SEE1203 – CONTROL SYSTEMS

UNIT I

SYSTEM CONCEPTS

TYPES OF SYSTEMS

Control systems are basically classified as –

- Open-loop control system
- Closed-loop control system

In open-loop system the control action is independent of output. In closed-loop system control action is somehow dependent on output. Each system has at least two things in common, a controller and an actuator (final control element). The input to the controller is called reference input. This signal represents the desired system output.Open-loop control system is used for very simple applications where inputs are known ahed of time and there is no disturbance. Here the output is sensitive to the changes in ***disturbance*** inputs. Disturbance inputs are undesirable inputs that tend to deflect the plant outputs from their desired values. They must be calibrated and adjusted at regular intervals to ensure proper operation. Closed-loop systems are also called feedback control systems. Feedback is the property of the closed-loop systems which permits the output to be compared with the input of the system so that appropriate control action may be formed as a function of inputs and outputs. Feedback systems has the following features:

- reduced effect of nonlinearities and distortion
- Increased accuracy
- Increased bandwidth
- Less sensitivity to variation of system parameters
- Tendency towards oscillations
- Reduced effects of external disturbances

The general block diagram of a control system is shown below.

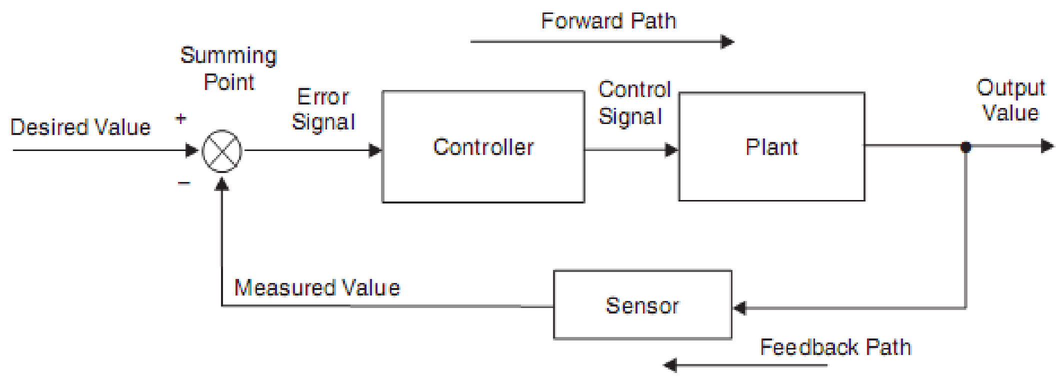


Figure: Closed-loop control system

Some Definitions

Reference input – It is the actual signal input to the control system.
Output (Controlled variable) – It is the actual response obtained from a control system.
Actuating error signal – It is the difference between the reference input and feedback signal.
Controller – It is a component required to generate control signal to drive the actuator.
Control signal – The signal obtained at the output of a controller is called control signal.
Actuator – It is a power device that produces input to the plant according to the control signal, so that output signal approaches the reference input signal.
Plant – The combination of object to be controlled and the actuator is called the plant.
Feedback Element – It is the element that provides a mean for feeding back the output quantity in order to compare it with the reference input.
Servomechanism – It is a feedback control system in which the output is mechanical position, velocity, or acceleration.

EXAMPLE OF CONTROL SYSTEMS

Toilet tank filling system:

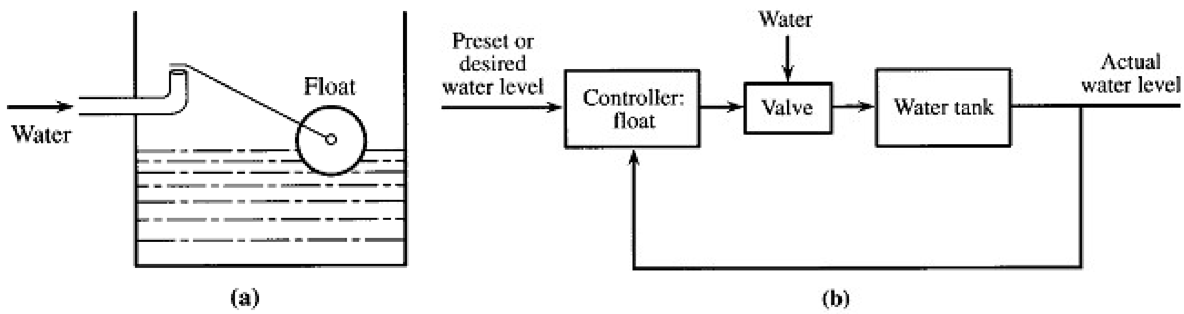


Figure: Toilet tank filling system

Position control system: [antenna]

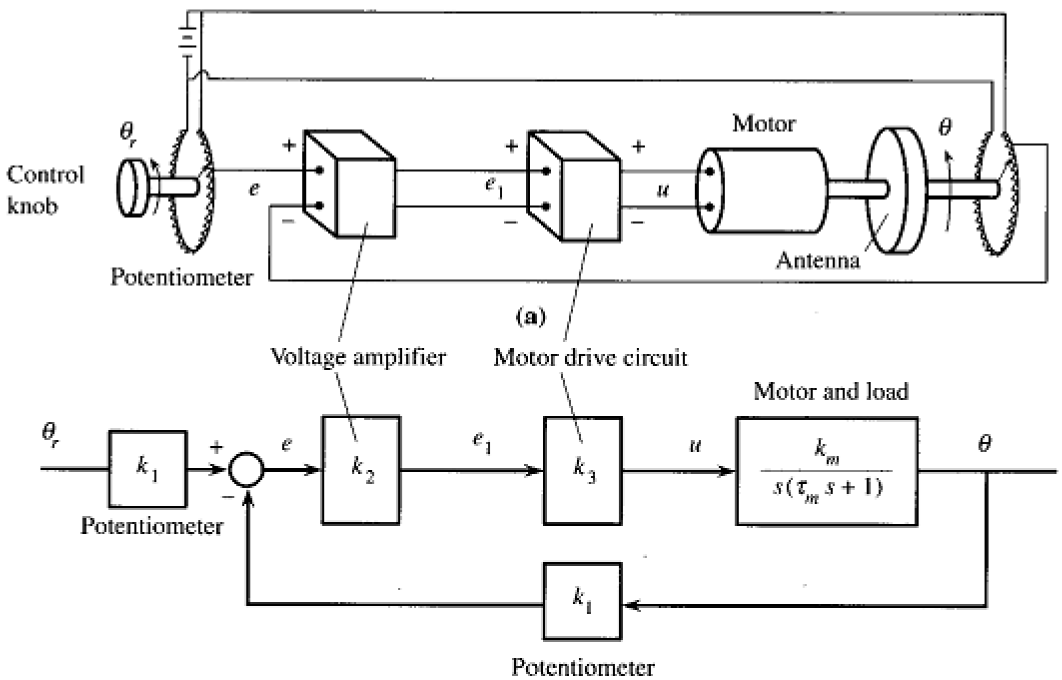


Figure: Position control system

Velocity control system: [audio/ video recorder]

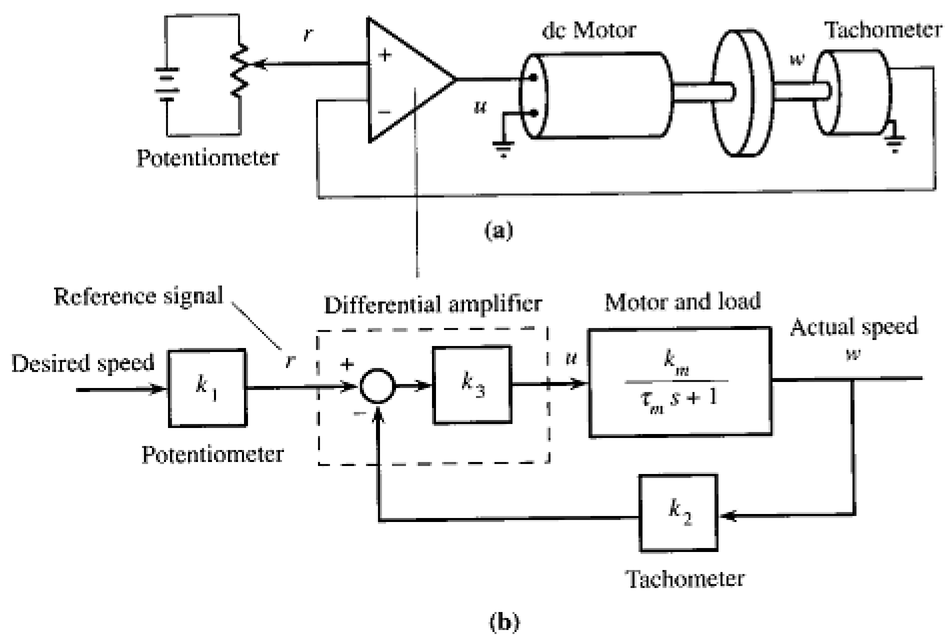
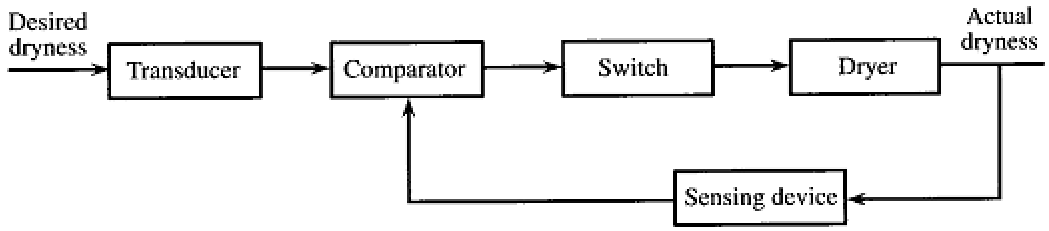


Figure: Velocity control system

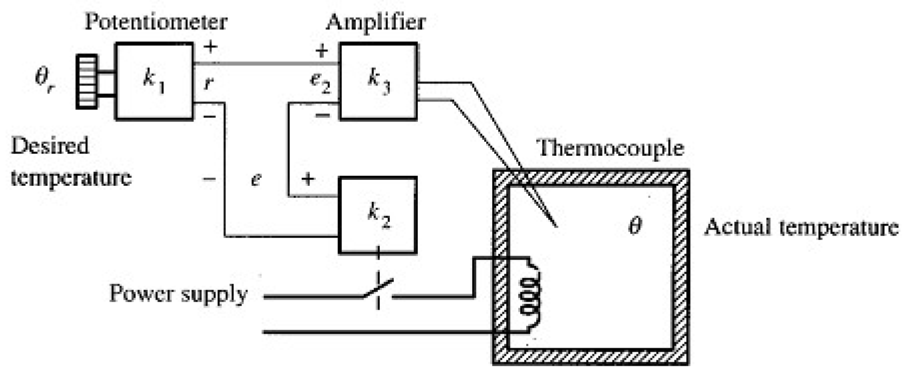
Clothes Dryer:



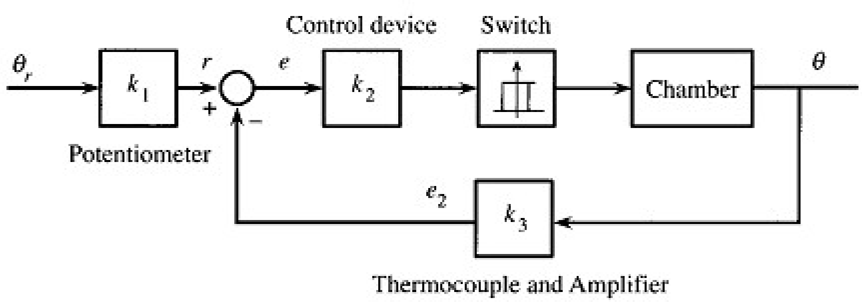
(b)

<u>Figure: Automatic dryer</u>

Temperature control system: [oven, refrigerator, house]



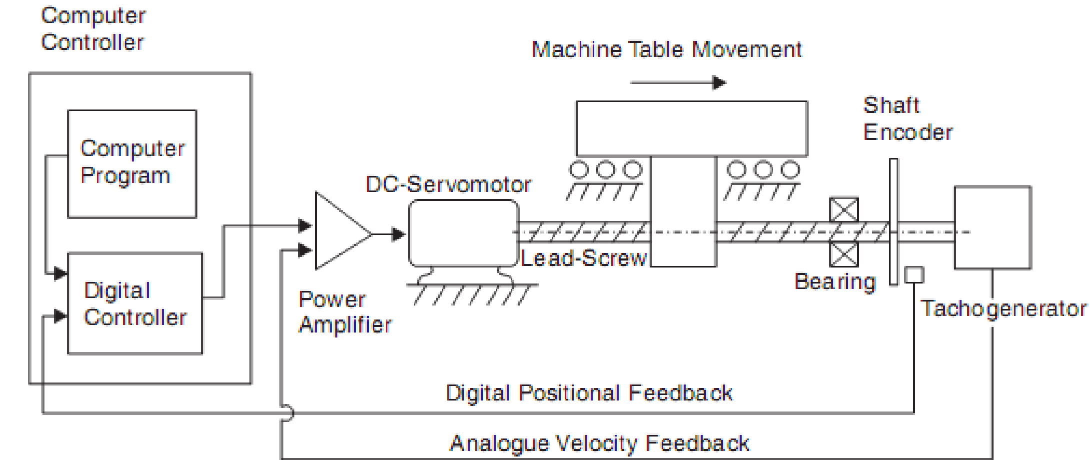
(a)



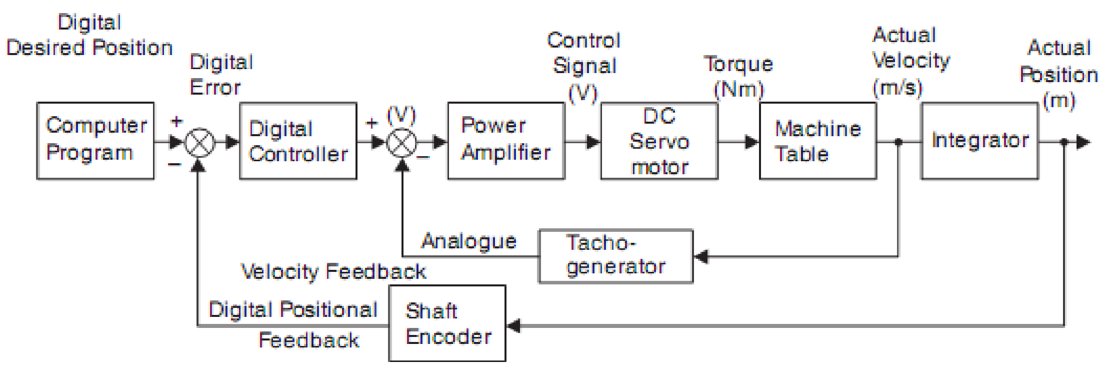
(b)

<u>Figure: Temperature control system</u>

Computer numerically controlled (CNC) machine tool:



(a)

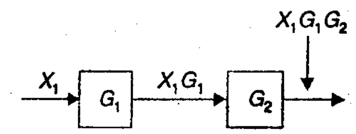
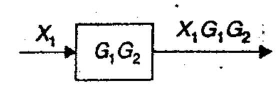
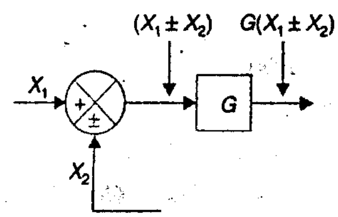
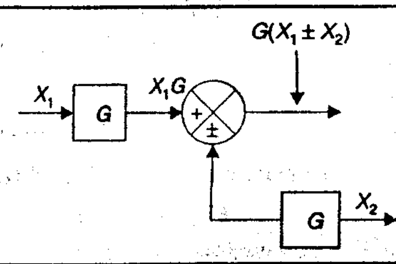
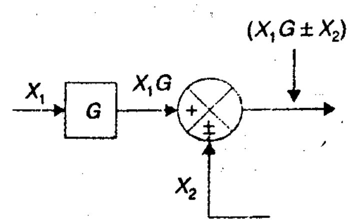
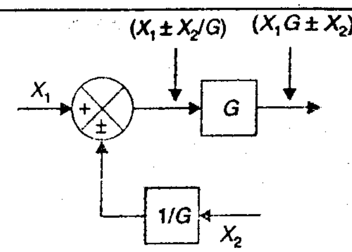
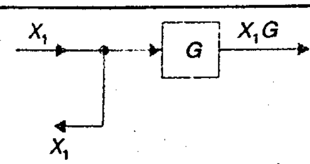
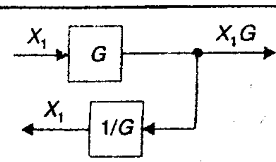
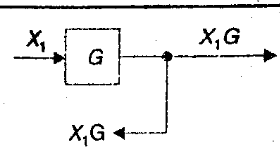
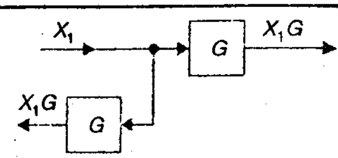
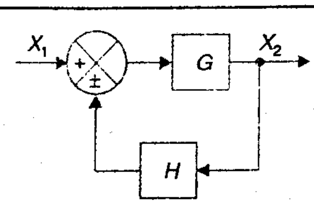
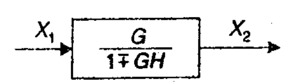


(b)

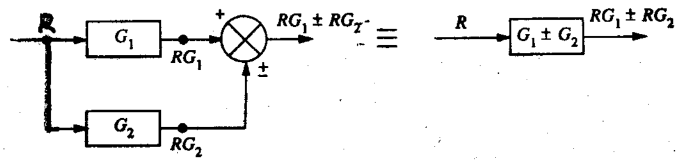
Figure: CNC machine tool control system

BLOCK DIAGRAM ALGEBRA

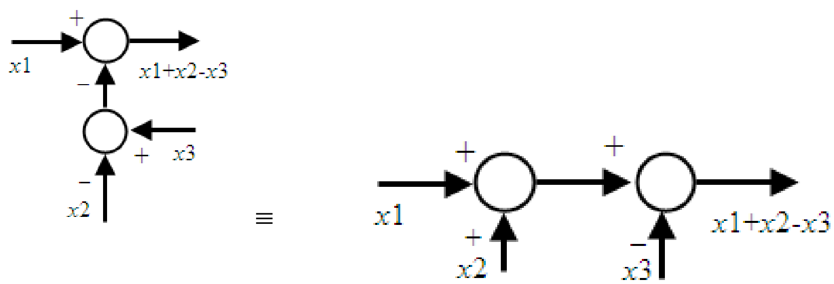
A complex system is represented by the interconnection of the blocks for individual elements. Evaluation of complex system requires simplification of block diagrams by block diagram rearrangement. Some of the important rules are given in figure below.

Rule	*Original diagram*	*Equivalent diagram*
1. Combining blocks in cascade		
2. Moving a summing point after a block		
3. Moving a summing point ahead of a block		
4. Moving a take off point after a block		
5. Moving a take off point ahead of a block		
6. Eliminating a feedback loop		

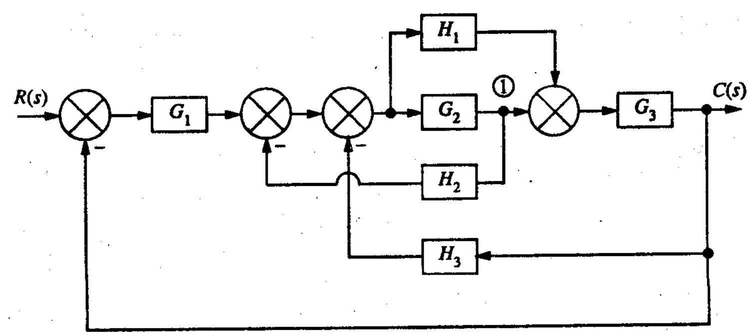
7. Combining Blocks in Parallel

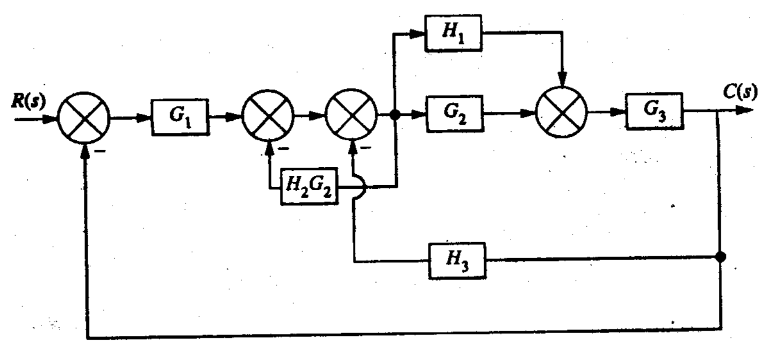


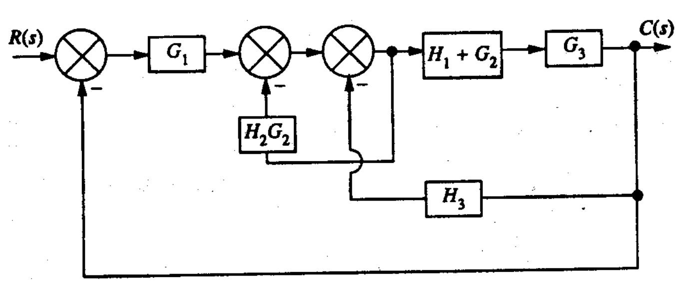
8. Moving summing point :

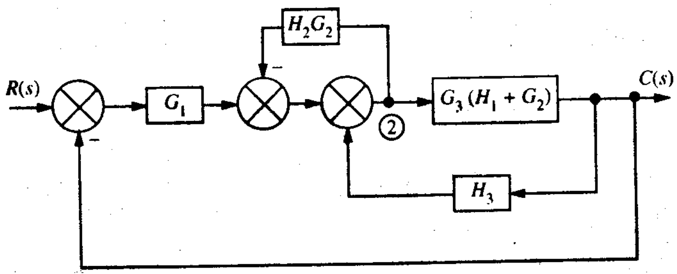


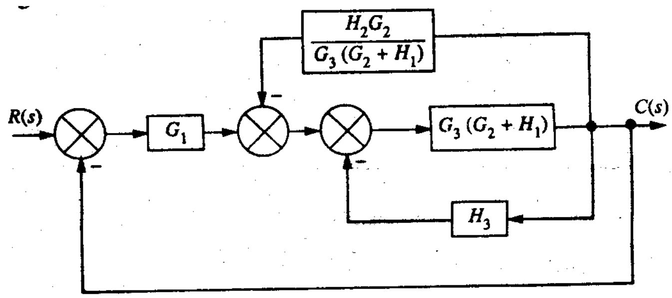
Example: Simplify the block diagram shown in Figure below.

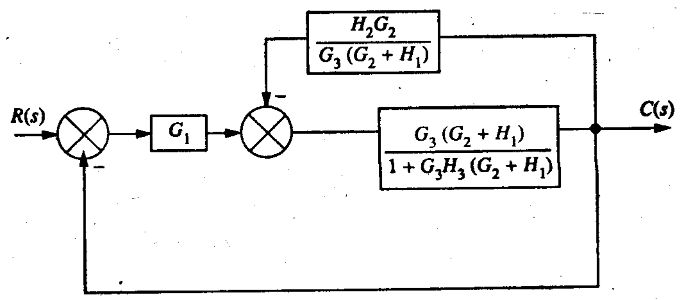


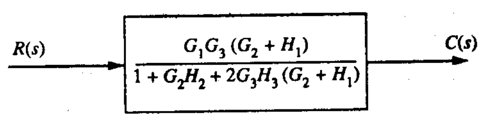




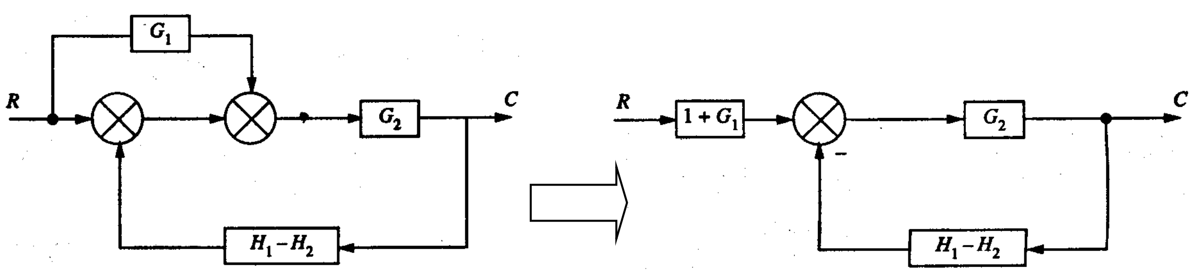


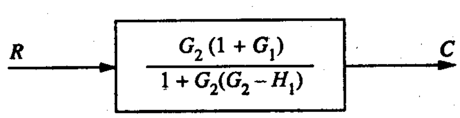






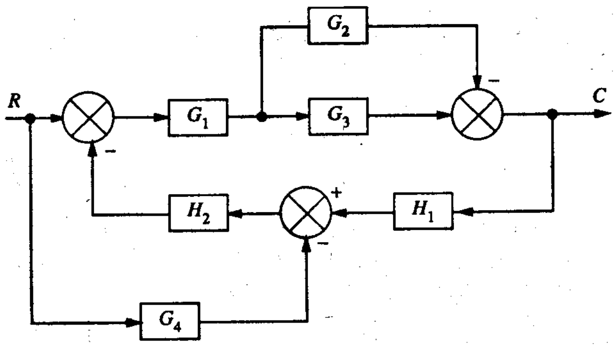
Example: Obtain the transfer function C/R of the block diagram shown in Figure below.



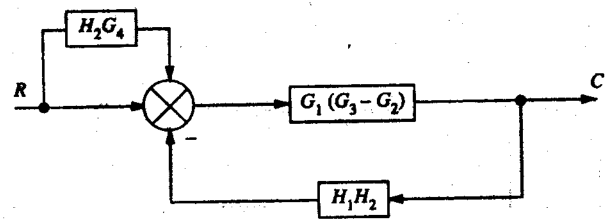


[Ans]

Example: Derive the transfer function of the system shown below.



(a)

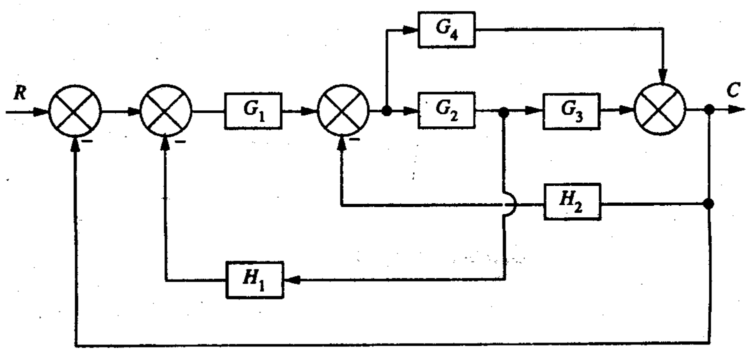


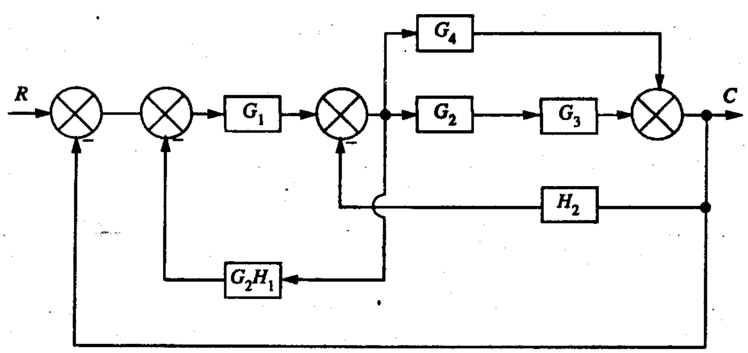
(b)

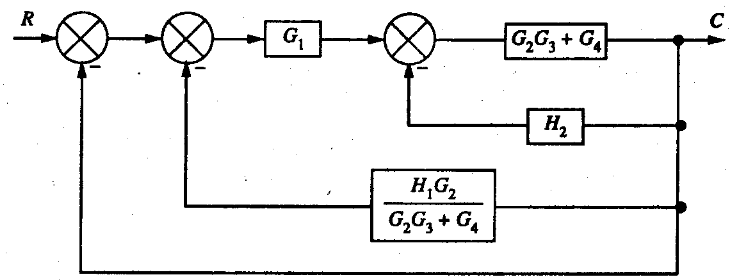
$$R \rightarrow \boxed{\frac{G_1(G_3 - G_2)(1 + H_2G_4)}{1 + G_1H_1H_2(G_3 - G_2)}} \rightarrow C$$

[Answer]

Example: Derive the transfer function of the system shown below.

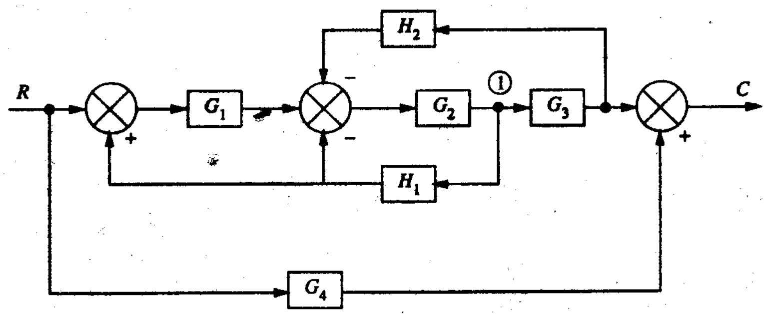


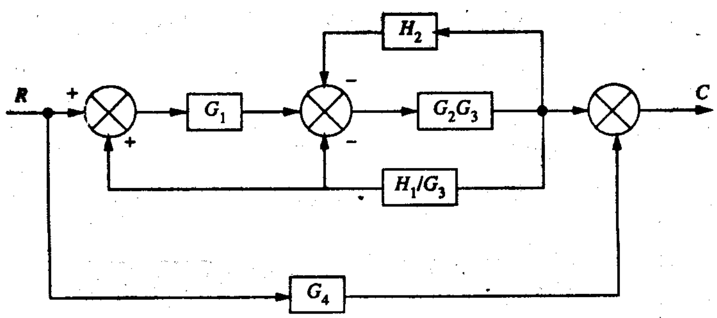




$$\frac{C}{R} = \frac{G_1 [G_2 G_3 + G_4)}{1 + H_2 [G_2 G_3 + G_4] + G_1 G_2 H_1 + G_1 [G_2 G_3 + G_4]}$$

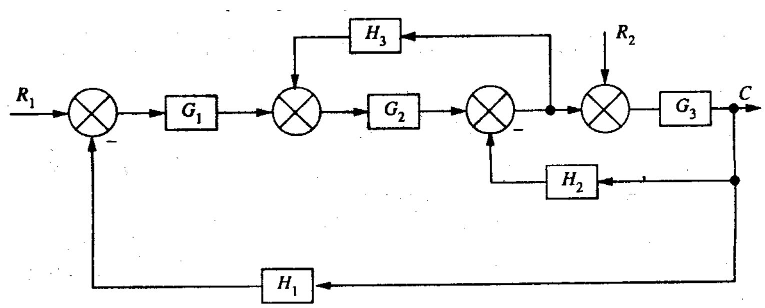
Example: Find the transfer function of the following system.



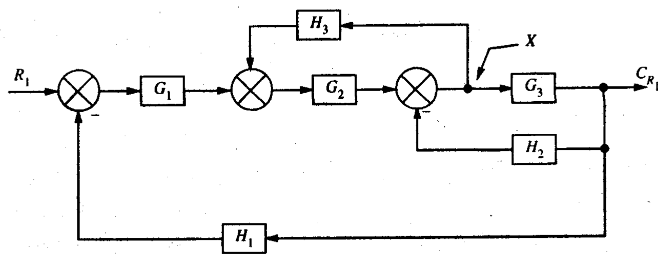


$$\frac{C}{R} = G_4 + \frac{G_1 G_2 G_3}{1 + H_2 G_2 G_3 + G_2 H_1 - H_1 G_1 G_2}$$ {Answer}

Example: Find the output of the system shown below.

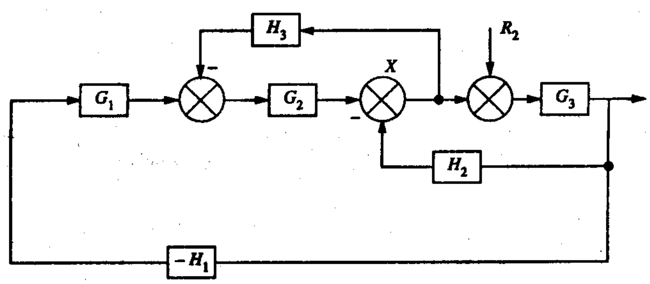


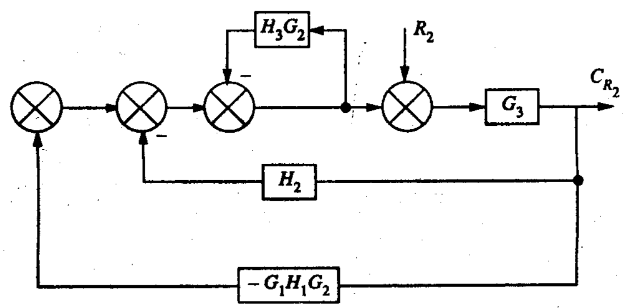
For Input R_1:



$$C_{R_1} = \left[\frac{G_1 G_2 G_3}{1 + G_3 H_2 + H_3 G_2 + G_1 G_2 G_3 H_1}\right] R_1 \qquad \text{...} (1)$$

For input R_2:





$$C_{R_2} = \left[\frac{G_3 [1 + G_2 H_3]}{1 + G_2 H_3 + G_3 \,[G_1 G_2 H_1 + H_2]}\right] R_2 \qquad \text{...} (2)$$

$$C = C_{R_1} + C_{R_2}$$

$$C = \left[\frac{G_1 G_2 G_3}{1 + G_3 H_2 + G_2 H_3 + G_1 G_2 G_3 H_1}\right] R_1 + \left[\frac{G_3 [1 + G_2 H_3]}{1 + G_2 H_3 + G_3 [G_1 G_2 H_1 + H_2]}\right] R_2$$ {Answer}

SIGNAL FLOW GRAPH

SFG is a diagram that represents a set of simultaneous linear algebraic equations which describe a system. Let us consider an equation, $Y = aX$. It may be represented graphically as,

where ‘a’ is called ***transmittance*** or transmission function.

Definitions in SFG

Node – A system variable, the value of which equals the sum of all incoming signals at the node.

Branch – A directed line segment joining two nodes.

Input/ Output node – node having only one outgoing/ incoming branch.

Path – A traversal of connected branches in the direction of branch arrows.

Forward path – A path from input to output node.

Loop – A closed path that originates and terminates on the same node.

Self-loop – A loop containing one branch.

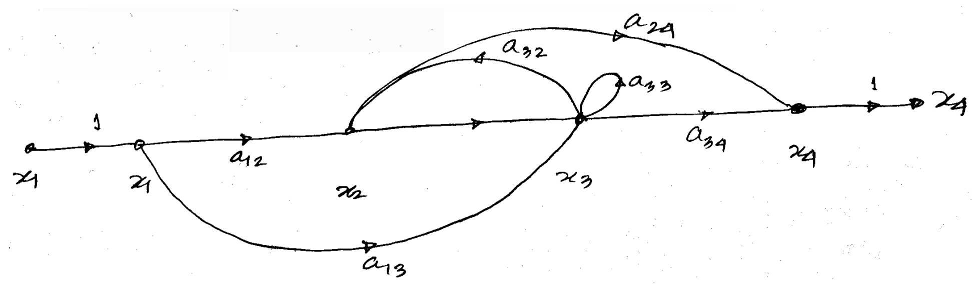
Non-touching loops – Loops which do not have a common node.

Gain – Transmittance of a branch.

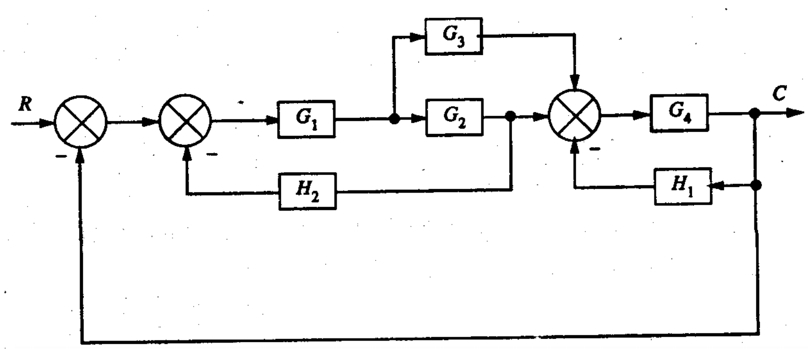
Construction of SFGs

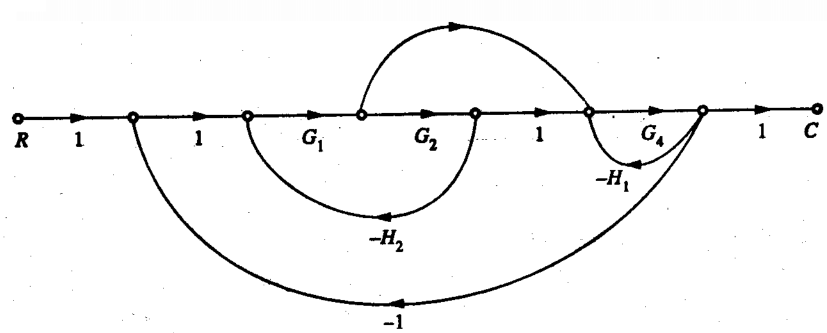
The SFG of a system can be constructed from the describing equations:

$$x_2 = a_{12}x_1 + a_{32}x_3$$
$$x_3 = a_{13}x_1 + a_{23}x_2 + a_{33}x_3$$
$$x_4 = a_{24}x_4 + a_{34}x_3$$



SFG from Block Diagram





Each variable in the block diagram becomes a node, and each block becomes a branch.

Mason’s Gain Formulae

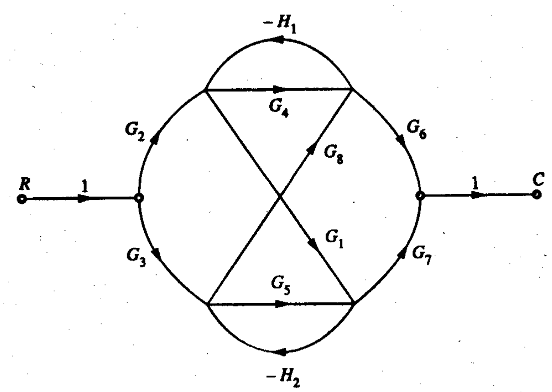
It is possible to write the overall transfer function of a system through inspection of SFG using Mason's gain formulae given by, $T = (\sum_i P_i \Delta_i)/\Delta$.

where T = overall gain of the system, P_i = path gain of *ith* forward path, Δ = determinant of SFG, Δ_i = value of Δ for that part of the graph not touching the *ith* forward path.

$\Delta = 1 - \sum_j P_{j1} + \sum_j P_{j2} - \sum_j P_{j3} + \cdots$ = 1 – [sum of loop gain of all individual loops] + [sum of all gain-products of two non-touching loops] – [sum of all gain-products of three non-touching loops] + ……;

P_{jk} = *j*th product of *k* non-touching loops .

Example



1. There are 6 forward paths with path gains

$$P_1 = G_2 G_4 G_6$$
$$P_2 = G_3 G_5 G_7$$
$$P_3 = G_2 G_1 \cdot G_7$$
$$P_4 = G_3 G_8 G_6$$
$$P_5 = -G_2 G_1 \cdot H_2 G_8 \cdot G_6$$
$$P_6 = -G_3 G_8 H_1 G_1 G_7$$

2. There are three individual loops with loop gains

$$P_{11} = -H_1 G_4$$
$$P_{21} = -H_2 G_5$$
$$P_{31} = G_1 H_2 G_8 H_1$$

3. There is only one combination of two non-touching loops

$$P_{12} = H_1 H_2 G_4 G_5$$

4. There are no combinations of more than two non-touching loops.

5. Hence,

$$\Delta = 1 - [-H_1 G_4 - H_2 G_5 + G_1 H_2 G_8 H_1] + [H_1 H_2 G_4 G_5]$$
$$= 1 - G_1 H_2 G_8 H_1 + H_2 G_5 - G_1 H_2 G_8 H_1 + H_1 H_2 G_4 G_5$$

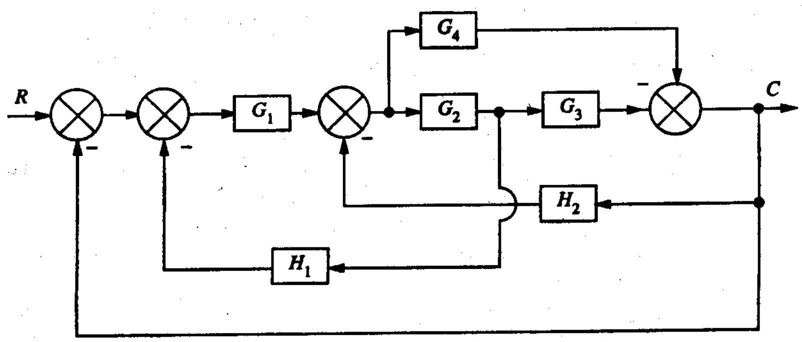
$\Delta_1 = 1 - (-H_2 G_5) = 1 + H_2 G_5$; $\Delta_2 = 1 - (-H_1 G_4) = 1 + H_1 G_4$;

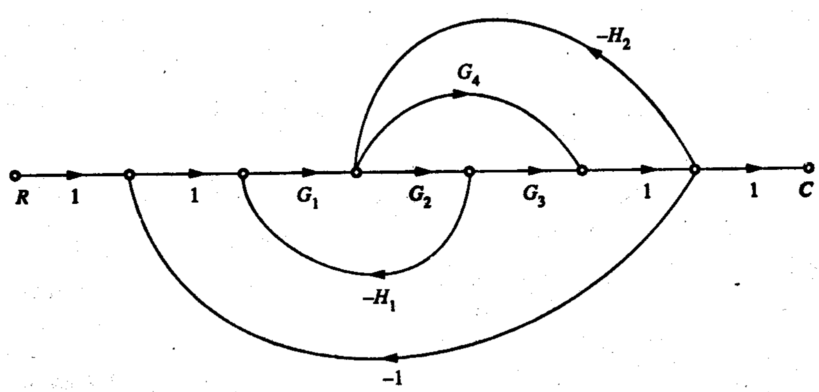
$\Delta_3 = \Delta_4 = \Delta_5 = \Delta_6 = 1$

Thus, $T = \dfrac{P_1\Delta_1 + P_2\Delta_2 + P_3\Delta_3 + P_4\Delta_4 + P_5\Delta_5 + P_6\Delta_6}{\Delta}$, where P_1, Δ_1, Δ etc. are derived before.

Example

Draw the SFG and determine C/ R for the block diagram shown in Figure below.





$$\frac{C}{R} = \frac{G_1 G_2 G_3 + G_1 G_4}{1 + G_1 G_2 G_3 + G_1 G_2 H_1 + G_2 G_3 H_2 + G_1 G_4 + G_4 H_2}$$ {Answer}

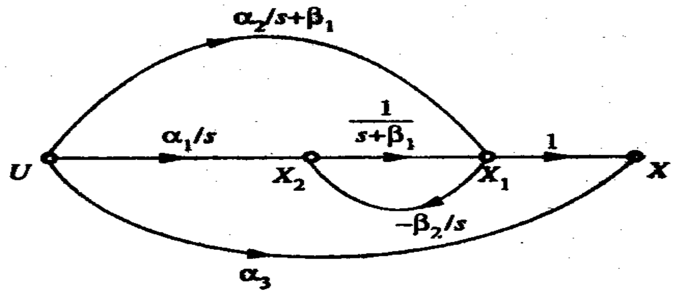
Example

For the system represented by the following equations, find the transfer function X(s)/U(s) by SFG technique.

$$
\begin{aligned}
x &= x_1 + \alpha_3 u \\
\dot{x}_1 &= -\beta_1 x_1 + x_2 + \alpha_2 u \\
\dot{x}_2 &= -\beta_2 x_1 + \alpha_1 u
\end{aligned}
$$

⟹ We need to Laplace transform the given sets of equations in order to represent differentiated variables.

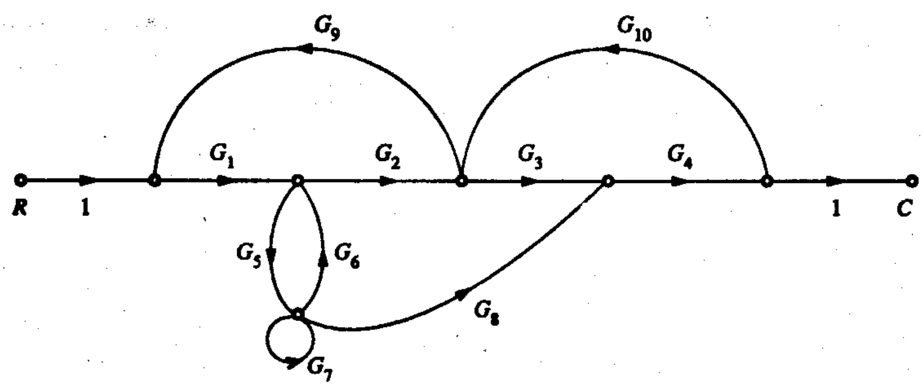
$$
\begin{aligned}
X &= X_1 + \alpha_3 U \\
X_1 &= \frac{1}{s+\beta_1} X_2 + \frac{\alpha_2}{s+\beta_1} u \\
X_2 &= -\frac{\beta_2}{s} X_1 + \frac{\alpha_1}{s} u
\end{aligned}
$$



$$\frac{X(s)}{U(s)} = \frac{\alpha_1 + \alpha_2 s + \alpha_3 \cdot [s^2 + \beta_1 s + \beta_2]}{s^2 + \beta_1 s + \beta_2}$$ {Answer}

Example

Using Mason's gain formulae find C/R of the SFG shown in Figure below.



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

$$= \frac{G_1 G_2 G_3 G_4 (1 - G_7) + G_1 G_5 G_8 G_4}{1 - [G_1 G_2 G_9 + G_3 G_4 G_{10} + G_1 G_5 G_8 G_4 G_{10} G_9 + G_5 G_6 + G_7] + [G_1 G_2 G_9 G_7 + G_3 G_4 G_{10} G_5 G_6 + G_3 G_4 G_{10} G_7]}$$

SEE1203 – CONTROL SYSTEMS

UNIT II

TIME RESPONSE ANALYSIS OF CONTROL SYSTEM

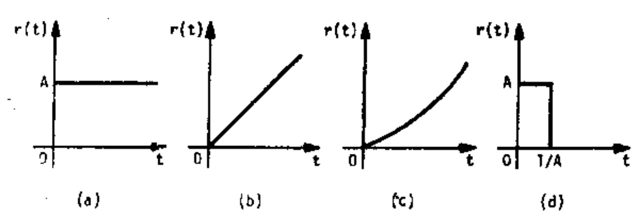
In time-domain analysis the response of a dynamic system to an input is expressed as a function of time. It is possible to compute the time response of a system if the nature of input and the mathematical model of the system are known.

Usually, the input signals to control systems are not known fully ahead of time. In a radar tracking system, the position and the speed of the target to be tracked may vary in a random fashion. It is therefore difficult to express the actual input signals mathematically by simple equations. The characteristics of actual input signals are a sudden shock, a sudden change, a constant velocity, and constant acceleration. The dynamic behavior of a system is therefore judged and compared under application of standard test signals – an impulse, a step, a constant velocity, and constant acceleration. Another standard signal of great importance is a sinusoidal signal.

The time response of any system has two components: transient response and the steady-state response. Transient response is dependent upon the system poles only and not on the type of input. It is therefore sufficient to analyze the transient response using a step input. The steady-state response depends on system dynamics and the input quantity. It is then examined using different test signals by final value theorem.

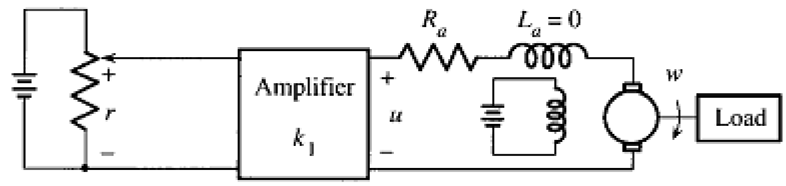
STANDARD TEST SIGNALS

a) Step signal:
b) Ramp signal:
c) Parabolic signal:
d) Impulse signal:



TIME-RESPONSE OF FIRST-ORDER SYSTEMS

Let us consider the armature-controlled dc motor driving a load, such as a video tape. The objective is to drive the tape at constant speed. Note that it is an open-loop system.



; If ,

;

is the steady-state final speed. If the desired speed is , choosing the motor will eventually reach the desired speed.

We are interested not only in final speed, but also in the speed of response. Here, is the time constant of motor which is responsible for the speed of response.

The time response is plotted in the Figure in next page. A plot of is shown, from where it is seen that, for the value of is less than 1% of its original value. Therefore, the speed of the motor will reach and stay within 1% of its final speed at 5 time constants.

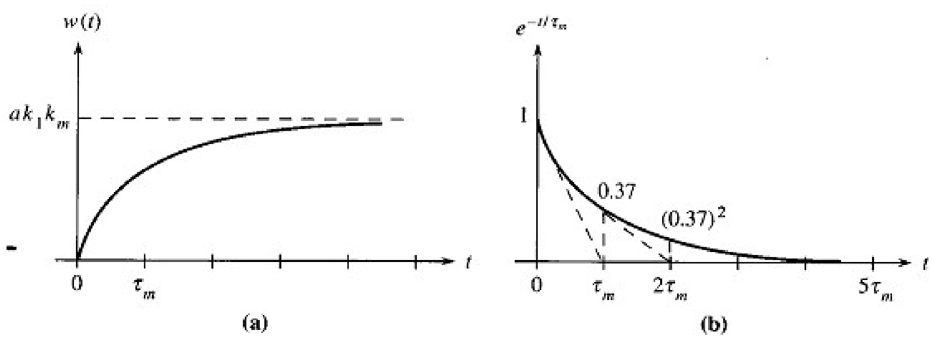
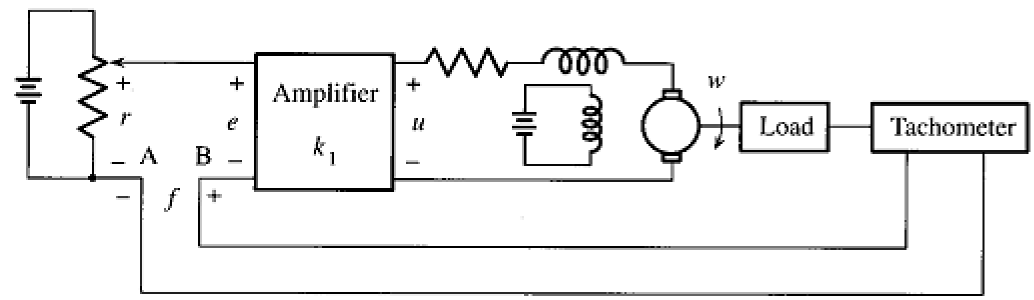
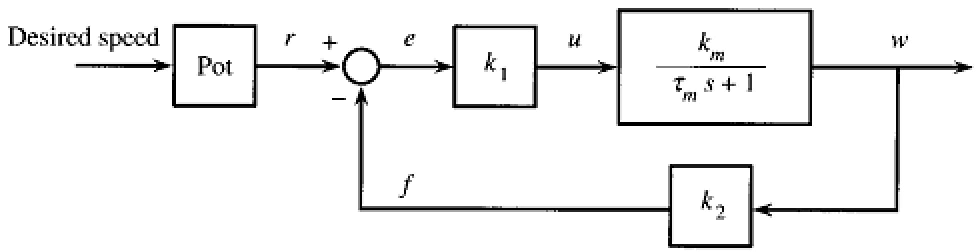


Figure: Time responses

Let us now consider the closed-loop system shown below.



(a)



(b)

Here,

where, and .

If , the response would be, .

If a is properly chosen, the tape can reach a desired speed. It will reach the desired speed in 5seconds. Here, . Thus, we can control the speed of response in feedback system.

Although the time-constant is reduced by a factor , in the feedback system, the motor gain constant is also reduced by the same factor. In order to compensate for this loss of gain, the applied reference voltage must be increased by the same factor.

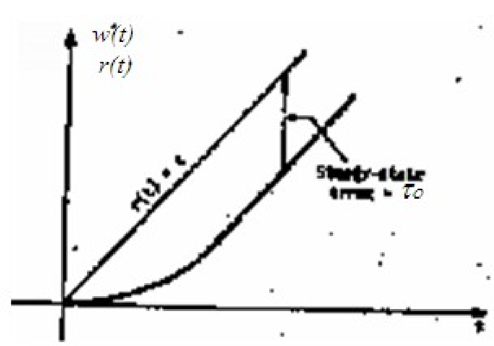
Ramp response of first-order system

Let, for simplicity. Then, . Also, let, .

Then, ;

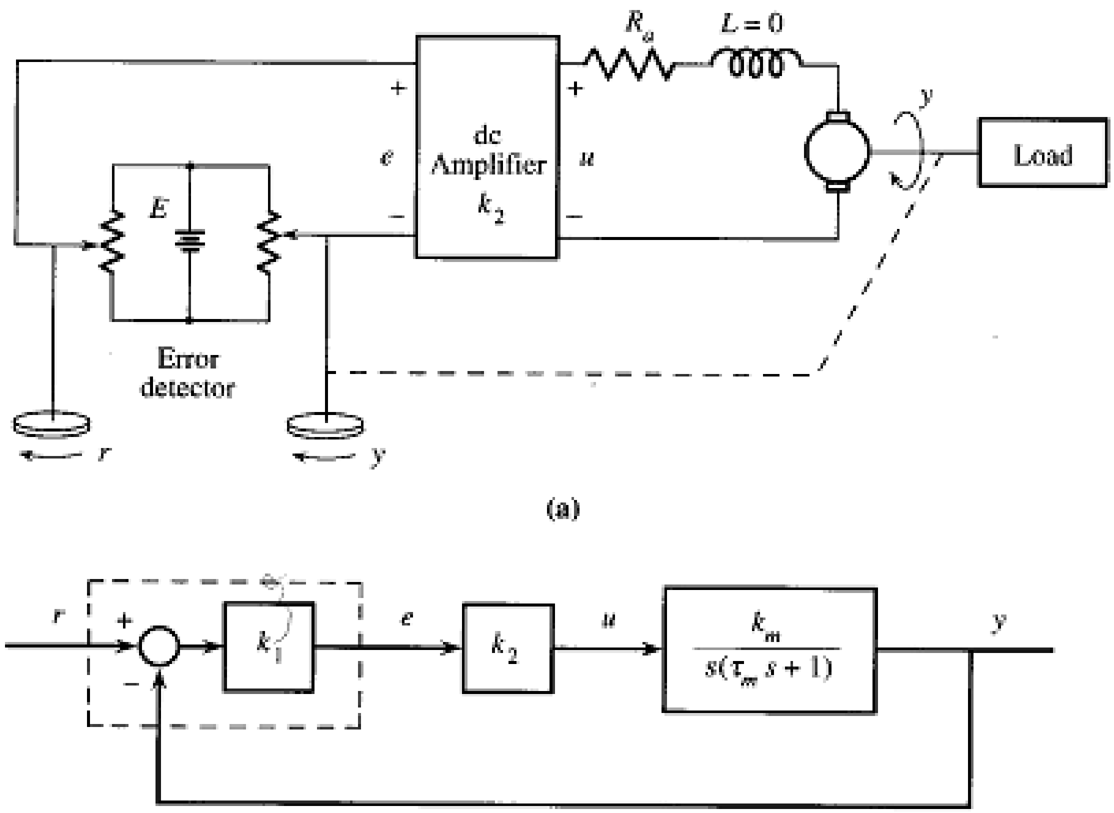
The error signal is,

Or,



Thus, the first-order system will track the unit ramp input with a steady-state error , which is equal to the time-constant of the system.

TIME-RESPONSE OF SECOND-ORDER SYSTEMS



(a)

(b)

Consider the antenna position control system. Its transfer function from *r* to *y* is,

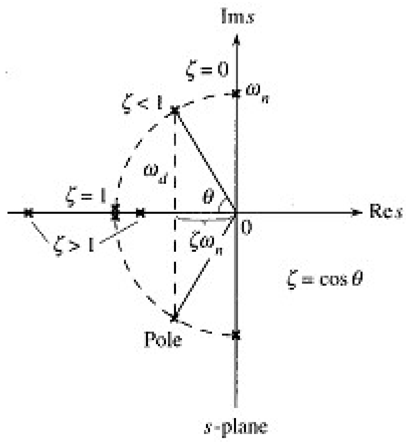
where, we define, and . The constant is called the *damping ratio* and is called the *natural frequency*. The system above is in fact a standard second order system.

The transfer function has two poles and no zero. Its poles are,

.

Here, is called the *damping factor*, is called *damped or actual frequency*.

The location of poles for different are plotted in Figure below. For, the two poles are purely imaginary. If , the two poles are complex conjugate. All possible cases are described in a table shown below.



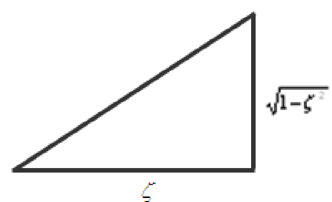
Damping Ratio	Poles	Remark
$\zeta = 0$	Pure imaginary	Undamped
$0 < \zeta < 1$	Complex conjugate	Underdamped
$\zeta = 1$	Repeated real poles	Critically damped
$\zeta > 1$	Two distinct real poles	Overdamped

Unit step response of second-order systems

Suppose, ;

Or,

Performing inverse Laplace transform,

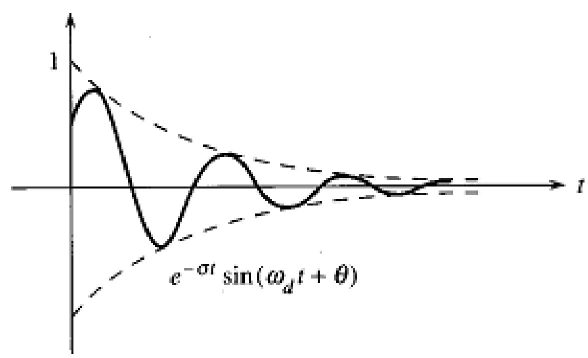


or,

or, ,where, and

or,

The plot of is shown in Figure.



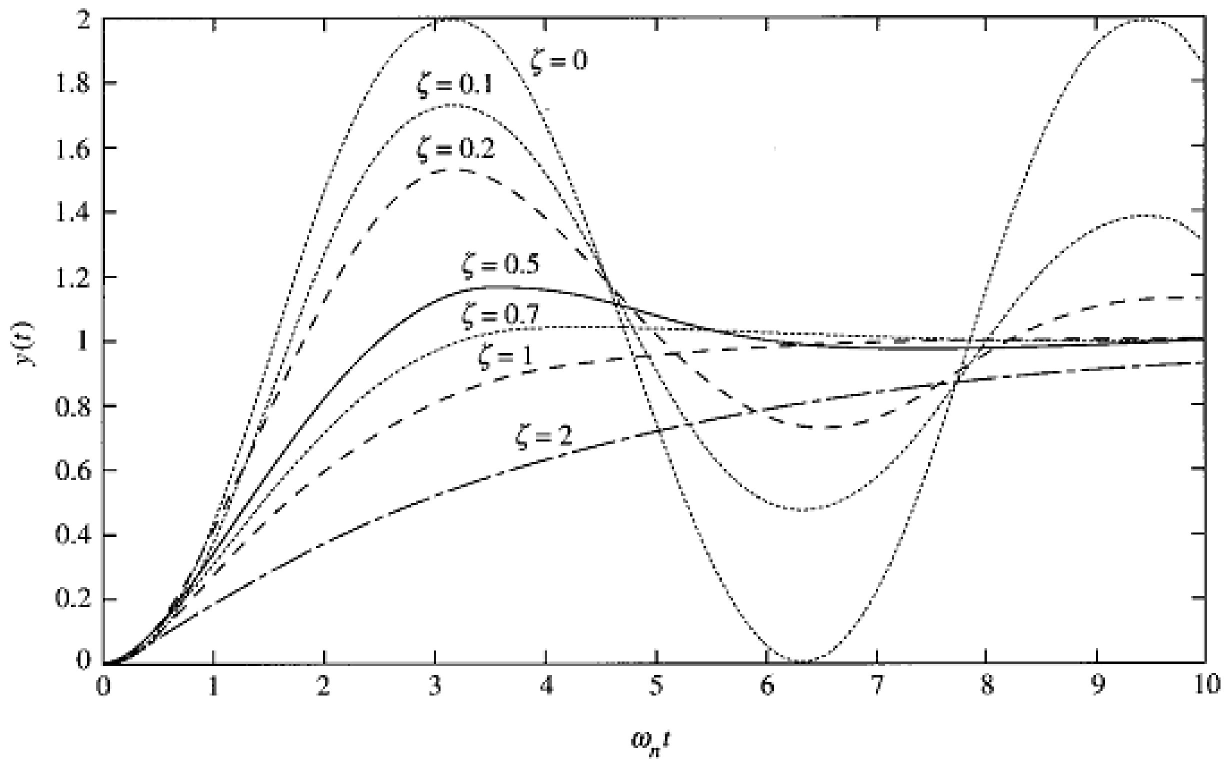
The steady-state response is,

Thus, the system has zero steady-state error.

The pole of dictates the response,

.

The response for different is shown in Figure below.

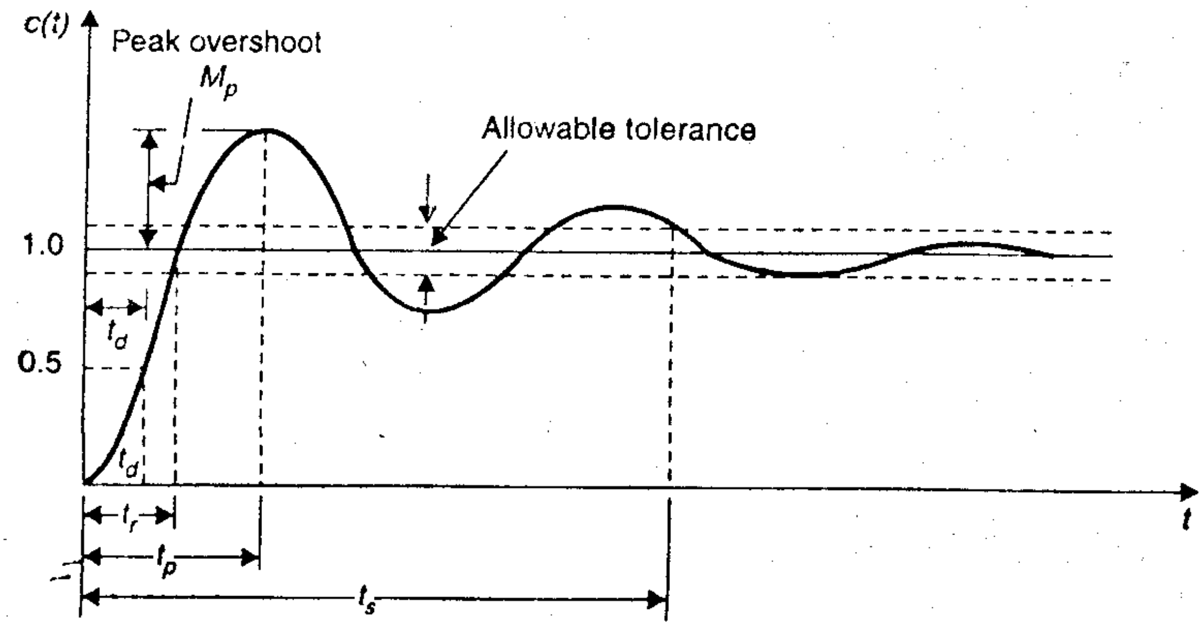


TIME DOMAIN SPECIALIZATION

Control systems are generally designed with damping less than one, i.e., oscillatory step response. Higher order control systems usually have a pair of complex conjugate poles with damping less than unity that dominate over the other poles. Therefore the time response of second- and higher-order control systems to a step input is generally of damped oscillatory nature as shown in Figure next (next page).

In specifying the transient-response characteristics of a control system to a unit step input, we usually specify the following:

1. Delay time,
2. Rise time,
3. Peak time,
4. Peak overshoot,
5. Settling time,
6. Steady-state error,



1. ***Delay time, :*** It is the time required for the response to reach 50% of the final value in first attempt.
2. ***Rise time, :*** It is the time required for the response to rise from 0 to 100% of the final value for the underdamped system.
3. ***Peak time, :*** It is the time required for the response to reach the peak of time response or the peak overshoot.
4. ***Settling time, :*** It is the time required for the response to reach and stay within a specified tolerance band (2% or 5%) of its final value.
5. ***Peak overshoot, :*** It is the normalized difference between the time response peak and the steady output and is defined as,

6. ***Steady-state error, :*** It indicates the error between the actual output and desired output as 't' tends to infinity.

.

Let us now obtain the expressions for the rise time, peak time, peak overshoot, and settling time for the second order system.

1. ***Rise time, :*** Put at ,,;.
2. ***Peak time, :*** Put and solve for ; .

,

Peak overshoot occurs at *k = 1*. .

3. ***Settling time, :*** For 2% tolerance band, , .

4. ***Steady-state error, :*** It is found previously that steady-state error for step input is zero.

Let us now consider ramp input, .

Then,

.

Therefore, the steady-state error due to ramp input is.

STEADY STATE ERRORS

The steady-state performance of a stable control system is generally judged by its steady-state error to step, ramp and parabolic inputs. For a unity feedback system,

,.

It is seen that steady-state error depends upon the input and the forward transfer function. The steady-state errors for different inputs are derived as follows:

ALGEBRIC CRITERIA

1. For unit-step input:

; is called position error constant.

2. For unit-ramp input:

; is called velocity error constant.

3. For unit-parabolic input:

; is called acceleration error const.

Types of Feedback Control System

The open-loop transfer function of a system can be written as,

If n = 0, the system is called type-0 system, if n = 1, the system is called type-1 system, if n = 2, the system is called type-2 system, etc. Steady-state errors for various inputs and system types are tabulated below.

Type of input	Steady-state error		
	Type-0 system	Type-1 system	Type-2 system
Unit-step	$1/(1+K_p)$	0	0
Unit-ramp	∞	$1/K_v$	0
Unit-parabolic	∞	∞	$1/K_a$
	$K_p = \lim_{s\to 0} G(s)$	$K_v = \lim_{s\to 0} sG(s)$	$K_a = \lim_{s\to 0} s^2G(s)$

ERROR CONSTANTS

The error constants for non-unity feedback systems may be obtained by replacing G(s) by G(s)H(s). Systems of type higher than 2 are not employed due to two reasons:

1. The system is difficult to stabilize.
2. The dynamic errors for such systems tend to be larger than those types-0, -1 and -2.

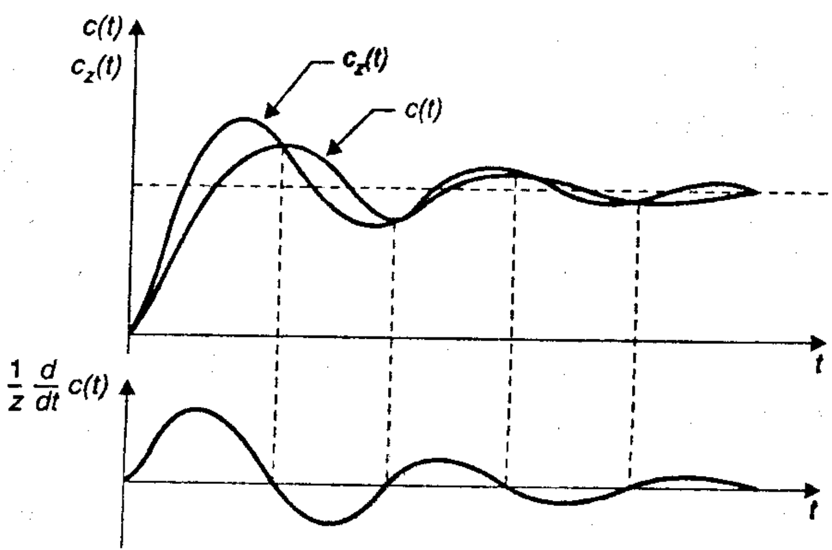
Effect of Adding a Zero to a System

Let a zero at s = -z be added to a second order system. Then we have,

.

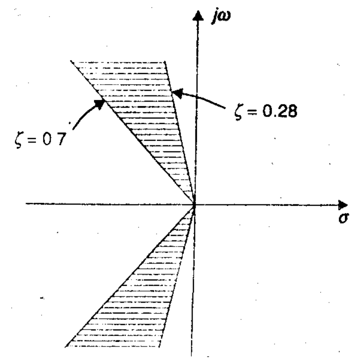
The multiplication term is adjusted to make the steady-state gain of the system unity. This gives $c_{ss} = 1$ when the input is unit step. Let $c_z(t)$ be the response of the system given by the above equation and $c(t)$ is the response without adding the pole. Manipulation of the above equation gives,

The effect of added derivative term is to produce a pronounced early peak to the system response which will be clear from the figure in the next page. Closer the zero to origin, the more pronounce the peaking phenomenon. Due to this fact, *the zeros on the real axis near the origin are generally avoided in design*. However, in a sluggish system the artful introduction of a zero at the proper position can improve the transient response. We can see from equation (03) that as z increases, i.e., the zero moves further into the left half of the s-plane, its effect becomes less pronounced.



Design Specifications of Second-order Systems

A control system is generally required to meet three time response specifications: steady-state accuracy, damping factor ζ (or peak overshoot, M_p) and settling time t_s. Steady-state accuracy requirement is met by suitable choice of *Kp, Kv, or Ka* depending on the type of the system. For most control systems ζ in the range of 0.7 – 0.28 (or peak overshoot of 5 – 40%) is considered acceptable. For this range of ζ, the closed-loop pole locations are restricted to the shaded region of the s-plane as shown in Figure.



For the antenna position control system,;;;. Here, is only the adjustable parameter. If we increase, will increase and thus settling time will decrease. At the same time, will decrease, this indicates the increase in peak overshoot. Thus by merely increasing gain, we cannot improve both transient and steady-state error specifications. We need to add additional components to the system. These are called compensators. It will allow improvement of both transient and steady-state specifications.

CONCEPTS OF STABILITY

BIBO stability: A system is said to be BIBO stable if for any bounded input, its output is also bounded. • Absolute stability: Stable /Unstable • Relative stability: Degree of stability (i.e. how far from instability) • A stable linear system described by a T.F. is such that all its poles have negative real parts

SEE1203 – CONTROL SYSTEMS

UNIT 3

THE CONCEPT OF STABILITY AND ROOT LOCUS TECHNIQUE

THE CONCEPT OF STABILITY

When a system is unstable, the output of the system may be infinite even though the input to the system was finite. This causes a number of practical problems. For instance, a robot arm controller that is unstable may cause the robot to move dangerously. Also, systems that are unstable often incur a certain amount of physical damage, which can become costly. Nonetheless, many systems are inherently unstable - a fighter jet, for instance, or a rocket at liftoff, are examples of naturally unstable systems. Although we can design controllers that stabilize the system, it is first important to understand what stability is, how it is determined, and why it matters.

The chapters in this section are heavily mathematical, and many require a background in linear differential equations. Readers without a strong mathematical background might want to review the necessary chapters in the Calculus and Ordinary Differential Equations books (or equivalent) before reading this material.

For most of this chapter we will be assuming that the system is linear, and can be represented either by a set of transfer functions or in state space. Linear systems have an associated characteristic polynomial, and this polynomial tells us a great deal about the stability of the system. Negativeness of any coefficient of a characteristic polynomial indicates that the system is either unstable or at most marginally stable. If any coefficient is zero/negative then we can say that the system is unstable. It is important to note, though, that even if all of the coefficients of the characteristic polynomial are positive the system may still be unstable. We will look into this in more detail below.

A system is defined to be **BIBO Stable** if every bounded input to the system results in a bounded output over the time interval $[t_0, \infty)$. This must hold for all initial times t_o. So long as we don't input infinity to our system, we won't get infinity output.

A system is defined to be **uniformly BIBO Stable** if there exists a positive constant *k* that is independent of t_0 such that for all t_0 the following conditions:

$$\|u(t)\| \leq 1$$

$$t \geq t_0$$

implies that

$$\|y(t)\| \leq k$$

There are a number of different types of stability, and keywords that are used with the topic of stability. Some of the important words that we are going to be discussing in this chapter, and the next few chapters are: **BIBO Stable**, **Marginally Stable**, **Conditionally Stable**, **Uniformly Stable**, **Asymptoticly Stable**, and **Unstable**. All of these words mean slightly different things.

Consider the system:

$$h(t) = \frac{2}{t}$$

We can apply our test, selecting an arbitrarily large finite constant M, and an arbitrary input x such that -M < x < M.

As M approaches infinity (but does not reach infinity), we can show that:

$$y_{-M} = \lim_{M \to \infty} \frac{2}{-M} = 0^-$$

And:

$$y_M = \lim_{M \to \infty} \frac{2}{M} = 0^+$$

So now, we can write out our inequality:

$$y_{-M} \le y_x \le y_M$$

$$0^- \le x < 0^+$$

And this inequality should be satisfied for all possible values of x. However, we can see that when x is zero, we have the following:

$$y_x = \lim_{x \to 0} \frac{2}{x} = \infty$$

Which means that x is between -M and M, but the value y_x is not between y_{-M} and y_M. Therefore, this system is not stable.

Poles and StabilitY

When the poles of the closed-loop transfer function of a given system are located in the right-half of the S-plane (RHP), the system becomes unstable. When the poles of the system are located in the left-half plane (LHP) and the system is not improper, the system is shown to be stable. A number of tests deal with this particular facet of stability: The**Routh-Hurwitz Criteria**, the **Root-Locus**, and the **Nyquist Stability Criteria** all test whether there are poles of the transfer function in the RHP. We will learn about all these tests in the upcoming chapters.

If the system is a multivariable, or a MIMO system, then the system is stable if and only if *every pole of every transfer function* in the transfer function matrix has a negative real part and every transfer function in the transfer function matrix is not improper. For these systems, it is possible to use the Routh-Hurwitz, Root Locus, and Nyquist methods described later, but these methods must be performed once for each individual transfer function in the transfer function matrix.

Let us remember our generalized feedback-loop transfer function, with a gain element of K, a forward path Gp(s), and a feedback of Gb(s). We write the transfer function for this system as:

$$H_{cl}(s) = \frac{KGp(s)}{1 + H_{ol}(s)}$$

Where H_{cl} is the closed-loop transfer function, and H_{ol} is the open-loop transfer function. Again, we define the open-loop transfer function as the product of the forward path and the feedback elements, as such:

$H_{ol}(s) = KGp(s)Gb(s)$ <---Note this definition now contradicts the updated definition in the "Feedback" section.

Now, we can define F(s) to be the **characteristic equation**. F(s) is simply the denominator of the closed-loop transfer function, and can be defined as such:

[Characteristic Equation]

$$F(s) = 1 + H_{ol} = D(s)$$

We can say conclusively that the roots of the characteristic equation are the poles of the transfer function. Now, we know a few simple facts:

1. The locations of the poles of the closed-loop transfer function determine if the system is stable or not
2. The zeros of the characteristic equation are the poles of the closed-loop transfer function.
3. The characteristic equation is always a simpler equation than the closed-loop transfer function.

These functions combined show us that we can focus our attention on the characteristic equation, and find the roots of that equation.

State-Space and Stability

As we have discussed earlier, the system is stable if the eigenvalues of the system matrix A have negative real parts. However, there are other stability issues that we can analyze, such as whether a system is *uniformly stable*, *asymptotically stable*, or otherwise. We will discuss all these topics in a later chapter.

Marginal Stability

When the poles of the system in the complex S-Domain exist on the complex frequency axis (the vertical axis), or when the eigenvalues of the system matrix are imaginary (no real part), the system exhibits oscillatory characteristics, and is said to be marginally stable. A marginally stable system may become unstable under certain circumstances, and may be perfectly stable under other circumstances.

ROUTH STABILITY CRITERION:

The Routh approximation method which has been suggested for the reduction of stable discrete time linear systems to guarantee stable models, uses the bilinear transformation. A stability theorem in the *z*-plane is presented which is shown to be an equivalent of the Routh criterion. An efficent method that avoids the bilinear transformation is presented by which the Routhdisrete models are derived directly in the *z*-plane.

In control system theory, the **Routh–Hurwitz stability criterion** is a mathematical test that is a necessary and sufficient condition for the stability of a linear time invariant (LTI)control system. The Routh test is an efficient recursive algorithm that English mathematician Edward John Routh proposed in 1876 to determine whether all the roots of thecharacteristic polynomial of a linear system have negative real parts.[1] German mathematician Adolf Hurwitz independently proposed in 1895 to arrange the coefficients of the polynomial into a square matrix, called the Hurwitz matrix, and showed that the polynomial is stable if and only if the sequence of determinants of its principal submatrices are all positive.[2] The two procedures are equivalent, with the Routh test providing a more efficient way to compute the Hurwitz determinants than computing them directly. A polynomial satisfying the Routh–Hurwitz criterion is called a Hurwitz polynomial.

The importance of the criterion is that the roots ***p*** of the characteristic equation of a linear system with negative real parts represent solutions e^{pt} of the system that are stable (bounded). Thus the criterion provides a way to determine if the equations of motion of a linear system have only stable solutions, without solving the system directly. For discrete systems, the corresponding stability test can be handled by the Schur–Cohn criterion, the Jury test and the Bistritz test. With the advent of computers, the criterion has become less widely used, as an alternative is to solve the polynomial numerically, obtaining approximations to the roots directly.

The Routh test can be derived through the use of the Euclidean algorithm and Sturm's theorem in evaluating Cauchy indices. Hurwitz derived his conditions differently.

Using Euclid's algorithm

The criterion is related to Routh–Hurwitz theorem. Indeed, from the statement of that theorem, we have $p - q = w(+\infty) - w(-\infty)$ where:

- *p* is the number of roots of the polynomial $f(z)$ with negative Real Part;
- *q* is the number of roots of the polynomial $f(z)$ with positive Real Part (let us remind ourselves that *f* is supposed to have no roots lying on the imaginary line);
- $w(x)$ is the number of variations of the generalized Sturm chain obtained from $P_0(y)$ and $P_1(y)$ (by successive Euclidean divisions) where $f(iy) = P_0(y) + iP_1(y)$ for a real *y*.

By the fundamental theorem of algebra, each polynomial of degree *n* must have *n* roots in the complex plane (i.e., for an *f* with no roots on the imaginary line, $p + q = n$). Thus, we have the condition that f is a (Hurwitz) stable polynomial if and only if $p - q = n$ (the proof is given below). Using the Routh–Hurwitz theorem, we can replace the condition on p and q by a condition on the generalized Sturm chain, which will give in turn a condition on the coefficients of f.

Using matrices

Let $f(z)$ be a complex polynomial. The process is as follows:

1. Compute the polynomials $P_0(y)$ and $P_1(y)$ such that $f(iy) = P_0(y) + iP_1(y)$ where y is a real number.
2. Compute the Sylvester matrix associated to $P_0(y)$ and $P_1(y)$.
3. Rearrange each row in such a way that an odd row and the following one have the same number of leading zeros.
4. Compute each principal minor of that matrix.
5. If at least one of the minors is negative (or zero), then the polynomial f is not stable.

Example

- Let $f(z) = az^2 + bz + c$ (for the sake of simplicity we take real coefficients) where $c \neq 0$ (to avoid a root in zero so that we can use the Routh–Hurwitz theorem). First, we have to calculate the real polynomials $P_0(y)$ and $P_1(y)$:

$$f(iy) = -ay^2 + iby + c = P_0(y) + iP_1(y) = -ay^2 + c + i(by).$$

Next, we divide those polynomials to obtain the generalized Sturm chain:

- $P_0(y) = ((-a/b)y)P_1(y) + c,$ yields $P_2(y) = -c,$
- $P_1(y) = ((-b/c)y)P_2(y),$ yields $P_3(y) = 0$ and the Euclidean division stops.

Notice that we had to suppose b different from zero in the first division. The generalized Sturm chain is in this case $(P_0(y), P_1(y), P_2(y)) = (c - ay^2, by, -c)$. Putting $y = +\infty$, the sign of $c - ay^2$ is the opposite sign of a and the sign of by is the sign of b. When we put $y = -\infty$, the sign of the first element of the chain is again the opposite sign of a and the sign of by is the opposite sign of b. Finally, $-c$ has always the opposite sign of c.

Suppose now that f is Hurwitz-stable. This means that $w(+\infty) - w(-\infty) = 2$ (the degree of f). By the properties of the function w, this is the same as $w(+\infty) = 2$ and $w(-\infty) = 0$. Thus, a, b and c must have the same sign. We have thus found the necessary condition of stability for polynomials of degree 2.

Routh–Hurwitz criterion for second, third, and fourth-order polynomials[edit]

In the following, we assume the coefficient of the highest order (e.g. a_2 in a second order polynomial) to be positive. If necessary, this can always be achieved by multiplication of the polynomial with -1.

- For a second-order polynomial, $P(s) = a_2 s^2 + a_1 s + a_0 = 0$, all the roots are in the left half plane (and the system with characteristic equation $P(s)$ is stable) if all the coefficients satisfy $a_n > 0$.
- For a third-order polynomial $P(s) = a_3 s^3 + a_2 s^2 + a_1 s + a_0 = 0$, all the coefficients must satisfy $a_n > 0$, and $a_2 a_1 > a_3 a_0$
- For a fourth-order polynomial $P(s) = a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0 = 0$, all the coefficients must satisfy $a_n > 0$, and $a_3 a_2 > a_4 a_1$ and $a_3 a_2 a_1 > a_4 a_1^2 + a_3^2 a_0$
- In general Routh stability criterion proclaims that all First column elements of Routh array is to be of same sign.

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

1) Part one (necessary condition for stability of the system): In this we have two conditions which are written below: (a) All the coefficients of the characteristic equation should be positive and real. (b) 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: (a) 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: a0 a2 a4 a6............ (b) 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: a1 a3 a5 a7........... (c) 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 a1 and a2 (where a2 is diagonally opposite element of next column) and then finally divide the result so obtain with a1. Mathematically we write as first element

(2) Second element : Multiply a0 with the diagonally opposite element of next to next column (i.e. a5) then subtract this from the product of a1 and a4 (where a4 is diagonally opposite element of next to next column) and then finally divide the result so obtain with a1. Mathematically we write as second elemenT

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 b1 with the diagonally opposite element of next column (i.e. a3) then subtract this from the product of a1 and b2 (where b2 is diagonally opposite element of next column) and then finally divide the result so obtain with b1. Mathematically we write as first element

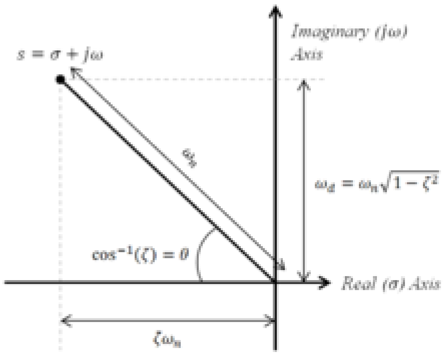
(2) Second element :Multiply b1 with the diagonally opposite element of next to next column (i.e. a5) then subtract this from the product of a1 and b3 (where b3 is diagonally opposite element of next to next column) and then finally divide the result so obtain with a1. Mathematically we write as second element

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:

In control theory and stability theory, **root locus analysis** is a graphical method for examining how the roots of a system change with variation of a certain system parameter, commonly a gain within a feedback system. This is a technique used as a stability criterion in the field of control systems developed by Walter R. Evans which can determinestability of the system. The root locus plots the poles of the closed loop transfer function in the complex S plane as a function of a gain parameter (see pole–zero plot).

Uses

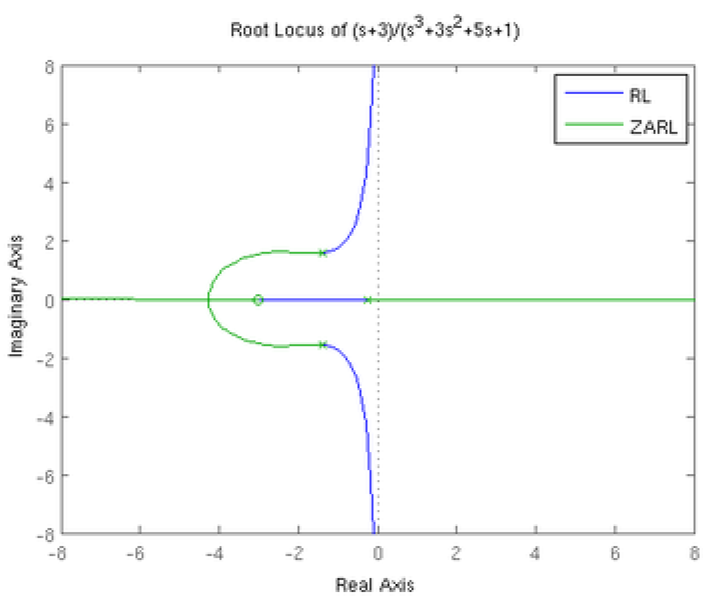


Effect of pole location on a second order system's natural frequency and damping ratio.

In addition to determining the stability of the system, the root locus can be used to design the damping ratio (ζ) and natural frequency (ω_n) of a feedback system. Lines of constant damping ratio can be drawn radially from the origin and lines of constant natural frequency can be drawn as arcs whose center points coincide with the origin. By selecting a point along the root locus that coincides with a desired damping ratio and natural frequency, a gain K can be calculated and implemented in the controller. More elaborate techniques of controller design using the root locus are available in most control textbooks: for instance, lag, lead, PI, PD and PID controllers can be designed approximately with this technique.

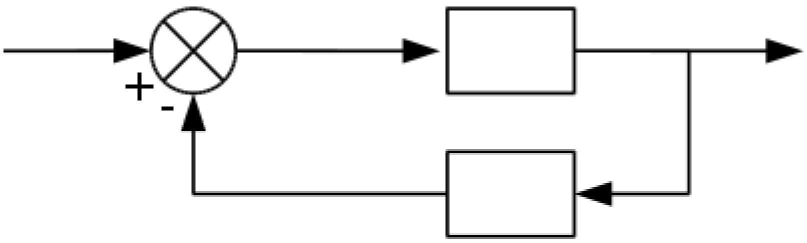
The definition of the damping ratio and natural frequency presumes that the overall feedback system is well approximated by a second order system; i.e. the system has a dominant pair of poles. This is often not the case, so it is good practice to simulate the final design to check if the project goals are satisfied.

Example



RL = root locus; ZARL = zero angle root locus

Suppose there is a feedback system whose input is the signal *X*(*s*) and output is *Y*(*s*). The feedback system forward path gain is *G*(*s*); the feedback path gain is *H*(*s*).



For this system, the overall transfer function is given by

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

Thus the closed-loop poles (roots of the characteristic equation) of the transfer function are the solutions to the equation 1 + *G*(*s*)*H*(*s*) = 0. The principal feature of this equation is that roots may be found wherever *G*(*s*)*H*(*s*) = -1.

In systems without pure delay, the product *G*(*s*)*H*(*s*) = -1 is a rational polynomial function and may be expressed as[2]

$$G(s)H(s) = \frac{K(s+z_1)(s+z_2)\cdots(s+z_m)}{(s+p_1)(s+p_2)\cdots(s+p_{m+n})}$$

where the $-z_i$ are the *m* zeros, the $-p_i$ are the *m* + *n* poles, and *K* is a scalar gain. Typically, a root locus diagram will indicate the transfer function's pole locations for varying values of *K*. A root locus plot will be all those points in the *s*-plane where *G*(*s*)*H*(*s*) = -1 for any value of *K*.

The factoring of *K* and the use of simple monomials means the evaluation of the rational polynomial can be done with vector techniques that add or subtract angles and multiply or divide magnitudes. The vector formulation arises from the fact that each monomial term in the factored *G*(*s*)*H*(*s*), (*s*−*a*) for example, represents the vector from *a* to *s*. The polynomial can be evaluated by considering the magnitudes and angles of each of these vectors. According to vector mathematics, the angle of the result is the sum of all the angles in the numerator add minus the sum of all the angles in the denominator. Similarly, the magnitude of the result is the product of all the magnitudes in the numerator divided by the product of all the magnitudes in the denominator. It turns out that the calculation of the magnitude is not needed because *K* varies; one of its values may result in a root. So to test whether a point in the *s*-plane is on the root locus, only the angles to all the open loop poles and zeros need be considered. A graphical method that uses a special protractor called a "Spirule" was once used to determine angles and draw the root loci.

From the function *T*(*s*), it can be seen that the value of *K* does not affect the location of the zeros. The root locus only gives the location of closed loop poles as the gain *K* is varied. The zeros of a system do not move.

Using a few basic rules, the root locus method can plot the overall shape of the path (locus) traversed by the roots as the value of *K* varies. The plot of the root locus then gives an idea of the stability and dynamics of this feedback system for different values of *K*.

Sketching root locus[edit]

- Mark open-loop poles and zeros
- Mark real axis portion to the left of an odd number of poles and zeros
- Find asymptotes

Let *P* be the number of poles and *Z* be the number of zeros:

$$P - Z = \text{number of asymptotes}$$

The asymptotes intersect the real axis at α (which is called the centroid) and depart at angle ϕ given by:

$$\phi_l = \frac{180^\circ + (l-1)360^\circ}{P-Z}, l = 1, 2, \ldots, P-Z$$

$$\alpha = \frac{\sum_P - \sum_Z}{P-Z}$$

where $\sum_P$ is the sum of all the locations of the poles, and $\sum_Z$ is the sum of all the locations of the explicit zeros.

- Phase condition on test point to find angle of departure
- Compute breakaway/break-in points

The breakaway points are located at the roots of the following equation:

$$\frac{dG(s)H(s)}{ds} = 0 \text{ or } \frac{d\overline{GH}(z)}{dz} = 0$$

Once you solve for *z*, the real roots give you the breakaway/reentry points. Complex roots correspond to a lack of breakaway/reentry.

z-plane versus *s*-plane

The root locus method can also be used for the analysis of sampled data systems by computing the root locus in the *z*-plane, the discrete counterpart of the *s*-plane. The equation $z = e^{sT}$ maps continuous *s*-plane poles (not zeros) into the *z*-domain, where *T* is the sampling period. The stable, left half *s*-plane maps into the interior of the unit circle of the *z*-plane, with the *s*-plane origin equating to $|z| = 1$ (because $e^0 = 1$). A diagonal line of constant damping in the *s*-plane maps around a spiral from (1,0) in the *z* plane as it curves in toward the origin. Note also that the Nyquist aliasing criteria is expressed graphically in the *z*-plane by the *x*-axis, where $\omega nT = \pi$.

The line of constant damping just described spirals in indefinitely but in sampled data systems, frequency content is aliased down to lower frequencies by integral multiples of the Nyquist frequency. That is, the sampled response appears as a lower frequency and better damped as well since the root in the *z*-plane maps equally well to the first loop of a different, better damped spiral curve of constant damping. Many other interesting and relevant mapping properties can be described, not least that *z*-plane controllers, having the property that they may be directly implemented from the *z*-plane transfer function (zero/pole ratio of polynomials), can be imagined graphically on a *z*-plane plot of the open loop transfer function, and immediately analyzed utilizing root locus.

Since root locus is a graphical angle technique, root locus rules work the same in the *z* and *s* planes.

The idea of a root locus can be applied to many systems where a single parameter *K* is varied. For example, it is useful to sweep any system parameter for which the exact value is uncertain in order to determine its behavior.

CONSTRUCTION OF ROOT LOCI:

To facilitate the application of the root-locus method for systems of higher order than 2nd, rules can be established. These rules are based upon the interpretation of the angle condition and the analysis of the characteristic equation. The rules presented aid in obtaining the root locus by expediting the manual plotting of the locus. But for automatic plotting using a computer these rules provide checkpoints to ensure that the solution is correct.

Though the angle and magnitude conditions can also be applied to systems having dead time, in the following we restrict to the case of the open-loop rational transfer functions according to Eq.

or

$$G_0(s) = k_0 \, \frac{b_0 + b_1 a + \ldots + b_{m-1} s^{m-1} + s^m}{a_0 + a_1 s + \ldots + a_{n-1} s^{n-1} + s^n} = k_0 \, \frac{N_0(s)}{D_0(s)} \; .$$

As this transfer function can be written in terms of poles and zeros s_{P_ν} and s_{Z_μ} ($\nu = 1, 2, \ldots n$; $\mu = 1, 2, \ldots, m$) $G_0(s)$ can be represented by their magnitudes and angles

$$G_0(s) = k_0 \frac{|s - s_{Z_1}| e^{j\varphi_{Z_1}} |s - s_{Z_2}| e^{j\varphi_{Z_2}} \ldots |s - s_{Z_m}| e^{j\varphi_{Z_m}}}{|s - s_{P_1}| e^{j\varphi_{P_1}} |s - s_{P_2}| e^{j\varphi_{P_2}} \ldots |s - s_{P_n}| e^{j\varphi_{P_n}}}$$

or

$$G_0(s) = k_0 \frac{\prod_{\mu=1}^{m} |s - s_{Z_\mu}|}{\prod_{\nu=1}^{n} |s - s_{P_\nu}|} e^{j\left(\sum_{\mu=1}^{m} \varphi_{Z_\mu} - \sum_{\nu=1}^{n} \varphi_{P_\nu}\right)} .$$

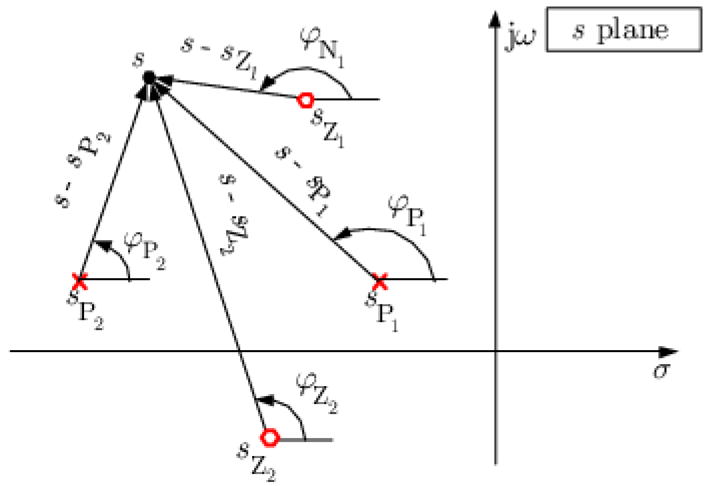
From Eq. (6.8) the *magnitude condition*

$$\frac{\prod_{\mu=1}^{m} |s - s_{Z_\mu}|}{\prod_{\nu=1}^{n} |s - s_{P_\nu}|} = \frac{1}{k_0}$$

and from Eq. the *angle condition*

$$\varphi(s) = \sum_{\mu=1}^{m} \varphi_{Z_\mu} - \sum_{\nu=1}^{n} \varphi_{P_\nu} = \pm 180^\circ (2k + 1) \quad \text{for} \quad k = 0, 1, 2, \ldots$$

follows. Here φ_{Z_μ} and φ_{P_ν} denote the angles of the complex values $(s - s_{Z_\mu})$ and $(s - s_{P_\nu})$, respectively. All angles are considered positive, measured in the counterclockwise sense. If for each point the sum of these angles in the s plane is calculated, just those particular points that fulfil the condition in Eq. are points on the root locus. This principle of constructing a root-locus curve - as shown in Figure is mostly used for automatic root-locus plotting.



Pole-zero diagram for construction of the root locus

In the following the most important *rules for the construction of root loci* for $k_0 > 0$ are listed:

Symmetry

As all roots are either real or complex conjugate pairs so that the root locus is symmetrical to the real axis.

Number of branches

The number of branches of the root locus is equal to the number of poles n of the open-loop transfer function.

Locus start and end points

The locus starting points ($k_0 = 0$) are at the open-loop poles and the locus ending points ($k_0 = \infty$) are at the open-loop zeros. $(n - m)$ branches end at infinity. The number of starting branches from a pole and ending branches at a zero is equal to the multiplicity of the poles and zeros, respectively. A point at infinity is considered as an equivalent zero of multiplicity equal to $n - m$.

Real axis locus

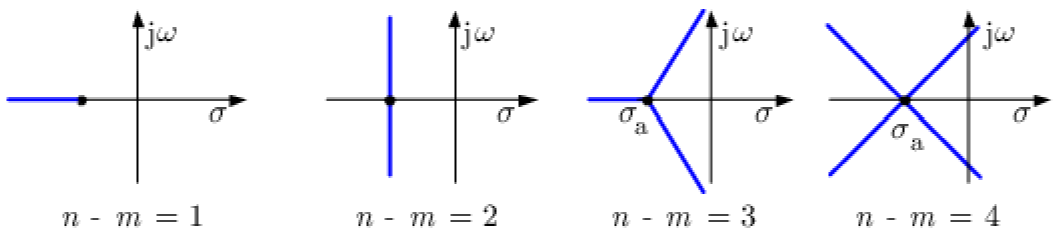
If the total number of poles and zeros to the right of a point on the real axis is odd, this point lies on the locus.

Asymptotes

There are $n - m$ asymptotes of the root locus with a slope of

$$\alpha_k = \arg s = \frac{\pm 180^\circ (2k+1)}{n-m} \ .$$

For $(n-m) = 1, 2, 3$ and 4 one obtains the asymptote configurations as shown in Figure 6.4.



Asymptote configurations of the root locus

Real axis intercept of the asymptotes

The real axis crossing $(\sigma_a, \mathrm{j}0)$ of the asymptotes is at

$$\sigma_a = \frac{1}{n-m} \left\{ \sum_{\nu=1}^{n} \mathrm{Re}\, s_{P_\nu} - \sum_{\mu=1}^{m} \mathrm{Re}\, s_{Z_\mu} \right\} \ .$$

Breakaway and break-in points on the real axis

At least one breakaway or break-in point $(\sigma_B, \mathrm{j}0)$ exists if a branch of the root locus is on the real axis between two poles or zeros, respectively. Conditions to find such real points are based on the fact that they represent multiple real roots. In addition to the characteristic equation for multiple roots the condition

$$\frac{\mathrm{d}}{\mathrm{d}s}[1 + G_0(s)] = \frac{\mathrm{d}}{\mathrm{d}s} G_0(s) = 0 \ .$$

must be fulfilled, which is equivalent to

$$\sum_{\nu=1}^{n} \frac{1}{s - s_{P_\nu}} = \sum_{\mu=1}^{m} \frac{1}{s - s_{Z_\mu}}$$

for $s = \sigma_B$. If there are no poles or zeros, the corresponding sum is zero.

Complex pole/zero angle of departure/entry

The angle of departure of pairs of poles with multiplicity r_{P_ϱ} is

$$\varphi_{P_\varrho,D} = \frac{1}{r_{P_\varrho}} \left\{ -\sum_{\substack{\nu=1 \\ \nu\neq\varrho}}^{n} \varphi_{P_\nu} + \sum_{\mu=1}^{m} \varphi_{Z_\mu} \pm 180^\circ(2k+1) \right\}$$

and the angle of entry of the pairs of zeros with multiplicity r_{Z_ϱ}

$$\varphi_{Z_\varrho,E} = \frac{1}{r_{Z_\varrho}} \left\{ -\sum_{\substack{\mu=1 \\ \mu\neq\varrho}}^{m} \varphi_{Z_\mu} + \sum_{\nu=1}^{n} \varphi_{P_\nu} \pm 180^\circ(2k+1) \right\} .$$

Rule 9 Root-locus calibration

The labels of the values of k_0 can be determined by using

$$k_0 = \frac{\prod_{\nu=1}^{n} |s - s_{P_\nu}|}{\prod_{\mu=1}^{m} |s - s_{Z_\mu}|} .$$

For $m = 0$ the denominator is equal to one.

Asymptotic stability

The closed loop system is asymptotically stable for all values of k_0 for which the locus lies in the left-half s plane. From the imaginary-axis crossing points the critical values $k_{0_{crit}}$ can be determined.

The rules shown above are for positive values of k_0. According to the angle condition of for negative values of k_0 some rules have to be modified. In the following these rules are numbered as above but labelled by a *.

Locus start and end points

The locus starting points ($k_0 = 0$) are at the open-loop poles and the locus ending points ($k_0 = -\infty$) are at the open-loop zeros. $(n-m)$ branches end at infinity. The number of starting branches from a pole and ending branches at a zero is equal to the multiplicity of the poles and zeros, respectively. A point at infinity is considered as an equivalent zero of multiplicity equal to $n-m$.

Real axis locus

If the total number of poles and zeros to the right of a point on the real axis is even including zero, this point lies on the locus.

Asymptotes

There are $n-m$ asymptotes of the root locus with a slope of

$$\alpha_k = \arg s = \frac{\pm 360° k}{n-m} .$$

Complex pole/zero angle of departure/entry

The angle of departure of pairs of poles with multiplicity r_{P_ϱ} is

$$\varphi_{\mathrm{P}_\varrho,\mathrm{D}} = \frac{1}{r_{\mathrm{P}_\varrho}} \left\{ -\sum_{\substack{\nu=1\\ \nu\neq\varrho}}^{n} \varphi_{\mathrm{P}_\nu} + \sum_{\mu=1}^{m} \varphi_{\mathrm{Z}_\mu} \pm 360° k \right\}$$

and the angle of entry of the pairs of zeros with multiplicity r_{Z_ϱ}

$$\varphi_{Z_{\varrho},E} = \frac{1}{r_{Z_{\varrho}}} \left\{ -\sum_{\substack{\mu=1 \\ \mu\neq\varrho}}^{m} \varphi_{Z_{\mu}} + \sum_{\nu=1}^{n} \varphi_{P_{\nu}} \pm 360^{\circ} k \right\} .$$

The root-locus method can also be applied for other cases than varying k_0. This is possible as long as $G_0(s)$ can be rewritten such that the angle condition according to Eq. and the rules given above can be applied. This will be demonstrated in the following two examples.

Given the closed-loop characteristic equation

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

the root locus for varying the parameter a_1 is required. The characteristic equation is therefore rewritten as

$$1 + a_1 \frac{s}{a_0 + a_2 s^2 + \ldots + s^n} = 0 .$$

This form then correspondents to the standard form

$$1 + G_0(s) = 1 + a_1 \frac{N_0(s)}{D_0(s)} = 0$$

to which the rules can be applied. ■

Given the closed-loop characteristic equation

$$s^3 + (3 + \alpha) s^2 + 2s + 4 = 0 ,$$

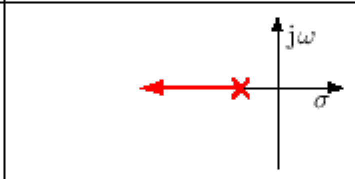
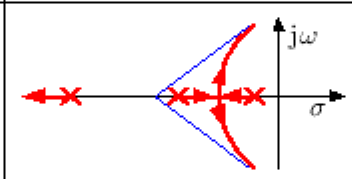
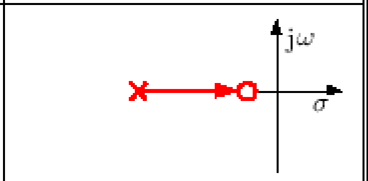
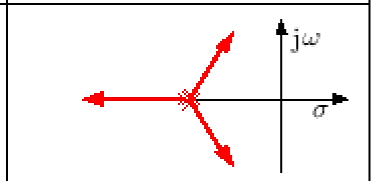
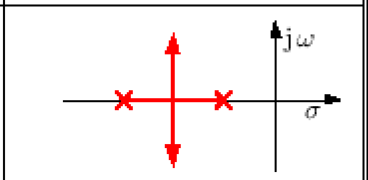
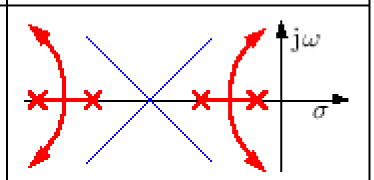
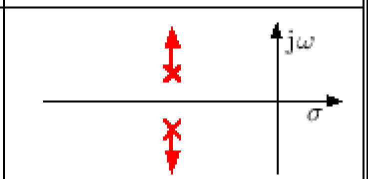
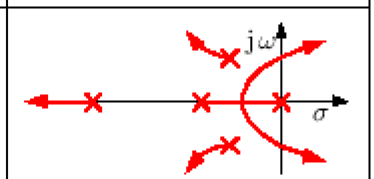
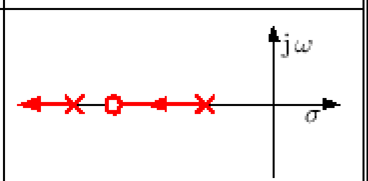
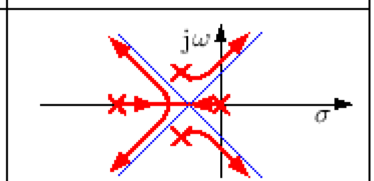
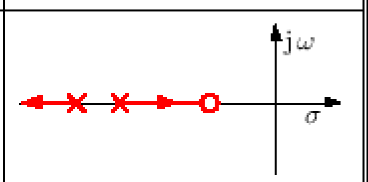
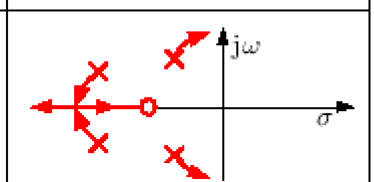
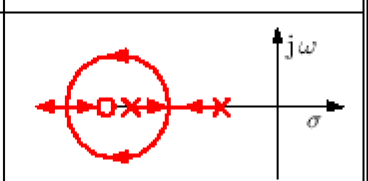
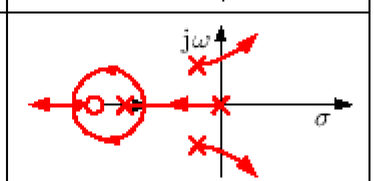
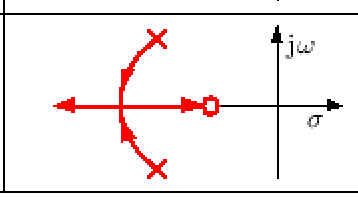
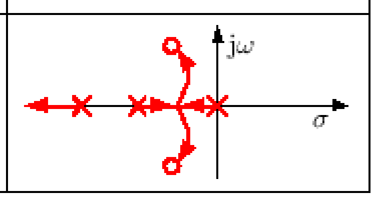
it is required to find the effect of the parameter α on the position of the closed-loop poles. The equation is rewritten into the desired form

$$1+\alpha\frac{s^2}{s^3+3s^2+2s+4}=0\ .$$

■

Using the rules 1 to 10 one can easily predict the geometrical form of the root locus based on the distribution of the open-loop poles and zeros. Table 6.2 shows some typical distributions of open-loop poles and zeros and their root loci.

Typical distributions of open-loop poles and zeros and the root loci

No.	root locus	No.	root locus
1		9	
2		10	
3		11	
4		12	
5		13	
6		14	
7		15	
8		16	

For the qualitative assessment of the root locus one can use a physical analogy. If all open-loop poles are substituted by a negative electrical charge and all zeros by a commensurate positive one and if a massless negative charged particle is put onto a point of the root locus, a movement is observed. The path that the particle takes because of the interplay between the repulsion of the poles and the attraction of the zeros lies just on the root locus.

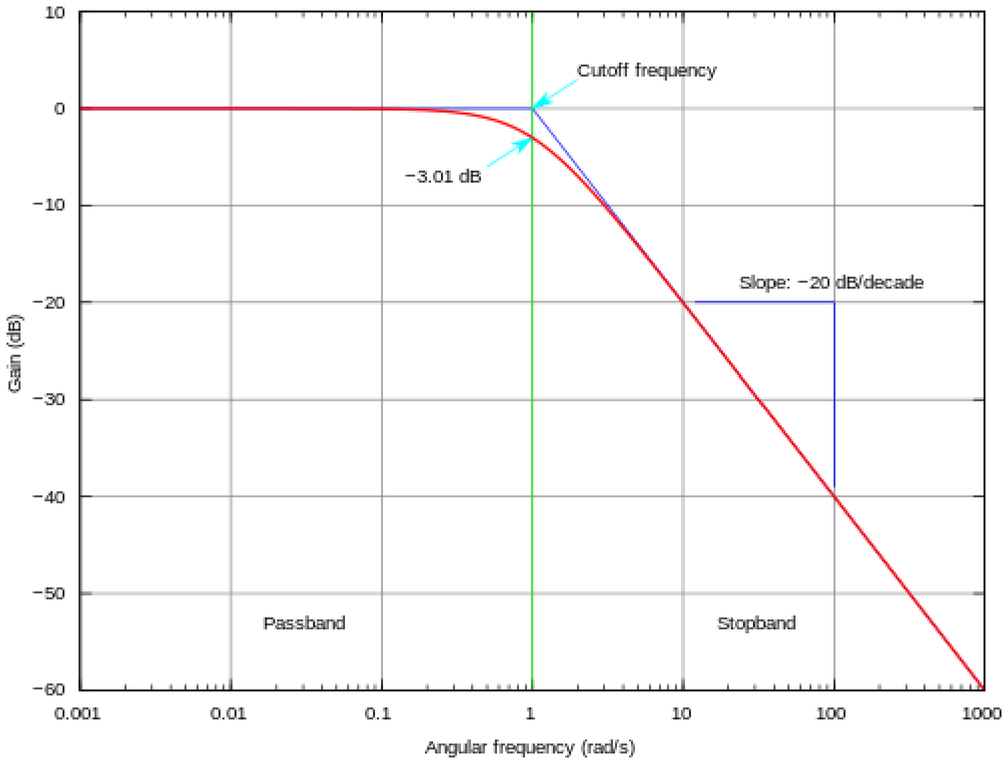
SEE1203 – CONTROL SYSTEMS

UNIT 4

FREQUENCY RESPONSE ANALYSIS

Frequency response is the quantitative measure of the output spectrum of a system or device in response to a stimulus, and is used to characterize the dynamics of the system. It is a measure of magnitude and phase of the output as a function of frequency, in comparison to the input. In simplest terms, if a sine wave is injected into a system at a given frequency, a linear system will respond at that same frequency with a certain magnitude and a certain phase angle relative to the input. Also for a linear system, doubling the amplitude of the input will double the amplitude of the output. In addition, if the system is time-invariant (so LTI), then the frequency response also will not vary with time. Thus for LTI systems, the frequency response can be seen as applying the system's transfer function to a purely imaginary number argument representing the frequency of the sinusoidal excitation.[1]

Two applications of frequency response analysis are related but have different objectives. For an audio system, the objective may be to reproduce the input signal with no distortion. That would require a uniform (flat) magnitude of response up to the bandwidth limitation of the system, with the signal delayed by precisely the same amount of time at all frequencies. That amount of time could be seconds, or weeks or months in the case of recorded media. In contrast, for a feedback apparatus used to control a dynamic system, the objective is to give the closed-loop system improved response as compared to the uncompensated system. The feedback generally needs to respond to system dynamics within a very small number of cycles of oscillation (usually less than one full cycle), and with a definite phase angle relative to the commanded control input. For feedback of sufficient amplification, getting the phase angle wrong can lead to instability for an open-loop stable system, or failure to stabilize a system that is open-loop unstable. Digital filters may be used for both audio systems and feedback control systems, but since the objectives are different, generally the phase characteristics of the filters will be significantly different for the two applications.



Nonlinear frequency response

If the system under investigation is nonlinear then applying purely linear frequency domain analysis will not reveal all the nonlinear characteristics. To overcome these limitations, generalized frequency response functions and nonlinear output frequency response functions have been defined that allow the user to analyze complex nonlinear dynamic effects.[2] The nonlinear frequency response methods reveal complex resonance, inter modulation, and energy transfer effects that cannot be seen using a purely linear analysis and are becoming increasingly important in a nonlinear world.

TIME DOMAIN AND FREQUENCY DOMAIN:

The frequency domainrefers to the analysis of mathematical functions or signals with respect to frequency, rather than time.[1] Put simply, a time-domain graph shows how a signal changes over time, whereas a frequency-domain graph shows how much of the signal lies within each given frequency band over a range of frequencies. A frequency-domain representation can also include information on the phase shift that must be applied to each sinusoid in order to be able to recombine the frequency components to recover the original time signal. the frequency domainrefers to the analysis of mathematical functions or signals with respect to frequency, rather than time.[1] Put simply, a time-domain graph shows how a signal changes over time, whereas a frequency-domain graph shows how much of the signal lies within each given frequency band over a range of frequencies. A frequency-domain representation can also include information on the phase shift that must be applied to each sinusoid in order to be able to recombine the frequency components to recover the original time signal. Time domain and frequency domain are two modes used to analyze data. Both time domain analysis and frequency domain analysis are widely used in fields such as electronics, acoustics, telecommunications, and many other fields.

- Frequency domain analysis is used in conditions where processes such as filtering, amplifying and mixing are required.
- Time domain analysis gives the behavior of the signal over time. This allows predictions and regression models for the signal.
- Frequency domain analysis is very useful in creating desired wave patterns such as binary bit patterns of a computer.
- Time domain analysis is used to understand data sent in such bit patterns over time.

Time Domain:

Time domain analysis is analyzing the data over a time period. Functions such as electronic signals, market behaviors, and biological systems are some of the functions that are analyzed using time domain analysis. For an electronic signal, the time domain analysis is mainly based on the voltage – time plot or the current – time plot. In a time domain analysis, the variable is always measured against time. There are several devices used to analyze data on a time domain basis. The cathode ray oscilloscope (CRO) is the most common device when analyzing electrical signals on a time domain.

Frequency Domain:

Frequency domain is a method used to analyze data. This refers to analyzing a mathematical function or a signal with respect to the frequency. Frequency domain analysis is widely used in fields such as control systems engineering, electronics and statistics. Frequency domain analysis is mostly used to signals or functions that are periodic over time. This does not mean that frequency domain analysis cannot be used in signals that are not periodic.

The most important concept in the frequency domain analysis is the transformation. Transformation is used to convert a time domain function to a frequency domain function and vice versa. The most common transformation used in the frequency domain is the Fourier transformations. Fourier transformation is used to convert a signal of any shape into a sum of infinite number of sinusoidal waves. Since analyzing sinusoidal functions is easier than analyzing general shaped functions, this method is very useful and widely used.

All signals have a frequency domain representation and Fourier detailed the theory that any real world waveform can be generated by the addition of sinusoidal waves. The following diagram shows an example of this process:

There are a number of different mathematical transforms which are used to analyze time functions and are referred to as frequency domain methods. The following are some most common transforms, and the fields in which they are used:

- Fourier series – repetitive signals, oscillating systems
- Fourier transform – nonrepetitive signals, transients
- Laplace transform – electronic circuits and control systems
- Z transform – discrete signals, digital signal processing .

GAIN AND PHASE MARGINS:

The gain margin is the amount of gain increase or decrease required to make the loop gain unity at the frequency Wgm where the phase angle is –**180°** (modulo 360°). In other words, the gain margin is 1/g if g is the gain at the –**180°** phase frequency. n electronic amplifiers, the **phase margin** (PM) is the difference between the phase and 180°, for an amplifier's output signal (relative to its input), at a certain frequency.

$$\mathrm{PM} = 180^\circ - |\Delta\phi|.$$

Typically the open-loop phase lag (relative to input) varies with frequency, progressively increasing to exceed 180°, at which frequency the output signal becomes inverted, or antiphase in relation to the input. The PM will be positive but decreasing at frequencies less than the frequency at which inversion sets in (at which PM = 0), and PM is negative (PM < 0) at higher frequencies. In the presence of negative feedback, a zero or negative PM at a frequency where the loop gain exceeds unity (1) guarantees instability. Thus positive PM is a "safety margin" that ensures proper (non-oscillatory) operation of the circuit. This applies to amplifier circuits as well as more generally, to active filters, under various load conditions (e.g. reactive loads). In its simplest form, involving ideal negative feedback *voltage* amplifiers with non-reactive feedback, the phase margin is measured at the frequency where the open-loop voltage gain of the amplifier equals the desired closed-loop DC voltage gain.[1]

More generally, PM is defined as that of the amplifier and its feedback network combined (the "loop", normally opened at the amplifier input), measured at a frequency where the loop gain is unity, and prior to the closing of the loop, through tying the output of the open loop to the input source, in such a way as to subtract from it.

In the above loop-gain definition, it is assumed that the amplifier input presents zero load. To make this work for non-zero-load input, the output of the feedback network needs to be loaded with an equivalent load for the purpose of determining the frequency response of the loop gain.

It is also assumed that the graph of gain vs. frequency crosses unity gain with a negative slope and does so only once. This consideration matters only with reactive and active feedback networks, as may be the case with active filters.

Phase margin and its important companion concept, gain margin, are measures of stability in closed-loop, dynamic-control systems. Phase margin indicates relative stability, the tendency to oscillate during its damped response to an input change such as a step function. Gain margin indicates absolute stability and the degree to which the system will oscillate, without limit, given any disturbance.

The output signals of all amplifiers exhibit a time delay when compared to their input signals. This delay causes a phase difference between the amplifier's input and output signals. If there are enough stages in the amplifier, at some frequency, the output signal will lag behind the input signal by one cycle period at that frequency. In this situation, the amplifier's output signal will be in phase with its input signal though lagging behind it by 360°, i.e., the output will have a phase angle of −360°. This lag is of great consequence in amplifiers that use feedback. The reason: the amplifier will oscillate if the fed-back output signal is in phase with the input signal at the frequency at which its open-loop voltage gain equals its closed-loop voltage gain and the open-loop voltage gain is one or greater. The oscillation will occur because the fed-back output signal will then reinforce the input signal at that frequency.[2] In conventional operational amplifiers, the critical output phase angle is −180° because the output is fed back to the input through an inverting input which adds an additional −180°.

In practice, feedback amplifiers must be designed with phase margins substantially in excess of 0°, even though amplifiers with phase margins of, say, 1° are theoretically stable. The reason is that many practical factors can reduce the phase margin below the theoretical minimum. A prime example is when the amplifier's output is connected to a capacitive load. Therefore, operational amplifiers are usuallycompensated to achieve a minimum phase margin of 45° or so. This means that at the frequency at which the open and closed loop gains meet, the phase angle is −135°. The calculation is: -135° - (-180°) = 45°. See Warwick[3] or Stout[4] for a detailed analysis of the techniques and results of compensation to insure adequate phase margins. See also the article "Pole splitting". Often amplifiers are designed to achieve a typical phase margin of 60 degrees. If the typical phase margin is around 60 degrees then the minimum phase margin will typically be greater than 45 degrees. A phase margin of 60 degrees is also a magic number because it allows for the fastest settling time when attempting to follow a voltage step input (a Butterworth design). An amplifier with lower phase margin will ring[nb 1] for longer and an amplifier with more phase margin will take a longer time to rise to the voltage step's final level.

A related measure is gain margin. While phase margin comes from the phase where the loop gain equals one, the gain margin is based upon the gain where the phase equals -180 degrees.

BODE PLOTS:

Bode plot is a graph of the frequency response of a system. It is usually a combination of a Bode magnitude plot, expressing the magnitude (usually in decibels) of the frequency response, and a Bode phase plot, expressing the phase shift. Both quantities are plotted against a horizontal axis proportional to thelogarithm of frequency. Given that the decibel is itself a logarithmic scale, the Bode amplitude plot is log–log plot, whereas the Bode phase plot is a lin-logplot.[1]

As originally conceived by Bode in the 1930s, the plot is only an asymptoticapproximation of the frequency response, using straight line segments.[2]However, with the advent of low cost computing, it is often taken nowadays to mean the precise plot of the actual frequency response.

NYQUIST STABILITY CRITERION:

In control theory and stability theory, the **Nyquist stability criterion**, discovered by Swedish-American electrical engineer Harry Nyquist at Bell Telephone Laboratoriesin 1932,[1] is a graphical technique for determining the stability of a dynamical system. Because it only looks at the Nyquist plot of the open loop systems, it can be applied without explicitly computing the poles and zeros of either the closed-loop or open-loop system (although the number of each type of right-half-plane singularities must be known). As a result, it can be applied to systems defined by non-rational functions, such as systems with delays. In contrast to Bode plots, it can handletransfer functions with right half-plane singularities. In addition, there is a natural generalization to more complex systems with multiple inputs and multiple outputs, such as control systems for airplanes.

The Nyquist criterion is widely used in electronics and control system engineering, as well as other fields, for designing and analyzing systems with feedback. While Nyquist is one of the most general stability tests, it is still restricted to linear, time-invariant (LTI) systems. Non-linear systems must use more complex stability criteria, such as Lyapunov or the circle criterion. While Nyquist is a graphical technique, it only provides a limited amount of intuition for why a system is stable or unstable, or how to modify an unstable system to be stable. Techniques like Bode plots, while less general, are sometimes a more useful design tool. In control theory and stability theory, the **Nyquist stability criterion**, discovered by Swedish-American electrical engineer Harry Nyquist at Bell Telephone Laboratoriesin 1932,[1] is a graphical technique for determining the stability of a dynamical system. Because it only looks at the Nyquist plot of the open loop systems, it can be applied without explicitly computing the poles and zeros of either the closed-loop or open-loop system (although the number of each type of right-half-plane singularities must be known). As a result, it can be applied to systems defined by non-rational functions, such as systems with delays. In contrast to Bode plots, it can handletransfer functions with right half-plane singularities. In addition, there is a natural generalization to more complex systems with multiple inputs and multiple outputs, such as control systems for airplanes.

The Nyquist criterion is widely used in electronics and control system engineering, as well as other fields, for designing and analyzing systems with feedback. While Nyquist is one of the most general stability tests, it is still restricted to linear, time-invariant (LTI) systems. Non-linear systems must use more complex stability criteria, such as Lyapunov or the circle criterion. While Nyquist is a graphical technique, it only provides a limited amount of intuition for why a system is stable or unstable, or how to modify an unstable system to be stable. Techniques like Bode plots, while less general, are sometimes a more useful design tool.

The Nyquist criterion

We first construct **the Nyquist contour**, a contour that encompasses the right-half of the complex plane:

- a path traveling up the $j\omega$ axis, from $0 - j\infty$ to $0 + j\infty$.
- a semicircular arc, with radius $r \rightarrow \infty$, that starts at $0 + j\infty$ and travels clock-wise to $0 - j\infty$.

The Nyquist contour mapped through the function $1 + G(s)$ yields a plot of $1 + G(s)$ in the complex plane. By the Argument Principle, the number of clock-wise encirclements of the origin must be the number of zeros of $1 + G(s)$ in the right-half complex plane minus the poles of $1 + G(s)$ in the right-half complex plane. If instead, the contour is mapped through the open-loop transfer function $G(s)$, the result is the Nyquist Plot of $G(s)$. By counting the resulting contour's encirclements of -1, we find the difference between the number of poles and zeros in the right-half complex plane of $1 + G(s)$. Recalling that the zeros of $1 + G(s)$ are the poles of the closed-loop system, and noting that the poles of $1 + G(s)$ are same as the poles of $G(s)$, we now state

The Nyquist Criterion:

Given a Nyquist contour Γ_s, let P be the number of poles of $G(s)$ encircled by Γ_s, and Z be the number of zeros of $1 + G(s)$encircled by Γ_s. Alternatively, and more importantly, Z is the number of poles of the closed loop system in the right half plane. The resultant contour in the $G(s)$-plane, $\Gamma_{G(s)}$ shall encircle (clock-wise) the point $(-1 + j0)$ N times such that $N = Z - P$.

- If the system is originally open-loop unstable, feedback is necessary to stabilize the system. Right-half-plane (RHP) poles represent that instability. For closed-loop stability of a system, the number of closed-loop roots in the right half of the s-plane must be zero. Hence, the number of counter-clockwise encirclements about $-1 + j0$ must be equal to the number of open-loop poles in the RHP. Any clockwise encirclements of the critical point by the open-loop frequency response (when judged from low frequency to high frequency) would indicate that the feedback control system would be destabilizing if the loop were closed. (Using RHP zeros to "cancel out" RHP poles does not remove the instability, but rather ensures that the system will remain unstable even in the presence of feedback, since the closed-loop roots travel between open-loop poles and zeros in the presence of feedback. In fact, the RHP zero can make the unstable pole unobservable and therefore not stabilizable through feedback. If the open-loop transfer function $G(s)$ has a zero pole of multiplicity l, then the Nyquist plot has a discontinuity at $\omega = 0$. During further analysis it should be assumed that the phasor travels l times clock-wise along a semicircle of infinite radius. After applying this rule, the zero poles should be neglected, i.e. if there are no other unstable poles, then the open-loop transfer function $G(s)$ should be considered stable.

- If the open-loop transfer function $G(s)$ is stable, then the closed-loop system is unstable for *any* encirclement of the point -1.

- If the open-loop transfer function $G(s)$ is *unstable*, then there must be one *counter* clock-wise encirclement of -1 for each pole of $G(s)$ in the right-half of the complex plane.
- The number of surplus encirclements (greater than N+P) is exactly the number of unstable poles of the closed-loop system.
- However, if the graph happens to pass through the point $-1 + j0$, then deciding upon even the marginal stability of the system becomes difficult and the only conclusion that can be drawn from the graph is that there exist zeros on the $j\omega$ axis.

POLAR PLOTS:

the polar coordinate system is a two-dimensional coordinate system in which each point on a plane is determined by a distance from a reference point and an anglefrom a reference direction. Angles in polar notation are generally expressed in either degrees or radians (2π rad being equal to 360°). Degrees are traditionally used in navigation, surveying, and many applied disciplines, while radians are more common in mathematics and mathematical physics.[9]

In many contexts, a positive angular coordinate means that the angle ϕ is measuredcounterclockwise from the axis.

Polar equation of a curve

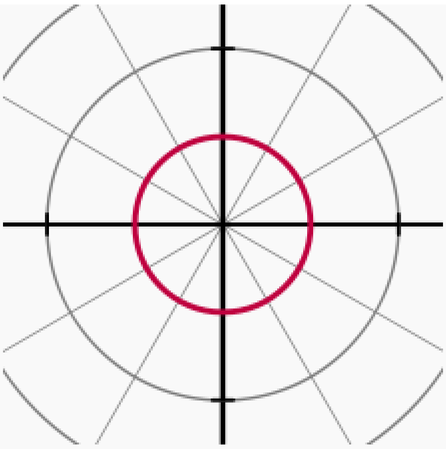
The equation defining an algebraic curve expressed in polar coordinates is known as a *polar equation*. In many cases, such an equation can simply be specified by defining *r* as afunction of ϕ. The resulting curve then consists of points of the form ($r(\phi)$, ϕ) and can be regarded as the graph of the polar function r.

Different forms of symmetry can be deduced from the equation of a polar function r. If $r(-\phi) = r(\phi)$ the curve will be symmetrical about the horizontal (0°/180°) ray, if $r(\pi - \phi) = r(\phi)$ it will be symmetric about the vertical (90°/270°) ray, and if $r(\phi - \alpha) = r(\phi)$ it will be rotationally symmetric by α clockwise and counterclockwise about the pole.

Because of the circular nature of the polar coordinate system, many curves can be described by a rather simple polar equation, whereas their Cartesian form is much more intricate. Among the best known of these curves are the polar rose, Archimedean spiral, lemniscate,limaçon, and cardioid.

For the circle, line, and polar rose below, it is understood that there are no restrictions on the domain and range of the curve.

Circle



A circle with equation $r(\phi) = 1$

The general equation for a circle with a center at (r_0, γ) and radius a is

$$r^2 - 2rr_0\cos(\varphi - \gamma) + r_0^2 = a^2.$$

This can be simplified in various ways, to conform to more specific cases, such as the equation

$$r(\varphi) = a$$

for a circle with a center at the pole and radius a.[14]

When r_0 = a, or when the origin lies on the circle, the equation becomes

$$r = 2a\cos(\varphi - \gamma).$$

In the general case, the equation can be solved for r, giving

$$r = r_0\cos(\varphi - \gamma) + \sqrt{a^2 - r_0^2\sin^2(\varphi - \gamma)},$$

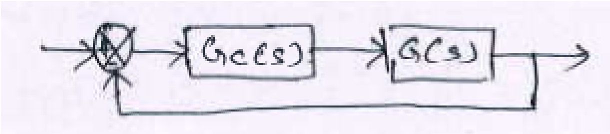
the solution with a minus sign in front of the square root gives the same curve

SEE1203 – CONTROL SYSTEMS

Unit V

COMPENSATION AND CONTROLLERS

PI, PD, PID Controllers



A Controller with transfer function $G_c(s)$ can be introduced in cascade with open loop transfer function, G(s) to modify the transient and steady state response of the system.

The different types of controllers employed in control system are the following:

1. Proportional Controller (P- Controller)
2. Proportional – Plus –Integral Controller (PI Controller)
3. Proportional – Plus – Derivative Controller (PD Controller)
4. Proportional – Plus – Derivative- Plus –Integral Controller (PID Controller)

Transfer Function of P – Controller, $G_c(s) = \frac{U(s)}{E(s)} = k_p$, k_p – Proportional gain

Transfer Function of PI – Controller, $G_c(s) = \frac{U(s)}{E(s)} = k_p + \frac{k_i}{s}$, k_i – Integral constant or gain

Transfer Function of PD – Controller, $G_c(s) = \frac{U(s)}{E(s)} = k_p + \frac{k_i}{s} + k_d s$

Procedure for Design of PD/PI/PID controller in frequency domain

Step 1: Determine the magnitude and phase of uncompensated open loop sinusoidal transfer function (ie $G(j\omega)$)

Let $A_1 = |G(j\omega)|$ at $\omega=\omega_1$

And $\phi_1 = \angle G(j\omega)$ at $\omega=\omega_1$

Step 2: Determine the phase margin of uncompensated system and the angle to be contributed by the controller to achieve the desired phase margin.

Let y_u = phase margin of uncompensated system

Y_d = Desired phase margin at ω_1

θ = Phase angle of the controller at $\omega=\omega_1$

Now $y_u = 180° + \phi_1$

$\theta = y_d - y_u$

Step 3: Determine the transfer function of the controller

a) PD Controller

Derivative Constant , $k_d = \dfrac{\sin\theta}{\omega_1 A_1}$

Proportional constant, $k_p = \dfrac{\cos\theta}{A_1}$

Transfer function of PD controller $G_c(s) = (k_p + k_d s) = k_p(1 + \dfrac{k_d}{k_p} s)$

b) PI Controller

Integral constant , $k_i = \dfrac{-\omega_1 \sin\theta}{A_1}$

Proportional constant, $k_p = \dfrac{\cos\theta}{A_1}$

Transfer function of PI controller $G_c(s) = (k_p + \dfrac{k_i}{s}) = \dfrac{k_i(1 + \dfrac{k_p}{k_i} s)}{s}$

c) PID Controller

Transfer function of PI controller $G_c(s) = (k_p + k_d s + \dfrac{k_i}{s}) = \dfrac{k_d(s^2 + \dfrac{k_p}{k_d} s + \dfrac{k_i}{k_d})}{s}$

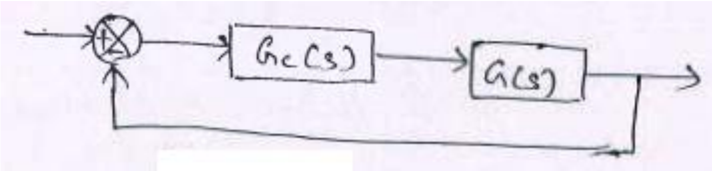
Evaluate K_i such that the compensated system satisfies the error requirement. For example if the compensated system is type 1 system then, $k_v = \underset{s \to 0}{Lt}\, sG_c(s)G(s)$ will give the value of k_i

Derivative constant $k_d = \dfrac{\sin\theta}{\omega_1 A_1} + \dfrac{k_i}{\omega^2}$

Proportional constant , $k_p = \dfrac{\cos\theta}{A_1}$

Step 4: Determine the open loop transfer function of compensation system

The transfer function of the controller is placed in series with G(s) as shown in fig



Open loop transfer function of compensated system, $G_o(s) = G_c(s) \times G(s)$

Step 5: Verify the design by calculating phase margin of compensated system.

Let $A_0 = |G_0(j\omega)|$ at $\omega=\omega_1$

and $\phi_0 = \angle G_0(j\omega)$ at $\omega=\omega_1$

Y_0 = phase margin of compensated system

Now $y_0 = 180° + \phi_0$

It can be observed that $A_0=1$ anf y_0 satisfies the specifications.

Example: Consider a unity feedback system with open loop transfer function, $G(s) = \frac{5}{s(s+0.5)(s+1)}$. Design a PD controller so that the phase margin of the system is 30° at a frequency of 1.2 rad/sec.

Solution:

Step 1:

$$G(s) = \frac{5}{s(s+0.5)(s+1)} = \frac{5}{s \times 0.5(1+\frac{s}{0.5})(1+s)} = \frac{10}{s(1+2s)(1+s)}$$

Put $s=j\omega$ in G(s)

$$G(j\omega) == \frac{10}{j\omega(1+2j\omega)(1+j\omega)} = \frac{10}{\omega\angle 90^\circ \sqrt{1+4\omega^2}\angle \tan^{-1} 2\omega \sqrt{1+\omega^2}\angle \tan^{-1}\omega}$$

$$|G(j\omega)| = \frac{10}{\omega\sqrt{1+4\omega^2}\sqrt{1+\omega^2}}$$

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

The gain crossover frequency of compensated system ω_1=1.21 rad/sec

$$|A_1| = \frac{10}{1.2\sqrt{1+4\times1.2^2}\,\sqrt{1+1.2^2}} = 2.052$$

$$\phi_1 = -90^\circ - \tan^{-1}(2\times1.2) - \tan^{-1}(1.2) = -207.5$$

To find y_u & ϕ

$Y_u = 180 + \phi_1 = 180 + (-207.5) = -27.5$

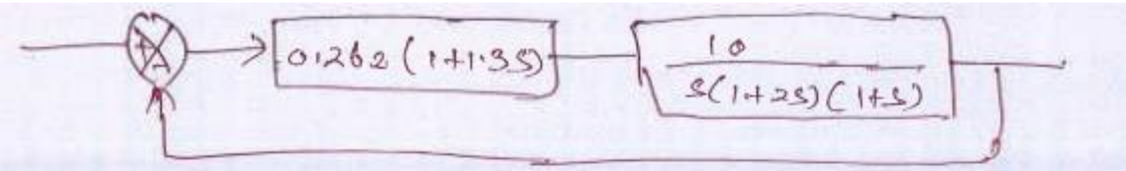
$\theta = y_d - y_u = 30 - (-27.5) = 57.3$

PD Controller

$$k_d = \frac{\sin\theta}{\omega_1 A_1} = \frac{\sin 57.5}{1.2\times 2.052} = 0.343$$

$$k_p = \frac{\cos\theta}{A_1} = \frac{\cos 57.5}{2.052} = 0.262$$

$$G_c(s) = (k_p + k_d s) = k_p(1 + \frac{k_d}{k_p}s) = 0.262(1 + \frac{0.343}{0.262}s) = 0.262(1+1.3s)$$



$$G_o(s) = G_c(s)\times G(s) = 0.262(1+1.32)\times\frac{10}{s(1+2s)(1+s)}$$

To verify the design

Put s=jω in G0(s)

$$G_0(j\omega) = \frac{2.62(1+j1.3\omega)}{j\omega(1+2j\omega)(1+j\omega)} = \frac{2.62\sqrt{1+1.69\omega^2}\angle\tan^{-1}1.3\omega}{\omega\angle 90^\circ\sqrt{1+4\omega^2}\angle\tan^{-1}2\omega\sqrt{1+\omega^2}\angle\tan^{-1}\omega}$$

$$A_0 = \frac{2.62\sqrt{1+1.69\omega^2}}{\omega\sqrt{1+4\omega^2}\,\sqrt{1+\omega^2}}$$

$$\phi_0 = \tan^{-1} 1.3\omega - 90^\circ - \tan^{-1} 2\omega - \tan^{-1}\omega$$

$$\omega = \omega_1, A = A_{01} = \frac{2.62\sqrt{1+1.69\times1.2^2}}{1.2\times\sqrt{1+4\times1.2^2}\sqrt{1+1.2^2}} = 1$$

$$\omega = \omega_1, \phi_0 = \phi_{01} = \tan^{-1}(1.3\times1.2) - 90 - \tan^{-1}(2\times1.2) - \tan^{-1}1.2 = -150$$

$$y_0 = 180 + \phi_{01} = 180 - 150 = 30^o$$

Phase margin of the compensated system is satisfactory, hence the design is acceptable.

Transfer function of PD controller $G_c(s)$ =0.262 (1+1.33)

$$G_0(s) = \frac{2.62(1+1.3s)}{s(1+2s)(1+s)}$$

Example 2:

Consider a unity feedback system with open loop transfer function $G(s) = \frac{100}{(s+1)(s+2))(s+10)}$. Design a PID controller, so that the phase margin of the system is 45° at a frequency of 4 rad/sec and the steady state error for unit ramp input is 0.1.

Solution

$$G(s) = \frac{100}{(s+1)(s+2))(s+10)} = \frac{100}{(1+s)\times2\times(1+\frac{s}{2})\times10\times(1+\frac{s}{10})} = \frac{5}{(1+s)(1+0.5s)(1+0.1s)}$$

Put s=jω in G(s)

$$G(j\omega) == \frac{5}{(1+j\omega)(1+j0.5\omega)(1+j0.1\omega)}$$

$$G(j\omega) == \frac{5}{\sqrt{1+\omega^2}\angle\tan^{-1}\omega\sqrt{1+0.25\omega^2}\angle\tan^{-1}0.5\omega\sqrt{1+0.01\omega^2}\angle\tan^{-1}0.1\omega}$$

$$|G(j\omega)| == \frac{5}{\sqrt{1+\omega^2}\sqrt{1+0.25\omega^2}\sqrt{1+0.01\omega^2}}$$

$$\angle G(j\omega) = -\tan^{-1}\omega - \tan^{-1}0.5\omega - \tan^{-1}0.1\omega$$

The gain crossover frequency of compensated system ω1=4rad/sec

Let, A_1=|G(jω)| at ω=ω_1

ϕ_1=∠G(jω) at ω=ω_1

$$A_1 = \frac{5}{\sqrt{1+4^2}\sqrt{1+0.25\times4^2}\sqrt{1+0.01\times4^2}} = 0.5$$

$$\phi_1 = -\tan^{-1}4 - \tan^{-1}(0.5\times4) - \tan^{-1}(0.1\times4) = -161^o$$

To find y_u & θ

Y_u=180+ϕ_1=180-161=19°

θ=y_d-y_u = 45-19 =26°

To find transfer function of PID controller

ess =0.1 for unit ramp input

$$k_v = \frac{1}{e_{ss}} = \frac{1}{0.1} = 10$$

$$k_v = \underset{s\to0}{Lt}\, sG_c(s)G(s)$$

$$G_c(s) = (k_p + k_d s + \frac{k_i}{s}) = \frac{k_d s^2 + k_p s + k_i)}{s}$$

$$G(s) = \frac{5}{(1+s)(1+0.5s)(1+0.1s)}$$

$$k_v = \underset{s\to0}{Lt}\, s\frac{k_d s^2 + k_p s + k_i}{s} \times \frac{5}{(1+s)(1+0.5s)(1+0.1s)} = 10$$

$$5k_i = 10; k_i = \frac{10}{5} = 2$$

$$k_d = \frac{\sin\theta}{\omega_1 A_1} + \frac{k_i}{\omega^2} = \frac{\sin(26)}{4\times 0.5} + \frac{2}{4^2} = 0.344$$

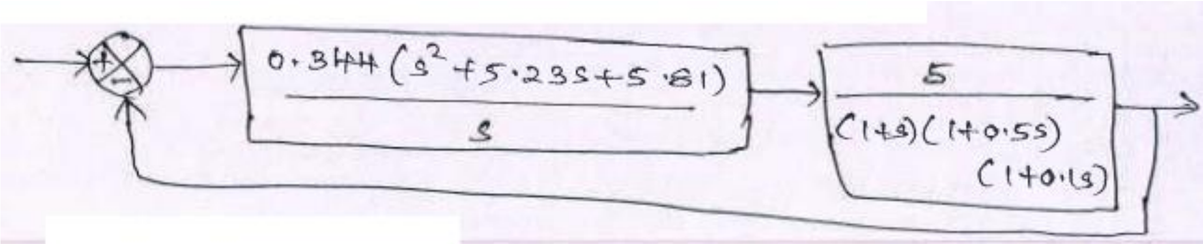
$$k_p = \frac{\cos\theta}{A_1} = \frac{\cos 26}{0.5} = 1.8$$

$$G_c(s) = (k_p + k_d s + \frac{k_i}{s}) = (1.8 + 0.344s + \frac{2}{s})$$

$$= \frac{0.344s^2 + 1.8s + 2}{s} = \frac{0.344(s^2 + \frac{18}{0.344}s + \frac{2}{0.344})}{s}$$

$$= \frac{0.344(s^2 + 5.23s + 5.81)}{s}$$

To find open loop transfer function of compensated system



$$G_0(s) = G_c(s) \times G(s)$$

$$= \frac{0.344(s^2 + 5.23s + 5.81)}{s} \times \frac{5}{(1+s)(1+0.5s)(1+0.1s)}$$

$$= \frac{1.72(s^2 + 5.23s + 5.81)}{s(1+s)(1+0.5s)(1+0.1s)}$$

To verify the design

Put s=jω in G_0(s)

$$G_0(j\omega) = \frac{1.72(-\omega^2 + j5.23\omega + 5.81)}{j\omega(1+j\omega)(1+j0.5\omega)(1+j0.1\omega)}$$

$$=\frac{1.72\sqrt{(5.81-\omega^2)+(5.23\omega^2)}\angle\tan^{-1}\frac{5.23\omega}{5.81-\omega^2}}{\omega\angle 90\sqrt{1+\omega^2}\angle\tan^{-1}\omega\sqrt{1+(0.5\omega)^2}\angle\tan^{-1}0.5\omega\sqrt{1+(0.1\omega)^2}\angle\tan^{-1}0.1\omega}$$

$$A_0=\frac{1.72\sqrt{(5.81-\omega^2)+(5.23\omega^2)}}{\omega\sqrt{1+\omega^2}\sqrt{1+(0.5\omega)^2}\sqrt{1+(0.1\omega)^2}}$$

$$\phi_0=\tan^{-1}\frac{5.23\omega}{5.81-\omega^2}-90-\tan^{-1}\omega-\tan^{-1}0.5\omega-\tan^{-1}0.1\omega for \omega\prec\sqrt{5.81}$$

$$=180^o+\tan^{-1}\frac{5.23\omega}{5.81-\omega^2}-90-\tan^{-1}\omega-\tan^{-1}0.5\omega-\tan^{-1}0.1\omega for \omega\succ\sqrt{5.81}$$

$$At\omega=\omega_1=A_0=A_{01}=\frac{1.72\sqrt{(5.81-4^2)+(5.23\times 4^2)}}{4\sqrt{1+4^2}\sqrt{1+(0.5\times 4)^2}\sqrt{1+(0.14\times 4)^2}}=1$$

$$At\omega=\omega_1,\phi_0=180^o+\tan^{-1}\frac{5.23\times 4}{5.81-4^2}-90-\tan^{-1}4-\tan^{-1}(0.5\times 4)-\tan^{-1}(0.1\times 4)=-135^o$$

$$y_0=180+\phi_{01}=180-135=45^o$$

The phase margin of the compensated system meets the given specification, hence the design is acceptable.

Compensators

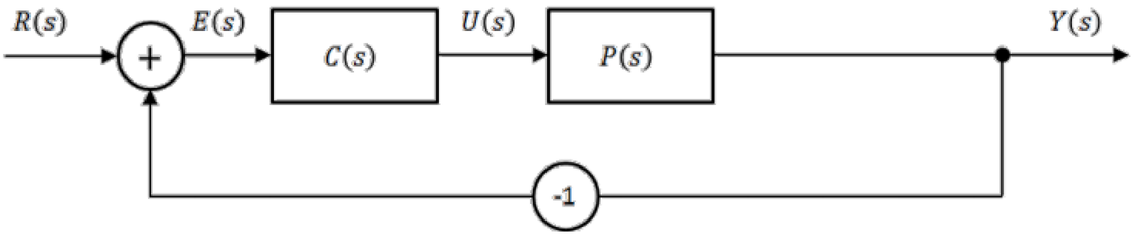
Both a lead compensator and a lag compensator have the same shape:

Lead compensators:

$$C(s)=K\frac{s+\frac{1}{\tau}}{s+\frac{1}{\alpha\tau}} \text{ with } 0<\alpha<1$$

Lag compensators:

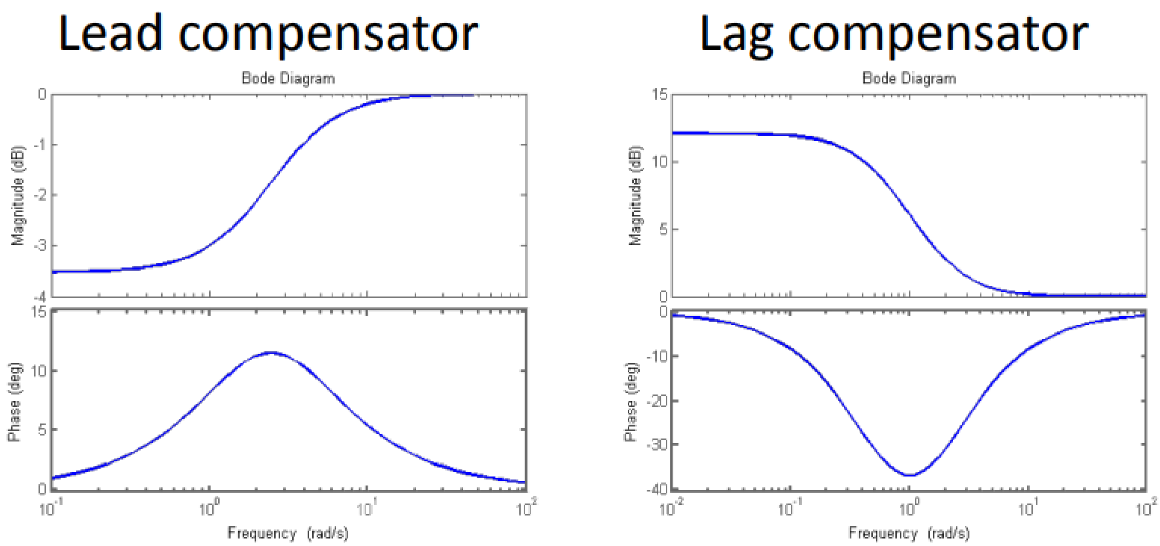
$$C(s)=K\frac{s+\frac{1}{\tau}}{s+\frac{1}{\beta\tau}} \text{ with } \beta>1$$



So they have a zero at $s = -1/\tau$ and a pole at $s = -1/\alpha\tau$ or $-1/\beta\tau$

For lead compensators the pole lies more to the left in the complex plane than the zero and vice versa for lag compensators

Their differences show themselves clearly by comparing their respective Bode plots:



Lead compensators: design with Bode plots

Required increase in phase gain: ϕ

To compensate for increase GCF due to $C(s) \Rightarrow \phi m = \phi + 5°$. This will determine α and τ.

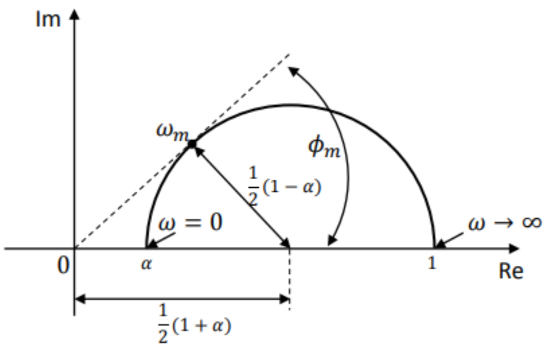
K will be used to tune the steady state error

Determination of α

From the polar plot, we find

$$\sin(\phi_m) = \frac{\frac{1}{2}(1-\alpha)}{\frac{1}{2}(1+\alpha)} = \frac{1-\alpha}{1+\alpha}$$

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



Polar plot of a lead compensator:
$\alpha(j\omega\tau + 1)/(j\omega\alpha\tau + 1)$ where $0 < \alpha < 1$

Determination of τ

From the Bode plot of the lead compensator, the maximal phase is obtained at the frequency of the geometric mean of

$1/\tau$ and $1/\alpha\tau$: $\quad \omega_m = \frac{1}{\sqrt{\alpha}\tau}$

Use the gain crossover frequency of $P(s)\,C(s)$ as ω_m:

$$|P(j\omega_m)C(j\omega_m)| = 1$$

$$|P(j\omega_m)|K\frac{\sqrt{1/\alpha\tau^2 + 1/\tau^2}}{\sqrt{1/\alpha\tau^2 + 1/\alpha^2\tau^2}} = |P(j\omega_m)|K\sqrt{\alpha} = 1$$

$$20\ \log(|P(j\omega_m)|) = -20\ \log(K\sqrt{\alpha})$$

So the value of ω_m can be determined from $P(s)$'s Bode plot,

Determination of K

Remember the steady state error for references of the shape $At^n\varepsilon(t)/n!$, with $\varepsilon(t)$ the step function

The error constants Kp, Kv and Ka as measures for the steady state error for a proportional (n = 0), linear (n = 1) and accelerating (n = 2) reference

So these error constants can be used to find proper values of

K: $\lim_{s\to 0} K\frac{s+\frac{1}{\tau}}{s+\frac{1}{\alpha\tau}}s^nP(s) = K\alpha\lim_{s\to 0} s^nP(s)$

Procedure for design of lead compensators using Bode plot

1. Find $K\alpha$ from steady-state requirement

2. Determine ϕ, the amount with which to increase the Phase Margin(PM); if the PM is OK, then don't need a lead compensator; a proportional controller with gain $K\alpha$ suffices

3. Add 5°, to get $\phi_m = \phi + 5°$ (if $\phi_m > 60°$, then more than one lead compensator)

4. Find α from this ϕ_m: $\alpha = \frac{1-\sin(\phi_m)}{1+\sin(\phi_m)}$, and hence also k

5. Find the desired ω_m by looking at the Bode plot of $P(s)$ and finding the frequency at which the gain equals $-20\log(K\sqrt{\alpha})$ dB

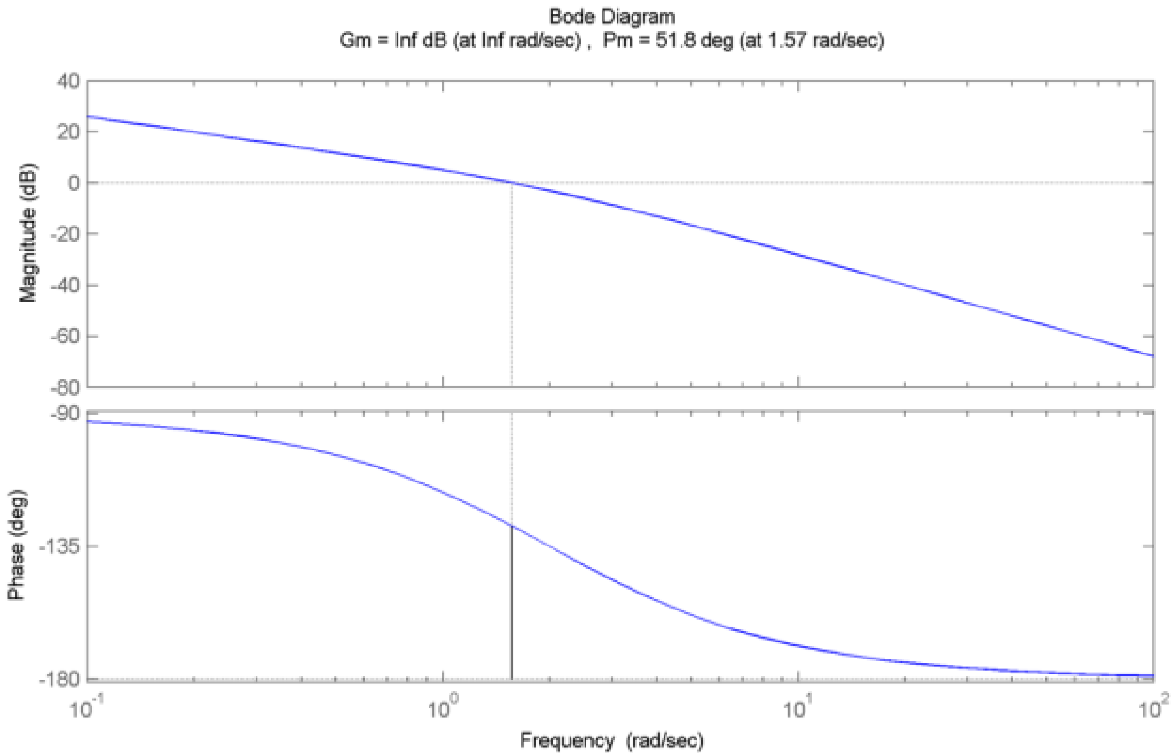
6. find τ as $1/\sqrt{\alpha}\omega_m$

7. Verify if the system behaves as desired.

Example

Given: system $P(s) = 4/s(s + 2)$ at $PM \geq 50°$ and a steady state error for a slope reference of $A/20$.

Solution:



1. Steady-state requirement: $K_v = 20/s$

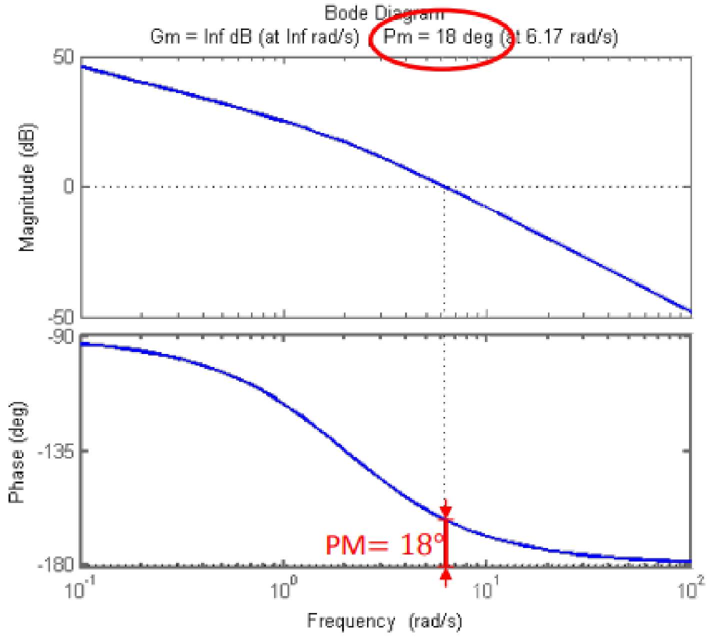
$$\lim_{s\to 0}(sP(s)C(s)) = \lim_{s\to 0}\left(s\frac{4}{s(s+2)}K\alpha\right) = 2K\alpha = 20$$

$$\Rightarrow K\alpha = 10$$

2. Phase margin of $K\alpha P(s) = 18°$
   $\Rightarrow \phi = 32°$

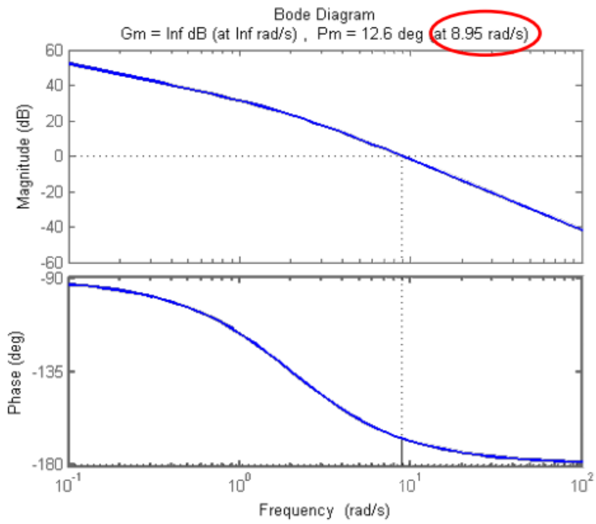
3. $\phi_m = \phi + 5° = 37°$

4. $\alpha = \frac{1-\sin(\phi_m)}{1+\sin(\phi_m)} = 0.24$
   $K = 10/\alpha = 42$



5. find ω_m = the frequency at which the gain is $-20\log(K\sqrt{\alpha})$ dB

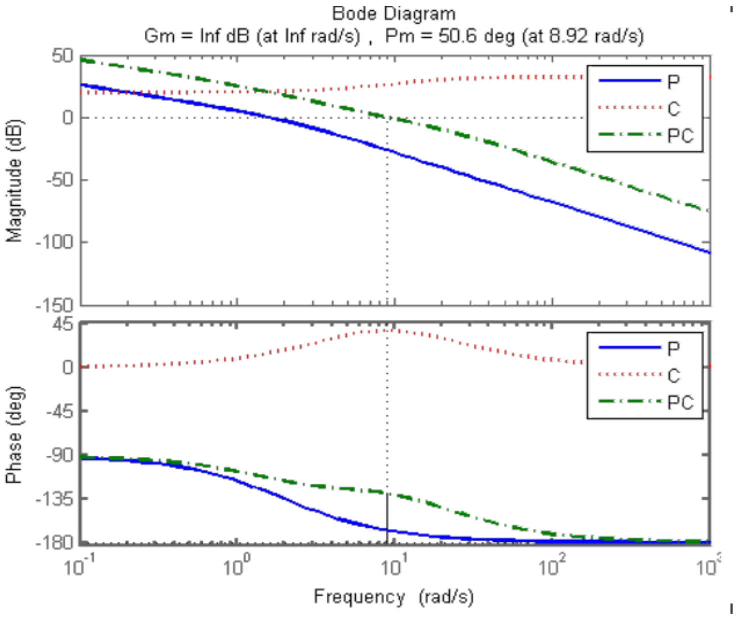
   GCF($P(s)K\sqrt{\alpha}$) = GCF($P(s)C(s)$)
   $\Rightarrow \omega_m = 9 \text{ rad}/s$



6. $\tau = \frac{1}{\omega_m\sqrt{\alpha}} = 0.23$

7. Verify!

- The new PM is indeed ≥ 50°
- The new ω_m is indeed 9 rad/s



Lag compensators:

Increase the stability and tune the steady-state error by increasing the phase at the crossover frequency

Impact lag compensator = lead compensator, but different approach

By decreasing the gain, the gain crossover frequency comes down to a frequency at which the corresponding phase is higher.

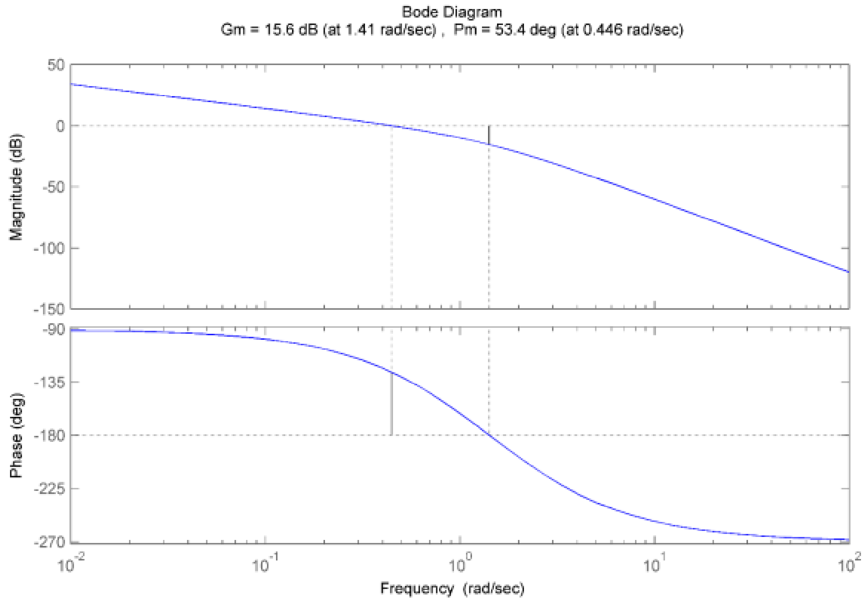
Procedure for design of lag compensators using Bode plot

1. Translate the steady-state requirement into a requirement on lim $s \rightarrow 0$ $C(s) = K\beta$ and verify whether a proportional controller with gain $K\beta$ would suffice
2. Read ω, the frequency at which the phase margin equals $-180° +$ your desired phase margin + 10°, off the Bode diagram; this allows us to compute $\tau = 10/\omega$
3. Read Q, the magnitude at ω off the Bode plot and determine $K = 1/Q$
4. Determine $\beta = K\beta/K$
5. Verify the behaviour of the resulting system

Example:

Given: system $P(s) = 1/s\,(s+1)(s+2)$ for $PM \geq 40°$ and a ramp input results in a steady state error of at most $A/5$, or $K_v = 5$/s

Solution:



1. Steady-state requirement $K_v = 5/s$

   $\lim_{s\to 0}\left(sC(s)P(s)\right) = \frac{1}{2}\lim_{s\to 0} C(s) = 5\ /s$

   $\Rightarrow K\beta = 10$

   Adding a gain of 10 = 20 dB to get the right steady state error, the phase gain would become negative which means the system would become unstable. So Lag compensator is necessary

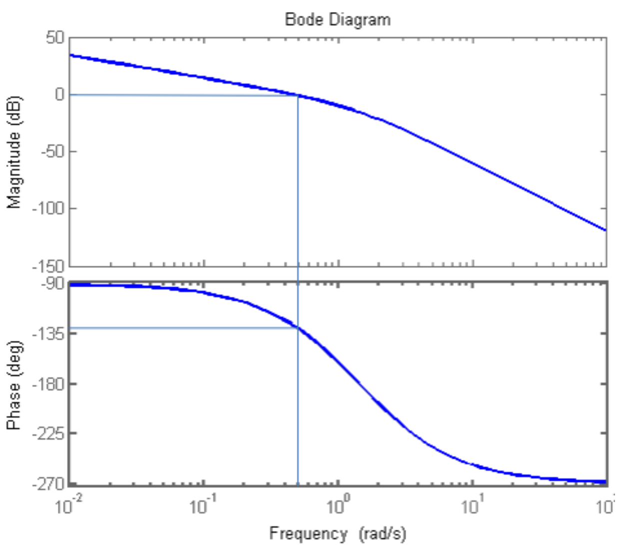
2. Determine ω: the frequency at which the phase equals $-180° + 40° + 10° = -130°$
   $\Rightarrow \omega \cong 0.5$ rad/s $\Rightarrow \tau = 20$

3. We can also read off $Q = 0$ dB $= 1$, which gives us $K = 1/Q = 1$

4. Thus we find $\beta = 10$

   This gives us the following compensator:

   $$C(s) = \frac{s+0.05}{s+0.005}$$



5. Verify!

- The new PM is indeed $\geq 40°$