GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)

Seshadri Rao Knowledge Village, Gudlavalleru – 521 356.

Department of Electrical and Electronics Engineering

CONTROL SYSTEMS

Unit - I

Introduction

Objectives:

To introduce the students to basic concepts of control systems, Classification of control systems, Feedback characteristics, Mathematical models – differential equations, Transfer function of different types of mechanical and electrical systems

Syllabus:

Concepts of Control Systems- Open Loop and closed loop control systems and their differences-Different examples of control systems- Classification of control systems, Feed-Back Characteristics, Effects of feedback. Mathematical models – Differential equations, Impulse Response and transfer functions.

Outcomes:

Students will be able to

- Determine the Mathematical model of any system using basic physical laws.
- Determine feedback characteristics of any given system
- Determine the Factors effecting feedback of the system
- Determine the Transfer function of Physical systems.

1.1Introduction

A control system is a system of devices or set of devices, that manages, commands, directs or regulates the behavior of other device(s) or system(s) to achieve desire results. In other words the **definition of control system** can be rewritten as a control system is a system, which controls other system. As the human civilization is being modernized day by day the demand of automation is increasing accordingly. Automation highly requires control of devices. In recent years, **control systems** plays main role in the development and advancement of modern technology and civilization. Practically every aspects of our day-to-day life is affected less or more by some control system. A refrigerator, an air conditioner, a geezer, an automatic iron, an automobile all are control system. These systems are also used in industrial process for more output. We find control system in quality control of products, weapons system, transportation systems, power system, space technology, robotics and many more. The **principles of control theory** are applicable to engineering and non engineering field both.

Requirement of Good Control System

Accuracy: Accuracy is the measurement tolerance of the instrument and defines the limits of the errors made when the instrument is used in normal operating conditions. Accuracy can be improved by using feedback elements. To increase accuracy of any control system error detector should be present in control system.

Sensitivity: The parameters of control system are always changing with change in surrounding conditions, internal disturbance or any other parameters. This change can be expressed in terms of sensitivity. Any control system should be insensitive to such parameters but sensitive to input signals only.

Noise: An undesired input signal is known as noise. A good control system should be able to reduce the noise effect for better performance.

Stability: It is an important characteristic of control system. For the bounded input signal, the output must be bounded and if input is zero then output must be zero then such a control system is said to be stable system.

Bandwidth: An operating frequency range decides the bandwidth of control system. Bandwidth should be large as possible for frequency response of good control system.

Speed: It is the time taken by control system to achieve its stable output. A good control system possesses high speed. The transient period for such system is very small.

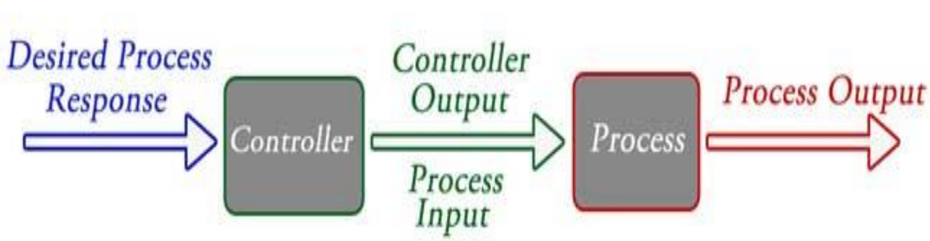
Oscillation: A small numbers of oscillation or constant oscillation of output tend to system to be stable.

1.2 Open Loop and Closed Loop Control Systems

Open Loop Control Systems*:*

A system in which the output has no effect on the control action is known as an open loop control system. For a given input the system produces a certain output. If there are any disturbances, the output changes and there is no adjustment of the input to bring back the output to the original value. A perfect calibration is required to get good accuracy and the system should be free from any external disturbances. No measurements are made at the output. A traffic control system is a good example of an open loop system. The signals change according

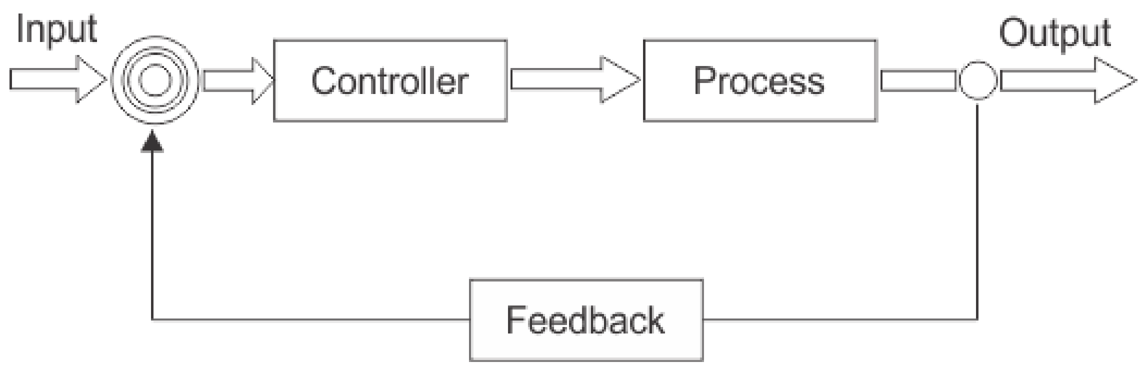
to a preset time and are not affected by the density of traffic on any road. A washing machine is another example of an open loop control system. The quality of wash is not measured; every cycle like wash, rinse and dry' cycle goes according to a preset timing.



Open Loop control system

Closed Loop Control Systems*:*

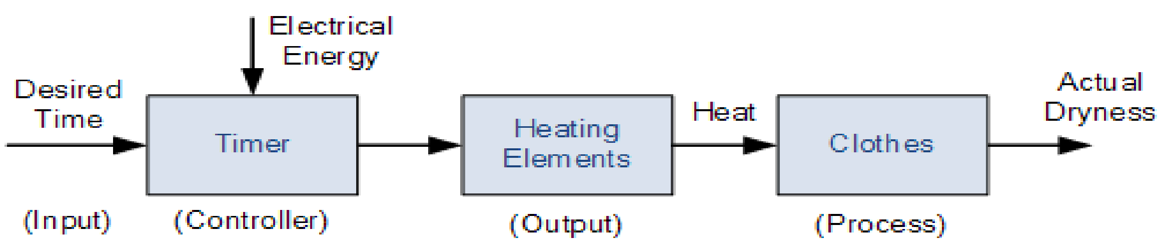
These are also known as feedback control systems. A system which maintains a prescribed relationship between the controlled variable and the reference input, and uses the difference between them as a signal to activate the control, is known as a feedback control system. The output or the controlled variable is measured and compared with the reference input and an error signal is generated. This is the activating signal to the controller which, by its action, tries to reduce the error. Thus, the controlled variable is continuously fedback and compared with the input signal. If the error is reduced to zero, the output is the desired output and is equal to the reference input signal.



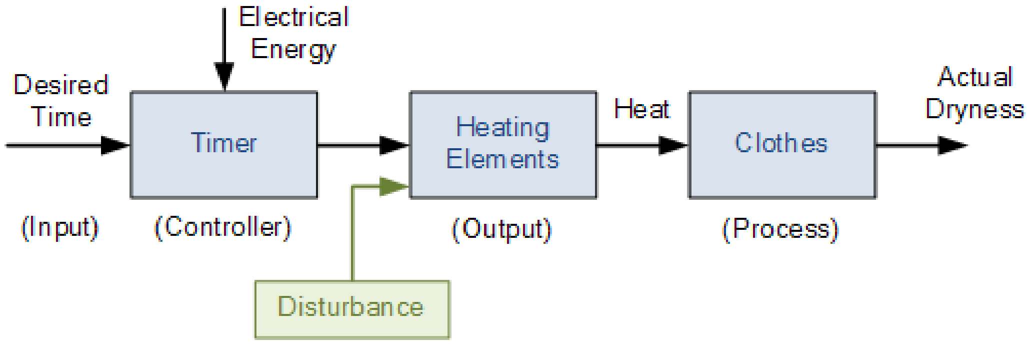
Closed Loop control system

1.2.1 Examples of control system:

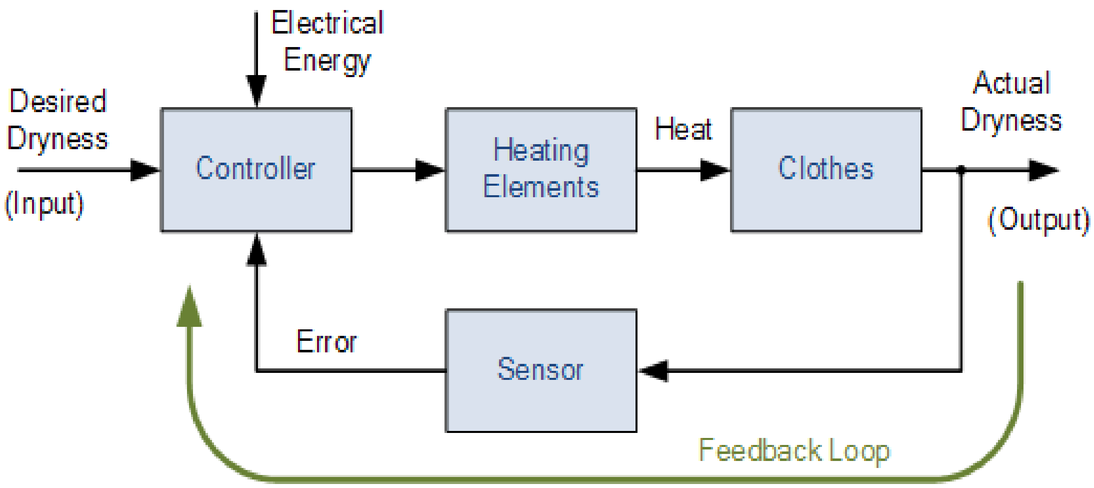
a) an electric clothes dryer: Depending upon the amount of clothes or how wet they are, a user or operator would set a timer (controller) to say 30 minutes and at the end of the 30 minutes the drier will automatically stop and turn-off even if the clothes are still wet or damp. In this case, the control action is the manual operator assessing the wetness of the clothes and setting the process (the drier) accordingly. In this example, the clothes dryer would be an open-loop system as it does not monitor or measure the condition of the output signal, which is the dryness of the clothes. So the accuracy of the drying process or success of drying the clothes will depend on the experience of the user (operator).



So an **Open-loop system**, also referred to as non-feedback system, is a type of continuous control system in which the output has no influence or effect on the control action of the input signal. In other words, in an open-loop control system the output is neither measured nor “fed back” for comparison with the input. Also, an open-loop system has no knowledge of the output condition so cannot self-correct any errors it could make when the preset value drifts, even if this results in large deviations from the preset value. Another disadvantage of open-loop systems is that they are poorly equipped to handle disturbances or changes in the conditions which may reduce its ability to complete the desired task. For example, the dryer door opens and heat is lost. The timing controller continues regardless for the full 30 minutes but the clothes are not heated or dried at the end of the drying process. This is because there is no information fed back to maintain a constant temperature.



So for example, consider our electric clothes dryer from the previous open-loop systems. Suppose we used a sensor or transducer (input device) to continually monitor the temperature or dryness of the clothes and feed a signal relating to the dryness back to the controller as shown below.



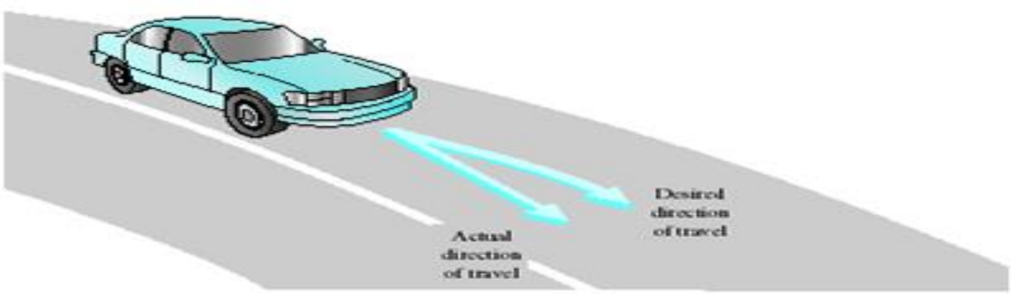
This sensor would monitor the actual dryness of the clothes and compare it with (or subtract it from) the input reference. The error signal (error = required dryness – actual dryness) is amplified by the controller, and the controller output makes the necessary correction to the heating system to reduce any error. For example if the clothes are too wet the controller may increase the temperature or drying time. Likewise, if the clothes are nearly dry it may reduce the temperature or stop the process so as not to overheat or burn the clothes, etc. Also, because a closed-loop system has some knowledge of the output condition, (via the sensor) it is better equipped to handle any system disturbances or changes in the conditions which may reduce its ability to complete the desired task. For example, as before, the dryer door opens and heat is lost. This time the deviation in temperature is detected by the feedback sensor and the controller self-corrects the error to maintain a constant temperature within the limits of the preset value. Or possibly stops the process and activates an alarm to inform the operator.

b) Driving a car implies controlling the vehicle to follow the desired path to arrive safely at a planned destination.

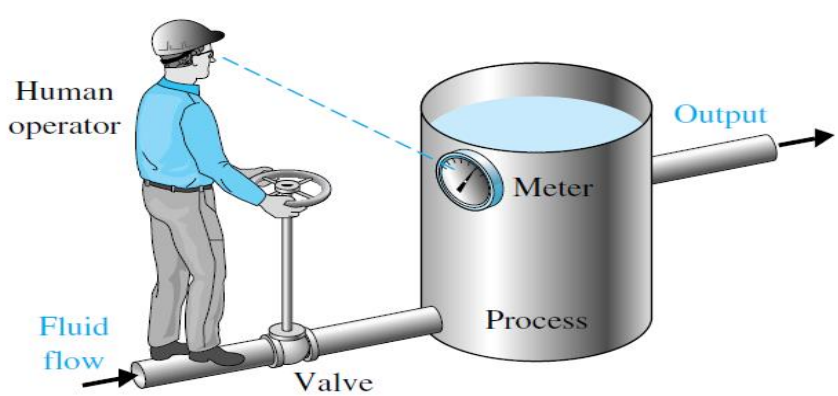
i. If you are driving the car yourself, you are performing manual control of the car.

ii. If you use design a machine, or use a computer to do it, then you have built an automatic control system.

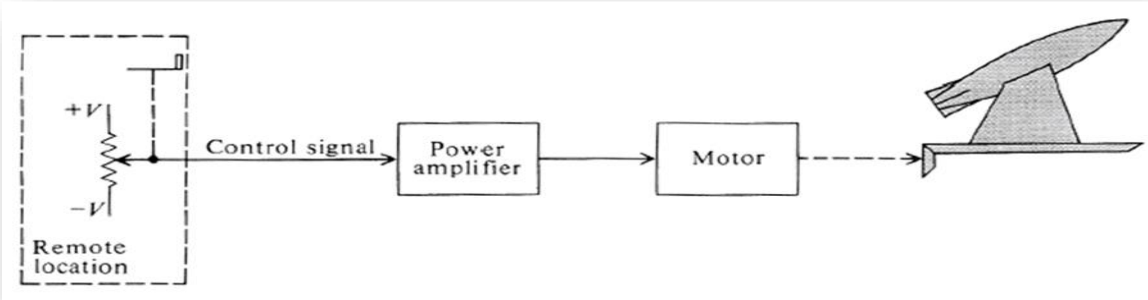
c) Automobile steering control system: The driver uses the difference between the actual and the desired direction of travel to generate a controlled adjustment of the steering wheel.

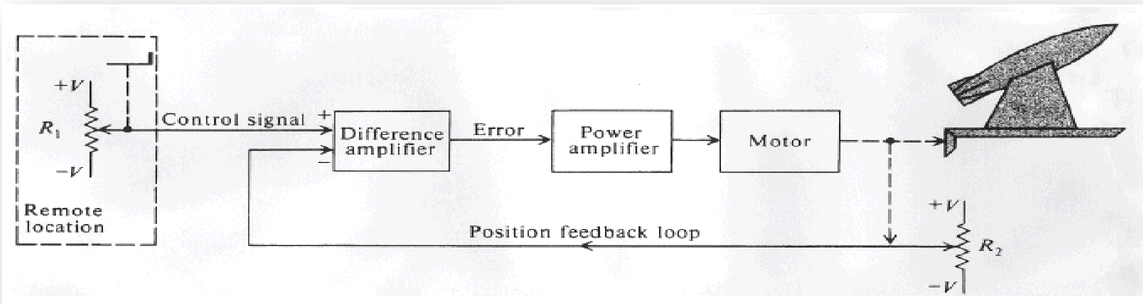


d) A manual control system for regulating the level of fluid in a tank by adjusting the output valve. The operator view the level of fluid through a port in the side of the tank.



e)Missile Launcher System:





1.2.2. Open Loop Vs Closed Loop Control Systems

The open loop systems are simple and easier to build. Open loop systems are cheaper and they should be preferred whenever there is a fixed relationship between the input and the output and there are no disturbances. Accuracy is not critical in such systems. Closed loop systems are more complex, use more number of elements to build and are costly. The stability is a major concern for closed loop systems. We have to ensure that the system is stable and will not cause undesirable oscillations in the output. The major advantage of closed loop system is that it is insensitive to external disturbances and variations in parameters. Comparatively cheaper components can be used to build these systems, as accuracy and tolerance do not affect the performance. Maintenance of closed loop systems is more difficult than open loop systems. Overall gain of the system is also reduced.

Open Loop Systems

Advantages

1. They are simple and easy to build.
2. They are cheaper, as they use less number of components to build.
3. They are usually stable.
4. Maintenance is easy.

Disadvantages

1. They are less accurate.
2. If external disturbances are present, output differs significantly from the desired value.

3. If there are variations in the parameters of the system, the output changes.

Closed Loop Systems

Advantages

1. They are more accurate.
2. The effect of external disturbance signals can be made very small.
3. The variations in parameters of the system do not affect the output of the system i.e. the output may be made less sensitive to variation is parameters. Hence forward path components can be of less precision. This reduces the cost of the system.
4. Speed of the response can be greatly increased.

Disadvantages

1. They are more complex and expensive
2. They require higher forward path gains.
3. The systems are prone to instability. Oscillations in the output many occur.
4. Cost of maintenance is high.

Comparison of Closed Loop And Open Loop Control System:

S. No.	Open loop control system	Closed loop control system
1	The feedback element is absent.	The feedback element is always present.
2	An error detector is not present.	An error detector is always present.
3	It is stable one.	It may become unstable.
4	Easy to construct.	Complicated construction.
5	It is an economical.	It is costly.
6	Having small bandwidth.	Having large bandwidth.
7	It is inaccurate.	It is accurate.
8	Less maintenance.	More maintenance.
9	It is unreliable.	It is reliable.
10	Examples: Hand drier, tea maker	Examples: Servo voltage stabilizer, perspiration

Transfer Function

The transfer function of a control system is defined as the ratio of the Laplace transform of the output variable to Laplace transform of the input variable assuming all initial conditions to be zero.

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

Procedure for determining the transfer function of a control system is as follows

1. We form the equations for the system

2. Now we take Laplace transform of the system equations, assuming initial conditions as zero.

3. Specify system output and input

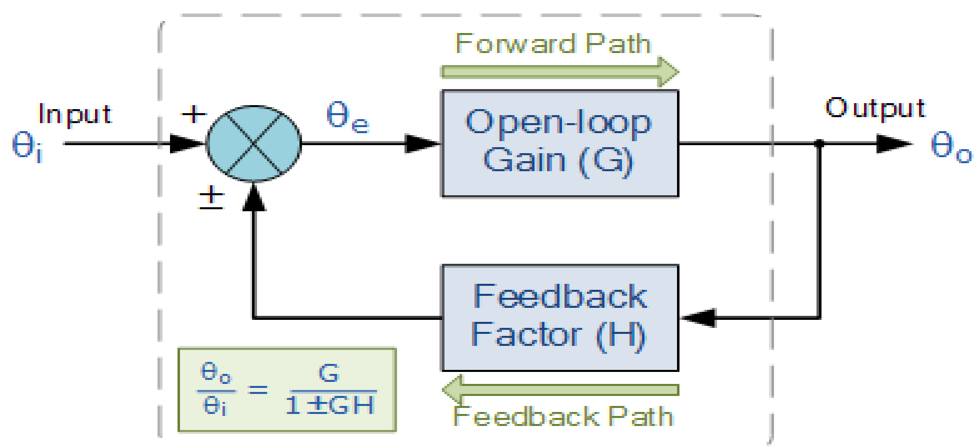
4. At the last we take the ratio of the Laplace transform of the output and the Laplace transform of the input which is the required transfer function

Methods of obtaining a Transfer function: There are major two ways of obtaining a transfer function for the control system. The ways are –

- **Block diagram method:** It is not convenient to derive a complete transfer function for a complex control system. Therefore the transfer function of each element of a control system is represented by a block diagram. Block diagram reduction techniques are applied to obtain the desired transfer function.

- **Signal Flow graphs:** The modified form of a block diagram is a signal flow graph. Block diagram gives a pictorial representation of a control system. Signal flow graph further shortens the representation of a control system.

Feedback System Block Diagram Model



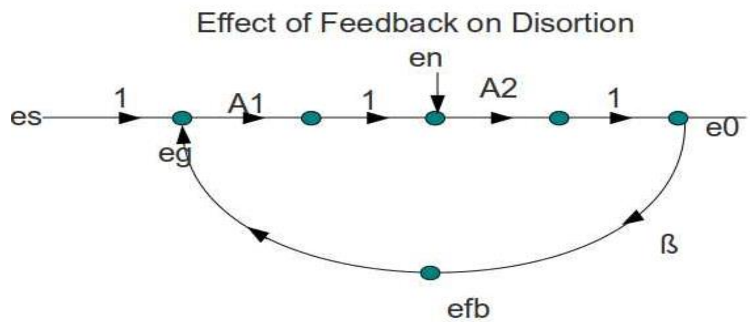
This basic feedback loop of sensing, controlling and actuation is the main concept behind a feedback control system and there are several good reasons why feedback is applied and used in electronic circuits:

- Circuit characteristics such as the systems gain and response can be precisely controlled.
- Circuit characteristics can be made independent of operating conditions such as supply voltages or temperature variations.
- Signal distortion due to the non-linear nature of the components used can be greatly reduced.
- The Frequency Response, Gain and Bandwidth of a circuit or system can be easily controlled to within tight limits.

Whilst there are many different types of control systems, there are just two main types of feedback control namely: **Negative Feedback** and **Positive Feedback**.

Effects of Feedback:

The effects of feedback in systems on other working parameters of the system has been analyzed. The parameter considered for analysis are gain, sensitivity, distortion, impedance and bandwidth. A feedback system has been shown in the figure. Effect of feedback on overall Gain:



From figure it is seen that the transfer function is given by the equation:

$$M = \frac{A}{(1-bA)}$$

Hence the feedback reduces the overall gain of the system by a factor of **(1-bA).**

The quantities A and B are function of frequency and can be adjusted to make the denominator greater than unity. Hence the gain increases for a particular frequency range and decreases for another frequency range.

Effect of Feedback on Sensitivity: Sensitivity is the extent to which the system responds to changes in parameters like gain, impedance, etc. Sensitivity is also said to be the ratio of the extent of change of one of the above mentioned parameter to a small change of the determining parameters. For example, if M= transfer function K= Determining Parameter Then the sensitivity (S) is given by: S= Percentage change in M/ Percentage change in K Following are the effect of feedback on Sensitivity. Feedback may reduce sensitivity with respect to certain parameters. Feedback does not affect variations of elements in the feedback path. Feedback reduces the sensitivity of the system based on variation of parameter in the forward path of the loop. Larger the loop gain 'Ab', more effective is the feedback in reducing sensitivity.

Effect of Feedback on Distortion: Feedback is used in communication systems to reduce noise and other distortion signals which it might pickup from extraneous sources. The place of insertion of the extraneous noise to the signal flow is the main factor that determines the extent to which the feedback reduces the effects due to distortion. Consider the figure which shows a signal flow graph of a system. A noise signal is inserted at the point shown. In the absence of the feedback, the output is given by

$$e_0 = A_1A_2es + A_2en = e_0s + e0n$$

where, e_0s = single component of the output. e_0n=The component of the output due to noise.

Output signal to noise ration = output due to signal/output due to noise.

Signal to noise ratio = $\frac{A_1A_2e_s}{A_2e_n} = \frac{A_1e_s}{e_n}$

Hence to increase the signal to noise ratio, either A1 and /or es is to be increased or en is to be decreased. If the system is aided by a feedback circuit, the output is given by:

$$e_0 = \frac{A_1A_2e_s}{(1 - A_1A_2b)} + \frac{A_2e_n}{(1 - A_1A_2b)}$$

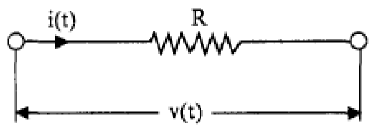
From the above equation it is clear that the noise component of the output has its gain reduced by a factor $(1 - A_1A_2b)$. Thus the noise is reduced and the overall distortion of the output is reduced.

Modeling of Electric systems, Translational and Rotational mechanical systems:

Electrical Systems:

Most of the electrical systems can be modelled by three basic elements : Resistor, inductor, and capacitor. Circuits consisting of these three elements are analysed by using Kirchhoff's Voltage law and Current law.

(a) Resistor*:* The circuit model of resistor is shown in Fig.

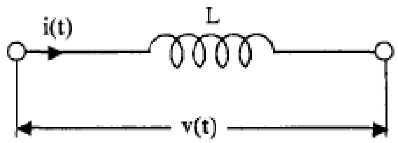


The mathematical model is given by the Ohm's law relationship,

$$V(t) = i(t)\,R$$

$$i(t) = V(t)/R$$

(b) Inductor: The circuit representation is shown in Fig.



The input output relations are given by Faraday's law,

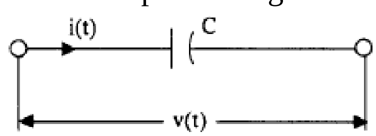
$$V(t) = L\,di(t)/dt$$

$$i(t) = \frac{1}{L}\int v\,dt$$

where Integral of v dt is known as the flux linkages. Thus

$$i(t) = \frac{\Psi(t)}{L}$$

(c) Capacitor: The circuit symbol of a capacitor is given in Fig.



$$v(t) = \frac{1}{C}\int i\,dt$$

$$i(t) = C\,\frac{dv}{dt}$$

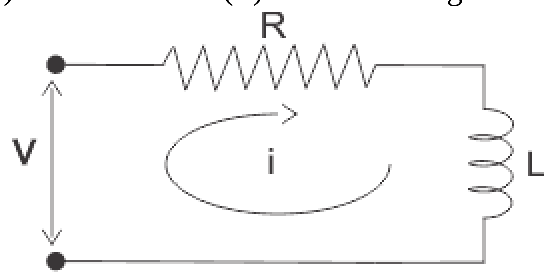
In eqn. idt is known as the charge on the capacitor and is denoted by *'q'*. Thus

$$q = \int i\,dt$$

$$v(t) = \frac{q(t)}{C}$$

Modeling of RL circuit:

The transfer function is generally expressed in Laplace Transform and it is nothing but the relation between input and output of a system. Let us consider a system consists of a series connected resistance (R) and inductance (L) across a voltage source (V).



In this circuit, the current 'i' is the response due to applied voltage (V) as cause. Hence the voltage and current of the circuit can be considered as input and output of the system respectively. From the circuit, we get, $V = iR + L\frac{di}{dt}$

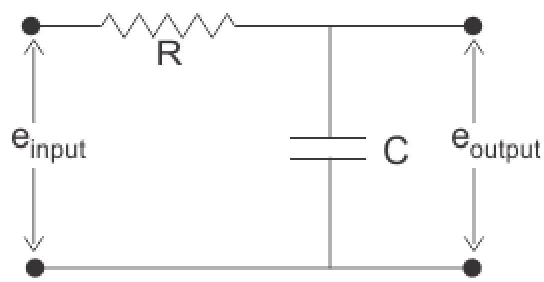
Now applying Laplace Transform, we get, $V(s) = RI(s) + L[sI(s) - i(0^{+})]$

$$[\because Initially\ inductor\ behaves\ as\ open, hence\ i(0^{+}) = 0]$$

$$\Rightarrow V(s) = I(s)[R + Ls]$$

$$\Rightarrow \frac{I(s)}{V(s)} = \frac{1}{R + Ls} = \frac{1/L}{s + R/L}$$

Modeling of RC circuit:



In the above network it is obvious that

$$e_{input} = Ri(t) + \frac{1}{C}\int i(t)dt$$

$$e_{output} = \frac{1}{C}\int i(t)dt$$

Let us assume, $\mathcal{L}[e_{input}(t)] = E_{input}(s), \mathcal{L}[e_{output}(t)] = E_{input}(s),\ \mathcal{L}[i(t)] = I(s)$

$$And\ then, \mathcal{L}\left[\int i\,(t)\right] = \frac{I(s)}{s} + \int i(0)dt = \frac{I(s)}{s} + 0 = \frac{F(s)}{s}$$

$$\left[\int i(0)dt = 0, as\ there\ is\ no\ curremt\ initially\ through\ the\ capacitor\right]$$

Taking the Laplace transform of above equations with considering the initial condition as zero, we get,

$$E_{input}(s) = RI(s) + \frac{1}{C}\cdot\frac{I(s)}{s} = I(s)\left[R + \frac{1}{Cs}\right]$$

$$E_{output}(s) = \frac{1}{C}\cdot\frac{I(s)}{s} = I(s)\frac{1}{Cs}$$

$$Transfer\ function\ of\ the\ network,$$

$$\frac{E_{output}(s)}{E_{input}(s)} = \frac{I(s)\frac{1}{Cs}}{I(s)\left[R + \frac{1}{Cs}\right]} = \frac{\frac{1}{Cs}}{\left[R + \frac{1}{Cs}\right]} = \frac{1}{RCs + 1}$$

Mechanical Systems:

Mechanical systems can be divided into two basic systems.

(a) Translational systems and (b) Rotational systems

We will consider these two systems separately and describe these systems in terms of three fundamental linear elements.

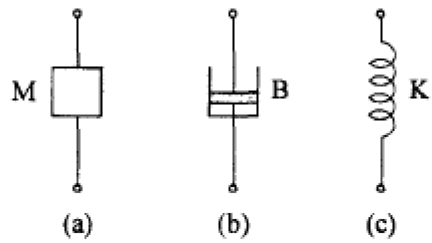
(a) Translational systems:

1. Mass: This represents an element which resists the motion due to inertia. According to Newton's second law of motion, the inertia force is equal to mass times acceleration.

$$f_M = M\,a = M\,(dv/dt) = M\,(dx^2/\,dt^2)$$

Where a, v and x denote acceleration, velocity and displacement of the body respectively.

Symbolically, this element is represented by a block as shown in fig . (a)



Passive linear elements of translational motion (a) Mass (b) Dash pot (c) Spring.

***2*. Dash pot:** This is an element which opposes motion due to friction. If the friction is viscous friction, the frictional force is proportional to *velocity*. This force is also known as damping force..

Thus we can write $f_B = Bv = B\,(dx/dt)$

Where B is the damping coefficient. This element is called as dash pot and is symbolically represented as in Fig.(b)

3. Spring: The third element which opposes motion is the spring. The restoring force of a spring is proportional to the displacement.

Thus $f_K = K\,x$

Where K is known as the stiffness of the spring or simply spring constant. The symbol used for this element is shown in Fig.(c)

(b) Rotational systems: Corresponding to the three basic elements of translation systems, there are three basic elements representing rotational systems.

1. Moment of Inertia: This element opposes the rotational motion due to Moment of inertia. The opposing inertia torque is given by,

$$T_I = Ja = J\frac{d\omega}{dt} = J\frac{d^2\theta}{dt^2}$$

Where *a, ω* and Θ are the angular acceleration, angular velocity and angular displacement respectively. J is known as the moment of inertia of the body.

2. Friction: The damping or frictional torque which opposes the rotational motion is given by,

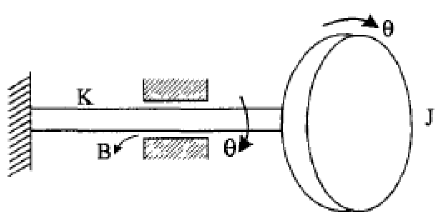
$$T_B = B\omega = B\frac{d\theta}{dt}$$

Where B is the rotational frictional coefficient.

3. Spring: The restoring torque of a spring is proportional to the angular displacement () and is given by,

$$T_K = K\theta$$

Where K is the stiffness of the spring. The three elements defined above are shown in fig below.

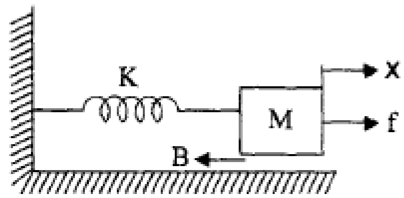


Rotational elements

Since the three elements of rotational systems are similar in nature to those of translational systems no separate symbols are necessary to represent these elements. Having defined the basic elements of mechanical systems, we must now be able to write differential equations for the system when these mechanical systems are subjected to external forces. This is done by using the **D' Alembert's** principle which is similar to the **Kirchhoff's laws in Electrical Networks**. Also, this principle is a modified version of Newton's second law of motion.

The D' Alembert's principle states that, "For anybody, the algebraic sum of externally applied forces and the forces opposing the motion in any given direction is zero". To apply this principle to anybody, a reference direction of motion is first chosen. All forces acting in this direction are taken positive and those against this direction are taken as negative. Let us apply this principle to a mechanical translation system shown in Fig.

A mass M is fixed to a wall with a spring K and the mass moves on the floor with a viscous friction. An external force f is applied to the mass. Let us obtain the differential equation governing the motion of the body.



A mechanical translational system

Let us take a reference direction of motion of the body from left to right. Let the displacement of the mass be *x*. We assume that the mass is a rigid body, ie, every particle in the body has the same displacement, *x*. Let us enumerate the forces acting on the body.

(a) external force = f

(b) resisting forces :

(i) Inertia force, $f_M = - M (d^2x/dt^2)$(ii) Damping force, $f_B = - B\ dx/dt$

(iii) Spring force, $f_K = - Kx$

Resisting forces are taken to be negative because they act in a direction opposite to the chosen reference direction. Thus, using D' Alemberts principle we have,

$$f - M\frac{d^2x}{dt^2} - B\frac{dx}{dt} - Kx \quad = 0$$

$$M\frac{d^2x}{dt^2} + B\frac{dx}{dt} + Kx \quad = f$$

This is the differential equation governing the motion of the mechanical translation system. The transfer function can be easily obtained by taking Laplace transform of above

$$\frac{X(s)}{F(s)} = \frac{1}{Ms^2 + Bs + K}$$

If velocity is chosen as the output variable,

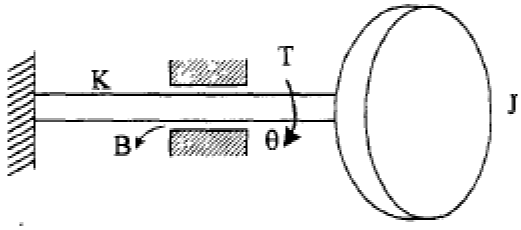
$$M\frac{du}{dt} + Bu + K\int u\,dt = f$$

Similarly, the differential equation governing the motion of rotational system can also be obtained. For the system in the following Fig., we have

$$J\frac{d^2\theta}{dt^2} + B\frac{d\theta}{dt} + K\theta \;= T$$

The transfer function of this system is given by

$$\frac{\theta(s)}{T(s)} = \frac{1}{Js^2 + Bs + K}$$



Mechanical rotational system

Unit - 2
Introduction

Objectives:

To determine transfer function of DC servomotor and AC servomotor and synchro transmitter and receiver. To determine transfer function using block diagram reduction technique, signal flow graph using Mason's gain formula.

Syllabus:

Transfer Function of DC Servo Motor – AC Servo Motor

Synchro Transmitter and receiver

Block Diagram representation of systems considering electrical systems

Block Diagram Algebra

Representation by signal flow graph

Reduction using Mason's gain formula

Tutorials - Problems

Outcomes:

Students will be able to

- Determine transfer function of DC servomotor and AC servomotor and synchro transmitter and receiver
- Determine transfer function using block diagram reduction technique
- Determine transfer function using signal flow graph with Mason's gain formula.

Modelling of Elements of Control Systems

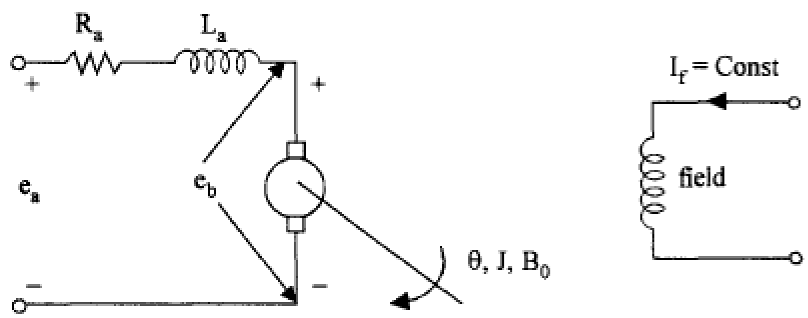
A feedback control system usually consists of several components in addition to the actual process. These are: error detectors, power amplifiers, actuators, sensors etc. Let us now discuss the physical characteristics of some of these and obtain their mathematical models.

DC Servo Motor

A DC servo motor is used as an actuator to drive a load. It is usually a DC motor oflow power rating. DC servo motors have a high ratio of starting torque to inertia and therefore they have a faster dynamic response. DC motors are constructed using rare earth permanent magnets which have high residual flux density and high coercivity. As no field winding is used, the field copper losses are zero and hence, the overall efficiency of the motor is high. The speed torque characteristic of this motor is flat over a wide range, as the armature reaction is negligible. Moreover speed is directly proportional to the armature voltage for a given torque. Armature of a DC servo motor is specially designed tohave low inertia. In some application DC servo motors are used with magnetic flux produced by field windings. The speed of PMDC motors can be controlled by applying variable armature voltage. These are called armature voltage controlled DC servo motors. Wound field DC motors can be controlled by either controlling the armature voltage or controlling the field current. Let us now consider modelling of these two types of DC servo motors.

(a) Armature controlled DC servo motor

The physical model of an armature controlled DC servo motor is given in Fig. 2.54.



The armature winding has a resistance Ra and inductance La' The field is produced either by a

permanent magnet or the field winding is separately excited and supplied with constant voltage so that the field current If is a constant. When the armature is supplied with a DC voltage of ea volts, the armature rotates and produces a back e.m.fe_b• The armature current $i_a$depends on the difference of $e_a$ande_b. The armature has a moment of inertia J, frictional coefficient Bo' The angular displacement of the motor is e.

The torque produced by the motor is given by,

$$T = K_T i_a$$

where K_T is the motor torque constant.

The back emf is proportional to the speed of the motor and hence

$$e_b = K_b \dot{\theta}$$

The differential equation representing the electrical system is given by,

$$R_a i_a + L_a \frac{di_a}{dt} + e_b = e_a$$

Taking Laplace transform of eqns.

$$T(s) = K_T I_a(s)$$
$$E_b(s) = K_b s \theta(s)$$
$$(R_a + s L_a) I_a(s) + E_b(s) = E_a(s)$$
$$I_a(s) = \frac{E_a(s) - K_b s \theta(s)}{R_a + sL_a}$$

The mathematical model of the mechanical system is given by,

$$J\frac{d^2\theta}{dt^2} + B_0\frac{d\theta}{dt} = T$$

Taking Laplace transform of eqn.

$$(Js^2 + B_o s)\ \theta(s) = T(s)$$

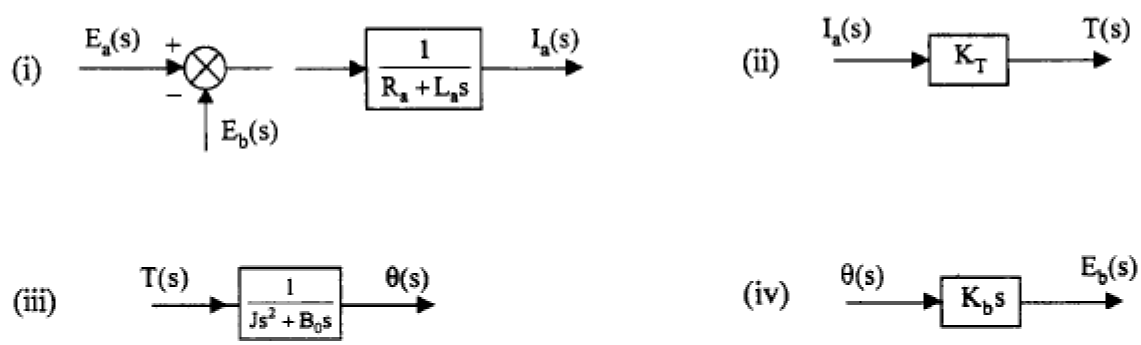
Using all the above equations we have

$$\theta(s) = K_T\ \frac{E_a(s) - K_b s\,\theta(s)}{(R_a + sL_a)(Js^2 + B_0 s)}$$

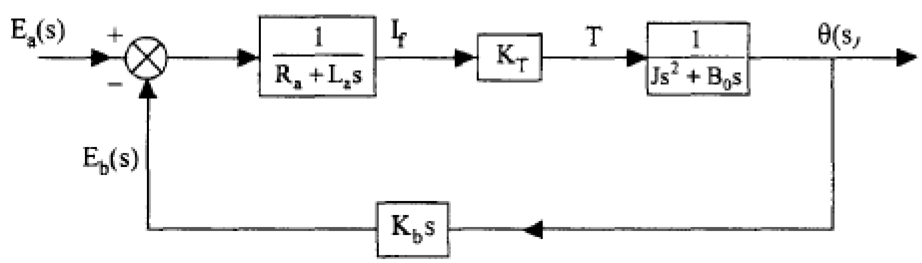
Solving for $\theta(s)$, we get

$$\theta(s) = \frac{K_T\ E_a(s)}{s[(R_a + sL_a)(Js + B_0) + K_T\ K_b]}$$

The block diagram representation of the armature controlled DC servo motor is developed in
steps, as shown in Fig.



Combining these blocks suitably we have the complete block diagram as shown in Fig



Usually the inductance of the armature winding is small and hence neglected. The overall transfer function, then becomes,

$$T(s) = \frac{\theta(s)}{E_a(s)} = \frac{K_T / R_a}{s\left[Js + B_0 + \frac{K_b K_T}{R_a}\right]}$$

$$= \frac{K_T / R_a}{s(Js + B)}$$

$B = B_0 + \frac{K_b K_T}{R_a}$ is the equivalent frictional coefficient.

It can be seen from eqn. the effect of back emf is to increase the effective frictionalcoeffcient thus providing increased dampingcan be written in another useful form known as time constant form, given by,

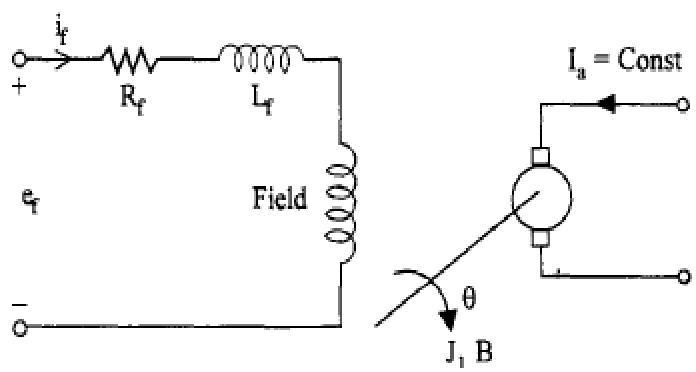
$$T(s) = \frac{K_M}{s(\tau_m s + 1)}$$

where $K_M = \frac{K_T}{R_a J}$ is the motor gain constant

and $\tau_m = \frac{J}{B}$ is the motor time constant

Armature controlled DC servo motors are used where power requirements are large and the
additional damping provided inherently by the back emf is an added advantage.

(b) Field controlled DC servo motor



Field controlled DC servo motors are economical where small size motors are required. For thefield circuit, low power servo amplifiers are sufficient and hence they are cheaper.

The electrical circuit is modelled as,

$$I_f(s) = \frac{E_f(s)}{R_f + L_f s} \qquad .....(2.98)$$

$$T(s) = K_T\, I_f(s) \qquad .....(2.99)$$

and

$$(Js^2 + B_0)\, \theta(s) = T(s) \qquad .....(2.100)$$

Combining eqns. (2.98), (2.99) and (2.100) we have

$$\frac{\theta(s)}{E_f(s)} = \frac{K_T}{s(Js + B_0)(R_f + L_f s)}$$

$$= \frac{K_T / R_f B_0}{s\left(\frac{J}{B_0}s + 1\right)\left(\frac{L_f}{R_f}s + 1\right)}$$

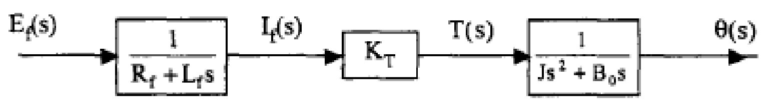
$$= \frac{K_m}{s(\tau_m s + 1)(\tau_f s + 1)} \qquad .....(2.101)$$

where $K_m = K_T/R_fB_0$ = motor gain constant

$\tau_m = J/B_0$ = motor time constant

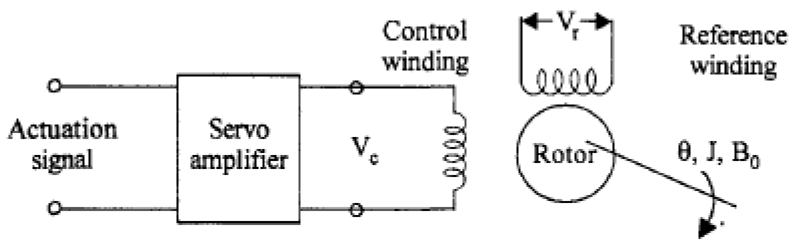
$\tau_f = L_f/R_f$ = field time constant

The block diagram is as shown in Fig. 2.58.



AC Servo Motors

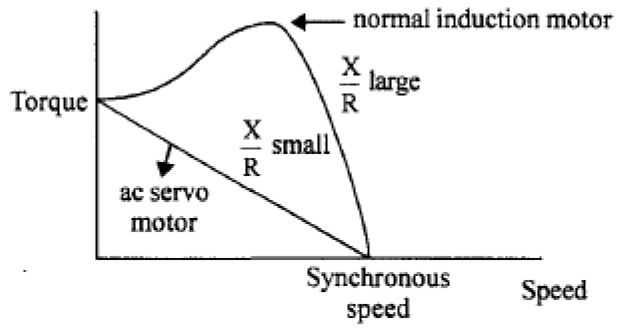
An AC servo motor is essentially a two phase induction motor with modified constructional features to suit servo applications. The schematic of a two phase ac servo motor is shown in Fig.



It has two windings displaced by 90° on the stator. One winding, called as reference winding, is

supplied with a constant sinusoidal voltage. The second winding, called control winding, is supplied with a variable control voltage which is dip laced by ± 90° out of phase from the reference voltage. The major differences between the normal induction motor and an AC servo motor are:

1. The rotor winding of an ac servo motor has high resistance (R) compared to its inductive reactance (X) so that its X/R ratio is very low. For a normal induction motor, X/R ratio is high so that the maximum torque is obtained in normal operating region which is around 5% of slip.The torque speed characteristics of a normal induction motor and an ac servo motor are shown in Fig.



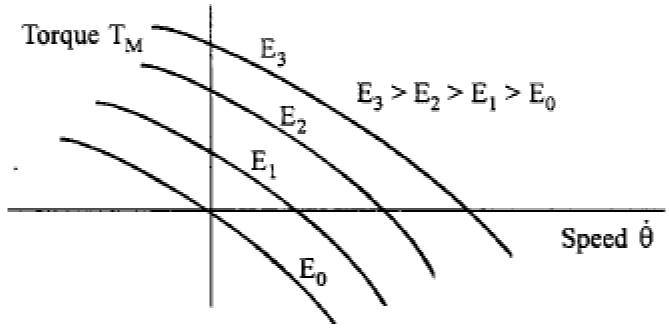
Torque speed characteristics of normal induction motor and ac servo motor

The Torque speed characteristic of a normal induction motor is highly nonlinear and has a positive slope for some portion of the curve. This is not desirable for control applications, as the positive slope makes the systems unstable. The torque speed characteristic of an ac servo motor is fairly linear and has negative slope throughout.

2. The rotor construction is usually squirrel cage or drag cup type for an ac servo motor. The diameter is small compared to the length of the rotor which reduces inertia of the moving parts. Thus it has good accelerating characteristic and good dynamic response.

3. The supply to the two windings of ac servo motor are not balanced as in the case of a normal induction motor. The control voltage varies both in magnitude and phase with respect to the constant reference voltage applied to the reference winding. The direction of rotation of the motor depends on the phase C± 90°) of the control voltage with respect to the reference voltage.

For different rms values of control voltage the torque speed characteristics are shown in Fig. The torque varies approximately linearly with respect to speed and also _control voltage. The torque speed characteristics can be linearised at the operating point and the transfer function of the motor can be obtained.



AC servo motor speed torque characteristics

The torque is a function of speed a and the control voltage E. Thus

$$T_M = f(\dot{\theta}, E) \qquad .....(2.102)$$

Expanding eqn. (2.102) in Taylor series around the operating point, $T_M = T_{MO}$, $E = E_0$ and $\dot{\theta} = \dot{\theta}_0$ and neglecting terms of order equal to and higher than two, we have

$$T_M = T_{MO} + \left.\frac{\partial T_M}{\partial E}\right|_{\substack{E=E_0\\ \dot{\theta}=\dot{\theta}_o}} (E - E_o) + \left.\frac{\partial T_M}{\partial \dot{\theta}}\right|_{\substack{E=E_0\\ \dot{\theta}=\dot{\theta}_o}} (\dot{\theta} - \dot{\theta}_o) \qquad .....(2.103)$$

or

$$T_M - T_{MO} = K(E - E_o) - B(\dot{\theta} - \dot{\theta}_o) \qquad .....(2.104)$$

where

$$K = \left.\frac{\partial T_M}{\partial E}\right|_{\substack{E=E_0\\ \dot{\theta}=\dot{\theta}_o}} \text{ and } B = -\left.\frac{\partial T_M}{\partial \dot{\theta}}\right|_{\substack{E=E_0\\ \dot{\theta}=\dot{\theta}_o}} \qquad .....(2.105)$$

(**Note :** Since $\frac{\partial T_M}{\partial \dot{\theta}}$ is negative, B will be positive)

Eqn. (2.104) can be written as

$$\Delta T_M = K\,\Delta E - B\,\Delta\dot{\theta} \qquad .....(2.106)$$

The mechanical equation is given by

$$J\,\Delta\ddot{\theta} + B_o\,\Delta\dot{\theta} = \Delta T_M = K\Delta E - B\Delta\dot{\theta} \qquad .....(2.107)$$

Taking Laplace transform of eqn. (2.107), we get the transfer function of an ac servo motor as,

$$\frac{\Delta\theta(s)}{\Delta E(s)} = \frac{K}{Js^2 + (B_o + B)s} \qquad .....(2.108)$$

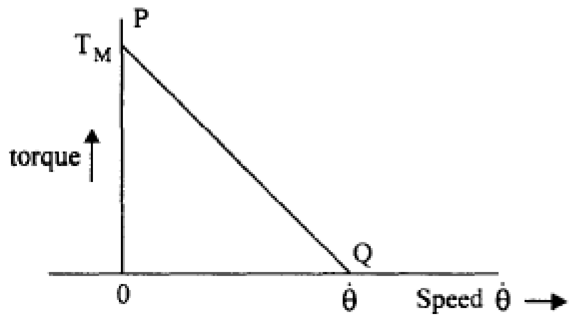
In time constant form we can write eqn. (2.108) as

$$\frac{\Delta\theta(s)}{\Delta E(s)} = \frac{K_m}{s(\tau_m s + 1)}$$

The constants K and B can be obtained by conducting a no load test and blocked rotor test on the ac servo motor at the rated control voltage Ec.

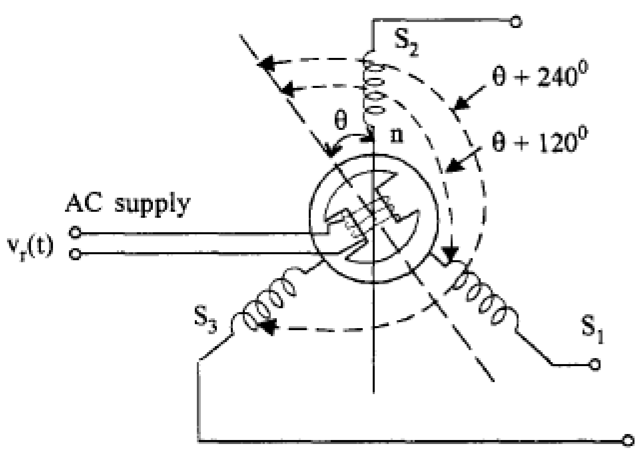
On no load $T_M = 0$ and on blocked rotor, $\theta = o$.

These two points are indicated as P and Q respectively on the diagram of Fig. 2.62. The line joining P and Q represents the approximate speed torque characteristic at rated control voltage.



Synchros

A commonly used error detector of mechanical positions of rotating shafts in AC control systems is the Synchro. It consists of two electro mechanical devices. 1. Synchro transmitter 2. Synchro receiver or control transformer. The principle of operation of these two devices is same but they differ slightly in their construction. The construction of a Synchro transmitter is similar to a 3 phase alternator. The stator consists of a balanced three phase winding and is star connected. The rotor is of dumbbell type -construction and is wound with a coil to produce a magnetic field. When an ac voltage is applied to the winding of the rotor, a magnetic field is produced. The coils in the stator link with this sinusoidal distributed magnetic flux and voltages are induced in the three coils due to transformer action. Thus the three voltages are in time phase with each other and the rotor voltage. The magnitudes of the voltages are proportional to the cosine of the angle between the rotor position and the respective coil axis. The position of the rotor and the coils are shown in Fig.



Synchro transmitter

If the voltages induced in the three coils are designated as V_{S1} ,$v_{S2}$and $V_{S3}$and if the rotor axis makes angle θ with the axis of S_I winding, we have,

$$v_R(t) = v_r \text{ Sin } \omega_r t$$

$$v_{s_{1n}} = KV_r \text{ Sin } \omega_r t \cos(\theta + 120) \quad .....(2.113)$$

$$v_{s_{2n}} = KV_r \text{ Sin } \omega_r t \cos\theta \quad .....(2.114)$$

$$v_{s_{3n}} = KV_r \text{ Sin } \omega_r t \cos(\theta + 240) \quad .....(2.115)$$

These are the phase voltages and hence the line voltages are given by,

$$v_{s_1 s_2} = v_{s_1 n} - v_{s_2 n} = \sqrt{3}\ KV_r \text{ Sin }(\theta + 240) \text{ Sin } \omega_r t \quad .....(2.116)$$

$$v_{s_2 s_3} = v_{s_2 n} - v_{s_3 n} = \sqrt{3}\ KV_r \text{ Sin }(\theta + 120) \text{ Sin } \omega_r t \quad .....(2.117)$$

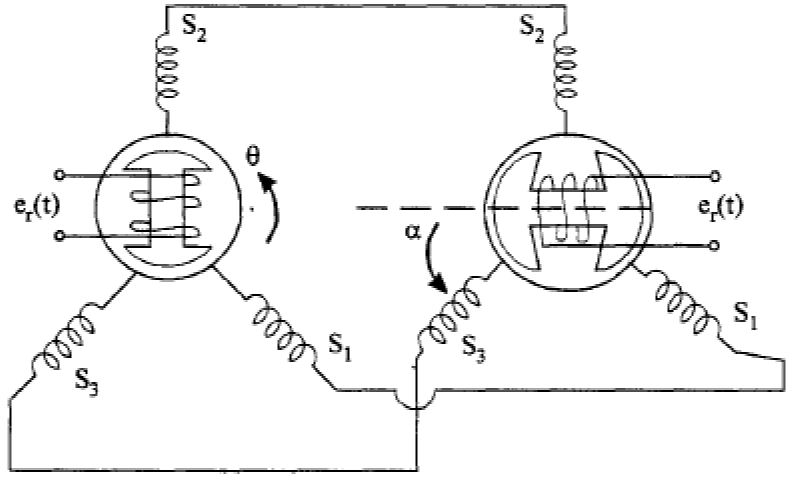
$$v_{s_3 s_1} = v_{s_3 n} - v_{s_1 n} = \sqrt{3}\ KV_r \text{ Sin } \theta \text{ Sin } \omega_r t \quad .....(2.118)$$

When θ= 0, the axis of the magnetic field coincides with the axis of coil S2 and maximum voltage is induced in it as seen from eqn. (2.114). For this position of the rotor, the voltage $V_{S_3 S_1}$ is zero, as given by eqn. (2.118). This position of the rotor is known as the 'Electrical Zero' of transmitter and is taken as reference for specifying the rotor position. In summary, it can be seen that the input to the transmitter is the angular position of the rotor and the set of three single phase voltages is the output. The magnitudes of these voltages depend on the angular position of the rotor as given in eqn. (2.116) to (2.118). Now consider these three voltages to be applied to the stator of a similar device called control transformer or synchro receiver. The construction of a control transformer is similar to that of the dumbbell in shape. Since the rotor is cylindrical, the air gap is uniform and the reluctance of the magnetic path is constant. This makes the output impedance of rotor to be a constant. Usually the rotor winding of control transformer is connected to an amplifier which requires signal with constant impedance for better performance. A synchro transmitter is usually required to supply several control transformers and hence the stator winding of control transformer is wound with higher impedance per phase.

Since the same currents flow through the stators of the synchro transmitter and receiver, the same pattern of flux distribution will be produced in the air gap of the control transformer. The control transformer flux axis is in the same position as that of the synchro transmitter. Thus the voltage induced in the rotor coil of control transformer is proportional to the cosine of the angle between the two rotors. Hence

$$e_r(t) = K_1 V_r \text{ Cos } \phi \sin \omega_r t \quad .....(2.119)$$

where ¢ is the angle between the two rotors. When ¢ = 90°, *e/t)* = 0 and the two rotors are at right angles. This position is known as the 'Electrical Zero' for the control transformer. In Fig. 2.64, a synchro transmitter receiver pair (usually called Synchro pair) connected as an error detector, is shown in the respective electrical zero positions.



If the rotor of the transmitter rotates through an angle θ in the anticlockwise direction, and the rotor of control transformer rotates by an angle α in the anticlockwise direction, the net angular displacement between the two is $(90 - \theta + \alpha)$. From eqn. (2.119),

$$e_r(t) = K_1 V_r \sin \omega_r t \cos (90 - \theta + \alpha)$$

$$= K_1 V_r \sin (\theta - \alpha) \sin \omega_r t \qquad .....(2.120)$$

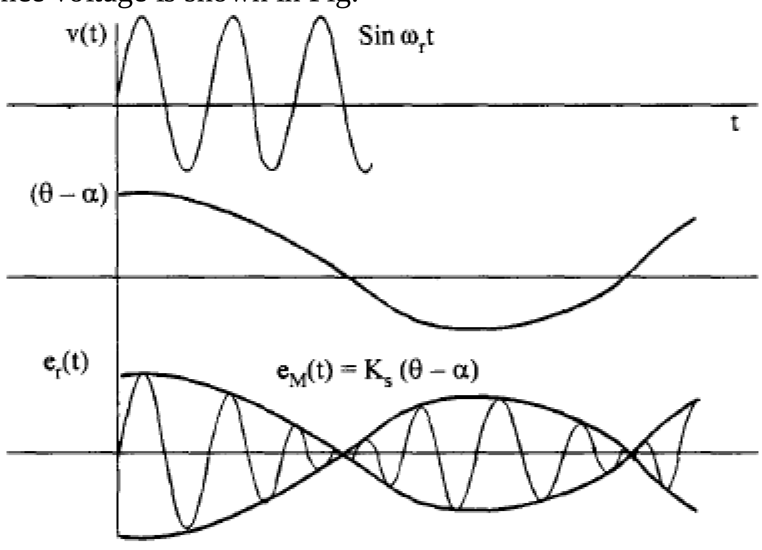
If $(\theta - \alpha)$ is small, which is usually the case,

$$e_r(t) = K_1 V_r (\theta - \alpha) \sin \omega_r t \qquad .....(2.121)$$

Thus the synchro pair acts as an error detector, by giving a voltage $e_r(t)$proportional to the difference in the angles of the two rotors. If the angular position of synchro transmitter is used as the reference position, the transformer rotor can be coupled to the load to indicate the error between the reference and the actual positions.

The waveform of error voltage for a given variation of difference in the angular positions together

with the reference voltage is shown in Fig.



Thus, we see that the output of the error detector is a modulated signal with the ac input to the

rotor of transmitter acting as carrier wave. The modulating signal *em(t)* is

$$e_m(t) = K_s (\theta - \alpha) \qquad .....(2.122)$$

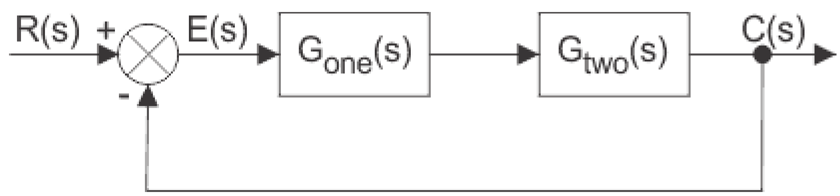
where Ks known as the sensitivity of the error detector. Thus the synchro pair is modelled by the

eqn. (2.122) when it is connected as an error detector.

Block Diagrams of Control System

The block diagram is to represent a control system in diagram form. In other words practical representation of a control system is its block diagram. It is not always convenient to derive the entire transfer function of a complex control system in a single function. It is easier and better to derive transfer function of control element connected to the system, separately. The transfer function of each element is then represented by a block and they are then connected together with the path of signal flow. For simplifying a complex control system, block diagrams are used. Each element of the control system is represented with a block and the block is the symbolic representation of transfer function of that element. A complete control system can be represented with a required number of interconnected such blocks. In the figure below, there are two elements with transfer function $G_{one}(s)$ and $G_{two}(s)$. Where $G_{one}(s)$ is the transfer function of first element and $G_{two}(s)$ is the transfer function of second element of the system.

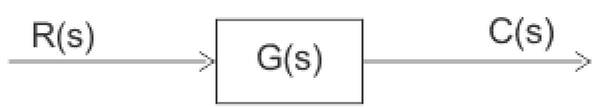
In addition to that, the diagram also shows there is a feedback path through which output signal C(s) is fed back and compared with the input R(s) and the difference between input and output E(s) = R(s) – C(s) is acting as actuating signal or error signal.



In each block of diagram, the output and input are related together by transfer function.

Where, transfer function $G(s) = \frac{C(s)}{R(s)}$

where, C(s) is the output and R(s) is the input of that particular block.



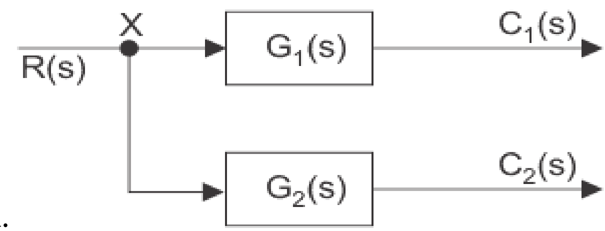
A complex control system consists of several blocks. Each of them has its own transfer function. But overall transfer function of the system is the ratio of transfer function of final output to transfer function of initial input of the system. This overall transfer function of the system can be obtained by simplifying the control system by combining this individual blocks, one by one. Technique of combining of these blocks is referred as **block diagram reduction technique**. For successful implementation of this technique, some rules for block diagram reduction to be followed. Let us discuss these rules, one by one for reduction of block diagram of control system.

If the transfer function of input of control system is R(s) and corresponding output is C(s), and the overall transfer function of the control system is G(s), then the control system can be represented as

$$Where,\ G(s) = \frac{C(s)}{R(s)}$$

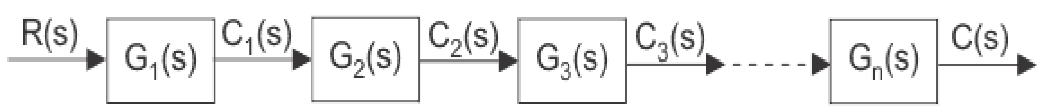
Take off Point of Block Diagram

when we need to apply one or same input to more than one blocks, we use **take off point**. A point is where the input gets more than one paths to propagate. This to be noted that the input does not get divided at a point, hence input propagates through all the paths connected to that point without affecting its value. Hence, by takeoff point same input signals can be applied to more than one systems or blocks. Representation of a common input signal to more than one blocks of control system is done by a common point as shown in the figure below with point X.



Cascade Blocks

When several systems or control blocks are connected in cascaded manner, the transfer function of the entire system will be the product of transfer function of all individual blocks. Here it also to be remembered that the output of any block will not be affected by the presence of other blocks in the cascaded system.



Now,

$$G_1(s) = \frac{C_1(s)}{R(s)},\ G_2(s) = \frac{C_2(s)}{C_1(s)},\ G_3(s) = \frac{C_3(s)}{C_2(s)},$$

$$\cdots\cdots\cdots\ G_n(s) = \frac{C(s)}{C_{n-1}(s)}$$

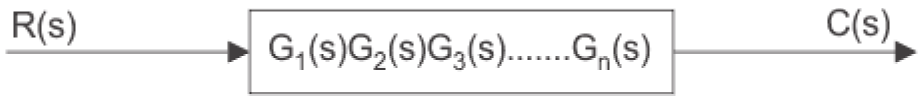
$$\therefore G(s)\ can\ be\ rewritten\ as,$$

$$G(s) = \frac{C(s)}{R(s)} = \frac{C_1(s)}{R(s)} \cdot \frac{C_2(s)}{C_1(s)} \cdot \frac{C_3(s)}{C_2(s)} \cdots \frac{C(s)}{C_{n-1}(s)}$$

$$= G_1(s)G_2(s)G_3(s) \cdot \cdots \cdot G_n(s)$$

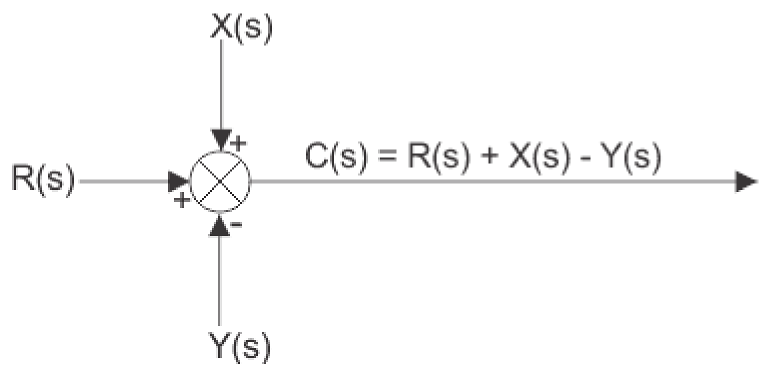
$$\therefore G(s) = G_1(s)G_2(s)G_3(s) \cdot \cdots \cdot G_n(s)$$

from the diagram it is seen that, Where, G(s) is the overall transfer function of cascaded control system.



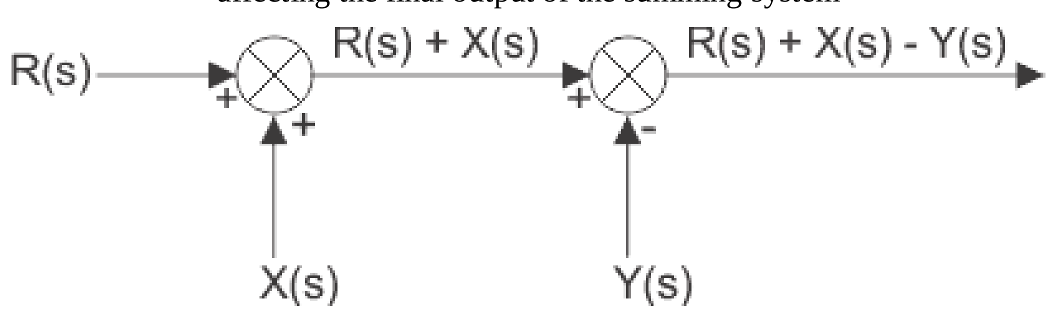
Summing Point of Block Diagram

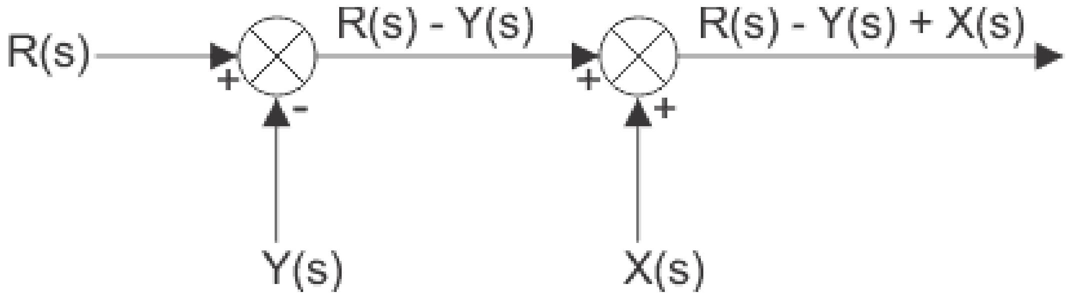
Instead of applying single input signal to different blocks as in the previous case, there may be such situation where different input signals are applied to same block. Here, resultant input signal is the summation of all input signals applied. Summation of input signals is represented by a point called summing point which is shown in the figure below by crossed circle. Here R(s), X(s) and Y(s) are the input signals. It is necessary to indicate the fine specifying the input signal entering a summing point in the block diagram of control system.



Consecutive Summing Point

A summing point with more than two inputs can be divided into two or more consecutive summing points, where alteration of the position of consecutive summing points does not effect the output of the signal. In other words - if there are more than one summing points directly inter associated, then they can be easily interchanged from their position without affecting the final output of the summing system

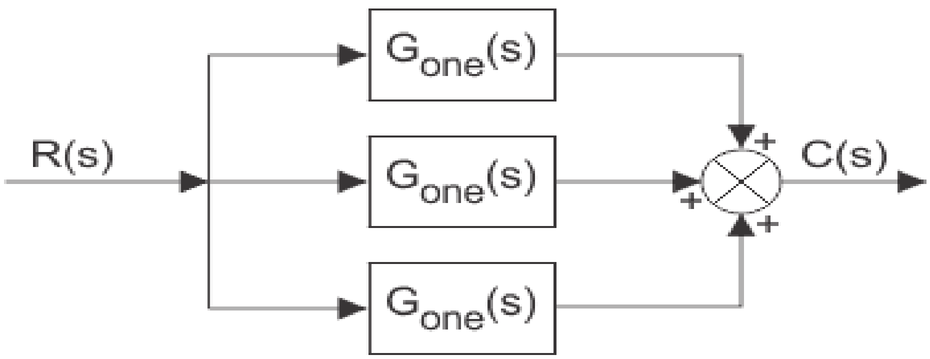




.

Parallel Blocks

When same input signal is applied different blocks and the output from each of them are added in a summing point for taking final output of the system then over all transfer function of the system will be the algebraic sum of transfer function of all individual blocks.



If C_{one}, C_{two} and C_{three} are the outputs of the blocks with transfer function G_{one}, G_{two} and G_{three}, then

$$G_{one}(s) = \frac{C_{one}(s)}{R(s)},\ G_{two}(s) = \frac{C_{two}(s)}{R(s)}\ and\ G_{three}(s) = \frac{C_{three}(s)}{R(s)}$$

$$From\ summing\ point$$

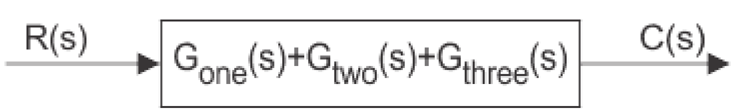
$$C(s) = C_{one}(s)+C_{two}(s)+C_{three}(s) = R(s)G_{one}(s)+R(s)G_{two}(s)+R(s)G_{three}(s)$$

$$\Rightarrow C(s) = R(s)\left[G_{one}(s) + G_{two}(s) + G_{three}(s)\right]$$

$$\Rightarrow \frac{C(s)}{R(s)} = G_{one}(s) + G_{two}(s) + G_{three}(s)$$

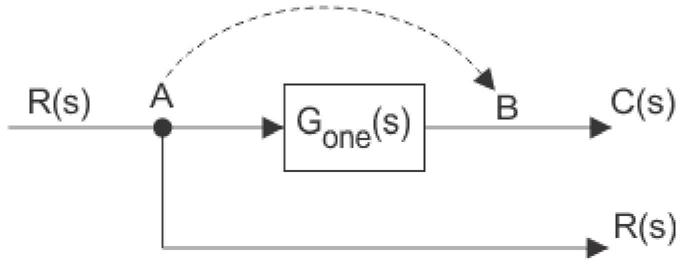
$$This\ is\ over\ all\ transfer\ function\ the\ system\ G(s),$$

$$\therefore G(s) = \frac{C(s)}{R(s)} = G_{one}(s) + G_{two}(s) + G_{three}(s)$$

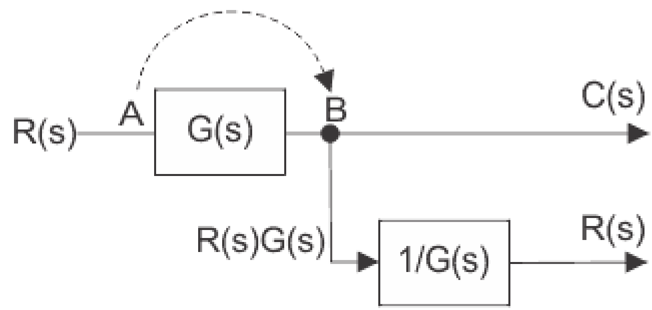


Shifting of Take off Point

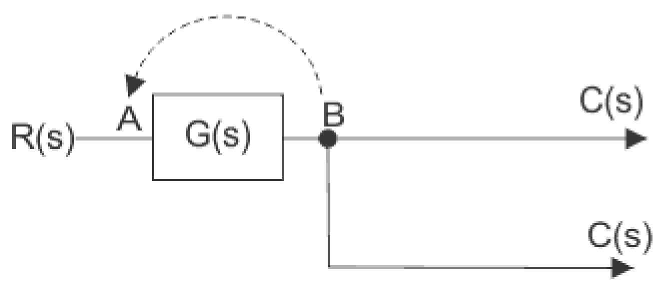
If same signal is applied to more than one system, then the signal is represented in the system by a point called take off point. Principle of **shifting of take off point** is that, it may be shifted either side of a block but final output of the branches connected to the take off point must be un-changed. The take off point can be shifted either sides of the block.



In the figure above the take off point is shifted from position A to B. The signal R(s) at take off point A will become G(s)R(s) at point B. Hence another block of inverse of transfer function G(s) is to be put on that path to get R(s) again.

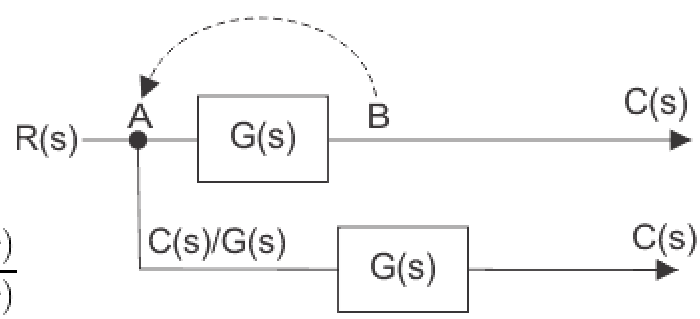


Now let us examine the situation when take off point is shifted before the block which was previously after the block.



Here the output is C(s) and input is R(s) and hence

$$G(s) = \frac{C(s)}{R(s)} \Rightarrow R(s) = \frac{C(s)}{G(s)}$$



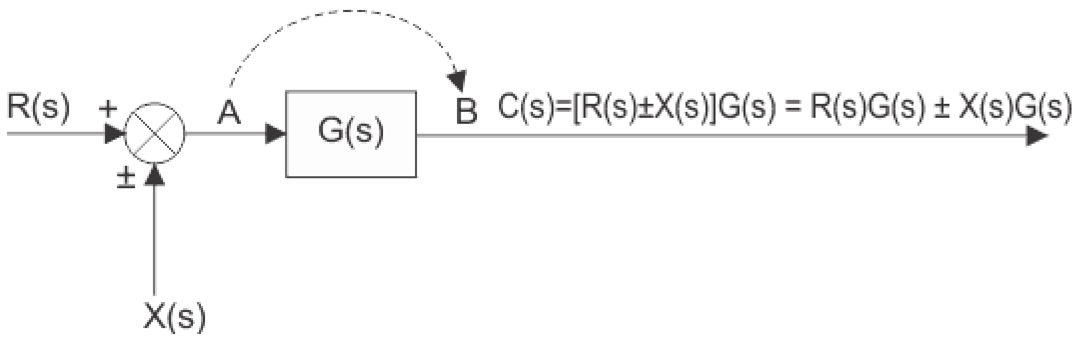
Here, we have to put one block of transfer function G(s) on the path so that output again comes as C(s).

$$G(s) \cdot \frac{C(s)}{G(s)} = C(s)$$

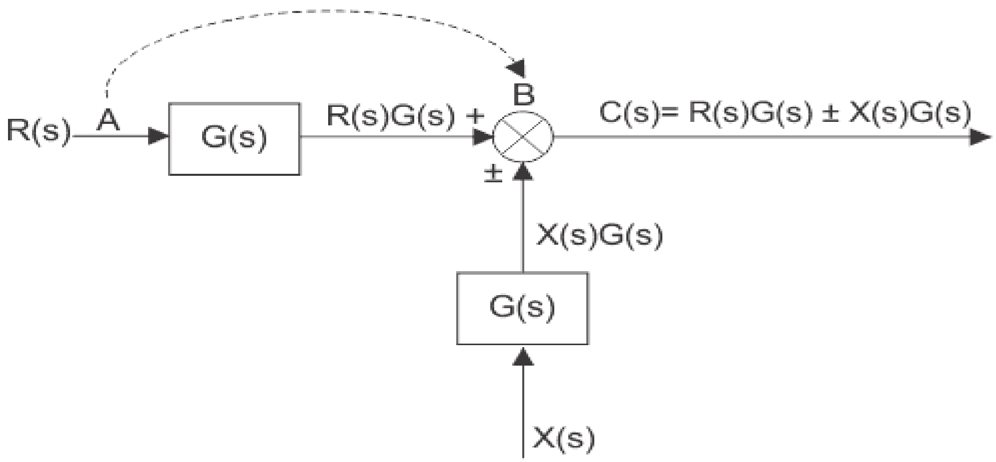
Shifting of Summing Point

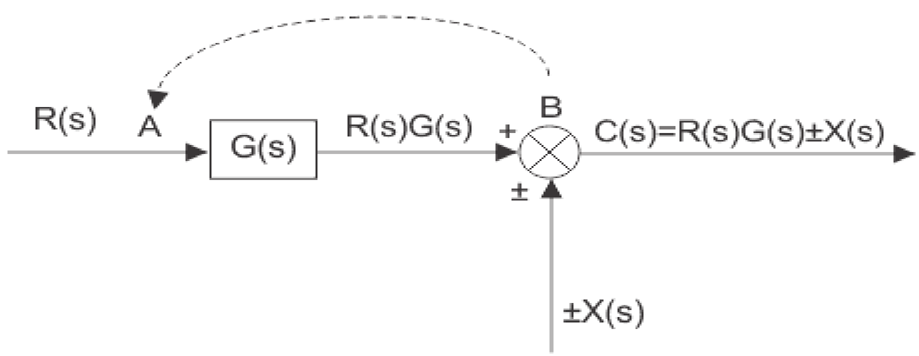
Let us examine the shifting of summing point from a position before a block to a position after a block. There are two input signals R(s) and ±X(s) entering in a summing point at position A. The output of the summing point is R(s) ± X(s). The resultant signal is the input of a control system block of transfer function G(s) and the final output of the system is

$$C(s) = [(R(s) \pm X(s)]\, G(s) = R(s)G(s) \pm X(s)G(s)$$



Hence, a summing point can be redrawn with input signals R(s)G(s) and ± X(s)G(s)

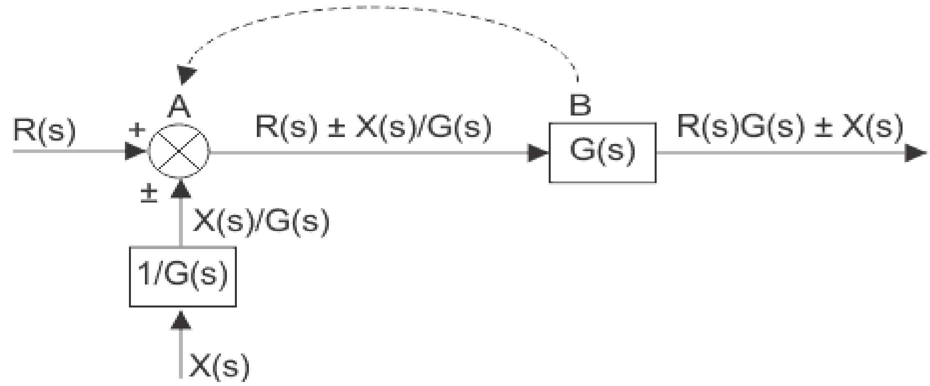




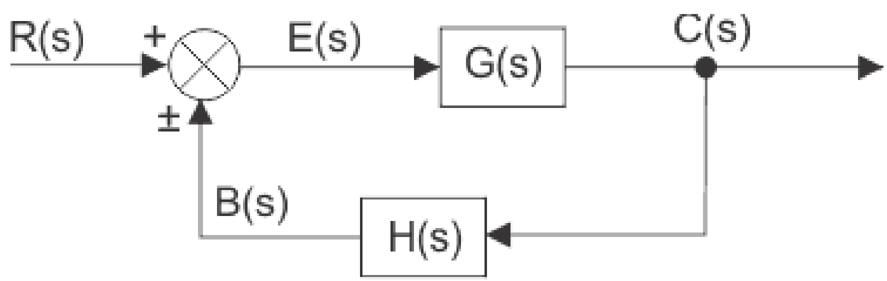
In the above block diagrams of control system output can be rewritten as

$$C(s) = R(s)G(s) \pm X(s) = G(s)\left[R(s) \pm \frac{X(s)}{G(s)}\right]$$

The above equation can be represented by a block of transfer function G(s) and input R(s)±X(s)/G(s) again R(s)±X(s)/G(s) can be represented with a summing point of input signal R(s) and ±X(s)/G(s) and finally it can be drawn as below.



Block Diagram of Closed Loop Control System



In a closed loop control system, a fraction of output is fed-back and added to input of the system. If H (s) is the transfer function of feedback path, then the transfer function of feedback signal will be B(s) = C(s)H(s). At summing point, the input signal R(s) will be added to B(s) and produces actual input signal or error signal of the system and it is denoted by E(s).

$$E(s) = R(s) \pm B(s) = R(s) \pm C(s)H(s)$$

$$Now,\ G(s) = \frac{C(s)}{E(s)} = \frac{C(s)}{R(s) \pm C(s)H(s)} [\because E(s) = R(s) \pm C(s)H(s)]$$

$Now,\ overall\ transfer\ function\ of\ the\ system\ is$

$$G'(s) = \frac{C(s)}{R(s)} = \frac{G(s)E(s)}{R(s)} [\because\ from\ above\ equation\ C(s) = G(s)E(s)]$$

$$\Rightarrow G'(s) = \frac{G(s)\ [R(s) \pm C(s)H(s)]}{R(s)} [\because\ from\ above\ equation\ E(s) = R(s) \pm C(s)H(s)]$$

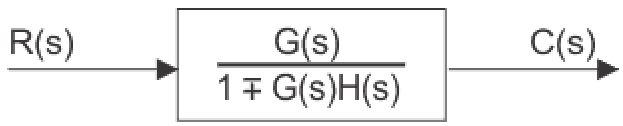
$$\Rightarrow G'(s) = G(s)\left[1 \pm \frac{C(s)H(s)}{R(s)}\right]$$

$$\Rightarrow G'(s) = G(s)\left[1 \pm G'(s)H(s)\right]\left[\because G'(s) = \frac{C(s)}{R(s)}\right]$$

$$\Rightarrow G'(s) = G(s) \pm G(s)G'(s)H(s)$$

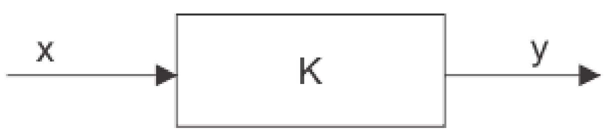
$$\Rightarrow G'(s)\left[1 \mp G(s)H(s)\right] = G(s)$$

$$\Rightarrow G'(s) = \frac{G(s)}{1 \mp G(s)H(s)}$$

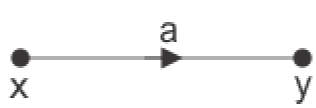


Signal Flow Graph of Control System

Signal flow graph of control system is further simplification of block diagram of control system. Here, the blocks of transfer function, summing symbols and take off points are eliminated by branches and nodes. The transfer function is referred as transmittance in signal flow graph. Let us take an example of equation y = Kx. This equation can be represented with block diagram as below

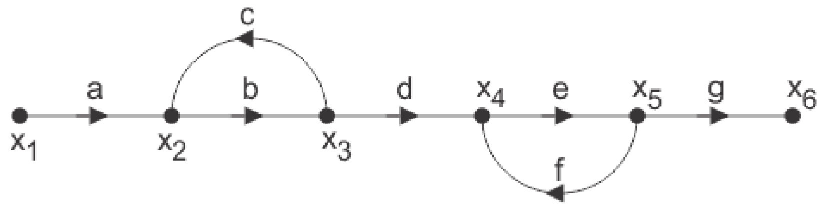


The same equation can be represented by signal flow graph, where x is input variable node, y is output variable node and a is the transmittance of the branch connecting directly these two nodes.



Rules for Drawing Signal Flow Graph

1. The signal always travels along the branch towards the direction of indicated arrow in the branch.
2. The output signal of the branch is the product of transmittance and input signal of that branch.
3. Input signal at a node is summation of all the signals entering at that node.
4. Signals propagate through all the branches, leaving a node.



$$x_2 = ax_1 + cx_3$$
$$x_3 = bx_2$$
$$x_4 = dx_3 + fx_5$$
$$x_5 = ex_4$$
$$x_6 = gx_5$$

$$\Rightarrow x_6 = gx_5 = gex_4 = ge\left(dx_3 + fx_5\right) = gedx_3 + gefx_5 = gedbx_2 + efx_6$$

$$= gedb\left(ax_1 + cx_3\right) + efx_6 = abdegx_1 + bcdegx_3 + efx_6$$

$$\Rightarrow \left(1 - ef\right)x_6 = abdegx_1 + bceg\left(x_4 - fx_5\right) = abdegx_1 + bcegx_4 - bcegfx_5$$

$$= abdegx_1 + bcgx_5 - bcefx_6 = abdegx_1 + bcx_6 - bcefx_6$$

$$\Rightarrow \left(1 - ef - bc + bcef\right)x_6 = abdegx_1$$

$$\Rightarrow \frac{x_6}{x_1} = \frac{abdeg}{1 - (bc + ef) + bcef}$$

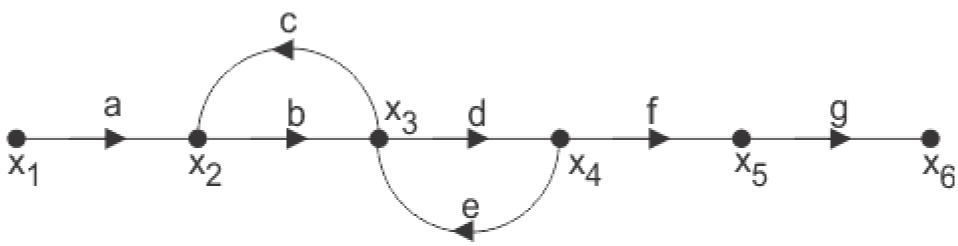
Simple Process of Calculating Expression of Transfer Function for Signal Flow Graph

- First, the input signal to be calculated at each node of the graph. The input signal to a node is summation of product of transmittance and the other end node variable of each of the branches arrowed towards the former node.

- Now by calculating input signal at all nodes will get numbers of equations which relating node variables and transmittance. More precisely, there will be one unique equation for each of the input variable node.

- By solving these equations we get, ultimate input and output of the entire signal flow graph of control system.

- Lastly by dividing inspiration of ultimate output to the expression of initial input we calculate the expiration of transfer function of that signal flow graph.



$x_2 = ax_1 + cx_3$
$x_3 = bx_2 + ex_4$
$x_4 = dx_3$
$x_5 = fx_4$
$x_6 = gx_5$
$\Rightarrow x_6 = gx_5 = fgx_4 = dfgx_3 = dfg\,(bx_2 + ex_4) = bdfgx_2 + dfgex_4$

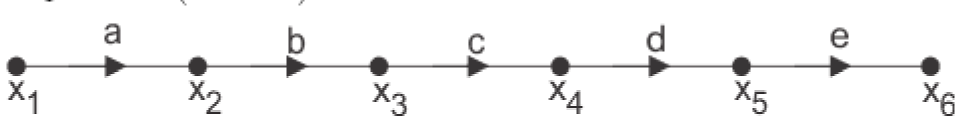
$= bdfg\,(ax_1 + cx_3) + dfgex_4 = abdfgx_1 + bcdfgx_3 + dfgex_4$

$\Rightarrow x_6 = abdfgx_1 + bcfgx_4 + dgex_5 = abdfgx_1 + bcfgx_4 + dgex_5$

$= abdfgx_1 + bcgx_5 + dex_6 = abdfgx_1 + bcx_6 + dex_6$

$\Rightarrow x_6 - (bcx_6 + dex_6) = abdfgx_1$

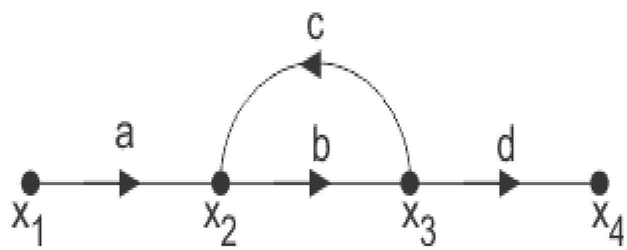
$$\Rightarrow \frac{x_6}{x_1} = \frac{abdfg}{1 - (bc + de)}$$



$x_2 = ax_1$
$x_3 = bx_2$
$x_4 = cx_3$
$x_5 = dx_4$
$x_6 = ex_5$

$\Rightarrow x_6 = ex_5 = edx_4 = edcx_3 = edcbx_2 = edcbax_1$

$$\frac{x_6}{x_1} = abcde$$

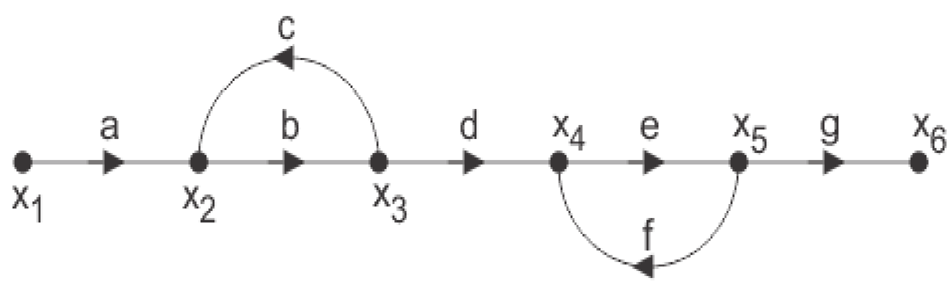


$x_2 = ax_1 + cx_3$
$x_3 = bx_2$
$x_4 = dx_3$

$$\Rightarrow x_4 = dx_3 = dbx_2 = db\left(ax_1 + cx_3\right) = abdx_1 + cbdx_3 = abdx_1 + cbx_4$$

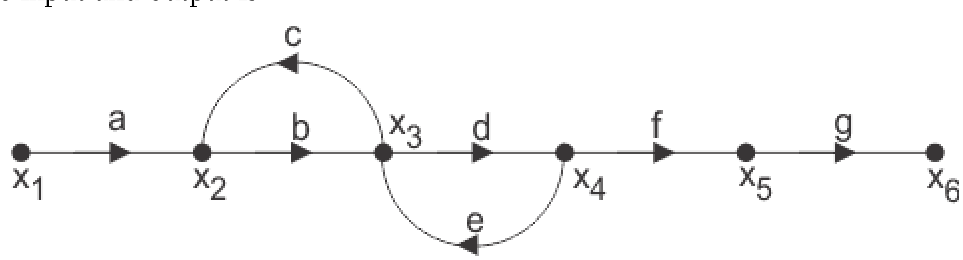
$$\Rightarrow \frac{x_4}{x_1} = \frac{abd}{1 - cb}$$

If P is the forward path transmittance between extreme input and output of a signal flow graph. L_1, L_2………………… loop transmittance of first, second,……..loop of the graph. Then for first signal flow graph of control system, the overall transmittance between extreme input and output is

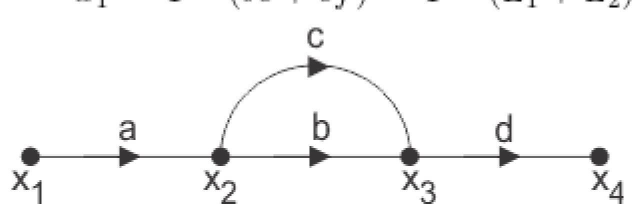


$$T = \frac{x_6}{x_1} = \frac{abdeg}{1 - (bc + ef) + bcef} = \frac{P}{1 - (L_1 + L_2) + L_1 L_2}$$

Then for second signal flow graph of control system, the overall transmittance between extreme input and output is



$$T = \frac{x_6}{x_1} = \frac{abdeg}{1 - (bc + ef)} = \frac{P}{1 - (L_1 + L_2)}$$



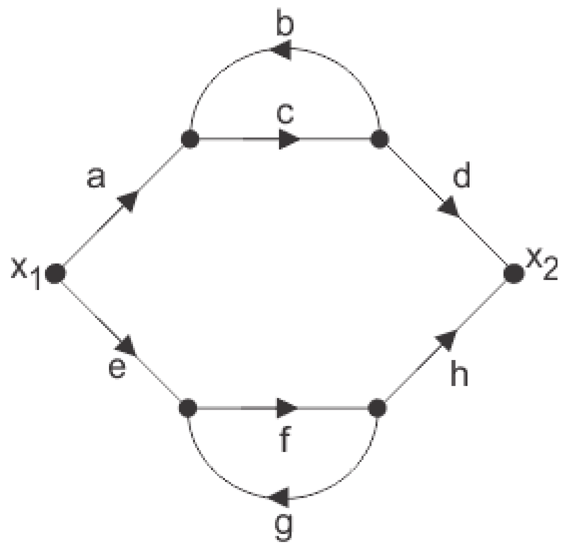
$$x_2 = ax_1$$
$$x_3 = bx_2 + cx_2$$
$$x_4 = dx_3$$

$$\Rightarrow x_4 = d\left(bx_2 + cx_2\right) = bdx_2 + cdx_2 = bdax_1 + cdax_1$$

$$\Rightarrow \frac{x_4}{x_1} = bda + cda = abd + acd$$

$$\therefore T = \frac{x_4}{x_1} = abd + acd = P_1 + P_2$$

Where P_1 & P_2 *forward path transmittance of two parallel path respectively.*



Here in the figure above, there are two parallel forward paths. Hence, overall transmittance of that signal flow graph of control system will be simple arithmetic sum of forward transmittance of these two parallel paths.

As the each of the parallel paths having one loop associated with it, the forward transmittances of these parallel paths are

$$T_1 = \frac{P_1}{1 - L_1} \; and \; T_2 = \frac{P_2}{1 - L_2} \; respectively.$$

Therefore overall transmittance of the signal flow graph is

$$T = T_1 + T_2 = \frac{P_1}{1 - L_1} + \frac{P_2}{1 - L_2}$$

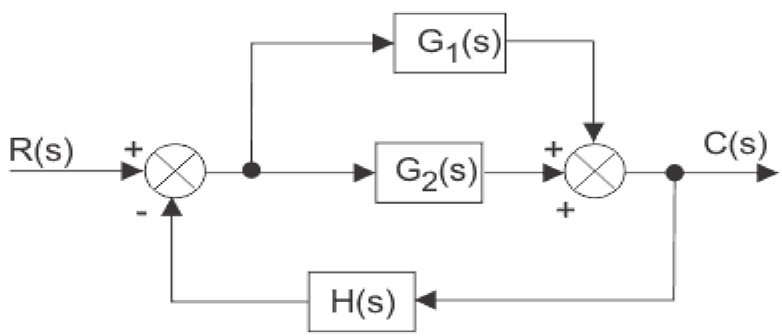
Mason's Gain Formula

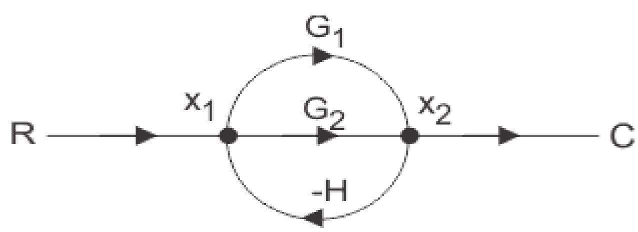
The overall transmittance or gain of signal flow graph of control system is given by Mason’s Gain Formula and as per the formula the overall transmittance is

$$T = \sum_{k=1}^{k} \frac{P_k \Delta_k}{\Delta}$$

Where, P_k is the forward path transmittance of k_{th} in path from a specified input is known to an output node. In arresting P_k no node should be encountered more than once. Δ is the graph determinant which involves closed loop transmittance and mutual interactions between non-touching loops. Δ = 1 - (sum of all individual loop transmittances) + (sum of loop transmittance products of all possible pair of non-touching loops) - (sum of loop transmittance products of all possible triplets of non-touching loops) + (……) - (……) Δ_k is the factor associated with the concerned path and involves all closed loop in the graph which are isolated from the forward path under consideration. The path factor Δ_k for the k_{th} path is equal to the value of grab determinant of its signal flow graph which exist after erasing the K_{th}path from the graph.

By using this formula one can easily determine the overall transfer function of control system by converting a block diagram of control system (if given in that form) to its equivalent signal flow graph. Let us illustrate the below given block diagram





$$P_1 = G_1,\ P_2 = G_2,\ L_1 = -G_1H,\ L_2 = -G_2H$$

$$\Delta_1 = 1,\ \Delta_2 = 1\ \because\ both\ loops\ L_1\ \&\ L_2\ touch\ both\ forward\ paths\ G_2\ \&\ G_1\ respectively$$

$$\Delta = 1 - \{L_1 + L_2\} = 1 - \{(-G_1H) + (-G_2H)\} = 1 + G_1H + G_2H$$

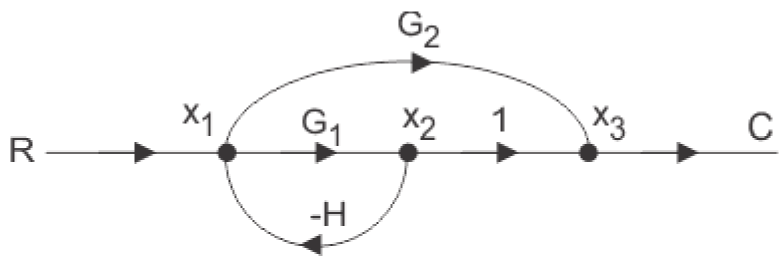
$$Now,\ T = \frac{C}{R} = \sum_{k=1}^{2} \frac{P_k\Delta_k}{\Delta} = \frac{P_1 \cdot \Delta_1 + P_2 \cdot \Delta_2}{\Delta} = \frac{G_1 + G_2}{1 + G_1H + G_2H}$$

$Here,\ x_1 = R - Hx_2$
$x_2 = x_1G_1 + x_1G_2$
$C = x_2$

$$\Rightarrow C = x_2 = x_1G_1 + x_1G_2 = x_1(G_1 + G_2) = (R - Hx_2)(G_1 + G_2)$$
$$= (R - HC)(G_1 + G_2)$$

$$\Rightarrow C(1 + G_1H + G_2H) = R(G_1 + G_2)$$

$$\Rightarrow T = \frac{C}{R} = \frac{G_1 + G_2}{1 + G_1H + G_2H}$$



$P_1 = G_1,\ P_2 = G_2,\ L_1 = -G_1H$

$\Delta_1 = 1,\ \Delta_2 = 1\ \because loop\ L_1\ touches\ both\ forward\ paths\ G_1\ \&\ G_2$

$$\Delta = 1 - \{L_1\} = 1 - \{(-G_1H)\} = 1 + G_1H$$

$$Now,\ T = \frac{C}{R} = \sum_{k=1}^{2} \frac{P_k\Delta_k}{\Delta} = \frac{P_1 \cdot \Delta_1 + P_2 \cdot \Delta_2}{\Delta} = \frac{G_1 + G_2}{1 + G_1H}$$

UNIT – III

TIME RESPONSE ANALYSIS

Objective:

To Familiarize with the Time Response analysis of various systems

To Familiarize the Concepts of Steady state errors.

Syllabus:

Standard test signals - Time response of first order systems – Characteristic Equation of Feedback control systems, Transient response of second order systems - Time domain specifications – Steady state response - Steady state errors and error constants – Effects of proportional derivative, proportional integral, proportional integral derivative systems.

Learning Outcomes:

Students will be able to

a) Analyze the First and Second Order Systems for different inputs.

b) Analyse the effect of P, PI, PD, PID Controllers.

c) Obtaining the Time- domain Specifications of Second order systems.

Time Response Analysis

In a control system, there may be some energy storing elements attached to it. Energy storing elements are generally inductors and capacitors in case of electrical system. Due to presence of these energy storing elements, if the energy state of the system is disturbed, it will take certain time to change from one energy state to another. The exact time taken by the system for changing one energy state to another, is known as transient time and the value and pattern voltages and currents during this period is known as transient response. A transient response is normally associated with an oscillation, which may be sustained or decaying in nature. The exact nature of the system depends upon the parameters of the system. Any system can be represented with a linear differential equation. The solution of this linear differential equation gives the response of the system. The representation of a control system by linear differential equation of functions of time and its solution is collectively called **time domain analysis of control system**.

Standard Test Signals:

1. **Step Function:**

Let us take an independent voltage source or a battery which is connected across a voltmeter via a switch, s. whenever the switch s is open, the voltage appears between the voltmeter terminals is zero. If the voltage between the voltmeter terminals is represented as v (t), the situation can be mathematically represented as

$$v(t) = 0 \;\; when \quad -\infty < t < 0$$

Now let us consider at t = 0, the switch is closed and instantly the battery voltage V volt appears across the voltmeter and that situation can be represented as,

$$v(t) = k \;\; when \;\; 0 < t < \infty$$

Combining the above two equations we get

$$\begin{aligned} v(t) &= 0 \;\; when \quad -\infty < t < 0 \\ \& \qquad &= k \;\; when \qquad 0 < t < \infty \end{aligned}$$

In the above equations if we put 1 in place of V, we will get a unit step function which can be defined as

$$\begin{aligned} u(t) &= 0 \;\; when \;\; t \le 0 \\ \& \qquad &= 1 \;\; when \;\; t \ge 0 \end{aligned}$$

Now let us examine the Laplace transform of unit step function. Laplace transform of any function can be obtained by multiplying this function by e^{-st} and integrating multiplied from 0 to infinity.

$$\pounds u(t) = \int_0^\infty u(t)e^{-st}dt = \int_0^\infty 1.e^{-st}dt = \left[\frac{e^{-st}}{-s}\right]_0^\infty = \frac{1}{s}$$

If input is R(s), then

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

2. **Ramp Function**

The function which is represented by an inclined straight line intersecting the origin is known as ramp function. That means this function starts from zero and increases or decreases linearly with time. A ramp function can be represented as,

$$\begin{aligned} r(t) &= 0 \quad when \; t < 0 \\ \& \qquad &= kt \quad when \; t > 0 \end{aligned}$$

Here in this above equation, k is the slope of the line.

Now let us examine the Laplace transform of ramp function. As we told earlier Laplace transform of any function can be obtained by multiplying this function by e^{-st} and integrating multiplied from 0 to infinity.

$$\pounds r(t) = \int_0^\infty r(t)e^{-st}dt = \int_0^\infty kt.e^{-st}dt = \frac{k}{s^2}$$
$$R(S) = \frac{k}{s^2}$$

3. Parabolic Function

Here, the value of function is zero when time t<0 and is quadratic when time t>0. A parabolic function can be defined as,

$$\begin{aligned} & p(t) = 0 \quad when\ t < 0 \\ \& \quad & = \frac{kt^2}{2} \quad when\ t > 0 \end{aligned}$$

Now let us examine the Laplace transform of parabolic function. As we told earlier Laplace transform of any function can be obtained by multiplying this function by e^{-st} and integrating multiplied from 0 to infinity.

$$\pounds p(t) = \int_0^\infty p(t)e^{-st}dt = \int_0^\infty \frac{kt^2}{2}.e^{-st}dt = \frac{k}{s^3}$$
$$R(S) = \frac{k}{s^3}$$

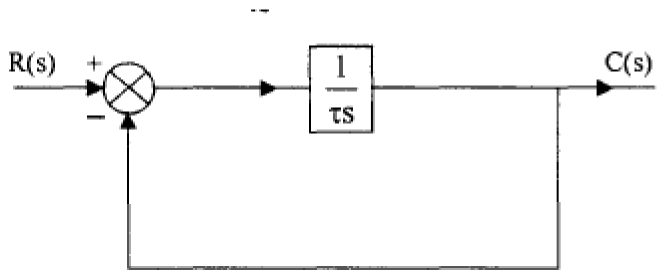
4. Impulse Function

Impulse signal is produced when input is suddenly applied to the system for infinitesimal duration of time. The waveform of such signal is represented as impulse function. If the magnitude of such function is unity, then the function is called unit impulse function. The first time derivative of step function is impulse function. Hence Laplace transform of unit impulse function is nothing but Laplace transform of first-time derivative of unit step function.

$$\pounds\left(Unit\ impulse\ function\right) = \pounds\frac{d}{dt}\left(Unit\ step\ function\right)$$
$$= s\pounds\left(Unit\ step\ function\right) = s.\frac{1}{s} = 1$$

Time Response of First Order Control Systems:

Consider a feedback system with G(s) = 1 / τs as shown in Fig



When the maximum power of s in the denominator of a transfer function is one, the transfer function represents a first order control system. Commonly, the first order control system can be represented as

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

Time Response for Step Function

Now a unit step input is given to the system, then let us analyze the expression of

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

$$\Rightarrow C(s) = \frac{1}{s}\cdot\frac{1}{sT+1}\left[\because R(s) = \frac{1}{s}\right] = \frac{T}{sT(sT+1)}$$

$$= T\left[\frac{sT+1-sT}{sT(sT+1)}\right] = T\left[\frac{1}{sT} - \frac{1}{(sT+1)}\right]$$

$$= \frac{1}{s} - \frac{T}{sT+1} = \frac{1}{s} - \frac{1}{s+\frac{1}{T}}$$

$$\therefore \pounds^{-1}\left[C(s)\right] = \pounds^{-1}\left[\frac{1}{s} - \frac{1}{s+\frac{1}{T}}\right]$$

$$\Rightarrow c(t) = \pounds^{-1}\left[\frac{1}{s}\right] - \pounds^{-1}\left[\frac{1}{s+\frac{1}{T}}\right] = 1 - e^{-t/T}$$

$$\left(\because \pounds^{-1}\left[\frac{1}{s}\right] = 1 \;\&\; \pounds^{-1}\left[\frac{1}{s+a}\right] = e^{-at}\right)$$

It is seen from the error equation that if the time approaching to infinity, the output signal reaches exponentially to the steady-state value of one unit. As the output is approaching towards input exponentially, the steady-state error is zero, when time approaches to infinity.

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

$$\therefore\ Steady\ State\ Error$$

$$= \lim_{t\to\infty}\left[1 - \left(1 - e^{-t/T}\right)\right] = \lim_{t\to\infty} e^{-t/T} = 0$$

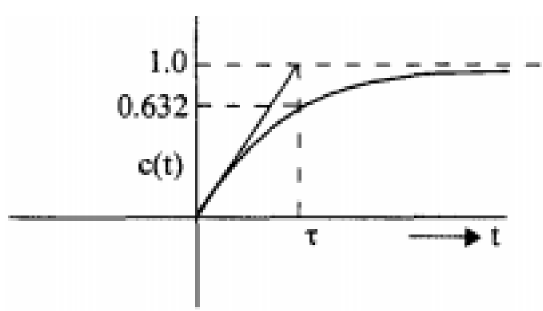
Let us put t = T in the output equation and then we get, $c(T) = 1 - e^{-T/T} = 1 - e^{-1} = 1 - 0.368 = 0.632$

This T is defined as the time constant of the response and the time constant of a response signal is that time for which the signal reaches to its 63.2 % of its final value.

Now if we put t = 4T in the above output response equation, then we get,

$$c(T) = 1 - e^{-4T/T} = 1 - e^{-4} = 1 - 0.018 = 0.982$$

When actual value of response, reaches to the 98% of the desired value, then the signal is said to be reached to its steady-state condition. This required time for reaching the signal to 98 % of its desired value is known as setting time and naturally setting time is four times of the time constant of the response. The condition of response before setting time is known as transient condition and condition of the response after setting time is known as steady-state condition. From this explanation it is clear that if the time constant of the system is smaller, the response of the system reaches to its steady-state condition faster.



Unit step response of a first order system

Time Response for Ramp Function

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

$$\Rightarrow C(s) = \frac{1}{s^2} \cdot \frac{1}{sT+1}\left[\because R(s) = \frac{1}{s^2}\right]$$

$$= \frac{1 - s^2T^2 + s^2T^2}{s^2(sT+1)} = \frac{(1+sT)(1-sT)}{s^2(sT+1)} + \frac{s^2T^2}{s^2(sT+1)}$$

$$= \frac{(1-sT)}{s^2} + \frac{T^2}{(sT+1)} = \frac{1}{s^2} - \frac{T}{s} + \frac{T^2}{(sT+1)}$$

$$= \frac{1}{s^2} - \frac{T}{s} + \frac{T}{(s+\frac{1}{T})}$$

$$\therefore c(t) = \pounds^{-1}\left[C(s)\right] = \pounds^{-1}\left[\frac{1}{s^2} - \frac{T}{s} + \frac{T}{(s+\frac{1}{T})}\right]$$

$$= \pounds^{-1}\left[\frac{1}{s^2}\right] - \pounds^{-1}\left[\frac{T}{s}\right] + \pounds^{-1}\left[\frac{T}{(s+\frac{1}{T})}\right]$$

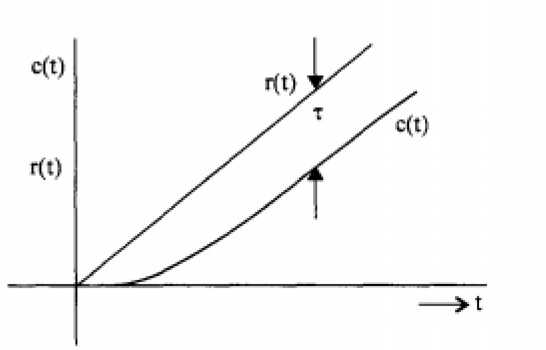
$$= t - T + Te^{-t/T}$$

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

$$\therefore Steady\ State\ Error$$

$$= \lim_{t\to\infty}\left[t - \left(t - T + Te^{-t/T}\right)\right] = \lim_{t\to\infty}\left(T - Te^{-t/T}\right)$$

$$= \lim_{t\to\infty} T\left(1 - e^{-t/T}\right) = T$$



Unit ramp response of a first order system

In this case during steady-state condition, the output signal lags behind input signal by a time equal to the time constant of the system. If the time constant of the system is smaller, the positional error of the response becomes lesser.

Time Response for Impulse Function

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

$$\Rightarrow C(s) = 1 \cdot \frac{1}{sT+1} [\because R(s) = 1]$$

$$= \frac{1}{T} \cdot \frac{1}{s + \frac{1}{T}}$$

$$\therefore c(t) = \pounds^{-1}[C(s)] = \pounds^{-1}\left[\frac{1}{T} \cdot \frac{1}{s+\frac{1}{T}}\right] = \frac{1}{T}e^{-t/T}$$

In the above explanation of time response of control system, we have seen that the step function is the first derivative of ramp function and the impulse function is the first derivative of step function. It is also found that the time response of step function is first derivative of time response of ramp function and time response of impulse function is first derivative of time response of step function.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(t - T + Te^{-t/T}\right) = 1 - 0 + \frac{-T}{T}e^{-t/T} = 1 - e^{-t/T}$$

$$and\ \frac{\mathrm{d}}{\mathrm{d}t}\left(1 - e^{-t/T}\right) = -\frac{1}{T}\left(-e^{-t/T}\right) = \frac{1}{T}e^{-t/T}$$

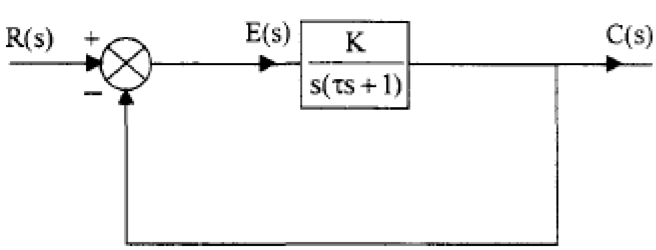
Definition of Final Value Theorem of Laplace Transform:

If f(t) and f'(t)both are Laplace Transformable and sF(s)has no pole in jw axis and in the R.H.P. (Right half Plane) then,

$$\lim_{s \to 0} sF(s) = \lim_{t \to \infty} f(t)$$

Time Response of Second-Order Control System

The order of a control system is determined by the power of s in the denominator of its transfer function. If the power of s in the denominator of transfer function of a control system is 2, then the system is said to be second-order control system.



$$T(s) = \frac{K/\tau}{s^2 + \frac{1}{\tau}s + K/\tau}$$

$$= \frac{\omega_n{}^2}{s^2 + 2\delta\omega_n s + \omega_n{}^2}$$

Where,

$$\omega_n = \sqrt{\frac{K}{\tau}} = \text{natural frequency}$$

$$\delta = \frac{1}{2\sqrt{K\tau}} = \text{damping factor}$$

The general expression of transfer function of a second order control system is given as

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

Here, ζ and ω_n are damping ratio and natural frequency of the system respectively and we will learn about these two terms in detail later on. Therefore, the output of the system is given as

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

If we consider a unit step function as the input of the system, then the output equation of the system is

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

$$Now,\ r(t) = 1\ or\ R(s) = \frac{1}{s}$$

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

$$= \frac{1}{s} \cdot \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \zeta^2\omega_n^2 + \omega_n^2 - \zeta^2\omega_n^2}$$

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

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

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

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

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

$$Putting,\ \omega_d = \omega_n\sqrt{1-\zeta^2}$$

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

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

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

Taking inverse Laplace transform of above equation, we get,

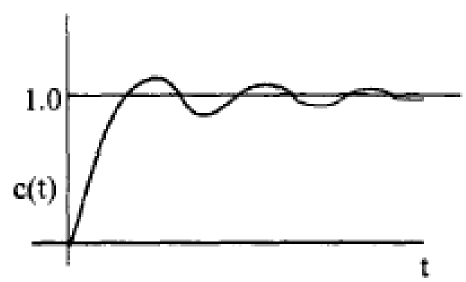
$$c(t) = 1 - e^{-\delta\omega_n t}\left[\cos\omega_n\sqrt{1-\delta^2}\ t + \frac{\delta}{\sqrt{1-\delta^2}}\sin\omega_n\sqrt{1-\delta^2}\ t\right]$$

$$c(t) = 1 - \frac{e^{-\delta\omega_n t}}{\sqrt{1-\delta^2}}\ \mathrm{Sin}\,(\omega_d t + \phi)$$

Where $\omega_d = \omega_n\sqrt{1-\delta^2}$

and $\tan\phi = \dfrac{\sqrt{1-\delta^2}}{\delta}$

This response is plotted in Fig. The response is oscillatory and as $t \sim \infty$, it approaches



unity.

Step response of an underdamped second order system.

Time Domain Specifications of a Second Order System

The performance of a system is usually evaluated in terms of the following qualities.

I. How fast it is able to respond to the input,

2. How fast it is reaching the desired output,

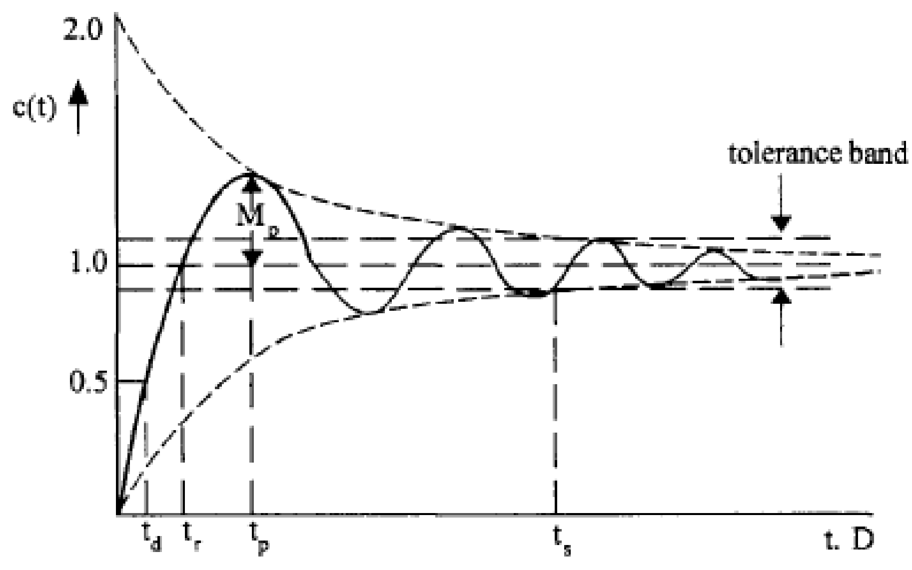
3. What is the error between the desired output and the actual output, once the transients die down and steady state is achieved,

4. Does it oscillate around the desired value, and

5. Is the output continuously increasing with time or is it bounded.

The last aspect is concerned with the stability of the system and we would require the system tobe stable. This aspect will be considered later. The first four questions will be answered in terms oftime domain specifications of the system based on its response to a unit step input. These are thespecifications to be given for the design of a controller for a given system. The step response of a typical underdamped second order system is plotted in Fig.

It is observed that, for an underdamped system, there are two complex conjugate poles. Usually,even if a system is of higher order, the two complex conjugate poles nearest to the *j OJ* - axis (calleddominant poles) are considered and the system is approximated by a second order system. Thus, indesigning any system, certain design specifications are given based on the typical underdamped step response shown as Fig.



Time domain specifications of a second order system

1. Delay time td*:* It is the time required for the response to reach 50% of the steady state value for the first time

2. Rise time tr: It is the time required for the response to reach 100% of the steady state value for under damped systems. However, for over damped systems, it is taken as the time required for the response to rise from 10% to 90% of the steadystate value.

3. Peak time tp*:* It is the time required for the response to reach the maximum or Peak value of the response.

4. Peak overshoot M_p: It is defined as the difference between the peak value of the response and the steady state value. It is usually expressed in percent of the steady state value. If the time for the peak is t_p percent peak overshoot is given by,

$$\text{Percent peak overshoot } M_p = \frac{c(t_p) - c(\infty)}{c(\infty)} \times 100.$$

For systems of type 1 and higher, the steady state value *c* (00) is equal to unity, the same as the input.

5. Settling time t_s: It is the time required for the response to reach and remain within a specified tolerance limits (usually ± 2% or ± 5%) around the steady state value.

6. Steady state error e_{ss}: It is the error between the desired output and the actual output as *t* ~ 00 or under steady state conditions. The desired output is given by the reference input *r (t)* and therefore,

$$e_{ss} = \underset{t \to \infty}{Lt} \, [r(t) - c(t)]$$

From the above specifications it can be easily seen that the time response of a system for a unit step input is almost fixed once these specifications are given. But it is to be observed that all the above specifications are not independent of each other and hence they have to be specified in such a way that they are consistent with others.

Let us now obtain the expressions for some of the above design specifications in terms of the damping factor and natural frequency.

1. ***Rise time (t_r)***

If we consider an underdamped system, from the definition of the rise time, it is the time required for the response to reach 100% of its steadystate value for the first time. Hence from eqn. (3.20).

$$C(t_r) = 1 = 1 - \frac{e^{-\delta\omega_n t_r}}{\sqrt{1-\delta^2}} \operatorname{Sin}\left(\omega_n\sqrt{1-\delta^2}\, t_r + \phi\right)$$

Or

$$\frac{e^{-\delta\omega_n t_r}}{\sqrt{1-\delta^2}} \operatorname{Sin}\left(\omega_n\sqrt{1-\delta^2}\, t_r + \phi\right) = 0$$

Since $\frac{e^{-\delta\omega_n t_r}}{\sqrt{1-\delta^2}}$ cannot be equal to zero,

$$\operatorname{Sin}(\omega_d t_r + \phi) = 0$$

$$\therefore \qquad \omega_d t_r + \phi = \pi$$

and

$$t_r = \frac{\pi - \phi}{\omega_n\sqrt{1-\delta^2}} = \frac{\pi - \tan^{-1}\frac{\sqrt{1-\delta^2}}{\delta}}{\omega_n\sqrt{1-\delta^2}}$$

2. ***Peak time (t_P)***

At the peak time, t_P, the response attains its maximum value and this can be obtained by differentiating $c(t)$ and equating it to zero. Thus,

$$\frac{dc(t)}{dt} = \frac{\delta\omega_n}{\sqrt{1-\delta^2}} e^{-\delta\omega_n t} \operatorname{Sin}(\omega_d t + \phi) - \frac{e^{-\delta\omega_n t}}{\sqrt{1-\delta^2}} \cos(\omega_d t + \phi).\ \omega_d = 0$$

Simplifying we have,

$$\delta \operatorname{Sin}(\omega_d t + \phi) - \sqrt{1-\delta^2} \cos(\omega_d t + \phi) = 0$$

This can be written as,

$$\cos\phi \sin(\omega_d t + \phi) - \sin\phi \cos(\omega_d t + \phi) = 0$$

where $$\tan\phi = \frac{\sqrt{1-\delta^2}}{\delta}$$

$\therefore$ $$\sin(\omega_d t + \phi - \phi) = \sin\omega_d t = 0$$

or $$\omega_d t = n\pi \qquad \text{for } n = 0, 1, 2, \ldots$$

Here

n = 0 Corresponds to its minimum value at t = 0

n = 1 Corresponds to its first peak value at $t = t_p$

n = 2 Corresponds to its first undershoot

n = 3 Corresponds to its second overshoot and so on

Hence for n = 1

$$t_p = \frac{\pi}{\omega_n\sqrt{1-\delta^2}}$$

3. *Peak overshoot* (M_p)

The peak overshoot is defined as

$$M_p = c(t_p) - 1$$

$$= 1 - \frac{e^{-\delta\omega_n t_p}}{\sqrt{1-\delta^2}} \sin(\omega_d t_p + \phi) - 1$$

$$= -\frac{e^{-\delta\omega_n t_p}}{\sqrt{1-\delta^2}} \sin\left(\omega_d \cdot \frac{\pi}{\omega_d} + \phi\right)$$

$$M_p = \frac{e^{-\frac{\delta\omega_n\pi}{\omega_n\sqrt{1-\delta^2}}}}{\sqrt{1-\delta^2}} \sin\phi \quad (\because \sin(\pi+\phi) = -\sin\phi)$$

$$= e^{\frac{-\pi\delta}{\sqrt{1-\delta^2}}} \qquad \left(\because \sin\phi = \sqrt{1-\delta^2}\right)$$

Hence, peak overshoot, expressed as a percentage of steady state value, is given by,

$$M_p = 100\, e^{\frac{-\pi\delta}{\sqrt{1-\delta^2}}}\ \% \qquad \ldots$$

4. ***Settling time (t_s)***

The time varying term in the step response, $c(t)$, consists of a product of two terms; namely, an exponentially delaying term, $\frac{e^{-\delta \omega_n t}}{\sqrt{1-\delta^2}}$ and a sinusoidal term, $Sin(\omega_d t + \phi)$. It is clear that the response is a decaying sinusoid, the envelop of which is given by $\frac{e^{-\delta \omega_n t}}{\sqrt{1-\delta^2}}$. Thus, the response reaches and remains within a given band, around the steadystate value, when this envelop crosses the tolerance band. Once this envelop reaches this value, there is no possibility of subsequent oscillations to go beyond these tolerane limits. Thus for a 2% tolerance band,

$$\frac{e^{-\delta \omega_n t_s}}{\sqrt{1-\delta^2}} = 0.02$$

For low values of δ, $\delta^2 << 1$ and therefore $e^{-\delta \omega_n t_s} \simeq 0.02$

$$\therefore \qquad t_s \simeq \frac{4}{\delta \omega_n} = 4\,\tau$$

where τ is the time constant of the exponential term.

5. ***Steady state error (e_{ss})***

For a type 1 system, considered for obtaining the design specifications of a second order control system, the steady state error for a step input is obviously zero. Thus

$$e_{ss} = \underset{t \to \infty}{Lt} \; 1 - c(t) = 0$$

The steady state error for a ramp input was obtained in eqn. (3.24) as $e_{ss} = \frac{2\delta}{\omega_n}$.

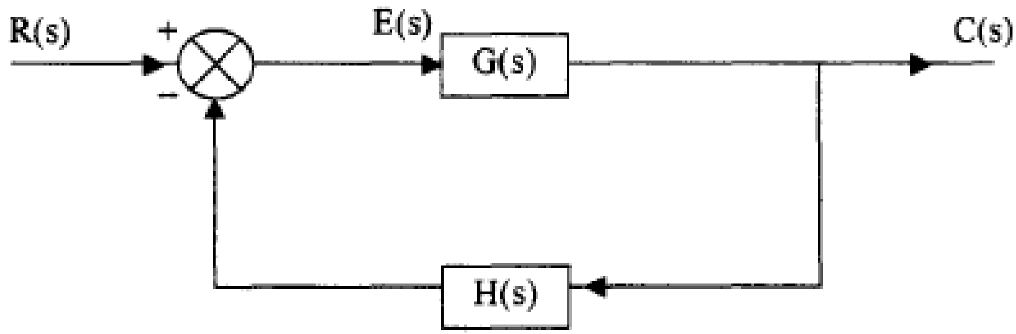
As the steadystate error, for various test signals, depends on the type of the system, it is dealt in the next section in detail.

Steady State Errors

One of the important design specifications for a control system is the steady state error. The steady state output of any system should be as close to desired output as possible. If it deviates from this desired output, the performance of the system is not satisfactory under steady state conditions. The steady state error reflects the accuracy of the system. Among many reasons for these errors, the most important ones are the type of input, the type of the system and the nonlinearities present in the system. Since the actual input in a physical system is often a random signal, the steady state errors are obtained for the standard test signals, namely, step, ramp and parabolic signals.

Error Constants

Let us consider a feedback control system shown in Fig.



Feedback Control System.

The error signal E (s) is given by

$$E(s) = R(s) - H(s)\,C(s)$$

But $\quad C(s) = G(s)\,E(s)$

From eqns. (3.31) and (3.32) we have

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

Applying final value theorem, we can get the steady state error e_{ss} as,

$$e_{ss} = \underset{s \to 0}{\text{Lt}}\; s\,E(s) = \underset{s \to 0}{\text{Lt}}\; \frac{sR(s)}{1+G(s)H(s)}$$

Above shows that the steady state error is a function of the input R(s) and the open loop transfer function G(s). Let us consider various standard test signals and obtain the steady state error for these inputs.

1. Unit step or position input.

For a unit step input, $R(s) = \frac{1}{s}$. Hence from eqn. (3.33)

$$e_s = \underset{s \to 0}{\text{Lt}}\; \frac{s.\frac{1}{s}}{1+G(s)H(s)}$$

$$= \frac{1}{1+\underset{s\to 0}{\text{Lt}}\; G(s)\,H(s)}$$

Let us define a useful term, *position error constant* K_p as,

$$K_p \triangleq \underset{s \to 0}{\text{Lt}}\; G(s)\,H(s)$$

In terms of the position error constant, e_{ss} can be written as,

$$e_{ss} = \frac{1}{1+K_p}$$

2. Unit ramp or velocity input.

For unit velocity input, $R(s) = \frac{1}{s^2}$ and hence,

$$e_{ss} = \underset{s \to 0}{Lt} \frac{s.\frac{1}{s}}{1+G(s)H(s)} = \underset{s \to 0}{Lt} \frac{1}{s+sG(s)H(s)}$$

$$= \frac{1}{\underset{s \to 0}{Lt}\ sG(s)H(s)}$$

Again, defining the *velocity error constant* K_v as,

$$K_v = \underset{s \to 0}{Lt}\ s\,G(s)\,H(s)$$

$$\therefore \qquad e_{ss} = \frac{1}{K_v}$$

3. Unit parabolic or acceleration input.

For unit acceleration input $R(s) = \frac{1}{s^3}$ and hence

$$e_{ss} = \underset{s \to 0}{Lt} \frac{s}{s^3[1+G(s)H(s)]} = \underset{s \to 0}{Lt} \frac{1}{s^2+s^2G(s)H(s)}$$

$$= \frac{1}{\underset{s \to 0}{Lt}\ s^2G(s)H(s)}$$

Defining the acceleration error constant K_a as,

$$K_a = \underset{s \to 0}{Lt}\ s^2\,G(s)\,H(s)$$

$$\therefore \qquad e_{ss} = \frac{1}{K_a}$$

For the special case of unity of feedback system, H (s)=1, and above equations are modified as,

$$K_p = \underset{s \to 0}{Lt}\ G(s)$$

$$K_v = \underset{s \to 0}{Lt}\ sG(s)$$

$$K_a = \underset{s \to 0}{Lt}\ s^2\,G(s)$$

In design specifications, instead of specifying the steady state error, it is a common practice to specify the error constants which have a direct bearing on the steady state error.

Dependence of Steady state Error on Type of the System

Let the loop transfer function G (s) H (s) or the open loop transfer function G (s) for a unity feedback system, be giv·en is time constant form.

$$G(s) = \frac{K(T_{z1}s+1)(T_{z2}s+1)----}{S^r(T_{p1}s+1)(T_{p2}s+1)----}$$

As $s \to 0$, the poles at the origin dominate the expression for $G(s)$. We had defined the type of a system, as the number of poles present at the origin. Hence the steady state error, which depends on $\underset{s \to 0}{Lt}\ G(s)$, $\underset{s \to 0}{Lt}\ s\,G(s)$ or $\underset{s \to 0}{Lt}\ s^2\,G(s)$, is dependent on the type of the system. Let us therefore obtain the steady state error for various standard test signals for type–0, type–1 and type–2 systems.

1. ***Type –0 system***

From eqn. (3.46) with $r = 0$, the error constants are given by

$$K_p = \underset{s \to 0}{Lt}\ G(s) \qquad = \underset{s \to 0}{Lt}\ \frac{K(\tau_{Z1}s+1)(\tau_{Z2}s+1)--}{(\tau_{p1}s+1)(\tau_{p2}s+1)--} = K$$

$$K_v = \underset{s \to 0}{Lt}\ s\,G(s) \qquad = \underset{s \to 0}{Lt}\ \frac{sK(\tau_{Z1}s+1)(\tau_{Z2}s+1)--}{(\tau_{p1}s+1)(\tau_{p2}s+1)--} = 0$$

Similarly $\qquad K_a = \underset{s \to 0}{Lt}\ s^2\,G(s) = 0$

The steady state errors for unit step, velocity and acceleration inputs are respectively,

$$e_{ss} = \frac{1}{1+K_P} = \frac{1}{1+K} \text{ (step input)}$$

$$e_{ss} = \frac{1}{K_v} = \infty \text{ (velocity input)}$$

$$e_{ss} = \frac{1}{K_a} = \infty \text{ (acceleration input)}$$

2. ***Type 1 system***

For type 1 system, $r = 1$ in eqn. (3.46) and

$$K_p = \underset{s \to 0}{Lt}\ G(s) \quad = Lt\ \frac{K}{s} = \infty$$

$$K_v = \underset{s \to 0}{Lt}\ s\,G(s) = \underset{s \to 0}{Lt}\ s.\ \frac{K}{s} = K$$

and

$$K_a = \underset{s \to 0}{Lt}\ s\,G(s) = \underset{s \to 0}{Lt}\ s^2.\ \frac{K}{s} = 0$$

The steady state error for unit step, unit velocity and unit acceleration inputs are respectively,

$$e_{ss} = \frac{1}{1+K_p} = \frac{1}{\infty} = 0 \qquad \text{(position)}$$

$$e_{ss} = \frac{1}{K_v} = \frac{1}{K} \qquad \text{(velocity)}$$

and

$$e_{ss} = \frac{1}{K_a} = \frac{1}{0} = \infty \qquad \text{(acceleration)}$$

3. ***Type 2-system***

For a type – 2 system $r = 2$ in eqn. (3.46) and

$$K_p = \underset{s \to 0}{Lt}\ G(s) \quad = \underset{s \to 0}{Lt}\ \frac{K}{s^2} = \infty$$

$$K_v = \underset{s \to 0}{Lt}\ s\,G(s) \quad = \underset{s \to 0}{Lt}\ \frac{sK}{s^2} = \infty$$

and

$$K_a = \underset{s \to 0}{Lt}\ s^2\,G(s) \quad = \underset{s \to 0}{Lt}\ \frac{s^2K}{s^2} = K$$

The steady state errors for the three test inputs are,

$$e_{ss} = \frac{1}{1+K_P} = \frac{1}{1+\infty} = 0 \quad \text{(position)}$$

$$e_{ss} = \frac{1}{K_v} = \frac{1}{\infty} = 0 \qquad \text{(velocity)}$$

and

$$e_{ss} = \frac{1}{K_a} = \frac{1}{K} \qquad \text{(acceleration)}$$

Thus a type zero system has a finite steady state error for a unit step input and is equal to

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

Where K is the positional error constant of system. It is customary to specify the gain of a type zero system by Kp rather than K.

Similarly, a type -1 system has a finite steady state error for a Ramp input only and is given by

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

K_v Velocity error constant.

Thus the gain of type -1 system in normally specified as Kv, A type -2 systems has a finite steady state error only for acceleration input and is given by

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

Where K_a is the acceleration error constant

As before, the gain of type -2 systems is specified as Ka rather than K.

Steady state errors for various inputs and type of systems:

The steady state errors, for various standard inputs for type - 0, type - 1 and type - 2 are summarized in Table

Standard input	Steadystate error e_{ss}		
	Type - 0 $K_p = \underset{s \to 0}{Lt}\ G(s)$	Type - 1 $K_v = \underset{s \to 0}{Lt}\ s\,G(s)$	Type - 2 $K_a = \underset{s \to 0}{Lt}\ s^2 G(s)$
Unit step	$\frac{1}{1+K_p}$	0	0
Unit velocity	∞	$\frac{1}{K_v}$	0
Unit acceleration	∞	∞	$\frac{1}{K_a}$

Types of Controllers:

A controller is one which compares controlled values with the desired values and has a function to correct the deviation produced.

Important uses of the controllers are:

1. Controllers improve steady state accuracy by decreasing the steady state errors.
2. As the steady state accuracy improves, the stability also improves.
3. They also help in reducing the offsets produced in the system.
4. Maximum overshoot of the system can be controlled using these controllers.

5. They also help in reducing the noise signals produced in the system.
6. Slow response of the over damped system can be made faster with the help of these controllers.

There are mainly two **types of controllers** and they are written below: **Continuous Controllers:** The main feature of continuous controllers is that the controlled variable (also known as the manipulated variable) can have any value within the range of controller's output. Now in the continuous controller's theory, there are three basic modes on which the whole control action takes place and these modes are written below. We will use the combination of these modes in order to have a desired and accurate output.

1. **Proportional controllers**.
2. **Integral controllers**.
3. **Derivative controllers**.

Combinations of these three controllers are written below:

4. Proportional and integral controllers.
5. Proportional and derivative controllers.

Now we will discuss each of these modes in detail.

Proportional Controllers

We cannot use **types of controllers** at anywhere, with each type controller, there are certain conditions that must be fulfilled. With **proportional controllers** there are two conditions and these are written below:

1. Deviation should not be large, it means there should be less deviation between the input and output.
2. Deviation should not be sudden.

Now we are in a condition to discuss proportional controllers, as the name suggests in a proportional controller the output (also called the actuating signal) is directly proportional to the error signal. Now let us analyze proportional controller mathematically. As we know in proportional controller output is directly proportional to error signal, writing this mathematically we have,

$$A(t) \propto e(t)$$

Removing the sign of proportionality we have,

$$A(t) = K_p \times e(t)$$

Where, K_p is proportional constant also known as controller gain. It is recommended that K_p should be kept greater than unity. If the value of K_p is greater than unity, then it will amplify the error signal and thus the amplified error signal can be detected easily.

Advantages of Proportional Controller

Now let us discuss some advantages of proportional controller.

1. Proportional controller helps in reducing the steady state error, thus makes the system more stable.
2. Slow response of the over damped system can be made faster with the help of these controllers.

Disadvantages of Proportional Controller

Now there are some serious disadvantages of these controllers and these are written as follows:

1. Due to presence of these controllers we some offsets in the system.
2. Proportional controllers also increase the maximum overshoot of the system.

Integral Controllers

As the name suggests in **integral controllers** the output (also called the actuating signal) is directly proportional to the integral of the error signal. Now let us analyze integral controller mathematically. As we know in an integral controller output is directly proportional to the integration of the error signal, writing this mathematically we have,

$$A(t) \propto \int_0^t e(t)dt$$

Removing the sign of proportionality we have,

$$A(t) = K_i \times \int_0^t e(t)dt$$

Where K_i is integral constant also known as controller gain. Integral controller is also known as reset controller.

Advantages of Integral Controller

Due to their unique ability they can return the controlled variable back to the exact set point following a disturbance that’s why these are known as reset controllers.

Disadvantages of Integral Controller

It tends to make the system unstable because it responds slowly towards the produced error.

Derivative Controllers

We never use **derivative controllers** alone. It should be used in combinations with other modes of controllers because of its few disadvantages which are written below:

1. It never improves the steady state error.
2. It produces saturation effects and also amplifies the noise signals produced in the system.

Now, as the name suggests in a derivative controller the output (also called the actuating signal) is directly proportional to the derivative of the error signal. Now let us analyze derivative controller mathematically. As we know in a derivative controller output is directly proportional to the derivative of the error signal, writing this mathematically we have,

$$A(t) \propto \frac{de(t)}{dt}$$

Removing the sign of proportionality we have,

$$A(t) = K_d \times \frac{de(t)}{dt}$$

Where, K_d is proportional constant also known as controller gain. Derivative controller is also known as rate controller.

Advantages of Derivative Controller

The major advantage of derivative controller is that it improves the transient response of the system.

Proportional and Integral Controller

As the name suggests it is a combination of proportional and an integral controller the output (also called the actuating signal) is equal to the summation of proportional and integral of the error signal. Now let us analyze proportional and integral controller mathematically. As we know in a proportional and integral controller output is directly proportional to the summation of proportional of error and integration of the error signal, writing this mathematically we have,

$$A(t) \propto \int_0^t e(t)dt + A(t) \propto e(t)$$

Removing the sign of proportionality we have,

$$A(t) = K_i \int_0^t e(t)dt + K_p e(t)$$

Where, K_i and k_p proportional constant and integral constant respectively.

Advantages and disadvantages are the combinations of the advantages and disadvantages of proportional and integral controllers.

Proportional and Derivative Controller

As the name suggests it is a combination of proportional and a derivative controller the output (also called the actuating signal) is equals to the summation of proportional and derivative of the error signal. Now let us analyze proportional and derivative controller mathematically. As we know in a proportional and derivative controller output is directly proportional to summation of proportional of error and differentiation of the error signal, writing this mathematically we have,

$$A(t) \propto \frac{de(t)}{dt} + A(t) \propto e(t)$$

Removing the sign of proportionality we have,

$$A(t) = K_d \frac{de(t)}{dt} + K_p e(t)$$

Where, K_d and k_p proportional constant and derivative constant respectively. Advantages and disadvantages are the combinations of advantages and disadvantages of proportional and derivative controllers

PID control stands for proportional plus derivative plus integral control. PID control is a feedback mechanism which is used in control system. This type of control is also termed as three term control. By controlling the three parameters - proportional, integral and derivative we can achieve different control actions for specific work. PID is considered to be the best controller in the control system family. Nicholas Minorsky published the theoretical analysis paper on PID controller. For PID control the actuating signal consists of proportional error signal added with derivative and integral of the error signal. Therefore, the actuating signal for PID control is -

$$e_a(t) = e(t) + T_d \frac{de(t)}{dt} + K_i \int e(t)dt$$

The Laplace transform of the actuating signal incorporating PID control is

$$E_a(s) = E(s) + sT_d E(s) + \frac{K_i}{s} E(s)$$

$$or,\ E_a(s) = E(s)\left[1 + sT_d + \frac{K_i}{s}\right]$$

There are some control actions which can be achieved by using any of the two parameters of the PID controller. Two parameters can work while keeping the third one to zero. So PID controller becomes sometimes PI (Proportion integral), PD (proportional derivative) or even P or I. The derivative term D is responsible for noise measurement while integral term is meant for reaching the targeted value of the system. In early days PID controller was used as mechanical device. These were pneumatic controllers as they were compressed by air. Mechanical controllers include spring, lever or mass. Many complex electronic systems are provided with PID control loop. In modern days PID controllers are used in PLC (programmable logic controllers) in the industry. The proportional, derivative and integral parameters can be expressed as - K_p, K_d and K_i. All these three parameters have effect on the closed loop control system. It affects rise time, settling time and overshoot and also the steady state error

Control Response	**Rise time**	**Settling time**	**Overshoot**	**Steady state error**
K_p	decrease	small change	increase	decrease
K_d	small change	decrease	decrease	no change
K_i	decrease	increase	increase	eliminate

PID control combines advantages of proportional, derivative and integral control actions. Let us discuss these control actions in brief. Proportional control : here actuating signal for the control action in a control system is proportional to the error signal. The error signal being the difference between the reference input signal and the feedback signal obtained from input.

PID controller have some limitations also apart from being one of the best controller in control action system. PID control is applicable to many control actions but it does not perform well in case of optimal control. Main disadvantage is the feedback path. PID is not provided with any model of the process. Other drawbacks are that PID is linear system and derivative part is noise sensitive. Small amount of noise can cause great change in the output.

UNIT-IV

Stability Analysis in S-Domain

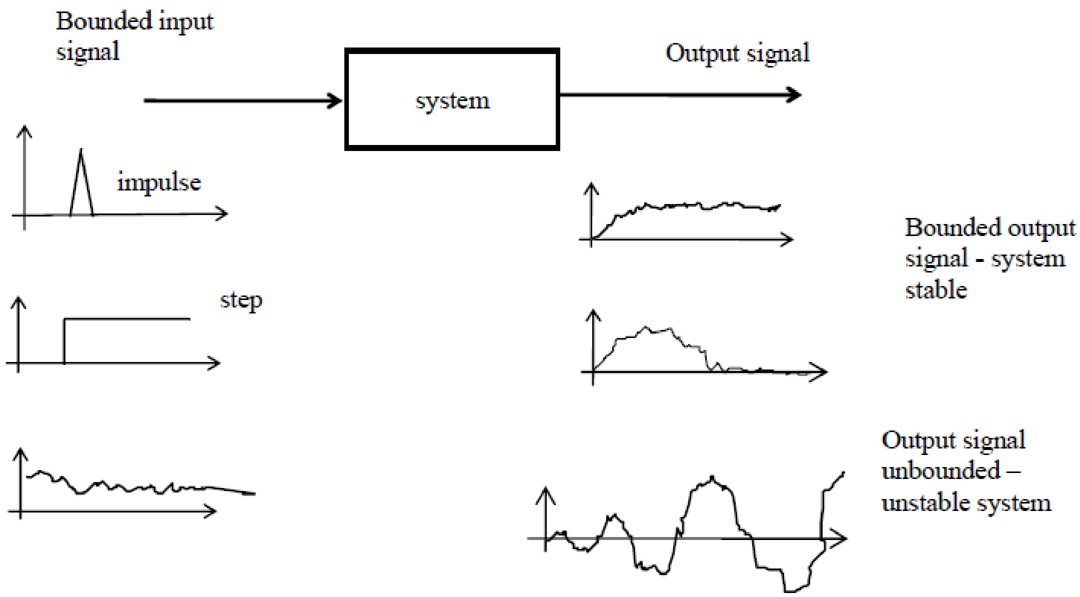
Stability:

Concept of Stability

Closed-loop feedback system is either stable or unstable. This type ofcharacterization is referred to as *absolute stability.*

Given that the system isstable, the degree of stability of the system is referred to as *relative stability*.

A *stable system* is defined as a system with bounded response to a boundedinput.



Theory of Network Synthesis
Network Functions

As the name suggests, in **theory of network synthesis** we are going to study about the synthesis of various networks which consists of both the active (resistors) and passive elements (inductors and capacitors).

In the frequency domain, **network functions** are defined as the quotient obtained by dividing the phasor corresponding to the circuit output by the phasor corresponding to the circuit input.

In simple words, **network functions** are the ratio of output phasor to the input phasor when phasors exists in frequency domain. The general form of network functions are given below:

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

Now with the help of the above general network function we are in position to describe the necessary conditions of the stability of all the network functions. There are three mains necessary conditions for the stability of these network functions and they are written below:

1. The degree of the numerator of F(s) should not exceed the degree of denominator by more than unity. In other words (m - n) should be less than or equal to one.

2. F(s) should not have multiple poles on the jω-axis or the y-axis of the pole-zero plot.

3. F(s) should not have poles on the right half of the s-plane.

Hurwitz Polynomial

If above all the stability criteria are fulfilled (i.e. we have stable network function) then the denominator of the F(s) is called the **Hurwitz polynomial**.

$$Let,\ F(s) = \frac{P(s)}{Q(s)}$$

Where, Q(s) is a **Hurwitz polynomial**.

Properties of Hurwitz Polynomials:

There are five important properties of Hurwitz polynomials and they are written below:

1. For all real values of s value of the function P(s) should be real.

2. The real part of every root should be either zero or negative.

3. Let us consider the coefficients of denominator of F(s) is b_n, $b_{(n-1)}$, $b_{(n-2)}$. . . . b_0. Here it should be noted that b_n, $b_{(n-1)}$, b_0 must be positive and b_n and $b_{(n-1)}$ should not be equal to zero simultaneously.

4. The continued fraction expansion of even to the odd part of the **Hurwitz polynomial** should give all positive quotient terms, if even degree is higher or the continued fraction expansion of odd to the even part of the Hurwitz polynomial should give all positive quotient terms, if odd degree is higher.

5. In case of purely even or purely odd polynomial, we must do continued fraction with the of derivative of the purely even or purely odd polynomial and rest of the procedure is same as mentioned in the point number (4).

From the above discussion we conclude one very simple result, **If all the coefficients of the quadratic polynomial are real and positive then that quadratic polynomial is always a Hurwitz polynomial.**

Positive Real Functions

Any function which is in the form of F(s) will be called as a **positive real function** if fulfill these four important conditions:

1. F(s) should give real values for all real values of s.

2. P(s) should be a Hurwitz polynomial.

3. If we substitute s = jω then on separating the real and imaginary parts, the real part of the function should be greater than or equal to zero, means it should be non negative. This most important condition and we will frequently use this condition in order to find out the whether the function is positive real or not.

4. On substituting s = jω, F(s) should posses simple poles and the residues should be real and positive.

Properties of Positive Real Function

There are four very important properties of **positive real functions** and they are written below:

1. Both the numerator and denominator of F(s) should be Hurwitz polynomials.
2. The degree of the numerator of F(s) should not exceed the degree of denominator by more than unity. In other words (m-n) should be less than or equal to one.
3. If F(s) is positive real function then reciprocal of F(s) should also be positive real function.
4. Remember the summation of two or more positive real function is also a positive real function but in case of the difference it may or may not be positive real function.

Following are the four necessary but not the sufficient conditions for the functions to be a positive real function and they are written below:

1. The coefficient of the polynomial must be real and positive.
2. The degree of the numerator of F(s) should not exceed the degree of denominator by more than unity. In other words (m - n) should be less than or equal to one.
3. Poles and zeros on the imaginary axis should be simple.
4. Let us consider the coefficients of denominator of F(s) is b_n, $b_{(n-1)}$, $b_{(n-2)}$. . . . b_0.Here it should be noted that b_n, $b_{(n-1)}$, b_0 must be positive and b_n and $b_{(n-1)}$ should not be equal to zero simultaneously.

Now there two necessary and sufficient conditions for the functions to be a positive real function and they are written below:

1. F(s) should have simple poles on the jω axis and the residues of these poles must be real and positive.
2. Summation of both numerator and denominator of F(s) must be a Hurwitz polynomial.

Routh Hurwitz Stability Criterion

If any pole of the system lies on the right hand side of the origin of the s plane, it makes the system unstable. On the basis of this condition A. Hurwitz and E.J.Routh started investigating the necessary and sufficient conditions of stability of a system. We will discuss two criteria for stability of the system. A first criterion is given by A. Hurwitz and this criterion is also known as **Hurwitz Criterion for stability** or **Routh Hurwitz Stability Criterion**.

Hurwitz Criterion

With the help of characteristic equation, we will make a number of Hurwitz determinants in order to find out the stability of the system. We define characteristic

equation of the system as
$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots\cdots\cdots\cdots a_{n-1} s^1 + a_n$$
Now there are n determinants for n^{th} order characteristic equation.
Let us see how we can write determinants from the coefficients of the characteristic equation. The step by step procedure for k^{th} order characteristic equation is written below:
Determinant one : The value of this determinant is given by |a1| where a1 is the coefficient of s^{n-1} in the characteristic equation.
Determinant two : The value of this determinant is given by
$$\begin{bmatrix} a_1 & a_3 \\ a_0 & a_2 \end{bmatrix}$$
Here number of elements in each row is equal to determinant number and we have determinant number here is two. The first row consists of first two odd coefficients and second row consists of first two even coefficients.
Determinant three : The value of this determinant is given by
$$\begin{bmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{bmatrix}$$
Here number of elements in each row is equal to determinant number and we have determinant number here is three. The first row consists of first three odd coefficients, second row consists of first three even coefficients and third row consists of first element as zero and rest of two elements as first two odd coefficients.
Determinant four: The value of this determinant is given by,
$$\begin{bmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{bmatrix}$$
Here number of elements in each row is equal to determinant number and we have determinant number here is four. The first row consists of first three four coefficients, second row consists of first four even coefficients, third row consists of first element as zero and rest of three elements as first three odd coefficients the fourth row consists of first element as zero and rest of three elements as first three even coefficients. By following the same procedure we can generalize the determinant formation. The general form of determinant is given below:
$$\begin{bmatrix} a_1 & a_3 & a_5 & a_7 & \cdot & \cdot & \cdot & a_{2k-1} \\ a_0 & a_2 & a_4 & a_6 & \cdot & \cdot & \cdot & a_{2k-2} \\ 0 & a_1 & a_3 & a_5 & \cdot & \cdot & \cdot & a_{2k-3} \\ 0 & a_0 & a_2 & a_4 & \cdot & \cdot & \cdot & a_{2k-4} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & ak \end{bmatrix}$$
Now in order to check the stability of the above system, calculate the value of each determinant. The system will be stable if and only if the value of each determinant is greater than zero, i.e. the value of each determinant should be positive. In all the other cases the system will not be stable.

Routh Stability Criterion

This criterion is also known as modified Hurwitz Criterion of stability of the system. We will study this criterion in two parts. Part one will cover necessary condition for stability of the system and part two will cover the sufficient condition for the stability of the system.

Let us again consider the characteristic equation of the system as

$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots\cdots\cdots\cdots a_{n-1} s^1 + a_n 1)$$

Part one (necessary condition for stability of the system): In this we have two conditions which are written below:

1. All the coefficients of the characteristic equation should be positive and real.
2. All the coefficients of the characteristic equation should be non zero.

2)Part two (sufficient condition for stability of the system): Let us first construct routh array. In order to construct the routh array follow these steps:

- The first row will consist of all the even terms of the characteristic equation.

Arrange them from first (even term) to last (even term). The first row is written below: a_0 a_2 a_4 a_6............

- The second row will consist of all the odd terms of the characteristic equation.

Arrange them from first (odd term) to last (odd term). The first row is written below: a_1 a_3 a_5 a_7...........

- The elements of third row can be calculated as:

(1) First element : Multiply a0 with the diagonally opposite element of next column (i.e. a3) then subtract this from the product of a_1 and a_2 (where a_2 is diagonally opposite element of next column) and then finally divide the result so obtain with a_1. Mathematically we write as first element

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

(2) Second element : Multiply a_0 with the diagonally opposite element of next to next column (i.e. a_5) then subtract this from the product of a_1 and a_4 (where a_4 is diagonally opposite element of next to next column) and then finally divide the result so obtain with a_1.

Mathematically we write as second element

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

Similarly, we can calculate all the elements of the third row. (d) The elements of fourth row can be calculated by using the following procedure:

(1) First element : Multiply b_1 with the diagonally opposite element of next column (i.e. a_3) then subtract this from the product of a_1 and b_2 (where b_2 is diagonally opposite element of next column) and then finally divide the result so obtain with b_1. Mathematically we write as first element

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

(2) Second element :Multiply b_1 with the diagonally opposite element of next to next column (i.e. a_5) then subtract this from the product of a_1 and b_3 (where b_3 is diagonally opposite element of next to next column) and then finally divide the result so obtain with a_1. Mathematically we write as second element

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

Similarly, we can calculate all the elements of the fourth row. Similarly, we can calculate all the elements of all the rows. Stability criteria if all the elements of the first column are positive then the system will be stable.

However if anyone of them is negative the system will be unstable. Now there are some special cases related to Routh Stability Criteria which are discussed below:

(1) Case one: If the first term in any row of the array is zero while the rest of the row has at least one non zero term. In this case we will assume a very small value (ε) which is tending to zero in place of zero. By replacing zero with (ε) we will calculate all the elements of the Routh array. After calculating all the elements we will apply the limit at each element containing (ε). On solving the limit at every element if we will get positive limiting value then we will say the given system is stable otherwise in all the other condition we will say the given system is not stable.

(2) Case second : When all the elements of any row of the Routh array are zero. In this case we can say the system has the symptoms of marginal stability. Let us first understand the physical meaning of having all the elements zero of any row. The physical meaning is that there are symmetrically located roots of the characteristic equation in the s plane. Now in order to find out the stability in this case we will first find out auxiliary equation. Auxiliary equation can be formed by using the elements of the row just above the row of zeros in the Routh array. After finding the auxiliary equation we will differentiate the auxiliary equation to obtain elements of the zero row. If there is no sign change in the new routh array formed by using auxiliary equation, then in this we say the given system is limited stable. While in all the other cases we will say the given system is unstable.

Root Locus Technique in Control System | Root Locus Plot:

Any physical system is represented by a transfer function in the form of

$$G(s) = k \times \frac{numerator\ of\ s}{denomerator\ of\ s}$$

We can find poles and zeros from G(s). The location of poles and zeros are crucial keeping view stability, relative stability, transient response and error analysis. When the system put to service stray inductance and capacitance get into the system, thus changes the location of poles and zeros. In **root locus technique in control system** we will evaluate the position of the roots, their locus of movement and associated information. These information will be used to comment upon the system performance.

Some of the advantages of root locus technique are written below.

Advantages of Root Locus Technique

1. Root locus technique in control system is easy to implement as compared to other methods.
2. With the help of root locus we can easily predict the performance of the whole system.
3. Root locus provides the better way to indicate the parameters.

Now there are various terms related to root locus technique that we will use frequently in this article.

1. Characteristic Equation Related to Root Locus Technique :

$$1 + G(s)H(s) = 0 \text{ is known as characteristic equation.}$$

Now on differentiating the characteristic equation and on equating dk/ds equals to zero, we can get break away points.

2. Break away Points : Suppose two root loci which start from pole and moves in opposite direction collide with each other such that after collision they start moving in different directions in the symmetrical way. Or the break away points at which multiple roots of the characteristic equation 1 + G(s)H(s)= 0 occur.

The value of K is maximum at the points where the branches of root loci break away. Break away points may be real, imaginary or complex.

3. Break in Point : Condition of break in to be there on the plot is written below : Root locus must be present between two adjacent zeros on the real axis.

4. Centre of Gravity : It is also known centroid and is defined as the point on the plot from where all the asymptotes start. Mathematically, it is calculated by the difference of summation of poles and zeros in the transfer function when divided by the difference of total number of poles and total number of zeros. Centre of gravity is always real & it is denoted by σ_A.

$$\sigma_A = \frac{(Sum\ of\ real\ parts\ of\ poles) - (Sum\ of\ real\ parts\ of\ zeros)}{N - M}$$

Where, N is number of poles and M is number of zeros.

5. Asymptotes of Root Loci : Asymptote originates from the center of gravity or centroid and goes to infinity at definite some angle. Asymptotes provide direction to the root locus when they depart break away points.

6. Angle of Asymptotes : Asymptotes makes some angle with the real axis and this angle can be calculated from the given formula,

$$Angle\ of\ asymptotes = \frac{(2p+1)\times 180}{N - M}$$

Where, p = 0, 1, 2 .... (N-M-1) N is the total number of poles M is the total number of zeros.

7. Angle of Arrival or Departure : We calculate angle of departure when there exists complex poles in the system. Angle of departure can be calculated as 180-{(sum of angles to a complex pole from the other poles)-(sum of angle to a complex pole from the zeros)}.

8. Intersection of Root Locus with the Imaginary Axis : In order to find out the point of intersection root locus with imaginary axis, we have to use Routh Hurwitz criterion. First, we find the auxiliary equation then the corresponding value of K will give the value of the point of intersection.

9. Gain Margin : We define gain margin as a by which the design value of the gain factor can be multiplied before the system becomes unstable. Mathematically it is given by the formula

$$Gain\ margin = \frac{Value\ of\ K\ at\ the\ imaginary\ axes\ cross\ over}{Design\ value\ of\ K}$$

10. Phase Margin : Phase margin can be calculated from the given formula:

$$Phase\ margin = 180 + \angle(G(jw)H(jw))$$

11. Symmetry of Root Locus : Root locus is symmetric about the x axis or the real axis.

How to determine the value of K at any point on the root loci ? Now there are two ways of determining the value of K, each way is described below.

1. Magnitude Criteria : At any points on the root locus we can apply magnitude criteria as, $|G(s)H(s)| = 1$
2. Using this formula we can calculate the value of K at any desired point.
3. Using Root Locus Plot : The value of K at any s on the root locus is given by

$$K = \frac{product\ of\ all\ of\ the\ vector\ lengths\ drawn\ from\ the\ poles\ of\ G(s)H(s)\ to\ s}{product\ of\ all\ of\ the\ vector\ lengths\ drawn\ from\ the\ zeros\ of\ G(s)H(s)\ to\ s}$$

Root Locus Plot

This is also known as root locus technique in control system and is used for determining the stability of the given system. Now in order to determine the stability of the system using the root locus technique we find the range of values of K for which the complete performance of the system will be satisfactory and the operation is stable. Now there are some results that one should remember in order to plot the root locus. These results are written below:

1. Region where root locus exists : After plotting all the poles and zeros on the plane, we can easily find out the region of existence of the root locus by using one simple rule which is written below,
Only that segment will be considered in making root locus if the total number of poles and zeros at the right hand side of the segment is odd.

2. How to calculate the number of separate root loci ? : A number of separate root loci are equal to the total number of roots if number of roots are greater than the number of poles otherwise number of separate root loci is equal to the total number of poles if number of roots are greater than the number of zeros.

Procedure to Plot Root Locus

Keeping all these points in mind we are able to draw the **root locus plot** for any kind of system. Now let us discuss the procedure of making a root locus.

1. Find out all the roots and poles from the open loop transfer function and then plot them on the complex plane.
2. All the root loci starts from the poles where k = 0 and terminates at the zeros where K tends to infinity. The number of branches terminating at infinity equals to the difference between the number of poles & number of zeros of G(s)H(s).
3. Find the region of existence of the root loci from the method described above after finding the values of M and N.
4. Calculate break away points and break in points if any.
5. Plot the asymptotes and centroid point on the complex plane for the root loci by calculating the slope of the asymptotes.
6. Now calculate angle of departure and the intersection of root loci with imaginary axis.
7. Now determine the value of K by using any one method that I have described above.

By following above procedure you can easily draw the **root locus plot** for any open loop transfer function.

8. Calculate the gain margin.
9. Calculate the phase margin.
10. You can easily comment on the stability of the system by using Routh array.

UNIT – V
FREQUENCY RESPONSE ANALYSIS

The frequency response is the steady state response (output) of a system when the input to the system is a sinusoidal signal. Consider a linear system, with sinusoidal input

$$r(t)=A\sin wt$$

Under steady- state, the system output as well as the signals at all other points in the system are sinusoidal. The steady- state output may be written as

$$c(t)= B\sin(wt+\Phi)$$

The magnitude and phase relation between the sinusoidal input and steady -state output of a system is termed the frequency response. In linear time- invariant systems, the frequency response is independent of the amplitude and phase of the input signal.
The frequency response test on a system or component is normally performed by keeping the amplitude A fixed and determining B and Φ for suitable range of frequencies.

The advantages of frequency response analysis are as follows:

1.The absolute and relative stability of the closed loop system can be estimated from the knowledge of their open loop frequency response.
2.The practical testing of systems can be easily carried with available sinusoidal signal generators and precise measurement equipment.
3.The transfer function of complicated systems can be determined experimentally by frequency response tests.
4.The design parameter adjustment of the open loop transfer function of a system for specified closed loop performance is carried out more easily in frequency domain.
5.The frequency response analysis and design can be extended to certain nonlinear control systems.
6.The effects of noise disturbance and parameter variations are relatively easy to visualize and asses through frequency response.

The draw backs of frequency response method as follows

1.For system with very large time constants, the frequency response test is cumbersome to perform, as the time required for the output to reach the steady-state for each frequency of the test signal is excessively long. Therefore, the frequency response test is not recommended for systems with very large time constants.
2.The frequency response test obviously cannot be performed on non-interruptible systems. Under such circumstances ,a single shot test is more convenient even though the computation of the transfer function from it gets involved.

FREQUENCY DOMAIN SPECIFICATIONS

The performance characteristics of a system in frequency domain are measured in terms of frequency domain specifications. the requirements of a system to be designed are usually specified in terms of these specifications.
The frequency domain specifications are
1.Resonant Peak, M_r
2.Resonanat Frequency, W_r
3.Bandwidth
4.Cut-off rate
4.Gain Margin
Phase Margin

RESONANT PEAK(M_r)

The Maximum value of the magnitude of a closed loop transfer function is called the resonant peak . A large resonant peak corresponds to a large overshoot in transient response.

RERSONANT FREQUENCY(W_r)

The frequency at which the resonant peak occurs is called resonant frequency. This is related to the frequency of oscillation in the step response and thus it is indicative of the speed of the time response.

BAND WIDTH

The Bandwidth is the range of frequencies for which the system gain is more than -3db. The frequency at which the gain is -3db is called cut-off frequency .Band width is usually defined for closed loop system and it transmits the signal s whose frequencies are less than the cut-off frequency. The Band width is a measure of the ability of a feedback system to produce the input signal ,noise rejection characteristics and rise time. A large bandwidth corresponds to a small rise time or fast response.

GAIN MARGIN

The gain margin, is defined as the reciprocal of the magnitude of openloop transfer function at phase crossover frequency.The frequency at which the phase of open loop transfer function is 180^0 is called the phase cross over frequency W_{pc}.

$$\text{Gain Margin } K_g = \frac{1}{|G(jw_{pc})|}$$

The gain margin in dbcanbe expressed as

$$K_g \text{ in db}=20\log K_g = 20\log \frac{1}{|G(jw_{pc})|} = -20\log |G(jw_{pc})|$$

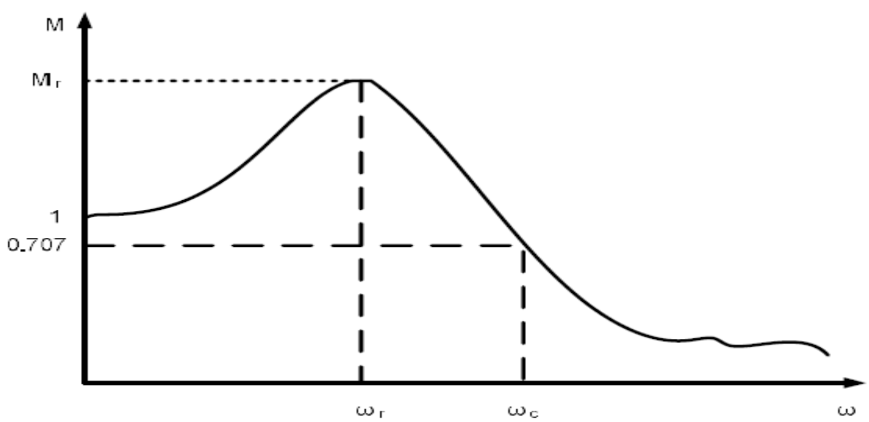
PHASE MARGIN

The phase margin (γ)

The phase margin γ , is that amount of additional phase lag at the gain cross over frequency required to bring the system to the verge of instability. The gain crossover frequency W_{gc} is the frequency at which the magnitude of the open loop transfer function is unity (or is the frequency at which the db magnitude is zero).

The phased margin γ is obtained by adding 180^0 to the phase angle Φ of the open loop transfer function at the gain cross over frequency.

Phase Margin $\gamma = 180^0 + \Phi_{gc}$, where $\Phi_{gc} = \angle G(Jw_{pc})$



Correlation between time and frequency response

For second order system

$$\boxed{\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (1)}$$

Putting $S=jw$

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

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

--------(2)

Let, $u = \frac{\omega}{\omega_n}$, then $\frac{C(j\omega)}{R(j\omega)} = \frac{1}{(1-u^2) + j2\zeta u}$

-------(3)

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

-----------(4)

Where,

$$|M(j\omega)| = \frac{1}{\sqrt{(1-u^2)^2 + (2\zeta u)^2}}$$

$$\theta = -\tan^{-1}\left(\frac{2\zeta u}{1-u^2}\right)$$

-------(5)

Now,

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

$$\omega_r = \omega_n\sqrt{1-2\zeta^2}$$

-----------(6)

$$\omega_b = \omega_n\sqrt{1-2\zeta^2+\sqrt{4\zeta^4-4\zeta^2+2}}$$

$$PM = -180^0 + \varphi$$

--------(7)

Where, $\varphi = \tan^{-1}\frac{2\zeta}{\sqrt{\sqrt{4\zeta^2+1}-2\zeta^2}}$

------------(8)

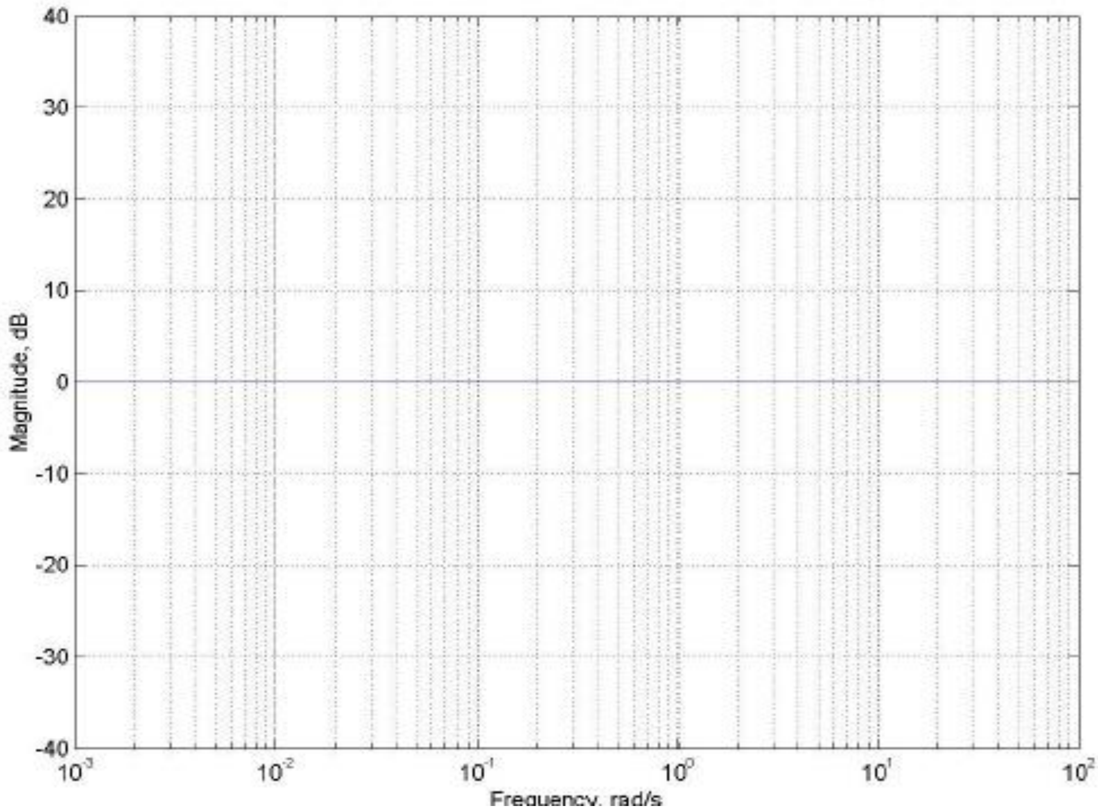
BODE PLOT

Basic of any frequency response is to plot magnitude M and angle Φ against input frequency *w*.When *'w'* is varied from 0 to ∞ there is wide range of variations in M and Φ and hence it becomes difficult to accommodate all such variations with linear scale. Hence H.W.Bode suggested the method in which logarithmic values of Magnitude and are to be plotted against logarithmic values of frequencies.

Bode plot consists of two graphs. One is a plot of the magnitudea sinusoidal transfer function versus log*w*. The other is a plot of the phase angle of a sinusoidal transfer function versus logw.

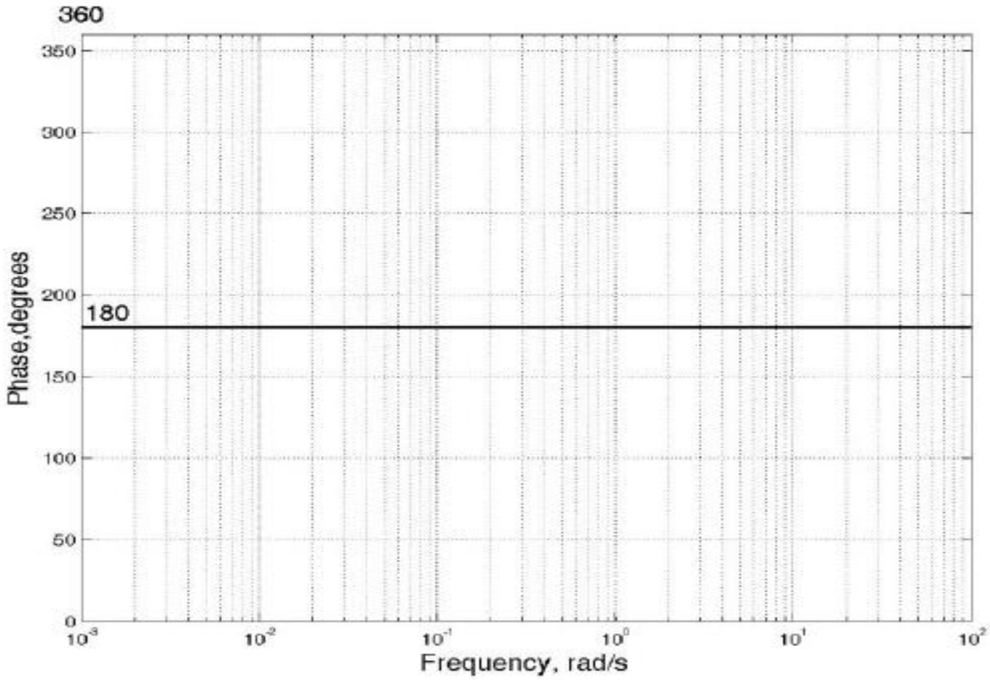
Therefore to sketch the magnitude plot, a knowledge of the magnitudevariations of individual factors is essential.

Magnitude plot and phase plot on a semi log graph sheet



Magnitude plot

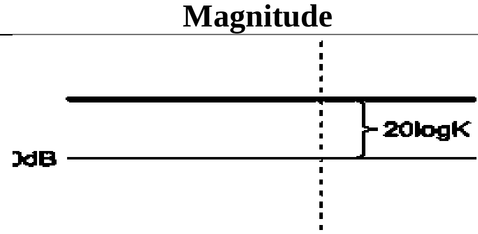
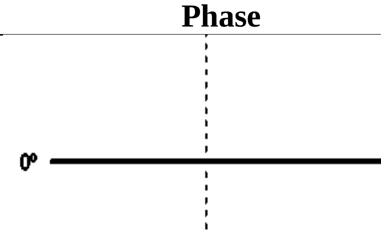
M=*20log|G(jw)H(jw)|* dB

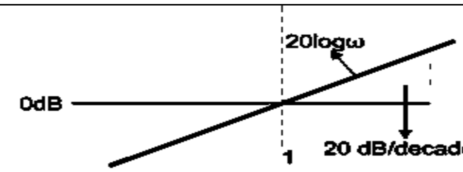
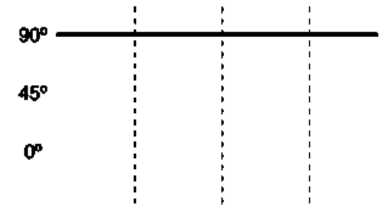
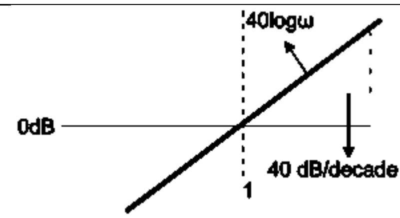
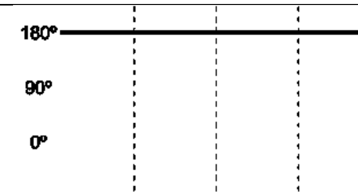
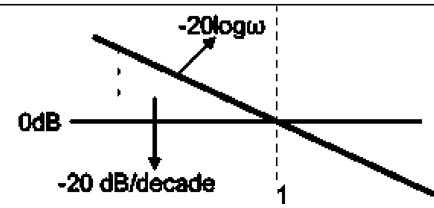
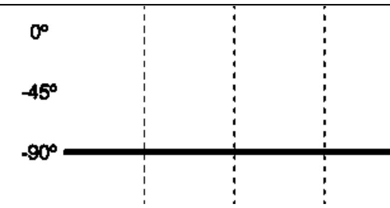
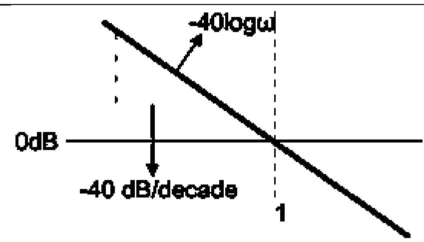
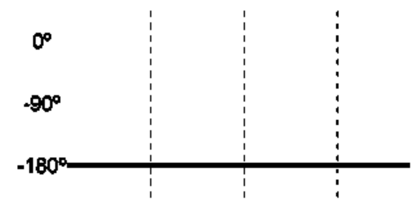
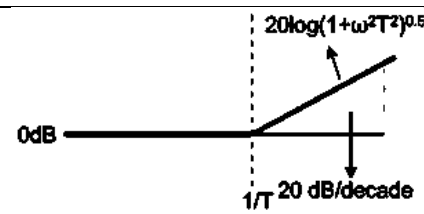
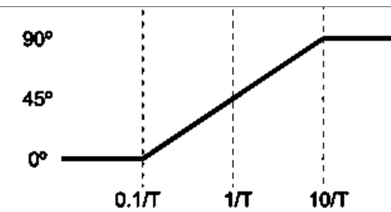
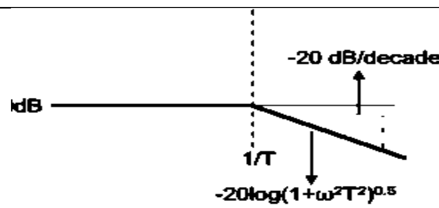
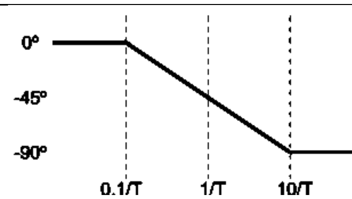


Phase Plot

Basic frequency response factors

No	Laplace term	Frequency response	Type of factor
1	K	K	Constant
2	S	*Jw*	Derivative Factor
3	1/S	*1/jw*	Integrating Factor
4	τ s+1	$(1+jw\tau)$	First order derivative factor
5	1/(τ s+1)	$1/(1+jw\tau)$	First order integral factor
6	$S^2+2\zeta w_n+w_n^2$	$w_n^2-w^2+2\zeta w_n w$	Second order derivative factor
7	$1/(S^2+2\zeta w_n+w_n^2)$	$1/(w_n^2-w^2+2\zeta w_n w)$	Second order integral factor

Factor	**Magnitude**	**Phase**
Gain, *K*		

$j\omega$		
$j\omega^2$		
$1/j\omega$		
$1/j\omega^2$		
$1+j\omega T$		
$1/(1+j\omega T)$		

The step by step procedure for plotting magnitude plot

Step1:

1.Convert the transfer function into Bode form or time constant form .The Bodefprm of the transfer function is

$$\backslash G(s) = \frac{K(1+ST_1)}{S(1+ST_2)(1+\frac{S^2}{w_{n^2}}+2\zeta\frac{s}{w_n})} \qquad \text{--------(1)}$$

$$G(s)= \frac{K(1+jwT_1)}{jw(1+jwT_2)(1-\frac{w^2}{w_{n^2}})(1-\frac{w^2}{w_{n^2}}+j2\zeta\frac{w}{w_n})} \qquad \text{-------(2)}$$

Step2:
List the corner frequencies in the increasing order and prepare a table as shown below.

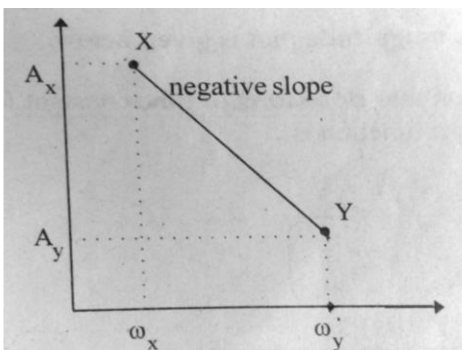
Term	Corner frequency rad/sec	Slope db/dec	Change in slope db/dec

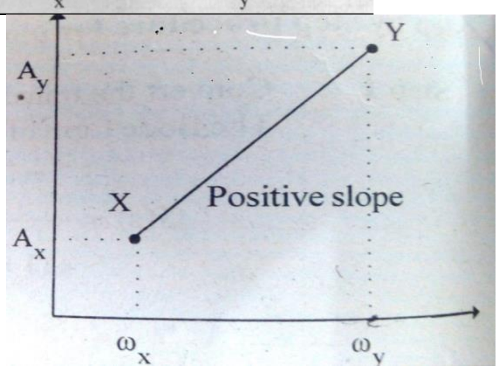
Step:3
In the above table enter K or $\frac{K}{(jw)^n}$ or K(Jw)^n as the first term and the other terms in the increasing order of corner frequencies .Then enter the corner frequency ,Slope contributed by each term and change in slope at every corner frequency.

Step4:
Then Calculate the gain (db magnitude) at every corner frequency one by one using the formula

$$\text{Gain at } \boldsymbol{W_y} = \text{Change in gain from } w_x \text{ to } w_y \pm \text{Gain at } w_x$$

$$= [\text{Slope from } w_x \text{to} w_y * \log \frac{w_x}{w_y}] + \text{Gain at } w_x$$





Step 5:Choose an arbitrary frequency which is greater than the highest corner frequency .calculate the gain at w_h by using the formula in step 4.
Step 6: In semi log graph sheet mark the required range of frequency on x-axis (Log scale) and the range of db on Y axis (Ordinary scale) after choosing proper scale.
Step 7: Mark all he points obtained in steps 3,4 and 5 on the graph and join the points by straight lines. Mark the slope at every part of the graph

Procedure For Phase Plot of Bode Plot

The phase plot is an exact plot and no approximations are made while drawing the phase plot. Hence the exact phase angles G(jw) are computed for various values of w and tabulated. The choice of frequencies are preferably the frequencies chosen for magnitude plot. Usually the magnitude plot and phase plot are drawn in a single semilog-sheet on a common frequency.

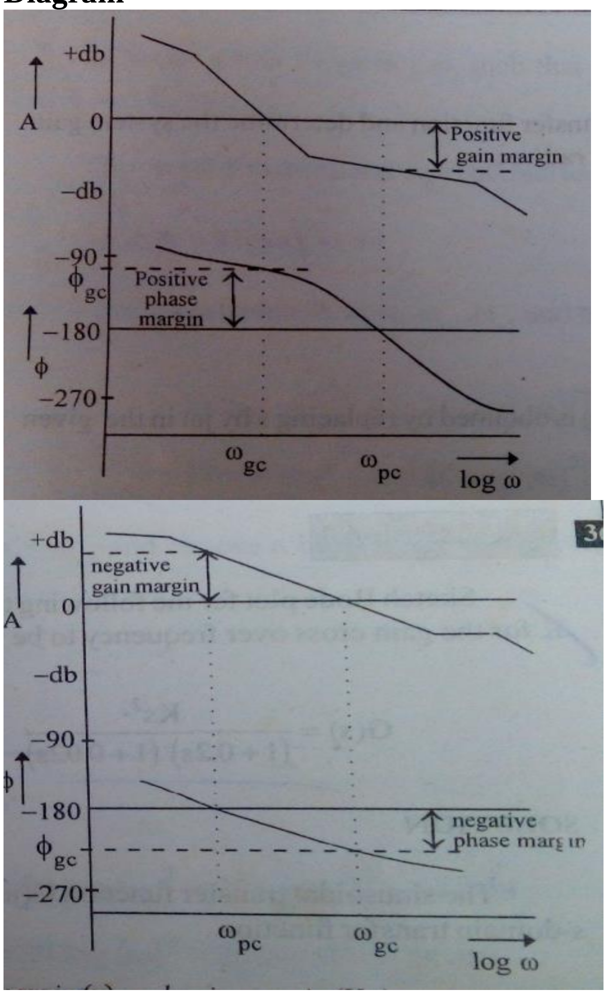
Take another y-axis in the graph where the magnitude plot is drawn and in this y-axis mark the desired range of phase angles after choosing proper scale. From the tabulated values of *w* and phase angles, mark all the points on the graph. Join the points by smooth line.

Determination of Gain Margin and Phase Margin

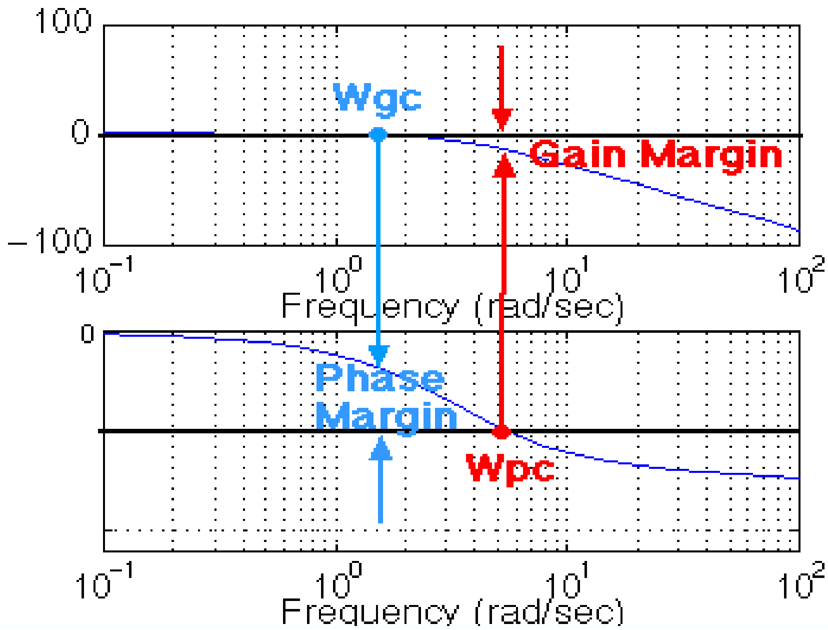
The gain margin in **db** is given by the negative of db magnitude of G(*jw*) at the phase cross over frequency ,$w_{pc.}$ The $w_{pc.}$ Is the frequency at which the phase of G(jw) is -180^0. If the db magnitude of G(jw) at w_{pc} is neagative then gain margin is positive and vice versa.

Let ϕ_{gc} be the phase angle of G(jw) at gain crossover frequency W_{gc} .The W_{gc} is the frequency at which the db magnitude of G(jw) is zero. Now the phase Margin, γ is given by $\gamma = 180^0 + \phi_{gc}$. If ϕ_{gc} is less negative than -180^0 then phase margin is positive and viceversa.

Diagram



Bode plot showing Phase margin and gain margin



Gain Adjustment in bode plot:

In the openloop transfer function G(jw) the constant K contributes only magnitude.

Hence by changing the value of K the system gain can be adjusted to meet the desired specifications. The desired specifications are gain margin, phase margin, gain crossover frequency, phase cross over frequency. In a system transfer function if the value of K require to be estimated to satisfy a desired specification then draw the bode plot of the system with K=1.the constant K

Can add 20logK to every point of magnitude plot and due to this addition the magnitude plot will shift vertically up or down. Hence shift the magnitude plot vertically up or down to meet ted desired specifications. Equate the vertical distance by which the magnitude plot is shifted to 20logK and solve for K.

Let x= change in db

Now 20log k = x: log k = x/20 $K=10^{(x/20)}$

POLAR PLOT

The polar plot of a sinusoidal transfer function G(jw) is a plot of the magnitude of G(jw) versus the phase angle of G(jw)on a polar coordinates as *w* varied from zero to infinity. Thus the polar plot is locus vectors |G(jw)|∠G(jw) as *w* varied from zero to infinity. In the polar plot the magnitude of G(jω) is plotted as the distance from the origin while phase angle is measured from positive real axis.

Steps to draw the polar plot

Step 1: Determine the T.F G(s)

Step 2: Put s=jω in the G(s) $|G(j\omega)|$

Step 3: At ω=0 &ω=∞ find by &

Step 4: At ω=0 &ω=∞ findby & $\lim_{\omega\to\infty}\angle G(j\omega)$

Step 5: Rationalize the function G(jω) and *separate* the real and imaginary parts

Step 6: Put Re [G(jω)]=0, determine the frequency at which plot intersects the Im axis and calculate intersection value by putting the above calculated frequency in G(jω)

Step 7: Put Im [G(jω)]=0, determine the frequency at which plot intersects the real axis and calculate intersection value by putting the above calculated frequency in G(jω)

Step 8: Sketch the Polar Plot with the help of above information .

Polar plot for TYPE-0 system.

$\lim_{\omega\to\infty}|G(j\omega)|$

Let

$$G(s)=\frac{K}{(1+sT_1)(1+sT_2)}\quad \angle G(j\omega)\qquad \lim_{\omega\to 0}\angle G(j\omega)$$

Step 1: Put s=jω

$\lim_{\omega\to 0}|G(j\omega)|$

Step 2: Taking the limit for magnitude of G(jω)

$$G(j\omega)=\frac{K}{(1+j\omega T_1)(1+j\omega T_2)}$$

$$=\frac{K}{\sqrt{1+(\omega T_1)^2}\sqrt{1+(\omega T_2)^2}}\angle-\tan^{-1}\omega T_1-\tan^{-1}\omega T_2$$

$$\lim_{\omega\to0}|G(j\omega)|=\frac{K}{\sqrt{1+(\omega T_1)^2}\sqrt{1+j(\omega T_2)^2}}=K$$

$$\lim_{\omega\to\infty}|G(j\omega)|=\frac{K}{\sqrt{1+(\omega T_1)^2}\sqrt{1+j(\omega T_2)^2}}=0$$

Step 3: Taking the limit of the Phase Angle of G(jω)

$$\lim_{\omega\to0}\angle G(j\omega)=\angle-\tan^{-1}\omega T_1-\tan^{-1}\omega T_2=0$$

$$\lim_{\omega\to\infty}\angle G(j\omega)=\angle-\tan^{-1}\omega T_1-\tan^{-1}\omega T_2=-180$$

Step 4: Separate the real and Imaginary part of G(jω)

$$G(j\omega)=\frac{K(1-\omega^2T_1T_2)}{1+\omega^2T_1^2+\omega^2T_2^2+\omega^4T_1T_2}-j\frac{K\omega(T_1+T_2)}{1+\omega^2T_1^2+\omega^2T_2^2+\omega^4T_1T_2}$$

Step 5: Put Re [G(jω)]=0

$$\frac{K(1-\omega^2T_1T_2)}{1+\omega^2T_1^2+\omega^2T_2^2+\omega^4T_1T_2}=0\Rightarrow\omega=\frac{1}{\sqrt{T_1T_2}}\ \&\ \omega=\infty$$

So When

$$\omega=\frac{1}{\sqrt{T_1T_2}}\Rightarrow G(j\omega)=\frac{K\sqrt{T_1T_2}}{T_1+T_2}\angle-90^0$$

$$\&\ \omega=\infty\Rightarrow G(j\omega)=0\angle-180^0$$

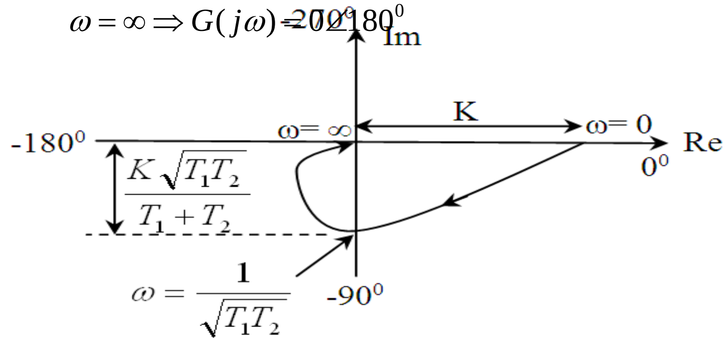
Step 6: Put Im [G(jω)]=0

$$\frac{K\omega(T_1+T_2)}{1+\omega^2T_1^2+\omega^2T_2^2+\omega^4T_1T_2}=0\Rightarrow\omega=0\ \&\ \pm\infty$$

So When

$$\omega=0\Rightarrow G(j\omega)=K\angle0^0$$

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



Polar plot of type zero system

Polar Plot for Type -1

Let

$$G(s)=\frac{K}{s(1+sT_1)(1+sT_2)}$$

Step 1: Put s=jω

$$G(j\omega)=\frac{K}{j\omega(1+j\omega T_1)(1+j\omega T_2)}$$

$$=\frac{K}{\omega\sqrt{1+(\omega T_1)^2}\sqrt{1+j(\omega T_2)^2}}\angle-90^0-\tan^{-1}\omega T_1-\tan^{-1}\omega T_2$$

Step 2: Taking the limit for magnitude of G(jω)

$$\lim_{\omega\to0}|G(j\omega)|=\frac{K}{\omega\sqrt{1+(\omega T_1)^2}\sqrt{1+j(\omega T_2)^2}}=\infty$$

$$\lim_{\omega\to\infty}|G(j\omega)|=\frac{K}{\omega\sqrt{1+(\omega T_1)^2}\sqrt{1+j(\omega T_2)^2}}=0$$

Step 3: Taking the limit of the Phase Angle of G(jω)

$$\lim_{\omega\to 0} \angle G(j\omega) = \angle -90^0 - \tan^{-1}\omega T_1 - \tan^{-1}\omega T_2 = -90^0$$

$$\lim_{\omega\to \infty} \angle G(j\omega) = \angle -90^0 - \tan^{-1}\omega T_1 - \tan^{-1}\omega T_2 = -270^0$$

Step 4: Separate the real and Im part of G(jω)

$$G(j\omega) = \frac{-\omega K(T_1+T_2)}{\omega+\omega^3(T_1^2+T_2^2+\omega^2T_1^2T_2^2)} + j\frac{j(K\omega^2T_1T_2-K)}{\omega+\omega^3(T_1^2+T_2^2+\omega^2T_1^2T_2^2)}$$

Step 5: Put Re [G(jω)]=0

$$\frac{-\omega K(T_1+T_2)}{\omega+\omega^3(T_1^2+T_2^2+\omega^2T_1^2T_2^2)} = 0 \Rightarrow \omega = \infty$$

So at

$$\omega = \infty \qquad \Rightarrow G(j\omega) = 0\angle -270^0$$

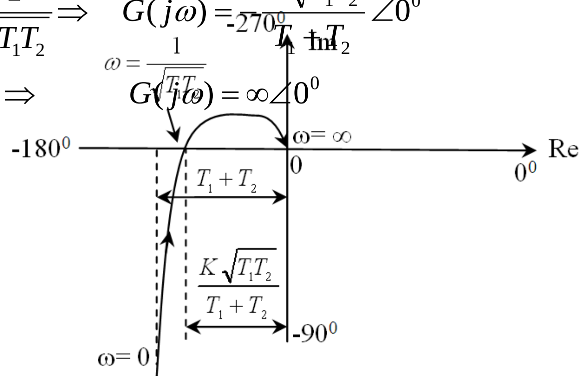
Step 6: Put Im [G(jω)]=0

$$\frac{j(K\omega^2T_1T_2-K)}{\omega+\omega^3(T_1^2+T_2^2+\omega^2T_1^2T_2^2)} = 0 \Rightarrow \omega = \frac{1}{\sqrt{T_1T_2}} \;\; \& \; \omega = \pm\infty$$

So When

$$\omega = \frac{1}{\sqrt{T_1T_2}} \Rightarrow \quad G(j\omega) = \frac{K\sqrt{T_1T_2}}{T_1+T_2}\angle 0^0$$

$$\omega = \infty \Rightarrow \qquad G(j\omega) = \infty\angle 0^0$$

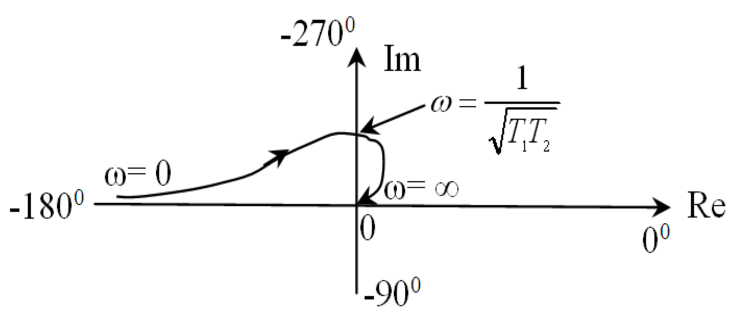


Polar plot for type -1 system

Polar plot for type 2

Let

$$G(s) = \frac{K}{s^2(1+sT_1)(1+sT_2)}$$



Polar plot for type -2 system

Determination of Gain Margin and Phase Margin From Polar Plot

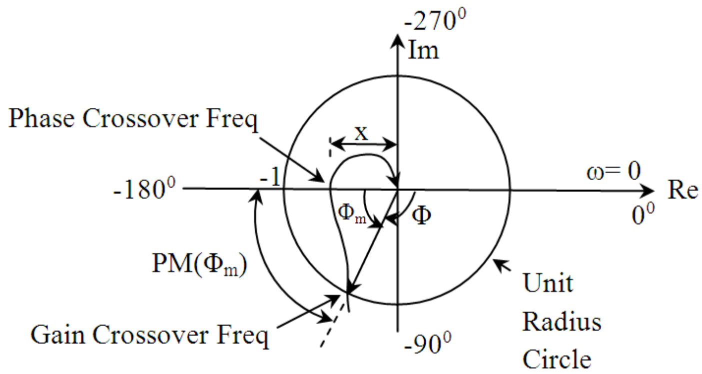
The **gain margin** is defined as there inverse of the magnitude of G(jw) at phase crossover frequency .The Phase crossover frequency is the frequency at which the phase of G(jw) is 180^0

Phase Crossover Frequency (ω_p) : The frequency where a polar plot intersects the negative real axis is called phase crossover frequency.

Gain Crossover Frequency (ω_g) : The frequency where a polar plot intersects the unit circle is called gain crossover frequency. so W_g

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

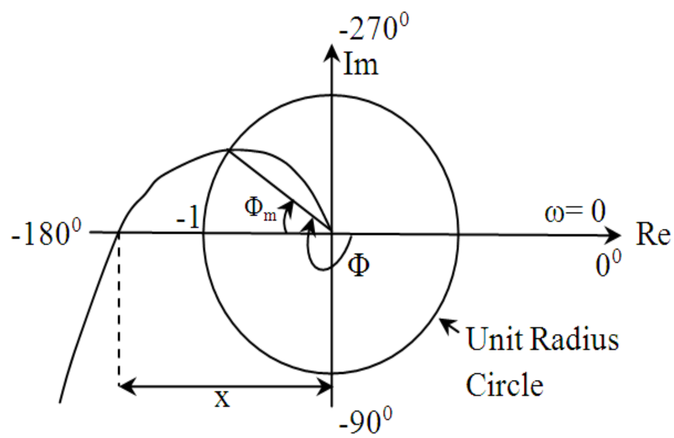
Let the polar plot cut the 180^0 at an axis at a point x and the magnitude circle passing through the point G_x.Now the gain margin $K_g=1/G_x$. if the point x lies within unity circle then the gain margin is positive otherwise negative. (If the polar plot is drawn in ordinary graph sheet using rectangular coordinates then the point x is the cutting point of G(jw) locus with negative real axis and $K_g=1/|G_x|$ is the magnitude corresponding to point x.



Stable system: +GM, +PM

Polar plot showing positive gain margin and phase margin

The phase Margin is defined as ,Phase Margin $=180^0 + \Phi$ where Φ is the phase angle of G(jw) at gain cross over frequency.



Unstable system: -GM, -PM

Polar plot showing Negative gain margin and phase margin

Gain Margin $K_g = 1/G_x$

Phase Margin = $180^0 + \Phi$

Nyquist Stability Criterion:

A frequency domain technique is developed in this chapter, which gives a simple way of determining the absolute stability of the system, and also defines and determines the relative stability of a system.

This frequency domain criterion is known as Nyquist Stability Criterion. This method relates the location of the closed loop poles of the system with the frequency response of the open loop system. It is a graphical technique and does not require the exact determination of the closed loop poles. Open loop frequency response can be obtained by subjecting the system to a sinusoidal input of constant amplitude and variable frequency and measuring the amplitude and phase angle of the output. The development of Nyquist

Criterion is based on a theorem due to Cauchy, 'the principle of argument' in complex variable theory.

Principle of Argument:

Consider a function of complex variable's', denoted by *F(s)*, which can be described as a quotient of
two polynomials. Assuming that the two polynomials can be factored, we have

$$F(s) = \frac{(s+Z_1)(s+Z_2).....(s+Z_m)}{(s+p_1)(s+p_2).....(s+p_n)} \qquad .... (1)$$

Since $s = a + jw$ is a complex variable, for any given value of *s*, *F(s)* is also complex and can be represented by $F(s) = u + jv$. For every point *s* in the s-plane at which *F(s)* and all its derivatives exist, i.e., for points at which *F(s)* is analytic, there is a corresponding point in the *F(s)* plane. It means that *F(s)* in eqn. (1) maps points in s plane at which *F(s)* is analytic into points in *F(s)* plane. In eqn. (1) $s = - P_1 = - P_2$ are the poles of the function *F(s)* and therefore the function goes to infinity at these points. These points are also called singular points of the function *F(s)*.

Now, consider a contour τ_s in s-plane as shown in Fig (a). Assume that this contour does not pass through any singular points of *F(s)*. Therefore, for every point on this contour, we can find a corresponding point in *F(s)* plane, or corresponding to the contour τ_s in s-plane there is a contour τ_f in *F(s)* plane as shown in Fig.

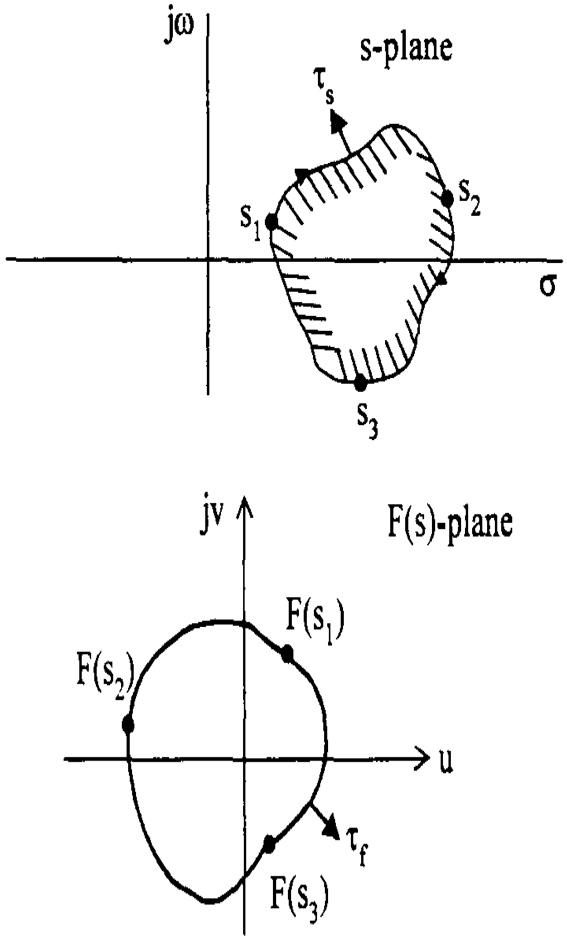


Fig 1.1(a)Arbitrary contour τ_s in s-plane not passing through singular points of F(s)

Fig 1.1(b)Corresponding F(s)-plane contour τ_f

Let us consider a closed contour and define the region to the right of the contour, when it is traversed in clockwise direction, to be enclosed by it. Thus, the shaded region in Fig. 1.1(a) is considered to be enclosed by the closed contour τ_s . Let us investigate some of the properties of the mapping of this contour on to F(s)-plane when τ_s encloses (a) a zero of *F(s)* (b) a pole of *F(s).*

Case a: **When $\tau_s$encloses a zero of *F(s):***

Let $s = -Z_j$ be encloses by the contour $\tau_s$as shown in Fig. (1.2) (a).

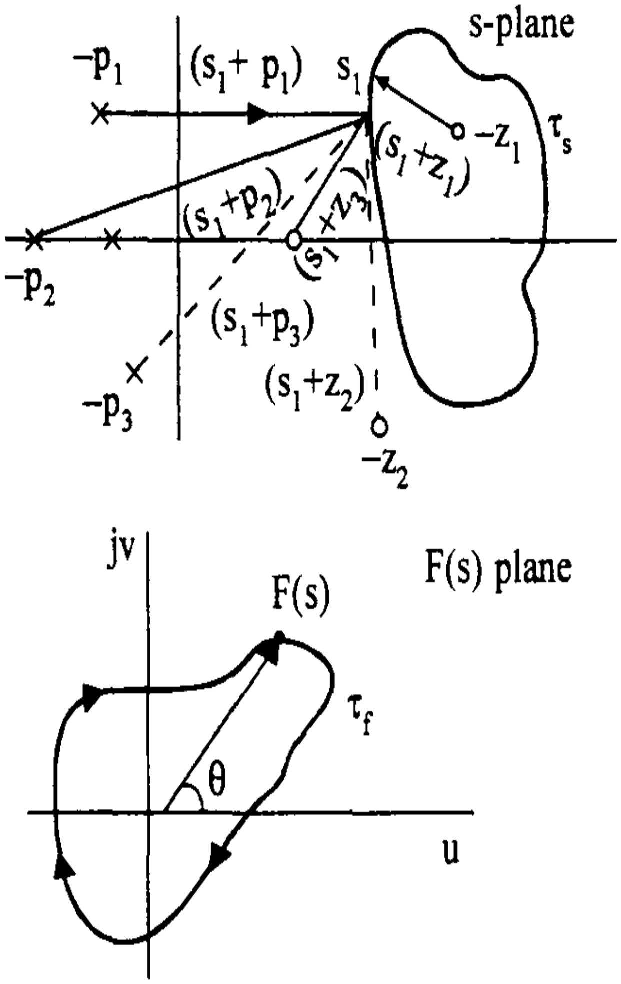


Fig. 1.2(a) Contour τ_s encloses one zero s = Z_1 of F(s)

Fig. 1.2 (b) Corresponding F(s) plane contour τ_f

For any point s = s_1 we have

$$F(s_1) = \frac{(s_1 + z_1)(s_1 + z_2).....(s_1 + z_m)}{(s_1 + p_1)(s_1 + p_2).....(s_1 + p_n)}$$

$$= \frac{\alpha_1\,\alpha_2.....\alpha_m}{\beta_1\,\beta_2.....\beta_m} \underline{/\theta_1 + \theta_2 + ...\theta_m - \phi_1 - \phi_2....\phi_n} \quad .....(1.2)$$

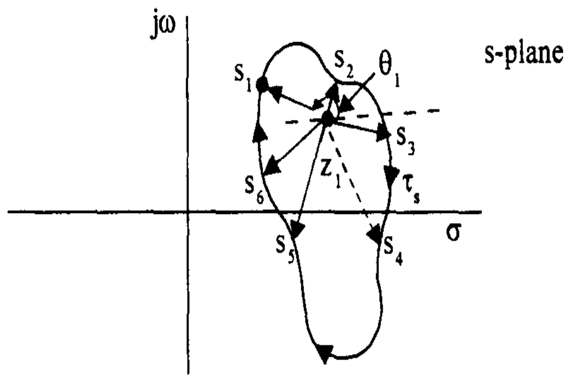
Where

$$\begin{array}{lll}
\alpha_1 = |s_1 + z_1|, & \alpha_2 = |s_1 + z_2|, \; ..... & \alpha_m = |s_1 + z_m| \\
\beta_1 = |s_1 + p_1|, & \beta_2 = |s_1 + p_2|, \; ..... & \beta_n = |s_1 + p_n| \\
\theta_1 = \underline{/(s_1 + z_1)}, & \theta_2 = \underline{/(s_1 + z_2)}, \; ..... & \theta_m = \underline{/(s_1 + z_m)} \\
\phi_1 = \underline{/(s_1 + p_1)}, & \phi_2 = \underline{/(s_1 + p_2)}, \; ..... & \phi_n = \underline{/(s_1 + p_n)}
\end{array}$$

In the development of Nyquist criterion, the magnitude of *F(s)* is not important, as we will see
later. Let us concentrate on the angle of *F(s).*

In Fig. 1.2(a), as the point *s* moves on the contour $\tau_s$in clockwise direction, and returns to the starting point, let us compute the total angle described by *F(s)* vector as shown in Fig. 1.2(b). The vector *(s + z_1)* contributes a total angle of -2πt to the angle of

F(s) as shown in Fig. 1.3(a) since the vector *(s + z_1)* makes one complete rotation. This is because the point $s = -z_1$ lies inside the contour.



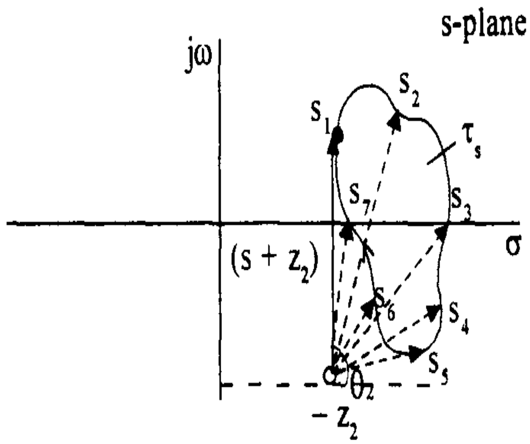


Fig. 1.3 (a) Angle contributed by (s + Z_1) to F(s) Fig. 1.3 (b) Angle contributed by (s + Z_2) to F(s)

The vector *(s + Z* $_2$) contributes zero net angle for one complete traversal of the point *s*on the contour $\tau_s$in s-plane as shown in Fig. 1 .3 (b). This is because the point $s = -z_2$lies outside the contour τ_s.

Similarly, all the zeros and poles which are not enclosed by the contour $\tau_s$contribute net zero angles to *F(s)* for one complete traversal of a point s on the contour τ_sThus the total angle contributed by all the poles and zeros of *F(s)* is equal to the angle contributed by the zero s = - z_1 which is enclosed by the contour. The *F(s)* vector describes an angle of - $2\pi$and therefore the tip of *F(s)* vector describes a closed contour about the origin of *F(s)* plane in the clockwise direction. Similarly, if *k* zeros are enclosed by the s-plane contour, the *F(s)* contour will encircle the origin *k* times in the clockwise direction. Two cases for $k = 2$ and $k = 0$ are shown in the Fig. 7.4 (a) and (b).

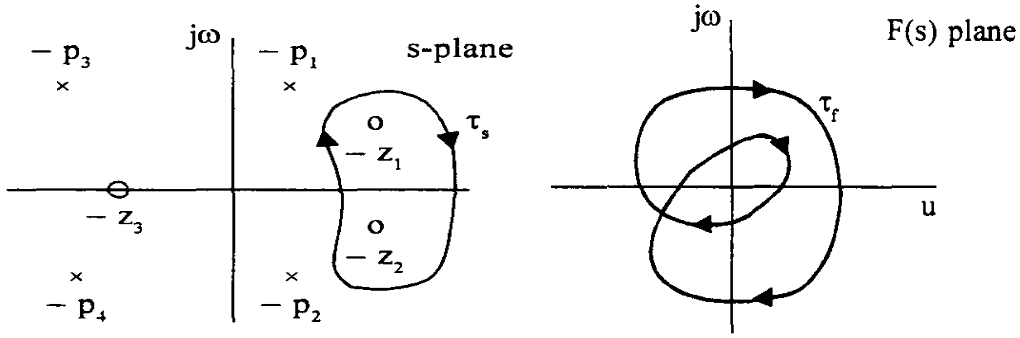


Fig. 1.4 (a) s-plane contour and F(s) plane contour for k = 2.

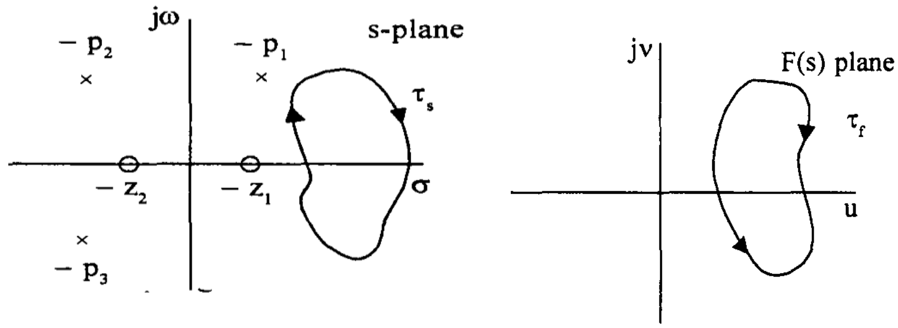


Fig. 1.4 (b) s-plane contour and F(s) plane contour for k = 0.

***Case (b)*: Whenτ_s encloses a pole of F(s):**

When a pole of *F(s)*, s = -$P_1$is enclosed by the contour τ_sa net angle of 2π is contributed by the vector (s + P_1) to *F(s)* as the factor (s + P_1) is in the denominator of *F(s).* Thus, the *F(s)* plane contour will encircle the origin once in the anticlockwise direction.

If s-plane contour $\tau_s$encloses P poles and Z zeros, the *F(s)* plane contour encircles the origin P times in the anticlockwise direction and Z times in the clockwise direction. In other words, it encircles the origin of *F(s)* plane (P - Z) times in the anticlockwise direction. The magnitude and hence the actual shape of the *F(s)* plane contour is not important, but the number of times the contour encircles the origin is important in the development of Nyquist stability criterion, as will be discussed in the next section. This relation between the number of poles and zeros enclosed by the closed s-plane contour $\tau_s$and the number of encirclements of *F(s)* plane contour $\tau_f$is known as the principle of Argument.

Nyquist Criterion:

For a feedback control system with loop transfer function given by :

$$G(s)\,H(s) = \frac{K(s+z_1)(s+z_2).....(s+z_m)}{(s+p_1)(s+p_2).....(s+p_n)} \quad m \le n$$

$$= K\,\frac{\prod_{i=1}^{m}(s+z_i)}{\prod_{j=1}^{n}(s+p_j)} \qquad ......(1.3)$$

The characteristic equation is given by

$$D(s) = 1 + G(s)\,H(s) = 1 + \frac{K \prod_{i=1}^{m}(s+z_i)}{\prod_{j=1}^{n}(s+p_j)} = 0$$

$$= \frac{\prod_{j=1}^{n}(s+p_j) + K\prod_{i=1}^{m}(s+z_i)}{\prod_{j=1}^{n}(s+p_j)} = 0 \qquad \text{....(1.4)}$$

The numerator of eqn. (1.4) is a polynomial of degree *n* and hence it can be factored and written as

$$D(s) = \frac{(s+z_1')(s+z_2').....(s+z_n')}{\prod_{j=1}^{n}(s+p_j)} = 0 \qquad \text{.....(1.5)}$$

Thus, it can be observed that:

1. The poles of the open loop system G(s) H(s) and poles of D(s) are the same (eqns. (1.3) and (1.5).
2. The roots of the characteristic equation D(s) = 0 are the zeros of D(s) given by $-z_1$, $-z_2$, ...$-z_n$in eqn. (1.5).
3. The closed loop system will be stable if all the poles of the closed loop system, i.e., all the roots of the characteristic equation lie in the left half of s-plane. In other words, no pole of the closed loop system should be in the right half of s-plane.
4. From eqn. (1.4) and (1.5), it is clear that even if some poles of open loop transfer function $(-P_1\ -P_2...\ -P_n)$ lie in right half of *s* plane, the closed loop poles, or the zeros of D(s) = 0 i.e. $s = -z_1, -z_2$etc. many all lie in the left half of s-p lane. Thus, even if the open loop system is unstable, the closed loop system may be stable.

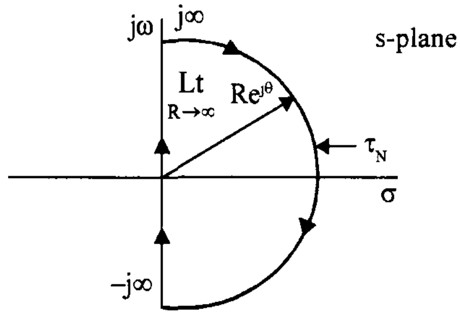
In order to determine the stability of a closed loop system.we have to find if any of the zeros of characteristic equation D(s) = 0 in eqn. (1.5) lie in the right half of s-plane. If we consider an s-plane contour enclosing the entire right half of s-plane, plot the D(s) contour and find the number of encirclements of the origin, we can find the number of poles and zeros of D(s) in the right half of s-plane.

Since the poles of D(s) are the same as open loop poles, the number of right half plane poles are known. Thus, we can find the number of zeros of D(s) i.e. the number of closed loop poles in the right half of s-plane. If this number is zero. Then the closed loop system is stable, otherwise the system is unstable.

Development of Nyquist Criterion:

Nyquist Contour:

Let us consider a closed contour, $\tau_N$which encloses the entire right half of s-plane as shown in FigThis contour is known as a Nyquist Contour. It consists of the entirejw-axis and a semicircle of infinite radius.



Nyquist Contour

On the jw-axis,

$$s = jw \text{ and } w \text{ varies from } -\infty \text{ to } +\infty$$

On the infinite semicircle,

$$s = \lim_{R\to\infty} Re^{j\theta} \ , \theta \text{ varies from } -\frac{\pi}{2} \text{ to } 0 \text{ to } +\frac{\pi}{2}$$

Thus the Nyquist Contour encloses the entire right half of s-plane and is traversed in the clockwise direction.

Nyquist Stability Criterion:

if

$$D(s) = \frac{\prod_{i=1}^{n}(s+z_i)}{\prod_{j=1}^{m}(s+p_j)} \quad .....(1.6)$$

is plotted for values of *s* on the Nyquist contour, the D(s) plane contour will encircle the origin N times in the counter clockwise direction, where

$$N=P-Z$$

and

P = number of poles of D(s) or the number of open loop poles in the right half of s-plane (R.H.S)

Z= Number of zeros of D(s) or the number of closed loop poles in the RHS.

If the closed loop system is stable,

Z = 0

Thus, for a stable closed loop system,

N = P

i.e., the number of counter clockwise encirclements of origin by the D(s) contour must be equal to the number of open loop poles in the right half of s-plane. Further, if the open loop system is stable, there are no poles of G(s) H(s) in the RHS and hence,

p = 0

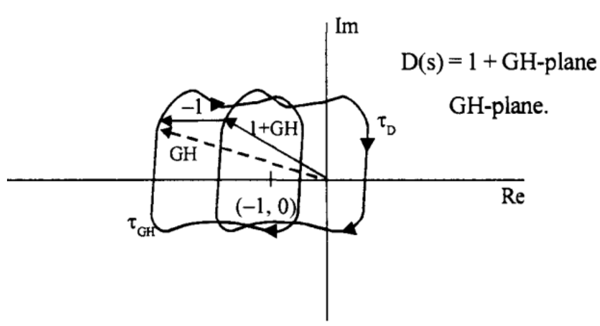
For stable closed loop system,

N = 0

i.e., the number of encirclements of the origin by the D(s) contour must be zero.

Also, observe that G(s) H(s) = [1 + G(s) H(s)] – 1

Thus G(s) H(s) contour and D(s) = 1 + G(s) H(s) differ by 1. If 1 is subtracted from D(s) = 1 + G(s) H (s) for every value of s on the Nyquist Contour, G(s) H(s) contour will be obtained and the origin of D(s) plane corresponds to the point (- 1, 0) of G(s) H(s) plane, this is shown graphically in Fig.



D(s) = 1 + G,H plane and GH plane Contours

If G(s) H(s) is plotted instead of 1 + G(s) H(s), the G(s) H(s) plane contour corresponding to the Nyquist Contour should encircle the (-1, j0) point P time in the counter clockwise direction, where P is the number of open loop poles in the RHS. The Nyquist Criterion for stability can now be stated as follows:

If the τ_{GH} Contour of the open loop transfer function G(s) H(s) corresponding to the Nyquist Contour in the s-plane encircles the (-1, j0) point in the counter clockwise direction, as many times as the number of poles of G(s) H(s) in the right half of s-p1ane, the closed loop system is stable. In the more common special case, where the open loop system is also stable, the number of these encirclements must be zero.

Nyquist Contour When Open Loop Poles Occur on jw-axis:

If G(s) H(s) has poles on the jw-axis, 1 + G(s) H(s) also has these poles on the jw-axis. As the NyquistContour passes through these jw-axis poles, this Contour is not suitable for the study of stability. No singulasitics of 1 + G(s) H(s) should lie on the s-plane Contour τ_s . In such cases a small semicircle is taken around these poles on the jw-axis towards the RHS so that these poles are bypassed. This is shown in Fig..

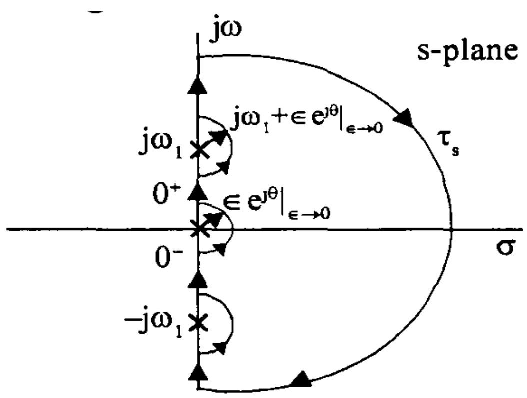


Fig. 1.7 Nyquist Contour when jro-axis poles are present

If an open loop pole is at $s = \pm j\omega_1$, near this point, s is taken to vary as given be eqn. (1.10).

s = $j\omega_1 + \lim_{\epsilon \to \infty} \epsilon e^{j\theta}$, θ *varies from* $-\frac{\pi}{2}$ to 0 to $+\frac{\pi}{2}$ ..... (1.10)

This describes the semicircle around the pole $s = j\omega_1$ in the anti clockwise direction. Let us now consider some examples illustrating how the Nyquist plots are constructed and the stability deduced.

State space analysis of continuous systems

Objectives:

- To familiarize with fundamental ideas for the analysis and design of control systems using state variable approach..
- Provides the detail discussion of for analysis of stability of control systems and testing of system either controllable or observable.

Learning Outcomes:

Students will be able to

a) Analyze the state model representation of the time invariant linear systems and solving the time invariant state equations

b) Observe the controllability and obervabilty of the system

Introduction:

The methods of analysis and the design of feedback control systems; such as root locus, and Bode and Nyquist plots, require the physical system to be modeled in the form of a transfer function. Although the transfer function approach of analysis is a simple and powerful technique, it suffers from certain drawbacks:

1. The transfer function model is only applicable to linear time-invariant systems.
2. A transfer function is only defined under zero initial conditions; meaning that the system is initially **at rest** and the time solution obtained is, in a general form, due to input only. However, for multi-input-multi-output, the systems are initially not at rest. Hence, as a conclusion, the transfer function model is generally restricted to single-input-single-output (SISO) systems.
3. A transfer function technique only reveals the system output for a given input and does not provide any information regarding the **internal state** of the system.

Thus, the classical design methods (root locus and frequency domain methods) based on the transfer function approach are **inadequate** and not convenient.

The limitations listed above made us feel the need for a more general mathematical representation of a control system, which, along with the output, yields information about the state of the system-variables at some predetermined points along the flow of signals. This leads to the development of the **state-variable approach**, which has the following advantages over the **classical approach:**

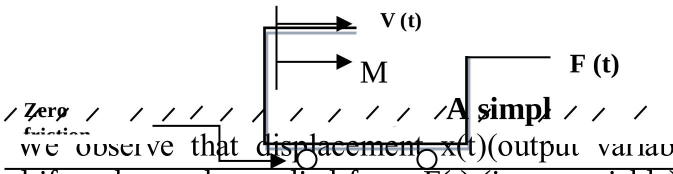
1. It is a direct **time-domain approach** which provides a basis for modern control theory and system optimization.
2. It is a very power full technique for the analysis and design of linear and non linear, time – invariant or time-varying multi-input-multi output systems.

3. The organization of the state variable approach is such that it is easily amenable to solution through digital computers.

Concept of state and state variables:

A mathematical abstraction to represent or model the dynamics of systems utilizes three types of variables called the input, the output and state variables.

Consider the mechanical system shown in figure 1, where in mass M is acted upon by the force F(t).



A simpl ical systems

We observe that displacement x(t)(output variable) at any time t≥t_0 can be computed if we know the applied force F(t) (input variable) from t=t_0 onwards, provided v(t_0) the initial velocity and x(t_0) the initial displacement are known. We may conceive of initial velocity and initial displacement as describing the status or state of the system at t=t_0. The state of the systems of **Fig: 1** at any time t is given variables x(t) and v(t) which are called the state variables of the system.

- **State:** The state of a dynamic system is the smallest set of variables and the knowledge of these variables at $t = t_0$ together with inputs for $t \geq t_0$ completely determines the behavior of the system at $t \geq t_0$. A compact and concise representation of the past history of the system can be termed as the state of the system.
- **State Variables:** The smallest set of variables that determine the state of the system are known as state variables.

The knowledge of capacitor voltage at t = 0 i.e., the initial voltage of the capacitor is a history dependent term and it forms a state variable. Similarly, initial current in an inductor is treated as state variable.

In state variable formulation of the system, the state variables are usually represented by x_1(t), x_2(t),……..,the inputs by u_1(t),u_2(t),………:and the outputs by the y_1(t),y_2(t),….the states space representation may be visualized in block diagram form as shown Fig:2.

For generality, we have depicted a system which has m inputs outputs and n state variable. For notational economy, the different variables may be represented by the input vector u (t), output vector y(t) and the state x(t); where

$$u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ \vdots \\ u_m(t) \end{bmatrix}_{m\times 1} \quad x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ \vdots \\ x_n(t) \end{bmatrix}_{n\times 1} \quad y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ \vdots \\ y_p(t) \end{bmatrix}_{p\times 1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

In fig:2 arrows have been used to represent vector quantities.

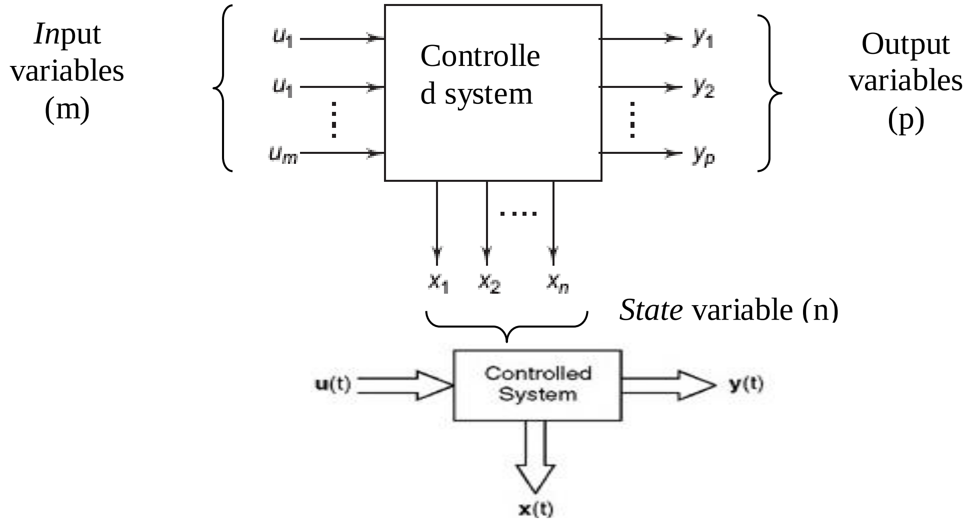


Fig: 2 Structure of a general control system

State Model of Linear Time Invariant systems or multi-input-multi output system:

The state and output equations constitute the state model of the system.

In state representation, the derivative of each state variable can be written as a linear combination of system states and inputs i.e.,

$$\left.\begin{aligned}
\dot{x}_1(t) &= a_{11}x_1(t) + a_{12}x_2(t) + \ldots\ldots + a_{1n}x_n(t) + b_{11}u_1(t) + b_{12}u_2(t) + \ldots\ldots + b_{1m}u_m(t) \\
\dot{x}_2(t) &= a_{21}x_1(t) + a_{22}x_2(t) + \ldots\ldots + a_{2n}x_n(t) + b_{21}u_1(t) + b_{22}u_2(t) + \ldots\ldots + b_{2m}u_m(t) \\
\vdots \quad & \quad \vdots \qquad\qquad \vdots \\
\dot{x}_n(t) &= a_{n1}x_1(t) + a_{n2}x_2(t) + \ldots\ldots + a_{nn}x_n(t) + b_{n1}u_1(t) + b_{n2}u_2(t) + \ldots\ldots + b_{1m}u_m(t)
\end{aligned}\right\} \quad \ldots(2$$

where coefficients $\boldsymbol{a}_{ij}$; $\boldsymbol{i = 1, 2, \ldots n}$; $\boldsymbol{j = 1, 2, \ldots n}$ and $\boldsymbol{b}_{ik}$; $\boldsymbol{i = l, 2, \ldots. n}$; $\boldsymbol{k = 1, 2, , m}$ are constants. Equations (2) may be written in vector-matrix form as

$$\dot{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)) \quad \textit{State Equation} \qquad \ldots\ldots\ldots\ldots (3)$$

$$where \quad \boldsymbol{x(t)} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}_{n\times l} \quad ; \text{ State Vector}$$

$$\boldsymbol{u(t)} = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{bmatrix}_{m\times l} \quad ; \text{ Input Vector}$$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ a_{n1} & a_{n2} & \cdots\cdots & a_{nn} \end{bmatrix}_{n\times n} \quad ; \text{ System Matrix}$$

$$and \quad \boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots\cdots & b_{1m} \\ b_{21} & b_{22} & \cdots\cdots & b_{2m} \\ & & \cdots\cdots & \\ b_{n1} & b_{n2} & \cdots\cdots & b_{nm} \end{bmatrix}_{n\times m} \quad ; \text{ Input Matrix}$$

Similarly, the output variables can be written as a linear combination of system states and inputs, i.e.

$$\left.\begin{aligned} y_1(t) &= c_{11}x_1(t) + c_{12}x_2(t) + \cdots\cdots + c_{1n}x_n(t) + d_{11}u_1(t) + d_{12}u_2(t) + \cdots\cdots d_{1m}u_m(t) \\ y_2(t) &= c_{21}x_1(t) + c_{22}x_2(t) + \cdots\cdots + c_{2n}x_n(t) + d_{21}u_1(t) + d_{22}u_2(t) + \cdots\cdots d_{2m}u_m(t) \\ y_p(t) &= c_{p1}x_1(t) + c_{p2}x_2(t) + \cdots\cdots + c_{pn}x_n(t) + d_{p1}u_1(t) + d_{p2}u_2(t) + \cdots\cdots d_{pm}u_m(t) \end{aligned}\right\} \quad \ldots\ldots\ldots$$

Where, coefficients c_{ij}; $i = 1, 2\ldots.. p$; $j = 1, 2,\ldots. n$ and d_{ik}; $i = 1, 2,\ldots. p$: $k = 1, 2,\ldots.. m$ are constants. Equations (4) may be written in vector-matrix form as

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)) \quad \text{Output Equation} \qquad \ldots\ldots\ldots\ldots$$

$$\text{where} \quad \boldsymbol{y(t)} = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_p(t) \end{bmatrix}_{p\times l} \quad ; \; Output\,Vector$$

$$\boldsymbol{u(t)} = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{bmatrix}_{m\times l} \quad ; \; Input\,Vector$$

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots\cdots & c_{1n} \\ c_{21} & c_{22} & \cdots\cdots & c_{2n} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ c_{p1} & c_{p2} & \cdots\cdots & c_{pn} \end{bmatrix}_{p\times n} \quad ; \; Output\,Matrix$$

$$\text{and} \quad \boldsymbol{D} = \begin{bmatrix} d_{11} & d_{12} & \cdots\cdots & d_{1m} \\ d_{21} & d_{22} & \cdots\cdots & d_{2m} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ d_{p1} & d_{p2} & \cdots\cdots & d_{pm} \end{bmatrix}_{p\times m} \quad ; \; Transmission\,Matrix$$

The state equation and output equation together constitute the state model of the system. Thus, the state model of a linear time-invariant MIMO system is given as

$$\dot{\boldsymbol{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)) \quad State\,Equation$$
$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)) \quad Output\,Equation$$

The state model of a linear, time-varying MIMO system is of the same form as given in above Equations except for the fact that the coefficients of the matrices **A, B, C**, and **D** are no longer constants but are the functions of time.

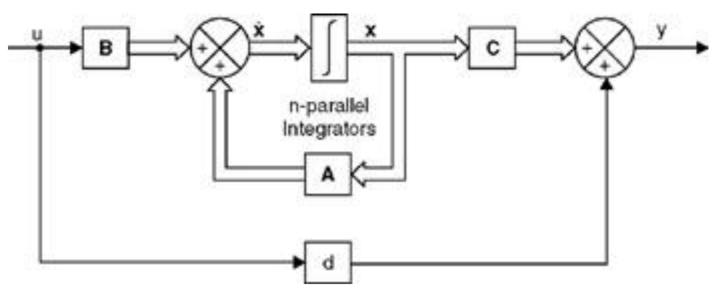


Fig: 3. Block diagram representation of the state model of a linear multi-input-multi output systems

State Model of Linear single-input-single output systems:

The transfer function analysis deals mainly with single- input-single output linear time invariant systems. Here we represent the transfer function in state variable form. If we let m =1 and p=1 in the state model of a multi-input output linear system. we obtain the following state model for a single-input-single output linear systems.

$$(1.12A)\dot{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)) \quad State\,Equation$$

$$(1.12B)\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + du(t)) \quad Output\,Equation$$

$$where \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}_{n\times l} \quad ; State\,Vector \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ a_{n1} & a_{n2} & \cdots\cdots & a_{nn} \end{bmatrix}_{n\times n} \quad ; System\,Matrix$$

$$\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}_{n\times 1} \quad ; \quad Input\,Matrix\,\mathbf{C} = [c_1\, c_2\, c_3 \ldots\ldots c_n]_{1\times n}; \quad Output\,Matrix$$

d = Transmission Constant

u(t) = Input or Control Variable (scalar)

and, *y(t)* = Output Variable (scalar).

The block-diagram representation of the state model of linear single-input-single-output system is shown in Figure 1.3.

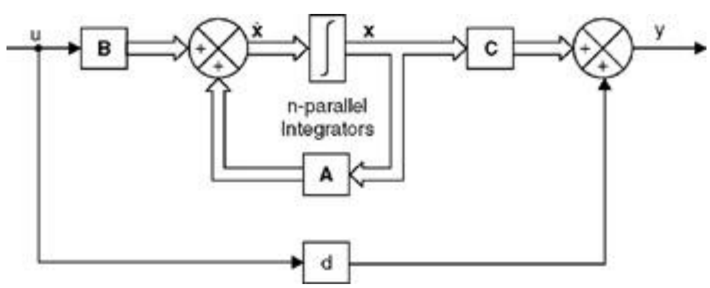


Figure 4: Block-diagram representation of the state model of the linear single-input-single-output system

State models for linear continuous-time systems:

Non uniqueness of state model:

The state equations of a system are not unique i.e., there exist more than one set of state variables in terms of which the system behavior can be completely described. Alternate representations are possible for any given system. Let us consider a state model of a system as given below.

$$\dot{x} = \mathbf{A}x + \mathbf{B}u \qquad \text{……………………..} (1)$$

$$y = \mathbf{C}x + \mathbf{D}u \qquad \text{…………….} (2)$$

Where in **x** is an n-dimensional state vector.

Consider now another n-dimensional state vector **z** such that

$$x = \mathbf{M}\, z \qquad \text{………………} (3)$$

Since **M** is a constant matrix, it follows that

$$\dot{x} = \mathbf{M}\, \dot{z} \qquad \text{……………} (4)$$

Substituting x and $\dot{x}$ in above equation (1)

$$\mathbf{M}\dot{z} = \mathbf{AM}z + \mathbf{B}u$$

Pre multiplying by $\boldsymbol{M^{-1}}$, we obtain

$$\dot{z} = \mathbf{M}^{-1}\mathbf{AM}z + \mathbf{M}^{-1}\mathbf{B}u = \mathbf{P}z + \mathbf{Q}u \qquad \text{………………} (5)$$

From equations (3)& (2)

$$y = \mathbf{CM}z + \mathbf{D}u = \mathbf{R}z + \mathbf{D}u \text{…………..} (6)$$

Where $\mathbf{P} = \mathbf{M}^{-1}\mathbf{AM}$

$\mathbf{Q} = \mathbf{M}^{-1}\mathbf{B}$

$\mathbf{R} = \mathbf{CM}$

And $\mathbf{D} = \mathbf{D}$

Equations (5) & (6) give another state model for a given system. Since **M** is assumed to be nonunique nonsingular matrix, the state model is also nonunique. It is important to note that the transformation matrix **M** must be non singular, i.e., $|\boldsymbol{M}| \neq 0$.If this were not the case, the inverse transformation would obviously not exist.

State-space representation using Phase variables in Conical Form (CF):

For any given system, there are essentially an infinite number of possible state space models that will give the identical input/output dynamics. Thus, it is desirable to have certain standardized state space model structures: these are the so-called **canonical forms**. Given a system transfer function, it is possible to obtain each of the canonical models. And, given any particular canonical form it is possible to transform it to another form. The phase variables state model is easily determined if the system model is already known in the differential equation/transfer function form.

The general form of an nth-order linear differential equation relating the output y(t) to the input u(t) of a linear continuous-time system is

$$Y^{(n)} + a_1\, y^{(n-1)} + \ldots + a_{n-1}\dot{y} + a_n y = b_0\, u^{(m)} + b_1\, u^{(m-1)} + \ldots + b_{m-1}\dot{u} + b_m u \qquad \text{………}(a)$$

Where for time-invariant system a_i 's and b_j's are constants, m and n are integers with $m \leq n$ and $y^{(n)}$

Where **u** is the input, **y** is the output and $\mathbf{y}^{(n)}$ represents the $\mathbf{n^{th}}$ derivative of **y** with respect to time. Taking the Laplace transform of both sides we get:

$$Y(s)(s^n + a_1 s^{n-1} + \dots + a_{n-1}s + a_n) = U(s)(b_0 s^m + b_1 s^{m-1} + \dots + b_{m-1}s + b_m),$$

This yields the transfer function as

$$\frac{Y(s)}{U(s)} = \frac{b0sm + b1sm-1 + \dots + bm-1s + bm}{s\,n + a1sn-1 + \dots + an-1s + an} \quad \dots\dots\dots\dots\dots\dots \text{(b)} \quad m \le n$$

Given the system having transfer function as defined in (b) above, we will define the controllable canonical and observable canonical forms. We shall obtain a state model from the transfer function with zero initial conditions and then relax the zero initial conditions to arbitrary initial conditions.

The phase variables are defined as those particular state variables which are obtained from one of the system variables and its derivatives. Often the variable used is the system output and the remaining state variables are then derivatives of the output.

Let us first consider a simple case, where the transfer function does not have zeros. Such a transfer function has the form

$$\frac{Y(s)}{U(s)} = \frac{b}{s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_{n-1}s + a_n} \quad \dots\dots\dots\dots\dots \text{(c)}$$

The transfer function in given Eq.(c) can also be represented by the differential equation as follows:

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1}y}{dt^{n-1}} + a_2 \frac{d^{n-2}y}{dt^{n-2}} + \cdots + a_{n-1}\frac{dy}{dt} + a_n y = bu \quad \dots\dots\dots\dots\dots\dots\text{(d)}$$

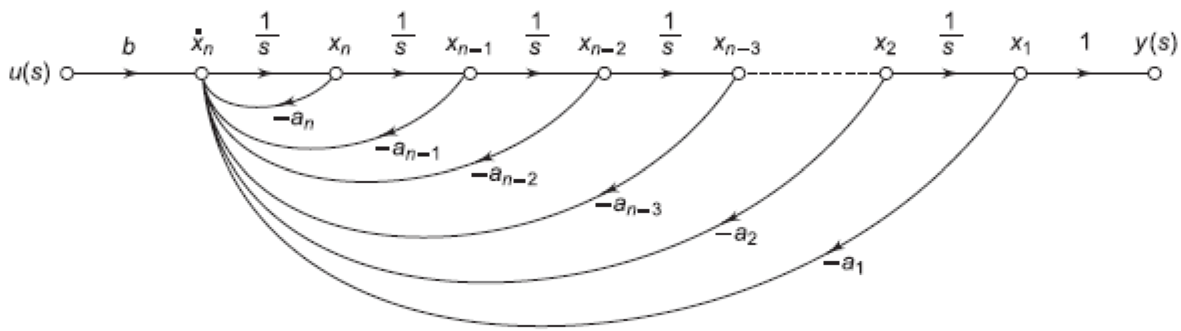
by letting

$$\begin{aligned} x_1 &= y \\ x_2 &= \dot{y} \\ &\dots\dots\dots\dots \\ x_n &= y^{(n-1)} \dots \quad \dots\dots\dots\dots \quad \text{(e)} \end{aligned}$$

Equation (d) is reduced to a set of n- first –order differential equations given bellow:

$$\begin{aligned} x_1 &= y \\ \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= -a_m x_1 - a_{n-1}x_2 - a_{n-2}x_3 \dots - a_1 x_1 + bu \dots\dots\dots\dots\text{(f)} \end{aligned}$$

signal flow diagram of above equation.



the above state equations can be written in vector-matrix form as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots\ldots 0 \\ 0 & 0 & 1 & \ldots\ldots 0 \\ . & . & . & \ldots\ldots 0 \\ . & . & . & \ldots\ldots 0 \\ . & . & . & \ldots\ldots 0 \\ 0 & . & . & \ldots\ldots 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \ldots -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \\ b \end{bmatrix} [u] \qquad \ldots\ldots\ldots\ldots(1)$$

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

The matrix **A** has all ones in the upper off-diagonal and other elements except the last row. The co-efficient of the last row has negative co-efficient. This form is known as the Bush or Comparison form. The other name is phase variable Controllable Comparison form (CCF).

The output being $y=x_1$ the output equation is given by

$$\mathbf{y = C\ x \ldots\ldots\ldots\ldots(2)}$$

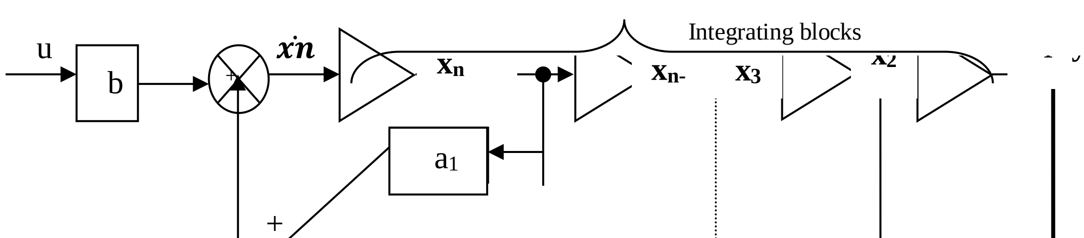
Where C=[1 0 …..0]………….

The matrix for output equation of the system is given below:

$$y = [1\ 0\ 0\ \ldots\ldots 0] \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_{n-1} \\ x_n \end{bmatrix} + [0]\ [u]$$

$$\therefore \qquad y = Cx + Du$$

The block diagram representation of the state model of equation (1) is drawn as



Feedback blocks

Phase Variable Formulations for Transfer function:

Let the transform of the system be

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^3 + b_1 s^2 + b_2 s + b_3}{s^3 + a_1 s^2 + a_2 s + a_3} = \frac{(b_0 + b_1 s^{-1} + b_2 s^{-2} + b_3 s^{-3})\, X(s)}{(1 + a_1 s^{-1} + a_2 s^{-2} + a_3 s^{-3})\, X(s)} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

Let $\quad Y(s) = (b_0 + b_1 s^{-1} + b_2 s^{-2} + b_3 s^{-3})X(s) \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$

and $\quad U(s) = (1 + a_1 s^{-1} + a_2 s^{-2} + a_3 s^{-3})X(s) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$

From Eq.3 it can be written that,

$X(s) = U(s) - (a_1 s^{-1} + a^2 s^{-2} + a_3 s^{-3})X(s) \quad \ldots\ldots.(4)$

The transfer functions and signal flow graph are related by masons gain formula, reproduced below.

$T(s) = \frac{1}{\Delta} \sum_k \quad \Delta_k P_k$

Where P_k= Path gain of the k^{th} forward path; Δ= 1-(sum of loop gains of all individual loops)+(sum of gain products of all possible combinations of two non-touching loops)-(sum of gain products of all possible combinations of three non-touching loops)+.....; Δ_k= the value of Δ for that part of the graph not touching the k^{th} forward path.

The signal flow graph of eq(1), we shall construct in such a way that all the feedback paths touch each other and all the forward paths touch the feedback paths. The Mason's gain formula then simplifies to

$$T(s) = \frac{\text{sum of forward path gains}}{1 - \text{sum of feedback path gains}} \ldots\ldots\ldots\ldots\ldots (5)$$

comparing Eq(1) and Eq(2) we observe that a signal flow graph of Eq (1) would consist of

i) Three feedback loops (touching each other) with gains $-a_i/s, -a_2/s^2$ and a_3/s^3;

ii) three forward paths which touch the loops and have gains $b_0, b_1/s, b_2/s^2$ and b_3/s^3

Phase variable format

The signal flow graph of Eq. (4) is shown in Fig:b

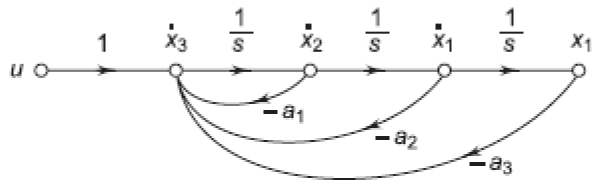


Fig. b

The SFG of <u>Eq. (2)</u> is shown in fig c.

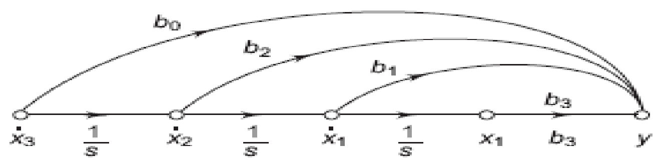


Fig. (c)

Combining fig(b)&(c) , the state diagram is shown in fig(d)

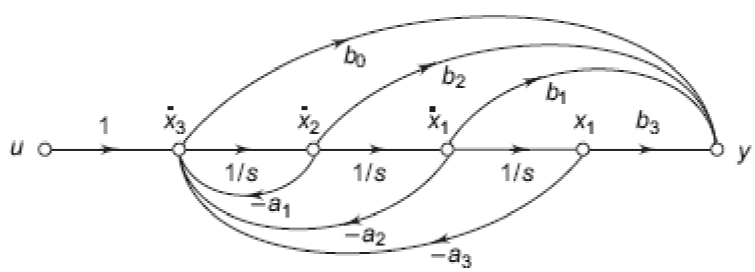


Fig. d

The relations are given below:

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = x_3$$
$$\dot{x}_3 = u - a_1x_3 - a_2x_2 - a_3x_1$$

Where x1, x2, x3 are the phase variables

The output y from the figure is found be

$$Y = b_3x_1 + b_2x_2 + b_1x3 + b_0u$$

The vector matrix form

$$\therefore \quad \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [u] \quad ....(6)$$

Bush form

$$Y = [b_3 \quad b_2 \quad b_1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b_0 u \; .... (7)$$

The direct feed forward term b_0u would be absent if m<n in T(s) of equation (1) i.e.

$$T(s) = \frac{b_0 s^2 + b_1 s + b_2}{s^3 + a_1 s^2 + a_2 s + a_3}$$

Input feed forward format:

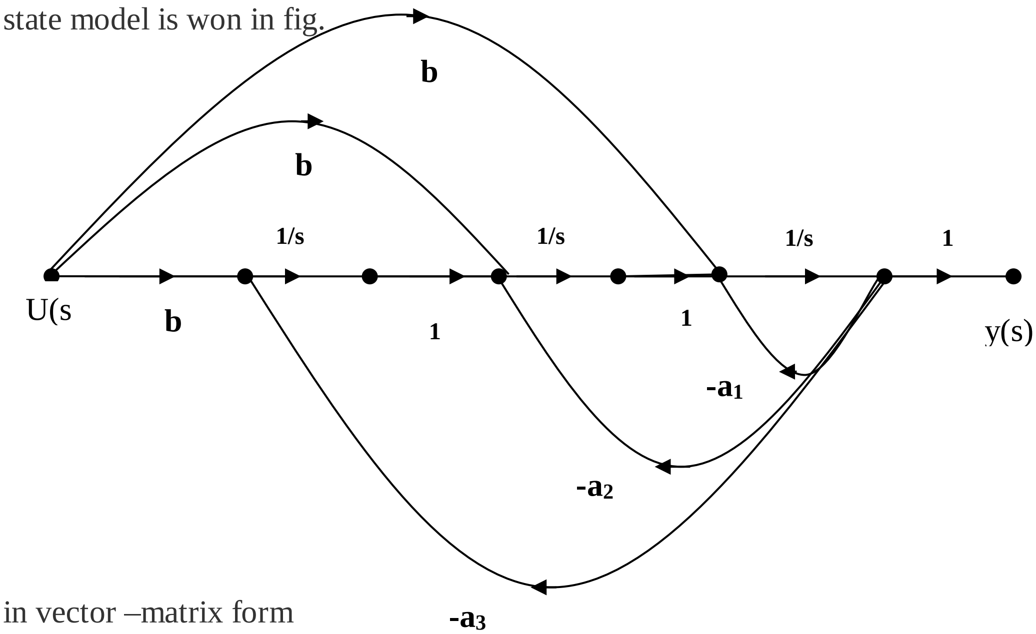
Consider T(s) with m<n. It has the form

$$T(s) = \frac{b_0 s^2 + b_1 s + b_2}{s^3 + a_1 s^2 + a_2 s + a_3}$$

It can be written as

$$T(s) = \frac{y(s)}{U(s)} = \frac{b_0/s + \frac{b_1}{s^2} + b_2/s^3}{1-(-\frac{a_1}{s} - \frac{a_2}{s^2} - \frac{a_3}{s^3})}$$

The denominator is constructed from state feed back as in the earlier formulation. The numerator is determined by three input fees forward. This alternative signal flow graph state model is won in fig.



in vector –matrix form

$$\frac{dx}{dt} = \begin{bmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} u$$

$$\dot{x} = Ax + Bu$$

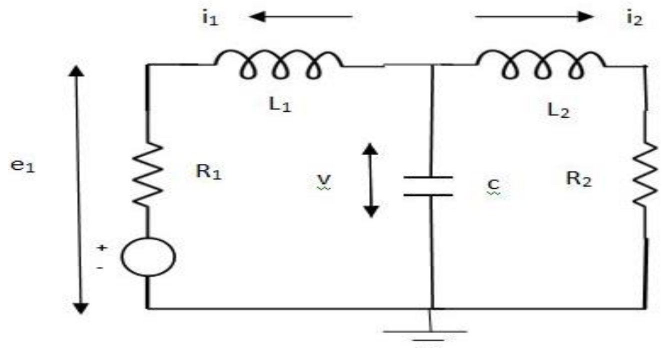
Or compare this A with bush form and observe the change in symmetry.

Output $y = x_1 = [1 \quad 0 \quad 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

From the above equations the phase variable formulation can be obtained by inspection from the transfer function and vice versa.

State space representation using physical variables:

The concept of state space representation is perhaps best introduced by considering an example. We shall consider state variable formulation for a simple electrical system which is an RLC network shown in fig (a).The network has a three energy storage elements: a capacitor C and two inductors L_1 and L_2.



Initial conditions v (0), $i_1(0)$,$i_2(0)$ together with the input signal e(t) for $t \leq 0$.then the selection of state variables would be

$$x_1 = v(t)$$
$$x_2 = i_1(t)$$
$$x_3 = i_2(t) \ldots\ldots\ldots\ldots\ldots\ldots..(1)$$

The differential equations governing the behavior of the RLC network are

$$i_1 + i_2 + c\,\frac{dv}{dt} = 0$$

$$L_1\frac{di1}{dt} + R_1 i_1 + e - v = 0$$

$$L_2\frac{di2}{dt} + R_2 i_2 - v = 0$$

From the above equations write these equations in form of state equation we get,

$$\frac{dv}{dt} = -\,(1/c)\, i_1 - (1/c)\, i_2$$

$$\frac{di1}{dt} = (1/L1)\, v - (R_1/L_1)\, i_1 - (1/L_1)\, e$$

$$\frac{di2}{dt} = (1/L_2)\, v - (R_2/L2)\, i_2$$

In terms of the state variables defined in Eq(1) and the input u(t) = e(t),then

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & -1/c & -1/c \\ -1/L_1 & -R_1/L_1 & 0 \\ 1/L_2 & 0 & -R_2/L_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ -1/L_1 \\ 0 \end{bmatrix} u \ldots\ldots.(2)$$

Assume that voltage across R_2 and current through R_2 are the output variables y1 and y_2 Respectively, the output equations are then given by

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & R_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \ldots\ldots\ldots\ldots..(3)$$

This concludes the state-space representation of the RLC network equations provide the state model of the systems.

Decomposition of transfer functions:

We know that a linear system can be modeled by differential equations, transfer function, or dynamic equations. All these methods are closely related. The state diagram is a useful tool that not only can lead to the solution of the state equations but also serves as a vehicle of transformation from one form of description to the others.

The block diagram of fig (1) shows, the relationship between the various ways of describing a linear system. And the starting for instance with the differential equation of a system, one can get to the solution by use of the transfer function method or the state equation method. And also shows that the majority of the relationships are bilateral, so a great deal of flexibility exists between the methods.

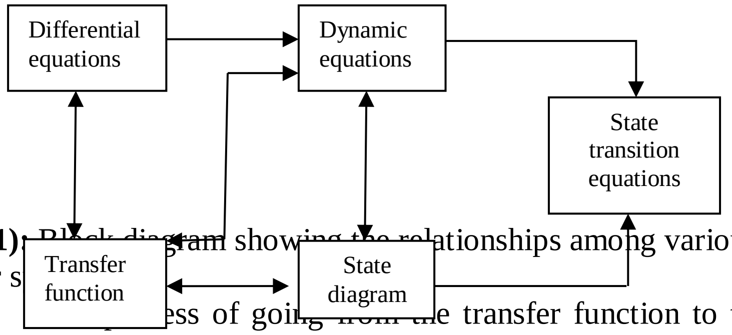


Fig (1): Block diagram showing the relationships among various methods of describing linear systems.

The process of going from the transfer function to the state diagram is called decomposition. In general there are three basic methods of decomposition of transfer functions: **direct decomposition**, **cascaded decomposition**, and **parallel decomposition**.

Direct decomposition:

First companion form: the direct decomposition is applied to an input-output transfer function that is not in factored form. Consider that the transfer function of an nth-order single-input-single output system between the input U(s) and the out Y(s) is

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

Here we have assumed that the order of the numerator and denominator are the same. To construct a state diagram from the transfer function, the following steps are to be followed.

1. Express the transfer function in negative powers of s. This is done by dividing each term present in the numerator and denominator of the transfer function by the highest power of s present in the denominator. This result in

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

2. Multiply and divide the transfer function eq.(1) by a X(s).This result in

$$\frac{Y(s)}{U(s)} = \frac{b_0 + b_1 s^{-1} + \ldots + b_{n-1} s^{-n+1} + b_n s^{-n} \quad X(s)}{1 + a_1 s^{-1} + \ldots + a_{n-1} s^{-n+1} + a_n s^{-n} \quad X(s)} \ldots\ldots\ldots\ldots\ldots (2)$$

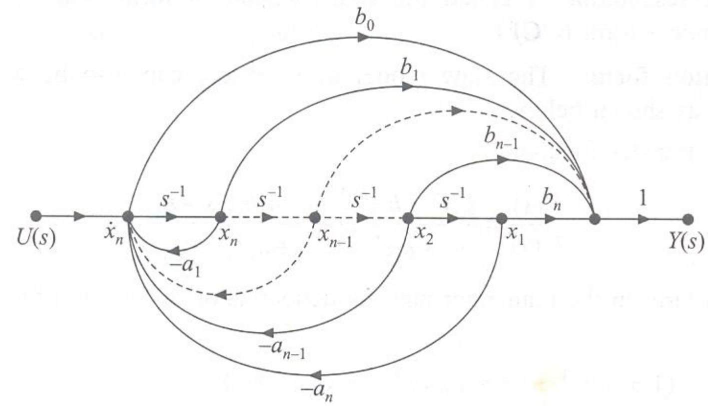
3. Equate the numerators and denominators on both sides of Eq (2):

$$Y(s) = b_0 + b_1 s^{-1} + \ldots + b_{n-1} s^{-n+1} + b_n s^{-n} \quad X(s) \ldots\ldots (3)$$
$$U(s) = 1 + a_1 s^{-1} + \ldots + a_{n-1} s^{-n+1} + a_n s^{-n} \quad X(s) \ldots\ldots (4)$$

4. Write the equations in cause-and –effect form. The equation for y(s) is already in cause-and effect form. Write the equation for U(s) in cause- and –effect form. Therefore,

$$X(s) = U(s) - a_1 s^{-1} X(s) - a_2 s^{-2} X(s) \ldots\ldots a_{n-1} s^{-n+1} X(s) - a_n s^{-n} X(s)$$

5. Construct the state diagram based on equations for Y(s) and X(s) as shown in fig (a).



Fig(a): state diagram for direct decomposition(first companion form)

The state variables $x_1(t)$, $x_2(t)$… $x_n(t)$ are defined as the outputs of the integrators and arranged in order from right to left on the state diagram. The state equations are obtained by expressing the first derivatives of the state variables in terms of the state variables and the input by applying signal flow graph gain formula.

From the state diagram, the dynamic equations are as follows:

$\dot{x}_1(t) = x_2(t)$

$\dot{x}_2(t) = x_3(t)$

$\vdots$

$\dot{x}_{n-1}(t) = x_n(t)$

$\dot{x}_n(t) = -a_n x_1(t) - a_{n-1}x_2(t) - \ldots\ldots -a_1 x_n(t) + u(t)$

and $y(t) = (b_n - a_n b_0)x_1(t) + (b_{n-1} - a_{n-1}b_0)x_2(t) + \ldots\ldots\ldots + (b_1 - a_1 b_0)x_n(t) + b_0 u(t)$

by writing them in matrix form, the state model is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots\ldots 0 \\ 0 & 0 & 1 & \ldots\ldots 0 \\ . & . & . & \ldots\ldots 0 \\ . & . & . & \ldots\ldots 0 \\ . & . & . & \ldots\ldots 0 \\ 0 & . & . & \ldots\ldots 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \ldots -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \\ b \end{bmatrix} [u]$$

$$y(t) = [(b_n - a_n b_0)\ (b_{n-1} - a_{n-1}b_0)\ldots\ldots\ (b_1 - a_1 b_0)] \begin{bmatrix} x_1(t) \\ x_2(t) \\ . \\ . \\ x_n(t) \end{bmatrix} + b_0\, u(t)$$

With usual notation, the state model is

$$\boldsymbol{\dot{x}(t) = A\, x(t) + B\, u(t)}$$
$$\boldsymbol{y(t) = C\, x(t) + D\, u(t)}$$

This form of representation is called the first companion form. This is also called the **Controllable canonical form (CCF).**

Second companion form:

The state model of a system can also be written in second companion form as shown below.

Consider the transfer function

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

Multiplying each term in the numerator and the denominator of Eq.1(a) by s^{-n} and cross multiplying

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

i.e, $y(s) = - a_1 s^{-1} + \ldots + a_{n-1} s^{-n+1} + a_n s^{-n})Y(s) + (b_0 + b_1 s^{-1} + \ldots + b_{n-1} s^{-n+1} + b_n s^{-n})U(s)$…..1(b)

Figure (b) shows the state diagram form of equation 1(b).the outputs of the integrators are designed as the state variables. However, unlike the usual convention, the state variables are assigned in descending order from right to left.

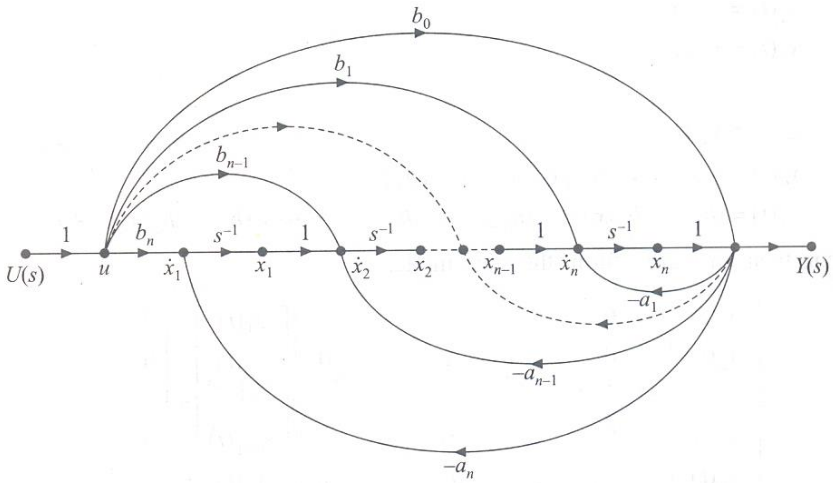


Fig: state diagram for second companion form

from the sate diagram of figure above fig, the dynamic equations of the system are as follows:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ . \\ . \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -a_n \\ 1 & 0 & 0 & -a_{n-1} \\ 0 & 1 & 0 & -a_{n-2} \\ . & . & . & . \\ 0 & 0 & 1 & -a_1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ . \\ x_n(t) \end{bmatrix} + \begin{bmatrix} b_n - a_n b_0 \\ b_{n-1} - a_{n-1} b_0 \\ b_{n-2} - a_{n-2} b_0 \\ . \\ b_1 - a_1 b_0 \end{bmatrix} u(t)$$

$y(t) = [\,0 \;\; 0 \ldots\ldots 1]\, x(t)$

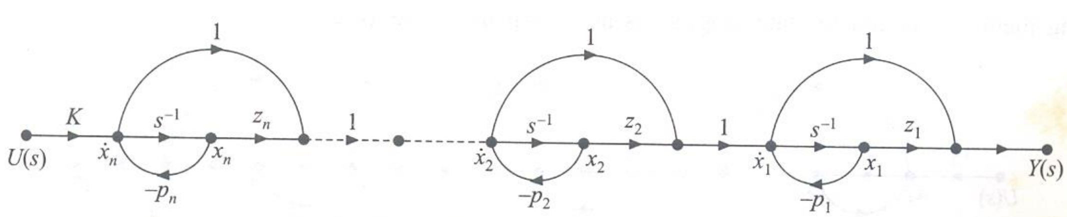
This form of representation is called the **observable canonical form (OCF).**

Cascaded decomposition:

Cascaded decomposition is applied to transfer functions that are written as products of simple first- or second-order components.

Consider the following transfer function which is the product of n first –order transfer functions.

$$\frac{Y(s)}{U(s)} = k\left(\frac{s+b_1}{s+p_1}\right)\left(\frac{s+b_2}{s+p_2}\right) \ldots \ldots . \left(\frac{s+b_n}{s+p_n}\right) \ldots\ldots\ldots(1)$$



let us take the product of two terms in first order transfer functions:

$$\frac{Y(s)}{U(s)} = k\,\frac{(s+b_1)(s+b_2)\cdots}{(s+a_1)(s+a_2)\cdots} \ldots\ldots\ldots\ldots\ldots \; (2)$$

Where $a_1, a_2, \ldots a_n$ and $b_1, b_2 \ldots\ldots\ldots b_n$ are all real constants.

Let us consider a general term for discussing the basic principle in cascade decomposition which is given below:

$$\frac{Y(s)}{X(s)} = \frac{s+b}{s+a} = \frac{1+bs^{-1}}{1+as^{-1}} = \frac{1+bs^{-1}}{1-(-as^{-1})} \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

The two forward paths between nodes $X(s)$ and $Y(s)$ is 1 and bs^{-1} and the loop gain is $-as^{-1}$. It touches both forward paths. This is shown in Fig.

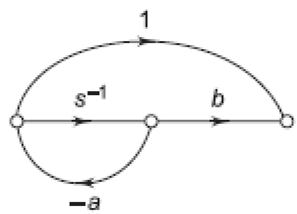


Fig. SFG

Let
$$\frac{Y(s)}{U(s)} = \frac{Y(s)}{X_1(s)} \cdot \frac{X_1(s)}{X_2(s)} \cdot \frac{X_2(s)}{U(s)}$$

where
$$\frac{Y(s)}{X_1(s)} = \frac{1+b_1 s^{-1}}{1+a_1 s^{-1}}$$
$$\frac{X_1(s)}{X_2(s)} = \frac{1+b_2 s^{-1}}{1+a_2 s^{-1}}$$

and
$$\frac{X_2(s)}{U(s)} = k$$

To obtain the state diagram, decompose each first –order transfer function by direct decomposition and connect all the state diagrams in cascaded as shown in below.

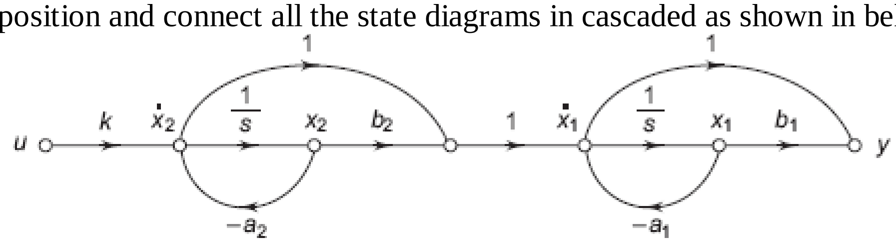


Fig. SFG of cascade decomposition of tow first-order transfer functions

State equations and output equations can be obtained from the state diagram by considering the output of each integrator as a state variable and expressing the first derivative of each state variable in terms of the state variables and the input as shown in below

$$\dot{x}_1 = b_2 x_2 + \dot{x}_2 - a_1 x_1$$

and
$$\dot{x}_2 = -a_2 x_2 + ku$$

$\therefore$
$$\dot{x}_1 = b_2 x_2 - a_2 x_2 + ku - a_1 x_1 = (b_2 - a_2) x_2 - a_1 x_1 + ku$$

$$\therefore \quad \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -a_1 & b_2 - a_2 \\ 0 & -a_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [k]\,[u]$$

$$\therefore \quad y = b_1 x_1 + \dot{x}_1 = b_1 x_1 + (b_2 - a_2) x_2 - a_1 x_1 + ku$$

$$= (b_1 - a_1) x_1 + (b_2 - a_2) x_2 + ku$$

$$\therefore \quad [y] = [b_1 - a_1 \;\; b_2 - a_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [k]\,[u]$$

The cascade decomposition has the advantage that the poles and zeros of the transfer function appear as isolated branch gains on the state diagram. This facilitates the study of the effects on the system when the poles and zeros are varied.

Parallel Decomposition:

when the denominator of the transfer function is in factored form, the transfer function may be expanded by partial-fraction.the resulting state diagram will consist of simple first- and second –order systems connected in parallel, which leads to state equations in canonical form or Jordan canonical form.

Consider an nth-order system represented by

Let $$\frac{Y(s)}{U(s)} = \frac{P(s)}{(s+a_1)(s+a_2)\cdots(s+a_n)} \quad \ldots\ldots\ldots.(1)$$

Where P(s) is a polynomial of order less than n and $a_1, a_2, \ldots\ldots. a_n$ are real and distinct. Although analytically $a_1, a_2, \ldots\ldots. a_n$ may be complex, realistically complex numbers are difficult to implement on a computer.

Expanding equation (1) into partial-fractions, we get

$$\frac{Y(s)}{U(s)} = \frac{d_1}{s+a_1} + \frac{d_2}{s+a_2} + \cdots + \frac{d_n}{s+a_n} \quad \ldots\ldots\ldots\ldots\ldots (2)$$

where $d_1, d_2, \ldots.., d_n$ are real constants. The state diagram of the system is drawn by the parallel combination of the state diagrams of each of the first-diagrams of each of the first-order terms in eq.(2) as shown in fig below.

$$Y(s) = \frac{d_1}{s+a_1} U(s) + \frac{d_2}{s+a_2} U(s) + \cdots + \frac{d_n}{s+a_n} U(s) \quad \ldots\ldots\ldots.(3)$$

$$= d_1 y_1(s) + d_2 y_2(s) + \ldots + d_n y_n(s) \quad \ldots\ldots\ldots(4)$$

where

$$\left. \begin{aligned} Y_1(s) &= \frac{1}{s+a_1} U(s) \\ Y_2(s) &= \frac{1}{s+a_2} U(s) \\ &. \\ &. \\ &. \\ Y_n(s) &= \frac{1}{s+a_n} U(s) \end{aligned} \right\}$$

The inverse Laplace Transformation of the above equations are

$$\left.\begin{aligned} \dot{y}_1 &= -a_1 y_1 + u \\ \dot{y}_2 &= -a_2 y_2 + u \\ &\;\vdots \\ \dot{y}_n &= -a_n y_n + u \end{aligned}\right\} \quad \ldots.(5)$$

Fig. Below shows the SFG of Eq.(5).

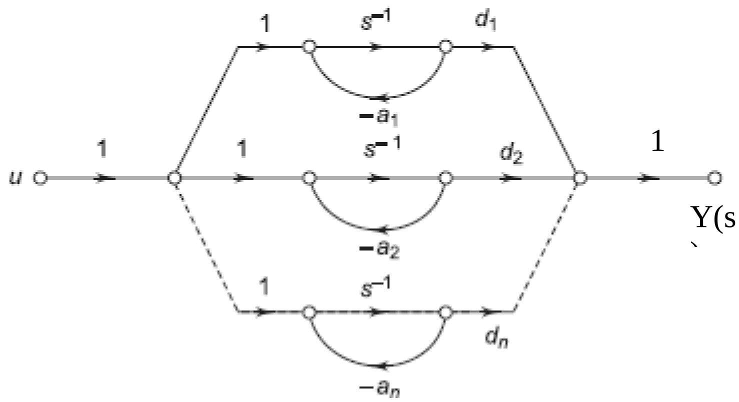


Fig. State diagram by parallel decomposition (only non-repeated poles).

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{y}_n \end{bmatrix} = \begin{bmatrix} -a_1 & 0 & \cdots\cdots\cdots 0 \\ 0 & -a_2 & \cdots\cdots\cdots 0 \\ \cdot & \cdot & \cdots\cdots\cdots 0 \\ \cdot & \cdot & \cdots\cdots\cdots 0 \\ \cdot & \cdot & \cdots\cdots\cdots 0 \\ 0 & 0 & \cdots\cdots\cdots -a_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} [u]$$

$\therefore \qquad \dot{y}_1 = \mathbf{A}y + \mathbf{B}u$

and inverse Laplace transform of (4) is

$y = d_1 y_1 + d_2 y_2 + \ldots + d_n y_n$

$$[y] = \begin{bmatrix} d_1 & d_2 & \cdots & d_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} + [0]\,[u]$$

$\therefore \qquad y = \mathbf{C}y + \mathbf{D}u$

The matrix A is in canonical form and the main diagonal elements are nothing but the Eigen values of the matrix A. this form of representation is called the diagonal canonical form (DCF).It is also called the normal form. This form of state model plays an important role in control theory. It has unique decoupled nature due to the fact that, the n first-order differential equations are completely independent of each other. The disadvantage of the canonical form is: the canonical variables like phase variables are not physical variables of the system.

If the transfer function has repeated poles, the A matrix will be in Jordan canonical form. Suppose the transfer function has three poles as $-p_1, -p_1, -p_1$, and two poles as $-p_2, -p_2$ and the remaining poles are distinct, then the partial-fraction expansion is

$$\frac{Y(s)}{U(s)} = \frac{K_1}{(s+p_1)^3} + \frac{K_2}{(s+p_1)^2} + \frac{K_3}{(s+p_1)^1} + \frac{K_4}{(s+p_2)^2} + \frac{K_5}{(s+p_2)^1} + \cdots \ldots\ldots\ldots\ldots. + \frac{K_n}{(s+p_n)}$$

The corresponding state diagram is as swon in below figure.
Choosing the outputs of the integrators as the state variables, the state equations are

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \\ \dot{x}_5(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} -p_1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -p_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -p_1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & -p_2 & 1 & \cdots & 0 \\ 0 & 0 & 0 & 0 & -p_2 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -p_n \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \\ \vdots \\ x_n(t) \end{bmatrix} + \begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ K_4 \\ K_5 \\ \vdots \\ K_n \end{bmatrix} u(t)$$

Observe the 3×3 and 2×2 Jordan blocks.

In the Jordan block, all the diagonal elements are the same (value of the repeated poles) and the elements just above the main diagonal are all 1s and all other elements are 0s.This form of representation is called the Jordan canonical form (JCF)

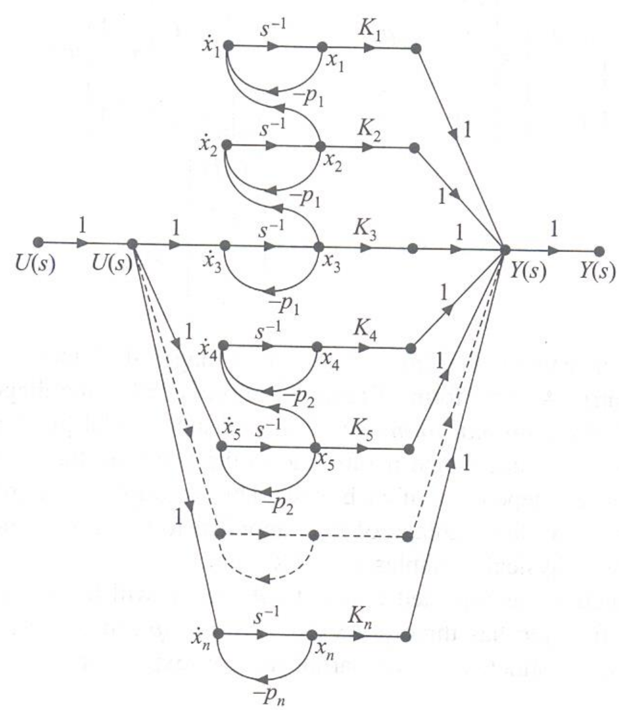


Figure: state diagram for parallel decomposition (repeated poles present).

Derivation of transfer functions from state model:

From the state model equations, we can derive the transfer function of the system. From definition of transfer function, we can write

$$\text{Transfer function} = \left.\frac{\text{Laplace transform of output}}{\text{Laplace transform of input}}\right|_{\text{Initial conditions zero}}$$

For a linear time-invariant system the dynamics can be written in vector differential form as follows:

$\dot{x}$ = Ax+Bu.......(a)

$y = \mathbf{C}x + \mathbf{D}u$ (b)

Taking Laplace transform of Eq. (a), we can write

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

$\therefore \quad [sI - \mathbf{A}]\,\mathbf{X}(s) = \mathbf{B}U(s)$ [Taking $x(0) = 0$]

$\therefore \quad \mathbf{X}(s) = [sI - \mathbf{A}]^{-1}\mathbf{B}U(s)$ (c)

Taking Laplace transform of Eq. (b), it can be written as

$$Y(s) = \mathbf{C}X(s) + \mathbf{D}U(s) = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}U(s) + \mathbf{D}U(s)$$

$$\text{Transfer function} = \frac{Y(s)}{U(s)} = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B} + \mathbf{D} \qquad \text{...........(d)}$$

If there is no direct coupling between input and output, $D = 0$. Eq. (d) becomes

$$\text{Transfer function} = \frac{Y(s)}{U(s)} = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}$$

Therefore, transfer function is the ratio of two polynomials of s. The denominator polynomial of s of the transfer function is known as characteristic equation. The characteristic equation is given by

$|s\mathbf{I} - \mathbf{A}| = 0$ **(e)**

Whose roots are the poles of the system also called **Eigen values.**

Diagonalization:

From the above discussions, we observed that the state model of a system is not unique. Depending on the choice of the state variables, the same system can be represented using many state models. The state variables selected may be physical variables, phase variables, canonical variables or some other variables.

From application point of view, the choice of physical variables as state variables for system representation is most useful, as these variables can be easily measured and used for control purposes. However, the corresponding state model in this case is generally not convenient for investigation of system properties and evaluation of time response. The phase variables in general are not physical variables of the system and therefore are not available for measurement and control purposes.

The canonical state model wherein A is in diagonal form is the most suitable for this purpose. So we discuss here the techniques for transforming a general state model into a canonical one. These techniques are generally referred to as diagonalization techniques.

Consider an nth-order-multi-input-multi-output state model

$\dot{x}$ = A x+ B u…..(1)

y = C x + D u……..(2)

Assume that the matrix a in this model is non diagonal. Let us define a new state vector z such that

$x = \mathbf{M}z$ …..(3)

Where M is a (n×n) non singular constant matrix. Under this transformation, the original state model modifies to

using (1)&(3) equations can be written as

$$\mathbf{M}\dot{z} = \mathbf{AM}z + \mathbf{B}u \quad \ldots\ldots\ldots \quad (4)$$

$$\therefore \quad \dot{z} = \mathbf{M}^{-1}\mathbf{AM}z + \mathbf{M}^{-1}\mathbf{B}u = \mathbf{P}z + \mathbf{Q}u \quad \ldots\ldots\ldots\ldots(5)$$

Using Eq (2) &(3) can be written as

y = **CM**z + **D**u = **R**z + **D**u…. (6)

If the matrix M can be selected such that $\mathbf{M}^{-1}\mathbf{AM}$ is a diagonalized matrix A, then the model given by (5)&(6) is a canonical state model. Under this condition, the matrix M is called the diagonalizing matrix or the modal matrix. Eq (5) & (6) can be written as

$\dot{z}$ = Pz + Qu

y = Rz + Du

where $\mathbf{P} = \mathbf{M}^{-1}\mathbf{AM}$ **= a diagonal matrix**

$\mathbf{Q} = \mathbf{M}^{-1}\mathbf{B}$

R = **CM**

and **D** = **D**

The determination of the diagonalization matrix is facilitated by the use of eigenvectors.

Eigen values and Eigenvectors:

let us now take

Ax = y

and it as transformation of n×1 vector x to n×1 vector y by n×n matrix operator A.

Here there exists a vector **x** such that matrix operator A transforms it to a vector **λx** (λ is a constant) i.e., to a vector having the same direction in state space as the vector **x**. such a vector **x** is a solution of the equation.

Ax = λx…..(a)

Above eq can be written as

(λI – A)x = 0….(b)

The set of homogenous eqs (b) have a nontrivial solution if and only if

$$|\boldsymbol{\lambda I - A}| = \mathbf{0} \ldots\ldots\text{ (c)}$$

This equation may be expressed in expanded form as

$$q(\lambda) = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \ldots\ldots + a_n = 0 \ldots\text{(d)}$$

The values of **λ** for which equ (c) is satisfied are called Eigen values of matrix A and equ (d) is called the characteristic equation corresponding to matrix A.

are often referred to as the Eigen values of the matrix **A**. we know that the roots of the characteristic equation

$$|\boldsymbol{sI - A}| = \mathbf{0}$$

$s^n + a_1s^{n-1} + a_2s^{n-2} + \ldots\ldots\ldots + a_n = 0$ in terms of **s** are the same as the poles of the closed loop transfer function. So it can be concluded that the Eigen values of A and the poles of the closed loop transfer function are the same. Hence system stability based upon the location of transfer function poles (i.e roots of the characteristic equation) are therefore valid for the Eigen values of the state model. Thus a state model is stable if all its Eigen values have negative real parts.

Some of the important properties of Eigen values are as follows:

- If the coefficients of A are all real, its Eigen values are either real or in complex conjugate pairs.
- The trace (sum of the main diagonal elements) of A is the sum of all the Eigen values of A.
- The Eigen values of the transpose of A i.e of A^T are the same as the Eigen values of A
- The Eigen values of the inverse of A, i.e. of A^{-1} are the inverse of the Eigen values of A.

Eigen vectors:

Let us take $\boldsymbol{\lambda = \lambda_i}$ satisfying equation (c) ,we have from equation (d)

$$(\boldsymbol{\lambda_i I - A)x = 0} \ldots\ldots\ldots\text{(e)}$$

Let $x = m_i$ be the solution of this equation. The solution m_i is called the eigen vector of A associated with eigen value $\boldsymbol{\lambda_i}$

Solution of (e) depends on the rank of matrix **(λ I – A).**If **r** is the rank of this matrix, then there are **(n-r**) independent solutions (Eigen vectors).

If the Eigen values of matrix A are all distinct, then the rank r of matrix **(λ I-A) is (n-1)** and hence we have only independent Eigen vector corresponding to any particular Eigen value $\boldsymbol{\lambda_i}$**.** This Eigen vector, may be obtained by taking cofactors of matrix $(\boldsymbol{\lambda_i I - A})$ along any row i.e.,

$$M_i = \begin{bmatrix} C_{k1} \\ C_{k2} \\ \cdot \\ C_{kn} \end{bmatrix}; \; k = 1,2,\ldots, \text{ or } n$$

where C_{ki} are the co-factors of matrix **(λ_i I - A).**

let $\mathbf{m_1, m_2, \ldots m_n}$ be the Eigen vectors corresponding to the Eigen value $\lambda_1, \lambda_2, \ldots\ldots \lambda_n$ respectively. Then we have

$$\mathbf{AM = A[m_1 \;\; m_2 \;\; \ldots\ldots\ldots m_n]}$$

$$\mathbf{= [Am_1 \;\; Am_2 \ldots\ldots\ldots Am_n]}$$

$$\mathbf{= [\lambda_1 m_1 \;\; \lambda_2 m_2 \ldots\ldots A_n m_n]}$$

$$\mathbf{= M\,p}$$

$$\textbf{Where p=} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \mathbf{M^{-1} A\, M}$$

The matrix constructed be placing the Eigen vectors (columns) together is therefore is a diagonalizing or modal matrix **M of A**, i.e.,

$$M = [\mathbf{m_1 \;\; m_2 \;\; \ldots\ldots\ldots m_n}]$$

Note that **A** and **p** have the same characteristic equation; therefore the Eigen values are invariant under this transformation.

When A is expressed in the comparison form given below,

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

Then the modal matrix can be shown to be a special matrix (called the Vannder Monde Mtrix)

$$V = \begin{bmatrix} 1 & 0 & \cdots & 1 \\ \boldsymbol{\lambda_1} & \boldsymbol{\lambda_2} & \ldots & \boldsymbol{\lambda_n} \\ \boldsymbol{\lambda_1}^2 & \boldsymbol{\lambda_2}^2 & \ldots & \boldsymbol{\lambda_n}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\lambda_1}^{n-1} & \boldsymbol{\lambda_2}^{n-2} & \ldots & \boldsymbol{\lambda_n}^{n-1} \end{bmatrix}$$

Solution of state equations:

For time invariant system, the state equation is given by

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

which is divided into two types as follows:

- homogeneous state equation
- non-homogeneous state equation

The equation is said to be **homogeneous**, if **A** is a constant matrix and **U** is a zero vector. In this case no control forces are applied. The equation is said to be **non-homogeneous,** if **A** is constant matrix and **U** is a non-zero vector. In the first case discharging a charged capacitor is an example while in the second case total response is an example.

Let us consider the state equation

$\dot{x}(t)$= Ax(t); x(0) = x_0……………(1)

which represents a homogenous linear system (unforced system) with constant coefficients.

By analogy with the scalar case,we assume a solution of the form

$$\mathbf{x(t) = a_0 + a_1 t + a_2 t^2 + \ldots\ldots\ldots a_i t^i + \ldots}$$

Where a_i are vector coefficients.

By substituting the assumed solution into equ(1) we get

$$\mathbf{a_1 + 2a_2 t + 3a_3 t^2 + \ldots\ldots\ldots = A\,(a_0 + a_1 t + a_2 t^2 + \ldots\ldots\ldots)}$$

The comparison of vector coefficients of equal powers of t, yield

$$\mathbf{a_1 = Aa_0}$$

$$\mathbf{a_2 = ½\, Aa_1 = \frac{1}{2!} A^2 a_0}$$

$$\mathbf{a_i = \frac{1}{i!} A^i a_0}$$

In the assumed solution, equating x (t = 0) = x_0, find that

$$\mathbf{a_0 = x_0}$$

The solution x(t) is thus found to be

$$\mathbf{x(t) = (\, I + At + \frac{1}{2!} A^2 t^2 + \ldots\ldots\ldots + \frac{1}{i!} A^i t^i + \ldots\ldots\ldots\ldots)x_0} \ \ldots\ldots\ldots\ldots\ldots\mathbf{(2)}$$

Each of the terms inside the brackets is an n × n matrix. Because of the similarity of the entity inside the brackets with a scalar exponential of Eq (2), we call is a matrix exponential. This may be written as

$$\mathbf{e^{At} = I + At + \frac{1}{2!} A^2 t^2 + \ldots\ldots\ldots + \frac{1}{i!} A^i t^i \ldots\ldots\ldots\ldots(3)}$$

The solution x(t) can now be written as

$$\mathbf{x\,(t) = e^{At} x_0 \ \ldots\ldots\ldots(4)}$$

From equ (4) it is observed that the initial state x_0 at t = 0,is driven into a state x(t) at time t.This is transition in state is carried out by the matrix exponential e^{At}. Because of this property, e^{At} is known as **state transition matrix** and is denoted by $\phi(t)$.

The solution of the **non homogenous** state equation (forced system):

$$\dot{x}(t) = \mathbf{A}x(t) + \mathbf{B}u(t) \quad ; x(0) = x_0$$

Re write this equation in this form

$$\boldsymbol{\dot{x}(t)} \textbf{- A x(t) = Bu(t)}$$

Multiplying both sides by e^{-At}, we can write

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

Integrating both sides with respect to **t** between the limits **0** and **t,** we get

$$\mathbf{e^{-At}\,x(t)\,\Big|_0^t = \int_0^t e^{-At}\,Bu(\tau)d\tau}$$

$$\mathbf{e^{-At}x(t) - x(0) = \int_0^t e^{-At}\,Bu(\tau)d\tau}$$

Now pre multiplying both sides by e^{At}, we have

$$x(t) = \underbrace{\mathbf{e^{-At}x(0)}}_{\text{Homogenous Solution}} + \underbrace{\int_{t_0}^{t} e^{-A(t-\tau)}\,\mathbf{Bu(\tau)d\tau}}_{\text{Forced solution}} \qquad \textbf{.......................(5)}$$

Properties of the state transition matrix:

- $\phi\mathbf{(0)} = e^{\mathbf{A0}} = I$

- $\phi(t) = e\mathbf{A}t = (e{-}\mathbf{A}t)^{-1} = [\phi(-t)]^{-1}$ **(or)** $\phi^{-1}(t) = \phi(-t)$

- $\phi(t_1 + t_2) = e^{(t1+t2)} = e^{\mathbf{A}t1}e^{\mathbf{A}t2} = \phi(t_1)\phi(t_2) = e^{\mathbf{A}t2}e^{\mathbf{A}t1} = \phi(t_2)\phi(t_1)$

- $[\phi(t)]^n = (e^{\mathbf{At}})^n = e^{\mathbf{A}nt} = \phi(nt)$

- $\phi(t_1 - t_2)\phi(t_2 - t_0) = e^{\mathbf{A}(t1 - t2)}\,e^{\mathbf{A}(t2 - t0)} = e^{(t1 - t2 + t2 - t0)} = \phi(t_1 - t_0)$

In terms of the state transition matrix $\phi(t)$,the solution of the forced system given by Eq (5) can be written as

$$x(t) = \phi(t)\,x)(0) + \int_0^{\mathbf{t}} \boldsymbol{\phi}(t-\tau) \quad \mathbf{Bu(\tau)d\tau}\textbf{........(6)}$$

Concept of Controllability and observability:

Controllability:

Definition**: A system is said to be controllable at time t0 if it is possible by means of an unconstrained control vector to transfer the system from any initial state to any other state in a finite interval of time. Controllability depends upon the system matrix A and the control influence matrix B.**

The concept of controllability involves the dependence of state variable of the system on the inputs. Consider a single input linear time invariant system

$$\dot{x} = \mathbf{A\,x + B\,u} \quad \ldots\ldots\ldots\ldots\ldots(1)$$

Where x = n - dimensional state vector;

U= control signal (control force);

A= n×n matrix, and **B**= n× 1 matrix

Let the initial system state be x(0) and the final desired state be $x(t_f)$.the system described be above equation (1) is controllable if it is possible to construct a control signal, which in finite time interval $0 < t < t_f$, will transfer the syatem state from x(0) to $x(t_f)$.

Let us take state solution,

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

The system described above equations completely controllable if state variable x can be transferred from initial state to final state in a finite time. In other words the system is controllable if it is possible to construct a control signal u(t) such that the following

Assuming $x(t_1)= 0$

$$0 = \mathbf{e}^{-At_1}x(0) + \int_0^{t_1} e^{-A(t_1-\tau)}\, \boldsymbol{Bu(\tau)d\tau}$$

$$X(0) = \int_0^{t_1} e^{-A\tau}\, \boldsymbol{Bu(\tau)d\tau}$$

$$e^{-A\tau} = \sum_{k=0}^{n-1} \alpha_k(\tau)A^k \quad \text{(Sylvester's formula)}$$

$$X(0) = \int_0^{t_1} e^{-A\tau}\, \boldsymbol{Bu(\tau)d\tau}$$

$$= \sum_{k=0}^{n-1} \mathbf{A^k B} \int_0^{t_1} \boldsymbol{\alpha_k(\tau)u(\tau)d\tau}$$

$$= \sum_{k=0}^{n-1} \mathbf{A^k B}\ \boldsymbol{\beta_k} \qquad \textbf{where } \boldsymbol{\beta_k} \triangleq \int_0^{t_1} \boldsymbol{\alpha_k(\tau)u(\tau)d\tau}$$

$$= [\mathbf{B\ AB\ A^2B\ldots\ldots A^{n-1}B}]\ [\boldsymbol{\beta_0\ \ \beta_1\ldots\ldots\ldots\ldots\beta_{n-1}}]^T$$

This system should have a non-trivial solution for $[\boldsymbol{\beta_0\ \ \beta_1\ldots\ldots\ldots\ldots\beta_{n-1}}]^T$

If the system is completely controllable if and only if the rank of the composite matrix is

$$\boxed{\mathbf{C_B \triangleq [\,B\ AB\ A^2B\ldots\ldots A^{n-1}B\,]}}$$

If there is no connection between a certain state and input, the state is not controllable as shown in Fig below

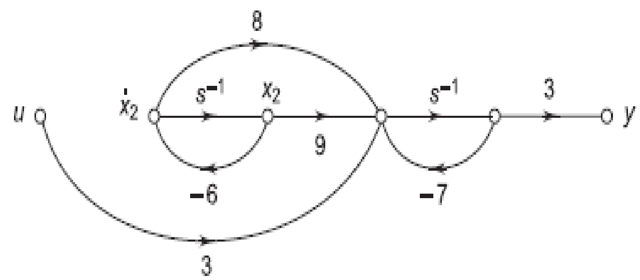


Fig. x_2 is not controllable

Condition for complete state controllability in the s-plane:

Condition for complete state controllability can be used in terms of transfer functions or transfer matrices. A necessary and sufficient condition for complete state controllability is that no cancellation occurs in the transfer function of transfer matrix. If cancellation occurs, the system cannot be controlled in the direction of the cancelled mode.

Output Controllability:

In practical design of a control system, we may want to control the output rather than the sate of complete state controllability is neither necessary nor sufficient for controlling the output of the system..

Hence, the system described by

$$\mathbf{\dot{x}(t) = A\,x(t) + B\,u(t)}$$
$$\mathbf{y(t) = C\,x(t) + D\,u(t)}$$

is said to be completely output controllable, if it is possible to construct an unconstrained controlled vector u(t) that will transfer any given initial output $y(t_0)$ to any final output $y(t_f)$ in a finite time interval $t_0 \le t \le t_f$.

$X \in R^n$, $U \in R^m$, $Y \in R^P$, then the rank of $\mathbf{C_B \triangleq [\,CB\ \ CAB\ \ CA^2B \ldots\ldots CA^{n-1}B\ \ D\,]}$ is p then the system is output controllable.

Note: The presence of DU term in the output equation always helps to establish output controllability.

Observability:

Definition**: A system is said to be *observable* at time *t0* if, with the system in state *X (t0)* ,it is possible to determine this state from the observation of the output over a finite interval of time. Observability depends upon the system matrix A and the output matrix C.**

A system is completely observable, if every state variable of the system affects some of the outputs. In other word, it is often desirable to obtain information on the sate variables from the measurements of the outputs and the inputs. If any one of the states cannot be observed from the measurements of the output, the state is said to be unobservable and the system is not completely observable or simply unobservable.

Figure below shows the state diagram of a linear system in which the state x_2 is not connected to the output y(t) in any way. Once we have measured y(t),we can observe the state x_1(t),since $x_1(t) = y(t)$.however, the state x_2(t) cannot be observed from the information on y(t).thus the system is unobservable.

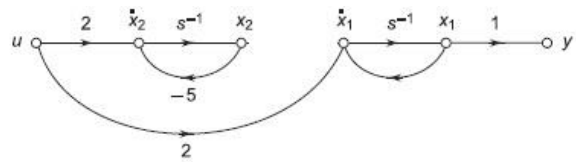


Fig: State diagram of an unobservable system.

A linear time – invariant system described be the state and output equations.

$$\dot{x}(t) = A\,x(t) + B\,u(t)$$
$$y(t) = C\,x(t) + D\,u(t)$$

is said to be completely observable, if every state $x(t_0)$ can be determined from the observation of the output y(t) over a finite time interval, $t_0 \le t \le t_1$.the system is therefore completely observable if every transition of the state eventually affects every element of the output vector.

The concept of observability is useful in solving the problem of reconstructing immeasurable state variables from measurable state variables in the minimum possible length of the kalman's test of observability is a general nth-order multi-input- multi-output linear time-invariant system described by

$$\dot{x}(t) = A\,x(t) + B\,u(t)$$
$$y(t) = C\,x(t) + D\,u(t)$$

is completely observable, if and only if, the rank of the observability matrix

$$Q_C \triangleq [\,C^T\ A^TC^T\ \ldots\ldots(A^T)^{n-1}C^T\,] \text{ is } n$$

OR The necessary and sufficient condition for observability is that the following matrix

$$Q_c = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix} \textbf{ must have rank } \boldsymbol{n}.$$