signal
- For a system subjected to shock inputs, an **impulse function** may be best.
-

Transient Response and Steady-State Response The time response of a control system consists of two parts:

- Transient response: which goes from the initial state to the final state

- Steady-state response: the manner in which the system output behaves as t approaches infinity

Thus the system response c(t) may be written as

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

Absolute Stability, Relative Stability, and Steady-State Error.

- **Absolute stability**: whether the system is stable or unstable.
- A control system is in **equilibrium** if, in the absence of any disturbance or input, the output stays in the same state.
- A linear time-invariant control system is **stable** if the output eventually comes back to its equilibrium state when the system is subjected to an initial condition.
- A linear time-invariant control system is **critically stable** if oscillations of the output continue forever.
- It is **unstable** if the output diverges without bound from its equilibrium state when the system is subjected to an initial condition.

FIRST-ORDER SYSTEMS

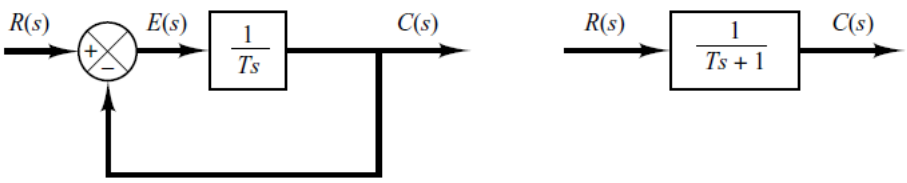


FIG: 2.1 Block diagram of a first-order system

Consider the first-order system shown in Figure 2.1

The input-output relationship is given by

$$\frac{C(s)}{R(s)} = \frac{1}{Ts + 1} \qquad \textbf{Eq. 2.1}$$

Let's analyse the system responses to such inputs as the unit-step, unit-ramp, and unit-impulse functions. The initial conditions are assumed to be zero.

Unit-Step Response of First-Order Systems.

Laplace transform of the unit-step function is *1/s*, substituting *R(s)=1/s* into Eq. 2.1 we obtain

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

Expanding C(s) into partial fractions gives

$$C(s) = \frac{1}{s} - \frac{T}{Ts+1} = \frac{1}{s} - \frac{1}{s+(1/T)} \qquad \textbf{Eq. 2.2}$$

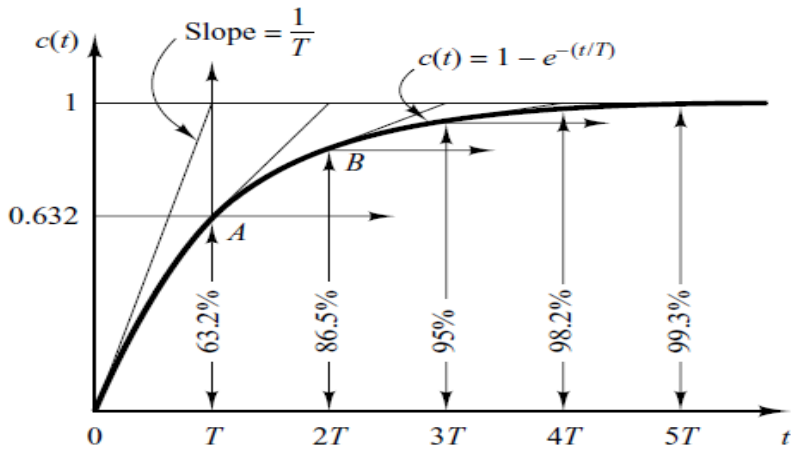
Taking the inverse Laplace transform of Eq. 2.2 we obtain

$$c(t) = 1 - e^{-t/T}, \qquad \text{for } t \geq 0 \qquad \textbf{Eq. 2.3}$$

Eq. 2.3 states that initially the output c(t) is zero i.e at t=0 and finally it becomes unity.

One important characteristic of such an exponential response curve c(t) is that at t=T the value of c(t) is 0.632, or the response c(t) has reached 63.2% of its total change. This may be easily seen by substituting t=T in c(t).That is,

$$c(T) = 1 - e^{-1} = 0.632$$



The smaller the time constant T, the faster the system response.

Another important characteristic of the exponential response curve is that the slope of the tangent line at t=0 is 1/T, since

$$\left.\frac{dc}{dt}\right|_{t=0} = \left.\frac{1}{T}e^{-t/T}\right|_{t=0} = \frac{1}{T} \qquad \textbf{Eq. 2.4}$$

In one time constant, the exponential response curve has gone from 0 to 63.2%of the final

value. In two time constants, the response reaches 86.5%of the final value.At t=3T, 4T, and 5T, the response reaches 95%, 98.2%, and 99.3%, respectively, of the final value. Thus, for t≥ 4T, the response remains within 2% of the final value.

Unit-Ramp Response of First-Order Systems

Laplace transform of the unit-ramp function is $1/s^2$, we obtain the output of the system of Figure: 2.1 as

$$C(s) = \frac{1}{Ts+1}\frac{1}{s^2}$$

Expanding C(s) into partial fractions gives

$$C(s) = \frac{1}{s^2} - \frac{T}{s} + \frac{T^2}{Ts+1} \qquad \textbf{Eq. 2.5}$$

Taking the inverse Laplace transform of Eq. 2.5, we obtain

$$c(t) = t - T + Te^{-t/T}, \qquad \text{for } t \geq 0 \qquad \textbf{Eq. 2.6}$$

The error signal e(t) is then

$$\begin{aligned} e(t) &= r(t) - c(t) \\ &= T(1 - e^{-t/T}) \end{aligned}$$

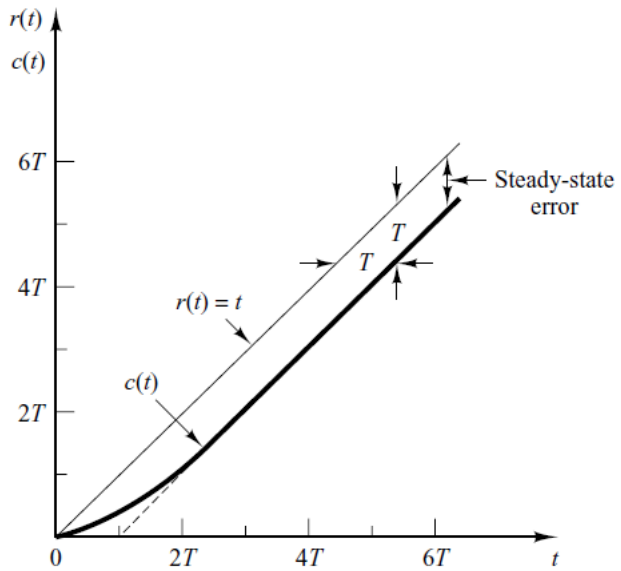


FIG: 2.3 Unit-ramp response of the system

As t approaches infinity, $e^{-t/T}$ approaches zero, and thus the error signal e(t) approaches T or

$$e(\infty) = T$$

The error in following the unit-ramp input is equal to T for sufficiently large t. The smaller the time constant T, the smaller the steady-state error in following the ramp input.

Unit-Impulse Response of First-Order Systems.

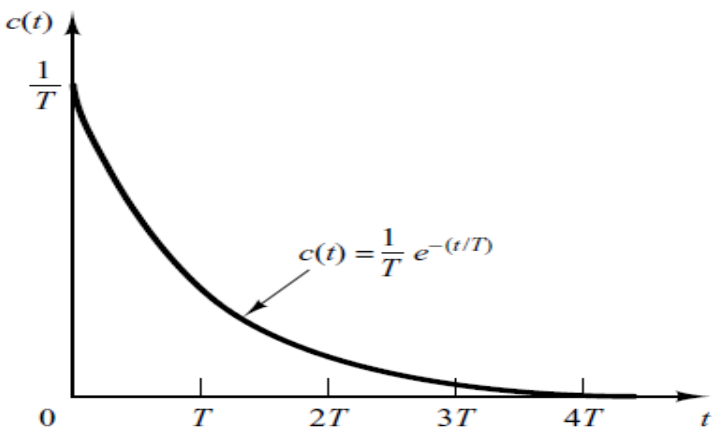
For the unit-impulse input, *R(s)=1* and the output of the system of Figure 2.1 can be obtained as

$$C(s) = \frac{1}{Ts + 1} \qquad \textbf{Eq. 2.7}$$

The inverse Laplace transform of Equation (5–7) gives

$$c(t) = \frac{1}{T} e^{-t/T}, \qquad \text{for } t \geq 0 \qquad \textbf{Eq. 2.8}$$

The response curve given by Eq. 2.8



Property of Linear Time-Invariant Systems

For the unit-ramp input the output c(t) is

$$c(t) = t - T + Te^{-t/T}, \qquad \text{for } t \geq 0$$

For the unit-step input, which is the derivative of unit-ramp input, the output c(t) is

$$c(t) = 1 - e^{-t/T}, \qquad \text{for } t \geq 0$$

For the unit-impulse input, which is the derivative of unit-step input, the output c(t) is

$$c(t) = \frac{1}{T} e^{-t/T}, \qquad \text{for } t \geq 0$$

Comparing the system responses to these three inputs clearly indicates that the response to the derivative of an input signal can be obtained by differentiating the response of the system to the original signal.

Step Response of Second-Order System

The closed-loop transfer function of a standard 2nd order system is

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \qquad \textbf{Eq. 2.9}$$

Where ω_n, the *undamped natural frequency*; and ξ, the *damping ratio* of the system.
The damping ratio ξ is the ratio of the actual damping to the critical damping.

The dynamic behavior of the second-order system can then be described in terms of two parameters ξ and ω_n.

- If 0< ξ <1, the closed-loop poles are complex conjugates and lie in the left-half s plane. The system is then called underdamped.
- the transient response is oscillatory, If ξ =0, the transient response does not die out.
- If ξ =1, the system is called critically damped.
- Overdamped systems correspond to ξ >1.

Underdamped case (0< ξ <1):

C(s)/R(s) can be written

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

The roots for a standard second order system can be calculated as $S_1, S_2 = -\xi\omega_n \pm j\omega_d$

Where $\omega_d = \omega_n\sqrt{1-\xi^2}$. The frequency ω_n is called the *damped natural frequency*. For

a unit-step input, C(s) can be written

$$C(s) = \frac{\omega_n^2}{(s^2 + 2\zeta\omega_n s + \omega_n^2)s} \qquad \textbf{Eq. 2.10}$$

The inverse Laplace transform of Eq. 2.10 can be obtained easily if C(s) is written in the following form:

$$\begin{aligned} C(s) &= \frac{1}{s} - \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\ &= \frac{1}{s} - \frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} - \frac{\zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} \end{aligned}$$

Referring to the basic Laplace inverse

$$\mathcal{L}^{-1}\left[\frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] = e^{-\zeta\omega_n t}\cos\omega_d t$$

$$\mathcal{L}^{-1}\left[\frac{\omega_d}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] = e^{-\zeta\omega_n t}\sin\omega_d t$$

Hence the inverse Laplace transform of Equation (5–11) is obtained as

$$\begin{aligned} \mathcal{L}^{-1}[C(s)] &= c(t) \\ &= 1 - e^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right) \\ &= 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\sin\left(\omega_d t + \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}\right), \qquad \text{for } t \geq 0 \qquad \textbf{Eq. 2.11} \end{aligned}$$

The error signal for this system is the difference between the input and output and is

$$\begin{aligned} e(t) &= r(t) - c(t) \\ &= e^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right), \qquad \text{for } t \geq 0 \end{aligned}$$

This error signal exhibits a damped sinusoidal oscillation. At steady state, or at t=∞, no error exists between the input and output.

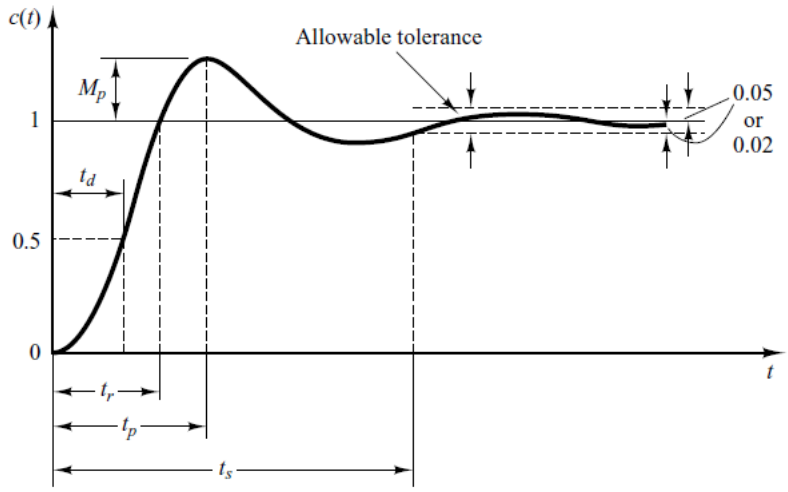
If the damping ratio ξ is equal to zero, the response becomes undamped and oscillations continue indefinitely. The response c(t) for the zero damping case may be obtained by substituting ξ =0

$$c(t) = 1 - \cos\omega_n t, \qquad \text{for } t \geq 0$$

Transient-Response Specifications.

transient-response characteristics of a control system to a unit-step input, the specifications are

- Delay time, t_d: The delay time is the time required for the response to reach half the final value the very first time.
- Rise time, t_r: The rise time is the time required for the response to rise from 10% to 90%, 5% to 95%, or 0% to 100% of its final value.
    - For underdamped second order systems, the 0%to 100%rise time is normally used.
    - For overdamped systems, the 10% to 90% rise time is commonly used.



- Peak time, t_p: The peak time is the time required for the response to reach the first peak of the overshoot.
- Maximum (percent) overshoot, *M*p: The maximum overshoot is the maximum peak value of the response curve measured from unity. If the final steady-state value of the response differs from unity, then it is common to use the maximum percent overshoot. It is defined by

$$\text{Maximum percent overshoot} = \frac{c(t_p) - c(\infty)}{c(\infty)} \times 100\%$$

- Settling time, t_s : The settling time is the time required for the response curve to reach and stay within a range about the final value of size specified by absolute percentage of the final value (usually 2% or 5%). The settling time is related to the largest time constant of the control system. Which percentage error criterion to use may be determined from the objectives of the system design in question.

Rise time t_r : Referring to Eq. 2.11 we obtain the rise time t_r by letting c(tr)=1.

$$c(t_r) = 1 = 1 - e^{-\zeta\omega_n t_r}\left(\cos\omega_d t_r + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t_r\right)$$

$e^{-\zeta\omega_n t_r} \neq 0$, Since we obtain from Equation the following equation:

$$\cos\omega_d t_r + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t_r = 0$$

Since $\omega_d = \omega_n\sqrt{1-\xi^2}$ and $\xi\omega_n = \sigma$, dividing the equation by $\sin\omega_d t_r$

$$\tan\omega_d t_r = -\frac{\sqrt{1-\zeta^2}}{\zeta} = -\frac{\omega_d}{\sigma}$$

Thus, the rise time t_r is

$$t_r = \frac{1}{\omega_d}\tan^{-1}\left(\frac{\omega_d}{-\sigma}\right) = \frac{\pi-\beta}{\omega_d}$$

Peak time t_p: Referring to Equation (5–12), we may obtain the peak time by differentiating c(t) with respect to time and letting this derivative equal zero. Since

$$\frac{dc}{dt} = \zeta\omega_n e^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right) + e^{-\zeta\omega_n t}\left(\omega_d\sin\omega_d t - \frac{\zeta\omega_d}{\sqrt{1-\zeta^2}}\cos\omega_d t\right)$$

and the cosine terms in this last equation cancel each other, *dc*/*dt*, evaluated at t=t_p, can be simplified to

$$\left.\frac{dc}{dt}\right|_{t=t_p} = (\sin\omega_d t_p)\frac{\omega_n}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t_p} = 0$$

This last equation yields the following equation:

$$\sin\omega_d t_p = 0$$

Or $\omega_d t_p = 0, \pi, 2\pi, 3\pi, \ldots$

Since the peak time corresponds to the first peak overshoot, $\omega_d t_p = \pi.$, Hence

$$t_p = \frac{\pi}{\omega_d}$$

The peak time tp corresponds to one-half cycle of the frequency of damped oscillation.

Maximum overshoot Mp: The maximum overshoot occurs at the peak time or at t=tp=p_vd.Assuming that the final value of the output is unity,Mp is obtained from Equation (5–12) as

$$\begin{aligned} M_p &= c(t_p) - 1 \\ &= -e^{-\zeta\omega_n(\pi/\omega_d)}\left(\cos\pi + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\pi\right) \\ &= e^{-(\sigma/\omega_d)\pi} = e^{-(\zeta/\sqrt{1-\zeta^2})\pi} \end{aligned}$$

The maximum percent overshoot is $e^{-(\sigma/\omega_d)\pi} \times 100\%.$

If the final value c(q) of the output is not unity, then we need to use the following equation:

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

Settling time ts : For an underdamped second-order system, the transient response is obtained from Equation (5–12) as

$$c(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\sin\left(\omega_d t + \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}\right), \qquad \text{for } t \geq 0$$

The speed of decay of the transient response depends on the value of the time constant 1/ $\xi\omega_n$.
For a given vn, the settling time ts is a function of the damping ratio z
The settling time corresponding to a ; 2% or ;5% tolerance band may be measured in terms of the time constant T=1/ $\xi\omega_n$. For 0<z<0.9, if the 2% criterion is used, ts is approximately four times the time constant of the system. If the 5% criterion is used, then ts is approximately three times the time constant.
For convenience in comparing the responses of systems, we commonly define the settling time ts to be

$$t_s = 4T = \frac{4}{\sigma} = \frac{4}{\zeta\omega_n} \qquad \text{(2\% criterion)}$$

$$t_s = 3T = \frac{3}{\sigma} = \frac{3}{\zeta\omega_n} \qquad \text{(5\% criterion)}$$

Application of initial and final value theorem

Module-II
Chapter-3
Concept of Stability

3.1 Routh-Hurwitz Criteria

- Under what conditions will a system become unstable?
- If it is unstable, how should we stabilize the system?
- Control system is stable if and only if all closed-loop poles lie in the left-half s plane.

$$\frac{C(s)}{R(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1} s + b_m}{a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n} = \frac{B(s)}{A(s)}$$

where the a's and b's are constants and m ≤ n.

3.1.1 Routh's stability criterion

- The number of closed-loop poles that lie in the right-half s-plane without having to factor the denominator polynomial.
- whether or not there are unstable roots in a polynomial equation without actually solving for them.
- Applies to polynomials with only a finite number of terms. When the criterion is applied to a control system, information about absolute stability can be obtained directly from the coefficients of the characteristic equation.

The procedure in Routh's stability criterion

1. Write the polynomial in s in the following form:

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

   where the coefficients are real quantities. We assume that an ≠ 0; that is, any zero root has been removed.

2. If any of the coefficients are zero or negative in the presence of at least one positive coefficient, a root or roots exist that are imaginary or that have positive real parts. **Therefore, in such a case, the system is not stable.**

3. If all coefficients are positive, arrange the coefficients of the polynomial in rows and columns according to the following pattern:

$$
\begin{array}{llllll}
s^n & a_0 & a_2 & a_4 & a_6 & \cdots \\
s^{n-1} & a_1 & a_3 & a_5 & a_7 & \cdots \\
s^{n-2} & b_1 & b_2 & b_3 & b_4 & \cdots \\
s^{n-3} & c_1 & c_2 & c_3 & c_4 & \cdots \\
s^{n-4} & d_1 & d_2 & d_3 & d_4 & \cdots \\
\cdot & \cdot & \cdot & & & \\
\cdot & \cdot & \cdot & & & \\
\cdot & \cdot & \cdot & & & \\
s^2 & e_1 & e_2 & & & \\
s^1 & f_1 & & & & \\
s^0 & g_1 & & & &
\end{array}
$$

The process of forming rows continues until we run out of elements. (The total number of rows is n+1.) The coefficients b_1, b_2, b_3, and so on, are evaluated as follows:

$$
b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1}
$$

$$
b_2 = \frac{a_1 a_4 - a_0 a_5}{a_1}
$$

$$
b_3 = \frac{a_1 a_6 - a_0 a_7}{a_1}
$$

$$
\cdot
$$

The evaluation of the b's is continued until the remaining ones are all zero. The same pattern of cross-multiplying the coefficients of the two previous rows is followed in evaluating the c's, d's, e's, and so on. That is,

$$
c_1 = \frac{b_1 a_3 - a_1 b_2}{b_1}
$$

$$
c_2 = \frac{b_1 a_5 - a_1 b_3}{b_1}
$$

$$
c_3 = \frac{b_1 a_7 - a_1 b_4}{b_1}
$$

$$
\cdot
$$

Routh's stability criterion states that the number of roots with positive real parts is equal to the number of changes in sign of the coefficients of the first column of the array.

Example

Consider the following polynomial:

$$s^4 + 2s^3 + 3s^2 + 4s + 5 = 0$$

Let us follow the procedure just presented and construct the array of coefficients. The first two rows can be obtained directly from the given polynomial. The remaining terms are obtained from these. If any coefficients are missing, they may be replaced by zeros in the array.

$$\begin{array}{c|ccc||c|ccc|l}
s^4 & 1 & 3 & 5 & s^4 & 1 & 3 & 5 & \\
s^3 & 2 & 4 & 0 & s^3 & \cancel{2} & \cancel{4} & \cancel{0} & \text{The second row is divided} \\
 & & & & & 1 & 2 & 0 & \text{by 2.} \\
s^2 & 1 & 5 & & s^2 & 1 & 5 & & \\
s^1 & -6 & & & s^1 & -3 & & & \\
s^0 & 5 & & & s^0 & 5 & & &
\end{array}$$

the number of changes in sign of the coefficients in the first column is 2. This means that there are two roots with positive real parts

3.1.2 Special Cases

1. If a first-column term in any row is zero, but the remaining terms are not zero or there is no remaining term, then the zero term is replaced by a very small positive number ε and the rest of the array is evaluated.

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

The array of coefficients is

$$\begin{array}{cc}
s^3 & 1 \quad 1 \\
s^2 & 2 \quad 2 \\
s^1 & 0 \approx \epsilon \\
s^0 & 2
\end{array}$$

If the sign of the coefficient above the zero (ε) is the same as that below it, it indicates that there are a pair of imaginary roots. Actually, has two roots at s = ±j.

2. If all the coefficients in any derived row are zero, it indicates that there are roots of equal magnitude lying radially opposite in the s plane—that is, two real roots with equal magnitudes and opposite signs and/or two conjugate imaginary roots.

Example

$$s^5 + 2s^4 + 24s^3 + 48s^2 - 25s - 50 = 0$$

The array of coefficients is

$$\begin{array}{lrrl} s^5 & 1 & 24 & -25 \\ s^4 & 2 & 48 & -50 \quad \leftarrow \text{Auxiliary polynomial } P(s) \\ s^3 & 0 & 0 & \end{array}$$

The terms in the s^3 row are all zero.

The auxiliary polynomial is then formed from the coefficients of the s^4 row

$$P(s) = 2s^4 + 48s^2 - 50$$

which indicates that there are two pairs of roots of equal magnitude and opposite sign

These pairs are obtained by solving the auxiliary polynomial equation P(s)=0. The derivative of P(s) with respect to s is

$$\frac{dP(s)}{ds} = 8s^3 + 96s$$

The terms in the s^3 row are replaced by the coefficients of the last equation

$$\begin{array}{lrrl} s^5 & 1 & 24 & -25 \\ s^4 & 2 & 48 & -50 \\ s^3 & 8 & 96 & \quad \leftarrow \text{Coefficients of } dP(s)/ds \\ s^2 & 24 & -50 & \\ s^1 & 112.7 & 0 & \\ s^0 & -50 & & \end{array}$$

There is one change in sign in the first column of the new array. Thus, the original equation has one root with a positive real part. By solving for roots of the auxiliary polynomial equation,

$$2s^4 + 48s^2 - 50 = 0$$

we obtain

$$s^2 = 1, \qquad s^2 = -25$$

$$s = \pm 1, \qquad s = \pm j5$$

3.2 Application of Routh's Stability Criterion to Control-System Analysis

Routh's stability criterion is of limited usefulness in linear control-system analysis, mainly because it does not suggest how to improve relative stability or how to stabilize an unstable system.

Let us determine the range of K for stability. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{K}{s(s^2 + s + 1)(s + 2) + K}$$

The characteristic equation is

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

The array of coefficients becomes

$$\begin{array}{cccc} s^4 & 1 & 3 & K \\ s^3 & 3 & 2 & 0 \\ s^2 & \frac{7}{3} & K & \\ s^1 & 2 - \frac{9}{7}K & & \\ s^0 & K & & \end{array}$$

For stability, K must be positive, and all coefficients in the first column must be positive. Therefore,

$$\frac{14}{9} > K > 0$$

When K = 14/9 the system becomes oscillatory and, mathematically, the oscillation is sustained at constant amplitude.

3.3 Relative Stability Analysis

Routh's stability criterion provides the answer to the question of absolute stability.

A useful approach for examining relative stability is to shift the s-plane axis and apply Routh's stability criterion. That is, we substitute

$$s = \hat{s} - \sigma \qquad (\sigma = \text{constant})$$

into the characteristic equation of the system, write the polynomial in terms of ŝ and apply Routh's stability criterion to the new polynomial in ŝ. The number of changes of sign in the first column of the array developed for the polynomial in ŝ is equal to the number of roots that are located to the right of the vertical line s=–σ. Thus, this test reveals the number of roots that lie to the right of the vertical line s=–σ.

3.4 Root Locus Method

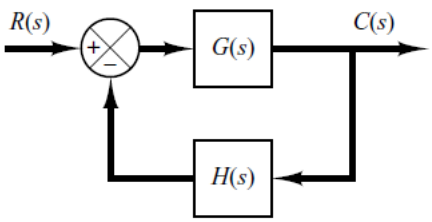
The basic characteristic of the transient response of a closed-loop system is closely related to the location of the closed-loop poles. If the system has a variable loop gain, then the location of the closed-loop poles depends on the value of the loop gain chosen.

- It is important, therefore, that the designer know how the closed-loop poles move in the s plane as the loop gain is varied.
- The *root-locus method*, is one in which the roots of the characteristic equation are plotted for all values of a system parameter. The roots corresponding to a particular value of this parameter can then be located on the resulting graph.
- The gain of the open-loop transfer function is the parameter to be varied through all values, from zero to infinity.

Angle and Magnitude Conditions

Consider the negative feedback system . The closed-loop transfer function is

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



The characteristic equation

$$1 + G(s)H(s) = 0$$

or

$$G(s)H(s) = -1$$

can be split into two equations by equating the angles and magnitudes of both sides, respectively, to obtain the following:

Angle condition:

$$\underline{/G(s)H(s)} = \pm 180^\circ(2k+1) \qquad (k = 0, 1, 2, \ldots)$$

Magnitude condition:

$$|G(s)H(s)| = 1$$

- A locus of the points in the complex plane satisfying the angle condition alone is the root locus.
- The roots of the characteristic equation (the closed-loop poles) corresponding to a given value of the gain can be determined from the magnitude condition.

3.4.1 Rules for Construction of Root Locus

1. Locate the open loop poles and zeros in the 's' plane.
2. Find the number of root locus branches.
   - The root locus branches start at the open loop poles and end at open loop zeros. So, the number of root locus branches **N** is equal to the number of finite open loop poles **P** or the number of finite open loop zeros **Z**, whichever is greater.
   - The number of root locus branches **N** as

$$N = P \quad \text{if} \quad P \geq Z$$

$$N = Z \quad \text{if} \quad P < Z$$

3. Identify and draw the **real axis root locus branches**.
   - If the angle of the open loop transfer function at a point is an odd multiple of 180^0, then that point is on the root locus.
   - If odd number of the open loop poles and zeros exist to the left side of a point on the real axis, then that point is on the root locus branch.
4. Find the centroid and the angle of asymptotes.

$$\alpha = \frac{\sum Real\ part\ of\ finite\ open\ loop\ poles - \sum Real\ part\ of\ finite\ open\ loop\ zeros}{P-Z}$$

5. The formula for the angle of **asymptotes θ** is

$$\theta = \frac{(2q+1)180^0}{P-Z}$$

Where, q=0,1,2,....,(P−Z)−1

6. Find the intersection points of root locus branches with an imaginary axis.

We can calculate the point at which the root locus branch intersects the imaginary axis and the value of **K** at that point by using the Routh array method.

- If all elements of any row of the Routh array are zero, then the root locus branch intersects the imaginary axis and vice-versa.
- Identify the row in such a way that if we make the first element as zero, then the elements of the entire row are zero. Find the value of **K** for this combination.
- Substitute this **K** value in the auxiliary equation. You will get the intersection point of the root locus branch with an imaginary axis.

7. Find Break-away and Break-in points.

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

Note − Break-away and break-in points exist only on the real axis root locus branches.

8. Follow these steps to find break-away and break-in points.

- Write K in terms of s from the characteristic equation 1+G(s)H(s)=0.
- Differentiate K with respect to s and make it equal to zero. Substitute these values of s in the above equation.
- The values of s for which the K value is positive are the **break points**.

9. Find the angle of departure and the angle of arrival.

The Angle of departure and the angle of arrival can be calculated at complex conjugate open loop poles and complex conjugate open loop zeros respectively.

- The formula for the **angle of departure** ϕ_d is

$$\phi_d = 180^0 - \phi$$

The formula for the **angle of arrival** ϕ_a is

$$\phi_a = 180^0 + \phi$$

Where,

$$\phi = \sum \phi_P - \sum \phi_Z$$

Chapter-4

Frequency response

The steady-state response of a system to a sinusoidal input.

The frequency of the input signal is varied over a certain range and study the resulting response.

4.1 A Bode diagram consists of two graphs:

- Plot of the logarithm of the magnitude of a sinusoidal transfer function
- Plot of the phase angle; both are plotted against the frequency on a logarithmic scale.

The standard representation of the logarithmic magnitude of G(jω) is 20 log |G(jω)|, where the base of the logarithm is 10.

The unit used in this representation of the magnitude is the decibel, usually abbreviated dB.

The main advantage of using the Bode diagram

- multiplication of magnitudes can be converted into addition.
- Furthermore, a simple method for sketching an approximate log-magnitude curve is available. It is based on asymptotic approximations.
- Such approximation by straight-line asymptotes is sufficient if only rough information on the frequency-response characteristics is needed.
- Expanding the low-frequency range by use of a logarithmic scale for the frequency is highly advantageous, since characteristics at low frequencies are most important in practical systems. Although it is not possible to plot the curves right down to zero frequency because of the logarithmic frequency (log0 = -∞), this does not create a serious problem.

4.2 Basic Factors of *G*(*j*ω)*H*(*j*ω).

1. Gain K
2. Integral and derivative factors $(j\omega)^{\mp 1}$
3. First-order factors $(1 + j\omega T)^{\mp 1}$
4. Quadratic factors $\left[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2\right]^{\mp 1}$

The Gain *K*.

- A number greater than unity has a positive value in decibels, while a number smaller than unity has a negative value.
- The log-magnitude curve for a constant gain K is a horizontal straight line at the magnitude of 20 logK decibels.
- The phase angle of the gain K is zero.
- The effect of varying the gain K in the transfer function is that it raises or lowers the log-magnitude curve of the transfer function by the corresponding constant amount, but it has no effect on the phase curve.
- As a number increases by a factor of 10, the corresponding decibel value increases by a factor of 20.

Example 20 log (K * 10) = 20 logK + 20

Similarly

$$20 \log AK * 10^{n}B = 20 \log K + 20n$$

Integral and Derivative Factors $(j\omega)^{\pm 1}$

The logarithmic magnitude of 1/jω in decibels is

$$20 \log \left| \frac{1}{j\omega} \right| = -20 \log \omega \text{ dB}$$

- The phase angle of 1/jω is constant and equal to –90°.
- If the log magnitude –20 logω dB is plotted against ω on a logarithmic scale, it is a straight line. To draw this straight line, we need to locate one point (0 dB, ω=1)on it.

(-20 log 10ω) dB = (-20 logω - 20) dB

- the slope of the line is –20 dB/decade

Similarly, the log magnitude of jω in decibels is

20 log **|jω|** = 20 logω dB

- The phase angle of jω is constant and equal to 90°.The log-magnitude curve is a straight line with a slope of 20 dB/decade.

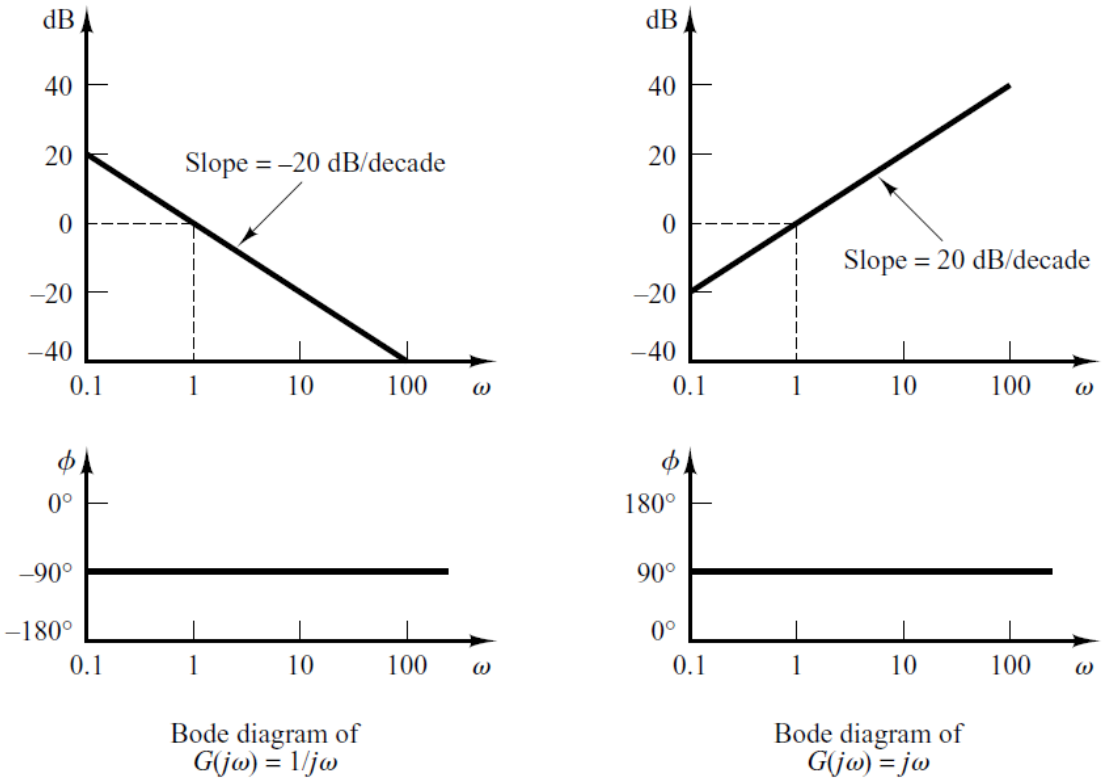
If the transfer function contains the factor (1/jω)n or (jω)n, the log magnitude becomes, respectively

$$20 \log \left| \frac{1}{(j\omega)^n} \right| = -n \times 20 \log |j\omega| = -20n \log \omega \text{ dB}$$

or

$$20 \log |(j\omega)^n| = n \times 20 \log |j\omega| = 20n \log \omega \text{ dB}$$

The slopes of the log-magnitude curves for the factors (1/jω)n and (jω)n are thus –20n dB/decade and 20n dB/decade, respectively. The phase angle of (1/jω)n is equal to –90°*n over the entire frequency range, while that of (jω)n is equal to 90°*n over the entire frequency range. The magnitude curves will pass through the point (0 dB, ω=1).



First-Order Factors (1+ *jωT*)$^{\pm 1}$. The log magnitude of the first-order factor 1/(1+jωT) is

$$20 \log \left| \frac{1}{1 + j\omega T} \right| = -20 \log \sqrt{1 + \omega^2 T^2} \text{ dB}$$

For low frequencies, such that ω<<1/T, the log magnitude may be approximated by

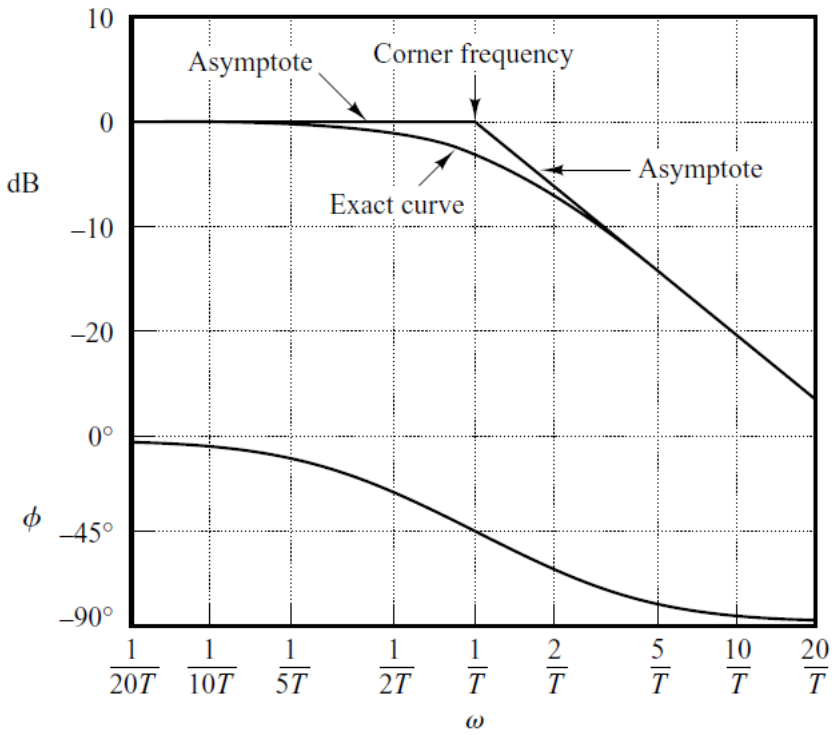
$$-20 \log \sqrt{1 + \omega^2 T^2} \doteqdot -20 \log 1 = 0 \text{ dB}$$

Thus, the log-magnitude curve at low frequencies is the constant 0-dB line. For high frequencies, such that ω>>1/T,

$$-20 \log \sqrt{1 + \omega^2 T^2} \doteqdot -20 \log \omega T \text{ dB}$$

At ω=1/T, the log magnitude equals 0 dB; at ω=10/T, the log magnitude is –20 dB. Thus, the value of –20 log ωT dB decreases by 20 dB for every decade of ω. For v >>1/T, the log-magnitude curve is thus a straight line with a slope of –20 dB/decade the logarithmic representation of the frequency-response curve of the factor 1/(1+jωT) can be approximated by

two straight-line asymptotes, one a straight line at 0 dB for the frequency range 0<ω<1/T and the other a straight line with slope –20 dB/decade for the frequency range 1/T<ω<∞.



The frequency at which the two asymptotes meet is called the *corner* frequency or *break* frequency. For the factor 1/(1+jωT), the frequency ω=1/T is the corner frequency, since at ω=1/T the two asymptotes have the same value.

The exact phase angle f of the factor 1/(1+jωT) is

$$\phi = -\tan^{-1}\omega T$$

At zero frequency, the phase angle is 0°. At the corner frequency, the phase angle is

$$\phi = -\tan^{-1}\frac{T}{T} = -\tan^{-1}1 = -45^\circ$$

At infinity, the phase angle becomes –90°.

Quadratic Factors [1 + 2ξ(*jω*/ω_n) + (*jω*/ω_n)^2]$^{\pm 1}$.

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

If $\xi>1$, this quadratic factor can be expressed as a product of two first-order factors with real poles. If $0< \xi <1$, this quadratic factor is the product of two complex conjugate factors. Asymptotic approximations to the frequency-response curves are not accurate for a factor with low values of ξ. This is because the magnitude and phase of the quadratic factor depend on both the corner frequency and the damping ratio ξ.

$$20 \log \left| \frac{1}{1 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + \left(j\frac{\omega}{\omega_n}\right)^2} \right| = -20 \log \sqrt{\left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\frac{\omega}{\omega_n}\right)^2}$$

for low frequencies such that $\omega \ll \omega_n$, the log magnitude becomes

$$-20 \log 1 = 0 \text{ dB}$$

The low-frequency asymptote is thus a horizontal line at 0 dB. For high frequencies such that $\omega \gg \omega_n$, the log magnitude becomes

$$-20 \log \frac{\omega^2}{\omega_n^2} = -40 \log \frac{\omega}{\omega_n} \text{ dB}$$

The equation for the high-frequency asymptote is a straight line having the slope –40 dB/decade, since

$$-40 \log \frac{10\omega}{\omega_n} = -40 - 40 \log \frac{\omega}{\omega_n}$$

The phase angle of the quadratic factor

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

4.3 General Procedure for Plotting Bode Diagrams

- Rewrite the sinusoidal transfer function $G(j\omega)H(j\omega)$ as a product of basic factors
- Identify the corner frequencies associated with these basic factors
- Draw the asymptotic log-magnitude curves with proper slopes between the corner frequencies.

- The phase-angle curve of G(jω)H(jω) can be drawn by adding the phase-angle curves of individual factors.

Example

Draw the Bode diagram for the following transfer function:

$$G(j\omega) = \frac{10(j\omega + 3)}{(j\omega)(j\omega + 2)[(j\omega)^2 + j\omega + 2]}$$

Solution:

To avoid any possible mistakes in drawing the log-magnitude curve, it is desirable to put G(jω) in the following normalized form, where the low-frequency asymptotes for the first-order factors and the second-order factor are the 0 dB line:

$$G(j\omega) = \frac{7.5\left(\frac{j\omega}{3} + 1\right)}{(j\omega)\left(\frac{j\omega}{2} + 1\right)\left[\frac{(j\omega)^2}{2} + \frac{j\omega}{2} + 1\right]}$$

This function is composed of the following factors:

$$7.5, \qquad (j\omega)^{-1}, \qquad 1 + j\frac{\omega}{3}, \qquad \left(1 + j\frac{\omega}{2}\right)^{-1}, \qquad \left[1 + j\frac{\omega}{2} + \frac{(j\omega)^2}{2}\right]^{-1}$$

- The corner frequencies of the third, fourth, and fifth terms are ω=3, ω=2, and $\omega = \sqrt{2}$ respectively. Note that the last term has the damping ratio of 0.3536.
- To plot the Bode diagram, the separate asymptotic curves for each of the factors. The composite curve is then obtained by algebraically adding the individual curves.

Note that when the individual asymptotic curves are added at each frequency, the slope of the composite curve is cumulative. Below ω = 2 the plot has the slope of –20 dB/decade. At the first corner frequency $\omega = \sqrt{2}$ the slope changes to –60 dB/decade and continues to the next corner frequency ω =2, where the slope becomes –80 dB/decade. At the last corner frequency ω =3, the slope changes to –60 dB/decade.

Start drawing the lowest-frequency portion of the straight line (that is, the straight line with the slope –20 dB/decade for $\omega = \sqrt{2}$). As the frequency is increased, we get the effect of the complex-conjugate poles (quadratic term) at the corner frequency $\omega = \sqrt{2}$.

The complex-conjugate poles cause the slopes of the magnitude curve to change from –20 to –60 dB/decade. At the next corner frequency, ω =2, the effect of the pole is to change the slope to –80 dB/decade. Finally, at the corner frequency ω =3, the effect of the zero is to change the slope from –80 to –60 dB/decade.

For plotting the complete phase-angle curve, the phase-angle curves for all factors have to be sketched. The algebraic sum of all phase-angle curves provides the complete phase-angle curve.

4.4 Stability Analysis using Bode Plots

From the Bode plots, we can say whether the control system is stable, marginally stable or unstable based on the values of these parameters.

- Gain cross over frequency and phase cross over frequency
- Gain margin and phase margin

Phase Cross over Frequency

The frequency at which the phase plot is having the phase of -180^0 is known as **phase cross over frequency**. It is denoted by ω_{pc}. The unit of phase cross over frequency is **rad/sec**.

Gain Cross over Frequency

The frequency at which the magnitude plot is having the magnitude of zero dB is known as **gain cross over frequency**. It is denoted by ω_{gc}. The unit of gain cross over frequency is **rad/sec**.

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

- If the phase cross over frequency ω_{pc} is greater than the gain cross over frequency ω_{gc}, then the control system is **stable**.
- If the phase cross over frequency ω_{pc} is equal to the gain cross over frequency ω_{gc}, then the control system is **marginally stable**.
- If the phase cross over frequency ω_{pc} is less than the gain cross over frequency ω_{gc}, then the control system is **unstable**.

Gain Margin

Gain margin GM is equal to negative of the magnitude in dB at phase cross over frequency.

$$GM=20\log(1/Mpc)=-20\log Mpc$$

Where, Mpc is the magnitude at phase cross over frequency. The unit of gain margin (GM) is **dB**.

Phase Margin

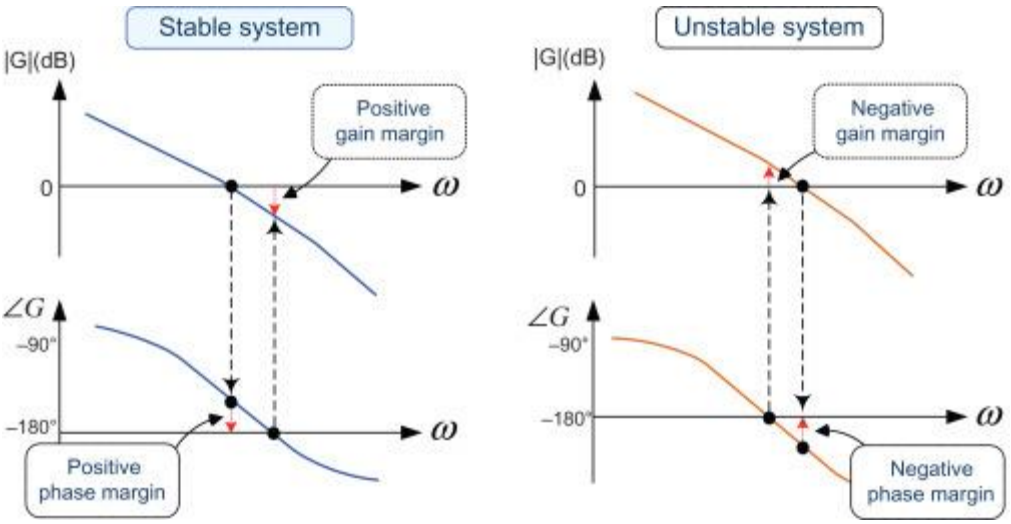
The formula for phase margin PM is

$$PM=1800+\phi_{gc}$$

Where, ϕ_{gc} is the phase angle at gain cross over frequency. The unit of phase margin is **degrees**.

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

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

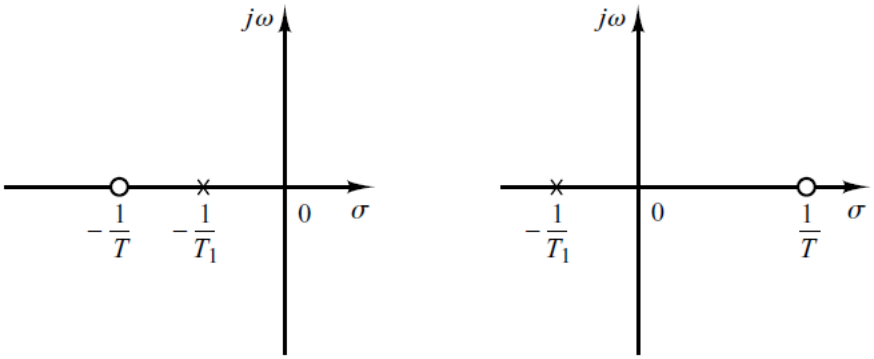


4.5 Minimum-Phase Systems and Nonminimum-Phase Systems

Transfer functions having neither poles nor zeros in the right-half s plane are minimum-phase transfer functions, whereas those having poles and/or zeros in the right-half s plane are nonminimum-phase transfer functions. Systems with minimum-phase transfer functions are called *minimum-phase* systems, whereas those with nonminimum-phase transfer functions are called *nonminimum-phase* systems.

It is noted that for a minimum-phase system, the transfer function can be uniquely determined from the magnitude curve alone. For a nonminimum-phase system, this is not the case.

Multiplying any transfer function by all-pass filters does not alter the magnitude curve, but the phase curve is changed.



$G_1(s) = \frac{1 + Ts}{1 + T_1 s}$ $\qquad$ $G_2(s) = \frac{1 - Ts}{1 + T_1 s}$

Pole–zero configurations of a minimum-phase system G_1(s) and nonminimum-phase system G_2(s).

4.6 Polar Plot

The polar plot of a sinusoidal transfer function G(jv) is a plot of the magnitude of G(jv) versus the phase angle of G(jv) on polar coordinates as v is varied from zero to infinity.

Thus, the polar plot is the locus of vectors @G(jv)@ /G(jv) as v is varied from zero to infinity.

The polar form of G(jω)H(jω) is

$$G(j\omega)H(j\omega)=|G(j\omega)H(j\omega)|\angle G(j\omega)H(j\omega)$$

4.6.1 Rules for Drawing Polar Plots

- substitute, s=jω in the open loop transfer function.
- Write the expressions for magnitude and the phase of G(jω)H(jω).
- Find the starting magnitude and the phase of G(jω)H(jω) by substituting ω=0. So, the polar plot starts with this magnitude and the phase angle.
- Find the ending magnitude and the phase of G(jω)H(jω) by substituting ω=∞. So, the polar plot ends with this magnitude and the phase angle.
- Check whether the polar plot intersects the real axis, by making the imaginary term of G(jω)H(jω) equal to zero and find the value(s) of ω.
- Check whether the polar plot intersects the imaginary axis, by making real term of G(jω)H(jω) equal to zero and find the value(s) of ω.

- For drawing polar plot more clearly, find the magnitude and phase of G(jω)H(jω) by considering the other value(s) of ω.

For the sinusoidal transfer function

$$G(j\omega) = \frac{1}{1 + j\omega T} = \frac{1}{\sqrt{1 + \omega^2 T^2}} \underline{/-\tan^{-1}\omega T}$$

the values of G(jv) at ω=0 and ω=1/T are, respectively

$$G(j0) = 1\underline{/0^\circ} \quad \text{and} \quad G\left(j\frac{1}{T}\right) = \frac{1}{\sqrt{2}}\underline{/-45^\circ}$$

If ω approaches infinity, the magnitude of G(jω) approaches zero and the phase angle approaches –90°.The polar plot of this transfer function is a semicircle as the frequency ω is varied from zero to infinity.The centre is located at 0.5 on the real axis, and the radius is equal to 0.5.

Let G(jω) = X + jY

Where

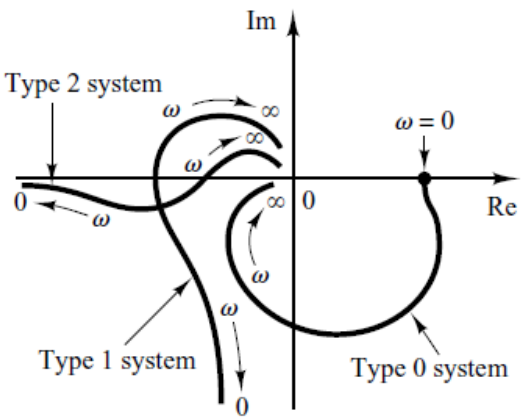
$$X = \frac{1}{1 + \omega^2 T^2} = \text{real part of } G(j\omega)$$

$$Y = \frac{-\omega T}{1 + \omega^2 T^2} = \text{imaginary part of } G(j\omega)$$

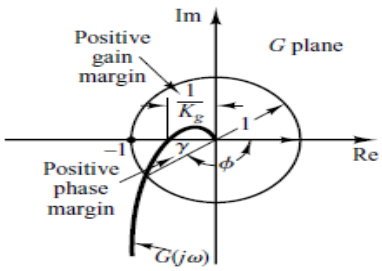
Then we obtain

$$\left(X - \frac{1}{2}\right)^2 + Y^2 = \left(\frac{1}{2}\frac{1 - \omega^2 T^2}{1 + \omega^2 T^2}\right)^2 + \left(\frac{-\omega T}{1 + \omega^2 T^2}\right)^2 = \left(\frac{1}{2}\right)^2$$

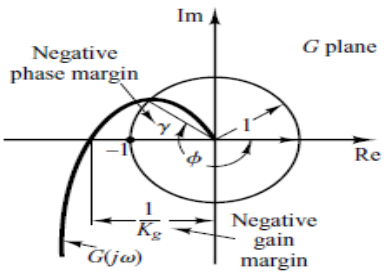
General Shapes of Polar Plots



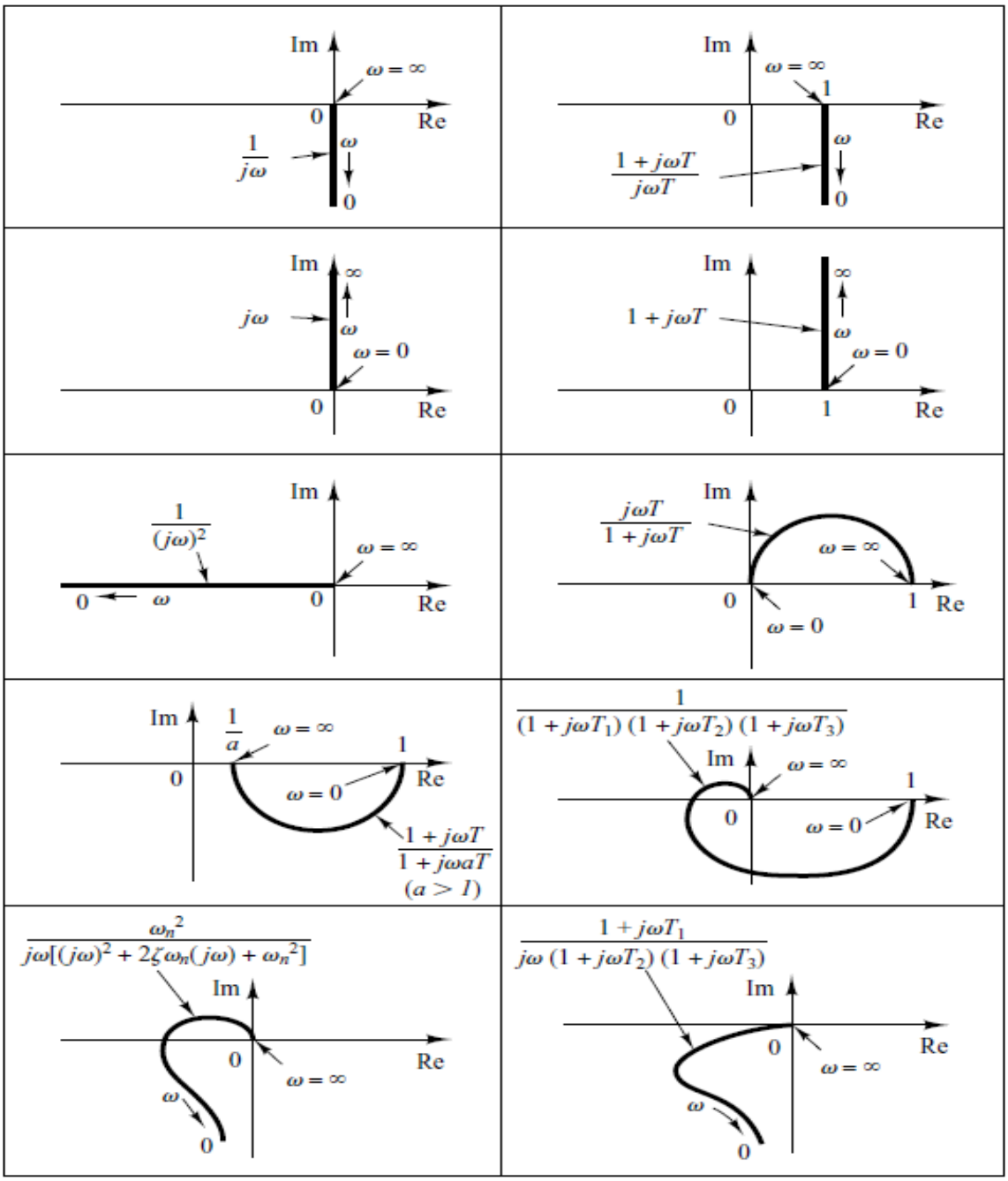
Stability from Polar Plot



Stable system



Unstable system



4.7 Nyquist Plot

Nyquist plots are the continuation of polar plots for finding the stability of the closed loop control systems by varying ω from $-\infty$ to ∞. That means, Nyquist plots are used to draw the complete frequency response of the open loop transfer function.

The Nyquist stability criterion works on the **principle of argument**. It states that if there are P poles and Z zeros are enclosed by the 's' plane closed path, then the corresponding G(s)H(s) plane must encircle the origin P−Z times. So, we can write the number of encirclements N as,

$$N=P-Z$$

- If the enclosed 's' plane closed path contains only poles, then the direction of the encirclement in the G(s)H(s) plane will be opposite to the direction of the enclosed closed path in the 's' plane.
- If the enclosed 's' plane closed path contains only zeros, then the direction of the encirclement in the G(s)H(s) plane will be in the same direction as that of the enclosed closed path in the 's' plane.

Let us now apply the principle of argument to the entire right half of the 's' plane by selecting it as a closed path. This selected path is called the **Nyquist** contour.

Closed loop control system is stable if all the poles of the closed loop transfer function are in the left half of the 's' plane. So, the poles of the closed loop transfer function are nothing but the roots of the characteristic equation. As the order of the characteristic equation increases, it is difficult to find the roots. So, let us correlate these roots of the characteristic equation as follows.

- The Poles of the characteristic equation are same as that of the poles of the open loop transfer function.
- The zeros of the characteristic equation are same as that of the poles of the closed loop transfer function.

We know that the open loop control system is stable if there is no open loop pole in the right half of the 's' plane.

$$\text{i.e., } P=0 \Rightarrow N=-Z$$

We know that the closed loop control system is stable if there is no closed loop pole in the right half of the ‘s’ plane.

$$\text{i.e., } Z=0 \Rightarrow N=P$$

Nyquist stability criterion states the number of encirclements about the critical point (1+j0) must be equal to the poles of characteristic equation, which is nothing but the poles of the open loop transfer function in the right half of the ‘s’ plane. The shift in origin to (1+j0) gives the characteristic equation plane

4.7.1 Rules for Drawing Nyquist Plots

Follow these rules for plotting the Nyquist plots.

- Locate the poles and zeros of open loop transfer function G(s)H(s) in ‘s’ plane.
- Draw the polar plot by varying ω from zero to infinity. If pole or zero present at s = 0, then varying ω from 0+ to infinity for drawing polar plot.
- Draw the mirror image of above polar plot for values of ω ranging from −∞ to zero (0^- if any pole or zero present at s=0).
- The number of infinite radius half circles will be equal to the number of poles or zeros at origin. The infinite radius half circle will start at the point where the mirror image of the polar plot ends. And this infinite radius half circle will end at the point where the polar plot starts.

After drawing the Nyquist plot, we can find the stability of the closed loop control system using the Nyquist stability criterion. If the critical point (-1+j0) lies outside the encirclement, then the closed loop control system is absolutely stable.

4.7.2 Stability Analysis using Nyquist Plots

From the Nyquist plots, we can identify whether the control system is stable, marginally stable or unstable based on the values of these parameters.

- Gain cross over frequency and phase cross over frequency
- Gain margin and phase margin

Phase Cross over Frequency

The frequency at which the Nyquist plot intersects the negative real axis (phase angle is 180^0) is known as the **phase cross over frequency**. It is denoted by ω_{pc}.

Gain Cross over Frequency

The frequency at which the Nyquist plot is having the magnitude of one is known as the **gain cross over frequency**. It is denoted by ω_{gc}.

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

- If the phase cross over frequency ω_{pc} is greater than the gain cross over frequency ω_{gc}, then the control system is **stable**.
- If the phase cross over frequency ω_{pc} is equal to the gain cross over frequency ω_{gc}, then the control system is **marginally stable**.
- If phase cross over frequency ω_{pc} is less than gain cross over frequency ω_{gc}, then the control system is **unstable**

Gain Margin

The gain margin GM is equal to the reciprocal of the magnitude of the Nyquist plot at the phase cross over frequency.

$$GM=1/Mpc$$

Where, Mpc is the magnitude in normal scale at the phase cross over frequency.

Phase Margin

The phase margin PMPM is equal to the sum of 180^0 and the phase angle at the gain cross over frequency.

$$PM=180^0+\phi_{gc}$$

Where, ϕ_{gc} is the phase angle at the gain cross over frequency.

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

- If the gain margin GM is greater than one and the phase margin PM is positive, then the control system is **stable**.
- If the gain margin GM is equal to one and the phase margin PM is zero degrees, then the control system is **marginally stable**.
- If the gain margin GM is less than one and / or the phase margin PM is negative, then the control system is **unstable**.

Module-III

Chapter 5

Control system design

5.1 ROOT-LOCUS APPROACH TO CONTROL-SYSTEMS DESIGN

In many practical situations because the plant may be fixed and not modifiable. Then we must adjust parameters other than those in the fixed plant.

In practice, the root-locus plot of a system may indicate that the desired performance cannot be achieved just by the adjustment of gain

The design problems, therefore, become those of improving system performance by insertion of a compensator.

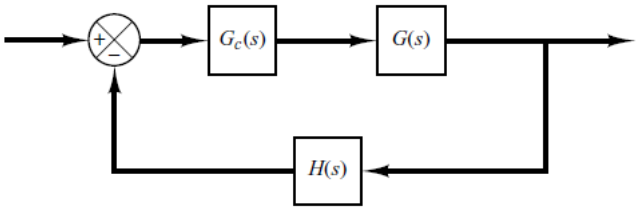
The design by the root-locus method is based on reshaping the root locus of the system by adding poles and zeros to the system's open-loop transfer function and forcing the root loci to pass through desired closed-loop poles in the s plane. The characteristic of the root-locus design is its being based on the assumption that the closed-loop system has a pair of dominant closed-loop poles. This means that the effects of zeros and additional poles do not affect the response characteristics very much.

in the design by the rootlocus method, the root loci of the system are reshaped through the use of a compensator so that a pair of dominant closed-loop poles can be placed at the desired location.

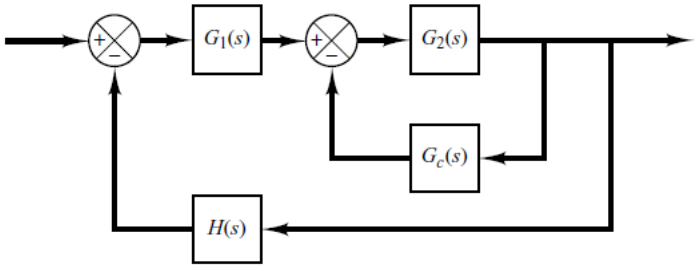
5.2 TYPES OF COMPENSATION

Series Compensation or Cascade Compensation

This is the most commonly used system where the controller is placed in series with the controlled process.

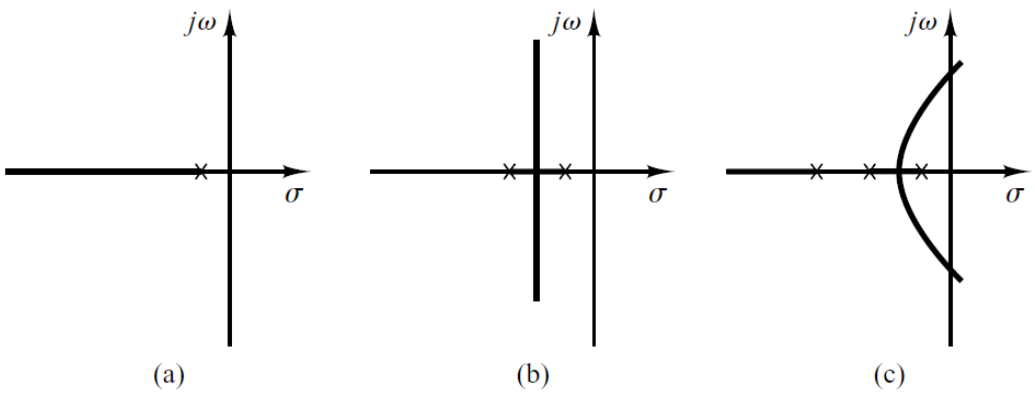


Feedback Compensation or Parallel Compensation



Effects of the Addition of Poles. The addition of a pole to the open-loop transfer function has the effect of pulling the root locus to the right, tending to lower the system's relative stability and to slow down the settling of the response. (Remember that the addition of integral control adds a pole at the origin, thus making the system less stable.)

Figure shows examples of root loci illustrating the effects of the addition of a pole to a single-pole system and the addition of two poles to a single-pole system.

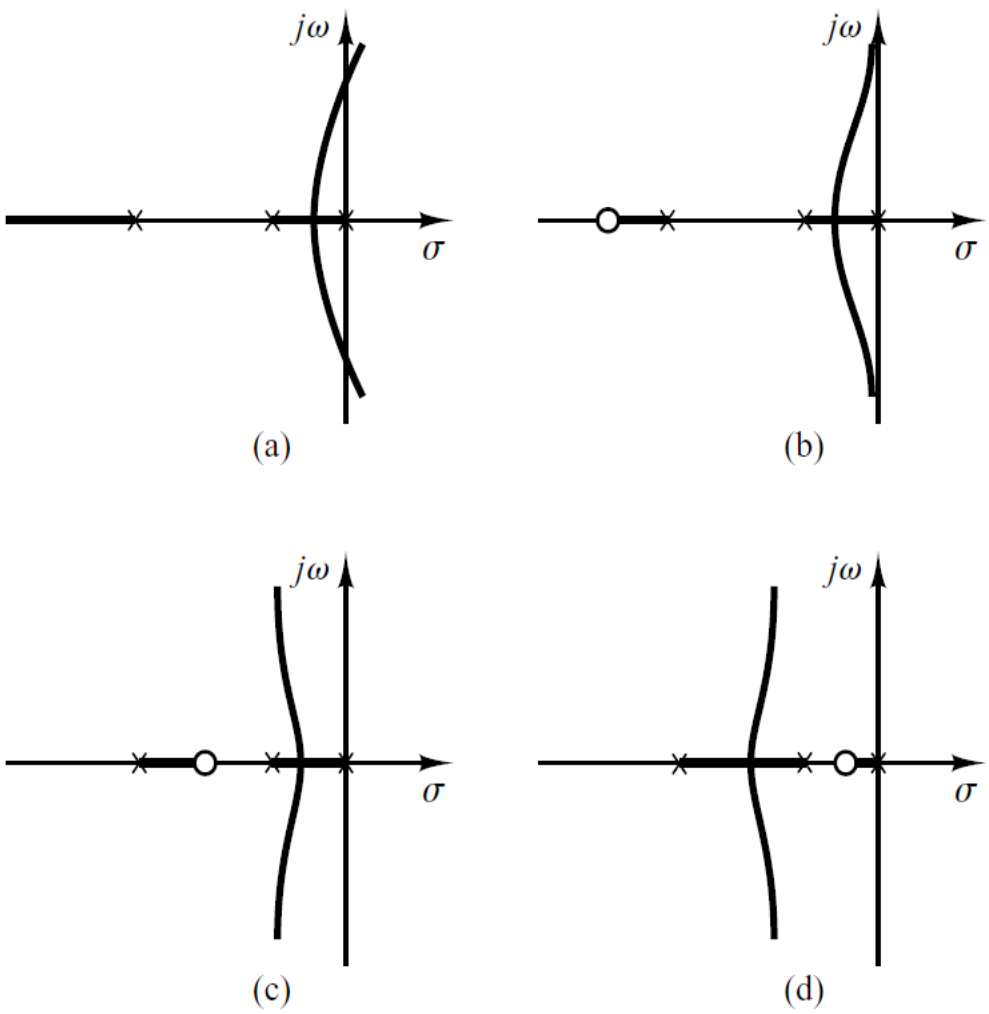


(a) Root-locus plot of a single-pole system;

(b) root-locus plot of a two-pole system;

(c) root-locus plot of a three-pole system.

Effects of the Addition of Zeros. The addition of a zero to the open-loop transfer function has the effect of pulling the root locus to the left, tending to make the system more stable and to speed up the settling of the response. (Physically, the addition of a zero in the feedforward transfer function means the addition of derivative control to the system. The effect of such control is to introduce a degree of anticipation into the system and speed up the transient response.)



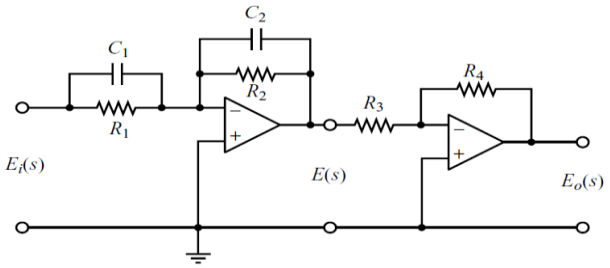
(a) Root-locus plot of a three-pole system; (

b), (c), and (d) root-locus plots showing effects of addition of a zero to the three-pole system.

5.3 LEAD COMPENSATION

The main problem then involves the judicious choice of the pole(s) and zero(s) of the compensator Gc(s) to have the dominant closed-loop poles at the desired location in the s plane so that the performance specifications will be met.

There are many ways to realize lead compensators and lag compensators, such as electronic networks using operational amplifiers, electrical *RC* networks, and mechanical spring-dashpot systems.



The transfer function for this circuit

$$\frac{E_o(s)}{E_i(s)} = \frac{R_2R_4}{R_1R_3}\frac{R_1C_1s+1}{R_2C_2s+1} = \frac{R_4C_1}{R_3C_2}\frac{s+\dfrac{1}{R_1C_1}}{s+\dfrac{1}{R_2C_2}}$$

$$= K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}}$$

where

$$T = R_1C_1, \qquad \alpha T = R_2C_2, \qquad K_c = \frac{R_4C_1}{R_3C_2}$$

The specifications are given in terms of time-domain quantities, such as the damping ratio and undamped natural frequency of the desired dominant closed-loop poles, maximum overshoot, rise time, and settling time.

Consider a design problem in which the original system either is unstable for all values of gain or is stable but has undesirable transient-response characteristics.

5.3.1 The procedures for designing a lead compensator

1. From the performance specifications, determine the desired location for the dominant closed-loop poles.
2. By drawing the root-locus plot of the uncompensated system (original system), ascertain whether or not the gain adjustment alone can yield the desired closed loop poles. If not, calculate the angle deficiency ϕ. This angle must be contributed by the lead compensator if the new root locus is to pass through the desired locations for the dominant closed-loop poles.
3. Assume the lead compensator $G_c(s)$ to be

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}}, \qquad (0<\alpha<1)$$

where a and T are determined from the angle deficiency. *K*c is determined from the requirement of the open-loop gain.

4. If static error constants are not specified, determine the location of the pole and zero of the lead compensator so that the lead compensator will contribute the necessary angle f. If no other requirements are imposed on the system, try to make the value of a as large as possible. A larger value of a generally results in a larger value of *K*v, which is desirable. Note that

$$K_v = \lim_{s\to 0} sG_c(s)G(s) = K_c\alpha \lim_{s\to 0} sG_c(s)$$

5. Determine the value of *K*c of the lead compensator from the magnitude condition
6. Once a compensator has been designed, check to see whether all performance specifications have been met. If the compensated system does not meet the performance specifications, then repeat the design procedure by adjusting the compensator pole and zero until all such specifications are met.

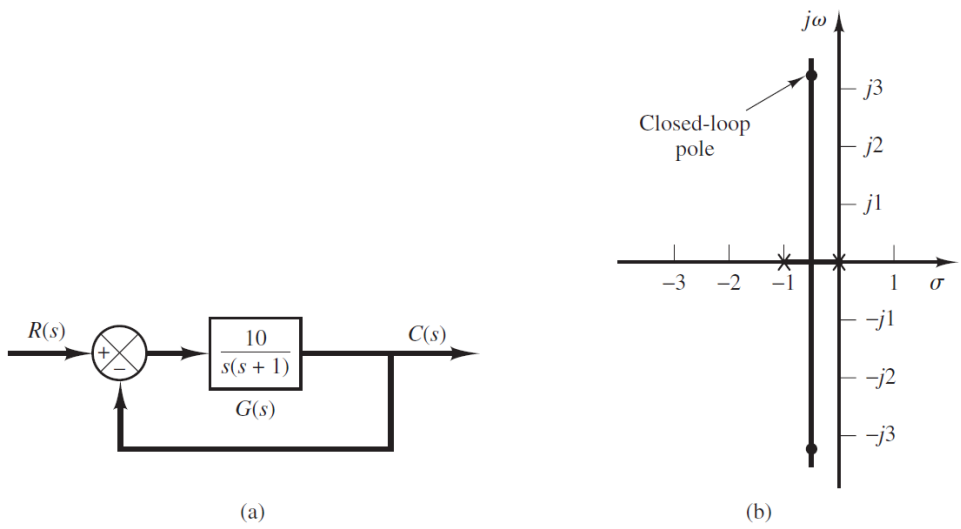
Example

Consider the position control system. The feedforward transfer function is

$$G(s) = \frac{10}{s(s+1)}$$

The closed-loop transfer function for the system is

$$\begin{aligned}\frac{C(s)}{R(s)} &= \frac{10}{s^2+s+10}\\ &= \frac{10}{(s+0.5+j3.1225)(s+0.5-j3.1225)}\end{aligned}$$



(a) Control system;

(b) root-locus plot

The closed-loop poles are located at

$$s = -0.5 \pm j3.1225$$

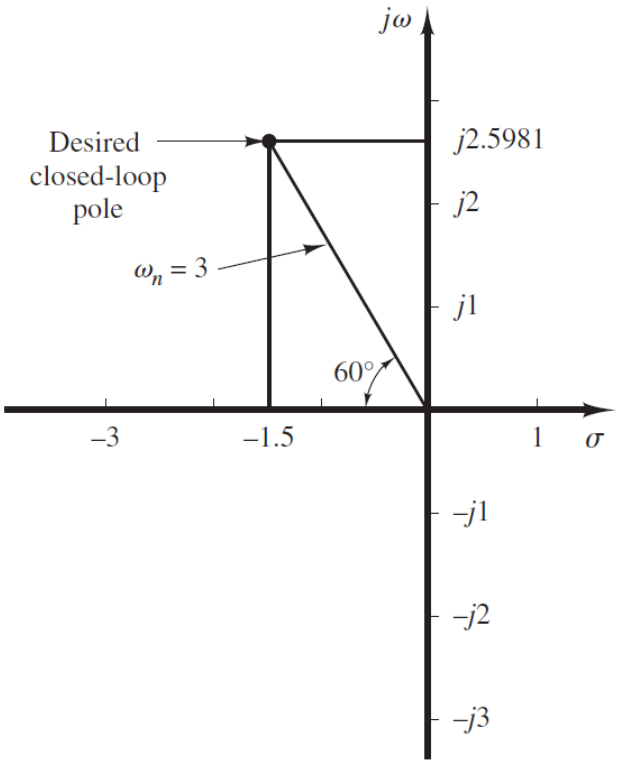
The damping ratio of the closed-loop poles is $\zeta = (1/2)/\sqrt{10} = 0.1581$. .The undamped natural frequency of the closed-loop poles is $\omega_n = \sqrt{10} = 3.1623$ rad/sec. Because the damping ratio is small, this system will have a large overshoot in the step response and is not desirable.

It is desired to design a lead compensator Gc(s) so that the dominant closed-loop poles have the damping ratio ξ=0.5 and the undamped natural frequency $\omega_n = 3\ rad/sec$.

The desired location of the dominant closed-loop poles can be determined from

$$\begin{aligned} s^2 + 2\zeta\omega_n s + \omega_n^2 &= s^2 + 3s + 9 \\ &= (s + 1.5 + j2.5981)(s + 1.5 - j2.5981) \end{aligned}$$

$$s = -1.5 \pm j2.5981$$



First, find the sum of the angles at the desired location of one of the dominant closed-loop poles with the open-loop poles and zeros of the original system, and determine the necessary angle ϕ to be added so that the total sum of the angles is equal to ±180(2K+1). The lead compensator must contribute this angle ϕ. (If the angle ϕ is quite large, then two or more lead networks may be needed rather than a single one.)

Assume that the lead compensator Gc(s) has the transfer function

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}}, \qquad (0<\alpha<1)$$

The angle from the pole at the origin to the desired dominant closed-loop pole at s = –1.5+j2.5981 is 120°.The angle from the pole at s=–1 to the desired closed-loop pole is 100.894°. Hence, the angle deficiency is

Angle deficiency=180° – 120° – 100.894°=–40.894°

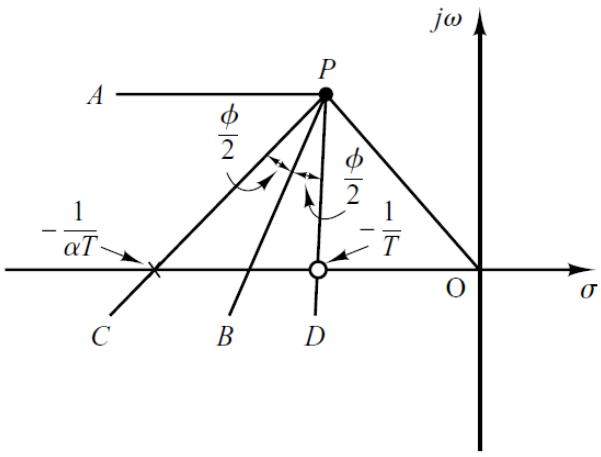
Deficit angle 40.894° must be contributed by a lead compensator

Method 1. a procedure to obtain a largest possible value for a. (Note that a larger value of a will produce a larger value of *K*v. In most cases, the larger the *K*v is, the better the system performance.)

First, draw a horizontal line passing through point *P*, the desired location for one of the dominant closed-loop poles. This is shown as line PA. Draw also a line connecting point P and the origin. Bisect the angle between the lines PA and PO. Draw two lines PC and PD that make angles ± ϕ/2 with the bisector PB. The intersections of PC and PD with the negative real axis give the necessary locations for the pole and zero of the lead networks. The compensator thus designed will make point P a point on the root locus of the compensated system. The open-loop gain is determined by use of the magnitude condition.

In the present system, the angle of G(s) at the desired closed-loop pole is

$$\angle \frac{10}{s(s+1)}\bigg|_{s=-1.5+j2.5981} = -220.894^\circ$$



Thus, if we need to force the root locus to go through the desired closed-loop pole, the lead compensator must contribute ϕ=40.894° at this point. By following the foregoing design procedure, we can determine the zero and pole of the lead compensator.

if we bisect angle APO and take 40.894°/2 each side, then the locations of the zero and pole are found as follows:

$$\text{zero at } s = -1.9432$$
$$\text{pole at } s = -4.6458$$

Thus, Gc(s) can be given as

$$G_c(s) = K_c \frac{s + \frac{1}{T}}{s + \frac{1}{\alpha T}} = K_c \frac{s + 1.9432}{s + 4.6458}$$

(For this compensator the value of a is a = 1.9432/4.6458 = 0.418.)

The value of Kc can be determined by use of the magnitude condition

$$\left| K_c \frac{s + 1.9432}{s + 4.6458} \frac{10}{s(s + 1)} \right|_{s=-1.5+j2.5981} = 1$$

$$K_c = \left| \frac{(s + 4.6458)s(s + 1)}{10(s + 1.9432)} \right|_{s=-1.5+j2.5981} = 1.2287$$

Hence, the lead compensator Gc(s) just designed is given by

$$G_c(s) = 1.2287 \frac{s + 1.9432}{s + 4.6458}$$

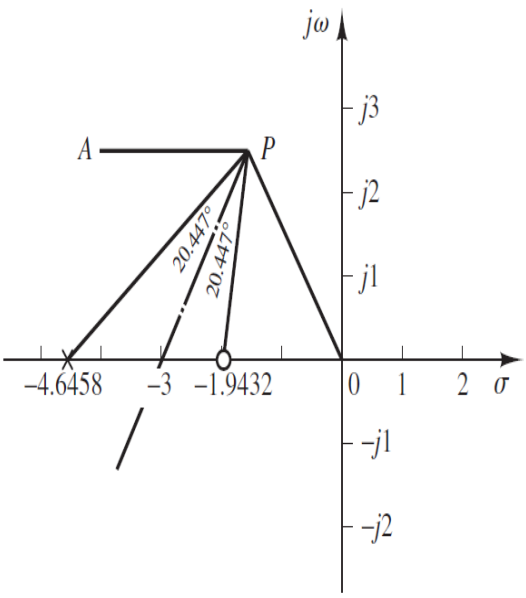
Then, the open-loop transfer function of the designed system becomes

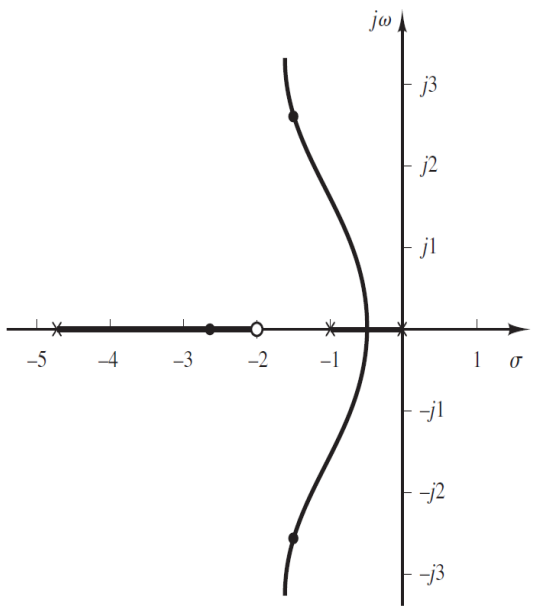
$$G_c(s)G(s) = 1.2287 \left(\frac{s + 1.9432}{s + 4.6458} \right) \frac{10}{s(s + 1)}$$

and the closed-loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{12.287(s + 1.9432)}{s(s + 1)(s + 4.6458) + 12.287(s + 1.9432)}$$

$$= \frac{12.287s + 23.876}{s^3 + 5.646s^2 + 16.933s + 23.876}$$





5.4 Design Procedures for Lag Compensation

1. Draw the root-locus plot for the uncompensated system whose open-loop transfer function is G(s). Based on the transient-response specifications, locate the dominant closed-loop poles on the root locus.
2. Assume the transfer function of the lag compensator to be given by

$$G_c(s) = \hat{K}_c\beta\frac{Ts + 1}{\beta Ts + 1} = \hat{K}_c\frac{s + \frac{1}{T}}{s + \frac{1}{\beta T}}$$

   Then the open-loop transfer function of the compensated system becomes
3. Evaluate the particular static error constant specified in the problem.
4. Determine the amount of increase in the static error constant necessary to satisfy the specifications.
5. Determine the pole and zero of the lag compensator that produce the necessary increase in the particular static error constant without appreciably altering the original root loci. (Note that the ratio of the value of gain required in the specifications and the gain found in the uncompensated system is the required ratio between the distance of the zero from the origin and that of the pole from the origin.)
6. Draw a new root-locus plot for the compensated system. Locate the desired dominant closed-loop poles on the root locus. (If the angle contribution of the lag network is very small that is a few degrees then the original and new root loci are almost identical. Otherwise, there will be a slight discrepancy between them. Then locate, on the new root locus, the desired dominant closed-loop poles based on the transient-response specifications.)
7. Adjust gain of the compensator from the magnitude condition so that the dominant closed-loop poles lie at the desired location which will be approximately 1.

Example

The feedforward transfer function is

$$G(s) = \frac{1.06}{s(s + 1)(s + 2)}$$

The closed-loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{1.06}{s(s+1)(s+2)+1.06}$$

$$= \frac{1.06}{(s+0.3307-j0.5864)(s+0.3307+j0.5864)(s+2.3386)}$$

The dominant closed-loop poles are

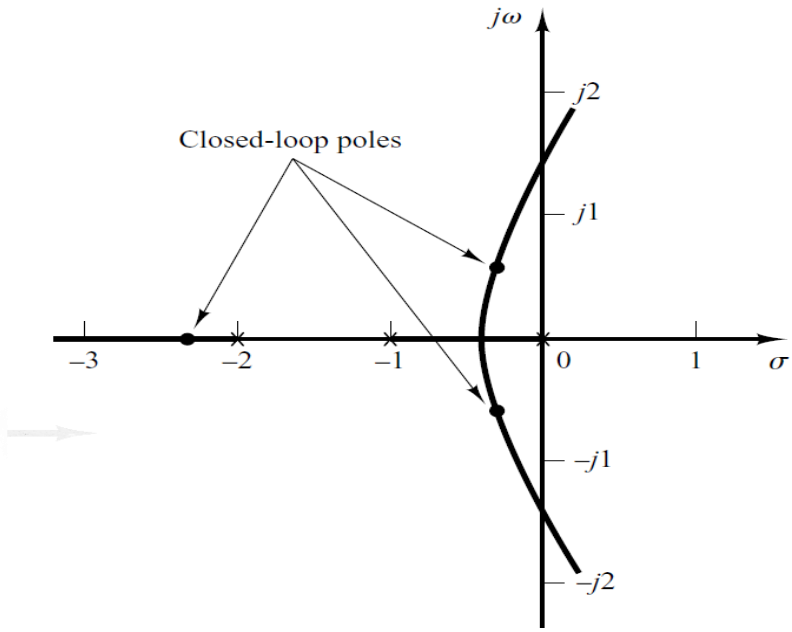
$$s = -0.3307 \pm j0.5864$$

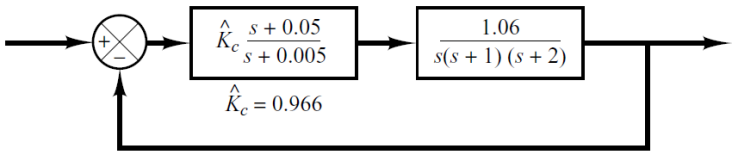
The damping ratio of the dominant closed-loop poles is ξ=0.491. The undamped natural frequency of the dominant closed-loop poles is 0.673 rad/sec. The static velocity error constant is 0.53 sec^{-1}.

It is desired to increase the static velocity error constant *K*v to about 5 sec^{-1} without appreciably changing the location of the dominant closed-loop poles.

To meet this specification, let us insert a lag compensator in cascade with the given feedforward transfer function. To increase the static velocity error constant by a factor of about 10, let us choose β=10 and place the zero and pole of the lag compensator at s=–0.05 and s=–0.005, respectively. The transfer function of the lag compensator becomes

$$G_c(s) = \hat{K}_c \frac{s+0.05}{s+0.005}$$





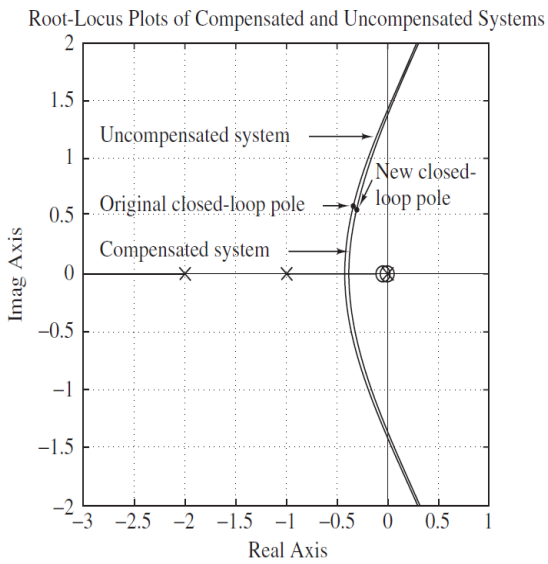
The angle contribution of this lag network near a dominant closed-loop pole is about 4°. Because this angle contribution is not very small, there is a small change in the new root locus near the desired dominant closed-loop poles.

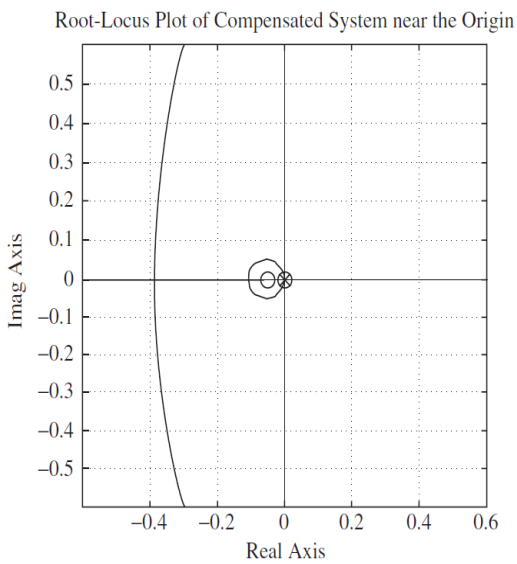
The open-loop transfer function of the compensated system then becomes

$$G_c(s)G(s) = \hat{K}_c \frac{s + 0.05}{s + 0.005} \frac{1.06}{s(s + 1)(s + 2)}$$

$$= \frac{K(s + 0.05)}{s(s + 0.005)(s + 1)(s + 2)}$$

$$K = 1.06\hat{K}_c$$





If the damping ratio of the new dominant closed-loop poles is kept the same, then these poles are obtained from the new root-locus plot as follows:

$s_1 = -0.31 + j0.55$, $s_2 = -0.31 - j0.55$

The open-loop gain K is determined from the magnitude condition as follows:

$$K = \left| \frac{s(s + 0.005)(s + 1)(s + 2)}{s + 0.05} \right|_{s=-0.31+j0.55}$$

$$= 1.0235$$

Then the lag compensator gain $\hat{K}_c$ is determined as

$$\hat{K}_c = \frac{K}{1.06} = \frac{1.0235}{1.06} = 0.9656$$

Thus the transfer function of the lag compensator designed is

$$G_c(s) = 0.9656\frac{s + 0.05}{s + 0.005} = 9.656\frac{20s + 1}{200s + 1}$$

Then the compensated system has the following open-loop transfer function:

$$G_1(s) = \frac{1.0235(s + 0.05)}{s(s + 0.005)(s + 1)(s + 2)}$$

$$= \frac{5.12(20s + 1)}{s(200s + 1)(s + 1)(0.5s + 1)}$$

The static velocity error constant K_v is

$$K_v = \lim_{s \to 0} sG_1(s) = 5.12 \text{ sec}^{-1}$$

5.5 CONTROL SYSTEMS DESIGN BY FREQUENCY RESPONSE APPROACH

The transient-response performance is specified in terms of the phase margin, gain margin, resonant peak magnitude (they give a rough estimate of the system damping); the gain crossover frequency, resonant frequency, bandwidth (they give a rough estimate of the speed of transient response); and static error constants (they give the steady-state accuracy).

There are basically two approaches in the frequency-domain design.

One is the polar plot approach and the other is the Bode diagram approach.

When a compensator is added, the polar plot does not retain the original shape, and, therefore, we need to draw a new polar plot, which will take time and is thus inconvenient. On the other hand, a Bode diagram of the compensator can be simply added to the original Bode diagram, and thus plotting the complete Bode diagram is a simple matter.

- Lead compensation essentially yields an appreciable improvement in transient response and a small change in steady-state accuracy. It may accentuate high-frequency noise effects.

- Lag compensation, on the other hand, yields an appreciable improvement in steady-state accuracy at the expense of increasing the transient-response time. Lag compensation will suppress the effects of high-frequency noise signals. Lag–lead compensation combines the characteristics of both lead compensation and lag compensation.

5.6 Lead Compensation Techniques

The primary function of the lead compensator is to reshape the frequency-response curve to provide sufficient phase-lead angle to offset the excessive phase lag associated with the components of the fixed system.

Assume the following lead compensator:

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}} \qquad (0<\alpha<1)$$

Define

$$K_c\alpha = K$$

Then

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

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = K\frac{Ts+1}{\alpha Ts+1}G(s) = \frac{Ts+1}{\alpha Ts+1}KG(s) = \frac{Ts+1}{\alpha Ts+1}G_1(s)$$

where

$$G_1(s) = KG(s)$$

1. Determine gain K to satisfy the requirement on the given static error constant.
2. Using the gain K thus determined, draw a Bode diagram of $G_1(j\omega)$, the gain adjusted but uncompensated system. Evaluate the phase margin.
3. Determine the necessary phase-lead angle to be added to the system. Add an additional 5° to 12° to the phase-lead angle required, because the addition of the lead compensator shifts the gain crossover frequency to the right and decreases the phase margin.
4. Determine the attenuation factor. Determine the frequency where the magnitude of the uncompensated system $G_1(j\omega)$ is equal to select this frequency as the new gain crossover frequency. This frequency corresponds to and the maximum phase shift ϕ_m occurs at this frequency.

5. Determine the corner frequencies of the lead compensator as follows:

$$\text{Zero of lead compensator:} \qquad \omega = \frac{1}{T}$$

$$\text{Pole of lead compensator:} \qquad \omega = \frac{1}{\alpha T}$$

6. Using the value of K determined in step 1 and that of a determined in step 4, calculate constant *K*c from

$$K_c = \frac{K}{\alpha}$$

7. Check the gain margin to be sure it is satisfactory. If not, repeat the design process by modifying the pole–zero location of the compensator until a satisfactory result is obtained.

Example

Consider the open-loop transfer function is

$$G(s) = \frac{4}{s(s+2)}$$

It is desired to design a compensator for the system so that the static velocity error constant *K*v is 20 sec^{-1}, the phase margin is at least 50°, and the gain margin is at least 10 dB.

We shall use a lead compensator of the form

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\frac{1}{T}}{s+\frac{1}{\alpha T}}$$

The compensated system will have the open-loop transfer function *G*c(s)G(s).

Define

$$G_1(s) = KG(s) = \frac{4K}{s(s+2)}$$

where $K = K_c\alpha$.

The first step in the design is to adjust the gain K to meet the steady-state performance specification or to provide the required static velocity error constant. Since this constant is given as 20 sec^{-1}, we obtain

$$K_v = \lim_{s\to 0} sG_c(s)G(s) = \lim_{s\to 0} s\frac{Ts+1}{\alpha Ts+1}G_1(s) = \lim_{s\to 0}\frac{s4K}{s(s+2)} = 2K = 20$$

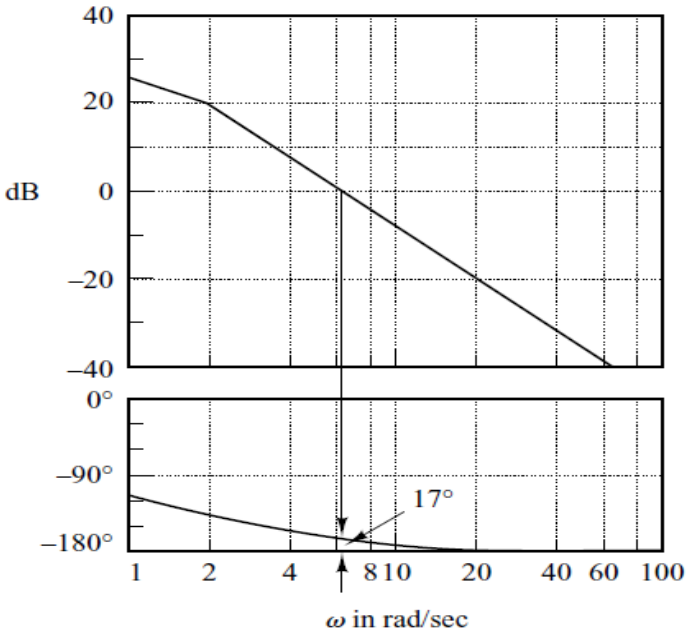
Or

K = 10

With K=10, the compensated system will satisfy the steady-state requirement.

We shall next plot the Bode diagram of

$$G_1(j\omega) = \frac{40}{j\omega(j\omega+2)} = \frac{20}{j\omega(0.5j\omega+1)}$$



From this plot, the phase and gain margins of the system are found to be 17° and +∞ dB, respectively. (A phase margin of 17° implies that the system is quite oscillatory. Thus, satisfying the specification on the steady state yields a poor transient-response performance.) The specification calls for a phase margin of at least 50°.We thus find the additional phase lead necessary to satisfy the relative stability requirement is 33°.To achieve a phase margin of 50° without decreasing the value of K, the lead compensator must contribute the required phase angle.

The gain crossover frequency will be shifted to the right. We must offset the increased phase lag of $G_1(j\omega)$ due to this increase in the gain crossover frequency. Considering the shift of the gain crossover frequency, we may assume that ϕ_m, the maximum phase lead required, is

approximately 38°. (This means that 5° has been added to compensate for the shift in the gain crossover frequency.)

since

$$\sin\phi_m = \frac{1-\alpha}{1+\alpha}$$

ϕ_m=38° corresponds to a=0.24.

determine the corner frequencies ω=1/T and ω=1/(αT) of the lead compensator

first note that the maximum phase-lead angle ϕ_m occurs at the geometric mean of the two corner frequencies, or v = 1_A1aTB

The amount of the modification in the magnitude curve at due to the inclusion of the term (Ts+1)/(αTs+1) is

$$\left|\frac{1+j\omega T}{1+j\omega\alpha T}\right|_{\omega=1/(\sqrt{\alpha}T)} = \left|\frac{1+j\dfrac{1}{\sqrt{\alpha}}}{1+j\alpha\dfrac{1}{\sqrt{\alpha}}}\right| = \frac{1}{\sqrt{\alpha}}$$

$$\frac{1}{\sqrt{\alpha}} = \frac{1}{\sqrt{0.24}} = \frac{1}{0.49} = 6.2\text{ dB}$$

and $|G_1(j\omega)|$=–6.2 dB corresponds to ω=9rad/sec. We shall select this frequency to be the new gain crossover frequency ω_c. Noting that this frequency corresponds to or

we obtain

$$\frac{1}{T} = \sqrt{\alpha}\omega_c = 4.41$$

$$\frac{1}{\alpha T} = \frac{\omega_c}{\sqrt{\alpha}} = 18.4$$

The lead compensator thus determined is

$$G_c(s) = K_c\frac{s+4.41}{s+18.4} = K_c\alpha\frac{0.227s+1}{0.054s+1}$$

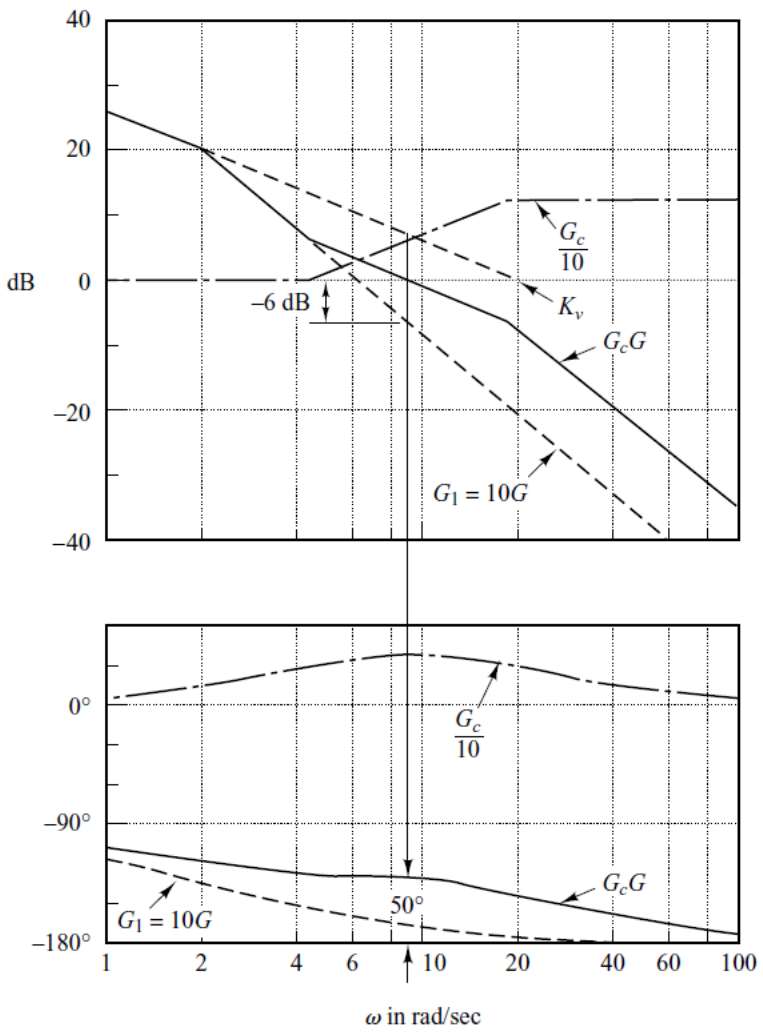
where the value of *K*c is determined as

$$K_c = \frac{K}{\alpha} = \frac{10}{0.24} = 41.7$$

Thus, the transfer function of the compensator becomes

$$G_c(s) = 41.7\frac{s + 4.41}{s + 18.4} = 10\frac{0.227s + 1}{0.054s + 1}$$

$$\frac{G_c(s)}{K}G_1(s) = \frac{G_c(s)}{10}10G(s) = G_c(s)G(s)$$



5.7 Lag Compensation Techniques

The primary function of a lag compensator is to provide attenuation in the high frequency range to give a system sufficient phase margin. The phase-lag characteristic is of no consequence in lag compensation.

The procedure for designing lag compensators

Assume the following lag compensator:

$$G_c(s) = K_c\beta\frac{Ts+1}{\beta Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\beta T}} \qquad (\beta > 1)$$

$$K_c\beta = K$$

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

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = K\frac{Ts+1}{\beta Ts+1}G(s) = \frac{Ts+1}{\beta Ts+1}KG(s) = \frac{Ts+1}{\beta Ts+1}G_1(s)$$

$G_1(s) = KG(s)$

1. Determine gain K to satisfy the requirement on the given static velocity error constant.
2. If the gain-adjusted but uncompensated system $G_1(j\omega) = KG(j\omega)$ does not satisfy the specifications on the phase and gain margins, then find the frequency point where the phase angle of the open-loop transfer function is equal to –180° plus the required phase margin. The required phase margin is the specified phase margin plus 5° to 12°. (The addition of 5° to 12° compensates for the phase lag of the lag compensator.) Choose this frequency as the new gain crossover frequency.
3. To prevent detrimental effects of phase lag due to the lag compensator, the pole and zero of the lag compensators must be located substantially lower than the new gain crossover frequency. Therefore, choose the corner frequency $\omega=1/T$ (corresponding to the zero of the lag compensator) 1 octave to 1 decade below the new gain crossover frequency. (If the time constants of the lag compensator do not become too large, the corner frequency $\omega=1/T$ may be chosen 1 decade below the new gain crossover frequency.)
4. Determine the attenuation necessary to bring the magnitude curve down to 0 dB at the new gain crossover frequency. Noting that this attenuation is $-20\log\beta$, determine the

value of β. Then the other corner frequency (corresponding to the pole of the lag compensator) is determined from ω=1/(βT).

5. Using the value of K determined in step 1 and that of b determined in step 4, calculate constant *K*c from

$$K_c = \frac{K}{\beta}$$

Example

Consider the open-loop transfer function is given by

$$G(s) = \frac{1}{s(s+1)(0.5s+1)}$$

It is desired to compensate the system so that the static velocity error constant *K*v is 5 sec^{-1}, the phase margin is at least 40°, and the gain margin is at least 10 dB.

We shall use a lag compensator of the form

$$G_c(s) = K_c\beta\frac{Ts+1}{\beta Ts+1} = K_c\frac{s+\frac{1}{T}}{s+\frac{1}{\beta T}} \qquad (\beta > 1)$$

$$K_c\beta = K$$

$$G_1(s) = KG(s) = \frac{K}{s(s+1)(0.5s+1)}$$

The first step in the design is to adjust the gain K to meet the required static velocity error constant.

Thus,

$$K_v = \lim_{s\to 0} sG_c(s)G(s) = \lim_{s\to 0} s\frac{Ts+1}{\beta Ts+1}G_1(s) = \lim_{s\to 0} sG_1(s)$$

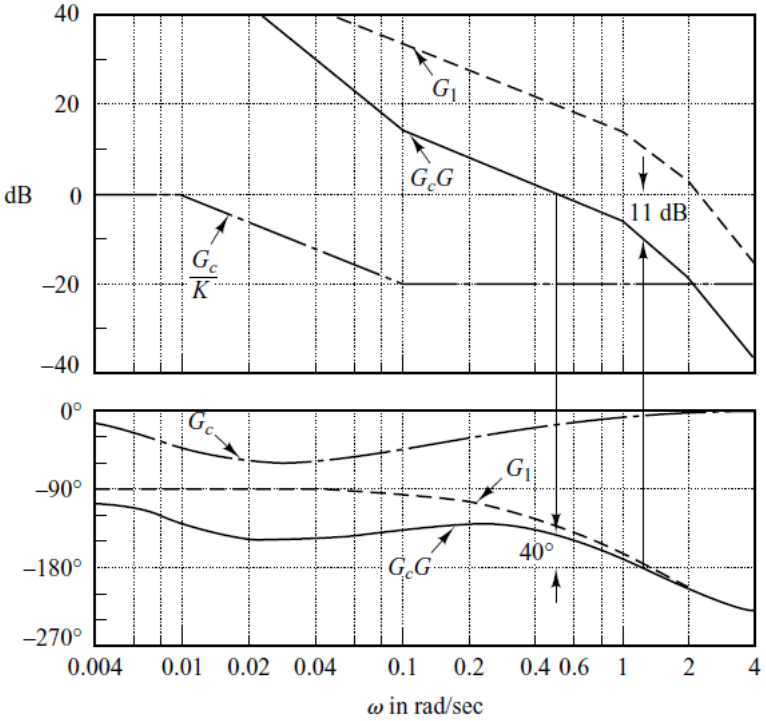
$$= \lim_{s\to 0}\frac{sK}{s(s+1)(0.5s+1)} = K = 5$$

$$K = 5$$

With K=5, the compensated system satisfies the steady-state performance requirement.

Plot the Bode diagram of

$$G_1(j\omega) = \frac{5}{j\omega(j\omega + 1)(0.5j\omega + 1)}$$



From this plot, the phase margin is found to be –20°, which means that the gain-adjusted but uncompensated system is unstable.

The addition of a lag compensator modifies the phase curve of the Bode diagram, we must allow 5° to 12° to the specified phase margin to compensate for the modification of the phase curve. Since the frequency corresponding to a phase margin of 40° is 0.7 rad/sec, the new gain crossover frequency (of the compensated system) must be chosen near this value.

To avoid overly large time constants for the lag compensator, we shall choose the corner frequency ω=1/T(which corresponds to the zero of the lag compensator) to be 0.1 rad/sec. Since this corner frequency is not too far below the new gain crossover frequency, the modification in the phase curve may not be small.

Hence, we add about 12° to the given phase margin as an allowance to account for the lag angle introduced by the lag compensator.

The required phase margin is now 52°.The phase angle of the uncompensated open-loop transfer function is –128° at about ω=0.5 rad/sec.

So we choose the new gain crossover frequency to be 0.5 rad/sec.

To bring the magnitude curve down to 0 dB at this new gain crossover frequency, the lag compensator must give the necessary attenuation, which in this case is –20 dB.

Hence,

$$20 \log \frac{1}{\beta} = -20$$

Or

$$\beta = 10$$

The other corner frequency $\omega=1/(\beta T)$, which corresponds to the pole of the lag compensator, is then determined as

$$\frac{1}{\beta T} = 0.01 \text{ rad/sec}$$

Thus, the transfer function of the lag compensator is

$$G_c(s) = K_c(10)\frac{10s + 1}{100s + 1} = K_c \frac{s + \frac{1}{10}}{s + \frac{1}{100}}$$

Since the gain K was determined to be 5 and b was determined to be 10, we have

$$K_c = \frac{K}{\beta} = \frac{5}{10} = 0.5$$

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = \frac{5(10s + 1)}{s(100s + 1)(s + 1)(0.5s + 1)}$$

The phase margin of the compensated system is about 40°, which is the required value. The gain margin is about 11 dB, which is quite acceptable. The static velocity error constant is 5 sec^{-1}, as required.

The compensated system, therefore, satisfies the requirements on both the steady state and the relative stability.

Chapter-6

PID Controllers

Proportional controllers: Pure gain or attenuation

Integral controllers: Integrate error

Derivative controllers: Differentiate error

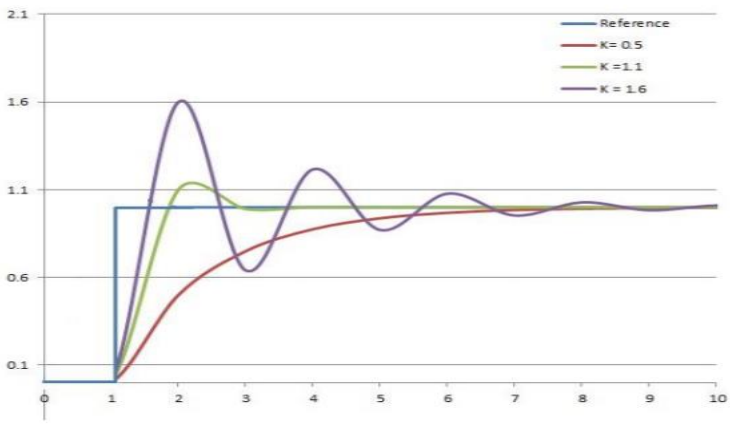
These values can be interpreted in terms of time: P depends on the present error, I on the accumulation of past errors, and D is a prediction of future errors, based on current rate of change

6.1 Proportional Term

The proportional term produces an output value that is proportional to the current error value. The proportional response can be adjusted by multiplying the error by a constant K_p, called the proportional gain constant. The proportional term is given by:

$$P_{out} = K_p e(t)$$

A high proportional gain results in a large change in the output for a given change in the error. If the proportional gain is too high, the system can become unstable. In contrast, a small gain results in a small output response to a large input error, and a less responsive or less sensitive controller. If the proportional gain is too low, the control action may be too small when responding to system disturbances. Tuning theory and industrial practice indicate that the proportional term should contribute the bulk of the output change.

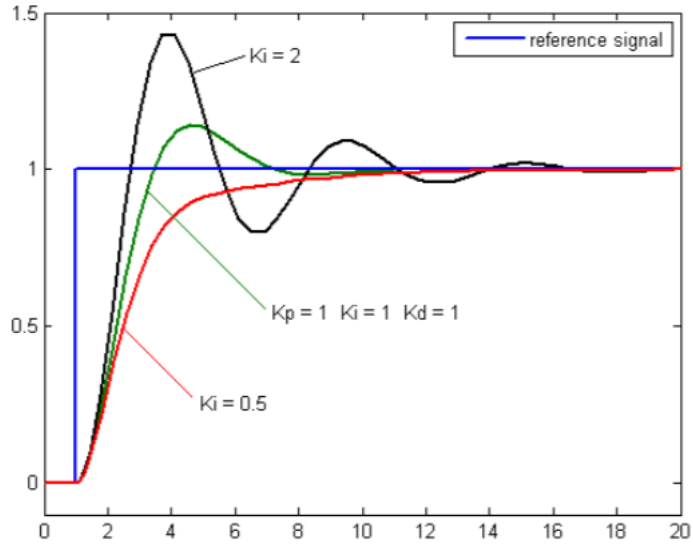


6.2 Integral Term

The contribution from the integral term is proportional to both the magnitude of the error and the duration of the error. The integral in a PID controller is the sum of the instantaneous error over time and gives the accumulated offset that should have been corrected previously. The accumulated error is then multiplied by the integral gain $\boldsymbol{K_i}$ and added to the controller output.

$$I_{out} = Ki \int_0^t e(\tau) d\tau$$

The integral term accelerates the movement of the process towards set-point and eliminates the residual steady-state error that occurs with a pure proportional controller. However, since the integral term responds to accumulated errors from the past, it can cause the present value to overshoot the set-point value.



6.3 Derivative Term

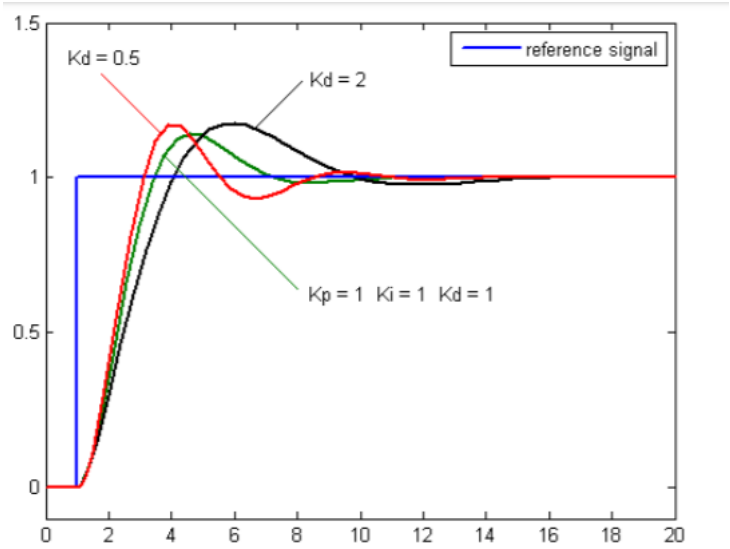
The derivative of the process error is calculated by determining the slope of the error over time and multiplying this rate of change by the derivative gain K_d. The magnitude of the contribution of the derivative term to the overall control action is termed the derivative gain, K_d.

The derivative term is given by

$$D_{out} = K_d \frac{d}{dt} e(t)$$

Derivative action predicts system behaviour and thus improves settling time and stability of the system. An ideal derivative is not causal, so that implementations of PID controllers include an additional low pass filtering for the derivative term, to limit the high frequency gain and noise. Derivative action is seldom used in practice though - by one estimate in only 20% of

deployed controllers- because of its variable impact on system stability in real-world applications.



Parameter	Rise Time	Overshoot	Settling Time	Steady-State Error	Stability
K_p	Decrease	Increase	Small Change	Decrease	Degrade
K_i	Decrease	Increase	Increase	Eliminate	Degrade
K_d	Minor Change	Decrease	Decrease	No Effect	Improve if K_d small

Module-IV
Chapter 7
State Space Analysis

7.1 Concepts of state variables

A modern complex system may have many inputs and many outputs, and these may be interrelated in a complicated manner. To analyze such a system, it is essential to reduce the complexity of the mathematical expressions, as well as to resort to computers for most of the tedious computations necessary in the analysis. The state-space approach to system analysis is best suited from this viewpoint.

1. The conventional control theory is completely based on the frequency domain approach while the modern control system theory is based on time domain approach.
2. In the conventional theory of control system we have linear and time invariant single input single output (SISO) systems only but with the help of theory of modern control system we can easily do the analysis of even nonlinear and time variant multiple inputs multiple outputs (MIMO) systems also.
3. In the modern theory of control system, the stability analysis and time response analysis can be done by both graphical and analytically method very easily.

Now **state space analysis of control system** is based on the modern theory which is applicable to all types of systems like single input single output systems, multiple inputs and multiple outputs systems, linear and nonlinear systems, time varying and time invariant systems.

The **state space model** of Linear Time-Invariant (LTI) system can be represented as,

$$\dot{X}=AX+BU$$

$$Y=CX+DU$$

The first and the second equations are known as **state equation** and **output equation** respectively.

Where,

- X and $\dot{X}$ are the state vector and the differential state vector respectively.
- U and Y are input vector and output vector respectively.
- A is the system matrix.
- B and C are the input and the output matrices.
- D is the feed-forward matrix.

7.2 Basic Concepts of State Space Model

State

It is a group of variables, which summarizes the history of the system in order to predict the future values (outputs).

It refers to smallest set of variables whose knowledge at $t = t_0$ together with the knowledge of input for $t \geq t_0$ gives the complete knowledge of the behaviour of the system at any time $t \geq t_0$.

State Variable

The number of the state variables required is equal to the number of the storage elements present in the system.

It refers to the smallest set of variables which help us to determine the state of the dynamic system. State variables are defined by $x_1(t)$, $x_2(t)$……..$x_n(t)$.

Examples − current flowing through inductor, voltage across capacitor

State Vector

It is a vector, which contains the state variables as elements.

Suppose there is a requirement of n state variables in order to describe the complete behaviour of the given system, then these n state variables are considered to be n components of a vector x(t). Such a vector is known as state vector.

State Space : It refers to the n dimensional space which has x_1 axis, x_2 axis ……x_n axis.

7.3 State-space representations of transfer-function systems

state-space representations in the controllable, observable, diagonal or Jordan canonical forms.

State-Space Representations in Canonical Forms

Consider a system defined by

$$\overset{(n)}{y} + a_1 \overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y = b_0 \overset{(n)}{u} + b_1 \overset{(n-1)}{u} + \cdots + b_{n-1}\dot{u} + b_n u$$

where u is the input and y is the output. This equation can also be written as

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n}$$

Controllable Canonical Form.

The following state-space representation is called a controllable canonical form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix} u$$

$$y = \left[b_n - a_n b_0 \;\vdots\; b_{n-1} - a_{n-1} b_0 \;\vdots\; \cdots \;\vdots\; b_1 - a_1 b_0 \right] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + b_0 u$$

Observable Canonical Form

The following state-space representation is called an observable canonical form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} b_n - a_n b_0 \\ b_{n-1} - a_{n-1} b_0 \\ \cdot \\ \cdot \\ \cdot \\ b_1 - a_1 b_0 \end{bmatrix} u$$

$$y = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} + b_0 u$$

Diagonal Canonical Form

Lets consider where the denominator polynomial involves only distinct roots.

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1} s + b_n}{(s + p_1)(s + p_2)\cdots(s + p_n)}$$

$$= b_0 + \frac{c_1}{s + p_1} + \frac{c_2}{s + p_2} + \cdots + \frac{c_n}{s + p_n}$$

The diagonal canonical form of the state-space representation of this system is given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} -p_1 & & & & 0 \\ & -p_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & -p_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} u$$

$$y = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + b_0 u$$

Jordan Canonical Form

Lets consider a TF where the denominator polynomial involves multiple roots.

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1} s + b_n}{(s + p_1)^3 (s + p_4)(s + p_5) \cdots (s + p_n)}$$

The partial-fraction expansion of this last equation becomes

$$\frac{Y(s)}{U(s)} = b_0 + \frac{c_1}{(s + p_1)^3} + \frac{c_2}{(s + p_1)^2} + \frac{c_3}{s + p_1} + \frac{c_4}{s + p_4} + \cdots + \frac{c_n}{s + p_n}$$

A state-space representation of this system in the Jordan canonical form is given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \left[\begin{array}{ccc|ccc} -p_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -p_1 & 1 & \vdots & & \vdots \\ 0 & 0 & -p_1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & -p_4 & & 0 \\ \cdot & & \cdot & & \cdot & \\ \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \\ 0 & \cdots & 0 & 0 & & -p_n \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} u$$

$$y = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + b_0 u$$

7.4 Eigenvalues of an *n*×*n* Matrix A.

The eigenvalues of an n×n matrix **A** are the roots of the characteristic equation

$$|\lambda \mathbf{I} - \mathbf{A}| = 0$$

The eigenvalues are also called the characteristic roots.

Consider, for example, the following matrix **A**:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}$$

The characteristic equation is

$$|\lambda \mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ 6 & 11 & \lambda + 6 \end{vmatrix}$$

$$= \lambda^3 + 6\lambda^2 + 11\lambda + 6$$

$$= (\lambda + 1)(\lambda + 2)(\lambda + 3) = 0$$

The eigenvalues of **A** are the roots of the characteristic equation, or –1, –2, and –3.

7.5 Diagonalization of *n*×*n* Matrix.

Note that if an n×n matrix **A** with distinct eigenvalues is given by

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$

the transformation **x**=**Pz**, where

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix}$$

$$\lambda_1, \lambda_2, \ldots, \lambda_n = n \text{ distinct eigenvalues of } \mathbf{A}$$

will transform $\mathbf{P}^{-1}\mathbf{AP}$ into the diagonal matrix, or

$$\mathbf{P}^{-1}\mathbf{AP} = \begin{bmatrix} \lambda_1 & & & & & 0 \\ & \lambda_2 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ 0 & & & & & \lambda_n \end{bmatrix}$$

If the matrix **A** involves multiple eigenvalues, then diagonalization is impossible. For example, if the 3×3 matrix **A**, where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix}$$

has the eigenvalues λ_1 , λ_2 , λ_3 , then the transformation **x=Sz**, where

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 1 \\ \lambda_1 & 1 & \lambda_3 \\ \lambda_1^2 & 2\lambda_1 & \lambda_3^2 \end{bmatrix}$$

will yield

$$\mathbf{S}^{-1}\mathbf{AS} = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

y=**CPz**

7.6 SOLUTION OF TIME-INVARIANT STATE EQUATION

Solution of Homogeneous State Equations.

Before we solve vector-matrix differential equations, let us review the solution of the scalar differential equation

$$\dot{x} = ax$$

In solving this equation, we may assume a solution x(t) of the form

$$x(t) = b_0 + b_1 t + b_2 t^2 + \cdots + b_k t^k + \cdots$$

By substituting this assumed solution, we obtain

$$\begin{aligned} &b_1 + 2b_2 t + 3b_3 t^2 + \cdots + kb_k t^{k-1} + \cdots \\ &= a(b_0 + b_1 t + b_2 t^2 + \cdots + b_k t^k + \cdots) \end{aligned}$$

If the assumed solution is to be the true solution, for any t.

Hence, equating the coefficients of the equal powers of t, we obtain

$$\begin{aligned} b_1 &= ab_0 \\ b_2 &= \frac{1}{2} ab_1 = \frac{1}{2} a^2 b_0 \\ b_3 &= \frac{1}{3} ab_2 = \frac{1}{3 \times 2} a^3 b_0 \\ &\cdot \\ &\cdot \\ &\cdot \\ b_k &= \frac{1}{k!} a^k b_0 \end{aligned}$$

The value of b_0 is determined by substituting t=0

$$x(0)=b_0$$

Hence, the solution x(t) can be written as

$$\begin{aligned} x(t) &= \left(1 + at + \frac{1}{2!}a^2t^2 + \cdots + \frac{1}{k!}a^kt^k + \cdots\right)x(0) \\ &= e^{at}x(0) \end{aligned}$$

We shall now solve the vector-matrix differential equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$$

where $\mathbf{x} = n$-vector
$\mathbf{A} = n \times n$ constant matrix

By analogy with the scalar case, we assume that the solution is in the form of a vector power series in t, or

$$\mathbf{x}(t) = \mathbf{b}_0 + \mathbf{b}_1 t + \mathbf{b}_2 t^2 + \cdots + \mathbf{b}_k t^k + \cdots$$

By substituting this assumed solution, we obtain

$$\begin{aligned} &\mathbf{b}_1 + 2\mathbf{b}_2 t + 3\mathbf{b}_3 t^2 + \cdots + k\mathbf{b}_k t^{k-1} + \cdots \\ &= \mathbf{A}(\mathbf{b}_0 + \mathbf{b}_1 t + \mathbf{b}_2 t^2 + \cdots + \mathbf{b}_k t^k + \cdots) \end{aligned}$$

If the assumed solution is to be the true solution, for all t. Thus, by equating the coefficients of like powers of t on both sides of Equation we obtain

$$\begin{aligned} \mathbf{b}_1 &= \mathbf{A}\mathbf{b}_0 \\ \mathbf{b}_2 &= \frac{1}{2}\mathbf{A}\mathbf{b}_1 = \frac{1}{2}\mathbf{A}^2\mathbf{b}_0 \\ \mathbf{b}_3 &= \frac{1}{3}\mathbf{A}\mathbf{b}_2 = \frac{1}{3 \times 2}\mathbf{A}^3\mathbf{b}_0 \\ &\vdots \\ \mathbf{b}_k &= \frac{1}{k!}\mathbf{A}^k\mathbf{b}_0 \end{aligned}$$

By substituting t=0, we obtain

$$\mathbf{x}(0)=\mathbf{b}_0$$

Thus, the solution **x**(t) can be written as

$$\mathbf{x}(t) = \left(\mathbf{I} + \mathbf{A}t + \frac{1}{2!}\mathbf{A}^2t^2 + \cdots + \frac{1}{k!}\mathbf{A}^kt^k + \cdots\right)\mathbf{x}(0)$$

The expression in the parentheses on the right-hand side of this last equation is an n×n matrix. Because of its similarity to the infinite power series for a scalar exponential, we call it the matrix exponential and write

$$\mathbf{I} + \mathbf{A}t + \frac{1}{2!}\mathbf{A}^2t^2 + \cdots + \frac{1}{k!}\mathbf{A}^kt^k + \cdots = e^{\mathbf{A}t}$$

In terms of the matrix exponential, the solution of Equation (9–28) can be written as

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0)$$

7.6.1 Laplace Transform Approach to the Solution of Homogeneous State Equations

Let us first consider the scalar case:

$$\dot{x} = ax$$

Taking the Laplace transform, we obtain

$$sX(s) - x(0) = aX(s)$$

Where X(s) is Laplace of x(t).

Solving for X(s) gives

$$X(s) = \frac{x(0)}{s-a} = (s-a)^{-1}x(0)$$

The inverse Laplace transform of this last equation gives the solution

$$x(t) = e^{at}x(0)$$

The foregoing approach to the solution of the homogeneous scalar differential equation can be extended to the homogeneous state equation:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t)$$

Taking the Laplace transform of both sides of Equation, we obtain

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s)$$

Where X(s) is Laplace of x(t).

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{x}(0)$$

So

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0)$$

The inverse Laplace transform of X(s) gives the solution x(t). Thus,

$$\mathbf{x}(t) = \mathscr{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}]\mathbf{x}(0)$$

Note that

$$(s\mathbf{I} - \mathbf{A})^{-1} = \frac{\mathbf{I}}{s} + \frac{\mathbf{A}}{s^2} + \frac{\mathbf{A}^2}{s^3} + \cdots$$

Hence, the inverse Laplace transform gives

$$\mathscr{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}] = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2t^2}{2!} + \frac{\mathbf{A}^3t^3}{3!} + \cdots = e^{\mathbf{A}t}$$

the solution of Equation is obtained as

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0)$$

7.6.2 State-Transition Matrix

We can write the solution of the homogeneous state equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$$

as

$$\mathbf{x}(t) = \mathbf{\Phi}(t)\mathbf{x}(0)$$

Where φ(t) is an n×n matrix and is the unique solution of

$$\dot{\mathbf{\Phi}}(t) = \mathbf{A}\mathbf{\Phi}(t), \qquad \mathbf{\Phi}(0) = \mathbf{I}$$

To verify this, note that

$$\mathbf{x}(0) = \mathbf{\Phi}(0)\mathbf{x}(0) = \mathbf{x}(0)$$

And

$$\dot{\mathbf{x}}(t) = \dot{\mathbf{\Phi}}(t)\mathbf{x}(0) = \mathbf{A}\mathbf{\Phi}(t)\mathbf{x}(0) = \mathbf{A}\mathbf{x}(t)$$

$$\mathbf{\Phi}(t) = e^{\mathbf{A}t} = \mathscr{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}]$$

Note that

$$\mathbf{\Phi}^{-1}(t) = e^{-\mathbf{A}t} = \mathbf{\Phi}(-t)$$

7.6.3 Properties of State-Transition Matrices

For the time-invariant system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$$

for which

$$\mathbf{\Phi}(t) = e^{\mathbf{A}t}$$

we have the following:

1. $\mathbf{\Phi}(0) = e^{\mathbf{A}0} = \mathbf{I}$
2. $\mathbf{\Phi}(t) = e^{\mathbf{A}t} = \left(e^{-\mathbf{A}t}\right)^{-1} = \left[\mathbf{\Phi}(-t)\right]^{-1}$ or $\mathbf{\Phi}^{-1}(t) = \mathbf{\Phi}(-t)$
3. $\mathbf{\Phi}(t_1 + t_2) = e^{\mathbf{A}(t_1+t_2)} = e^{\mathbf{A}t_1}e^{\mathbf{A}t_2} = \mathbf{\Phi}(t_1)\mathbf{\Phi}(t_2) = \mathbf{\Phi}(t_2)\mathbf{\Phi}(t_1)$
4. $\left[\mathbf{\Phi}(t)\right]^n = \mathbf{\Phi}(nt)$
5. $\mathbf{\Phi}(t_2 - t_1)\mathbf{\Phi}(t_1 - t_0) = \mathbf{\Phi}(t_2 - t_0) = \mathbf{\Phi}(t_1 - t_0)\mathbf{\Phi}(t_2 - t_1)$

7.6.4 Solution of Nonhomogeneous State Equations

We shall begin by considering the scalar case

$$\dot{x} = ax + bu$$

rewrite Equation

$$\dot{x} - ax = bu$$

Multiplying both sides of this equation by e^{-at}, we obtain

$$e^{-at}\left[\dot{x}(t) - ax(t)\right] = \frac{d}{dt}\left[e^{-at}x(t)\right] = e^{-at}bu(t)$$

Integrating this equation between 0 and t gives

$$e^{-at}x(t) - x(0) = \int_0^t e^{-a\tau}bu(\tau)\,d\tau$$

or

$$x(t) = e^{at}x(0) + e^{at}\int_0^t e^{-a\tau}bu(\tau)\,d\tau$$

The first term on the right-hand side is the response to the initial condition and the second term is the response to the input u(t).

Let us now consider the nonhomogeneous state equation described by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

where $\mathbf{x} = n$-vector
$\mathbf{u} = r$-vector
$\mathbf{A} = n \times n$ constant matrix
$\mathbf{B} = n \times r$ constant matrix

By writing Equation as

$$\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t) = \mathbf{B}\mathbf{u}(t)$$

and premultiplying both sides of this equation by $e^{-\mathbf{A}t}$, we obtain

$$e^{-\mathbf{A}t}[\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t)] = \frac{d}{dt}[e^{-\mathbf{A}t}\mathbf{x}(t)] = e^{-\mathbf{A}t}\mathbf{B}\mathbf{u}(t)$$

Integrating the preceding equation between 0 and t gives

$$e^{-\mathbf{A}t}\mathbf{x}(t) - \mathbf{x}(0) = \int_0^t e^{-\mathbf{A}\tau}\mathbf{B}\mathbf{u}(\tau)\,d\tau$$

or

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau$$

Equation can also be written as

$$\mathbf{x}(t) = \mathbf{\Phi}(t)\mathbf{x}(0) + \int_0^t \mathbf{\Phi}(t-\tau)\mathbf{B}\mathbf{u}(\tau)\,d\tau$$

$\mathbf{\Phi}(t) = e^{\mathbf{A}t}$.

The solution **x**(t) is clearly the sum of a term consisting of the transition of the initial state and a term arising from the input vector.

7.6.5 Laplace Transform Approach to the Solution of Nonhomogeneous State Equations

The solution of the nonhomogeneous state equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

The Laplace transform of this last equation yields

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s)$$

Or

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{x}(0) + \mathbf{B}\mathbf{U}(s)$$

Premultiplying both sides of this last equation by $(s\mathbf{I}-\mathbf{A})^{-1}$, we obtain

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)$$

Solving

$$\mathbf{X}(s) = \mathscr{L}[e^{\mathbf{A}t}]\mathbf{x}(0) + \mathscr{L}[e^{\mathbf{A}t}]\mathbf{B}\mathbf{U}(s)$$

The inverse Laplace transform of this last equation can be obtained by use of the convolution integral as follows:

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau$$

Solution in Terms of x(t_0) Thus far we have assumed the initial time to be zero. If, however, the initial time is given by t0 instead of 0, then the solution to Equation must be modified to

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau$$

7.7 CONTROLLABILITY

A system is said to be controllable at time t_0 if it is possible by means of an unconstrained control vector to transfer the system from any initial state $\mathbf{x}(t_0)$ to any other state in a finite interval of time.

Consider the continuous-time system.

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u$$

The system described is said to be state controllable at $t=t_0$ if it is possible to construct an unconstrained control signal that will transfer an initial state to any final state in a finite time interval $t_0 \leq t \leq t_1$ If every state is controllable, then the system is said to be completely state controllable.

If the system is completely state controllable, then, given any initial state **x**(0), Equation must be satisfied. This requires that the rank of the n×n matrix be n.

$$\left[\mathbf{B} \;\vdots\; \mathbf{AB} \;\vdots\; \cdots \;\vdots\; \mathbf{A}^{n-1}\mathbf{B}\right]$$

The system given is completely state controllable if and only if the vectors **B**, **AB**, … , $\mathbf{A}^{n-1}\mathbf{B}$ are linearly independent, or the n×n matrix

$$\left[\mathbf{B} \;\vdots\; \mathbf{AB} \;\vdots\; \cdots \;\vdots\; \mathbf{A}^{n-1}\mathbf{B}\right]$$

is of rank n.

7.7.1 Condition for Complete State Controllability in the *s* Plane

The condition for complete state controllability can be stated in terms of transfer functions or transfer matrices.

It can be proved that a necessary and sufficient condition for complete state controllability is that no cancellation occur in the transfer function or transfer matrix. If cancellation occurs, the system cannot be controlled in the direction of the cancelled mode.

7.7.2 Output Controllability

Complete state controllability is neither necessary nor sufficient for controlling the output of the system. For this reason, it is desirable to define separately complete output controllability. Consider the system described by

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$
$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du}$$

where $\mathbf{x}$ = state vector (n-vector)
$\mathbf{u}$ = control vector (r-vector)
$\mathbf{y}$ = output vector (m-vector)
$\mathbf{A} = n \times n$ matrix
$\mathbf{B} = n \times r$ matrix
$\mathbf{C} = m \times n$ matrix
$\mathbf{D} = m \times r$ matrix

The system is said to be completely output controllable if it is possible to construct an unconstrained control vector **u**(t) that will transfer any given initial output **y**(t_0) to any final

output $\mathbf{y}(t_1)$ in a finite time interval $t_0 \leq t \leq t_1$ is completely output controllable if and only if the $m \times (n+1)r$ matrix

$$\left[\mathbf{CB} \;\vdots\; \mathbf{CAB} \;\vdots\; \mathbf{CA}^2\mathbf{B} \;\vdots\; \cdots \;\vdots\; \mathbf{CA}^{n-1}\mathbf{B} \;\vdots\; \mathbf{D}\right]$$

is of rank m.

7.8 OBSERVABILITY

The system is said to be completely observable if every state $\mathbf{x}(t_0)$ can be determined from the observation of $\mathbf{y}(t)$ over a finite time interval, $t_0 \leq t \leq t_1$.

The concept of observability is useful in solving the problem of reconstructing unmeasurable state variables from measurable variables in the minimum possible length of time.

The concept of observability is very important because, in practice, the difficulty encountered with state feedback control is that some of the state variables are not accessible for direct measurement, with the result that it becomes necessary to estimate the unmeasurable state variables in order to construct the control signals.

If the system is completely observable, then, given the output $\mathbf{y}(t)$ over a time interval $0 \leq t \leq t_1$ $\mathbf{x}(0)$ is uniquely determined It can be shown that this requires the rank of the $nm \times n$ matrix

$$\begin{bmatrix} \mathbf{C} \\ \hdashline \mathbf{CA} \\ \hdashline \cdot \\ \cdot \\ \cdot \\ \hdashline \mathbf{CA}^{n-1} \end{bmatrix}$$

to be n.

7.8.1 Conditions for Complete Observability in the s Plane.

The conditions for complete observability can also be stated in terms of transfer functions or transfer matrices.

The necessary and sufficient conditions for complete observability is that no cancellation occur in the transfer function or transfer matrix. If cancellation occurs, the cancelled mode cannot be observed in the output.

7.9 Pole Placement Control Design

Assumptions:

- The system is completely state controllable.
- The state variables are measurable and are available for feedback.
- Control input is unconstrained

Pole Placement Control Design Objective: The closed loop poles should lie at which are their 'desired locations'. Difference from classical approach: Not only the "dominant poles", but "all poles" are forced to lie at specific desired locations.

Necessary and sufficient condition: The system is completely state controllable.

Consider a control system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u$$
$$y = \mathbf{C}\mathbf{x} + Du$$

Where **x** = state vector (n-vector)

D = constant (scalar)

C = 1 × n constant matrix

B = n × 1 constant matrix

A = n × n constant matrix

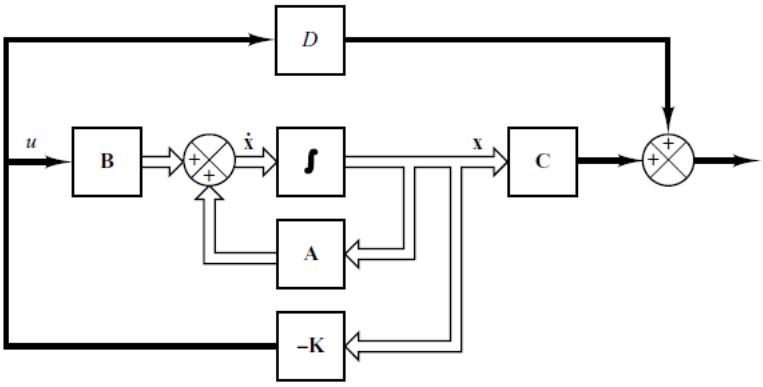
u = control signal (scalar)

y = output signal (scalar)

choose the control signal to be

u = **-Kx**

This means that the control signal *u* is determined by an instantaneous state. Such a scheme is called state feedback. The 1×n matrix **K** is called the state feedback gain matrix



This closed-loop system has no input. Its objective is to maintain the zero output.

Because of the disturbances that may be present, the output will deviate from zero. The nonzero output will be returned to the zero-reference input because of the state feedback scheme of the system. Such a system where the reference input is always zero is called a regulator system.

Substituting Equation

$$\dot{\mathbf{x}}(t) = (\mathbf{A} - \mathbf{BK})\mathbf{x}(t)$$

The solution of this equation is given by

$$\mathbf{x}(t) = e^{(\mathbf{A}-\mathbf{BK})t}\mathbf{x}(0)$$

where **x**(0) is the initial state caused by external disturbances. The stability and transient response characteristics are determined by the eigenvalues of matrix **A-BK**