KIITPOLYTECHNIC

LECTURE NOTES

ON

CONTROL SYSTEMS ENGINEERING

Compiled by

Mr. Gautam Kumar Mahto
(Lecturer in Department of Electrical Engineering, KIIT Polytechnic, BBSR)

Disclaimer

This document does not claim any originality and cannot be used as a substitute for prescribed textbooks. The information presented here is merely a collection by the committee members for their respective teaching assignments. Various sources of freely available materials from internet were consulted for preparing this document. The ownership of the information lies with the respective authors or institutions. Further, this document is not intended to be used for commercial purpose and the committee members are not accountable for any issues, legal or otherwise, arising out of use of this document. The committee members make no representations or warranties with respect to the accuracy or completeness of the contents of this document and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose.

CONTENTS

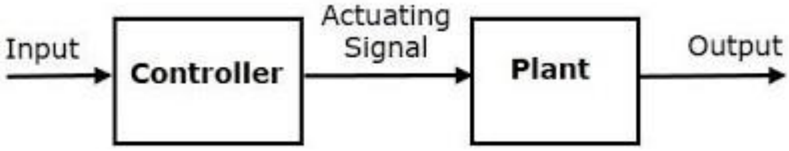
S.No	Chapter Name	Page No
1	Fundamental of Control System	4-9
2	Transfer Functions	10-13
3	Control system Components & Mathematical Modelling of Physical System	14-20
4	Block Diagram & Signal Flow Graphs (SFG)	21-36
5	Time Domain Analysis of Control Systems	37-51
6	Feedback Characteristics of Control Systems	52-55
7	Stability Concept & Root Locus Method	56-66
8	Frequency-Response Analysis & Bode Plot	67-114
9	State Variable Analysis	115-124

Unit-1: Fundamental of Control System

A Control System is a system in which the output is controlled by varying the input. Here a desired response by controlling the output.

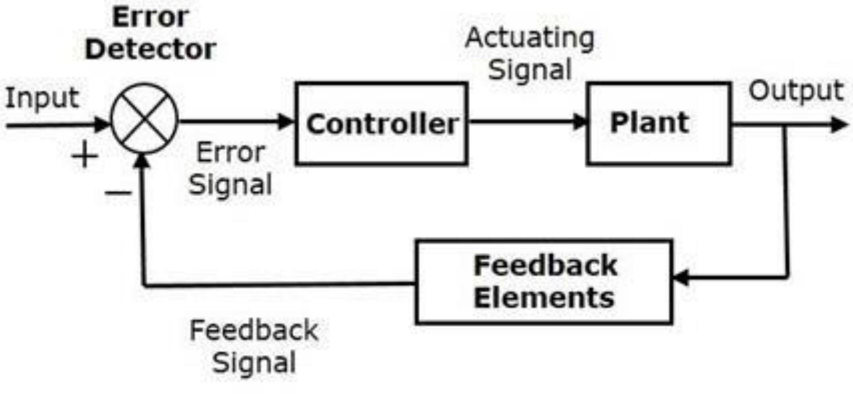
Open loop control system

An open-loop control system is a system in which the control action is independent of the desired output signal. In this system, the output signal is not compared with the input signal which means there is no feedback signal in this system. The open-loop control system is also known as a non-feedback control system or control system without feedback.


Input
Controller
Actuating Signal
Plant
Output

Closed loop control system

A closed loop control system is a system in which control action is dependent on the desired output. In this system, the output signal is compared with the reference input signal, and an error signal is produced then this error signal is fed to the controller to reduce the error to obtain the desired output.


Error Detector
Input
+
−
Error Signal
Controller
Actuating Signal
Plant
Output
Feedback Elements
Feedback Signal

Linear and Non-linear system

Linear system: A system is known as linear if and only if it possesses both homogeneity and superposition properties. Superposition implies that an input r_1 (t) gives an output c_1 (t) and another input r_2 (t) gives the output c_2 (t). If two inputs are applied together then the output will be the sum of two outputs:

$R_1(t) + R_2(t) = C_1(t) + C_2(t)$

If our input-output relationship is a straight line passing through the origin, then the system obeys the superposition property. The straight line passing through origin means that the output is zero (0) for zero (0) input.

If the input increases for any system K time from r_1 (t) to Kr_1 (t) then the magnitude of the output is also increased from c_1 (t) to Kc_1 (t) then this property is known as homogeneity. This property is a necessary condition for a system to be linear.

Non-Linear System: Non-linear system does not satisfy the superposition principle or homogeneity property, or it is the system whose output is not directly proportional to its input. Here, the stability of the non-linear system depends upon the input and initial status of the system.

In a linear system, if the input is sinusoidal and starts increasing, then the output will also increase but the form will remain the same.

However, in a non-linear system, the form may change with changes in the magnitude of the input. It means if the input is sinusoidal then the output is non-sinusoidal, i.e., the non-linear system.

Time-Variant and Invariant Control System

The system whose parameter vary with time is known as a time-varying control system and the system whose parameter does not vary with time is called as a time-invariant control system.

Continuous data and discrete data control system

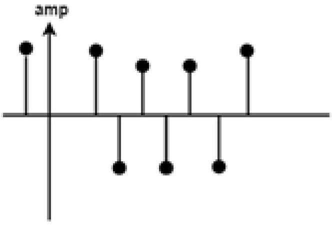
In a continuous system, all system variables are the function of continuous time variable 't.' At any time't' they are dependent on time thus they are called continuous data control system.

In discrete data control system, if the signal is not continuously varying with time but it is in the form of pulses, the controlled system is called discrete data control system.

It is of two types

1. Sampled
2. Digital

If the signal is in the form of pulse data, the system is called a sampled data control system. The sampled form is shown in the below-drawn diagram.



If the signal is in the form of digital code, the system is called a digital control system.

Dynamic and Static system

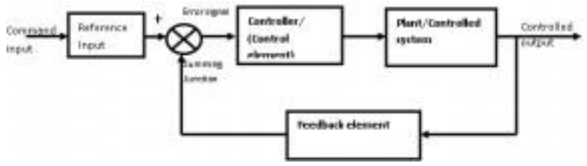
If in any system the input does not change with the time then the output will also not change with time such system is known as a static system. For example, an electric circuit with resistances.

If the output of the system is a function of time even when the input is constant, such system is called Dynamic system like R, L, C circuit because inductance and capacitance are energy storing devices.

Effects of Feedback

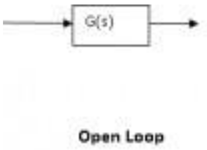
The closed loop control systems are also called feedback control system. Figure shows the block diagram of closed loop(feedback control) system with more details the various terminologies used are as given below,

1. Reference Input /Command Input: The input to the entire system is called reference input.
2. Input transducer: This converts the form of the input to the form used by the controller
3. The output transducer: This measures the output response and converts it into the form used by the controller.
4. Feedback Path: It is the return path from the output to the summing junction. It consists of a feedback element which feeds back the output to the error detector
5. Error detector: The error detector receives the measured signal from output and compares it with the reference input. The difference of two signals produces the error signal also called as the actuating signal
6. Control Element/controller: it regulates the output according to the signal obtained from the error detector.

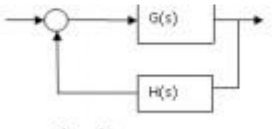


Effect of Feedback on gain:

Consider a system with overall gain G(s) as shown in figure with feedback function H(S) then the overall gain of the system with negative feedback is given by G(s)/[1+ G(s)H(s)]. Hence for a negative feedback the gain G(S) is reduced by a factor = 1/[1+ G(s)H(s)]



Open Loop



Closed Loop

Effect of feedback on stability:

The feedback in a system can improve stability also it can be harmful if not applied properly

Effect of feedback on sensitivity:

A good control system should always be very insensitive to the parameter variations and it should be sensitive to commands

Hence feedback can increase or decrease the sensitivity of the system.

Effect of feedback on external disturbance or noise:

Feedback reduces the effect of noise and disturbances on the system performance.

In general the feedback also has effects on performance parameters such as bandwidth, impedance, transient response and frequency response, 1) Increases accuracy, 2) Increases the bandwidth and 3) Stabilizes the unstable system.

Standard Test Signals

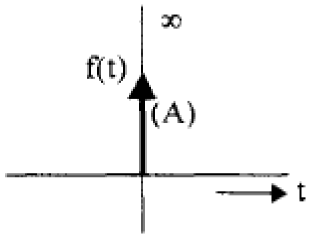
These standard input signals are

- an impulse,
- a step,
- a ramp and
- a parabolic input.

By using above standard test signals of control systems, analysis and design of control systems are carried out, defining certain performance measures for the system.

Impulse Signal

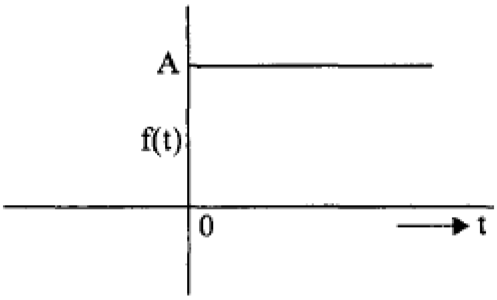
In bellow an impulse signal is shown in Fig



The impulse function is zero for all t not equal to 0 and it is infinity at $t = 0$. It rises to infinity at $t = 0^-$ and comes back to zero at $t = 0^+$ enclosing a finite area. If this area is A it is called as an impulse function of strength A. If A = 1 it is called a unit impulse function. Thus an impulse signal is denoted by $f(t)$ = A (t).

Step Signal

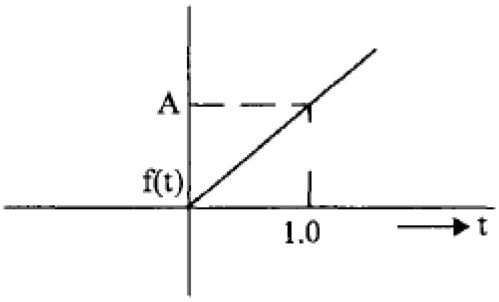
In bellow a step signal is shown in Fig.



It is zero for $t < 0$ and suddenly rises to a value A at $t = 0$ and remains at this value for $t > 0$: It is denoted by *f(t) = Au (t).* If A = 1, it is called a *unit step function.*

Ramp signal

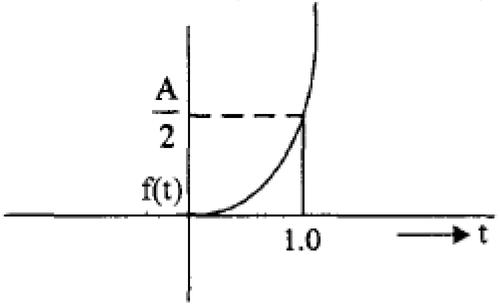
In bellow a ramp signal is shown in Fig



It is zero for $t < 0$ and uniformly increases with a slope equal to A. It is denoted by *f (t) = At.* If the slope is unity, then it is called a *unit ramp signal.*

Parabolic signal

In bellow a parabolic signal is shown in Fig



A parabolic signal is denoted by *f (t)* = .If A is equal to unity then it is known as a *unit parabolic signal.*

Servomechanism

Servomechanism is a powered mechanism producing motion or forces at a higher level of energy than the input level, e.g. in the brakes and steering of large motor vehicles, especially where feedback is employed to make the control automatic.

A servomechanism, sometimes also called ***Servo***, is an automatic device that uses error-sensing negative feedback to correct the performance of a mechanism and is defined by its function.

Components of Servomechanism:

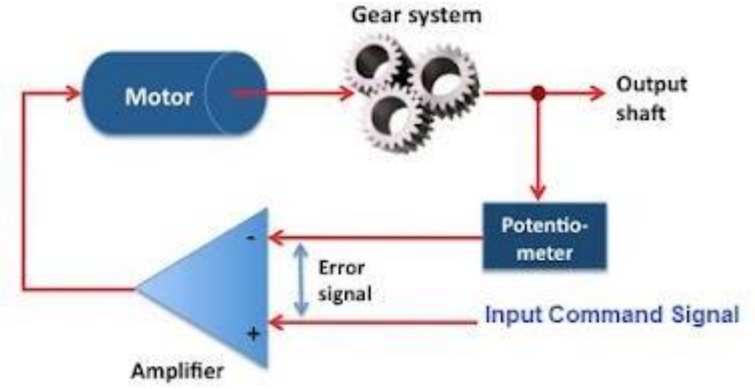
A servo system mainly consists of three basic components

- A controlled device

- A output sensor
- A feedback system

Working of Servomechanism:

Servomechanism is an automatic closed loop control system. Here instead of controlling a device by applying variable input signal, the device is controlled by a feedback signal generated by comparing output signal and reference input signal.



Carefully observe the figure above and think. When reference input signal or command signal is applied to the system, it is compared with output reference signal of the system produced by output sensor, and a third signal produced by feedback system. This third signal acts as input signal of controlled device. This input signal to the device presents as long as there is a logical difference between reference input signal and output signal of the system. After the device achieves its desired output, there will be no longer logical difference between reference input signal and reference output signal of the system. Then, third signal produced by comparing theses above said signals will not remain enough to operate the device further and to produce further output of the system until the next reference input signal or command signal is applied to the system.

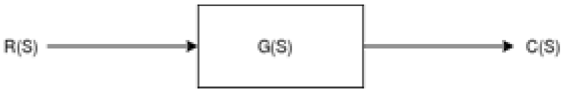
Hence the primary task of a servomechanism is to maintain the output of a system at the desired value in the presence of disturbances.

Unit-2: Transfer Functions

The transfer function of a system is defined as the ratio of Laplace transform of output to the **Laplace transform of input where all the initial** conditions are zero.

Input
Transfer Function
Output

$$Transferfunction = \frac{Laplace\ transform\ of\ output\ variable}{Laplace\ transform\ of\ input\ variable}$$



$$T(S) = \frac{C(S)}{R(S)} = G(S)$$

Where,

1. T(S) = Transfer function of the system.
2. C(S) = output.
3. R(S) = Reference output.
4. G(S) = Gain.

Steps to obtain transfer function -

Step-1 Write the differential equation.

Step-2 Find out Laplace transform of the equation assuming 'zero' as an initial condition.

Step-3 Take the ratio of output to input.

Step-4 Write down the equation of G(S) as follows -

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

$$=\frac{b_{\square}S^{m}+b_{m-1}s^{m-1}+\ldots\ldots\ldots b_1S+b_0}{a_nS^{n}+a_{n-1}S^{n-1}+\ldots\ldots\ldots a_1S+a_0} \qquad \text{-----Eq.1}$$

Here, a and b are constant, and S is a complex variable

Characteristic equation of a transfer function -

Here, the characteristic equation of a linear system can be obtained by equating the denominator to the polynomial of a transfer function is zero. Thus the characteristic equation of the transfer function of Eq.1 will be:

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

Advantages of Transfer function

1. If transfer function of a system is known, the response of the system to any input can be determined very easily.

2. A transfer function is a mathematical model and it gives the gain of the system.

3. Since it involves the Laplace transform, the terms are simple algebraic expressions and no differential terms are present.

4. Poles and zeroes of a system can be determined from the knowledge of the transfer function of the system.

Disadvantages of Transfer function

1. Transfer function does not take into account the initial conditions.

2. The transfer function can be defined for linear systems only.

3. No inferences can be drawn about the physical structure of the system

Poles and Zeros of a transfer function -

Consider the equation 1, the numerator and denominator can be factored in m and n terms respectively:

$$\frac{C(s)}{R(s)} = G(s) = \frac{K(s+Z_1)(s+Z_2)(as^2+bs+c)}{(s+p_1)(s+p_2)(As^2+Bs+C)}$$

Where, $K = \frac{b_m}{a_n}$ is known as the gain factor and's' is the complex frequency.

Poles

Poles are the frequencies of the transfer function for which the value of the transfer function becomes zero.

Zeros

Zeros are the frequencies of the transfer function for which the value of the transfer function becomes zero.

We will apply Sridhar Acharya method to find the roots of poles and zeros -

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

If any poles or zeros coincide then such poles and zeros are called multiple poles or multiple zeros.

If the poles and zeros do not coincide then such poles and zeros are called simple poles or simple zeros.

For example-

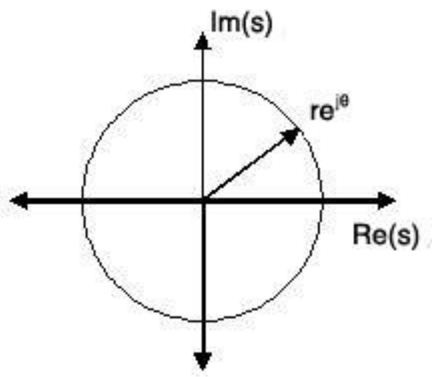
Find the transfer function of the following function

$$G(S) = \frac{50(S+3)}{S(S+2)(S+4)^2}$$

The zeros of the function are S = -3 and the poles of the function are S = 0, S = -2, and multiple poles at S = -4 i.e. the pole of order 2 at S = -4.

The S-Plane

Once the poles and zeros have been found for a given Laplace Transform, they can be plotted onto the S-Plane. The S-plane is a complex plane with an imaginary and real axis referring to the complex-valued variable z. The position on the complex plane is given by $re^{j\theta}$ and the angle from the positive, real axis around the plane is denoted by θ. When mapping poles and zeros onto the plane, poles are denoted by an "x" and zeros by an "o". The below figure shows the S-Plane, and examples of plotting zeros and poles onto the plane can be found in the following section.



Examples of Pole/Zero Plots

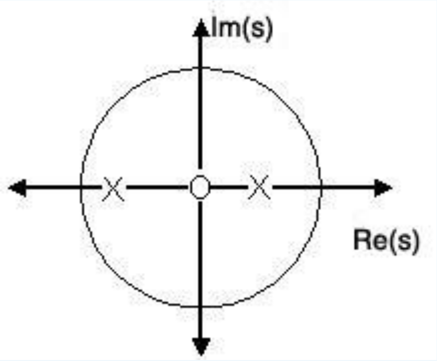
This section lists several examples of finding the poles and zeros of a transfer function and then plotting them onto the S-Plane.

Example: Simple Pole/Zero Plot

$H(s)=s/[(s-1/2)(s+3/4)]$

The zeros are: $\{0\}$

The poles are: $\{1/2, -3/4\}$



Using the zeros and poles found from the transfer function, the one zero is mapped to zero and the two poles are placed at $1/2$ and $-3/4$

Unit-3: Control system Components & Mathematical Modelling of Physical System

Components of Control System:

Feedback controls are widely used in modern automated systems. A feedback control system consists of five basic components: (1) input, (2) process being controlled, (3) output, (4) sensing elements, and (5) controller and actuating devices. These five components are illustrated in Figure 1. The term close-loop feedback control is often used to describe this kind of system.

The input to the system is the reference value, or set point, for the system output. This represents the desired operating value of the output. Using the previous example of the heating system as an illustration, the input is the desired temperature setting for a room. The process being controlled is the heater (e.g., furnace). In other feedback systems, the process might be a manufacturing operation, the rocket engines on a space shuttle, the automobile engine in cruise control, or any of a variety of other processes to which power is applied. The output is the variable of the process that is being measured and compared to the input; in the above example, it is room temperature.

The sensing elements are the measuring devices used in the feedback loop to monitor the value of the output variable. In the heating system example, this function is normally accomplished using a bimetallic strip. This device consists of two metal strips joined along their lengths. The two metals possess different thermal expansion coefficients; thus, when the temperature of the strip is raised, it flexes in direct proportion to the temperature change. As such, the bimetallic strip is capable of measuring temperature. There are many different kinds of sensors used in feedback control systems for automation.

The purpose of the controller and actuating devices in the feedback system is to compare the measured output value with the reference input value and to reduce the difference between them. In general, the controller and actuator of the system are the mechanisms by which changes in the process are accomplished to influence the output variable. These mechanisms are usually designed specifically for the system and consist of devices such as motors, valves, solenoid switches, piston cylinders, gears, power screws, pulley systems, chain drives, and other mechanical and electrical components. The switch connected to the bimetallic strip of the thermostat is the controller and actuating device for the heating system. When the output (room temperature) is below the set point, the switch turns on the heater. When the temperature exceeds the set point, the heat is turned off.

Mathematical Modelling and Representation of a Physical System:

A physical system is a system in which physical objects are connected to perform an objective. We cannot represent any physical system in its real form. Therefore, we have to make assumptions for analysis and synthesis of systems. An idealized physical system is called a physical model. A physical system can be modelled in different ways depending upon the problem and required accuracy with which we have to deal.

There are two types of physical system:

1. Mechanical system.
2. Electrical system.

Mathematical modelling of the Mechanical system:

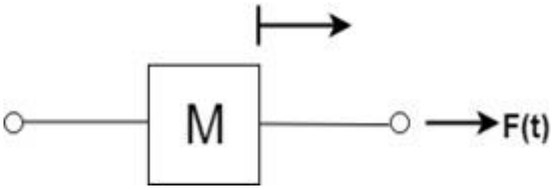
There are two types of mechanical systems which we have to study.

1. Translational or Linear system.
2. Rotational system.

Translational or Linear system:

The motion that takes place along a straight line is called a translational motion. There are three different types of forces that we have to study.

1. Inertia Force:



Consider a body of mass 'M' and acceleration 'a' then according to newton's second law of motion:

1. F (t) = Ma (t)

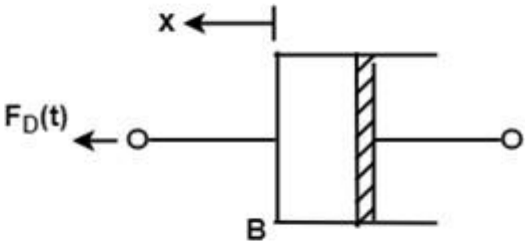
In terms of velocity

$$F(t) = M\frac{dv(t)}{dt}$$

In terms of displacement the above equation can be expressed as

$$F(t) = M\frac{d^2x(t)}{dt^2}$$

2. Damping Force:

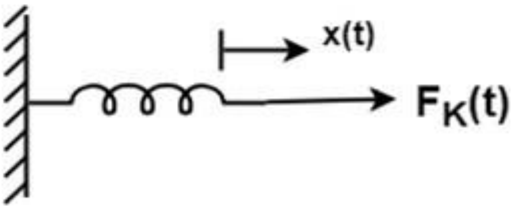


For viscous friction, we assume that the damping force is proportional to the velocity.

$$F_D(t) = Bv(t) = B\frac{d}{dt}x(t)$$

Where B = Damping Coefficient Unit of B = N/m/sec.

3. Spring Force:



A spring has potential energy. The restoring force of a spring is proportional to the displacement.

$F_K(t)$ a $x(t) = Kx(t)$

$F_k(t) = k\int v(t)dt$

Where

K = spring constant of stiffness

Unit of K = N/m

Rotational System:

When the motion of a body takes place about a fixed axis, this type of motion is known as rotational motion. There are three types of torques that resist the rotational motion.

1. Inertia Torque: The property of an element that stores the kinetic energy of rotational motion is called inertia (J). The inertia Torque T_1 is the product of the moment of inertia J and angular acceleration α (t).

$$T_1(t) = J\alpha(t)$$

$$T_1(t) = J\frac{d}{dt}\omega(t)$$

$$T_1(t) = J\frac{d^2}{dt^2}\theta(t)$$

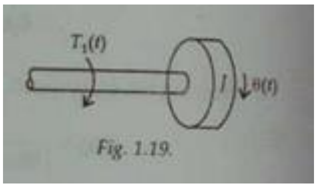


Fig. 1.19.

Where

ω(t) = angular veloci

θ(t) = angular displacement

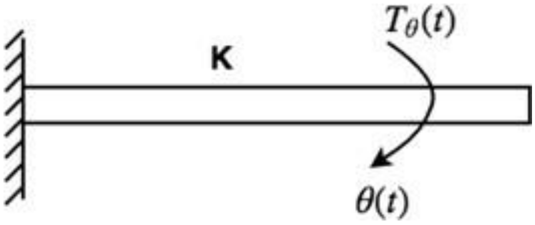
Unit of Torque = N-m

S.No.	Translational	Rotational
1	Force, F	Torque, T
2	Acceleration, a	angular acceleration, α
3	Velocity, v	angular velocity, ω
4	Displacement, x	angular displacement, θ
5	Mass, M	Moment of inertia, J
6	Damping Coefficient, B	Rotational damping Coefficient, B
7	Stiffness	torsional stiffness

2. Damping Torque: The product of angular velocity ω and damping coefficient B is known as Damping Torque $T_D(t)$.

1. TD(t) = Bω(t)
2.
3. TD(t) = B d/dt θ(t)

3. Spring Torque:



Spring Torque Tθ(t) is the product of torsional stiffness and angular displacement.

1. Tθ(t) = Kθ(t)

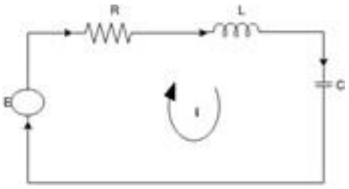
Unit of K = N-m/rad.

By comparing equations, we get an analogous system:

Mathematical modelling of the Electrical system:

Analogous System

Consider a series RLC circuit.



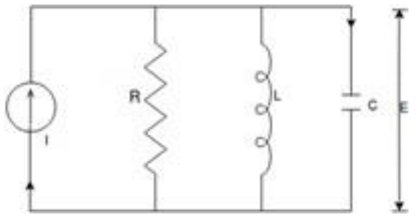
Apply Kirchhoff's voltage law

$$E = Ri + L\frac{di}{dt} + \frac{1}{C}\int i dt \qquad \text{..................................eq. 1}$$

In terms of charge equation becomes

$$E = R\frac{dq}{dt} + L\frac{d^2q}{dt} + \frac{1}{C}q \qquad \text{..................................eq. 2}$$

Now consider a parallel RLC circuit



Now apply Kirchhoff's current law

$$I = \frac{E}{R} + \frac{1}{L}\int E\,dt + C\frac{dE}{dt} \qquad \text{...............................eq. 3}$$

In terms of magnetic flux linkage, the above equation becomes Since

$$\phi = \int E\,dt$$

$$I = \frac{1}{R}\left(\frac{d\Phi}{dt}\right) + \frac{1}{L}\Phi + C\frac{d^2\Phi}{dt^2} \qquad \text{...............................eq. 4}$$

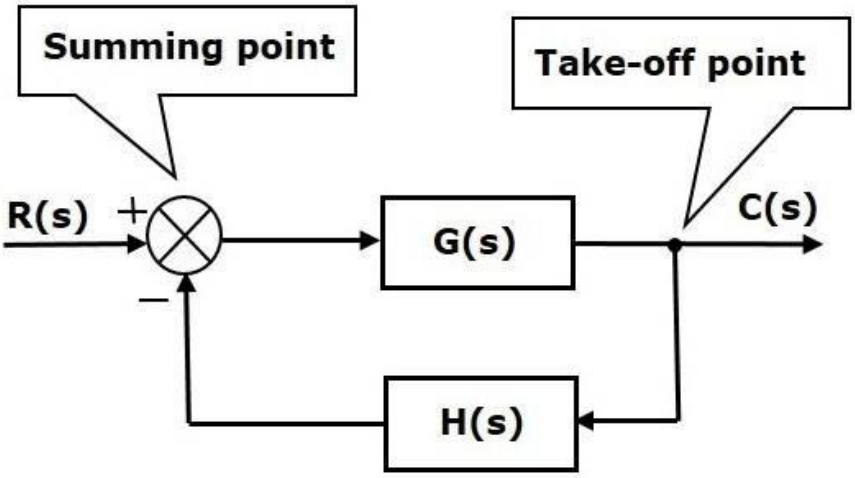
Force Current Analogy

S.No.	**Mechanical Translational System**	**Electrical System**
1	Force (F)	Voltage (E)
2	Mass (M)	Inductance (L)
3	Stiffness(K) (Elastance, 1/K)	Reciprocal of Capacitance (1/C) Capacitance (C)
4	Damping Coefficient (B)	Resistance R
5	Displacement (x)	Charge (q)

Unit-4: Block Diagram & Signal Flow Graphs (SFG)

Basic Elements of Block Diagram

The basic elements of a block diagram are a block, the summing point and the take-off point. Let us consider the block diagram of a closed loop control system as shown in the following figure to identify these elements.

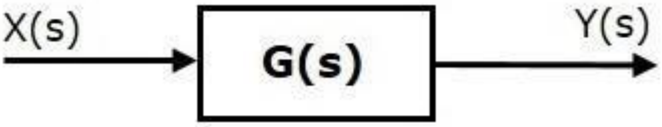


The above block diagram consists of two blocks having transfer functions G(s) and H(s). It is also having one summing point and one take-off point. Arrows indicate the direction of the flow of signals. Let us now discuss these elements one by one.

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

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

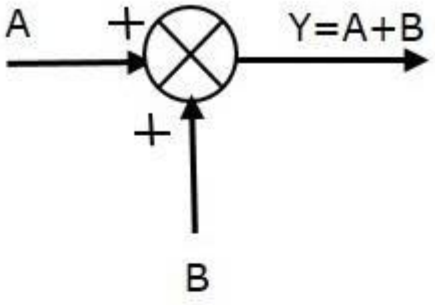
Output of the block is obtained by multiplying transfer function of the block with input.

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. Let us see these three operations one by one.

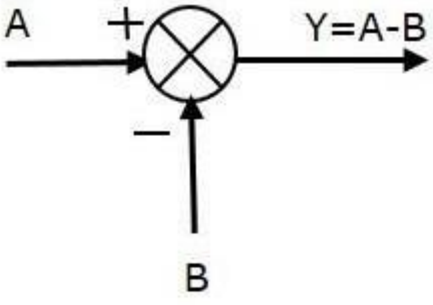
The following figure shows the summing point with two inputs (A, B) and one output (Y). Here, the inputs A and B have a positive sign. So, the summing point produces the output, Y as **sum of A and B.**

i.e.,Y = A + B.



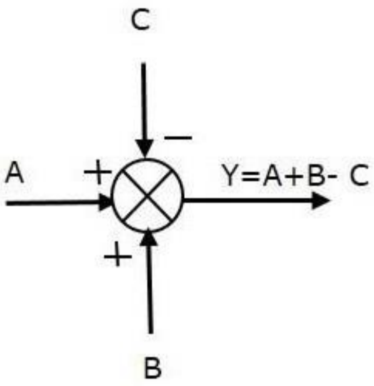
The following figure shows the summing point with two inputs (A, B) and one output (Y). Here, the inputs A and B are having opposite signs, i.e., A is having positive sign and B is having negative sign. So, the summing point produces the output **Y** as the **difference of A and B**.

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



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

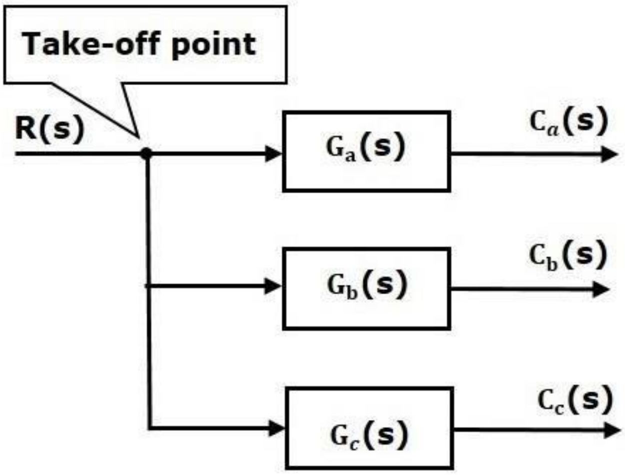
Y = A + B + (−C) = A + B − C.



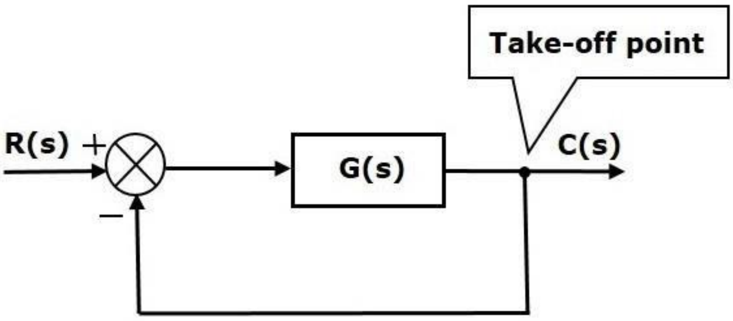
Take-off Point

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.

In the following figure, the take-off point is used to connect the same input, R(s) to two more blocks.


Take-off point
R(s)
$G_a(s)$
$C_a(s)$
$G_b(s)$
$C_b(s)$
$G_c(s)$
$C_c(s)$

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

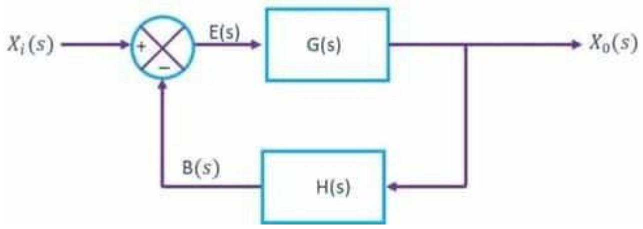

Take-off point
R(s)
+
−
G(s)
C(s)

Canonical Form of Closed Loop Control System:

Let,

- $X_i(s)$→Input signal
- $X_0(s)$→Output signal
- G(s)→Transfer function of the open-loop control system
- H(s)→Gain of the elements in the feedback path
- B(s)→The output of feedback elements
- E(s)→Error signal

Then the closed-loop control system can be represented as shown below.



From the above diagram

$G(s) = X_0(s)/E(s)$

$\Rightarrow E(s) = X_0(s)/G(s)$

E(s) can also be written as

$E(s) = X_i(s) - B(s)$

$\Rightarrow X_0(s)/G(s) = X_i(s) - B(s)$

From the diagram,

$H(s) = B(s)/X_0(s)$

$\Rightarrow X_0(s)/G(s) = X_i(s) - H(s)\ X0(s)$

$\Rightarrow X_0(s) = G(s)\ Xi(s) - G(s)\ H(s)\ X0(s)$

$\Rightarrow X_0(s)\ [1+G(s)\ H(s)] = G(s)\ Xi(s)$

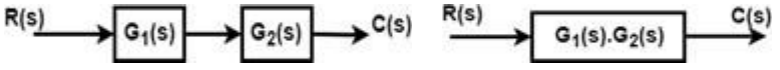
$\Rightarrow X_0(s)/X_i(s) = G(s)/1+G(s)\ H(s)$

This is the standard form of representing the closed loop control system with negative feedback.

Block diagram reduction rules:

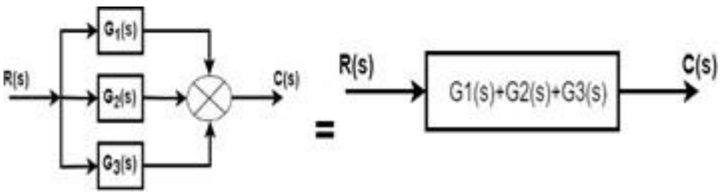
Rule No.1. Blocks in Cascade

When two or more blocks are connected **in series**, then the resultant block is the product of the individual blocks.


R(s)
G₁(s)
G₂(s)
C(s)
R(s)
G₁(s).G₂(s)
C(s)

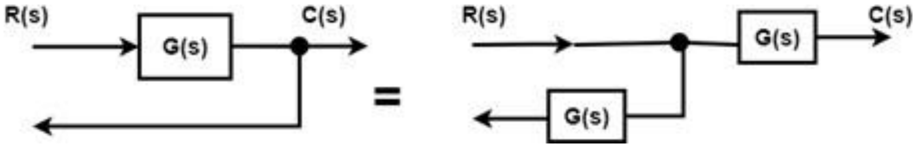
Rule No.2 Blocks in parallel

When two or more blocks are connected in parallel, then the resultant block is the sum of the individual blocks.


R(s)
G₁(s)
G₂(s)
G₃(s)
C(s)
=
R(s)
G1(s)+G2(s)+G3(s)
C(s)

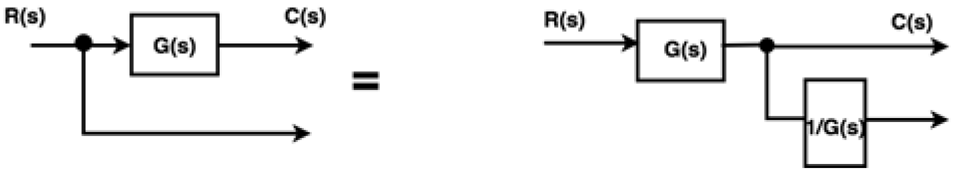
Rule No.3 Moving a take-off point ahead of a block

When the take-off point is moved ahead of a block (before the block), then the same transfer function is introduced in the take-off point branch.

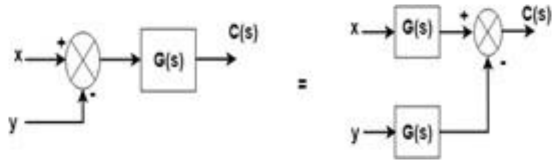

R(s)
G(s)
C(s)
=
R(s)
G(s)
C(s)
G(s)

Rule No.4 Moving the take-off point after the block

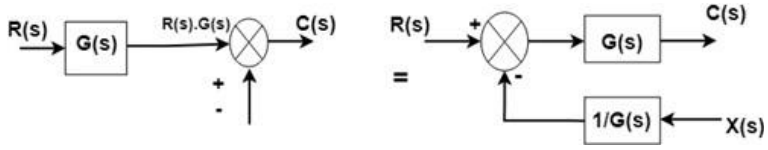
When the take-off point is moved after the block, then a block with reciprocal of a transfer function is introduced in the take-off point branch.


R(s)
G(s)
C(s)
=
R(s)
G(s)
C(s)
1/G(s)

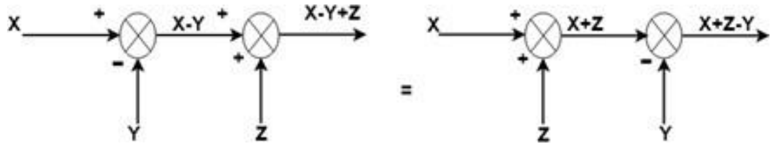
Rule No.5 Moving a summing point beyond the block


x
+
-
y
G(s)
C(s)
=
x
G(s)
+
-
C(s)
y
G(s)

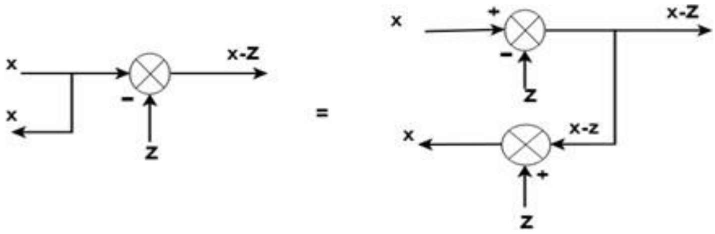
Rule No.6:Moving a summing point ahead of a block


R(s)
G(s)
R(s).G(s)
C(s)
+
-
R(s)
+
-
=
G(s)
C(s)
1/G(s)
X(s)

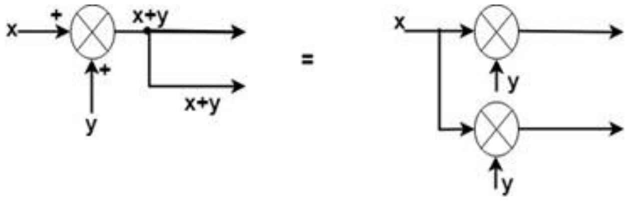
Rule No.7: Interchanging two summing points


X
+
X-Y
+
X-Y+Z
-
+
Y
Z
=
X
+
X+Z
X+Z-Y
+
-
Z
Y

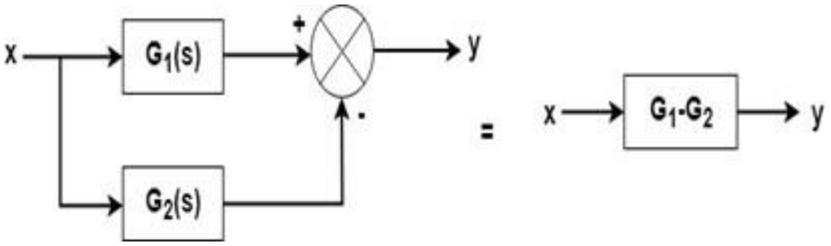
Rule No.8:Moving a take-off point beyond a summing point


x
x
x-Z
-
z
=
x
+
x-Z
-
z
x
x-z
+
z

Rule No.9: Moving a take-off point ahead of a summing point

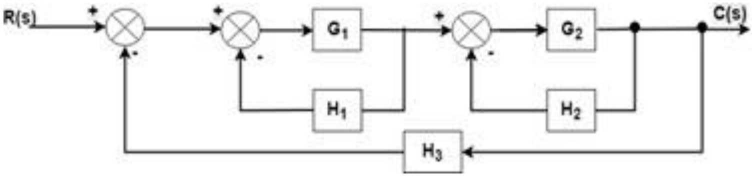


Rule No.10: Eliminating a forward loop


x
G1(s)
+
y
-
G2(s)
=
x
G1-G2
y

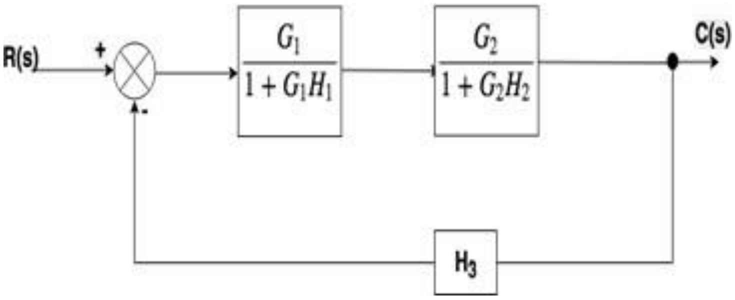
Example

Find the transfer function of the following by block reduction technique.

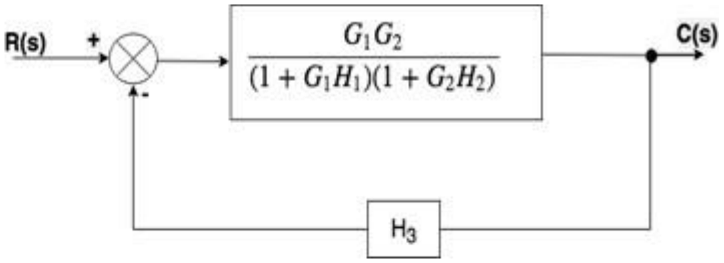

R(s)
+
-
+
-
G1
H1
+
-
G2
H2
C(s)
H3

Solution

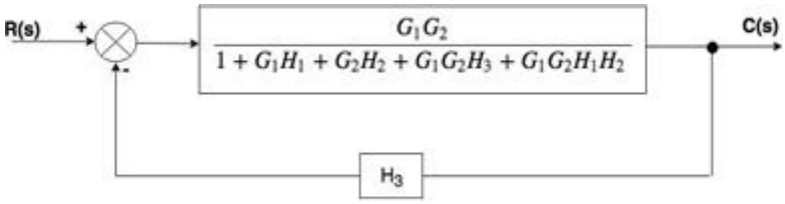
Step 1: There are two internal closed loops. Firstly, we will remove this loop.


R(s)
+
-
G1 / 1 + G1H1
G2 / 1 + G2H2
C(s)
H3

Step 2: When the two blocks are in a cascade or series we will use rule no.1.



Step 3: Now we will solve this loop.



Step 4:

$$\frac{C(s)}{R(s)} = \frac{G_1G_2}{1+G_1H_1+G_2H_2+G_1G_2H_3+G_1G_2H_1H_2}$$

Ans.

Signal Flow Graph (SFG):

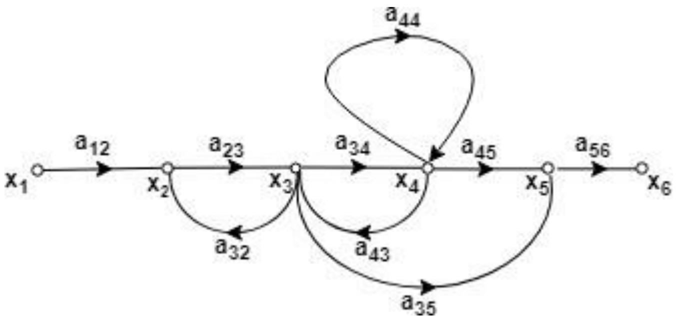
Introduction:

Block diagram reduction is the excellent method for determining the transfer function of the control system. However, in a complicated system, it is very difficult and time-consuming process that is why an alternate method, i.e., SFG was developed by S.J Mason which relates the input and output system variables graphically. In the signal flow graph, the transfer function is referred to as transmittance.

Characteristics of SFG:

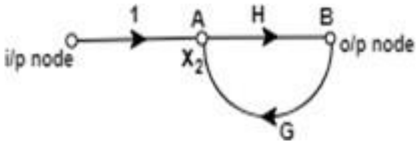
SFG is a graphical representation of the relationship between the variables of a set of linear algebraic equations. It doesn't require any reduction technique or process.

- It represents a network in which nodes are used for the representation of system variable which is connected by direct branches.
- SFG is a diagram which represents a set of equations. It consists of nodes and branches such that each branch of SFG having an arrow which represents the flow of the signal.
- It is only applicable to the linear system.

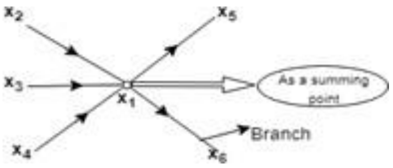

a44
a12
a23
a34
a45
a56
x1
x2
x3
x4
x5
x6
a32
a43
a35

Terms used in SFG

Node: It represents the system variable which equals to the sum of all signals. Outgoing signal from the node does not affect the value of node variables.


1
A
H
B
i/p node
X2
o/p node
G

Branch: Branch is defined as a path from one node to another node, in the direction indicated by the branch arrow.


x2
x5
x3
x1
As a summing point
x4
Branch
x6

Node as a summing point:

x_1 = Summing point

$x_1 = x_2+x_3+x_4$

Node as a transmitting (outgoing) point:

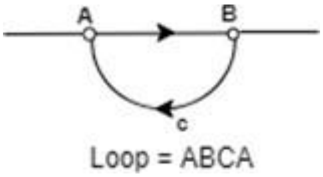
$x_1 = x_5+x_6$

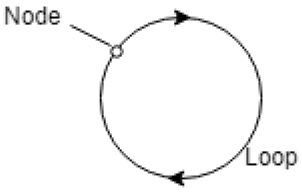
Input node or source: It is the node which have only outgoing branches.

Output node or sink: It is a node which has only incoming branches.

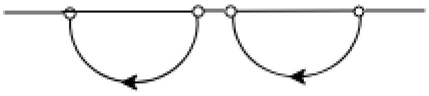
Forward Path: It is a path from an input node to an output node in the direction of branch arrow.

Loop: It is a path that starts and ends at the same node.

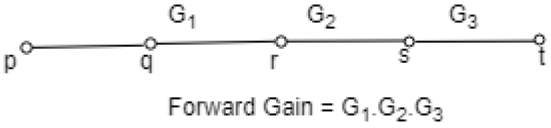




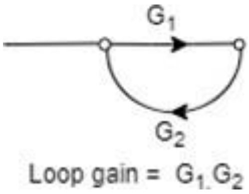
Non-touching loop: Loop is said to be non-touching if they do not have any common node.



Forward path gain: A product of all branches gain along the forward path is called Forward path gain.



Loop Gain: Loop gain is the product of branch gain which travels in the loop.


G1
G2
Loop gain = G1.G2

Construction of SFG and Mason Gain Formula:

The SFG of a system is constructed by the following equations -

Example

Consider a system described by following sets of equations

$x_2 = a_{12}x_1+a_{32}x_3+a_{42}x_4+a_{52}x_5$

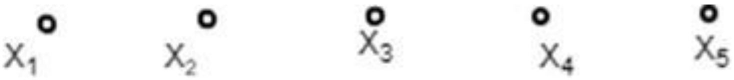
$x_3 = a_{23}x_2$

$x_4 = a_{34}x_3+a_{44}x_1$

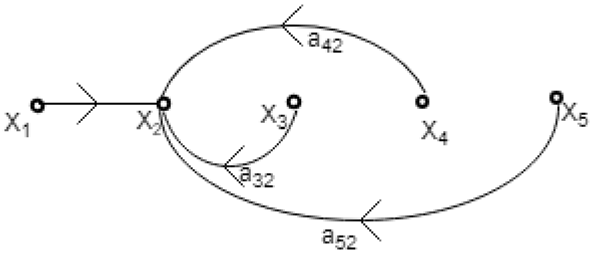
$x_5 = a_{35}x_3+a_{45}x_4$

Where x1 is input and x_5 is output.

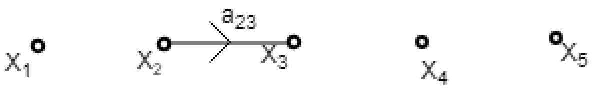
Step1 - First step is to draw all the nodes.


X1
X2
X3
X4
X5

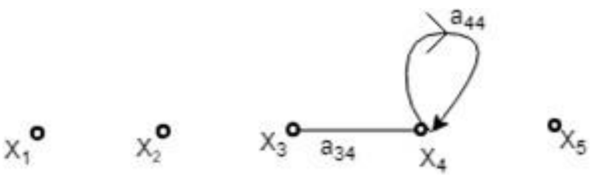
Step2 - Draw the SFG for equation (1)


a42
X1
X2
X3
X4
X5
a32
a52

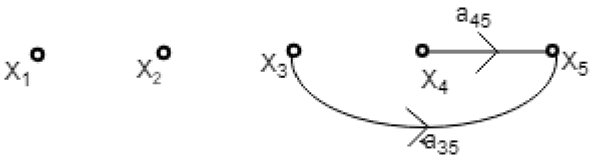
Step3 - Draw the SFG for equation (2)


X_1
X_2
a_{23}
X_3
X_4
X_5

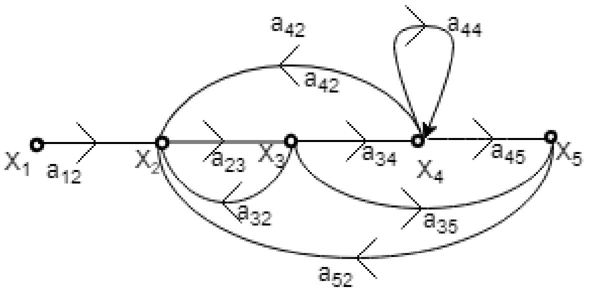
Step4 - Draw the SFG for equation (3)


a_{44}
X_1
X_2
X_3
a_{34}
X_4
X_5

Step5 - Draw the SFG for equation (4)


a_{45}
X_1
X_2
X_3
X_4
X_5
a_{35}

Step6 - Now draw the complete signal flow graph with the help of the above graph.


a_{42}
a_{44}
a_{42}
X_1
a_{12}
X_2
a_{23}
X_3
a_{34}
X_4
a_{45}
X_5
a_{32}
a_{35}
a_{52}

MASON'S GAIN FORMULA

The relation between an input variable and an output variable of a signal flow graph is given by Mason's Gain Formula.

For determination of the overall system, the gain is given by:

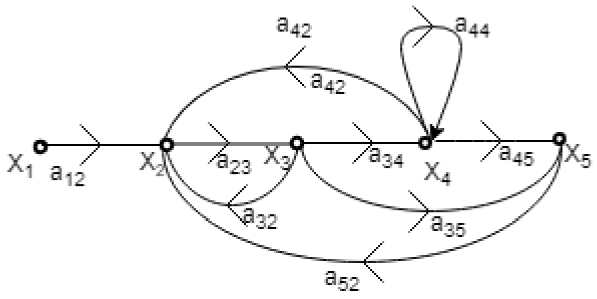
$$T = \frac{\sum_{K-1}^{n} P_K \Delta_K}{\Delta} = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta}$$

Where,

P_k = forward path gain of the K^{th} forward path.

Δ = 1 - [Sum of the loop gain of all individual loops] + [Sum of gain products of all possible of two non-touching loops] + [Sum of gain products of all possible three non-touching loops] + ...

Δ_k = The value of Δ for the path of the graph is the part of the graph that is not touching the K^{th} forward path.



Forward Path

From the above SFG, there are two forward paths with their path gain as -

$$P_1 = a_{12}\, a_{23} a_{34} a_{45}$$

$$p_2 = a_{12} a_{23} a_{35}$$

Loop

There are 5 individual loops in the above SFG with their loop gain as -

$$L_1 = a_{23}a_{32}$$

$$L_2 = a_{23}a_{34}a_{42}$$

$$L_3 = a_{44}$$

$$L_4 = a_{23}a_{34}a_{45}a_{52}$$

$$L_5 = a_{23}a_{35}a_{52}$$

Non-Touching Loops

There are two possible combinations of the non-touching loop with loop gain product as -

$$L_1.L_2 = a_{23}a_{32}a_{44}$$

$$L_5L_3 = a_{23}a_{52}a_{35}a_{44}$$

In above SFG, there are no combinations of three non-touching loops, 4 non-touching loops and so on.

Where,

$$\Delta = 1 - [L_1 + L_2 + L_3 + L_4 + L_5] + [L_1 . L_3 + L_5 . L_3]$$

$$\Delta = 1 - [a_{23}a_{32} + a_{23}a_{34}a_{42} + a_{44} + a_{23}a_{34}a_{45}a_{52} + a_{23}a_{35}a_{52}] + [a_{23}a_{32}a_{44} + a_{23}a_{35}a_{52}a_{44}]$$

$$P_1 = a_{12}\, a_{23}\, a_{34}\, a_{45}$$

$$P_2 = a_{12}\, a_{23}\, a_{35}$$

$$\Delta_1 = 1$$

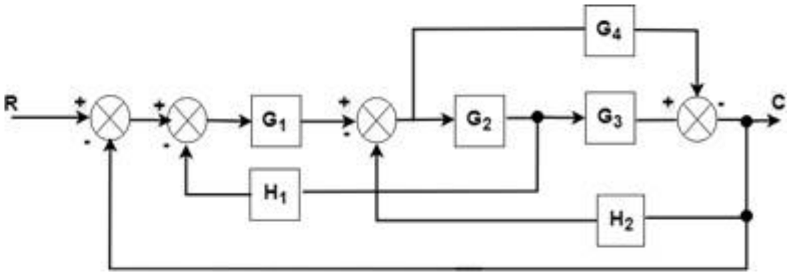
$$\Delta_2 = 1 - a_{44}$$

$$T = \frac{P_1\Delta_1 + P_2\Delta_2}{\Delta}$$

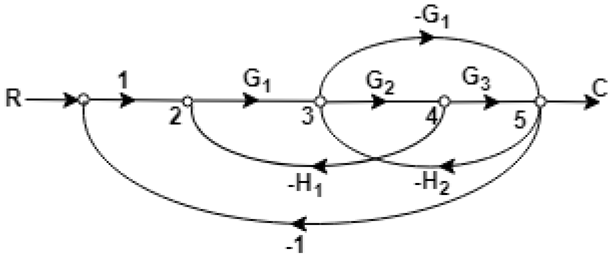
$$T = \frac{a_{12}a_{23}a_{34}a_{45}-1+\ a_{12}a_{23}a_{35}(\,1-a_{44}\,)}{[1-a_{23}a_{32}+a_{23}a_{34}a_{42}+a_{44}a_{23}a_{34}a_{45}a_{52}+a_{23}a_{35}a_{52}+a_{23}a_{32}a_{44}+a_{23}a_{35}a_{52}a_{44}\,]}$$

Example

Draw the Signal Flow Diagram and determine C/R for the block diagram shown in the figure.



The signal flow graph of the above diagram is drawn below



The gain of the forward paths

P_1 = G1G2G3   Δ1 = 1

$P_2 = -G_1G_4 \qquad \Delta_2 = 1$

Individual loops

$L_1 = -\ G_1G_2H_1$

$L_2 = -G_2G_3H_2$

$L_3 = -G_1G_2G_3$

$L_4 = G_1G_4$

$L_5 = G_4H_2$

Non touching Loops = 0

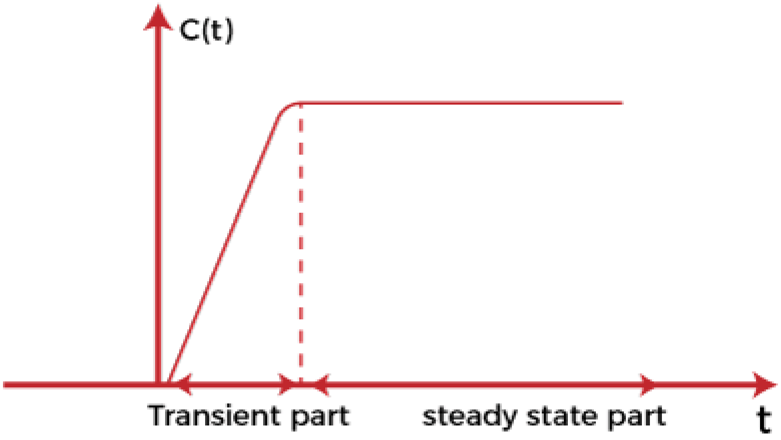
$$T = \frac{C}{R} = \frac{P_1\Delta_1 - P_2\Delta_2}{\Delta}$$

$$\boxed{\frac{C}{R} = \frac{G_1G_2G_3 - G_1G_4}{1 + G_1G_2H_1 + G_2G_3H_2 + G_1G_2G_3 - G_1G_4 - G_4H_2}}$$

Unit-5: Time Domain Analysis of Control Systems

Control system- Time response Analysis

The primary purpose of the time response analysis is to evaluate the system's performance with respect to time. The time-response graph is shown below:

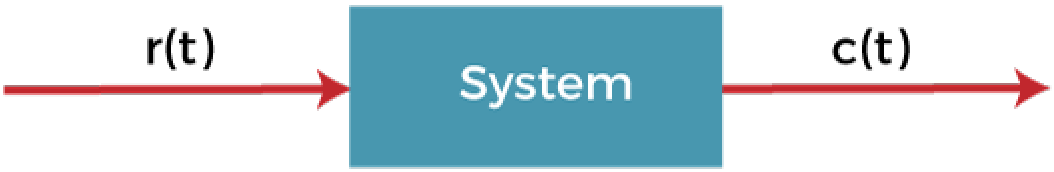


It comprises of two parts, **transient part** and the **steady state part**.

After applying an input to the control system, the output takes some time to reach the steady condition. The response during this stage is known as **transient response** and constitutes the transient part of the graph, as shown above. The graph when achieves the steady start after the transient part is known a **steady part**.

To describe a system, we need to develop the relationship between the inputs and output of the system that are the functions of time. The most common model used to describe such behavior is known as the **differential equation**. The analysis of the system can be done with the help of a differential equation by applying different inputs to it.

The common input signal is shown below:



System with input and output

A test signal r(t) is applied as the input to the system that results in the response c(t). The input signal of a system can take many forms.

Transient Response

It is a part of the time response that reaches **0** (zero) when the time becomes very large. In the graph analysis containing poles and zeroes, the poles lying on the left half of the s-plane gives the transient response. We can also say that it is a part of the response where output continuously increases or decreases. The transient response is also known as the **temporary part** of the response.

Or

Transient response is defined as the change in the response of the system from the equilibrium state.

For example,

The switching time of a bipolar transistor

The characteristics BJT or Bipolar Junction transistor depicts the transient nature.

Steady-state response

The response that comes after the transient response is called the steady-state response. In the graph analysis containing poles and zeroes, the poles on the **imaginary axis** give the steady-state response. We can also say that it is a part of the response where output remains **constant**. The output can also vary periodically with constant amplitude and frequency. The steady-state response is also known as the steady-state part of the response. It is a function of the input signal and hence also known as the **forced response** of the system.

Let's discuss some examples where we will find the transient and steady-state terms of the given equation.

Examples

Example 1: 5 + 2e^-t

Solution:

Here, the transient part of the equation is 2e^-t because as t approaches to infinity, the term becomes 0. Hence, 2e^-t is the transient term. In the case of the first term 5, it will remain same when t approaches infinity. Hence, 5 is the steady-state term of the equation.

Example 2: 10 + 5e^t

Solution:

Here, the first term, 10, is the steady-state term of the equation because it will remain the same when t approaches infinity. In the case of the second term, 5e^t, the result is infinity when t approaches infinity. Hence, it is not a transient term. It is because something to the power infinity is always infinity.

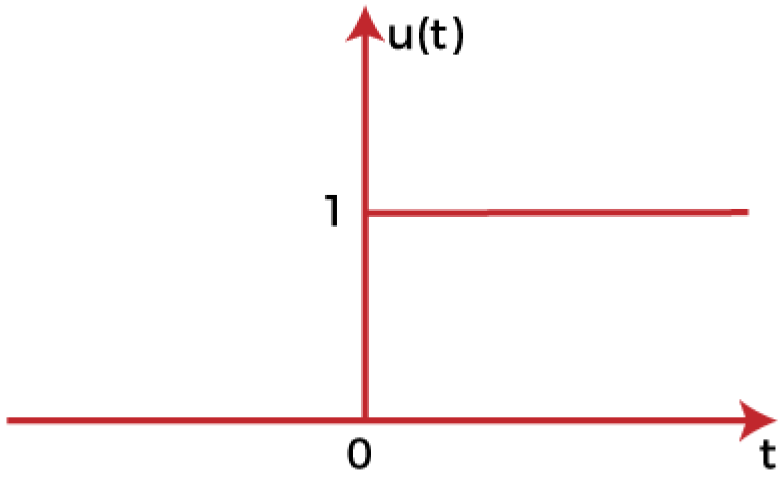
So, there is only a steady-state term in the equation.

Standard signals

1. Step Input signal
2. Ramp input signal
3. Sinusoidal input signal
4. Impulse input signal

Step Input signal

For the positive value, the step input shows constant values of the time. It has zero value for the negative value of the time signal. The initial value of the signal is and the transition is in the form of step size with a constant value. If the constant value of the signal is 1, it is called **step input signal**, which is represented as:



Step signal

The value of the signal is:

0 for t=0 and

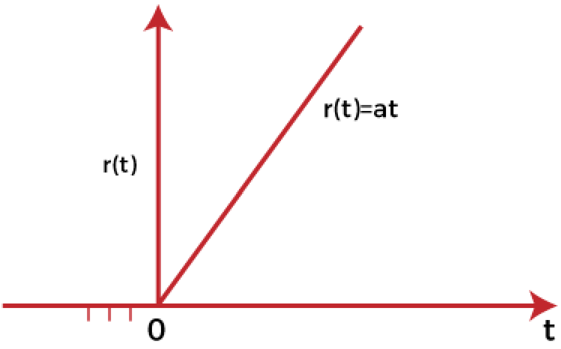
1 for t>0

The graph is a function of one variable named t.

Ramp input signal

The graph of the ramp input signal is in the shape of **ramp**. It depicts the linear increase that begins at some specific point. The value of the ramp signal shows the constant change with respect to time. The value of the signal is 0 for negative values. It means that it shows the output for positive inputs.

The ramp function is represented as:



Ramp Input

The value of the signal is:

At for t>0 and

0 for t<0

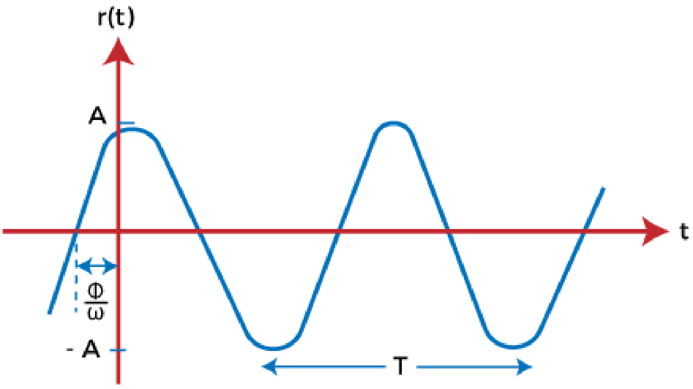
If the value of A is 1 when t>0. The signal is known as unit ramp signal.

Sinusoidal input signal

A sinusoidal input is one whose oscillations can be described by an equation in the form of **sine**. The response of the linear process to a sinusoidal is sinusoidal. The signal is given by:

$$r(t) = A\sin(\omega t + \emptyset)$$

The sinusoidal signal in the control system is represented as:



Sinusoidal Signal

The sine wave starts from zero, covers positive value, reach zero, covers negative values, and again reaches zero, as shown above.

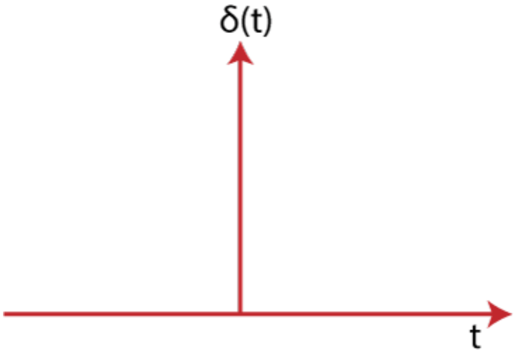
Impulse Input signal

The impulse signal is a type of high amplitude signal and has a very short duration. It means that the magnitude approaches ***infinity*** when the time reaches zero. Thus, we can say that the value of the signal is infinity at t =0. Otherwise, its value is 0.

$$\int_{-\infty}^{\infty} \delta(t)dt = 1$$

Its integration from -infinity to infinity is 1, as shown above.

It is a physical non-existing signal, which is defined based on the area concept. It is not based on the amplitude concept. The impulse input signal is represented as:



Transient and Steady State Analysis of Linear Time Invariant (LTI) Systems

Time Response Analysis

When the energy state of any system is disturbed, and the disturbances occur at input, output or both ends, then it takes some time to change from one state to another state. This time that is required to change from one state to another state is known as **transient time** and the value of current and voltage during this period is called **transient response**.

Depending upon the parameters of the system, the transient may have oscillations which may be either sustained or decaying in nature.

Thus time response of a control system is divided into two parts-

1. Transient response analysis.
2. Steady State Analysis.

Transient State response

It deals with the nature of the response of a system when subjected to an input.

Steady State Analysis

It deals with the estimation of the magnitude of steady-state error between input and output.

Different Type of Standard Test Signals

The various inputs or disturbances affecting the performance of a system are mathematically represented as a standard test signal.

- Step signal (sudden input)
- Ramp Signal (velocity type of input)
- Parabolic Signal (type of acceleration input)
- Impulse signal (sudden shock)

NOTE

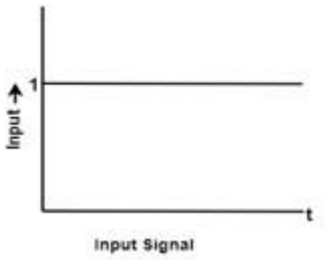
- Step signal and impulse signal are bounded input signal.
- Ramp signal and parabolic signal are an unbounded input signal.
- Step signal, a ramp signal, and periodic signal are for time domain analysis. Only an impulse signal is essential for steady-state analysis.

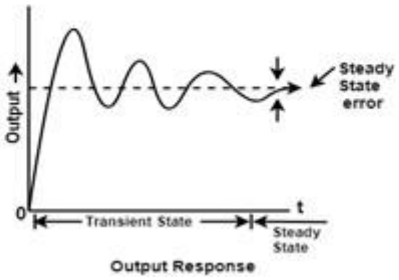
Characteristics of Time-domain Analysis

- Every transfer function representing the control system is of a particular type of order.
- The steady state analysis depends upon the type of the system.
- The type of the system is determined from open loop transfer function G(S).H(S)

Transient Time: The time required to change from one state to another is called the transient time.

Transient Response: The value of current and voltage during the time change is called transient response.



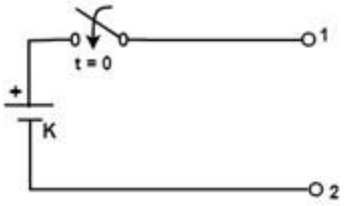


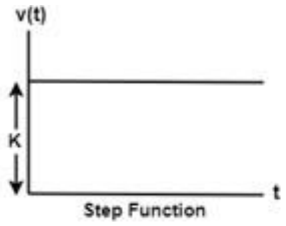
So, we can say that the transient response is the part of the response which goes to zero as time increases and the steady-state response is the part of the total response after transient has died. If the steady-state response is the part of the output does not match with the input then the system has a steady state error.

Test input signal for transient analysis

For the analysis of the time response of a control system, the following input signals are used.

Step Function





A unit step function is denoted by u(t) and is defined as

u(t) = 0 ; t=0 and
= 1; t>0

Laplace Transform:

$$£f(t) = \int_0^\infty f(t)e^{-st} = \int_0^\infty u(t)e^{-st}\,dt$$

$$= \int_0^\infty 1e^{-st}\,dt = \frac{1}{s}$$

Step function is also called displacement function. If input is R(S), then R(s) = 1/s

Ramp Function

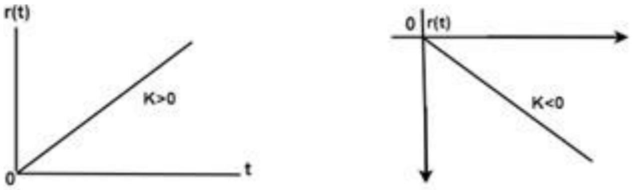


Fig: Ramp Function

This function starts from the origin and linearly decreases or increases with time as shown in the figure above.

Let r(t) be the ramp function then

r(t) = 0 ; t<0
= Kt ; t>0

Where 'K' is the slope of the line, for a positive value of 'K' the slope is upward, and the slope is downward for the negative value of 'K.'

Laplace transform

$$\pounds r(t) = \int_0^{\infty} r(t)e^{-st}\,dt = \int_0^{\infty} Kte^{-st}\,dt = \frac{K}{s^2}$$

$$R(S) = \frac{K}{s^2}$$

Parabolic Function

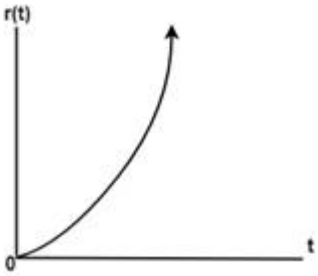


Fig: Parabolic function

The value of r(t) is zero when t<0 and is a quadratic function of time when t>0.

Therefore, $r(t) = 0 \quad ; t<0$

$= (Kt^2)/2 \quad ; \; t>0$

Where 'K' is constant for unit parabolic function K = 1. The unit parabolic function is defined as

$r(t) = 0 \quad ; \; t<0$

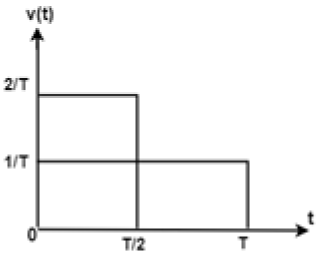
$= t^2/2 \quad ; t>0$

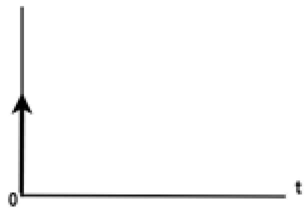
Laplace Transform

$$\pounds r(t) = \int_0^{\infty} r(t)e^{-st}\, dt \; = \; \int_0^{\infty} Kte^{-st}\, dt \; = \; \frac{K}{s^2}$$

$$R(S) = \frac{K}{s^2}$$

Impulse Function





A unit impulse function is defined as

$$\delta(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases} \qquad \int_{-\infty}^{\infty} \delta(t)dt = 1$$

Thus we can say that impulse function has zero value everywhere except at t=0 where the amplitude is infinite.

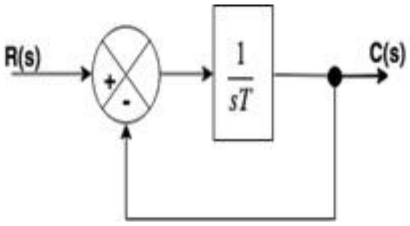
$\delta(t) = u(t)$

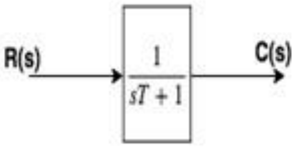
$£\,\delta(t) = £\frac{d}{dt}[u(t)]$

$= s£(\text{unit step function})$

$= s.\frac{1}{s} = 1$

Time Response of first order system





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

$$\frac{\frac{1}{ST}}{1 + \frac{1}{ST} \times 1}$$

$$Transfer\ function = \frac{C(s)}{R(s)} = \frac{1}{1 + ST}$$

In the above transfer function, the power of 's' is the one in the denominator. That is why the above transfer function is of the first order, and the system is said to be the **first order system**.

Response of 1st order system when the input is unit step -

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

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

For Unit Step,

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

$$C(s) = \frac{1}{(ST+1)}.\frac{1}{S}$$

$$C(s) = \frac{1}{S} - \frac{1}{(ST+1)}$$

Now, the partial fraction of above equation will be:

$$C(s) = \frac{1}{S} - \frac{1}{(S+\frac{1}{T})}$$

Taking the inverse Laplace of above equation is:

$$u(t) - e^{\frac{-t}{T}}$$

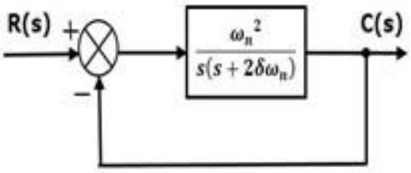
$1 - e^{\frac{-t}{T}}$ (t=T)

$$1 - e^{\frac{-T}{T}}$$

$$1 - e^{-1} = 1 - 0.368 = 0.632$$

Where T is known as time constant of the system and it is defined as the time required for the signal to attain 63.2 % of final or steady state value. Time constant means how fast the system reaches the final value. As smaller the time constant, as faster is the system response. If time constant is larger, system goes to move slowly.

Time response of second order system



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

$$G(S) = \frac{\omega_n^2}{s(s+2\xi\omega_n)} \quad , \qquad \text{H(S) =1}$$

$$\frac{C(S)}{R(S)} = \frac{\omega_n^2/s(s+2\xi\omega_n)}{1+\frac{\omega_n^2}{s(s+2\xi\omega_n)}}$$

$$\frac{C(S)}{R(S)} = \frac{\omega_n^2}{s^2+2\xi\omega_n s+\omega_n^2} \qquad \text{...........eq.1}$$

In the above transfer function, the power of 's' is two in the denominator. That is why the above transfer function is of a second order, and the system is said to be the **second order system**.

Time response of second order system with unit step

From equation 1

$$\text{C (S)} = \frac{\omega_n^2}{S^2+2\xi\omega_n S+\omega_n^2}\,.\,R(S)$$

For unit step the input is

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

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

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

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

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

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

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

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

Now putting, $\omega_d = \ \omega_n\sqrt{1 - \ \xi^2}$

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

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

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

Now taking the inverse Laplace of above equation

$$£^{-1}[C(s)] = £^{-1}\left[\frac{1}{s} - \frac{s + \xi\omega_n}{(s + \xi\omega_n)^2 + \omega_d{}^2} - \frac{\xi\omega_n}{\omega_d} \cdot \frac{\omega_d}{(s + \xi\omega_n)^2 + \omega_d{}^2}\right]$$

$$£^{-1}\left[\frac{1}{s}\right] - \ £^{-1}\left[\frac{s + \xi\omega_n}{(s + \xi\omega_n)^2 + \omega_d{}^2}\right] - £^{-1}\left[\frac{\xi\omega_n}{\omega_d} \cdot \frac{\omega_d}{(s + \xi\omega_n)^2 + \omega_d{}^2}\right]$$

$$\therefore \ c(t) = 1 - e^{-\xi\omega_n t} . \cos\omega_d t - \ \frac{\xi\omega_n}{\omega_d} \ . \, e^{-\xi\omega_n t} . \sin\omega_d t$$

$\because £^{-1}\left[\frac{1}{s}\right] = 1, £^{-1}\left[\frac{s+\alpha}{(s+\alpha)^2+ \omega^2}\right] = \ e^{-\alpha t}\cos\omega t,$

$$£^{-1}\left[\frac{\omega}{(s + \alpha)^2 + \ \omega^2}\right] = e^{-\alpha t} \sin\omega t$$

This equation can also be written as

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

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

$$[\text{ Let, } \xi = \cos\emptyset, hence, \sqrt{1-\xi^2} = \sin\emptyset\ ]$$

$$\therefore\quad c(t) = 1 - \frac{e^{-\xi\omega_n t}}{\sqrt{1-\xi^2}}(\sin\emptyset\cos\omega_d t + \cos\emptyset\sin\omega_d t)$$

$$= 1 - \frac{e^{-\xi\omega_n t}}{\sqrt{1-\xi^2}}\sin(\omega_d t + \emptyset)$$

The error signal is given by e (t) = r (t) –c (t), and hence

$$e(t) = \frac{e^{-\xi\omega_n t}}{\sqrt{1-\xi^2}}\sin(\omega_d t + \emptyset)$$

Transient response specification of second order system

The performance of the control system are expressed in terms of transient response to a unit step input because it is easy to generate initial condition basically are zero.

Following are the common transient response characteristics:

1. Delay Time.
2. Rise Time.
3. Peak Time.
4. Maximum Peak.
5. Settling Time.
6. Steady State error.

Delay Time

The time required for the response to reach 50% of the final value in the first time is called the delay time.

Rise Time

The time required for response to rising from 10% to 90% of final value, for an overdamped system and 0 to 100% for an underdamped system is called the rise time of the system.

Peak Time

The time required for the response to reach the 1st peak of the time response or 1st peak overshoot is called the Peak time.

Maximum overshoot

The difference between the peak of 1st time and steady output is called the maximum overshoot. It is defined by

$$\text{Maximum percent overshoot} = \frac{C(t_p)-C(\infty)}{C(\infty)}$$

$$\%M_P = e^{\frac{\pi\xi}{\sqrt{1-\xi^2}}} \times 100$$

Settling Time (t_s)

The time that is required for the response to reach and stay within the specified range (2% to 5%) of its final value is called the settling time.

$$t_s = \frac{4}{\xi\omega_n}$$

Steady State Error (e_{ss})

The difference between actual output and desired output as time't' tends to infinity is called the steady state error of the system.

$$e_{ss} = \lim_{t\to\infty}[\,r(t) - C(t)]$$

Unit-6: Feedback Characteristics of Control Systems

Controllers

The various types of controllers are used to improve the performance of control systems. In this chapter, we will discuss the basic controllers such as the proportional, the derivative and the integral controllers.

Proportional Controller

The proportional controller produces an output, which is proportional to error signal.

u(t) a e(t)

$\Rightarrow u(t) = K_P^{e(t)}$

Apply Laplace transform on both the sides -

$U(s) = K_P^{E(s)}$

$U(s)/E(s) = K_P$

Therefore, the transfer function of the proportional controller is K_P.

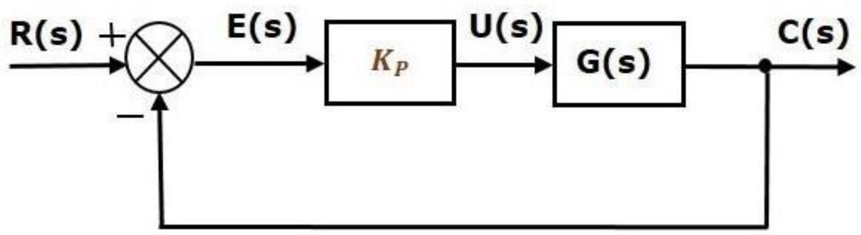
Where,

U(s) is the Laplace transform of the actuating signal u(t)

E(s) is the Laplace transform of the error signal e(t)

K_P is the proportionality constant

The block diagram of the unity negative feedback closed loop control system along with the proportional controller is shown in the following figure.



The proportional controller is used to change the transient response as per the requirement.

Derivative Controller

The derivative controller produces an output, which is derivative of the error signal.

$u(t)=K_D\,[de(t)/\,dt\,]$

Apply Laplace transform on both sides.

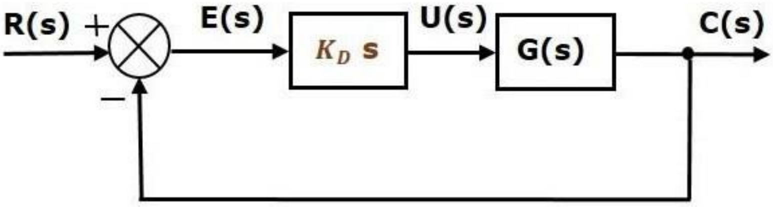
$U(s) = K_D\, sE(s)$

$U(s)/E(s)=K_D s$

Therefore, the transfer function of the derivative controller is K_D^s.
Where, K_D is the derivative constant.

Therefore, the transfer function of the derivative controller is K_D^s.
Where, K_D is the derivative constant.

The block diagram of the unity negative feedback closed loop control system along with the derivative controller is shown in the following figure.



The derivative controller is used to make the unstable control system into a stable one.

Integral Controller

The integral controller produces an output, which is integral of the error signal.

$u(t)= K_I \int e(t)\, dt$

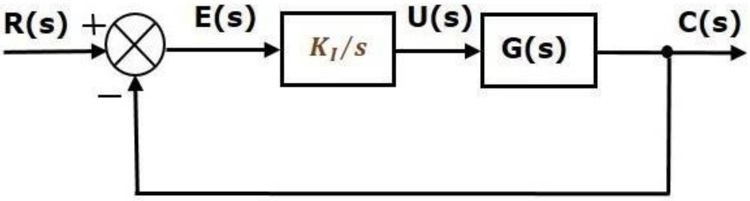
Apply Laplace transform on both the sides -

$U(s)= [K_I\, E(s)]/s$

$U(s) / E(s)=[K_I]/s$

Therefore, the transfer function of the integral controller is K_I/s.
Where, K_I is the integral constant.

The block diagram of the unity negative feedback closed loop control system along with the integral controller is shown in the following figure.



The integral controller is used to decrease the steady state error.

Let us now discuss about the combination of basic controllers.

Proportional Derivative (PD) Controller

The proportional derivative controller produces an output, which is the combination of the outputs of proportional and derivative controllers.

$$u(t) = K_P^{e(t)} + K_D\,[de(t) / dt]$$

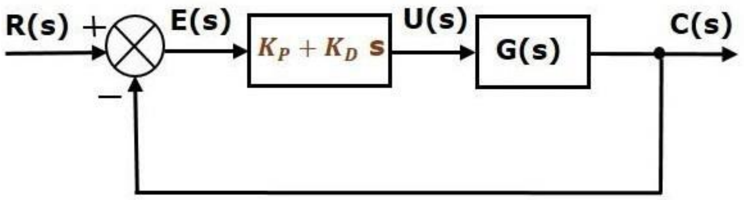
Apply Laplace transform on both sides -

$$U(s) = (K_P + K_D^{s})E(s)$$

$$U(s)/E(s) = K_P + K_D^{s}$$

Therefore, the transfer function of the proportional derivative controller is $K_P + K_D^{s}$

The block diagram of the unity negative feedback closed loop control system along with the proportional derivative controller is shown in the following figure.



The proportional derivative controller is used to improve the stability of control system without affecting the steady state error.

Proportional Integral (PI) Controller

The proportional integral controller produces an output, which is the combination of outputs of the proportional and integral controllers.

$$u(t) = K_P^{e(t)} + K_I \int e(t)\, dt$$

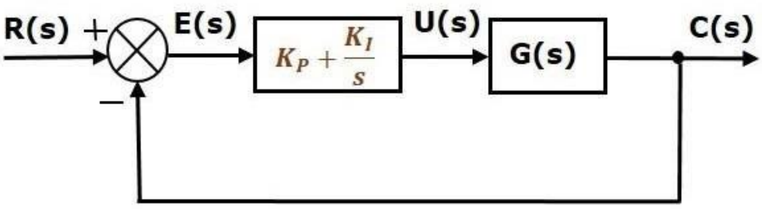
Apply Laplace transform on both sides -

$$U(s) = (K_P + K_I/s)\, E(s)$$

$$U(s)/E(s) = K_P + K_I/s$$

Therefore, the transfer function of proportional integral controller is $K_P + K_I/s$.

The block diagram of the unity negative feedback closed loop control system along with the proportional integral controller is shown in the following figure.



The proportional integral controller is used to decrease the steady state error without affecting the stability of the control system.

Proportional Integral Derivative (PID) Controller

The proportional integral derivative controller produces an output, which is the combination of the outputs of proportional, integral and derivative controllers.

$$u(t)=K_P e(t)+K_I\int e(t)dt+K_D\, de(t)/dt$$

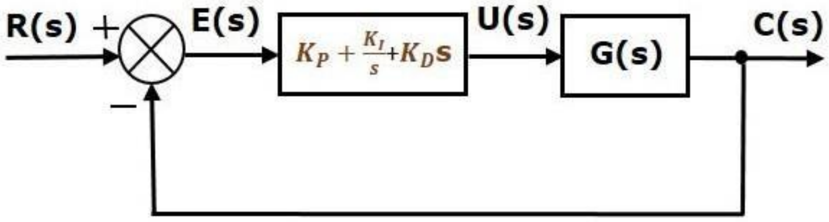
Apply Laplace transform on both sides -

$$U(s) = (K_P+K_I/s+K_D s)\, E(s)$$

$$U(s)/E(s) = K_P+K_I/s+K_D s$$

Therefore, the transfer function of the proportional integral derivative controller is $K_P+K_I/s+K_D s$

The block diagram of the unity negative feedback closed loop control system along with the proportional integral derivative controller is shown in the following figure.



The proportional integral derivative controller is used to improve the stability of the control system and to decrease steady state error.

Unit-7: Stability Concept & Root Locus Method

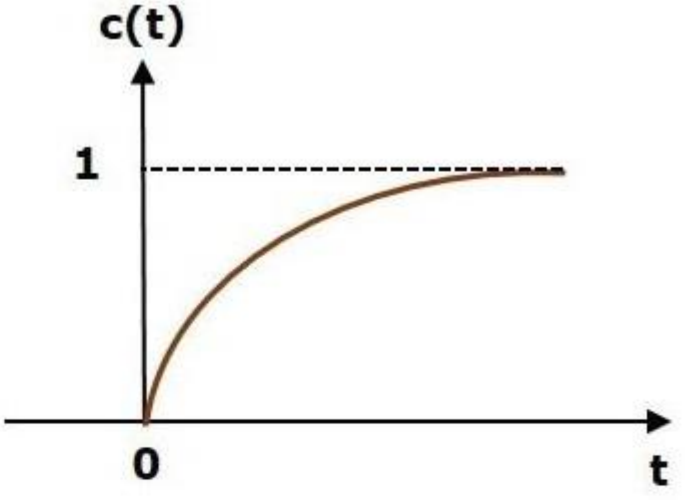
Stability

Stability is an important concept. In this chapter, let us discuss the stability of system and types of systems based on stability.

What is Stability?

A system is said to be stable, if its output is under control. Otherwise, it is said to be unstable. A **stable system** produces a bounded output for a given bounded input.

The following figure shows the response of a stable system.



This is the response of first order control system for unit step input. This response has the values between 0 and 1. So, it is bounded output. We know that the unit step signal has the value of one for all positive values of **t** including zero. So, it is bounded input. Therefore, the first order control system is stable since both the input and the output are bounded.

Types of Systems based on Stability

We can classify the systems based on stability as follows.

- Absolutely stable system
- Conditionally stable system
- Marginally stable system

Routh- Hurwitz Criterion

Before discussing the Routh-Hurwitz Criterion, firstly we will study the stable, unstable and marginally stable system.

1. **Stable System**: If all the roots of the characteristic equation lie on the **left** half of the 'S' plane then the system is said to be a stable system.
2. **Marginally Stable System**: If all the roots of the system lie on the imaginary axis of the 'S' plane then the system is said to be marginally stable.
3. **Unstable System**: If all the roots of the system lie on the **right** half of the 'S' plane then the system is said to be an unstable system.

Statement of Routh-Hurwitz Criterion

Routh Hurwitz criterion states that any system can be stable if and only if all the roots of the first column have the same sign and if it does not has the same sign or there is a sign change then the number of sign changes in the first column is equal to the number of roots of the characteristic equation in the right half of the s-plane i.e. equals to the number of roots with positive real parts.

Necessary but not sufficient conditions for Stability

We have to follow some conditions to make any system stable, or we can say that there are some necessary conditions to make the system stable.

Consider a system with characteristic equation:

$$a_0 s^m + a_1 s^{m-1} + \ldots\ldots\ldots\ldots + a_m = 0$$

1. All the coefficients of the equation should have the same sign.
2. There should be no missing term.

If all the coefficients have the same sign and there are no missing terms, we have no guarantee that the system will be stable. For this, we use **Routh Hurwitz Criterion** to check the stability of the system. If the above-given conditions are not satisfied, then the system is said to be unstable. This criterion is given by A. Hurwitz and E.J. Routh.

Advantages of Routh- Hurwitz Criterion

1. We can find the stability of the system without solving the equation.
2. We can easily determine the relative stability of the system.
3. By this method, we can determine the range of K for stability.
4. By this method, we can also determine the point of intersection for root locus with an imaginary axis.

Limitations of Routh- Hurwitz Criterion

1. This criterion is applicable only for a linear system.
2. It does not provide the exact location of poles on the right and left half of the S plane.
3. In case of the characteristic equation, it is valid only for real coefficients.

The Routh- Hurwitz Criterion

Consider the following characteristic Polynomial

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

When the coefficients a_0, a_1 ...a_n are all of the same sign, and none is zero.

Step 1: Arrange all the coefficients of the above equation in two rows:

Row 1	a_0	a_2	a_4	…………
Row 2	a_1	a_3	a_5	…………

Step 2: From these two rows we will form the third row:

Row 1	a_0	a_2	a_4	…………
Row 2	a_1	a_3	a_5	…………
Row 3	b_1	b_3	b_5	……….

Where,

$$b_1 = -\frac{1}{a_1}\begin{vmatrix} a_0 & a_2 \\ a_1 & a_3 \end{vmatrix} = -\frac{a_0 a_3 - a_1 a_2}{a_1}$$

$$b_3 = -\frac{1}{a_1}\begin{vmatrix} a_0 & a_4 \\ a_1 & a_5 \end{vmatrix} = -\frac{a_0 a_5 - a_1 a_5}{a_1}$$

Step 3: Now, we shall form fourth row by using second and third row:

Row 1	a_0	a_2	a_4	………..
Row 2	a_1	a_3	a_5	………..
Row 3	b_1	b_3	b_5	……….
Row 4	c_1	c_3	c_5	……….

Where,

$$c_1 = -\frac{1}{b_1}\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} = -\frac{a_1 b_3 - b_1 a_3}{b_1}$$

$$c_3 = -\frac{1}{b_1}\begin{vmatrix} a_1 & a_5 \\ b_1 & b_5 \end{vmatrix} = -\frac{a_1 b_5 - b_1 a_5}{b_1}$$

Step 4: We shall continue this procedure of forming a new rows:

Example

Check the stability of the system whose characteristic equation is given by

$s^4 + 2s^3 + 6s^2 + 4s + 1 = 0$

Solution

Obtain the arrow of coefficients as follows

s^4	1	6	1
s^3	2	4	
s^2	4	1	
s^1	3.5		
s^0			

$$b_1 = -\frac{1}{2}\begin{vmatrix} 1 & 6 \\ 2 & 4 \end{vmatrix},$$

$$b_1 = \frac{-(1 \times 4 - 2 \times 6)}{2} = \frac{-(-8)}{2} = 4$$

$$c_1 = -\frac{1}{4}\begin{vmatrix} 2 & 4 \\ 4 & 1 \end{vmatrix} = 3.5$$

$$b_2 = -\frac{1}{2}\begin{vmatrix} 1 & 1 \\ 2 & 0 \end{vmatrix} = 1$$

$$d_1 = -\frac{1}{3.5}\begin{vmatrix} 4 & 1 \\ 3.5 & 0 \end{vmatrix} = 1$$

Since all the coefficients in the first column are of the same sign, i.e., positive, the given equation has no roots with positive real parts; therefore, the system is said to be stable.

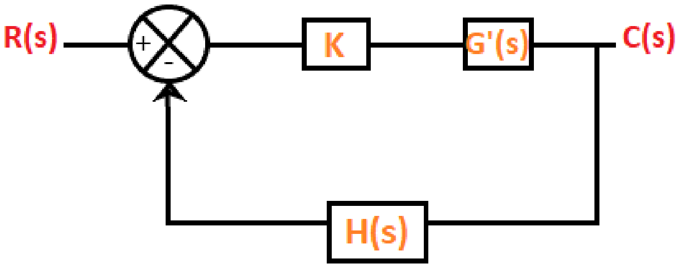
Basic concepts of root locus:

In the previous sections, we have studies that the stability of a system. It depends on the **location of the roots** of the characteristic equation. We can also say that the stability of the system depends on the location of **closed-loop poles.** Such knowledge of the movement of the poles in the s-plane when the parameters are varied is important. The minor changes in the parameters can greatly help in the system designing. The nature of the system's transient response is closely related to the location of the poles in the s-plane.

We have also studied the Routh Hurwitz criteria that describe the stability of the algebraic equation. If any of the term in the first column of the Roth table possesses a sign change, the system tends to become unstable.

The root locus method was introduced by **W.R Evans** in 1948. Root locus is a graphical method in which the movement of poles in the s-plane can be located when a specific parameter is varied from 0 to infinity. The parameter assumed to be varied is generally the **gain** of the system.

Consider the below closed loop system.



The equation of a closed loop system is given by:

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

Where,

G(s) is the gain of the transfer function

H(s) is the feedback gain

In the case of root locus, the gain K is also assumed as part of the closed-loop system. K is known as system gain or the gain in the forward path. The characteristic equation after including the forward gain can be represented as:

$1 + KG'(s)H(s) = 0$

Where,

$G(s) = KG'(s)$

When the system includes the variable parameter K, the roots of the closed loop system are now dependent on the values of 'K.'

The value of 'K' variable can vary in two cases, as shown below:

1. -∞ to +∞
2. 0 to +∞

In the first case, for every different value (integer or decimal) of K, we will get separate set of locations of the roots. If all such locations are joined, the resulting plot is defined as the root locus. We can also define root locus as the locus of the closed loop poles obtained when the system gain 'K' is varied from **-infinity** to **infinity.**

When the K varies from **zero to infinity,** the plot is called the direct root locus. If the system gain 'K' varies from -infinity to zero, the plot thus obtained is known as inverse root locus. The gain K is generally assumed from zero to infinity unless specially stated.

Rules of root locus

Here, we will discuss the six basic rules required to plot a root locus. We will also discuss an example that will help us to implement the rules easily.

Rule Number 1

We know that the root of the equation can be real or complex or a combination of both. The root locus is generally symmetric about the real axis. Thus, the plot needs to be symmetric about the real axis of the s-plane.

Rule Number 2

The transfer function of the system is generally represented by G(s)H(s), where H(s) is the feedback path. Let us suppose the open-loop transfer function to be the same G(s)H(s) and the poles and zeroes are P and Z.

There are two conditions where the poles can be greater than the number of zeroes, or the number of zeroes can be greater than the number of poles in the given characteristic equation.

Let the number of branches in the root locus is N. Both the cases arise when we plot the root locus. The default conditions are given for each case that helps in determining the number of branches terminating or approaching infinity.

Case 1: P > Z

In the above case, we assume that the number of branches in the root locus will equal the number of open-loop poles. It is because the numbers of poles are greater here.

(N = P)

Branches, in this case, will start from the location of the open-loop poles. Here, out of a number of branches at P, the Z number of branches will terminate at the location of open-loop zeroes. The remaining branches (P - Z) will approach infinity.

For example,

Let P = 3, and Z = 1

Then,

The number of root locus branches = 3 = no. of poles

P - Z = 3 - 1 = 2

It means that 3 branches will start from the location of open loop poles

No. of branches terminating at the open loop zero location = 1

No. of branches approaching to infinity = P - Z = 2

Case 2: Z > P

In the above case, we assume that the number of branches in the root locus will be equal to the number of open loop zeroes. It is because the numbers of zeroes are greater here. (N = Z)

Branches in this case will terminate at the finite location of the open loop zeroes. Here, out of number of branches at Z, the P number of branches will start at the location of open loop poles. The remaining branches (Z - P) will approach to finite zeroes originating from the infinity.

For example,

Let P = 1, and Z = 3

Then,

The number of root locus branches = 3 = no. of zeroes

Z - P = 3 - 1 = 2

It means that 1 branch will start from the location of open loop poles

No. of branches terminating at the finite open loop pole location = 3 = all the root locus branches

No. of branches originating from infinity = Z - P = 2

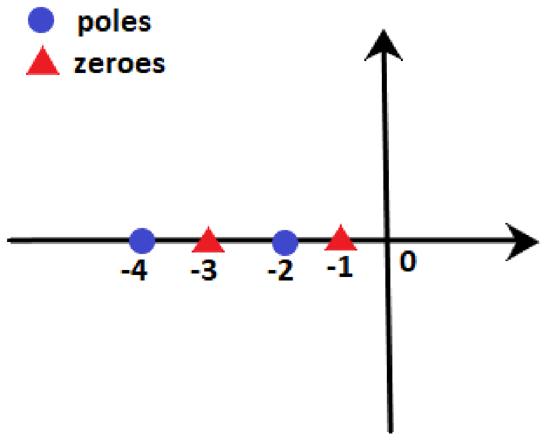
Rule Number 3

A point on the root locus is said to exist if the sum of the open loop poles and zeroes on the real axis towards the right hand side is **odd** with respect to that point.

For example,

Let poles be -2 and -4, and zeroes are -1 and -3. We need to find that points -2.2 and -3.4 lies on the root locus or not.

We know that -2.2 lies between -2 and -3, while point -3.4 lies between -3 and -4.

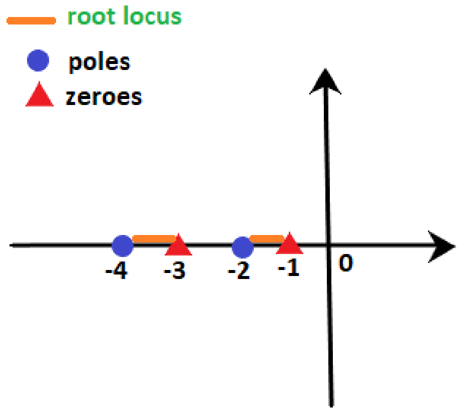


We know that -2.2 lie between -2 and -3, while point -3.4 lies between -3 and -4.

Point -2.2: At point -2.2, the sum of poles and zeroes on the right-hand side is 2, i.e. 1 pole and 1 zero. It means that the sum is even. According to rule number 3, the sum should be odd. Hence, point -2.2 does not lie on the root locus. We can also say that any point between -2 and -3 will not lie on the root locus.

Point -3.4: At point -3.4, the sum of poles and zeroes on the right-hand side is 3, i.e. 1 pole and 2 zeroes. It means that the sum is odd. According to rule number 3, the sum should be odd. Hence, point -3.4 lie on the root locus. We can also say that any point between -3 and -4 will lie on the root locus.

The area of the root locus between the points is shown below:



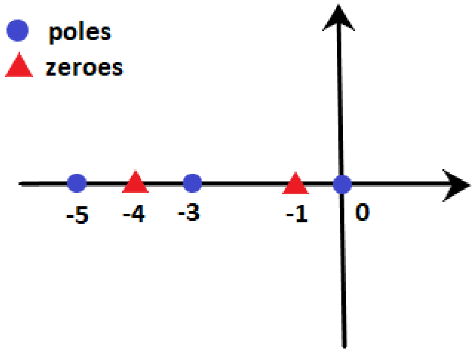
Here, orange line represents the area where root locus lies.

Let's consider an example for better understanding.

Example: G(s)H(s) = K(s + 1)(s + 4)/ s(s + 3)(s + 5). Find on which sections of the real axis, the root locus exists.

Solution: We know that the denominator signifies the poles, and the numerator signifies the zeroes. Thus, 0, -3, and -5 are the poles, and -1 and -4 are the zeroes per the given transfer function. It means that there are 3 poles and 2 zeroes.

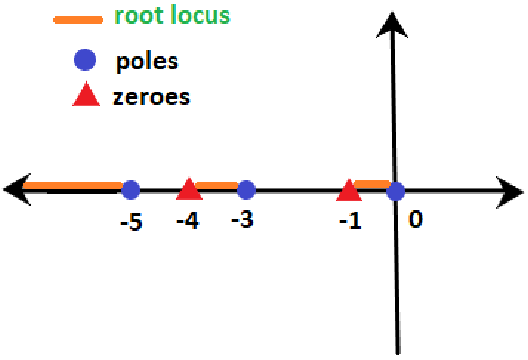
These poles and zeroes on the real axis will appear as:



According to rule number 3,

- The sections between **0 and -1** (for example, point -0.4) contain only one pole and no zeroes on the right-hand side. It means that the sum is odd (i.e. 1). So, it exists on the root locus.
- The sections between **-1 and -3** (for example, point -2.1) contain only one pole and one zero on the right-hand side. It means that the sum of poles and zeroes is even (i.e. 2). So, it does not exist on the root locus.
- The sections between **-3 and -4** (for example, point -3.5) contain two poles and one zero on the right-hand side. It means that the sum is odd (i.e. 3). So, it exists on the root locus.
- The sections between **-4 and -3** (for example, point -4.3) contain two poles and two zeroes on the right-hand side. It means that the sum of poles and zeroes is even (i.e. 4). So, it does not exist on the root locus.
- The sections greater than **-5** (for example, point -8.6) contain three poles and two zeroes on the right-hand side. It means that the sum is odd (i.e. 5). So, it exists on the root locus.

Thus, the line marked with orange depicts the sections where the root locus exists. It is shown below:



Rule Number 4

We have already discussed that (P - Z) provides the number of branches approaching infinity for the given transfer function. The information about such branches approaching infinity is defined under rule number 4, known as **asymptotes.** The angle of such asymptotes is given by:

= (2q + 1)180 / P - Z

Where,

q = 0, 1, 2, 3, 4 ... (P - Z - 1)

These are always symmetric about the real axis.

Rule Number 5

Rule number 4 describes the guidelines or information about the branches approaching infinity, known as the asymptotes. But, the angles are insufficient to plot the root locus, and the location of such branches in the s-plane is equally important, defined by rule number 5.

Centroid is a point where the asymptotes intersect at a common point on the real axis. It can be calculated as:

$$\sigma = \frac{\sum \text{Real part of poles of } G(s)H(s) - \sum \text{Real part of zeroes of } G(s)H(s)}{P - Z}$$

Note: The value of the centroid is always real, which can be positive or negative. It can be a part of root locus or sometimes not.

Rule Number 6

The last rule is the breakaway point. It is also a point on the root locus where multiple roots of the given equation occurs. It is calculated for a specific value of system gain K.

Or

It can be defined as a point on the root locus where two or more roots occur for a particular value of K.

The root locus branches always leave breakaway points at an angle of 180/n.

Where,

N = number of branches approaching at the breakaway point.

The value of the angle can be positive or negative.

Let's discuss some predictions about the existence of the breakaway points:

- There exists atleast one breakaway point between the adjacent placed poles if the section between the two poles lies on the root locus. For example,

  G(s)H(s) = K/s(s + 3)

The above transfer function has two poles at 0 and -3. According to rule number 3, the point on the section between 0 and -3 (for example, point -2.2) has one pole and no zero on the right-hand side. It signifies that the sum of zeroes and poles is 1, i.e. odd. Thus, the section between 0 and -3 exists of the root locus.

Hence, there must exist a minimum of one breakaway point in between them.

Unit-8: Frequency-Response Analysis & Bode Plot

Basic concepts of frequency response

The frequency of the input signal is varied over a specific range, and the system's output is studied. The change in the system's output response with respect to the varied input is known as the system's frequency response.

The frequency response is represented as T(JW), and it comprises of phase function and the magnitude function. They are also known as the system's frequency response, which can be evaluated both for the open-loop and closed-loop systems.

The open-loop transfer function is given by:

$G(s) = G(jw)$ Where, $s = jw$

$|G(s)|$ or $|G(jw)|$represents the magnitude of the transfer function.

$\angle G(jw)$ represents the phase of the transfer function.

Here, we will consider the response of the system with the input at different frequencies. The frequency response is a steady state response of the system to a sinusoidal input signal. For example, if a system has sinusoidal input, the output will also be sinusoidal. The changes can occur in the magnitude and the phase shift.

Let $G(s) = 1/(Ts + 1)$

It is the transfer function in the time-constant form. We are assuming all the parameters in the steady state.

$G(jw) = 1/(Tjw + 1)$

Let the input be Asin(wt).

$|G(jw)| = K/(1 + w^2T^2)^{1/2}$

It represents the magnitude.

The angle G(s) is given by:

$\tan^{-1}(Tw)$

The above analysis tells that the steady state output (Css) when the input was Asinwt is given by:

$Css(t) = A.K.\sin(wt - \tan^{-1}(Tw))/(1 + w^2T^2)^{1/2}$

Advantages of frequency response

The advantages of the frequency response method are as follows:

- It includes simple calculations.
- The frequency response method is easy to implement in the designs of the control system. It also helps us to find the stability of the system.
- It provides the stability analysis of the system without the need for any complex and time-consuming processes.
- The frequency response and the step response of the system are closely related. One known parameter gives us the idea of the other parameter.
- We can obtain the frequency response of the given control system without the knowledge of the transfer function.
- The stability analysis of the system can be performed even if it incorporates moderate degree of non-linearity.
- We can also apply the frequency response on the system that has irrational transfer function. For example, $^{e-2Ts}$.
- It involves simple and inexpensive apparatus.
- In case of the complex cases of the control system, it is better to use the Nyquist plot technique. It is the only method to analyze the stability in such conditions.
- The effect of the noise disturbance can be easily analyzed.
- The adjustment and performance of the closed loop system using the frequency response is easy as compared to the time domain.

Disadvantages of frequency response

The disadvantages of the frequency response method are as follows:

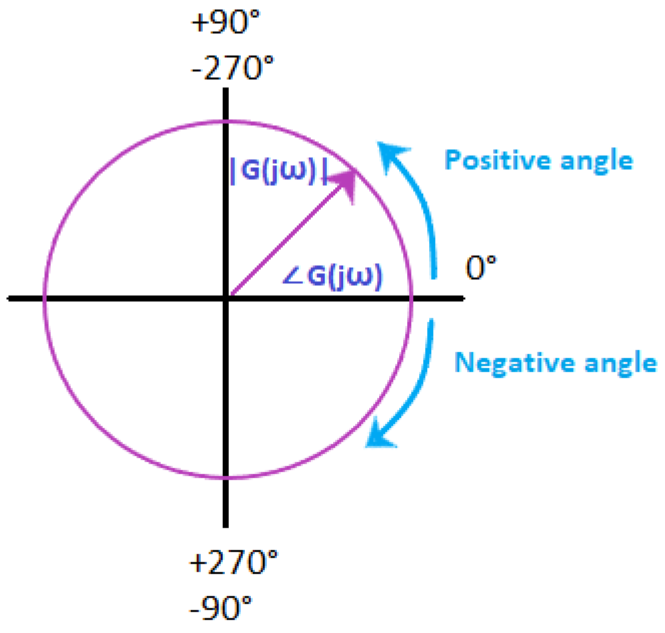
- The frequency response method works better with the linear system. The result in the cases of non-linear systems or the system with moderate non-linearity does not show the exact results. Hence, it is generally applied only to linear systems.
- The practical method to obtain the frequency response is time-consuming.
- There is a relation between the frequency response and the step response, but it is not exact as expected. But, the exact relation is possible if we use the Fourier transformation to describe it, which is difficult to apply due to complex calculations.

Polar plot

The polar plot of a sinusoidal transfer function is the plot of the magnitude $G(j\omega)$ versus the phase angle of $G(j\omega)$ on the **polar coordinates.** The frequency in the polar plot is varied from **zero** to **infinity.** The polar plot is drawn on the polar sheet, which is the form of a graph, and the graph consists of concentric circles and radial lines.

The concentric circles on the polar sheet graph represent the magnitude, and the radial lines represent phase angles. Each point on the graph displays information about the magnitude and the phase angle.

The example of a polar graph is shown below:



The positive angle in a polar graph is measured in the anti-clockwise direction, while the negative angle is measured clockwise direction. Both the angles are measured with respect to the reference point, i.e., 0 degree axis.

The transfer function $G(j\omega)$ in the rectangular form can be written as:

$G(j\omega) = GR(j\omega) + GI(j\omega)$

Where,

$G_R(j\omega)$ is the real part of the transfer function $G(j\omega)$

$G_I(j\omega)$ is the imaginary part of the transfer function $G(j\omega)$

As discussed, the angular frequency in the polar plot varies from zero to infinity. We should not get confused with the Nyquist plot, which is the extension of the polar plot. The angular frequency (ω) varies from zero to infinity, while the Nyquist plot varies from a negative value of infinity to positive infinity.

The primary advantage of the polar plot is that it depicts the **frequency response** characteristics of a system over the entire frequency range in a single plot. Since everything seems to be a single block, it fails to show the contributions of each factor of the open-loop transfer function.

The **integral** factor of the polar plot is given by:

$G(s) = 1/s$

Where,

$S = j\omega$

s is the transfer function

$G(j\omega) = 1/j\omega$

It is the negative imaginary axis.

The **derivative** factor of the polar plot is given by:

$G(j\omega) = j\omega$

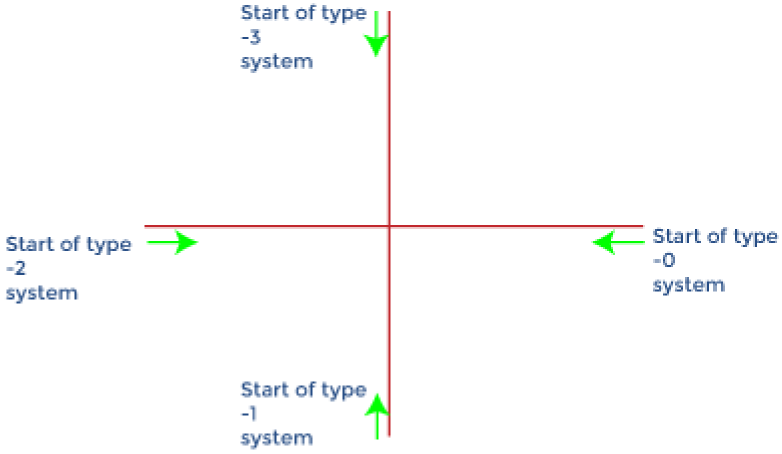
It is the positive imaginary axis.

Effect of addition of pole and zero

- The addition of pole in the polar plot will shift its end by -90 degrees.
- The addition of zero in the polar plot will shift its end by +90 degrees.

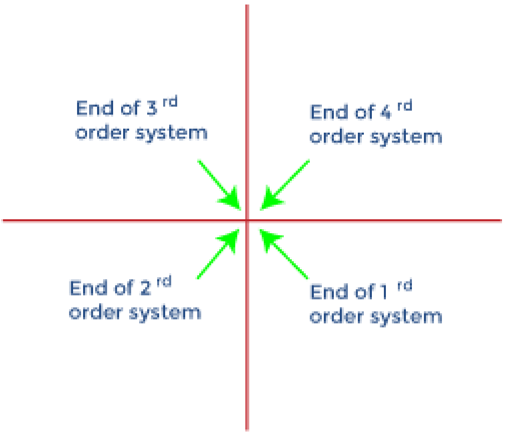
Type and order of the system

The type of the system in the polar plot determines the quadrant at which the polar plot starts. The start of the polar plot of all poles is shown in the below diagram.


Start of type -3 system
Start of type -2 system
Start of type -0 system
Start of type -1 system

Start of polar plot of all pole minimum phase system.

The order of the system in the polar plot determines the quadrant at which the polar plot ends. The end of the polar plot of all poles is shown in the below diagram.


End of 3 rd order system
End of 4 rd order system
End of 2 rd order system
End of 1 rd order system

Polar plot of standard transfer function

Here, we will discuss the polar plots of various standard transfer functions.

Type 0

Order: 1

Let, $G(s) = 1/(1 + sT)$

Put, $s= j\omega$

$G(j\omega) = 1/(1 + j\omega T)$

The above transfer function in the form of magnitude and angle can be represented as:

$G(j\omega) = 1/[(1 + \omega^2 T^2)^{1/2}\angle\tan^{-1}\omega T]$

If we consider the angle part in the numerator, we need to insert a negative sign due to the transition from the denominator to the numerator or vice-versa.

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

Let's find the value of the above function at zero and infinity.

When, $\omega = 0$

$G(j\omega) = 1\angle 0/1$

$G(j\omega) = 1\angle 0$

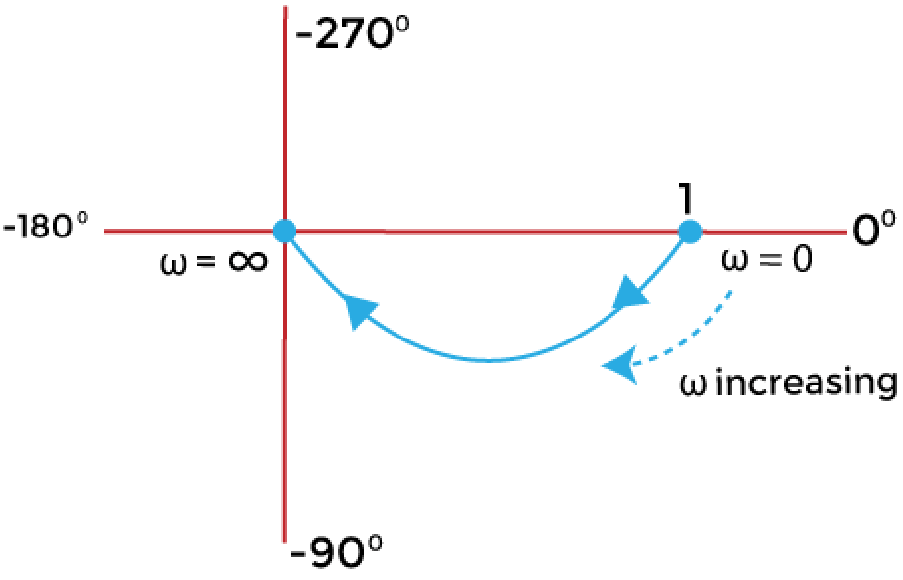
It is because $\tan^{-1}0 = 0$

When, ω = infinity

$G(j\omega) = 0\angle-90$

It is because $\tan^{-1}\infty = 90$ degrees

The polar plot at value 0 and infinity will appear as:



Order 2:

Since, it is a 2 order system; the function includes the highest derivative variable (s) with the power 2.

Let, $G(s) = 1/(1 + sT_1)(1 + sT_2)$

Put, $s= j\omega$

$G(j\omega) = 1/(1 + j\omega T_1)(1 + j\omega T_2)$

The above transfer function in the form of magnitude and angle can be represented as:

$G(j\omega) = 1/[(1 + \omega^2 T_1{}^2)^{1/2}\angle\tan^{-1}\omega T_1][(1 + \omega^2 T_2{}^2)^{1/2}\angle\tan^{-1}\omega T_2]$

If we consider the angle part in the numerator, we need to insert a negative sign due to the transition from the denominator to the numerator, as shown below:

$G(j\omega) = 1\angle(-\tan^{-1}\omega T_1 - \tan^{-1}\omega T_2) /(1 + \omega^2 T_1{}^2)^{1/2}(1 + \omega^2 T_2{}^2)^{1/2}$

Let's find the value of the above function at zero and infinity.

When, $\omega = 0$

$G(j\omega) = 1\angle(-0 - 0)/1x1$

$G(j\omega) = 1\angle 0$

It is because $\tan^{-1}0 = 0$

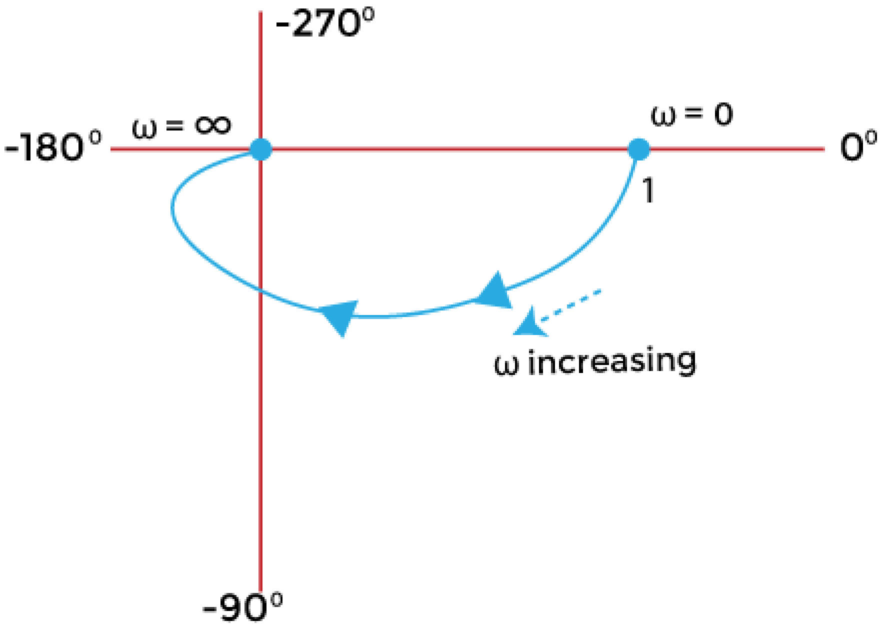
When, ω = infinity

$G(j\omega) = 0\angle(-90 - 90)/1$

$G(j\omega) = 0\angle -180$

It is because $\tan^{-1}\infty = 90$ degrees

The polar plot at value 0 and infinity will appear as:



Type 1

Order: 1

Let, G(s) = 1/s

Put, s= jω

G(jω) = 1/jω

G(jω) = 1/(ω∠90)

If we consider the angle part in the numerator, we need to insert a negative sign due to the transition from the denominator to the numerator, as shown below:

G(jω) =1∠-90 /ω

Let's find the value of the above function at zero and infinity.

When, ω = 0

G(jω) =∞ ∠-90

It is because the function is directly divided by ω. 1/0 = infinity.

$\tan^{-1}0 = 0$

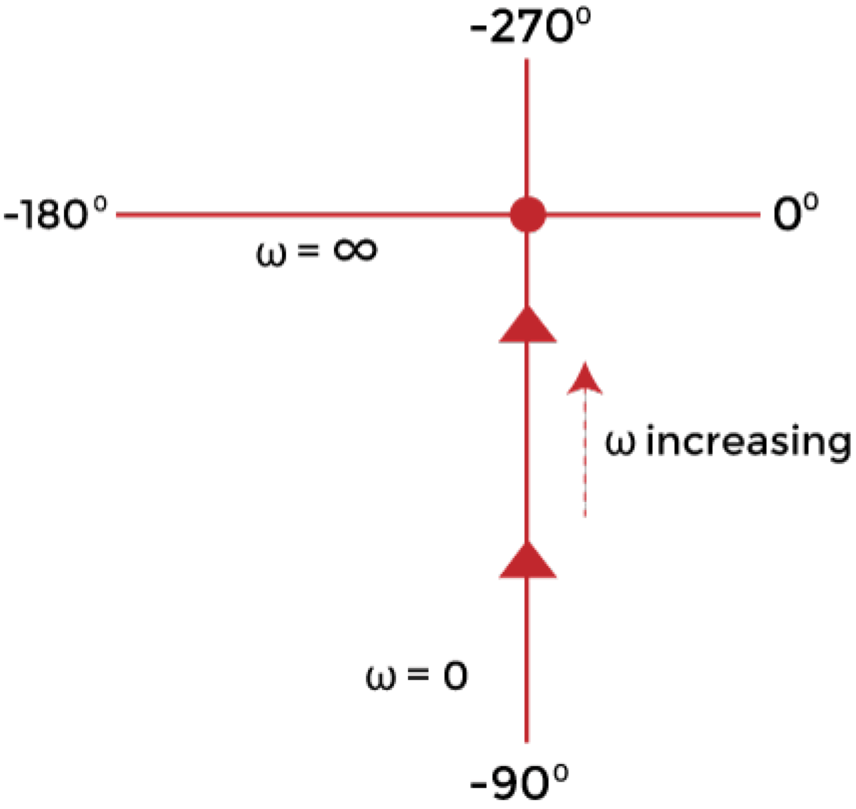
When, ω = infinity

G(jω) = 0∠-90

It is because the function is directly divided by ω. 1/infinity = zero.

$\tan^{-1} \infty = 90$ degrees

The polar plot at value 0 and infinity will appear as:



Order: 2

Let, G(s) = 1/s(1 + sT)

Put, s= jω

G(jω) = 1/ jω(1 + jωT)

The above transfer function in the form of magnitude and angle can be represented as:

$G(j\omega) = 1/\omega\angle 90\ [(1 + \omega^2 T^2)^{1/2}\angle\tan^{-1}\omega T]$

If we consider the angle part in the numerator, we need to insert a negative sign due to the transition from the denominator to the numerator or vice-versa.

$G(j\omega) = 1\angle(-90 - \tan^{-1}\omega T) / \omega(1 + \omega^2 T^2)^{1/2}$

Let's find the value of the above function at zero and infinity.

When, ω = 0

G(jω) =∞ ∠-90

It is because the function is directly divided by ω. 1/0 = infinity.

$\tan^{-1}0 = 0$

When, ω = infinity

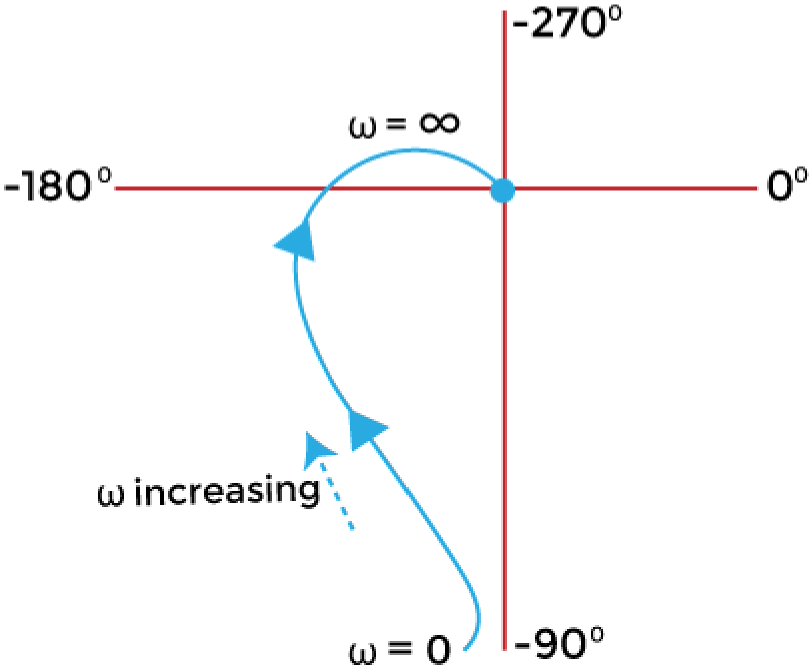
G(jω) = 0∠(-90-90)

G(jω) = 0∠-180

It is because the function is directly divided by ω. 1/infinity = zero.

$\tan^{-1} \infty = 90$ degrees

The polar plot at value 0 and infinity will appear as:



Order: 3

Let, $G(s) = 1/s(1 + sT_1)(1 + sT_2)$

Put, s= jω

$G(j\omega) = 1/ j\omega (1 + j\omega T_1) (1 + j\omega T_2)$

The above transfer function in the form of magnitude and angle can be represented as:

$G(j\omega) = 1/\ \omega\angle 90\ [(1 + \omega^2\ T_1\ ^2)^{1/2}\angle \tan^{-1}\omega T_1][(\ 1 + \omega^2\ T_2\ ^2)^{1/2}\angle \tan^{-1}\omega T_2]$

If we consider the angle part in the numerator, we need to insert a negative sign due to the transition from the denominator to the numerator, as shown below:

$G(j\omega) = 1\angle(-90 - \tan^{-1}\omega T_1 - \tan^{-1}\omega T_2)\ /\omega(1 + \omega^2\ T_1^2)^{1/2}(1 + \omega^2\ T_2^2)^{1/2}$

Let's find the value of the above function at zero and infinity.

When, $\omega = 0$

$G(j\omega) = \infty\ \angle -90$

It is because the function is directly divided by ω. 1/0 = infinity.

$\tan^{-1}0 = 0$

When, ω = infinity

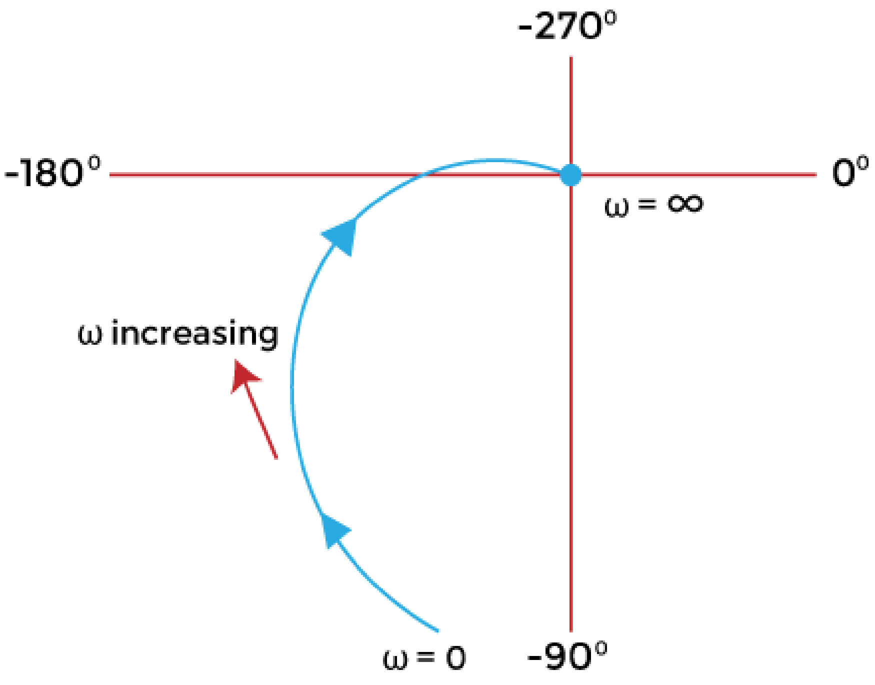
$G(j\omega) = 0\angle(-90 -90 -90)$

$G(j\omega) = 0\angle -270$

It is because the function is directly divided by ω. 1/infinity = zero.

$\tan^{-1}\infty = 90$ degrees

The polar plot at value 0 and infinity will appear as:



Type 2

Order: 4

Let, $G(s) = 1/s^2(1 + sT_1)(1 + sT_2)$

Put, $s = j\omega$

$G(j\omega) = 1/ j^2 \omega^2 (1 + j\omega T_1)(1 + j\omega T_2)$

The above transfer function in the form of magnitude and angle can be represented as:

$G(j\omega) = 1/ \omega^2\angle -180 \, [(1 + \omega^2 T_1{}^2)^{1/2}\angle \tan^{-1}\omega T_1][(1 + \omega^2 T_2{}^2)^{1/2}\angle \tan^{-1}\omega T_2]$

If we consider the angle part in the numerator, we need to insert a negative sign due to the transition from the denominator to the numerator, as shown below:

$G(j\omega) = 1\angle(-180-\tan^{-1}\omega T_1 - \tan^{-1}\omega T_2) /(1 + \omega^2 T_1{}^2)^{1/2}(1 + \omega^2 T_2{}^2)^{1/2}$

Let's find the value of the above function at zero and infinity.

When, $\omega = 0$

$G(j\omega) = \infty \angle(-180 - 0 - 0)/1x1$

$G(j\omega) = \infty \angle -180$

It is because the function is directly divided by ω. 1/0 is equal to infinity.

$\tan^{-1}0 = 0$

When, ω = infinity

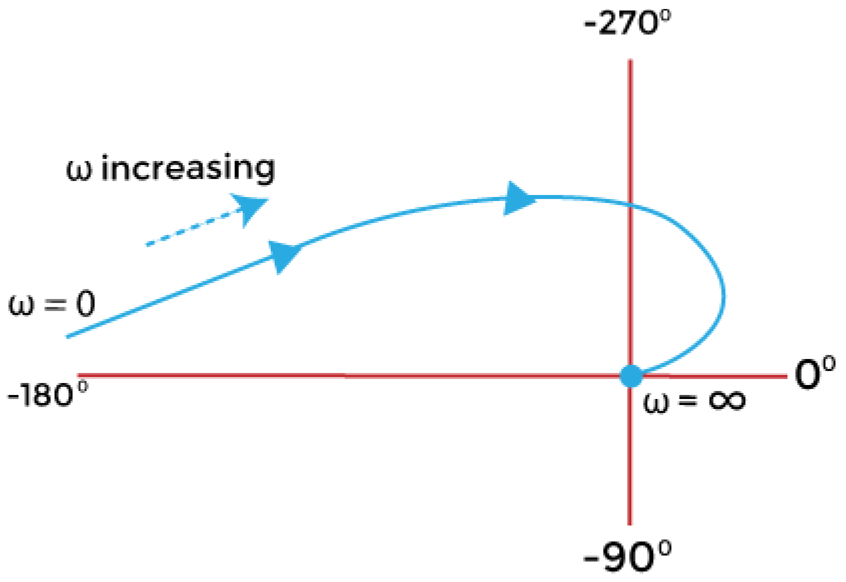
$G(j\omega) = 0\angle(-180 - 90 - 90)/1$

$G(j\omega) = 0\angle -360$

It is because the function is directly divided by ω. 1/infinity is equal to zero.

$\tan^{-1} \infty = 90$ degrees

The polar plot at value 0 and infinity will appear as:



Gain margin and phase margin of polar plot

The gain margin is Kg. It is given by:

$Kg = 1/G_B$

G_B is the value at a point (B) on the magnitude circle that cuts the 180 degree axis. The gain margin is positive if the point B lies within unit circle. Otherwise, it is negative.

The phase margin is given by:

Y = 180 + theta

Where,

Theta is the phase angle of G(jω) at the gain crossover frequency. The phase angle is calculated when the magnitude curve line intersects with the external unity circle. The line drawn from that intersection point to the end of the graph determines the angle theta. It can be positive or negative.

Both the phase margin and the gain margin can be better understood with the help of an example.

Examples

Let's discuss an example of the polar plot.

Example: The open loop transfer function of a unity feedback system is given by $G(s) = 1/s(s + 1)(2s + 1)$. Sketch the polar plot and also determine the gain margin and the phase margin.

Solution: The transfer function is given by:

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

The above function clearly depicts that the system is of type 1 and order 3. It is in the form:

$G(s) = 1/s(1 + sT_1)(1 + sT_2)$

Put, $s = j\omega$

$G(s) = 1/j\omega(j\omega + 1)(2j\omega + 1)$

The above transfer function in the form of magnitude and angle can be represented as:

$G(j\omega) = 1/\omega\angle 90\,[(1 + \omega^2)^{1/2}\angle\tan^{-1}\omega][(1 + \omega^2 4)^{1/2}\angle\tan^{-1}\omega 2]$

If we consider the angle part in the numerator, we need to insert a negative sign due to the transition from the denominator to the numerator, as shown below:

$G(j\omega) = 1\angle(-90 - \tan^{-1}\omega - \tan^{-1} 2\omega) / \omega(1 + \omega^2)^{1/2}(1 + 4\omega^2)^{1/2}$

Now, let us separate the magnitude and angle terms from the above equation.

$|G(j\omega)| = 1/\omega((1 + \omega^2)(1 + 4\omega^2))^{1/2}$

$|G(j\omega)| = 1/\omega(1 + 4\omega^2 + \omega^2 + 4\omega^4)^{1/2}$

$|G(j\omega)| = 1/\omega(1 + 5\omega^2 + 4\omega^4)^{1/2}$

Angle $G(j\omega) = -90 - \tan^{-1}\omega - \tan^{-1} 2\omega$

We know the value of the above function at zero and infinity.

When, $\omega = 0$

$G(j\omega) = \infty\angle -90$

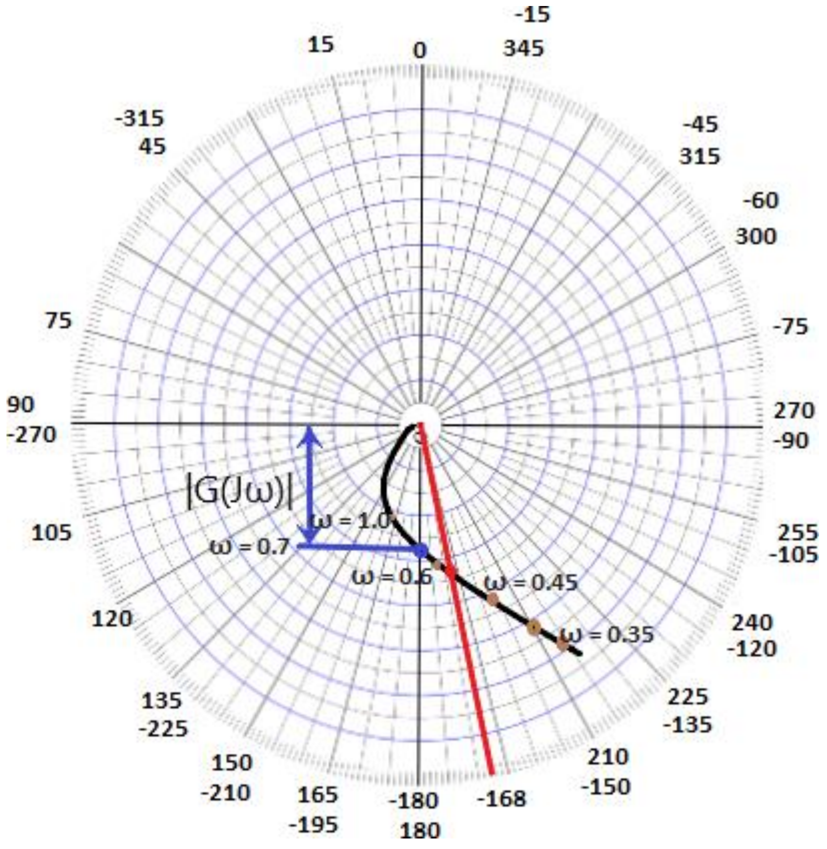
When, ω = infinity

$G(j\omega) = 0\angle -270$

Let's find the magnitude and phase of $G(j\omega)$ at different frequencies.

Radians/ second (ω)	0.35	0.5	0.6	0.7	1.0
|G(jω)|	2.22	1.2	0.9	0.7	0.3
Angle G(jω)	-144	-162	-171	-180	-198

The polar plot is shown below:



Here, every two lines have a gap of **15 degrees.** We have specified both the positive and negative angle value at a point. It is based on the concept that the positive angles are measured anti-clockwise and the negative angles are measured clockwise. As we start from the 0 angle in the clockwise direction, we can notice the increase in the negative values of the angle. Similarly, as we begin from the 0 angle in the clockwise direction, we can notice the rise in the positive values of the angle.

Let's calculate the gain margin and the phase margin.

We can see in the polar plot that the magnitude circle cuts the 180 degree axis at point 0.7. Hence, it will be the value of G_B.

The gain margin is $Kg = 1/G_B$

$Kg = 1/0.7$

$Kg = 1.428$

The phase margin is given by:

Y = 180 + theta

We can clearly see the point marked with the red color. It is the intersection point of the magnitude curve with the unity circle. The line drawn from the intersection point (marked in red) determines the theta angle, which is equal to (-168) degrees.

So, phase angle is equal to

Y = 180 - 168

Y = 12 degrees

Thus, the gain margin is 1.428 and the phase angle is 12 degrees.

Nyquist plot

The extension of the polar plot is known as the **Nyquist plot.** The frequency in the case of the Nyquist plot varies from **-infinity to infinity.** The primary difference between the polar plots and the Nyquist plot is that the polar plots are based on frequencies range from zero to infinity, while the Nyquist plot also deals with negative frequencies.

The Nyquist criteria help us determine the closed-loop system's stability from the frequency response of the open-loop poles and plot.

We know that F(s) is a function of s. The polynomial in the numerator and denominator of the system in terms of s can be represented as:

F(s) = (s - z1)(s - z2)... (s - zm)/(s - p1)(s - p2)... (s - pn)

The roots of the numerator, when equated to zero, determine the **zeroes** of the system, and the root of the denominator determines the **poles** of the system. It means that the given function has m number of zeroes and n number of poles. The numerical value of n is usually greater or equal to m.

S in the function is a complex variable, and it is given by **$\sigma + j\omega$.** Thus, F(s) is also a complex function that can be represented in the form **u + jv.**

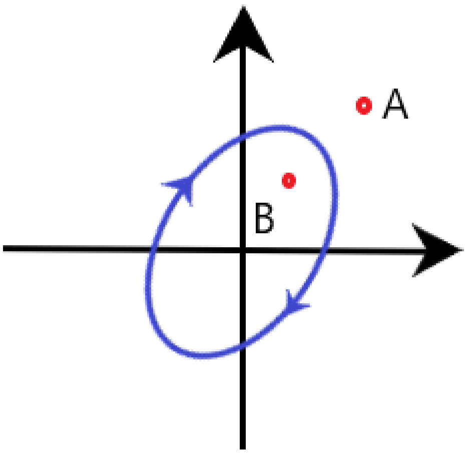
It means that for every point of s in the s-plane at which the F(s) is analytic, there exists a corresponding point in the F(s) plane. The function f(s) maps into the f(s) plane. There is a contour that maps on the contour on the other side.

In the Nyquist plot, we will detect the presence of the **closed-loop system** poles in the right half of the s-plane to determine the system's stability. It is because the Nyquist plot relates the open loop frequency response (given by G(jω)H(jω)) to the number of poles and zeroes of 1 + G(s)H(s) that lie in the right-half of the s-plane.

Contour in s-plane

The direction of the contour in the s-plane can either enclose or encircle the point in the s-plane, and the point can be a zero or a pole. First, let's discuss the concept of encircle and enclose because both terms are useful while implementing the Nyquist stability criterion.

Encircled: If a point is said to lie inside the closed path, it is said to be encircled. It is shown below:



Here, point B is encircled in the closed path in a clockwise direction, while the point A lies outside the path.

- **Enclosed:** If a point lies to the right side of path when the path is traversed in a specific direction, it is said to be enclosed by a closed path. Let's consider two closed path in the clockwise and anticlockwise direction, as shown below:

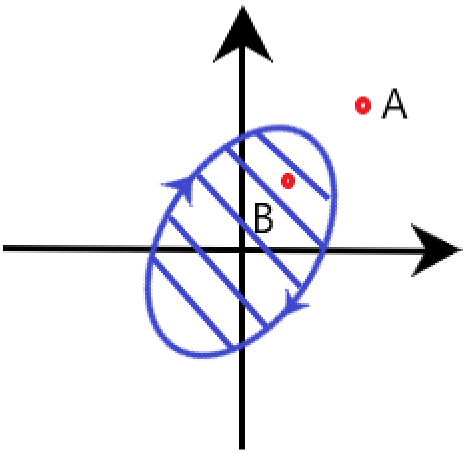


Figure 1

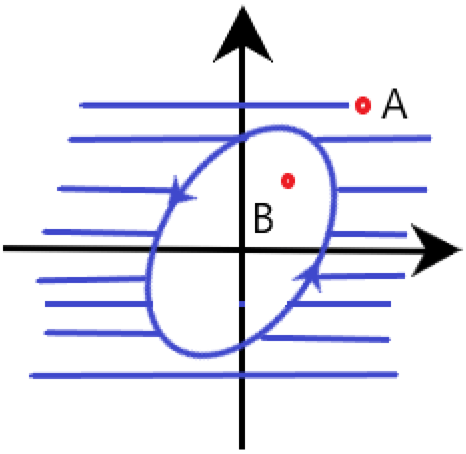


Figure 2

The shaded region represents the region enclosed by the closed path. In the first figure, point B lies to the right when traversed in the clockwise direction, while A does not. Thus, **point B** is said to be enclosed by a closed path. Similarly, point A lies to the right when traversed in the anticlockwise direction in the second figure, while B does not. Thus, **point A** is said to be enclosed by a closed path.

Nyquist stability criteria

The closed loop transfer function is given by:

C(s)/R(s) = G(s)/(1 + G(s)H(s))

Where,

C(s) is the output of the given control system

R(s) in the input of the given control system

H(s) is the feedback path

G(s) is the forward path of a system

The characteristic equation of the system is given by the condition 1 + G(s)H(s) = 0.

We know that G(s)H(s) in terms of zeroes and poles is given by:

$$G(s)H(s) = \frac{K(s+z_1)(s+z_2)(s+z_3)\ldots(s+z_m)}{(s+p_1)(s+p_2)(s+p_3)\ldots(s+p_n)}$$

The value of m is less or equal to n. It means that for an ideal control system, the number of zeroes is always less than or equal to the number of poles in a given control system.

Let, F(s) = 1 + G(s)H(s)

Putting the value of G(s)H(s) in the above equation, we get:

$$F(s) = \frac{K(s+z_1)(s+z_2)(s+z_3)\ldots(s+z_m)}{(s+p_1)(s+p_2)(s+p_3)\ldots(s+p_n)} + 1$$

$$F(s) = \frac{(s+p_1)(s+p_2)(s+p_3)\ldots(s+p_n) + K(s+z_1)(s+z_2)(s+z_3)\ldots(s+z_m)}{(s+p_1)(s+p_2)(s+p_3)\ldots(s+p_n)}$$

$$F(s) = \frac{K(s+z_1')(s+z_2')(s+z_3')\ldots(s+z_m')}{(s+p_1)(s+p_2)(s+p_3)\ldots(s+p_n)}$$

Thus, z1', z2'... zn' are the zeroes of the function F(s).

Now, let's combine the values of G(s), 1 + G(s)H(s) in the transfer function, which is given by:

C(s)/R(s) = G(s)/(1 + G(s)H(s))

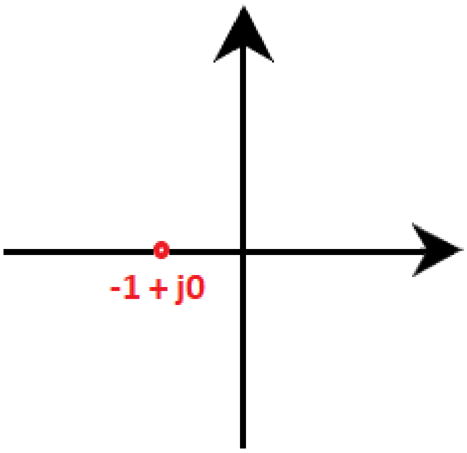
$$\frac{C(s)}{R(s)} = \frac{\dfrac{K(s+z_1)(s+z_2)(s+z_3)\ldots(s+z_m)}{(s+p_1)(s+p_2)(s+p_3)\ldots(s+p_n)}}{1 + \dfrac{K(s+z_1)(s+z_2)(s+z_3)\ldots(s+z_m)}{(s+p_1)(s+p_2)(s+p_3)\ldots(s+p_n)}}$$

$$\frac{C(s)}{R(s)} = \frac{K(s+z_1)(s+z_2)(s+z_3)\ldots(s+z_m)}{(s+z_1')(s+z_2')(s+z_3')\ldots(s+z_n')}$$

z1', z2', z3', and zn' are the poles of the above transfer function.

The Nyquist stability criterion is based on the point -1 + j0 to determine the stability of the closed loop system. It is because the contour of the function F(s) with respect to the origin of the plane is same as the contour of the F(s) -1 plane with respect to the point -1 + j0.

The point -1 + j0 on the axis will appear as:



Let's discuss the Nyquist stability criteria in terms of encirclement, anticlockwise encircle, and clockwise encirclement.

Encirclement of point -1 + j0
There should be no encirclement of point -1 + j0. We know that the system is stable if the poles are present on the left half of the s-plane. Here, no encirclement means that the system is stable if there are no poles on the right side of the s-plane. The poles present on the right half of the s-plane makes the system unstable.

Anticlockwise encirclements of point -1 + j0
The anticlockwise encirclements of the point -1 + j0 are equal to the number of poles present in the right half of the s-plane. If such encirclements are not equal to the number of poles, the system becomes unstable.
For example,
A given system has two poles. For the system to be stable, the encirclements of the point -1 + j0 should also be two.

Clockwise encirclements of point -1 + j0
There should be no clockwise encirclements of the point -1 + j0 in the Nyquist plot to stabilize the system. If such encirclements are present in the plot, the system is always unstable.

Stability analysis to draw the Nyquist plot

Here, we will discuss the steps that will help in the stability analysis to draw the Nyquist plot.

1. **Determining the poles and zeroes**

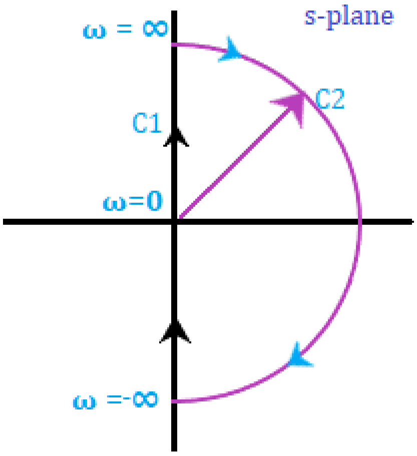
   From the given transfer function, we need to determine the poles and zeroes to check the valid points.

2. **Selecting a Nyquist plot**

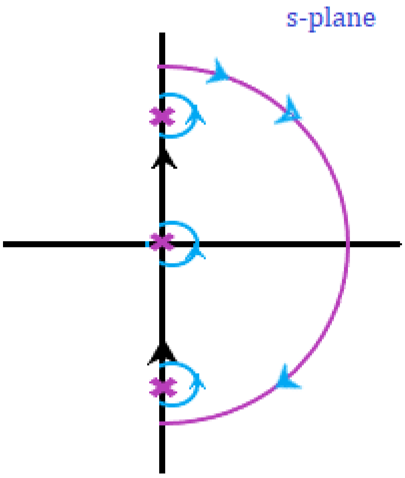
   We need to select a Nyquist plot that should enclose all the poles and zeroes present on the right-half of the s-plane except the singular points. The singular points are considered as the points lying on the imaginary axis. Hence, these points are avoided.

   If the transfer function has a pole at zero or origin, the contour encloses all the poles and zeroes except the origin.

   If there are no poles on the imaginary axis, the contour will appear as:



Both the avoided points (origin and imaginary) in the contour are shown below:

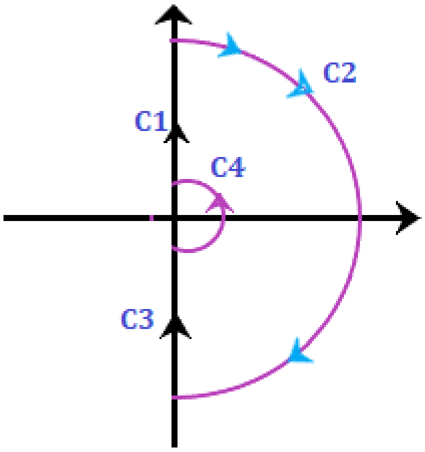


For mapping, the contour needs to be analytical. Let's first discuss about mapping.

3. Mapping a contour

A singular point (point on the imaginary axis) is not analytical. Hence, it is generally avoided. The Nyquist contour is mapped to determine the encirclement of the point -1 + j0. The contour is drawn based on the transfer function is G(s)H(s).

There are four sections C1, C2, C3, and C4. The Nyquist criterion pot is divided into four sections so that the process can be easily carried out section wise. At last, all the sections are combined to produce the desired Nyquist plot. Let the four sections be:



The order of these sections is somewhat similar as shown above.

4. Range for the first section C1

We know that the Nyquist range is from -infinity to infinity, and the value of **ω** in section C1 ranges from **0** to **infinity.** The contour will be drawn in G(s)H(s) plane with respect to the above range, and it will be the locus plot of G(jω)H(jω).

We can use various methods of mapping the first section C1, such as sketching the locus from the given transfer function, finding the frequency by equating the real and imaginary parts to zero, using the polar plot method (type and order of the system), and separating the magnitude and the phase..

5. Range for the second section C2

The second section is generally of infinite radius. Here, it is a semi-circle of infinite radius. The range of the second section is from -90 degrees to +90 degrees. It can be obtained by:

$$\lim_{R\to\infty} Re^{j\theta}$$

$$\theta = -\frac{\pi}{2} \text{ to } +\frac{\pi}{2}$$

If the transfer function is in the form G(s)H(s) = K(1 + sT)/s(1 + sT1)(1 + sT2), the term (1 + sT) can be assumed as sT.

The value of the transfer function at the specific value of theta will be:

$$\text{At } \theta = \frac{\pi}{2}, G(s)H(s) = 0e^{-j\frac{\pi}{2}(n-m)}$$

$$\text{At } \theta = -\frac{\pi}{2}, G(s)H(s) = 0e^{j\frac{\pi}{2}(n-m)}$$

Where,

N is the number of poles and m is the number of zeroes. The value of n is greater than or equal to m.

6. **Range for the second section C3**

   In the third section C3, the value of **ω** ranges from -infinity to zero. The locus of the third section is just the inverse of the polar plot of G(jω)H(jω). We can also say that it is the inverse of the first section. The resulted plot will be the mirror image of the polar with respect to the real axis.

7. **Range for the second section C4**

   The argument of the forth section varies from-90 degrees to 90 degrees. The Nyquist contour of this section has a semicircle of zero radii. The mapping of the section C4 is given by the condition:

$$\lim_{R \to 0} Re^{-j\theta}$$

$$\theta = -\frac{\pi}{2} \text{ to } +\frac{\pi}{2}$$

If the transfer function is in the form G(s)H(s) = K(1 + sT)/s^y(1 + sT1)(1 + sT2), the term (1 + sT) can be assumed as 1.

The value of the transfer function at the specific value of theta will be:

$$\text{At } \theta = \frac{\pi}{2}, G(s)H(s) = \infty e^{j\frac{\pi}{2}y}$$

$$\text{At } \theta = -\frac{\pi}{2}, G(s)H(s) = \infty e^{-j\frac{\pi}{2}y}$$

Where,

Y is the number of poles.

Note: If there are no poles at the origin of the Nyquist plot, the section C4 will be absent.

Let's discuss an example of the Nyquist plot for better understanding. We will follow the steps discussed above.

Examples

Consider the below example:

Example: Draw the Nyquist plot for the system whose open loop transfer function is given by:

G(s)H(s) = K/s(s + 2)(s + 10)

Also determine the range of K for which the system is stable.

Solution: The transfer function is given by:

G(s)H(s) = K/s(s + 2)(s + 10)

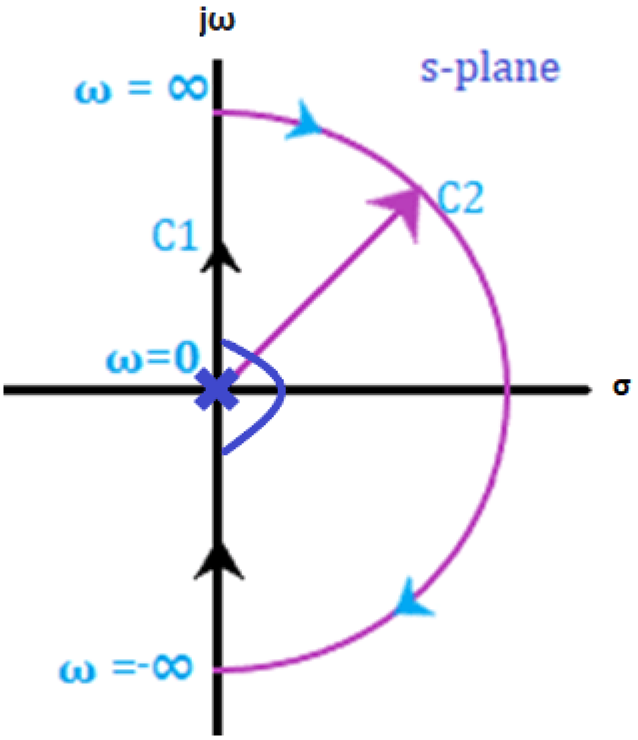
Step 1: Determining the poles and zeroes.

Since, the transfer function has no numerator in terms of s, there are no zeroes present in the function. There are only three poles.

G(s)H(s) = K/sx20(s/2 + 1)(s/10 + 1)

G(s)H(s) = 0.05K/s(1 + 0.5s)(1 + 0.1s)

The open loop function has a pole at the origin, which is shown below:



Now, we will perform the mapping of all four sections separately. At last, all four sections will be combined to produce the desired results.

Step 2: Mapping of section C1

The value of **ω** in section C1 ranges from 0 to infinity. The contour will be drawn in G(s)H(s) plane with respect to the above range will be the locus plot of G(jω)H(jω).

$G(s)H(s) = 0.05K/s(1 + 0.5s)(1 + 0.1s)$

$G(j\omega)H(j\omega) = 0.05K/ j\omega(1 + 0.5j\omega)(1 + 0.1j\omega)$

$G(j\omega)H(j\omega) = 0.05K/ j\omega(1 + 0.6j\omega - 0.05\omega^2)$

$G(j\omega)H(j\omega) = 0.05K/-0.6\ \omega^2 + j\omega(1 - 0.05\omega^2)$

The corresponding imaginary term will be 0 and the resulted frequency will be the crossover phase frequency, which is given by:

$\omega(1 - 0.05\omega^2) = 0$

$1 - 0.05\omega^2 = 0$

$\omega = 4.472$ radians/second = crossover phase frequency

Now, we will put the value of the above frequency in the real part of the G(jω)H(jω), which is given by:

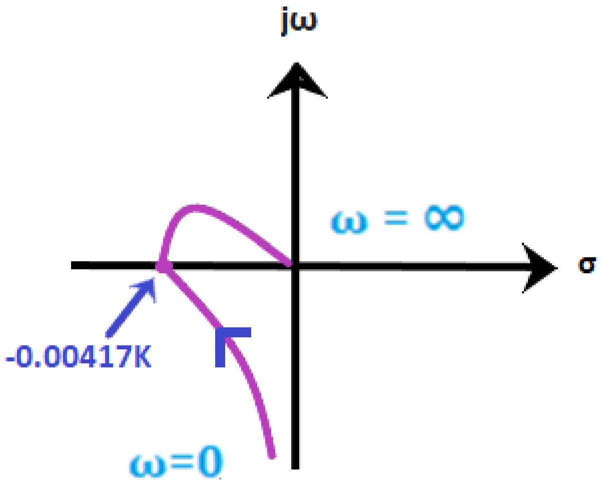
$G(j\omega)H(j\omega) = 0.05K/-0.6\ \omega^2$

$G(j\omega)H(j\omega) = 0.05K/-0.6x\ 4.472\ x\ 4.472$

$G(j\omega)H(j\omega) = -0.00417K$

From the given transfer function, we can easily determine that the system is of type 1 and order 3.

The polar plot will start at -90 degrees, crosses the point -0.00417K on the real axis, as shown below:



Step 3: Mapping of section C2

The range of the second section is from -90 degrees to +90 degrees. It can be obtained by:

$$\lim_{R\to\infty} Re^{j\theta}$$

$$\theta = -\frac{\pi}{2} \text{ to } +\frac{\pi}{2}$$

If the transfer function is in the form G(s)H(s) = 0.05K/s(1 + 0.5s)(1 + 0.1s), the term (1 + sT) can be assumed as sT. It is given by:

G(s)H(s) = 0.05K/s x 0.5s x 0.1s

G(s)H(s) = $0.05K/0.05s^3$

G(s)H(s) = K/s^3

Let, the value of s be:

$$s = \lim_{R\to\infty} Re^{j\theta}$$

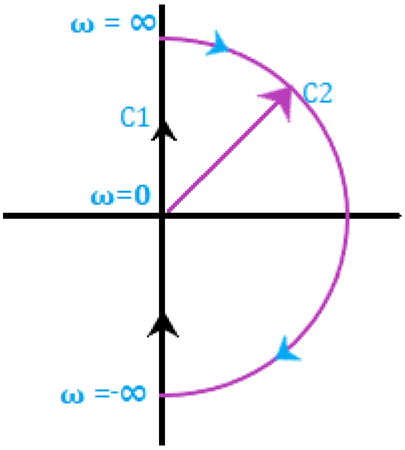
$$G(s)H(s) = 0e^{-j3\theta}$$

The value of the transfer function at the specific value of theta will be:

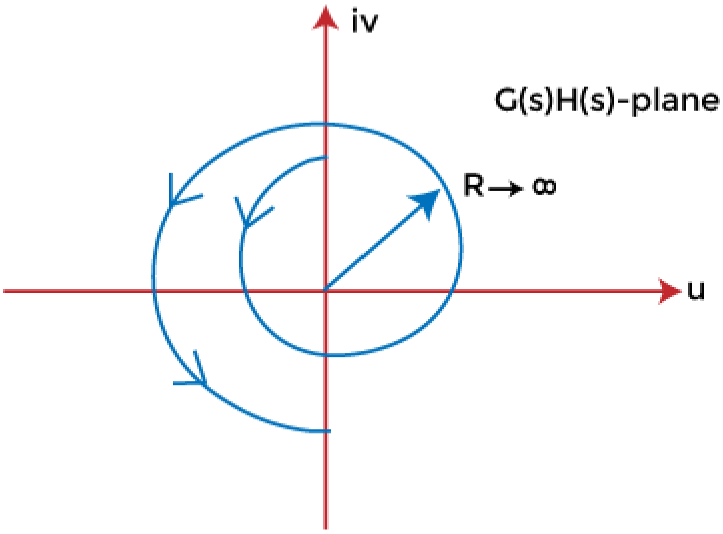
$$\text{At } \theta = \frac{\pi}{2}, G(s)H(s) = 0e^{-j3\frac{\pi}{2}}$$

$$\text{At } \theta = -\frac{\pi}{2}, G(s)H(s) = 0e^{j\frac{3\pi}{2}}$$

Thus, the section C2 varying from -270 degrees to 270 degrees will appear as the plane shown in the below image:



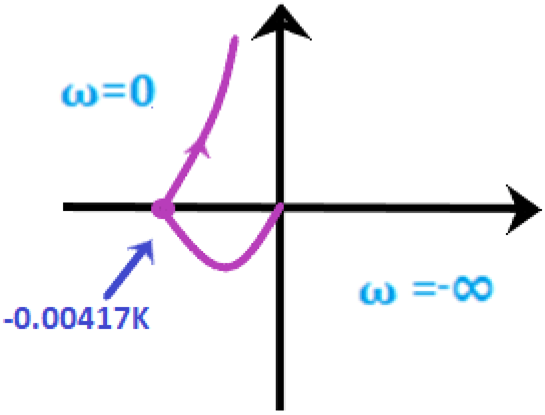
Section C2 in s-plane



Mapping of section C2 in G(s)H(s) plane

Step 4: Mapping of section C3

We know that the section C3 is simply the inverse of section C1. Thus, the locus plot of the section C3 will be the increase on the real axis at the same point, as shown below:



Step 5: Mapping of section C4

Let's plot the C4 section of the Nyquist plot.

Let's plot the C4 section of the Nyquist plot.

The argument of the fourth section varies from-90 degrees to 90 degrees. The Nyquist contour of this section has a semicircle of zero radii. The condition gives the mapping of the section C4:

$$\lim_{R\to 0} Re^{-j\theta}$$

$$\theta = -\frac{\pi}{2} \text{ to } +\frac{\pi}{2}$$

If the transfer function is in the form G(s)H(s) = 0.05K/s(1 + 0.5s)(1 + 0.1s), the term (1 + sT) can be assumed as 1. It is given by:

G(s)H(s) = 0.05K/s x 1 x 1

G(s)H(s) = 0.05K/s

Let, the value of s be:

$$s = \lim_{R\to\infty} Re^{j\theta}$$

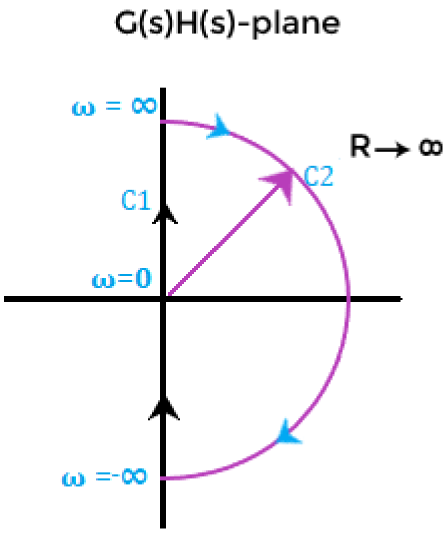
$$G(s)H(s) = \infty e^{-j\theta}$$

The value of the transfer function at the specific value of theta will be:

$$\text{At } \theta = \frac{\pi}{2}, G(s)H(s) = \infty e^{-j\frac{\pi}{2}y}$$

$$\text{At } \theta = -\frac{\pi}{2}, G(s)H(s) = \infty e^{j\frac{\pi}{2}y}$$

We can say that the section C4 is mapped in the s-plane as a circular arc of infinite radius. The locus is shown below:



Mapping of section C4 in G(s)H(s) plane

Step 6: Stability analysis

We will find the value of K when the contour passes through the point (-1 + j0).

-0.00417K = -1

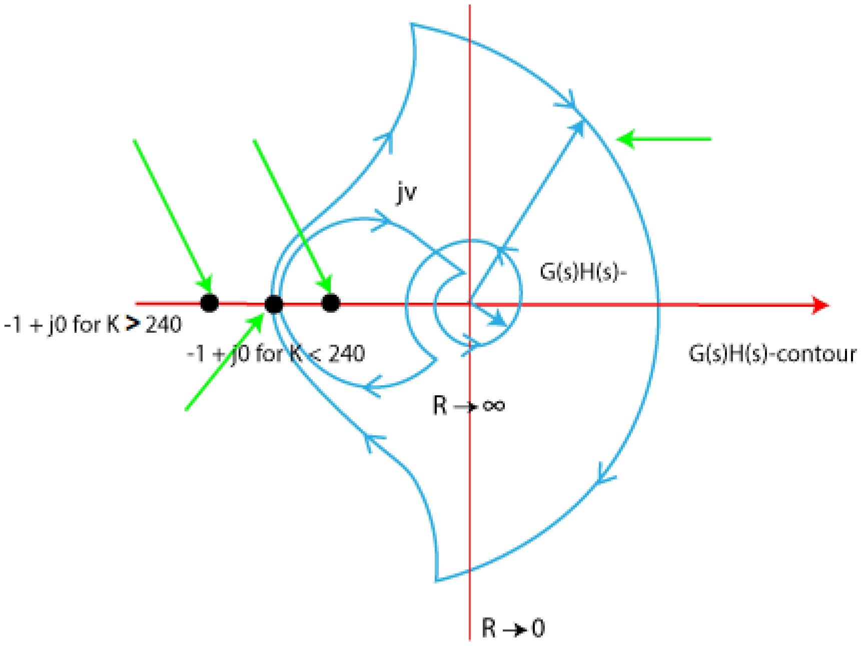
The limiting value of K will be:

K = 1/0.00417

K = 240

Step 7: Complete Nyquist plot

The complete Nyquist plot after combining all the above four sections are shown below:



Nyquist plot of $G(s)\,H(s) = \frac{K}{s(s+2)(s+10)}$

We will find the stability at the two values of K.

K< 240

When the value of K is less than 240, the contour does not cross the real axis, and the point -1 + j0 is not encircled. There are no poles on the right half of the s-plane. Hence, the system at this value of K is stable.

K>240

When the value of K is greater than 240, the contour crosses the real axis, and the point -1 + j0 is encircled two times in the clockwise direction. There are no poles on the right half of the s-plane. Hence, the system at this value of K is unstable.

Thus, for the stability of the given transfer function, the value of K is **0 < K < 240.**

Advantages of the Nyquist plot

The advantages of the Nyquist plot are as follows:

- It can determine the stability of the control system.
- It is better than the root locus in terms of time delay. It means that the Nyquist plot can easily handle the time delay in the system.
- It provides a way to use the bode plots.
- It can find the frequency response of the open-loop transfer function.

- It can find the number of poles present on the right side of the s-plane.
- It can also find the relative stability of the system.

Disadvantages of the Nyquist plot

The disadvantages of the Nyquist plot are as follows:

- It uses some complex mathematical techniques.
- It cannot determine the absolute stability of the system.
- It does not provide the exact information of the number of poles present on the right side of the s-plane.

Bode plot

It is a frequency response plot that contains two graphs, **magnitude** and **phase.** The first plot is the magnitude plot of sinusoidal transfer function versus log w, and the other graph represents the phase angle. It can be drawn both for the open-loop and closed-loop system. It is generally drawn for the **open-loop system** because it conveniently determines the stability and other related parameters.

Bode plot helps us determine the **system's stability** and provides us with a way to **improve** that stability. The standard representation of the Bode plot of the open-loop system is given by:

20 log|G(jω)|

It represents the **logarithmic** magnitude of the function G(s) or G(jω). Here, the base of logarithmic is 10. The unit represents the magnitude of the logarithmic function **G(jω)** in decibels or db. The curve is not drawn on a simple graph paper; and instead, it is drawn on the **semilog paper** that uses the **frequency, phase angle,** and **magnitude** for plotting. The log scale or abscissa is used for the frequency, and the linear scale or ordinate for the phase angle and magnitude.

The logarithmic function allows the various values in the forms of multiple to be added. For example,

Log ab = log a + log b

Thus, the bode plot has an advantage of converting the multiplication of magnitudes in addition. For example,

Let G(jω) = K/ jω (1 + jωT)

The magnitude of G(jω) is $K/[\ \omega\ (1 + \omega^2\ T^2)^{1/2}]$

The phase angle of G(jω) is -90 - ?$\tan^{-1}\omega T$

The magnitude in the form of decibels can be expressed as:

$20 \log [K/ \omega (1 + \omega^2 T^2)^{1/2}]$

$= 20 \log K/ \omega + 20 \log 1/(1 + \omega^2 T^2)^{1/2}$

$= 20 \log K/ \omega - 20 \log (1 + \omega^2 T^2)^{1/2}$

We have inserted the negative sign due to the inverse of the logarithmic value.

It means that log 1/a can be written as $\log (a)^{-1}$, which is equal to **-log a.**

Thus, the above equation depicts that the magnitude when expressed in terms of db. It can easily convert the multiplicative terms to add, meaning that the individual factors of the given transfer function can be added. We will also discuss an example to draw a bode plot later in the topic.

Constant gain K in Bode plot

The constant gain is represented by K. It is also known as the system gain in case of root locus and transfer functions.

Let, $G(s) = K$

$G(s) = G(j\omega) = K$

The angle assumed for the constant gain is 0 degrees.

In terms of decibels, we can represent the system gain as:

$A = |G(j\omega)| = 20\log K$

If the value of K is negative, the angle thus will be 180 degrees.

At the magnitude of the 20log K, the magnitude of constant gain K will be a horizontal straight line.

There are three range of K.

- $K > 1$
- $K = 1$
- $0 < K < 1$

When the value of K is greater than 1, 20log K is positive.

When the value of K is 1, 20log K is zero. It is because the value of log 1 is 0.

When the value of K is greater than 0 and less than 1, 20log K is negative.

Integral factor in Bode plot

Both the integral factor and the derivative factor affect the slope of the magnitude curve. Let's first discuss the integral factor in the Bode plot.

Let the integral factor be K/ jω. The integral factor is represented by constant gain divided by the s, and it can be of one order or more.

$G(s) = K/s$

$G(s) = K/ j\omega$

Its magnitude and phase angles are K/ω and -90 degrees.

Its gain in decibels = $A = |G(j\omega)| = 20\log (K/\omega)$

When $\omega = 0.1$ K, $A = 20\log (K/0.1K) = 20\log (10) = 20$ db

It is because the value of log 10 is 1.

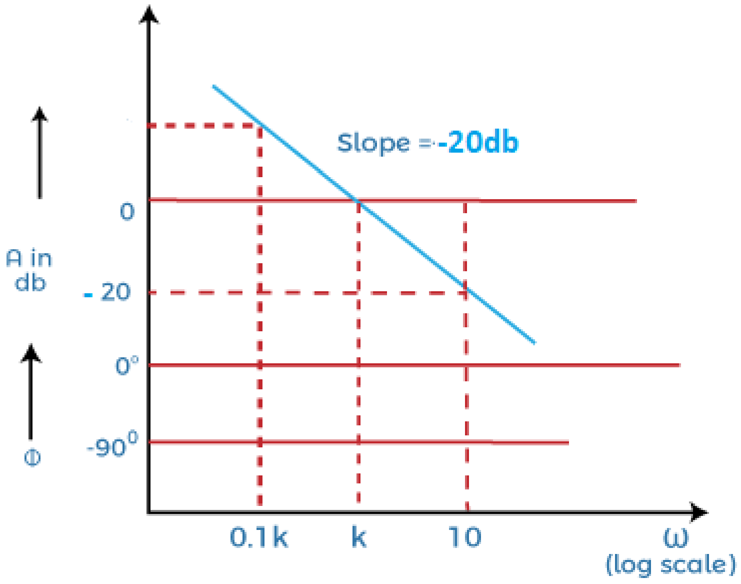
When $\omega = K$, $A = 20\log (K/K) = 20\log (1) = 0$ db

When $\omega = 10$ K, $A = 20\log (K/10K) = 20\log (10)^{-1} = -20$ db

It is because the value of $\log a^{-n}$ is $-n \log a$.

From the above analysis for different values of ω, we can find the magnitude of the integral factor. It will be a straight line with a slope of -20 db. It will also pass through 0 db, when the value of ω is K.

The plot of an integral factor K/ jω is shown below:



Bode plot of integral factor, $\frac{k}{(j\omega)}$

Let's consider another example of an integral factor in the multiples of n. It is given by:

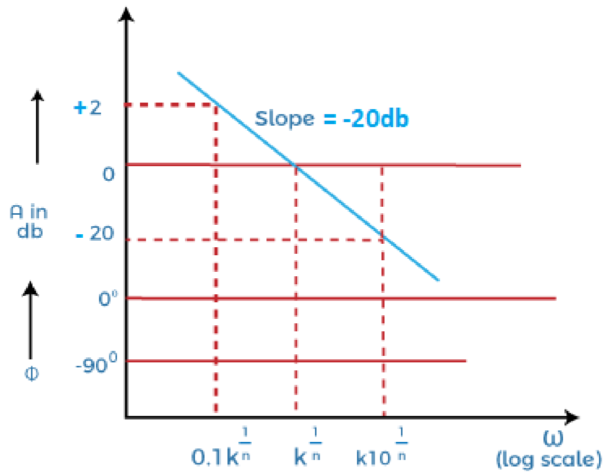
$G(s) = K/s^n$

$G(s) = K/ j\omega^n$

Its magnitude and phase angle are K/ω^n and -90n degrees.

Its gain in decibels = $A = |G(j\omega)| = 20\log (K/\omega^n)$

The plot of an integral factor in the multiples of n is shown below:



Bode plot of integral factor, $K / (j\omega)^n$

Derivative factor in the Bode plot

The derivative factor consists of the terms containing s in the numerator itself. Let's consider a derivative factor represented by K jω.

G(s) = Ks = K jω

Its magnitude and phase angles are Kω and 90 degrees.

It is because the term jω is in the numerator. Thus, the angle will be in positive degrees.

Its gain in decibels = A = |G(jω)| = 20log (Kω)

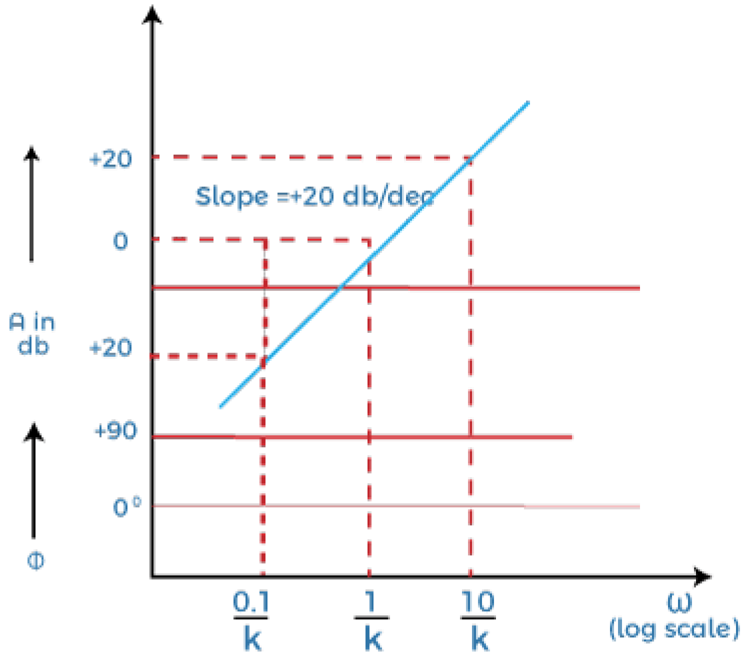
When ω = 0.1/K, A = 20log (K x 0.1/K) = 20log (0.1) = 20log $(10)^{-1}$ = -20 db

It is because the value of log a^{-n} is -n loga.

When ω = 1/K, A = 20log (K/K) = 20log (1) = 0 db

When ω = 10/ K, A = 20log (K x 10/ K) = 20log (10) = 20 db

The plot of the above derivative factor is shown below:



Bode plot of derivative factor, k x jω

It is because the value of log 10 is 1.

From the above analysis for different values of ω, we can find the magnitude of the derivative factor. It will be a straight line with a slope of +20 db. It will also pass through 0 db, when the value of ω is 1/K.

Let's consider another example of a derivative factor in the multiples of n. It is given by:

$G(s) = Ks^n$

$G(s) = Kj\omega^n$

Its magnitude and phase angle are $K\omega^n$ and 90n degrees.

Its gain in decibels = $A = |G(j\omega)| = 20\log (K\omega^n)$

Let's discuss the first order factors in numerator and denominator of the transfer function.

First order factor in the numerator

The first order factor in the numerator or a derivative factor is given by:

$G(s) = 1 + sT$

$G(j\omega) = 1 + j\omega T$

$|G(j\omega)| = (1 + \omega^2 T^2)^{1/2}$

Angle = $?\tan^{-1}\omega T$

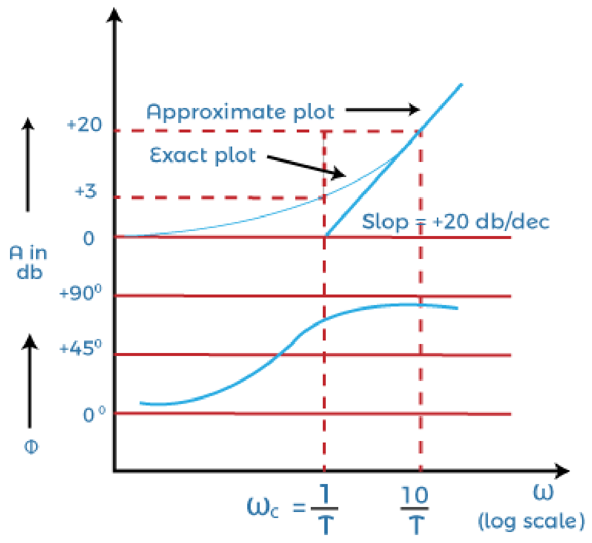
The gain in decibels is given by:

$A = 20 \log (1 + \omega^2 T^2)^{1/2}$

The value of gain at very low frequencies (<<1) will be equal to 20 log 1 = 0 dB.

The value of gain at high frequencies (>>1) will be equal to 20 log ωT.

The Bode plot for the first order function given above is shown below:



We have approximated the magnitude of the plot in two straight lines. The first line is a **straight horizontal line** with very low frequencies (0 db), while the other **slanting slope** refers to the high frequencies (20 log ωT). It would have a slope of +20db.

The two straight lines are the **asymptotes** of the exact curve. The frequency at which these two asymptotes meet is known as the **corner frequency,** and it is also known as break frequency.

For the above-given function, the corner frequency will arise when ω = 1/T. The phase angle will vary from **0 degrees** to **90 degrees** with frequency variation between zero and infinity.

The magnitude approximated at the corner frequency is around + 3dB, and it is generally considered the loss in db at this corner frequency.

First order factor in the denominator

The first order factor in the denominator or an integral factor is given by:

$G(s) = 1/1 + sT$

$G(j\omega) = 1/ 1 + j\omega T$

$|G(j\omega)| = 1/ (1 + \omega^2 T^2)^{1/2}$

Angle = ?$-\tan^{-1}\omega T$

The gain in decibels is given by:

$A = 20 \log 1/(1 + \omega^2 T^2)^{1/2}$

$A = 20 \log (1 + \omega^2 T^2)^{-1/2}$

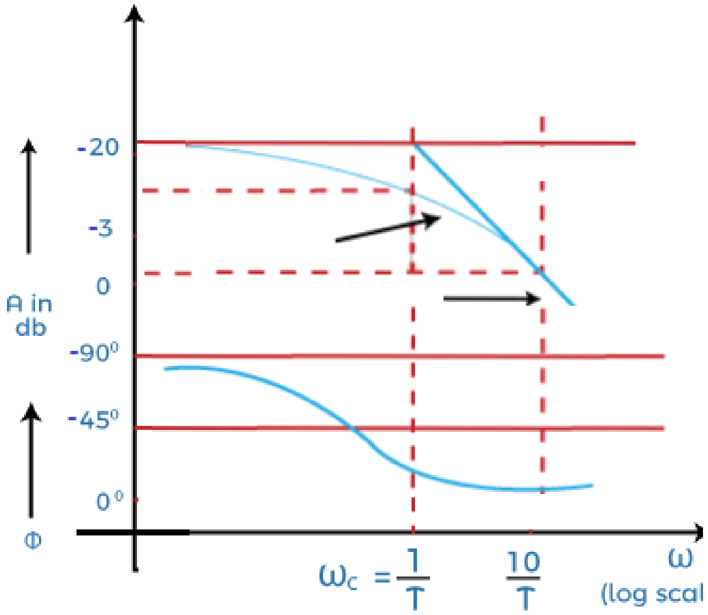
$A = -20 \log (1 + \omega^2 T^2)^{1/2}$

It is because the value of $\log a^{-n}$ is -n loga.

The value of gain at very low frequencies (<<1) will be equal to -20 log 1 = 0 dB.

The value of gain at high frequencies (>>1) will be equal to -20 log ωT.

The Bode plot for the first order function given above is shown below:



We have approximated the magnitude of the plot in two straight lines. The first line is a straight horizontal line with the very low frequencies (0 db), while the other slanting slope refers to the high frequencies (-20 log T). It would have a slope of -20db.

Note: The derivative factor in the transfer function generally represents to the positive slope, while the integral factor represents the negative slope in the bode plot.

The two straight lines are the asymptotes of the exact curve.

For the above-given function, the corner frequency will arise when ω = 1/T. The corner frequency divides the curve of the Bode plot into two regions, i.e., low frequency and high frequency. The phase angle will vary from 0 degrees to -90 degrees with frequency variation between zero and infinity.

The magnitude approximated at the corner frequency is around - 3dB, and it is generally considered the loss in db at this corner frequency. **Note:** The two curves in the Bode plot is the magnitude curve and the phase angle curve. The first curve that represents the value in decibels represents the magnitude curve, while the second curve represents the phase angle curve.

Quadratic factor in the denominator

It is the essential part of the Bode plot to understand. It means the factor in the denominator in present in the form of quadratic equation.

Let's consider a quadratic equation, which is given by:

$G(s) = \omega_n^2/(s^2 + 2a\, \omega_n s + \omega_n^2)$

Where,

a is the damping ratio

ω_n is the natural frequency

We can write the above transfer function as:

$G(s) = 1/\ (1 + 2as/\ \omega_n + (s/\ \omega_n)^2$

Put, $s = j\omega$

$G(j\omega) = 1/\ (1 + 2a\ j\omega/\ \omega_n + (j\omega\ /\ \omega_n)^2$

$G(j\omega) = 1/(1 - (\omega\ /\ \omega_n)^2 + j2a\ \omega/\ \omega_n$

The magnitude is given by:

$|G(j\omega)\ |= 1/[(1 - (\omega\ /\ \omega_n)^4) + 4a^2\ \omega^2/\ \omega_n^2]^{1/2}$

The phase angle is given by:

Angle $G(j\omega) = \tan^{-1}[(2a\ j\omega/\omega_n)/(1 - (\omega\ /\ \omega_n)^2)$

Let's find the gain of the above transfer function in decibels.

$A = 20\log G(j\omega)$

$A = 20\ \log 1/[(1 - (\omega\ /\ \omega_n)^4\ )+ 4a^2\ \omega^2/\ \omega_n^2]^{1/2}$

The factor $G(j\omega)$ is present in the denominator. To bring it in numerator, we will insert a negative sign. It is because log (1/a) is equal to -log a.

So, the gain will be:

$A = -20\ \log\ [(1 - (\omega\ /\ \omega_n)^4\ )+ 4a^2\ \omega^2/\ \omega_n^2]^{1/2}$

Let's find the value of gain at different frequencies.

The value of gain at very low frequencies ($\omega << \omega_n$) is equal to:

$A = -20\ \log\ [(1 - (\omega^2\ /\ \omega_n^2(2 - 4a^2)+ \omega^4/\ \omega_n^4]^{1/2}$

$A = -20\ \log 1 = 0$

The value of gain at very high frequencies ($\omega >> \omega_n$) is equal to:

$A = -20\ \log\ [(1 - (\omega^2\ /\ \omega_n^2(2 - 4a^2)+ \omega^4/\ \omega_n^4]^{1/2}$

$A = -20\ \log\ [\omega^4/\ \omega_n^4]^{1/2}$

$A = -20\ \log \omega^2\ /\ \omega_n^2$

$A = -20\ \log\ [\omega/\ \omega_n]^2$

We know that $\log a^2 = 2 \log a$

$A = -20 \times 2\log [\omega/ \omega_n]$

$A = -40 \log [\omega/ \omega_n]$

At, $\omega= \omega_n$

$A = -40 \log 1 = 0$ db

At, $\omega= 100\omega_n$

$A = -40 \log 100$

$A = -40 \log 10^2$

$A = -40 \times 2 \log 10$

$A = -80 \log 10 = -80$ db

At, $\omega= 10\omega_n$

$A = -40 \log 10 = -40$

Thus we can conclude that the magnitude plot of the quadratic factor can be approximated by two straight lines with a scale of 0 db and -40 db.

Let' find the value of angle G(jω) at different values of frequency.

Angle $G(j\omega) = \tan^{-1}[(2a\ j\omega/\omega_n)/(1 - (\omega / \omega_n)^2)$

At, $\omega = \omega_n$

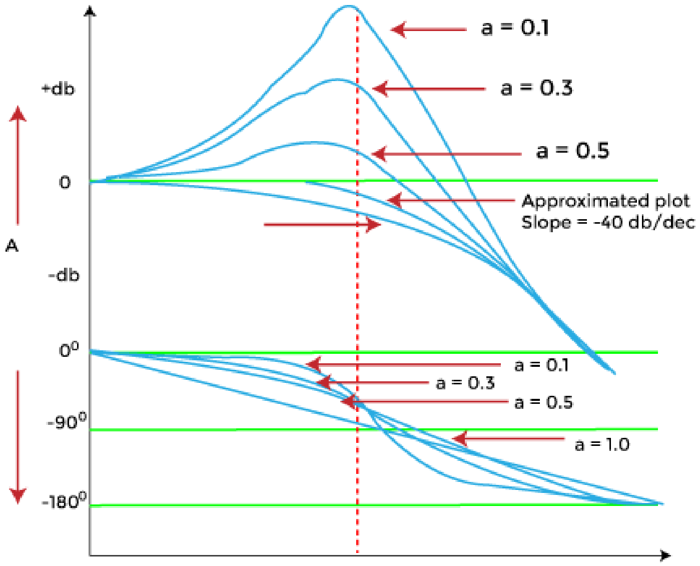
Angle $G(j\omega) = \tan^{-1}(2a/0) = -90$ degrees.

It is because the value of $\tan^{-1}$(infinity) is -90 degrees.

At $\omega = 0$, the value of angle is 0

At ω = infinity, the value of angle tends to -180 degrees.

The plot of the quadratic factor is denominator is shown below:



Bode plot of quadratic factor in denominator

The two curves are independent of the damping ratio a. The height of the resonant peak is inversely proportional to the damping ratio. It means that the decrease in the value of the damping ratio will increase the resonant peak.

Procedure to draw the magnitude curve and phase angle curve of the bode plot

Before proceeding, let's discuss the main conclusions obtained from the topic discussed above.

- The constant gain K, derivative, and integral factor in the bode plot contributes gain at all the frequencies.
- The higher order (two and more) factors in the transfer function contribute gain only when the frequency is greater than the corner frequency. The natural frequency in case of quadratic factor is assumed as corner frequency.

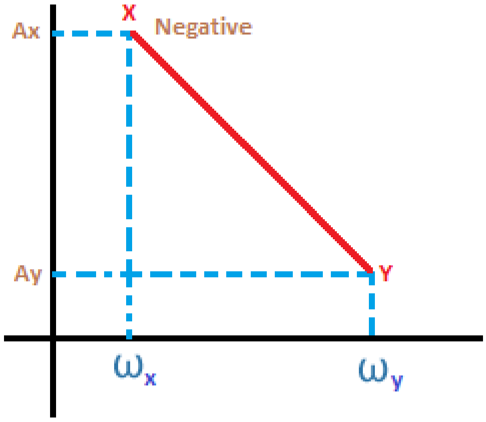
We have also discussed the impact of the derivative and integral factor on the positive and negative slope of the Bode plot.

Let's discuss the steps that will help us to draw the bode plot.

1. Convert the given transfer function into the time constant form, which is as follows: It helps us to easily find the corner frequencies.
2. Arrange the corner frequencies in the increasing order. The format of the table is shown below:

Term	Corner frequency	Slope (decibels)	Change in slope

3. The 'Term' specifies the separate constant gain, derivative factor, and the integral factor, such as, K, K/jω, or K (jω). The 'change in slope' includes the change in the value of the slope at each corer frequency.
4. Select a frequency, which has the lower value compared to the lowest corner frequency. Now, calculate the value of the terms at that selected frequency.
5. Find the gain in decibels one by one at every corner frequency. The formula to find the gain is given by:
   Gain at ω_y = **(Slope from ωx to ωy x log ωy/ωx] + gain at ωx**
   The frequencies ωx and ωy are shown below:



6. Now, we will perform step 4 again at a different frequency, which is greater than the highest corner frequency.
7. Take a semilog graph paper and first mark the dimension on the axis to draw the plot. The magnitude will appear on the y-axis and the frequency range will appear on the x-axis.
8. Mark the points calculated in the above steps on the semilog graph paper and join the points. The points on the graph paper should be joined only with straight lines. We can also mark the slope at every point on the graph.
9. The plot is assumed as the approximated plot. The exact plot requires the exact calculation at every corner frequency.

Consider an example.

Let the transfer function be $\mathbf{G(s) = K\,(1 + sT_1)^2/s^2(1 + sT_2)(\,1 + sT_3)}$

Put, s = jω

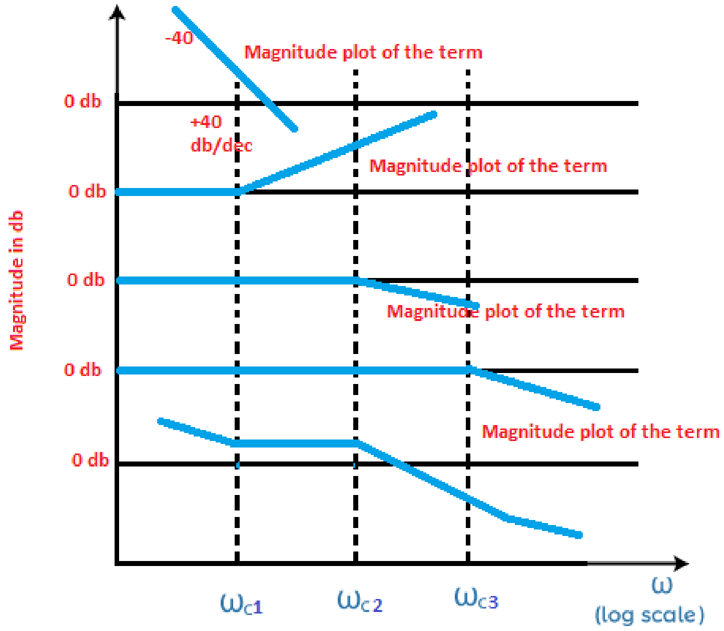
$G(j\omega) = K\,(1 + j\omega\, T_1)^2/(\,j\omega\,)^2(1 + j\omega T_2)(\,1 + j\omega T_3)$

The value of T1, T2, 3 in the increasing order will be T2 < T3 < T1.

There will be three corner frequencies at 1/T1, 1/T2, and 1/T3.

In the given transfer function $G(s) = K\,(1 + sT_1)^2/s^2(1 + sT_2)(\,1 + sT_3)$, there are four terms. Thus, we will first draw the magnitude plot of these four independent terms. The four terms are $K/\,j\omega$, $K(1 + sT_1)$, $1/(1 + sT_2)$, and $1/(\,1 + sT_3)$. The first, third, and the fourth term represent the integral term with the slope of -20db. The second term is the derivative term that represents the positive slope.

The final plot will be the combined plot of all the four plots, as shown below:



The above Bode plot shows that the first two terms have a slope of **40,** while the other two has a slope of **20.** The first two terms are in the form of squares in the transfer function, where log a2 is always equal to '**2 log a.**' Thus, it will appear with a slope of 40. As discussed above, the negative and positive terms with the slope are due to the integral and derivative factors.

Let's discuss how the value is combined.

A particular value at a region will be the result of the combined values at the above four sections. For example, the section between ωc1 and ωc3 will have the value by adding all section values (-40db + 40db + 0 + 0 = 0db).

Phase margin and gain margin of a Bode plot

The phase angle is calculated at various values of frequency. The frequencies are generally the same frequencies selected for plotting the magnitude section of a Bode plot. The phase and magnitude sections are drawn in a single semilog sheet with a common frequency scale for easy plotting. It also saves time.

The relation between the phase margin and the phase angle is given by:

Y = 180 + a

Where,

Y is the phase margin

a is the phase angle

It is calculated at the gain crossover frequency.

If the phase cross over frequency is negative, the gain margin of the system is positive. Similarly, if the phase cross over frequency is positive, the gain margin of the system is negative.

Let's discuss an example that will help us to clearly understand the concept of Bode plot, its steps, and the plotting on the semilog paper.

Example: Sketch the Bode plot of the following transfer function $G(s) = Ks^2/(1 + 0.2s)(1 + 0.02s)$.

Also determine the system gain K and the gain crossover frequency to be 5 radians/second.

Solution:

$G(s) = Ks^2/(1 + 0.2s)(1 + 0.02s)$

Step 1: Finding the corner frequencies

Here, we will first find the value of $G(j\omega)$.

$G(j\omega) = K\, j^2\omega^2/(1 + 0.2j\omega)(1 + 0.02j\omega)$

There will be two corner frequencies at T1 and T2, which are 0.2 and 0.02.

Thus, the corner frequencies are 1/0.2 = 5 radians/second and 1/0.02 = 50 radians/second.

Step 2: Preparing the table

As discussed in the above steps, we will prepare a table. There are three independent factors, $j^2\omega^2$, $1/(1 + 0.2j\omega)$, and $1/(1 + 0.02j\omega)$. So, the table will include three terms, as shown below:

Term	Corner frequency	Slope (decibels)	Change in slope
$j^2\omega^2$		+40db	
$1/(1 + 0.2j\omega)$	5 radians/second	-20db	40 - 20 = 20 db
$1/(1 + 0.02j\omega)$	50 radians/second	-20db	20 - 20 = 0 db

The first term will have a slope of 40 due to its square. The change in slope will include the sum of the slope from the top to bottom. The first row will have no change in the slope due to no term above it.

Step 3: We will find a frequency lower than the first corner frequency + a frequency higher than the second corner frequency.

The gain A = |G(jω)| = 20 log | $j^2\omega^2$|

We will find the value of gain at selected low frequency, two corner frequencies, and the high frequency.

Let's select the frequency 0.5, which is lower than the first corner frequency.

At, ω = low frequency, the value of gain is:

A = |G(jω)| = 20 log | $j^2\omega^2$|

A = 20 log | 0.5^2|

A = 20 log 0.25

A = 12 db

At, ω = first corner frequency (5 db), the value of gain is:

A = |G(jω)| = 20 log | $j^2\omega^2$|

A = 20 log | 5^2|

A = 20 log 25 or 40 log 5

A = 28 db

Now, at the second corner frequency and the selected higher corner frequency, we will use the formula:

Gain at ω_y **= (Slope from ωx to ωy x log ωy/ωx] + gain at ωx**

At ω_{c2} = (Slope from ωc1 to ωc2 x log ωc2/ωc1] + gain at ω (ω = ωc1)

Where,

ωc1 = first corner frequency

ωc2 = second corner frequency

Slope from ωc1 to ωc2 = change in slope of the second row

At, ω = second corner frequency (50 db), the value of gain is:

Gain = 20 log 50/5 + 28 = 20 log 10 + 28 = 20 + 28 = 48 db

At ω_h = (Slope from ωc2 to ωh x log ωh /ωc2] + gain at ω (ω = ωc2)

Where,

ωc1 = first corner frequency

ωc2 = second corner frequency

ωh = selected higher frequency

Slope from ωc1 to ωc2 = change in slope of the second row

Let the selected higher frequency be 100 db.

At, ω = second corner frequency (50 db), the value of gain is:

Gain = 0 log 100/50 + 28 = 0 log 2 + 48 = 0 + 48 = 48 db

Step 4: Phase plot

The phase angle of the given transfer function is:

Phase angle = 180 - $\tan^{-1} 0.2\omega$ - $\tan^{-1} 0.02\omega$

The term jω has an angle of 90 degrees. It square would result in the angle of 180 degrees.

Let's find the value of phase angles at different values of ω. It is shown in the below table:

ω radians/sec	$\tan^{-1} 0.2\omega$ degrees	$\tan^{-1} 0.02\omega$ degrees	Phase angle	Points for reference in the phase plot
0.5	5.7	0.6	173.7	E
1	11.3	1.1	167.6	F
5	45	5.7	129.3	G
10	63.4	11.3	105.3	H
50	84.3	45	50.7	I
100	87.1	63.4	29.5	J

We can consider the approximated value for the given values of the phase angle.

Step 5: Calculation of the value K

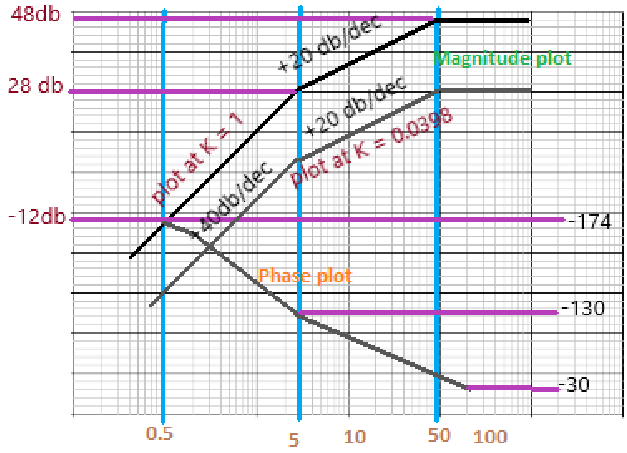
The value of gain K at frequency 5 radians per second is 28 db. Hence, at every point of magnitude plot is obtained by shifting the plot wit K = -1 by -28db downwards.

Thus, the term -28db is contributed by the term K.

20 log K = -28db

Log K = -28/20 = 0.0398

The magnitude plot and the phase plot are shown below:



Exact value of the plot

We know that the exact plot is 3 db less than the actual plot. Thus, for the exact plot, the term -28 db+ 3db = -25db is contributed by the term K.

20 log K = -25db

Log K = -25/20 = 0.0562

At last, let's discuss some of its advantages and disadvantages.

Advantages of the Bode Plot

The advantages of the Bode Plot are as follows:

- It covers low frequency as well as high frequency.
- The Bode plot provides the relative stability of the system in terms of the gain margin and phase margin.

- It can be drawn both for the closed loop system and open-loop system.
- It also provides us a method to improve the stability of the system.
- It converts the multiplication of magnitudes into addition.
- The sketching of bode plots is derived from a simple method.
- The information loss is very less.
- It consists of two graphs that eliminate the confusion between the magnitude and the phase plots.
- Bode plot is based on the asymptotic approximations that use straight line segments for plotting on the semilog graph.

Disadvantages of the Bode Plot

The disadvantages of the Bode Plot are as follows:

- The fundamental behavior of the system does not change frequently. Hence, the bode plot is not sensitive to the changes occurring in the measuring system.
- It does clearly predict the individual contribution of different factors in the given transfer functions.

Unit-9: State Variable Analysis

State Space Model

The process by which the state of a system is determined is called state variable analysis.

Advantages of State Space Techniques

1. This technique can be used for linear or nonlinear, time-variant or time-invariant systems.
2. It is easier to apply where Laplace transform cannot be applied.
3. The nth order differential equation can be expressed as 'n' equation of first order.
4. It is a time domain method.
5. As this is time domain method, therefore this method is suitable for digital computer computation.
6. On the basis of the given performance index, this system can be designed for an optimal condition.

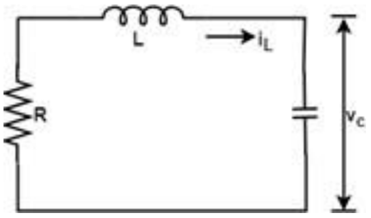
State Space representation of electrical system:

Consider an RLC network,

At time t = 0
Current = $i_L(0)$
Capacitor Voltage = $V_c(0)$

Thus, the state of the network at time t=0 is specified by the inductor current and capacitor voltage.

Therefore $i_L(0)$ and $V_c(0)$ is called the initial state of the network and the pair $i_L(t)$, $V_c(t)$ is called the state of the network at 't'. The variable iL and vc are called the state variables of the network.



Apply KVL

$$Ri_L + L\frac{di_L}{dt} + V_c = 0 \qquad \text{...........Eq.1}$$

Also,

$$i_c = i_L = C\frac{dV_c}{dt} \qquad \text{............Eq.2}$$

From eq.1

$$\frac{di_L}{dt} = -\frac{R}{L}i_L - \frac{1}{L}V_c \quad \text{...............Eq.3}$$

$$\frac{dv_c}{dt} = \frac{1}{C}i_L \quad \text{...............Eq.4}$$

This type of equation is called State equation. And the variables present in this equation are called state variables.

Eq. 3 and Eq. 4 can be written in matrix form as

$$\begin{bmatrix} \frac{di_L}{dt} \\ \frac{dv_c}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{R}{L} & -\frac{1}{L} \\ \frac{1}{C} & 0 \end{bmatrix} \begin{bmatrix} V_c \\ i_L \end{bmatrix} \qquad \text{.............................Eq.5}$$

Let $\quad x(t) \,\underline{\Delta}\, \begin{bmatrix} i_L \\ V_c \end{bmatrix}$ and $A = \begin{bmatrix} -\frac{R}{L} & -\frac{1}{L} \\ \frac{1}{C} & 0 \end{bmatrix}$

Then equation 5 can be written as

$$\frac{d}{dt}x(t) = A\,x(t)$$

Or, $\dot{x}(t) = Ax(t)$

The general form of state equation is

1. $\dot{x}(t)=Ax(t)+Bu(t)$ Eq.6
1. $y(t)=C\,x(t)+D\,u(t)$ Eq.7

Y = n- dimensional output vector

U = r-dimensional control vector or input vector

A = n × n system matrix

B = n × r control matrix

C = n × n output matrix

When there is no direct connection between input and output in that case D u(t) is not taken.

Y(t) = C x(t)

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \vdots \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots\dots\dots & a_{1n} \\ a_{21} & a_{22} & \dots\dots\dots & a_{2n} \\ \vdots & & & \\ \vdots & & & \\ a_{n_m} & & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots\dots & b_{1r} \\ \vdots & & & \\ b_{n1} & & & b_{nr} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_n \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \dots\dots & C_{1n} \\ \vdots & & & \\ C_{m1} & \dots. & \dots\dots & C_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

State Space Representation of nth order differential Equation

For the nth order differential equation

Example 1

A system is described by the differential equation

$$\boldsymbol{\frac{d^3y}{dt^3} + 6\frac{d^2y}{dt^2} + 11\frac{dy}{dt} + 10y = 8u(t)}$$

Where y is the output and u is the input to the system. Obtain the state space representation of the system.

Solution

$x_1 = y$, $x_2 = \dot{y}$ and $x_3 = \ddot{y}$ then

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = x_3$$
$$\dot{x}_3 = 8\,u(t) - 10x_1 - 11x_2 - 6x_3$$

And the last equation will be

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -10 & -11 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 8 \end{bmatrix} u(t)$$

Now comparing Eq.6 and above equation

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -10 & -11 & -6 \end{bmatrix},\; B = \begin{bmatrix} 0 \\ 0 \\ 8 \end{bmatrix},\; x(t) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

State variable analysis: Examples with explanation

Let's first discuss a small introduction about the state variables before proceeding with the examples.

State means a **set of variables** whose knowledge helps us to predict the **behaviour** of the control system. The state variable with the input functions provides the future state and the output of the system. The elements of the state variable analysis include **state vector, state space, state equations,** and the **state variable representation.**

The state variables are the n variables needed to describe the system's behaviour, and the state axis is the area where coordinate axes are described. Similarly, state equations are the set of n equations, and state variable representation is described in terms of state variables, such as x1(t), x2(t), and so on. The system's input, output, initial conditions, etc., is described in terms of these elements.

Here, we will discuss the examples in the form of Multiple Choice Questions because they will help us in the competitive exams.

Example 1: A system is described by the state equation X' = AX + BU. The output is given by Y = CX, where

$$A = \begin{bmatrix} -4 & -1 \\ 3 & -1 \end{bmatrix} \text{ and } B = \begin{matrix} 1 \\ 1 \end{matrix}$$

And C = [1 0]

The transfer function of the system is:

1. $s/s^2 + 5s + 7$
2. $2s/s^2 + 5s + 7$
3. $1/s^2 + 5s + 7$
4. $s/s^2 + 3s + 5$

Answer: (a) $s/s^2 + 5s + 7$

Explanation: The equation to find the transfer function of the system is given by:

$C\,[SI - A]^{-1}B + D$

Where,

I is always an identity matrix.

Computing the values of A, B, I, and C in the above equation, we get:

Transfer function = $C\,[SI - A]^{-1}B + D$

$$= [1\ 0]\begin{bmatrix} s+4 & 1 \\ -3 & s+1 \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [1\ 0]\,\frac{1}{[(s+4)(s+1)+3]}\begin{bmatrix} s+4 & 1 \\ -3 & s+1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [1\ 0]\,\frac{1}{[s^2+5s+4+3]}\begin{bmatrix} s+4 & 1 \\ -3 & s+1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [1\ 0]\,\frac{1}{[s^2+5s+7]}\begin{bmatrix} s+4 & 1 \\ -3 & s+1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= [1\ 0]\,\frac{1}{[s^2+5s+7]}\begin{bmatrix} s \\ s+7 \end{bmatrix}$$

$= s/s^2 + 5s + 7$

Thus, the correct answer is an option (a).

Example 2: Find the sum of the Eigen values of the matrix given below:

$$\begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ -1 & 0 & 0 & 4 \end{bmatrix}$$

1. 15
2. 12
3. -10
4. 10

Answer: (d) 10

Explanation: The shortcut approach to find the sum of the Eigen values is to add the elements present in the diagonal of the given matrix.

The diagonal elements are 2, 1, 3, and 4. It sum is equal to 2 + 1 + 3 + 4 = 10

Hence, the correct answer is 10.

Example 3: A system represented by the equations is given by:

$$X' = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

And Y = [1 0]

Find the equivalent transfer function of the system.

1. $s/s^2 + 3s + 3$
2. $1/s^2 + 3s + 2$
3. $5/s^2 + 5s + 2$
4. $2/s^2 + 3s + 5$

Answer: (b) $1/s^2 + 3s + 2$

Explanation: The equation to find the transfer function of the system is given by:

$C\,[SI - A]^{-1}B + D$

Where,

I is always an identity matrix.

Computing the values of A, B, I, and C in the above equation, we get:

Transfer function = C [SI - A]$^{-1}$B + D

$$= [1\ 0]\begin{bmatrix} s & 1 \\ -2 & s+3 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= [1\ 0]\frac{1}{[s(s+3)+2]}\begin{bmatrix} s & 1 \\ -2 & s+3 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= [1\ 0]\frac{1}{[s^2+3s+2]}\begin{bmatrix} s & 1 \\ -2 & s+3 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$= 1/s^2 + 3s + 2$

Example 4: The observability condition in the state space representation can be determined with the condition given by:

1. $[C^T\ A^TC^T]$
2. $[B^T\ A^TB^T]$
3. $[A^T\ A^TC^T]$
4. None of these

Answer: (a) $[C^T\ A^TC^T]$

Explanation: The observability condition can be determined from the condition $[C^T\ A^TC^T]$.

Example 5: The system model given by the equation is:

$$X' = \begin{bmatrix} 0 & 1 \\ 2 & -3 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

Y '= [1 1] x

1. Controllable
2. Observable
3. Both (a) and (b)
4. Neither (a) nor (b)

Answer: (c) Both (a) and (b)

Explanation: We know that the equation is represented in the form of AX + BU.

By comparing, we get:

$$A = \begin{bmatrix} 0 & 1 \\ 2 & -3 \end{bmatrix} x \text{ and } B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

C = [1 1]

The controllability of the given state equation can be checked by forming the matrix [B AB]. The observability of the given state equation can be checked by forming the matrix [C^T A^TC^T]. If the determinant of these matrixes is not equal to zero, it is said to observable or controllable.The matrix so formed for the controllability is given by:

[B AB]

$$= \begin{bmatrix} 0 & 1 \\ 1 & -3 \end{bmatrix}$$

The determinant of the above matrix is: 0 x -3 - 1 = -1

It is not equal to 0. Hence, the matrix is controllable.

Now, let's check for observability.

The matrix so formed for the observability is given by:

[C^T A^TC^T]

$$= \begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix}$$

The determinant of the above matrix is: 1 x -2 - 2 = -4

It is not equal to 0. Hence, the matrix is observable.

Thus, the system model described by the given state equation is both controllable and observable.

Example 6: The state of a system equation can be written in the form of:

1. First order differential equation
2. Second order differential equation
3. Third order differential equation
4. None of the above

Answer: (a) First order differential equation

Explanation: The state of a system equation is written in the form of first order differential equation using the state variables x1, x2, x3 ? xn.

Example 7: A LTI system is described by the state model given by:

$$X' = \begin{bmatrix} x1 \\ x2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, y = [1\ 2] \begin{bmatrix} x1 \\ x2 \end{bmatrix}$$

Find if the system is:

1. Completely controllable
2. Completely observable
3. Not completely controllable but completely observable
4. Not completely observable but completely controllable

Answer: (c) Not completely controllable but completely observable

Explanation: The controllability of the given state equation can be checked by forming the matrix [B AB]. The observability of the given state equation can be checked by forming the matrix [C^T A^TC^T].

We will use the above two conditions to determine controllability and observability.

We can check either the determinant or the rank of the formed matrix. If the determinant of these matrixes is not equal to zero, it is said to observable or controllable. If the rank of the matrix is equal to its order, it is said to be observable or controllable.

Let's start.

By comparing the given state model, the parameters are:

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$$

$$A^T = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$C^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Now, the matrix [B AB] is given by:

$$\begin{bmatrix} 0 & 0 \\ 1 & -2 \end{bmatrix}$$

The determinant of the above matrix is equal to 0. Hence, it is not completely controllable. We can also say that the rank of the matrix is not equal to its order (2).

Now, the matrix [C^T $A^T C^T$] is given by:

$$\begin{bmatrix} 1 & -1 \\ 2 & -4 \end{bmatrix}$$

The determinant of the above matrix is not equal to 0. Hence, it is completely observable. We can also say that the rank of the matrix is equal to its order (2).

Thus, the given state model is completely observable but not completely controllable.

References:

1. https://www.tutorialspoint.com/
2. https://www.javatpoint.com/
3. Control Systems by Principles and Design by Gopal. M,-Tata McGraw-Hill
4. Control System Engg by P Ramesh Babu& R. Anandanatarajan –SCITECH
5. Modern Control Engineering by Nagrath& Gopal-New Age International, New Delhi