CHAPTER 12

CLOSED-LOOP CONTROL SYSTEM ANALYSIS

Suhada Jayasuriya
Department of Mechanical Engineering
Texas A&M University
College Station, Texas

1 INTRODUCTION **444**
1.1 Closed-Loop versus Open-Loop Control 445
1.2 Supervisory Control Computers 445
1.3 Hierarchical Control Computers 445
1.4 Direct Digital Control (DDC) 446
1.5 Hybrid Control 446
1.6 Real Systems with Digital Control 447

2 LAPLACE TRANSFORMS **448**
2.1 Single-Sided Laplace Transform 448
2.2 Transforming LTI Ordinary Differential Equations 450
2.3 Transfer Function 451
2.4 Partial-Fraction Expansion and Inverse Transform 452
2.5 Inverse Transform by a General Formula 456

3 BLOCK DIAGRAMS **456**
3.1 Block Diagram Reduction 457
3.2 Transfer Functions of Cascaded Elements 458

4 *z*-TRANSFORMS **459**
4.1 Single-Sided *z*-Transform 461
4.2 Poles and Zeros in the *z*-Plane 461
4.3 *z*-Transforms of Some Elementary Functions 461
4.4 Some Important Properties and Theorems of the *z*-Transform 462
4.5 Pulse Transfer Function 463
4.6 Zero- and First-Order Hold 464

5 CLOSED-LOOP REPRESENTATION **468**
5.1 Closed-Loop Transfer Function 468
5.2 Open-Loop Transfer Function 469
5.3 Characteristic Equation 469
5.4 Standard Second-Order Transfer Function 471
5.5 Step Input Response of a Standard Second-Order System 472
5.6 Effects of an Additional Zero and an Additional Pole 474

6 STABILITY **475**
6.1 Routh–Hurwitz Stability Criterion 475
6.2 Polar Plots 481
6.3 Nyquist Stability Criterion 481

7 STEADY-STATE PERFORMANCE AND SYSTEM TYPE **482**
7.1 Step Input 483
7.2 Ramp Input 487
7.3 Parabolic Input 488
7.4 Indices of Performance 489
7.5 Integral-Square-Error (ISE) Criterion 489
7.6 Integral of Time-Multiplied Absolute-Error (ITAE) Criterion 490
7.7 Comparison of Various Error Criteria 490

8 SIMULATION FOR CONTROL SYSTEM ANALYSIS **490**
8.1 Analog Computation 492
8.2 Digital Computation 498
8.3 Hybrid Computation 501

REFERENCES **501**

BIBLIOGRAPHY **501**

Reprinted from *Instrumentation and Control,* Wiley, New York, 1990, by permission of the publisher.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

1 INTRODUCTION

The field of control has a rich heritage of intellectual depth and practical achievement. From the water clock of Ctesibius in ancient Alexandria, where feedback control was used to regulate the flow of water, to the space exploration and the automated manufacturing plants of today, control systems have played a very significant role in technological and scientific development. James Watt's flyball governor (1769) was essential for the operation of the steam engine, which was, in turn, a technology fueling the Industrial Revolution. The fundamental study of feedback begins with James Clerk Maxwell's analysis of system stability of steam engines with governors (1868). Giant strides in our understanding of feedback and its use in design resulted from the pioneering work of Black, Nyquist, and Bode at Bell Labs in the l920s. Minorsky's work on ship steering was of exceptional practical and theoretical importance. Tremendous advances occurred during World War II in response to the pressing problems of that period. The technology developed during the war years led, over the next 20 years, to practical applications in many fields.

Since the 1960s, there have been many challenges and spectacular achievements in space. The guidance of the Apollo spacecraft on an optimized trajectory from the earth to the moon and the soft landing on the moon depended heavily on control engineering. Today, the shuttle relies on automatic control in all phases of its flight. In aeronautics, the aircraft autopilot, the control of high-performance jet engines, and ascent/descent trajectory optimization to conserve fuel are typical examples of control applications. Currently, feedback control makes it possible to design aircraft that are aerodynamically unstable (such as the X-29) so as to achieve high performance. The National Aerospace Plane will rely on advanced control algorithms to fly its demanding missions.

Control systems are providing dramatic new opportunities in the automotive industry. Feedback controls for engines permit federal emission levels to be met, while antiskid braking control systems provide enhanced levels of passenger safety. In consumer products, control systems are often a critical factor in performance and thus economic success. From simple thermostats that regulate temperature in buildings to the control of the optics for compact disk systems, from garage door openers to the head servos for computer hard disk drives, and from artificial hearts to remote manipulators, control applications have permeated every aspect of life in industrialized societies.

In process control, where systems may contain hundreds of control loops, adaptive controllers have been available commercially since 1983. Typically, even a small improvement in yield can be quite significant economically. Multivariable control algorithms are now being implemented by several large companies. Moreover, improved control algorithms also permit inventories to be reduced, a particularly important consideration in processing dangerous material. In nuclear reactor control, improved control algorithms can have significant safety and economic consequences. In power systems, coordinated computer control of a large number of variables is becoming common. Over 30,000 computer control systems have been installed in the United States alone. Again, the economic impact of control is vast.

Accomplishments in the defense area are legion. The accomplishments range from the antiaircraft gunsights and the bombsights of World War II to the missile autopilots of today and to the identification and estimation techniques used to track and designate multiple targets.

A large body of knowledge has come into existence as a result of these developments and continues to grow at a very rapid rate. A cross section of this body of knowledge is collected in this chapter. It includes analysis tools and design methodologies based on classical techniques. In presenting this material, some familiarity with general control system design principles is assumed. As a result, detailed derivations have been kept to a minimum with appropriate references.

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

1.1 Closed-Loop versus Open-Loop Control

In a closed-loop control system the output is measured and used to alter the control inputs applied to the plant under control. Figure 1 shows such a closed-loop system.

If output measurements are not utilized in determining the plant input, then the plant is said to be under open-loop control. An open-loop control system is shown in Fig. 2.

An open-loop system is very sensitive to any plant parameter perturbations and any disturbances that enter the plant. So an open-loop control system is effective only when the plant and disturbances are exactly known. In real applications, however, neither plants nor disturbances are exactly known a priori. In such situations open-loop control does not provide satisfactory performance. When the loop is closed as in Fig. 1 any plant perturbations or disturbances can be sensed by measuring the outputs, thus allowing plant input to be altered appropriately.

The main reason for closed-loop control is the need for systems to perform well in the presence of uncertainties. It can reduce the sensitivity of the system to plant parameter variations and help reject or mitigate external disturbances. Among other attributes of closed-loop control is the ability to alter the overall dynamics to provide adequate stability and good tracking characteristics. Consequently, a closed-loop system is more complex than an open-loop system due to the required sensing of the output. Sensor noise, however, tends to degrade the performance of a feedback system, thus making it necessary to have accurate sensing. Therefore sensors are usually the most expensive devices in a feedback control system.

1.2 Supervisory Control Computers[1]

With the advent of digital computer technology computers have found their way into control applications. The earliest applications of digital computers in industrial process control were in a so-called supervisory control mode, and the computers were generally large-scale mainframes (minis and micros had not yet been developed). In supervisory control the individual feedback loops are locally controlled by typical analog devices used prior to the installation of the computer. The main function of the computer is to gather information on how the entire process is operating, then feed this into an overall process model located in the computer memory, and then periodically send signals to the set points of individual local analog controllers so that the overall performance of the system is enhanced. A conceptual block diagram of such supervisory control is shown in Fig. 3.

1.3 Hierarchical Control Computers[1]

To continuously achieve optimum overall performance, the function of the supervisory control computer can be broken up into several levels and operated in a hierarchical manner.

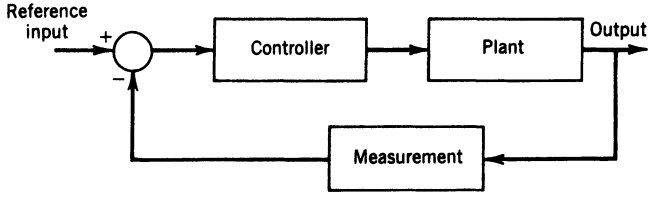


Figure 1 Closed-loop system configuration.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

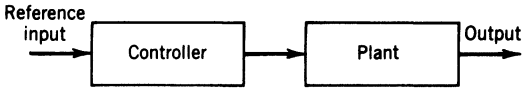


Figure 2 Open-loop system configuration.

Such a multilevel approach becomes useful in controlling very complex processes, with large numbers of control loops with many inputs and controlled variables. In the hierarchical approach the system is subdivided into a hierarchy of simpler control design problems rather than attempting direct synthesis of a single comprehensive controller for the entire process. Thus the controller on a given "level of control" can be less complex due to the existence of lower level controllers that remove frequently occurring disturbances. At each higher level in the hierarchy, the complexity of the control algorithm increases, but the required frequency of execution decreases. Such a system is shown in Fig. 4.

1.4 Direct Digital Control (DDC)

In direct digital control all the analog controllers at the lowest level or intermediate levels are replaced by a central computer serving as a single time-shared controller for all the individual feedback loops. Conventional control modes such as PI or PID (proportional integral differential) are still used for each loop, but the digital versions of the control laws for each loop reside in software in the central computer. In DDC the computer input and output are multiplexed sequentially through the list of control loops, updating each loop's control action in turn and holding this value constant until the next cycle. A typical DDC configuration is shown in Fig. 5.

1.5 Hybrid Control

Combinations of analog and digital methods based on individual loop controllers is known as hybrid control. It is clear that any real control system would have both analog and digital features. For example, in Fig. 5, the plant outputs and inputs are continuous (or analog) quantities. The inputs and outputs of the digital computer are nevertheless digital. The com-

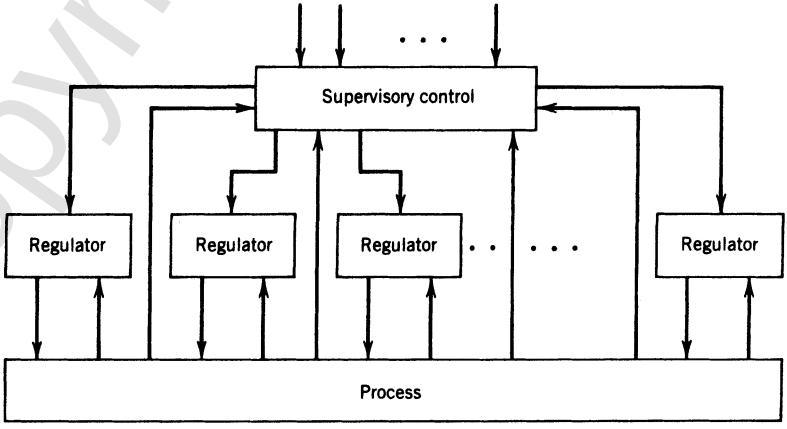


Figure 3 Supervisory control configuration.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

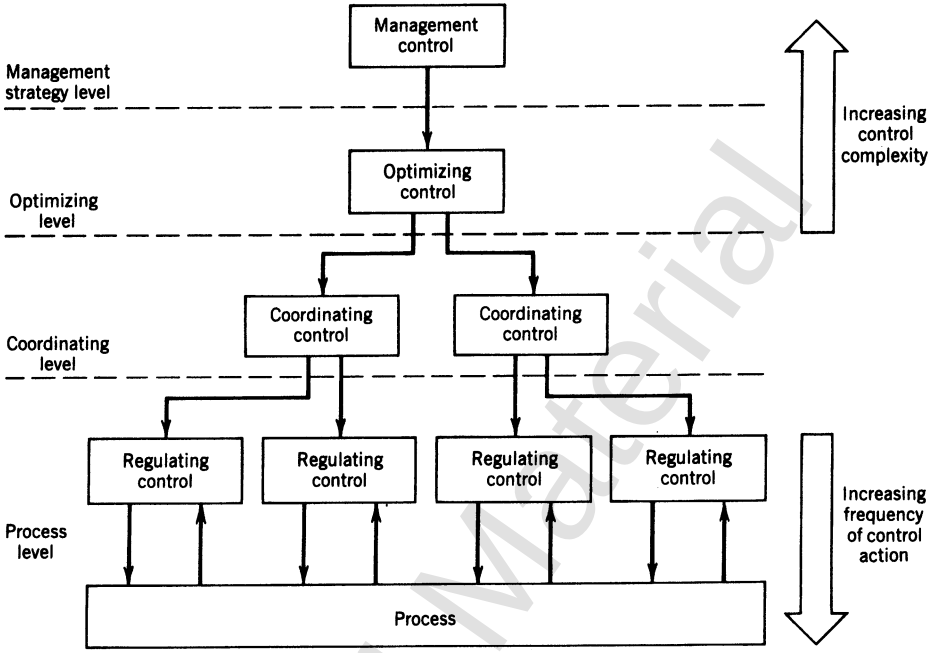


Figure 4 Multilevel hierarchical control structure.

puter processes only numbers. All control algorithms, however, need not be implemented digitally. As an example, for most feedback control loops, the PID algorithm implemented with analog circuitry is superior to the PID algorithm implemented in digital software. The derivative mode, for instance, requires very high resolution to approach the quality of analog performance. Thus it is sometimes advantageous to use hybrid systems for controller implementation.

1.6 Real Systems with Digital Control

Real systems are typically mixtures of analog and digital signals. Typical plant inputs and outputs are analog signals; that is, actuators and sensors are analog systems. If a digital

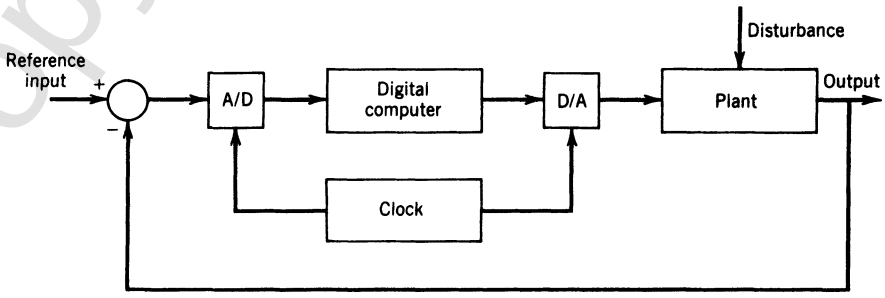


Figure 5 DDC configuration.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

computer is included in the control loop, then digitized quantities are needed for their processing. These are usually accomplished by using analog-to-digital (A/D) converters. To use real-world actuators, computations done as numbers need to be converted to analog signals by employing digital-to-analog (D/A) converters. One of the main advantages of digital control is the ease with which an algorithm can be changed online by altering the software rather than hardware.

2 LAPLACE TRANSFORMS

Often in designing control systems, linear time-invariant (LTI) differential equation representations of physical systems to be controlled are sought. These are typically arrived at by appropriately linearizing the nonlinear equations about some operating point. The general form of such an LTI ordinary differential equation representation is

$$a_n \frac{d^n y(t)}{dt^n} + a_{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + a_1 \frac{dy(t)}{dt} + a_0 y(t)$$
$$= b_m \frac{d^m}{dt^m} u(t) + b_{m-1} \frac{d^{m-1}}{dt^{m-1}} u(t) + \cdots + b_1 \frac{du(t)}{dt} + b_0 u(t) \qquad (1)$$

where $y(t)$ = output of the system
$u(t)$ = input to the system
t = time
a_j, b_j = physical parameters of the system

and $n \geq m$ for physical systems.

The ability to transform systems of the form given by Eq. (1) to algebraic equations relating the input to the output is the primary reason for employing Laplace transform techniques.

2.1 Single-Sided Laplace Transform

The Laplace transform $\mathcal{L}[f(t)]$ of the time function $f(t)$ defined as

$$f(t) = \begin{cases} 0 & t < 0 \\ f(t) & t \geq 0 \end{cases}$$

is given by

$$\mathcal{L}[f(t)] = F(s) = \int_0^\infty e^{-st} f(t)\, dt \qquad t > 0 \qquad (2)$$

where s is a complex variable $(\sigma + j\omega)$. The integral of Eq. (2) cannot be evaluated in a closed form for all $f(t)$, but when it can, it establishes a unique pair of functions, $f(t)$ in the time domain and its companion $F(s)$ in the s domain. It is conventional to use uppercase letters for s functions and lowercase for t functions.

Example 1 Determine the Laplace transform of the unit step function $u_s(t)$:

$$u_s(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}$$

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

By definition

$$\mathcal{L}[u_s(t)] = U_s = \int_0^\infty e^{-ts}1\ dt = \frac{1}{s}$$

Example 2 Determine the Laplace transform of $f(t)$:

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{-\alpha t} & t \geq 0 \end{cases}$$

$$\mathcal{L}[f(t)] = F(s) = \int_0^\infty e^{-ts}e^{-\alpha t}\ dt = \frac{1}{s + \alpha}$$

Example 3 Determine the Laplace transform of the function $f(t)$ given by

$$f(t) = \begin{cases} 0 & t < 0 \\ t & 0 \leq t \leq T \\ T & T \leq t \end{cases}$$

By definition

$$\begin{aligned} F(s) &= \int_0^\infty e^{-ts}f(t)\ dt \\ &= \int_0^T e^{-ts}t\ dt + \int_T^\infty e^{-ts}T\ dt \\ &= \frac{1}{s^2} - \frac{e^{-Ts}}{s^2} = \frac{1}{s^2}\,(1 - e^{-Ts}) \end{aligned}$$

In transforming differential equations, entire equations need to be transformed. Several theorems useful in such transformations are given next without proof.

T1. *Linearity Theorem*

$$\mathcal{L}[\alpha f(t) + \beta g(t)] = \alpha\mathcal{L}[f(t)] + \beta\mathcal{L}[g(t)] \tag{3}$$

T2. *Differentiation Theorem*

$$\mathcal{L}\left[\frac{df}{dt}\right] = sF(s) - f(0) \tag{4}$$

$$\mathcal{L}\left[\frac{d^2f}{dt^2}\right] = s^2F(s) - sf(0) = \frac{df}{dt}\,(0) \tag{5}$$

$$\mathcal{L}\left[\frac{d^nf}{dt^n}\right] = s^nF(s) - s^{n-1}f(0) - s^{n-2}\frac{df}{dt}\,(0) - \cdots - \frac{d^{n-1}}{dt^{n-1}}\,(0) \tag{6}$$

T3. *Translated Function* (Fig. 6)

$$\mathcal{L}[f(t - \alpha)u_s(t - \alpha)] = e^{-\alpha s}F(s + \alpha) \tag{7}$$

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

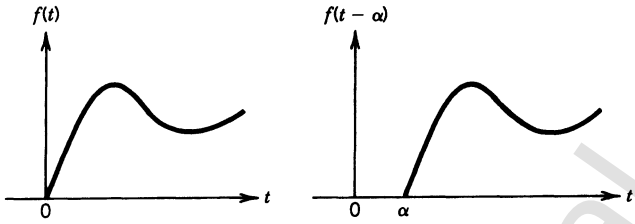


Figure 6 Plots of $f(t)$ and $f(t - \alpha)u_s(t - \alpha)$.

T4. *Multiplication of $f(t)$ by $e^{-\alpha t}$*

$$\mathcal{L}[e^{-\alpha t}f(t)] = F(s + \alpha) \tag{8}$$

T5. *Integration Theorem*

$$\mathcal{L}[\int f(t)\,dt] = \frac{F(s)}{s} + \frac{f^{-1}(0)}{s} \tag{9}$$

where $f^{-1}(0) = \int f(t)\,dt$ evaluated at $t = 0$.

T6. *Final-Value Theorem.* If $f(t)$ and $df(t)/dt$ are Laplace transformable, if $\lim_{t\to\infty} f(t)$ exists, and if $F(s)$ is analytic in the right-half s-plane including the $j\omega$ axis, except for a single pole at the origin, then

$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s) \tag{10}$$

T7. *Initial-Value Theorem.* If $f(t)$ and $df(t)/dt$ are both Laplace transformable, and if $\lim_{s\to\infty} sF(s)$ exists, then

$$f(0) = \lim_{s\to 0} sF(s) \tag{11}$$

Example 4 The time function of Example 3 can be written as

$$f(t) = tu_s(t) - (t - T)u_s(t - T) \tag{12}$$

and

$$\mathcal{L}[f(t)] = \mathcal{L}[tu_s(t)] - \mathcal{L}[(t - T)u_s(t - T)] \tag{13}$$

But

$$\mathcal{L}[tu_s(t)] = \frac{1}{s^2} \tag{14}$$

By using Eqs. (14) and (7) in Eq. (13), we get

$$F(s) = \frac{1}{s^2} - e^{-Ts}\frac{1}{s^2} = \frac{1}{s^2}(1 - e^{-Ts})$$

2.2 Transforming LTI Ordinary Differential Equations

The Laplace transform method yields the complete solution (the particular solution plus the complementary solution) of linear differential equations. Classical methods of finding the

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

complete solution of a differential equation require the evaluation of the integration constants by use of the initial conditions. In the case of the Laplace transform method, initial conditions are automatically included in the Laplace transform of the differential equation. If all initial conditions are zero, then the Laplace transform of the differential equation is obtained simply by replacing d/dt with s, d^2/dt^2 with s^2, and so on.

Consider the differential equation

$$5\frac{d^2y}{dt^2} + 3y = f(t) \tag{15}$$

with $\dot{y}(0) = 1$, $y(0) = 0$. By taking the Laplace transform of Eq. (15), we get

$$\mathcal{L}\left[5\frac{d^2y}{dt^2} + 3y\right] = \mathcal{L}[f(t)]$$

Now using Theorems T1 and T2 (page 449) yields

$$5[s^2Y(s) - sy(0) - y(0)] + 3Y(s) = F(s)$$

Thus

$$Y(s) = \frac{F(s) + 5}{5s^2 + 3} \tag{16}$$

For a given $f(t)$, say $f(t) = u_s(t)$, the unit step function is given as

$$F(s) = \frac{1}{s} \tag{17}$$

Substituting Eq. (17) in Eq. (16) gives

$$Y(s) = \frac{5s + 1}{s(5s^2 + 3)}$$

Then $y(t)$ can be found by computing the inverse Laplace transform of $Y(s)$. That is,

$$y(t) = \mathcal{L}^{-1}[Y(s)] = \mathcal{L}^{-1}\left\{\frac{5s + 1}{s(5s^2 + 3)}\right\}$$

This will be discussed in Section 2.4.

2.3 Transfer Function

The transfer function of a LTI system is defined to be the ratio of the Laplace transform of the output to the Laplace transform of the input under the assumption that all initial conditions are zero.

For the system described by the LTI differential equation (1),

$$\frac{\mathcal{L}[\text{output}]}{\mathcal{L}[\text{input}]} = \frac{\mathcal{L}[y(t)]}{\mathcal{L}[u(t)]} = \frac{Y(s)}{U(s)}$$

$$= \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{18}$$

It should be noted that the transfer function depends only on the system parameters characterized by a_i's and b_i's in Eq. (18). The highest power of s in the denominator of the

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

transfer function defines the order n of the system. The order n of a system is typically greater than or equal to the order of the numerator polynomial m.

Equation (18) can be further written in the form

$$\frac{Y(s)}{U(s)} = G(s) = \frac{b_m \prod_{j=1}^{m} (s + z_i)}{a_n \prod_{j=1}^{n} (s + p_j)} \tag{19}$$

The values of s making Eq. (19) equal to zero are called the system zeros, and the values of s making Eq. (19) go to ∞ are called poles of the system. Hence $s = -z_i$, $i = 1, \ldots, m$, are the system zeros and $s = -p_j$, $j = 1, \ldots, n$, are the system poles.

2.4 Partial-Fraction Expansion and Inverse Transform

The mathematical process of passing from the complex variable expression $F(s)$ to the time-domain expression $f(t)$ is called an inverse transformation. The notation for the inverse Laplace transformation is $\mathcal{L}^{-1}$, so that

$$\mathcal{L}^{-1}[F(s)] = f(t)$$

Example 5 The inverse Laplace transform of $1/s$ is the unit step function:

$$\mathcal{L}[u_s(t)] = \frac{1}{s}$$

Hence

$$\mathcal{L}^{-1}\left(\frac{1}{s}\right) = u_s(t)$$

Time functions for which Laplace transforms are found in a closed form can be readily inverse Laplace transformed by writing them in pairs. Such pairs are listed in Table 1. When transform pairs are not found in tables, other methods have to be used to find the inverse Laplace transform. One such method is the partial-fraction expansion of a given Laplace transform.

If $F(s)$, the Laplace transform of $f(t)$, is broken up into components

$$F(s) = F_1(s) + F_2(s) + \cdots + F_n(s)$$

and if the inverse Laplace transforms of $F_1(s)$, $F_2(s)$, . . . , $F_n(s)$ are readily available, then

$$\begin{aligned} \mathcal{L}^{-1}[F(s)] &= \mathcal{L}^{-1}[F_1(s)] + \mathcal{L}^{-1}[F_2(s)] + \cdots + \mathcal{L}^{-1}[F_n(s)] \\ &= f_1(t) + f_2(t) + \cdots + f_n(t) \end{aligned} \tag{20}$$

where $f_1(t)$, $f_2(t)$, . . . , $f_n(t)$ are the inverse Laplace transforms of $F_1(s)$, $F_2(s)$, . . . , $F_n(s)$, respectively.

For problems in control systems, $F(s)$ is frequently in the form

$$F(s) = \frac{B(s)}{A(s)}$$

where $A(s)$ and $B(s)$ are polynomials in s and the degree of $B(s)$ is equal to or higher than that of $A(s)$.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

Table 1 Laplace Transform Pairs

No.	$f(t)$	$F(s)$
1	Unit impulse $\delta(t)$	1
2	Unit step $u_s(t)$	$\frac{1}{s}$
3	t	$\frac{1}{s^2}$
4	e^{-at}	$\frac{1}{s+a}$
5	te^{-at}	$\frac{1}{(s+a)^2}$
6	$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$
7	$\cos \omega t$	$\frac{s}{s^2+\omega^2}$
8	$t^n \ (n = 1, 2, 3, \ldots)$	$\frac{n!}{s^{n+1}}$
9	$t^n e^{-at} \ (n = 1, 2, 3, \ldots)$	$\frac{n!}{(s+a)^{n+1}}$
10	$\frac{1}{b-a}(e^{-at} - e^{-bt})$	$\frac{1}{(s+a)(s+b)}$
11	$\frac{1}{b-a}(be^{-bt} - ae^{-at})$	$\frac{s}{(s+a)(s+b)}$
12	$\frac{1}{ab}\left[1 + \frac{1}{a-b}(be^{-at} - ae^{-bt})\right]$	$\frac{1}{s(s+a)(s+b)}$
13	$e^{-at} \sin \omega t$	$\frac{\omega}{(s+a)^2+\omega^2}$
14	$e^{-at} \cos \omega t$	$\frac{s+a}{(s+a)^2+\omega^2}$
15	$\frac{1}{a^2}(at - 1 + e^{-at})$	$\frac{1}{s^2(s+a)}$
16	$\frac{\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1-\zeta^2}\, t$	$\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$
17	$\frac{-1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1-\zeta^2}\, t - \phi)$ $\phi = \tan^{-1}\frac{\sqrt{1-\zeta^{*2}}}{\zeta}$	$\frac{s}{s^2+2\zeta\omega_n s+\omega_n^2}$
18	$1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1-\zeta^2}\, t + \phi)$ $\phi = \tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta}$	$\frac{\omega_n^2}{s(s^2+2\zeta\omega_n s+\omega_n^2)}$

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

If $F(s)$ is written as in Eq. (19) and if the poles of $F(s)$ are distinct, then $F(s)$ can always be expressed in terms of simple partial fractions as follows:

$$F(s) = \frac{B(s)}{A(s)} = \frac{\alpha_1}{s + p_1} + \frac{\alpha_2}{s + p_2} + \cdots + \frac{\alpha_n}{s + p_n} \tag{21}$$

where the α_j's are constant. Here α_j is called the residue at the pole $s = -p_j$. The value of α_j is found by multiplying both sides of Eq. (21) by $s + p_j$ and setting $s = -p_j$, giving

$$\alpha_j = \left[(s + p_j) \frac{B(s)}{A(s)} \right]_{s=p_j} \tag{22}$$

Noting that

$$\mathcal{L}^{-1}\left[\frac{1}{s + p_j} \right] = e^{-p_j t}$$

the inverse Laplace transform of Eq. (21) can be written as

$$f(t) = \mathcal{L}^{-1}[F(s)] = \sum_{j=1}^{n} \alpha_j e^{-p_j t} \tag{23}$$

Example 6 Find the inverse Laplace transform of

$$F(s) = \frac{s + 1}{(s + 2)(s + 3)}$$

The partial-fraction expansion of $F(s)$ is

$$F(s) = \frac{s + 1}{(s + 2)(s + 3)} = \frac{\alpha_1}{s + 2} + \frac{\alpha_2}{s + 3}$$

where α_1 and α_2 are found by using Eq. (22) as follows:

$$\alpha_1 = \left[(s + 2) \frac{s + 1}{(s + 2)(s + 3)} \right]_{s=-2} = -1$$

$$\alpha_2 = \left[(s + 3) \frac{s + 1}{(s + 2)(s + 3)} \right]_{s=-3} = 2$$

Thus

$$\begin{aligned} f(t) &= \mathcal{L}^{-1}\left[\frac{-1}{s + 2} \right] + \mathcal{L}^{-1}\left[\frac{2}{s + 3} \right] \\ &= -e^{-2t} + 2e^{-3t} \qquad t \geq 0 \end{aligned}$$

For partial-fraction expansion when $F(s)$ involves multiple poles, consider

$$F(s) = \frac{K \prod_{i=1}^{m} (s + z_i)}{(s + p_1)^r \prod_{j=r+1}^{n} (s + p_j)} = \frac{B(s)}{A(s)}$$

where the pole at $s = -p_1$ has multiplicity r. The partial-fraction expansion of $F(s)$ may then be written as

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$F(s) = \frac{B(s)}{A(s)} = \frac{\beta_1}{s + p_1} + \frac{\beta_2}{(s + p_1)^2} + \cdots + \frac{\beta_r}{(s + p_1)^r} + \sum_{j=r+1}^{n} \frac{\alpha_j}{s + p_j} \tag{24}$$

The α_j's can be evaluated as before using Eq. (22). The β_i's are given by

$$\beta_r = \left[\frac{B(s)}{A(s)}(s + p_1)^r\right]_{s=-p_1}$$

$$\beta_{r-1} = \left\{\frac{d}{ds}\left[\frac{B(s)}{A(s)}(s + p_1)^r\right]\right\}_{s=-p_1}$$

$$\vdots$$

$$\beta_{r-j} = \frac{1}{j!}\left\{\frac{d^j}{ds^j}\left[\frac{B(s)}{A(s)}(s + p_1)^r\right]\right\}_{s=-p_1}$$

$$\vdots$$

$$\beta_1 = \frac{1}{(r-1)!}\left\{\frac{d^{r-1}}{ds^{r-1}}\left[\frac{B(s)}{A(s)}(s + p_1)^r\right]\right\}_{s=-p_1}$$

The inverse Laplace transform of Eq. (24) can then be obtained as follows:

$$f(t) = \left[\frac{\beta_r}{(r-1)!}t^{r-1} + \frac{\beta_{r-1}}{(r-2)!}t^{r-2} + \cdots + \beta_2 t + \beta_1\right]e^{-p_1 t} + \sum_{j=r+1}^{n} \alpha_j e^{-p_1 t} \qquad t \geq 0 \tag{25}$$

Example 7 Find the inverse Laplace transform of the function

$$F(s) = \frac{s^2 + 2s + 3}{(s + 1)^3}$$

Expanding $F(s)$ into partial fractions, we obtain

$$F(s) = \frac{B(s)}{A(s)} = \frac{\beta_1}{s + 1} = \frac{\beta_2}{(s + 1)^2} + \frac{\beta_3}{(s + 1)^3}$$

where β_1, β_2, and β_3 are determined as already shown:

$$\beta_3 = \left[\frac{B(s)}{A(s)}(s + 1)^3\right]_{s=-1} = 2$$

similarly $\beta_2 = 0$ and $\beta_1 = 1$. Thus we get

$$f(t) = \mathcal{L}^{-1}\left[\frac{1}{s + 1}\right] + \mathcal{L}^{-1}\left[\frac{2}{(s + 1)^3}\right]$$

$$= e^{-t} + t^2 e^{-t} \qquad t \geq 0$$

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

2.5 Inverse Transform by a General Formula

Given the Laplace transform $F(s)$ of a time function $f(t)$, the following expression holds true:

$$f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(s)e^{ts}\, ds \tag{26}$$

where c, the abscissa of convergence, is a real constant and is chosen larger than the real parts of all singular points of $F(s)$. Thus the path of integration is parallel to the $j\omega$ axis and is displaced by the amount c from it (see Fig. 7). This path of integration is to the right of all singular points. Equation (26) need be used only when other simpler methods cannot provide the inverse Laplace transformation.

3 BLOCK DIAGRAMS

A control system consists of a number of components. To show the functions performed by each component, we commonly use a diagram called a *block diagram.*

In a block diagram, all system variables are linked to each other through blocks. The block is a symbol for the mathematical operation on the input signal to the block that produces the output. The transfer functions of the components are usually entered in the blocks, with blocks connected by arrows to indicate the direction of the flow of signals. A basic assumption in block diagrams is that there is no loading between blocks. Figure 8 shows a simple block diagram with two blocks in cascade. The arrowheads pointing toward blocks indicate inputs and those pointing away indicate outputs. These arrows are referred to as signals. The following basic components allow the generation of many complex block diagrams.

Addition. Signal addition is represented by a summing point, as shown in Fig. 9.

It should be noted that there can be only one signal leaving a summing point though any number of signals can enter a summing point. For the summing point shown in Fig. 9, we have

$$-x_1 + x_2 + x_3 = y$$

The sign placed near an arrow indicates whether the signal is to be added or subtracted.

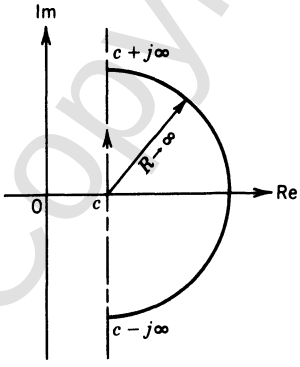


Figure 7 Path of integration.

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

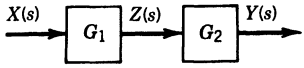


Figure 8 Two blocks in cascade.

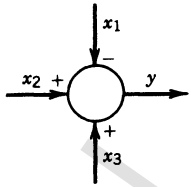


Figure 9 Summing point.

Multiplication. Multiplication is denoted by a symbol as shown in Fig. 10. Here the input X and the output Y are related by the expression

$$Y = GX$$

Takeoff Point. If a signal becomes an input to more than one element, then a takeoff point as shown in Fig. 11 is employed.

A typical block diagram using these elements is shown in Fig. 12.

In the block diagram of Fig. 12 it is assumed that G_2 will have no back reaction (or loading) on G_1; G_1 and G_2 usually represent two physical devices. If there are any loading effects between the devices, it is necessary to combine these components into a single block. Such a situation is given in Section 3.2.

3.1 Block Diagram Reduction

Any number of cascaded blocks representing nonloading components can be replaced by a single block, the transfer function of which is simply the product of the individual transfer functions. For example, the two elements in cascade shown in Fig. 12 can be replaced by a single element $G_0 = G_1G_2$. Some fundamental block diagram reduction rules are shown in Fig. 13.

In simplifying a block diagram a general rule is to first move takeoff points and summing points to form internal feedback loops as shown in Fig. 13*d*. Then remove the internal feedback loops and proceed in the same manner.

Example 8 Reduce the block diagram shown in Fig. 14.

First, to eliminate the loop $G_3G_4H_1$, we move H_2 behind block G_4 and therefore obtain Fig. 15*a*. Eliminating the loop $G_3G_4H_1$, we obtain Fig. 15*b*. Then eliminating the inner loop containing H_2/G_4, we obtain Fig. 15*c*. Finally, by reducing the loop containing H_3, we obtain the closed-loop system transfer function as shown in Fig. 15*d*.

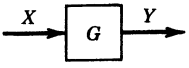


Figure 10 Multiplication $Y = GX$.

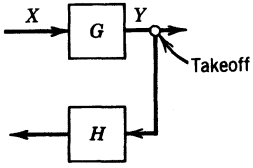


Figure 11 Takeoff point.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

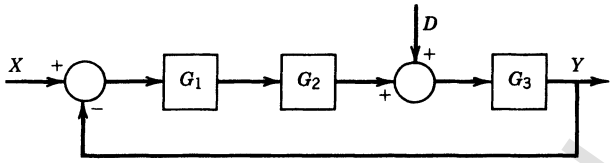


Figure 12 Typical block diagram.

3.2 Transfer Functions of Cascaded Elements[2]

Many feedback systems have components that load each other. Consider the system of Fig. 16. Assume that e_i is the input and e_o is the output. In this system the second stage of the circuit (R_2C_2 portion) produces a loading effect on the first stage (R_1C_1 portion). The equations for this system are

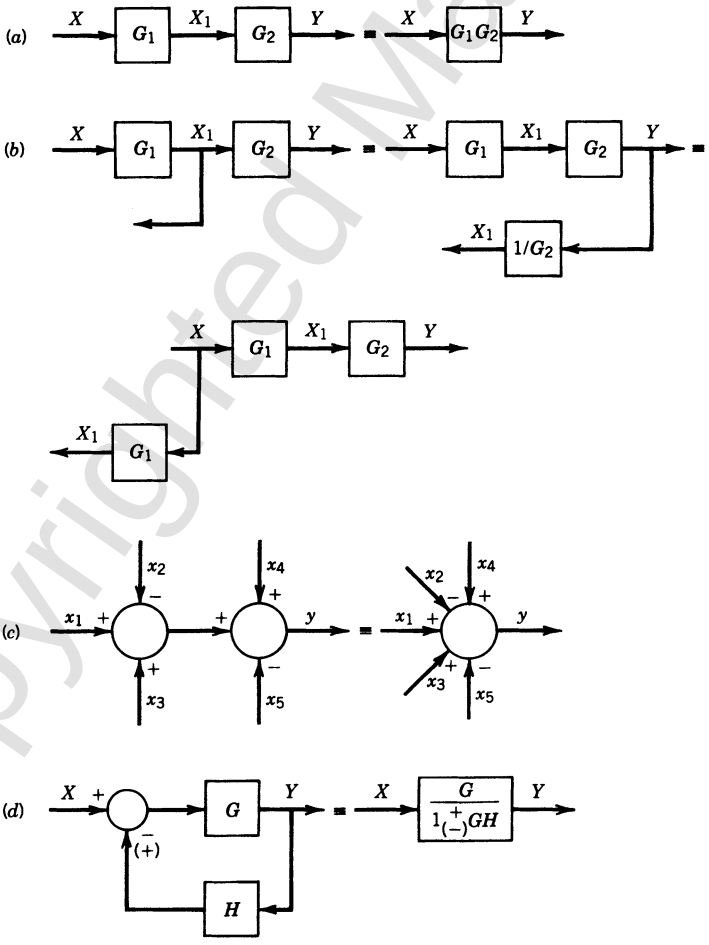


Figure 13 Basic block diagram reduction rules.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

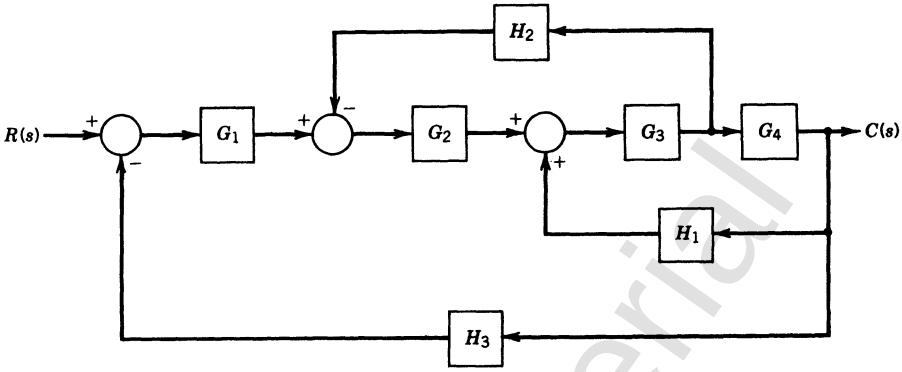


Figure 14 Multiple-loop feedback control system.

$$\frac{1}{C_1}\int (i_1 - i_2)\,dt + R_1 i_1 = e_i \tag{27}$$

and

$$\frac{1}{C_1}\int (i_1 - i_2)\,dt + R_2 i_2 = -\frac{1}{C_2}\int i_2\,dt = -e_o \tag{28}$$

Taking the Laplace transform of Eqs. (27) and (28), assuming zero initial conditions and simplifying yield

$$\frac{E_o(s)}{E_i(s)} = \frac{1}{R_1C_1R_2C_2s^2 + (R_1C_1 + R_2C_2 + R_1C_2)s + 1} \tag{29}$$

The term R_1C_2s in the denominator of the transfer function represents the interaction of two simple RC circuits.

This analysis shows that if two RC circuits are connected in cascade so that the output from the first circuit is the input to the second, the overall transfer function is not the product of $1/(R_1C_1s + 1)$ and $1/(R_2C_2s + 1)$. The reason for this is that when we derive the transfer function for an isolated circuit, we implicitly assume that the output is unloaded. In other words, the load impedance is assumed to be infinite, which means that no power is being withdrawn at the output. When the second circuit is connected to the output of the first, however, a certain amount of power is withdrawn and then the assumption of no loading is violated. Therefore, if the transfer function of this system is obtained under the assumption of no loading, then it is not valid. Chapter 2 deals with this type of problem in greater detail.

4 z-TRANSFORMS

One of the mathematical tools commonly used to deal with discrete-time systems is the z-transform. The role of the z-transform in discrete-time systems is similar to that of the Laplace transform in continuous-time systems. Laplace transforms allow the conversion of linear ordinary differential equations with constant coefficients into algebraic equations in s. The z-transformation transforms linear difference equations with constant coefficients into algebraic equations in z.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

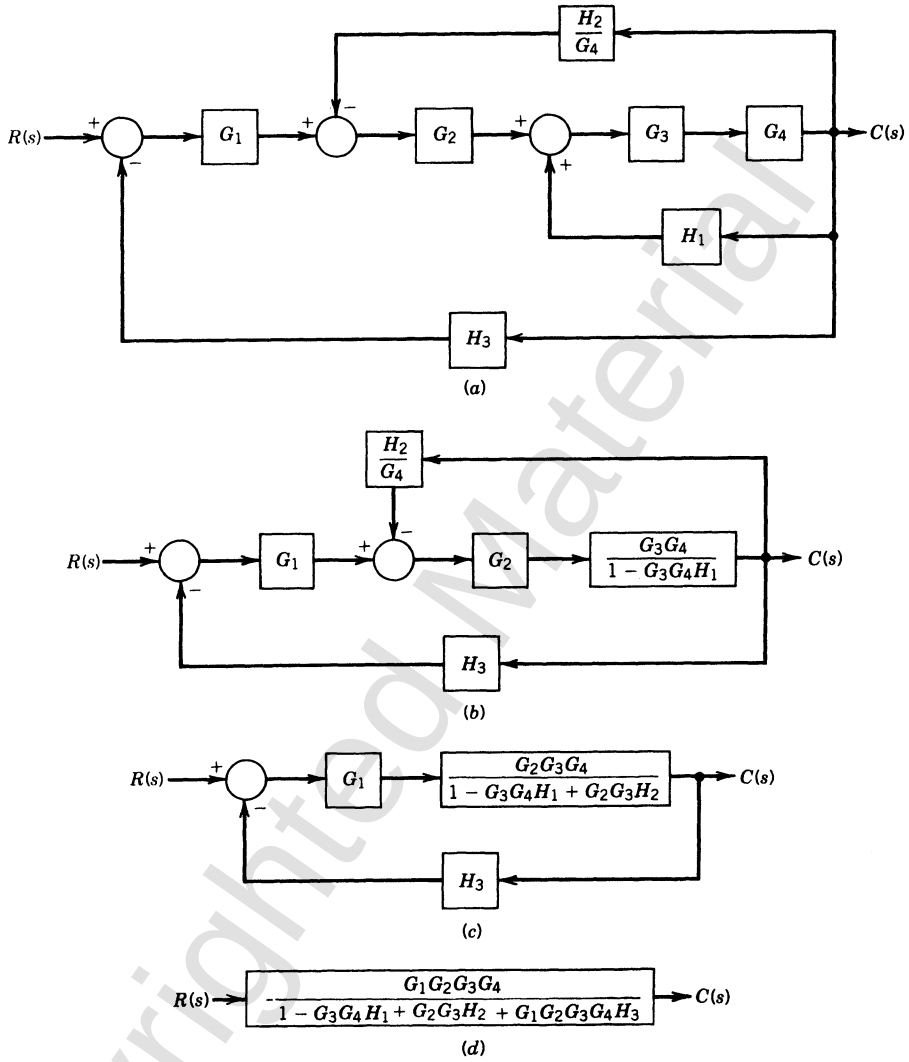


Figure 15 Block diagram reduction of the system of Fig. 14.

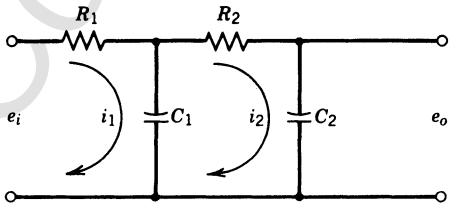


Figure 16 Electrical system.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

4.1 Single-Sided z-Transform

If a signal has discrete values $f_0, f_1, \ldots, f_k, \ldots$, we define the z-transform of the signal as the function

$$F(z) = \mathfrak{Z}\{f(k)\}$$

$$= \sum_{k=0}^{\infty} f(k)z^{-k} \qquad r_0 \le |z| \le R_0 \tag{30}$$

and is assumed that one can find a range of values of the magnitude of the complex variable z for which the series of Eq. (30) converges. This z-transform is referred to as the one-sided z-transform. The symbol $\mathfrak{Z}$ denotes "the z-transform of." In the one-sided z-transform it is assumed $f(k) = 0$ for $k < 0$, which in the continuous-time case corresponds to $f(t) = 0$ for $t < 0$.

Expansion of the right-hand side of Eq. (30) gives

$$F(z) = f(0) + f(1)z^{-1} + f(2)z^{-2} + \cdots + f(k)z^{-k} + \cdots \tag{31}$$

The last equation implies that the z-transform of any continuous-time function $f(t)$ may be written in the series form by inspection. The z^{-k} in this series indicates the instant in time at which the amplitude $f(k)$ occurs. Conversely, if $F(z)$ is given in the series form of Eq. (31), the inverse z-transform can be obtained by inspection as a sequence of the function $f(k)$ that corresponds to the values of $f(t)$ at the respective instants of time. If the signal is sampled at a fixed sampling period T, then the sampled version of the signal $f(t)$ given by $f(0), f(1), \ldots, f(k)$ correspond to the signal values at the instants $0, T, 2T, \ldots, kT$.

4.2 Poles and Zeros in the z-Plane

If the z-transform of a function takes the form

$$F(z) = \frac{b_0 z^m + b_1 z^{m-1} + \cdots + b_m}{z^n + a_1 z^{n-1} + \cdots + a_n} \tag{32}$$

or

$$F(z) = \frac{b_0(z - z_1)(z - z_2) \cdots (z - z_m)}{(z - p_1)(z - p_2) \cdots (z - p_n)}$$

then p_i's are the poles of $F(z)$ and z_i's are the zeros of $F(z)$.

In control studies, $F(z)$ is frequently expressed as a ratio of polynomials in z^{-1} as follows:

$$F(z) = \frac{b_0 z^{-(n-m)} + b_1 z^{-(n-m+1)} + \cdots + b_m z^{-n}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}} \tag{33}$$

where z^{-1} is interpreted as the unit delay operator.

4.3 z-Transforms of Some Elementary Functions

Unit Step Function

Consider the unit step function

$$u_s(t) = \begin{cases} 1 & t \ge 0 \\ 0 & t < 0 \end{cases}$$

whose discrete representation is

Copyrighted material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$u_s(k) = \begin{cases} 1 & k \geq 0 \\ 0 & k < 0 \end{cases}$$

From Eq. (30)

$$\mathfrak{Z}\{u_s(k)\} = \sum_{k=0}^{\infty} 1 \cdot z^{-k} = \sum_{k=0}^{\infty} z^{-k} = 1 + z^{-1} + z^{-2} + \cdots$$

$$= \frac{1}{1 - z^{-1}} = \frac{z}{z - 1} \tag{34}$$

Note that the series converges for $|z| > 1$. In finding the z-transform, the variable z acts as a dummy operator. It is not necessary to specify the region of z over which $\mathfrak{Z}\{u_s(k)\}$ is convergent. The z-transform of a function obtained in this manner is valid throughout the z-plane except at the poles of the transformed function. The $u_s(k)$ is usually referred to as the unit step sequence.

Exponential Function

Let

$$f(t) = \begin{cases} e^{-at} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

The sampled form of the function with sampling period T is

$$f(kT) = e^{-akT} \qquad k = 0, 1, 2, \ldots$$

By definition

$$F(z) = \mathfrak{Z}\{f(k)\}$$

$$= \sum_{k=0}^{\infty} f(k) z^{-k}$$

$$= \sum_{k=0}^{\infty} e^{-akt} z^{-k}$$

$$= 1 + e^{-at} z^{-1} + e^{-2at} z^{-2} + \cdots$$

$$= \frac{1}{1 - e^{-at} z^{-1}}$$

$$= \frac{z}{z - e^{-at}}$$

4.4 Some Important Properties and Theorems of the z-Transform

In this section some useful properties and theorems are stated without proof. It is assumed that the time function $f(t)$ is z-transformable and that $f(t)$ is zero for $t < 0$.

P1. If

$$F(z) = \mathfrak{Z}\{f(k)\}$$

then

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$\mathfrak{Z}\{af(k)\} = aF(z) \tag{35}$$

P2. If $f_1(k)$ and $g_1(k)$ are z-transformable and α and β are scalars, then $f(k) = \alpha f_1(k) + \beta g_1(k)$ has the z-transform

$$F(z) = \alpha F_1(z) + \beta G_1(z) \tag{36}$$

where $F_1(z)$ and $G_1(z)$ are the z-transforms of $f_1(k)$ and $g_1(k)$, respectively.

P3. If

$$F(z) = \mathfrak{Z}\{f(k)\}$$

then

$$\mathfrak{Z}\{a^k f(k)\} = F\left(\frac{z}{a}\right) \tag{37}$$

T1. *Shifting Theorem.* If $f(t) \equiv 0$ for $t < 0$ and $f(t)$ has the z-transform $F(z)$, then

$$\mathfrak{Z}\{f(t + nT)\} = z^{-n}F(z)$$

and

$$\mathfrak{Z}\{f(t + nT)\} = z^n\left[F(z) - \sum_{k=0}^{n-1} f(kT)z^{-k}\right] \tag{38}$$

T2. *Complex Translation Theorem.* If

$$F(z) = \mathfrak{Z}\{f(t)\}$$

then

$$\mathfrak{Z}\{e^{-at}f(t)\} = F(ze^{at}) \tag{39}$$

T3. *Initial-Value Theorem.* If $F(z) = \mathfrak{Z}\{f(t)\}$ and if $\lim_{z\to\infty} F(z)$ exists, then the initial value $f(0)$ of $f(t)$ or $f(k)$ is given by

$$f(0) = \lim_{z\to\infty} F(z) \tag{40}$$

T4. *Final-Value Theorem.* Suppose that $f(k)$, where $f(k) \equiv 0$ for $k < 0$, has the z-transform $F(z)$ and that all the poles of $F(z)$ lie inside the unit circle, with the possible exception of a simple pole at $z = 1$. Then the final value of $f(k)$ is given by

$$\lim_{k\to\infty} f(k) = \lim_{z\to 1} [(1 - z^{-1})F(z)] \tag{41}$$

4.5 Pulse Transfer Function

Consider the LTI discrete-time system characterized by the following linear difference equation:

$$y(k) + a_1 y(k - 1) + \cdots + a_n y(k - n) = b_0 u(k) + b_1 u(k - 1) + \cdots + b_m u(k - m) \tag{42}$$

where $u(k)$ and $y(k)$ are the system's input and output, respectively, at the kth sampling or at the real time kT; T is the sampling period. To convert the difference equation (42) to an algebraic equation, take the z-transform of both sides of Eq. (42) by definition:

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$\mathcal{Z}\{y(k)\} = Y(z) \tag{43a}$$

$$\mathcal{Z}\{u(k)\} = U(z) \tag{43b}$$

By referring to Table 2, the z-transform of Eq. (42) becomes

$$Y(z) + a_1 z^{-1}(Y(z) + \cdots + a_n z^{-n} y(z) = b_0 U(z) + b_1 z^{-1} U(z) + \cdots + b_m z^{-m} U(z)$$

or

$$[1 + a_1 z^{-1} + \cdots + a_n z^{-1}]Y(z) = [b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}]U(z)$$

which can be written as

$$\frac{Y(z)}{U(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}} \tag{44}$$

Consider the response of the linear discrete-time system given by Eq. (44), initially at rest when the input $u(t)$ is the delta "function" $\delta(kT)$,

$$\delta(kT) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

since

$$\mathcal{Z}\{\delta(kT)\} = \sum_{k=0}^{\infty} \delta(kT) z^{-k} = 1$$

$$U(z) = \mathcal{Z}\{\delta(kT)\} = 1$$

and

$$Y(z) = \frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}} = G(z) \tag{45}$$

Thus $G(z)$ is the response of the system to the delta input (or unit impulse) and plays the same role as the transfer function in linear continuous-time systems. The function $G(z)$ is called the pulse transfer function.

4.6 Zero- and First-Order Hold

Discrete-time control systems may operate partly in discrete time and partly in continuous time. Replacing a continuous-time controller with a digital controller necessitates the conversion of numbers to continuous-time signals to be used as true actuating signals. The process by which a discrete-time sequence is converted to a continuous-time signal is called data hold.

In a conventional sampler, a switch closes to admit an input signal every sample period T. In practice, the sampling duration is very small compared with the most significant time constant of the plant. Suppose the discrete-time sequence is $f(kT)$; then the function of the data hold is to specify the values for a continuous equivalent $h(t)$ where $kT \leq t < (k + 1)T$. In general, the signal $h(t)$ during the time interval $kT < t < (k + 1)T$ may be approximated by a polynomial in τ as follows:

$$h(kT + \tau) = a_n \tau^n + a_{n-1} \tau^{n-1} + \cdots + a_1 \tau + a_0 \tag{46}$$

where $0 \leq \tau < T$. Since the value of the continuous equivalent must match at the sampling instants, one requires

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

Table 2 Table of z-Transforms[a]

No.	$\mathcal{F}(s)$	$f(nT)$	$F(z)$
1	—	$1,\ n = 0;\ 0,\ n \neq 0$	1
2	—	$1,\ n = k;\ 0,\ n \neq k$	z^{-k}
3	$\dfrac{1}{s}$	$1(nT)$	$\dfrac{z}{z-1}$
4	$\dfrac{1}{s^2}$	nT	$\dfrac{Tz}{(z-1)^2}$
5	$\dfrac{1}{s^3}$	$\dfrac{1}{2!}(nT)^2$	$\dfrac{T^2}{2}\dfrac{z(z+1)}{(z-1)^3}$
6	$\dfrac{1}{s^4}$	$\dfrac{1}{3!}(nT)^3$	$\dfrac{T^3}{6}\dfrac{z(z^2+4z+1)}{(z-1)^4}$
7	$\dfrac{1}{s^m}$	$\lim\limits_{a\to 0}\dfrac{(-1)^{m-1}}{(m-1)!}\dfrac{\delta^{m-1}}{\delta a^{m-1}}e^{-unT}$	$\lim\limits_{a\to 0}\dfrac{(-1)^{m-1}}{(m-1)!}\dfrac{\delta^{m-1}}{\delta a^{m-1}}\dfrac{z}{z-e^{-aT}}$
8	$\dfrac{1}{s+a}$	e^{-anT}	$\dfrac{z}{z-e^{-aT}}$
9	$\dfrac{1}{(s+a)^2}$	nTe^{-anT}	$\dfrac{Tze^{-aT}}{(z-e^{-aT})^2}$
10	$\dfrac{1}{(s+a)^3}$	$\dfrac{1}{2}(nT)^2e^{-anT}$	$\dfrac{T^2}{2}(e^{-aT})\dfrac{z(z+e^{-aT})}{(z-e^{-aT})^3}$
11	$\dfrac{1}{(s+a)^m}$	$\dfrac{(-1)^{m-1}}{(m-1)!}\dfrac{\delta^{m-1}}{\delta a^{m-1}}(e^{-anT})$	$\dfrac{(-1)^{m-1}}{(m-1)!}\dfrac{\delta^{m-1}}{\delta a^{m-1}}\dfrac{z}{z-e^{-aT}}$
12	$\dfrac{a}{s(s+a)}$	$1-e^{-anT}$	$\dfrac{z(1-e^{-aT})}{(z-1)(z-e^{-aT})}$
13	$\dfrac{a}{s^2(s+a)}$	$\dfrac{1}{a}(anT-1+e^{-anT})$	$\dfrac{z[(aT-1+e^{-aT})z+(1-e^{-aT}-aTe^{-aT})]}{a(z-1)^2(z-e^{-aT})}$
14	$\dfrac{b-a}{(s+a)(s+b)}$	$(e^{-anT}-e^{-bnT})$	$\dfrac{(e^{-aT}-e^{-bT})z}{(x-e^{-aT})(z-e^{-bT})}$
15	$\dfrac{s}{(s+a)^2}$	$(1-anT)e^{-anT}$	$\dfrac{z[z-e^{-aT}(1+aT)]}{(z-e^{-aT})^2}$
16	$\dfrac{a^2}{s(s+a)^2}$	$1-e^{-anT}(1+anT)$	$\dfrac{z[z(1-e^{-aT}-aTe^{-aT})+e^{-2aT}-e^{-aT}+aTe^{-aT}]}{(z-1)(z-e^{-aT})^2}$
17	$\dfrac{(b-a)s}{(s+a)(s+b)}$	$be^{-bnT}-ae^{-anT}$	$\dfrac{z[z(b-a)-(be^{-aT}-ae^{-bT})]}{(z-e^{-aT})(z-e^{-bT})}$
18	$\dfrac{a}{s^2+a^2}$	$\sin anT$	$\dfrac{z\sin aT}{z^2-(2\cos aT)z+1}$
19	$\dfrac{s}{s^2+a^2}$	$\cos anT$	$\dfrac{z(z-\cos aT)}{z^2-(2\cos aT)z+1}$
20	$\dfrac{s+a}{(s+a)^2+b^2}$	$e^{-anT}\cos bnT$	$\dfrac{z(z-e^{-aT}\cos bT)}{z^2-2e^{-aT}(\cos bT)z+e^{-2eT}}$
21	$\dfrac{b}{(s+a)^2+b^2}$	$e^{-anT}\sin bnT$	$\dfrac{ze^{-aT}\sin bT}{z^2-2e^{-aT}(\cos bT)z+e^{-2aT}}$
22	$\dfrac{a^2+b^2}{s((s+a)^2+b^2)}$	$1-e^{-anT}\left(\cos bnT+\dfrac{a}{b}\sin bnT\right)$	$\dfrac{z(Az+B)}{(z-1)(z^2-2e^{-aT}(\cos bT)z+e^{-2aT})}$ $A = 1-e^{-aT}\cos bT-\dfrac{a}{b}e^{-aT}\sin bT$ $B = e^{-2aT}+\dfrac{a}{b}e^{-aT}\sin bT-e^{-aT}\cos bT$

[a] $\mathcal{F}(s)$ is the Laplace transform of $f(t)$ and $F(z)$ is the transform of $f(nT)$. Unless otherwise noted, $f(t) = 0$, $t < 0$, and the region of convergence of $F(z)$ is outside a circle $r < |z|$ such that all poles of $F(z)$ are inside r.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$h(kT) = f(kT)$$

Hence Eq. (46) can be written as

$$h(kT + \tau) = a_n\tau^n + a_{n-1}\tau^{n-1} + \cdots + a_1\tau + f(kT) \tag{47}$$

If an nth-order polynomial as in Eq. (47) is used to extrapolate the data, then the hold circuit is called an nth-order hold. If $n = 1$, it is called a first-order hold (the nth-order hold uses the past $n + 1$ discrete data $f[(k - n)T]$). The simplest data hold is obtained when $n = 0$ in Eq. (47), that is, when

$$h(kT + \tau) = f(kT) \tag{48}$$

where $0 \leq \tau < T$ and $k = 0, 1, 2, \ldots$. Equation (48) implies that the circuit holds the amplitude of the sample from one sampling instant to the next. Such a data hold is called a zero-order hold. The output of the zero-order hold is shown in Fig. 17.

Zero-Order Hold

Assuming that the sampled signal is 0 for $k < 0$, the output $h(t)$ can be related to $f(t)$ as follows:

$$\begin{aligned} h(t) &= f(0)[u_s(t) - u_s(t - T)] + f(T)[u_s(t - T) - u_s(t - 2T)] \\ &\quad + f(2T)[u_s(t - 2T)] - u_s(t - 3T)] + \cdots \\ &= \sum_{k=0}^{\infty} f(kT)\{u_s(t - kT) - u_s[t - (k + 1)T]\} \end{aligned} \tag{49}$$

Since $\mathcal{L}|u_s(t - kT)| = e^{-kTs}/s$, the Laplace transform of Eq. (49) becomes

$$\begin{aligned} \mathcal{L}[h(t)] = H(s) &= \sum_{k=0}^{\infty} f(kT)\,\frac{e^{-kTs} - e^{-(k+1)Ts}}{s} \\ &= \frac{1 - e^{-Ts}}{s} \sum_{k=0}^{\infty} f(kT)e^{-kTs} \end{aligned} \tag{50}$$

The right-hand side of Eq. (50) may be written as the product of two terms:

$$H(s) = G_{h0}(s)F^*(s) \tag{51}$$

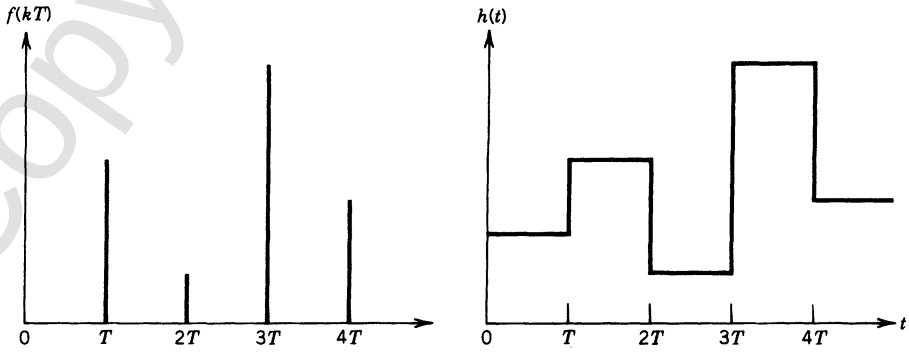


Figure 17 Input $f(kt)$ and output $h(t)$ of the zero-order hold.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

where

$$G_{h0}(s) = \frac{1 - e^{-Ts}}{s}$$

$$F^*(s) = \sum_{k=0}^{\infty} f(kT)e^{-kTs} \tag{52}$$

In Eq. (51), $G_{h0}(s)$ may be considered the transfer function between the output $H(s)$ and the input $F^*(s)$ (see Fig. 18). Thus the transfer function of the zero-order hold device is

$$G_{h0}(s) = \frac{1 - e^{-Ts}}{s} \tag{53}$$

First-Order Hold

The transfer function of a first-order hold is given by

$$G_{h1}(s) = \frac{1 - e^{-Ts}}{s}\frac{Ts + 1}{T} \tag{54}$$

From Eq. (47) for $n = 1$,

$$h(kT + \tau) = a_1\tau + f(kT) \tag{55}$$

where $0 \leq \tau < T$ and $k = 0, 1, 2, \ldots$. By using the condition $h[(k - 1)T] = f[(k - 1)T]$ the constant a_1 can be determined as follows:

$$h[(k - 1)T] = -a_1T + f(kT) = f[(k - 1)T]$$

or

$$a_1 = \frac{f(kT)_f(k - 1)T]}{T}$$

Hence Eq. (55) becomes

$$h(kT + \tau) = f(kT) + \frac{f(kT) - f[(k - 1)T]}{T}\tau \tag{56}$$

where $0 \leq \tau < T$. The extrapolation process of the first-order hold is based on Eq. (56) and is a piecewise-linear signal as shown in Fig. 19.

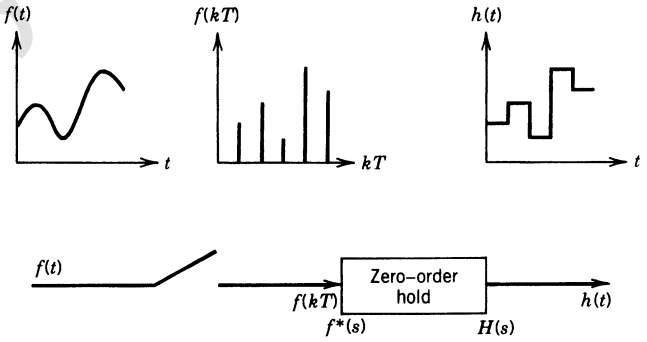


Figure 18 Sampler and zero-order hold.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

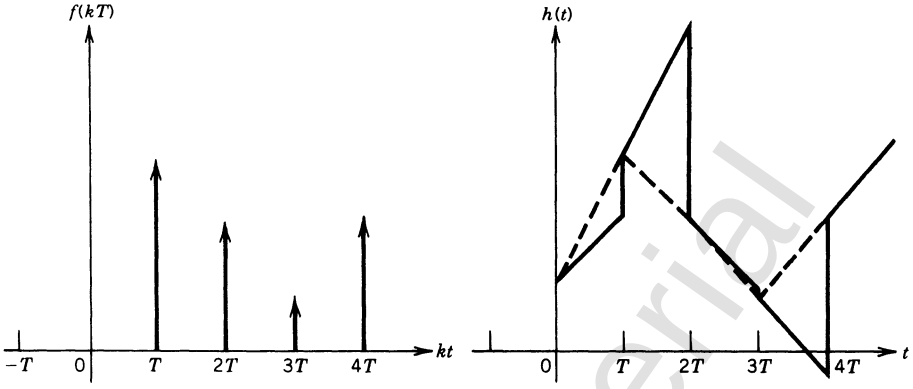


Figure 19 Input $f(kt)$ and output $h(t)$ of a first-order hold.

5 CLOSED-LOOP REPRESENTATION

The typical feedback control system has the feature that some output quantity is measured and then compared with a desired value, and the resulting error is used to correct the system output. A block diagram representation of a closed-loop or feedback system is shown in Fig. 20.

In this figure, r is the reference input, w is a disturbance, and y is the output. Transfer functions G_p, H, and G_c denote, respectively, the plant dynamics, sensor dynamics, and controller. The influence of r and w on the output y can be determined using elementary block diagram algebra as

$$Y(s) = \frac{G_c G_p}{1 + G_c G_p H} R(s) + \frac{G_p}{1 + G_c G_p H} W(s) \tag{57}$$

where $R(s) = \mathcal{L}[r(t)]$, $W(s) = \mathcal{L}[w(t)]$, and $Y(s) = \mathcal{L}[y(t)]$.

5.1 Closed-Loop Transfer Function

In Eq. (57), if $W(s) = 0$, then

$$Y(s) = \frac{G_c G_p}{1 + G_c G_p H} R(s)$$

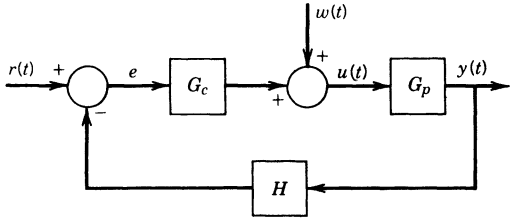


Figure 20 Closed-loop system with disturbance.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

or alternatively a transfer function called the closed-loop transfer function between the reference input and the output is defined:

$$\frac{Y(s)}{R(s)} = G_{\text{cl}}(s) = \frac{G_c G_p}{1 + G_c G_p H} \tag{58}$$

5.2 Open-Loop Transfer Function

The product of transfer functions within the loop, namely $G_c G_p H$, is referred to as the open-loop transfer function or simply the loop transfer function:

$$G_{\text{ol}} = G_c G_p H \tag{59}$$

5.3 Characteristic Equation

The overall system dynamics given by Eq. (57) is primarily governed by the poles of the closed-loop system or the roots of the closed-loop characteristic equation (CLCE):

$$1 + G_c G_p H = 0 \tag{60}$$

It is important to note that the CLCE is simply

$$1 + G_{\text{ol}} = 0 \tag{61}$$

This latter form is the basis of root-locus and frequency-domain design techniques discussed in Chapter 7.

The roots of the characteristic equation are referred to as poles. Specifically, the roots of the open-loop characteristic equation (OLCE) are referred to as open-loop poles and those of the closed loop are called closed-loop poles.

Example 9 Consider the block diagram shown in Fig. 21. The open-loop transfer function is

$$G_{\text{ol}} = \frac{K_1(s + 1)(s + 4)}{s(s + 2)(s + 3)}$$

The closed-loop transfer function of the system is

$$G_{\text{cl}} = \frac{K_1(s + 1)(s + 4)}{s(s + 2)(s + 3) + K_1(s + 1)(s + 4)}$$

The OLCE is the denominator polynomial of G_{ol} set equal to zero. Hence

$$\text{OLCE} \equiv s(s + 2)(s + 3) = 0$$

and the open-loop poles are 0, −2, and −3.

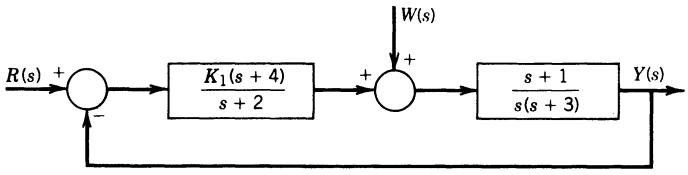


Figure 21 Closed-loop system.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

The CLCE is

$$s(s + 2)(s + 3) + K_1(s + 1)(s + 4) = 0$$

and its roots are the closed-loop poles appropriate for the specific gain value K_1.

If the transfer functions are rational (i.e., are ratios of polynomials), then they can be written as

$$G(s) = \frac{K \prod_{i-1}^{m} (s + z_i)}{\prod_{j-1}^{n} (s + p_j)} \tag{62}$$

When the poles of the transfer function $G(s)$ are distinct, $G(s)$ may be written in partial-fraction form as

$$G(s) = \sum_{j=1}^{n} \frac{A_j}{S + p_j} \tag{63}$$

Hence

$$g(t) = \mathcal{L}^{-1}[G(s)] = \mathcal{L}^{-1}\left(\sum_{j=1}^{n} \frac{A_j}{s + p_j}\right)$$

$$= \sum_{j=1}^{n} A_j e^{-p_j t} \tag{64}$$

Since a transfer function $G(s) = \mathcal{L}[\text{output}]/\mathcal{L}[\text{input}]$, $g(t)$ of of Eq. (64) is the response of the system depicted by $G(s)$ for a unit impulse $\delta(t)$, since $\mathcal{L}[\delta(t)] = 1$.

The impulse response of a given system is key to its internal stability. The term *system* here is applicable to any part of (or the whole) closed-loop system.

It should be noted that the zeros of the transfer function only affect the residues A_j. In other words, the contribution from the corresponding transient term $e^{-p_j t}$ may or may not be significant depending on the relative size of A_j. If, for instance, a zero $-z_k$ is very close to a pole $-p_l$, then the transient term $Ae^{-p_j t}$ would have a value close to zero for its residue A_l. As an example, consider the unit impulse response of the two systems:

$$G_1(s) = \frac{1}{(s + 1)(s + 2)} = \frac{1}{s + 1}\frac{1}{s + 2} \tag{65}$$

$$G_2(s) = \frac{(s + 1.05)}{(s + 1)(s + 2)} = \frac{0.05}{s + 1} + \frac{0.95}{s + 2} \tag{66}$$

From Eq. (65), $g_1(t) = e^{-t} - e^{-2t}$, and from Eq. (66), $g_2(t) = 0.05e^{-t} + 0.95e^{-2t}$.

Note that the effect of the term e^{-t} has been modified from a residue of 1 in G_1 to a residue of 0.05 in G_2. This observation helps to reduce the order of a system when there are poles and zeros close together. In $G_2(s)$, for example, little error would be introduced if the zero at -1.05 and the pole at -1 are neglected and the transfer function is approximated by

$$G_2(s) = \frac{1}{s + 2}$$

From Eq. (64) it can be observed that the shape of the impulse response is determined primarily by the pole locations. A sketch of several pole locations and corresponding impulse responses is given in Fig. 22.

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

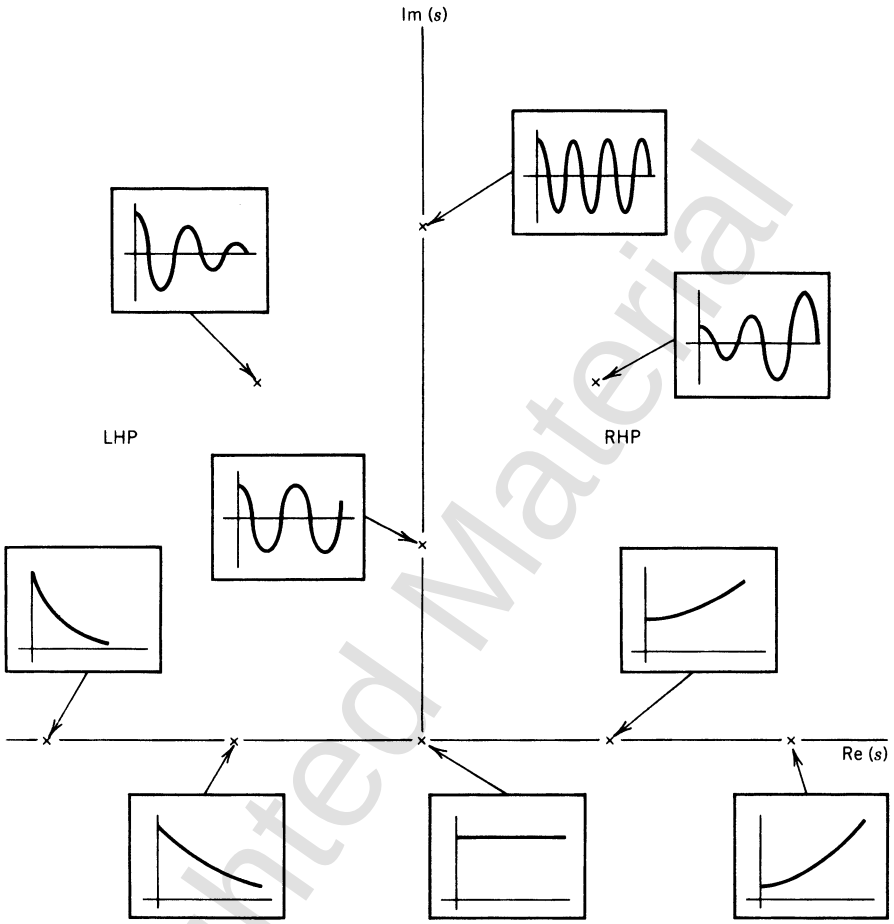


Figure 22 Impulse responses associated with pole locations. LHP, RHP: left- and right-half plane.

The fundamental responses that can be obtained are of the form $e^{-\alpha t}$ and $e^{\lambda t}$ sin βt with $\beta > 0$. It should be noted that for a real pole its location completely characterizes the resulting impulse response. When a transfer function has complex-conjugate poles, the impulse response is more complicated.

5.4 Standard Second-Order Transfer Function

A standard second-order transfer function takes the form

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{1}{s^2/\omega_n^2 + 2\zeta(s/\omega_n) + 1} \tag{67}$$

Parameter ζ is called the damping ratio, and ω_n is called the undamped natural frequency. The poles of the transfer function given by Eq. (67) can be determined by solving its characteristic equation:

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0 \tag{68}$$

giving the poles

$$s_{1,2} = -\zeta\omega_n \pm j\sqrt{1 - \zeta^2}\omega_n = -\sigma \pm j\omega_d \tag{69}$$

where $\sigma = \zeta\omega_n$ and $\omega_d = \omega_n\sqrt{1 - \zeta^2}$, $\zeta < 1$.

The two complex-conjugate poles given by Eq. (69) are located in the complex s-plane, as shown in Fig. 23.

It can be easily seen that

$$|s_{1,2}| = \sqrt{\zeta^2\omega_n^2 + \omega_d^2} = \omega_n \tag{70}$$

and

$$\cos\beta = \zeta \qquad 0 \le \beta \le \tfrac{1}{2}\pi \tag{71}$$

When the system has no damping, $\zeta = 0$, the impulse response is a pure sinusoidal oscillation. In this case the undamped natural frequency ω_n is equal to the damped natural frequency ω_d.

5.5 Step Input Response of a Standard Second-Order System

When the transfer function has complex-conjugate poles, the step input response rather than the impulse response is used to characterize its transients. Moreover, these transients are almost always used as time-domain design specifications. The unit step input response of the second-order system given by Eq. (67) can be easily shown to be

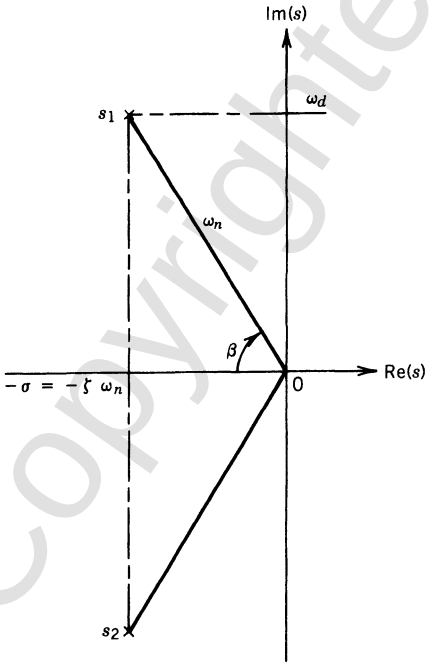


Figure 23 Compelx-conjugate poles in the s-plane.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$y(t) = 1 - e^{\sigma T}\left(\cos \omega_d t + \frac{\sigma}{\omega_d} \sin \omega_d t\right) \tag{72}$$

where $y(t)$ is the output.

A typical unit step input response is shown in Fig. 24.

From the time response of Eq. (72) several key parameters characterizing the shape of the curve are usually used as time-domain design specifications. They are the rise time t_r, the settling time t_s, the peak overshoot M_p, and the peak time t_p.

Rise Time t_r

The time taken for the output to change from 10% of the steady-state value to 90% of the steady-state value is usually called the rise time. There are other definitions of rise time.[2] The basic idea, however, is that t_r is a characterization of how rapid the system responds to an input.

A rule of thumb for t_r is

$$t_r \simeq \frac{1.8}{\omega_n} \tag{73}$$

Settling Time t_s

This is the time required for the transients to decay to a small value so that $y(t)$ is almost at the steady-state level. Various measures of "smallness" are possible: 1, 2, and 5% are typical.

$$\begin{aligned} &\text{For 1\% settling:} \quad t_s \simeq \frac{4.6}{\zeta\omega_n} \\ &\text{For 2\% settling:} \quad t_s \simeq \frac{4}{\zeta\omega_n} \\ &\text{For 5\% settling:} \quad t_s \simeq \frac{3}{\zeta\omega_n} \end{aligned} \tag{74}$$

Peak Overshoot M_p

The peak overshoot is the maximum amount by which the output exceeds its steady-state value during the transients. It can be easily shown that

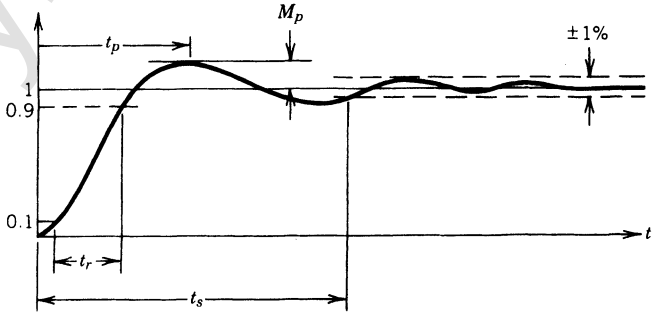


Figure 24 Definition of the rise time t_r, settling time t_s, and overshoot time M_p.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$M_p = e^{-\pi\zeta/\sqrt{1-\zeta^2}} \tag{75}$$

Peak Time t_p

This is the time at which the peak occurs. It can be readily shown that

$$t_p = \frac{\pi}{\omega_d} \tag{76}$$

It is important to note that given a set of time-domain design specifications, they can be converted to an equivalent location of complex-conjugate poles. Figure 25 shows the allowable regions for complex-conjugate poles for different time-domain specifications.

For design purposes the following synthesis forms are useful:

$$\omega_n \geq \frac{1.8}{t_r} \tag{77a}$$

$$\zeta \geq 0.6(1 - M_p) \qquad \text{for} \qquad 0 \leq \zeta \leq 0.6 \tag{77b}$$

$$\sigma \geq \frac{4.6}{t_s} \tag{77c}$$

5.6 Effects of an Additional Zero and an Additional Pole[2]

If the standard second-order transfer function is modified due to a zero as

$$G_1(s) = \frac{(s/\alpha\zeta\omega_n + 1)\omega_n^2}{s^2 + 2\zeta\omega_n + \omega_n^2} \tag{78}$$

or due to an additional pole as

$$G_2(s) = \frac{\omega_n^2}{(s/\alpha\zeta\omega_n + 1)(s^2 + 2\zeta\omega_n s + \omega_n^2)} \tag{79}$$

it is important to know how close the resulting step input response is to the standard second-order step response. Following are several features of these additions:

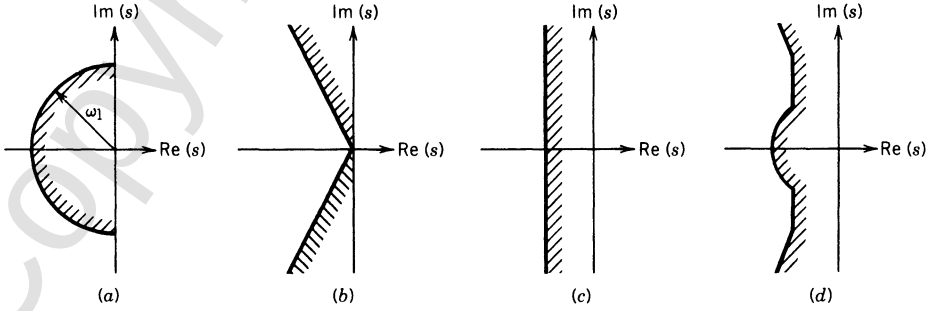


Figure 25 Graphs of regions in the s-plane for certain transient requirements to be met: (a) rise time requirements, (ω_n); (b) overshoot requirements, (ζ); (c) settling time requirements, (ζ); (d) composite of (a), (b), and (c).

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

1. For a second-order system with no finite zeros, the transient parameters can be approximated by Eq. (77).
2. An additional zero as in Eq. (78) in the left-half plane will increase the overshoot if the zero is within a factor of 4 of the real part of the complex poles. A plot is given in Fig. 26.
3. An additional zero in the right-half plane will depress the overshoot (and may cause the step response to undershoot). This is referred to as a non-minimum-phase system.
4. An additional pole in the left-half plane will increase the rise time significantly if the extra pole is within a factor of 4 of the real part of the complex poles. A plot is given in Fig. 27.

6 STABILITY

We shall distinguish between two types of stabilities: external and internal stability. The notion of external stability is concerned with whether or not a bounded input gives a bounded output. In this type of stability we notice that no reference is made to the internal variables of the system. The implication here is that it is possible for an internal variable to grow without bound while the output remains bounded. Whether or not the internal variables are well behaved is typically addressed by the notion of internal stability. Internal stability requires that in the absence of an external input the internal variables stay bounded for any perturbations of these variables. In other words internal stability is concerned with the response of the system due to nonzero initial conditions. It is reasonable to expect that a well-designed system should be both externally and internally stable.

The notion of asymptotic stability is usually discussed within the context of internal stability. Specifically, if the response due to nonzero initial conditions decays to zero asymptotically, then the system is said to be asymptotically stable. A LTI system is asymptotically stable if and only if all the system poles lie in the open left-half-plane (i.e., the left-half s-plane excluding the imaginary axis). This condition also guarantees external stability for LTI systems. So in the case of LTI systems the notions of internal and external stability may be considered equivalent.

For LTI systems, knowing the locations of the poles or the roots of the characteristic equation would suffice to predict stability. The Routh–Hurwitz stability criterion is frequently used to obtain stability information without explicitly computing the poles for LTI. This criterion will be discussed in Section 6.1.

For nonlinear systems, stability cannot be characterized that easily. As a matter of fact, there are many definitions and theorems for assessing stability of such systems. A discussion of these topics is beyond the scope of this handbook. Interested reader may refer to Ref. 3.

6.1 Routh–Hurwitz Stability Criterion

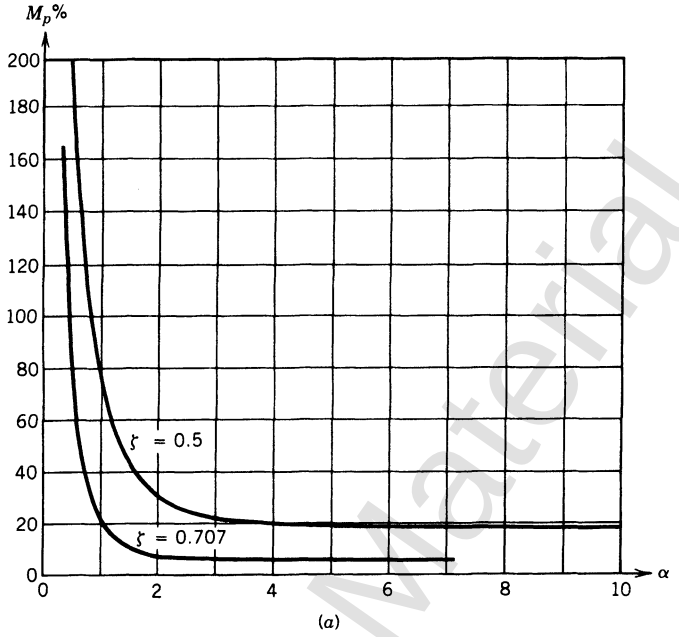
This criterion allows one to predict the status of stability of a system by knowing the coefficients of its characteristic polynomial. Consider the characteristic polynomial of an nth-order system:

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

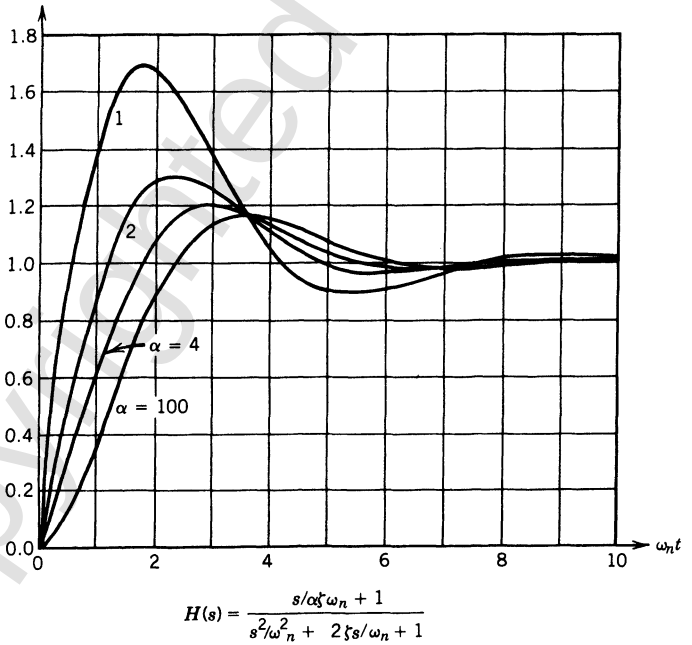
A necessary condition for asymptotic stability is that all the coefficients $\{a_i\}$'s be positive. If any of the coefficients are missing (i.e., are zero) or negative, then the system will have

Copyrighted material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com



(a)



$$H(s) = \frac{s/\alpha\zeta\omega_n + 1}{s^2/\omega_n^2 + 2\zeta s/\omega_n + 1}$$

(b)

Figure 26 Effect of an extra zero on a standard second-order system: (*a*) normalized rise time versus α; (*b*) step response versus α, for $\zeta = 0.5$.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

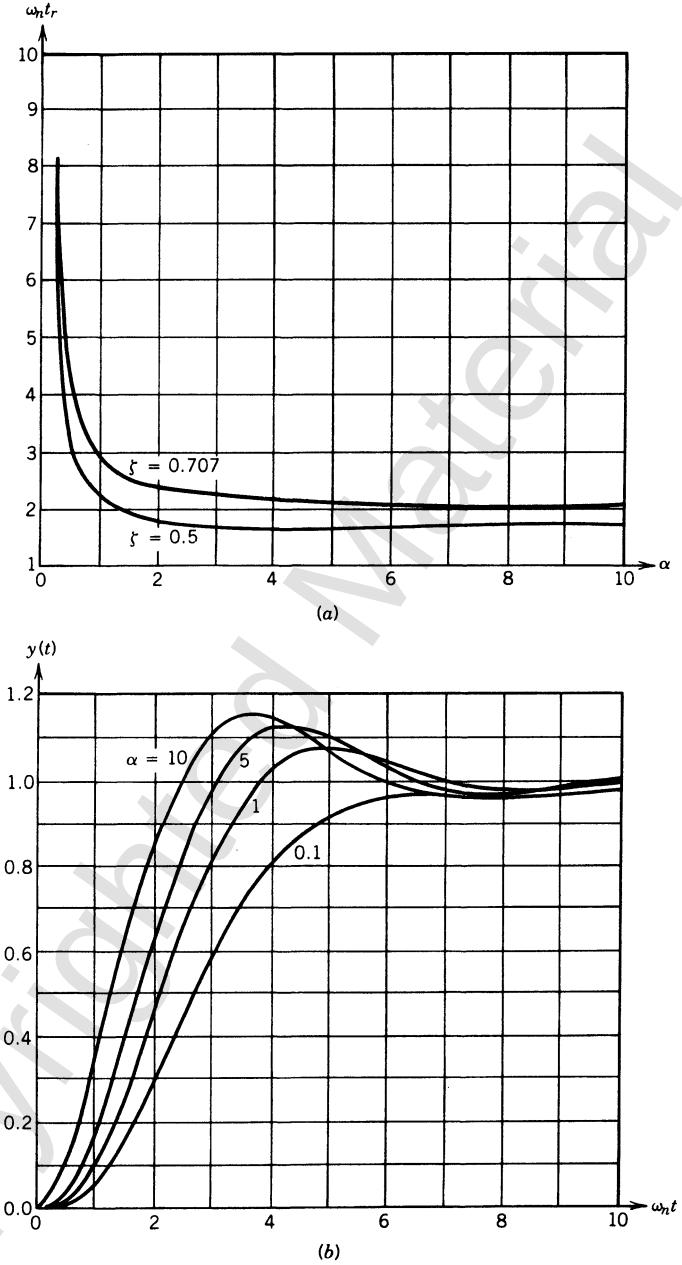


Figure 27 Effect of an additional pole on a standard second-order system: (*a*) normalized rise time versus α; (*b*) step response versus α, for $\zeta = 0.5$.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

poles located in the closed right-half plane. If the necessary condition is satisfied, then the so-called Routh array needs to be formed to make conclusions about the stability. A necessary and sufficient condition for stability is that all the elements in the first column of the Routh array be positive.

To determine the Routh array, the coefficients of the characteristic polynomial are arranged in two rows, each beginning with the first and second coefficients and followed by even-numbered and odd-numbered coefficients as follows:

$$\begin{array}{llllll} s^n & a_n & a_{n-2} & a_{n-4} & \cdots \\ s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & \cdots \end{array}$$

The following rows are subsequently added to complete the Routh array:

$$\begin{array}{llllll} s^n & a_n & a_{n-2} & a_{n-4} & \cdots \\ s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ s^{n-2} & b_1 & b_2 & b_3 & \cdots \\ s^{n-3} & c_1 & c_2 & c_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \\ s^2 & * & * & & \\ s & * & & & \\ s^0 & & & & \end{array}$$

where the elements from the third row on are computed as follows:

$$b_1 = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}} \qquad c_1 = \frac{b_1 a_{n-3} - a_{n-1}b_2}{b_1}$$

$$b_2 = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}} \qquad c_2 = \frac{b_1 a_{n-5} - a_{n-1}b_3}{b_1}$$

$$b_3 = \frac{a_{n-1}a_{n-6} - a_n a_{n-7}}{a_{n-1}} \qquad c_3 = \frac{b_1 a_{n-7} - a_{n-1}b_4}{b_1}$$

Note that the elements of the third row and of rows thereafter are formed from the two previous rows using the two elements in the first column and other elements for successive columns. Normally there will be $n + 1$ elements in the first column when the array is completed.

The Routh–Hurwitz criterion states that the number of roots of $P(s)$ with positive real parts is equal to the number of changes of sign in the first column of the array. This criterion requires that there be no changes of sign in the first column for a stable system. A pattern of $+, -, +$ is counted as two sign changes, one going from $+$ to $-$ and another from $-$ to $+$.

If the first term in one of the rows is zero or if an entire row is zero, then the standard Routh array cannot be formed and the use of special techniques, described in the following discussion, become necessary.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

Special Cases

1. Zero in the first column, while some other elements of the row containing a zero in the first column are nonzero
2. Zero in the first column, and the other elements of the row containing the zero are also zero

CASE 1: In this case the zero is replaced with a small positive constant $\epsilon > 0$, and the array is completed as before. The stability criterion is then applied by taking the limits of entries of the first column as $\epsilon \to 0$. For example, consider the following characteristic equation:

$$s^5 + 2s^4 + 2s^3 + 4s^2 + 11s + 10 = 0$$

The Routh array is then

1	2	11	
2	4	10	
0	6	0	
ϵ	6	0	first column zero replaced by ϵ
c_1	10		
d_1	0		

where

$$c_1 = \frac{4\epsilon - 12}{\epsilon} \quad \text{and} \quad d_1 = \frac{6c_1 - 10\epsilon}{c_1}$$

As $\epsilon \to 0$, we get $c_1 \simeq -12/\epsilon$, and $d_1 \simeq 6$. There are two sign changes due to the large negative number in the first column. Therefore the system is unstable, and two roots lie in the right-half plane. As a final example consider the characteristic polynomial

$$P(s) = s^4 + 5s^3 + 7s^2 + 5s + 6$$

The Routh array is

1	7	6	
5	5		
6	6		
ϵ	$\leftarrow$		Zero replaced by $\epsilon > 0$
6			

If $\epsilon > 0$, there are no sign changes. If $\epsilon < 0$, there are two sign changes. Thus if $\epsilon = 0$, it indicates that there are two roots on the imaginary axis, and a slight perturbation would drive the roots into the right-half plane or the left-half plane. An alternative procedure is to define the auxiliary variable

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

and convert the characteristic polynomial so that it is in terms of z. This usually produces a Routh array with nonzero elements in the first column. The stability properties can then be deduced from this array.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

CASE 2: This case corresponds to a situation where the characteristic equation has equal and opposite roots. In this case if the ith row is the vanishing row, an auxiliary equation is formed from the previous i_{-1} row as follows:

$$P_1(s) = \beta_1 s^{i+1} + \beta_2 s^{i-1} + \beta_3 s^{i-3} + \cdots$$

where the $\{\beta_i\}$'s are the coefficients of the $(i-1)$th row of the array. The ith row is then replaced by the coefficients of the derivative of the auxiliary polynomial, and the array is completed. Moreover, the roots of the auxiliary polynomial are also roots of the characteristic equation. As an example, consider

$$s^5 + 2s^4 + 5s^3 + 10s^2 + 4s + 8 = 0$$

for which the Routh array is

s^5	1	5	4
s^4	2	10	8
s^3	0	0	
s^3	8	20	
s^2	5	8	
s^1	7.2		
s^0	8		

Auxiliary equation: $2s^4 + 10s^2 + 8 = 0$

There are no sign changes in the first column. Hence all the roots have nonpositive real parts with two pairs of roots on the imaginary axis, which are the roots of

$$2s^4 + 10s^2 + 8 = 0$$
$$= 2(s^2 + 4)(s^2 + 1)$$

Thus equal and opposite roots indicated by the vanishing row are $\pm j$ and $\pm 2j$.

The Routh–Hurwitz criterion may also be used in determining the range of parameters for which a feedback system remains stable. Consider, for example, the system described by the CLCE

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

The corresponding Routh array is

s^4	1	3	K
s^3	3	2	
s^2	$\frac{7}{3}$	K	
s^1	$s - (9/7)K$		
s^0	K		

If the system is to remain asymptotically stable, we must have

$$0 < K < \tfrac{14}{9}$$

The Routh–Hurwitz stability criterion can also be used to obtain additional insights. For instance information regarding the speed of response may be obtained by a coordinate transformation of the form $s + a$, where $-a$ characterizes rate. For this additional detail the reader is referred to Ref. 4.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

6.2 Polar Plots

The frequency response of a system described by a transfer function $G(s)$ is given by

$$y_{ss}(t) = X|G(j\omega)| \sin[\omega t + \underline{/G(j\omega_1)}]$$

where $X \sin \omega t$ is the forcing input and $y_{ss}(t)$ is the steady-state output. A great deal of information about the dynamic response of the system can be obtained from a knowledge of $G(j\omega)$. The complex quantity $G(j\omega)$ is typically characterized by its frequency-dependent magnitude $|G(j\omega)|$ and the phase $\underline{/G(j\omega)}$. There are many different ways in which this information is handled. The manner in which the phasor $G(j\omega)$ traverses the $G(j\omega)$-plane as a function of ω is represented by a polar plot. This is represented in Fig. 28. The phasor OA corresponds to a magnitude $|G(j\omega_1)|$ and the phase $\underline{/G(j\omega_1)}$. The locus of point A as a function of the frequency is the polar plot. This representation is useful in stating the Nyquist stability criterion.

6.3 Nyquist Stability Criterion

This is a method by which the closed-loop stability of a LTI system can be predicted by knowing the frequency response of the 1oop transfer function. A typical CLCE may be written as

$$1 + GH = 0$$

where GH is referred to as the 1oop transfer function.

The Nyquist stability criterion[2] is based on a mapping of the so-called Nyquist contour by the loop transfer function. Figure 29*a* shows the Nyquist contour, and Fig. 29*b* shows its typical mapped form in the GH plane. Let C_1 be the contour in the s-plane to be mapped. Utilizing this information and the poles of GH, the Nyquist stability criterion can be stated as follows: A closed-loop system is stable if and only if the mapping of C_1 in the GH plane encircles the $(-1,0)$ point N number of times in the counterclockwise direction, where N is the number of poles of GH with positive real parts.

Clearly, if GH does not have any poles in the right-half plane, then C_1 should not encircle the $(-1,0)$ point if the closed-loop system is to be stable.

If there are poles of GH on the imaginary axis, then the Nyquist contour should be modified with appropriate indentations, as shown in Fig. 30. An example is given to map such an indented contour with the corresponding loop transfer function. Consider

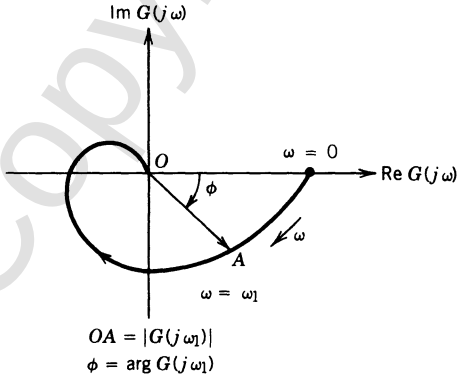


Figure 28 Polar plot.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

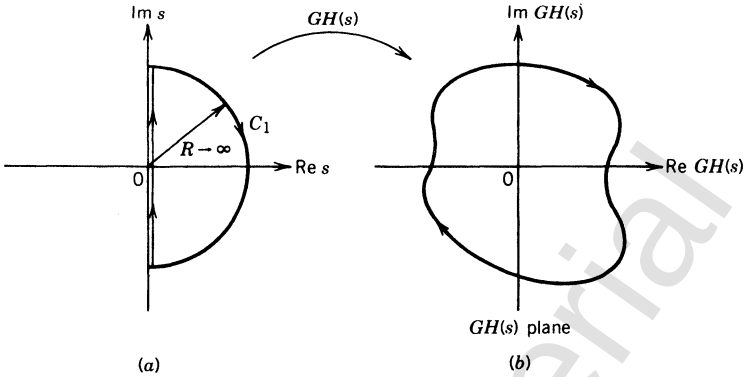


Figure 29 Nyquist contour and its mapping by $GH(s)$.

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

The contour Γ in the s-plane is shown in Fig. 31a, where an indentation is affected by a small semicircle of radius ϵ, where $\epsilon > 0$. When mapped by $GH(s)$, the contour of Fig. 31b is obtained. To effect the mapping, the points on the semicircles are represented as $s = \epsilon e^{j\phi}$ on the small indentation with $\phi\epsilon[-\pi/2, \pi/2]$ and $\epsilon \to 0$ and $s = Re^{j\theta}$ with $\theta\epsilon[-\pi/2, \pi/2]$ and $R \to \infty$ on the infinite semicircle. We observe that the $(-1,0)$ point is not encircled. Since $GH(s)$ does not have any poles in the right-half s-plane from the Nyquist stability criterion, it follows that the closed-loop system is stable. In Table 3 several loop transfer functions with the appropriate Nyquist contour and mapped contours are given.

7 STEADY-STATE PERFORMANCE AND SYSTEM TYPE

In the design of feedback control systems the steady-state performance is also of importance in many instances. This is in addition to the stability and transient performance requirements. Consider the closed-loop system shown in Fig. 32.

Here $G_p(s)$ is the plant and $G_c(s)$ is the controller. When steady-state performance is important, it is advantageous to consider the system error due to an input. From the block diagram

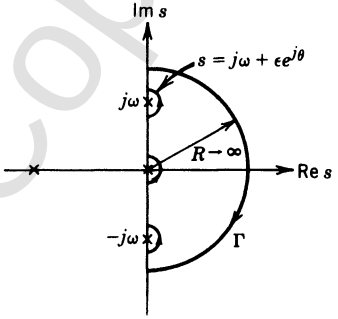


Figure 30 Nyquist contour with indentations.

Copyrighted material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

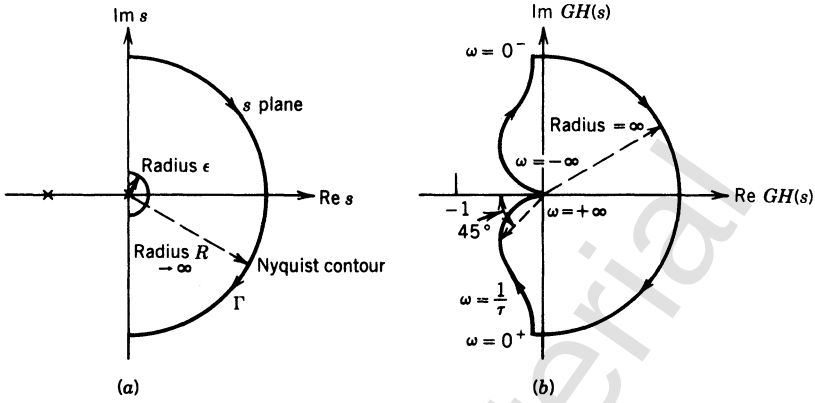


Figure 31 Nyquist contour and mapping for $GH(s) = K/s(\tau s + 1)$.

$$E(s) = R(s) - Y(s) \tag{80}$$

$$Y(s) = G_c(s)G_p(s)E(s) \tag{81}$$

Substituting Eq. (81) in (80) yields

$$\frac{E(s)}{R(s)} = \frac{1}{1 + G_c(s)G_p(s)} \tag{82}$$

called the error transfer function. It is important to note that the error dynamics are described by the same poles as those of the closed-loop transfer function. Namely, the roots of the characteristic equation $1 + G_c(s)G_p(s) = 0$.

Given any input $r(t)$ with $\mathcal{L}[r(t)] = R(s)$, the error can be analyzed by considering the inverse Laplace transform of $E(s)$ given by

$$e(t) = \mathcal{L}^{-1}\left[\frac{1}{1 + G_c(s)G_p(s)}\,R(s)\right] \tag{83}$$

For the system's error to be bounded, it is important to first assure that the closed-loop system is asymptotically stable. Once the closed-loop stability is assured, the steady-state error can be computed by using the final-value theorem. Hence

$$\lim_{t\to\infty} e(t) = e_{ss} = \lim_{s\to 0} sE(s) \tag{84}$$

By substituting for $E(s)$ in Eq. (84)

$$e_{ss} = \lim_{s\to 0} s\,\frac{1}{1 + G_c(s)G_p(s)}\,R(s) \tag{85}$$

7.1 Step Input

If the reference input is a step of magnitude c, then from Eq. (85)

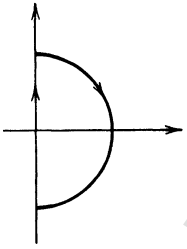
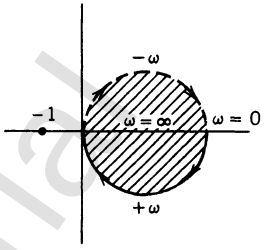
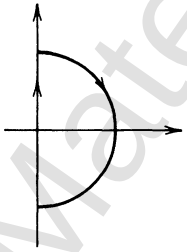
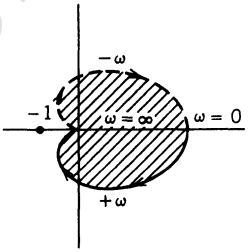
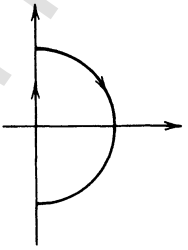
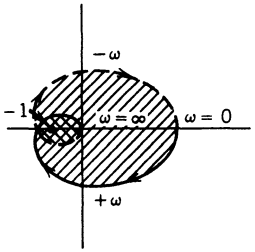
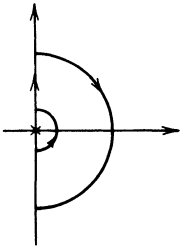
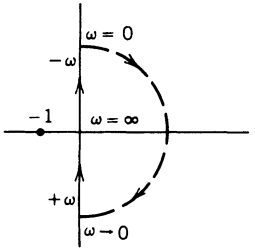
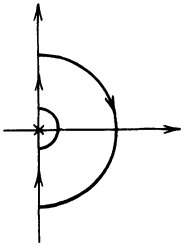
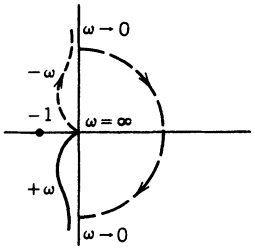
$$e_{ss} = \lim_{s\to 0} s\,\frac{1}{1 + G_c(s)G_p(s)}\,\frac{c}{s} = \frac{c}{1 + G_cG_p(0)} \tag{86}$$

Equation (86) suggests that to have small steady-state error, the low-frequency gain of the open-loop transfer function $G_cG_p(0)$ must be very large. It is typical to define

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

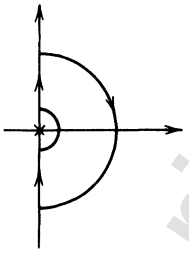
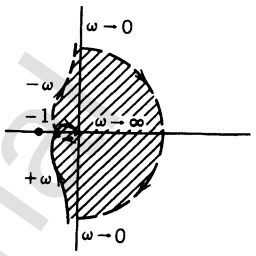
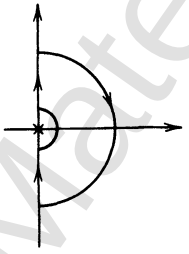
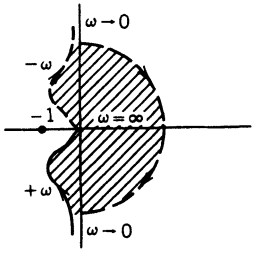
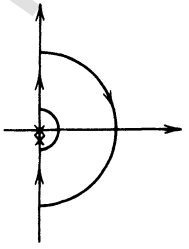
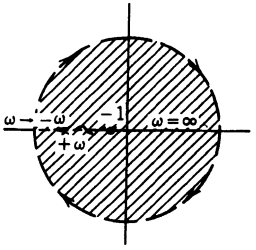
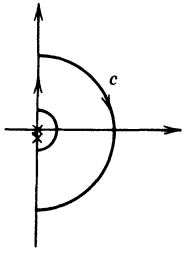
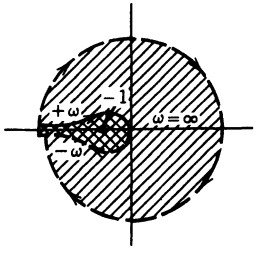
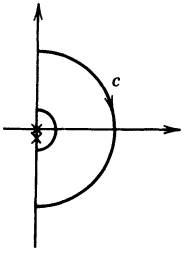
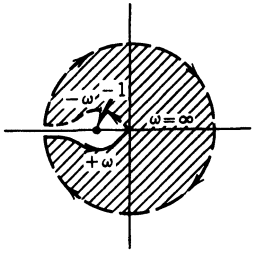
Table 3 Loop Transfer Functions

No.	$G(s)$	Nyquist Contour	Polar Plot
1.	$\frac{K}{s\tau_1 + 1}$		
2.	$\frac{K}{(s\tau_1 + 1)(s\tau_2 + 1)}$		
3.	$\frac{K}{(s\tau_1 + 1)(s\tau_2 + 1)(s\tau_3 + 1)}$		
4.	$\frac{K}{s}$		
5.	$\frac{K}{(s\tau_1 + 1)}$		

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

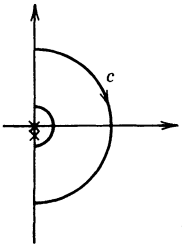
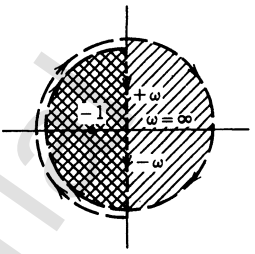
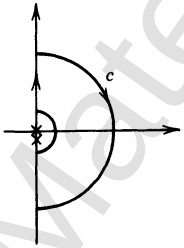
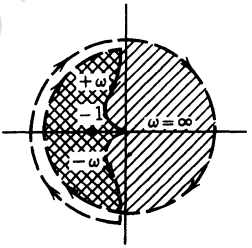
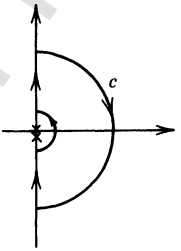
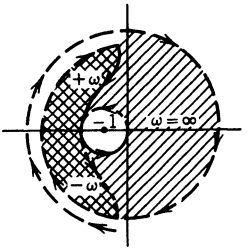
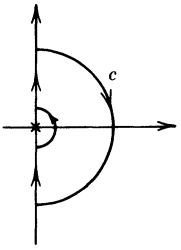
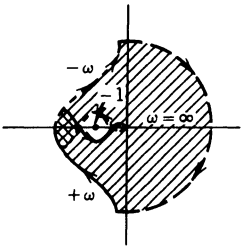
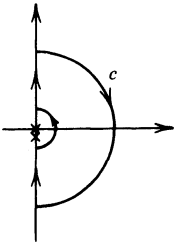
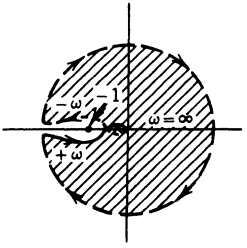
Table 3 (*Continued*)

No.	$G(s)$	Nyquist Contour	Polar Plot
6.	$\dfrac{K}{s(s\tau_1 + 1)(s\tau_2 + 1)}$		
7.	$\dfrac{K(s\tau_a + 1)}{s(s\tau_1 + 1)(s\tau_2 + 1)}$ $\tau_a < \dfrac{\tau_1\tau_2}{\tau_1 + \tau_2}$ $\tau_a < \dfrac{\tau_1\tau_2}{\tau_1 + \tau_2}$		
8.	$\dfrac{K}{s^2}$		
9.	$\dfrac{K}{s^2(s\tau_1 + 1)}$		
10.	$\dfrac{K(s\tau_1 + 1)}{s^2(s\tau_1 + 1)}$ $\tau_2 > \tau_1$ $\tau_a > \tau_1$		

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

Table 3 (*Continued*)

No.	$G(s)$	Nyquist Contour	Polar Plot
11.	$\frac{K}{s^3}$		
12.	$\frac{K(s\tau_a + 1)}{s^3}$		
13.	$\frac{K(s\tau_a + 1)(s\tau_b + 1)}{s^3}$		
14.	$\frac{K(s\tau_a + 1)(s\tau_b + 1)}{s(s\tau_1 + 1)(s\tau_2 + 1)(s\tau_3 + 1)(s\tau_4 + 1)}$		
15.	$\frac{K(s\tau_a + 1)}{s^2(s\tau_1 + 1)(s\tau_2 + 1)}$		

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

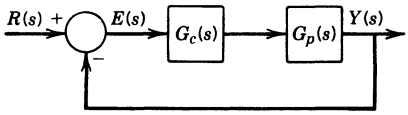


Figure 32 Closed-loop configuration.

$$K_p = G_c G_p(0) \tag{87}$$

as the position error constant. With this definition the steady-state error due to a step input of magnitude c can be written as

$$e_{ss} = \frac{c}{1 + K_p} \tag{88}$$

Thus a high value of K_p, corresponds to a low steady-state error. If the steady-state error is to be zero, then $K_p = \infty$. The only way that $K_p = \infty$ is if the open-loop transfer function has at least one pole at the origin, that is, $G_c G_p(s)$ must be of the form

$$G_c G_p(s) = \frac{1}{s^N} \frac{\prod_{i-1}^{m} (s + z_i)}{\prod_{j=1}^{n} (s + p_j)} \tag{89a}$$

where $N \geq 1$. When $N > 1$,

$$G_c G_p(0) = \frac{1}{0} \frac{\prod_{i-1}^{m} z_i}{\prod_{j=1}^{n} p_j} \to \infty \tag{89b}$$

Hence, it can be concluded that for the steady-state error due to a step input to be zero, the open-loop transfer function must have at least one free integrator. The value of N specifies the type of system. If $N = 1$ it is called a type I, when $N = 2$ it is called a type II system, and so on. So to get zero steady-state error for a step input, the system loop transfer function must be at least type I.

7.2 Ramp Input

If the reference input is a ramp $ctu_s(t)$, where $u_s(t)$ is the unit step, then from Eq. (85)

$$e_{ss} = \lim_{s \to 0} s \frac{1}{1 + G_c G_p(s)} \frac{c}{s^2} = \lim_{s \to 0} \frac{c}{s G_c G_p(s)} \tag{90}$$

From Eq. (90) for small steady-state errors $\lim_{s \to 0} sG_c G_p(s) = K_v$ must be large, where K_v is the velocity error constant and

$$e_{ss} = \frac{c}{K_v} \tag{91}$$

As in the case with the step input, for e_{ss} to be small, K_v must be very large. For zero steady-state error with a ramp input, $K_v = \infty$. From Eq. (90) it is clear that for $K_v = \infty$, $G_c G_p(s)$ must be at least type II. Thus

$$K_v = \lim_{s \to 0} sG_c G_p(s) = \lim_{s \to 0} s \frac{1}{s^2} = \frac{\prod_{i=1}^{m} (s + z_i)}{\prod_{j=1}^{n} (s + p_j)} = \infty$$

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

7.3 Parabolic Input

If the reference input is a parabolic input of the form

$$r(t) = c\,\frac{t^2}{2}\,u_s(t)$$

then the steady-state error becomes

$$e_{ss} = \lim_{s\to 0} \frac{c}{s^2 G_c G_p(s)} \tag{92}$$

From Eq. (92) for small steady-state errors

$$K_a = \lim_{s\to 0} s^2 G_c G_p(s) \tag{93}$$

must be very large, where K_a is the parabolic error constant or the acceleration error constant. For zero steady-state error due to a parabolic input, $K_a = \infty$. Therefore the system open-loop transfer function must be at least type III.

Table 4 shows the steady-state errors in terms of the error constants K_p, K_v, and K_a.

In steady-state performance considerations it is important to guarantee both closed-loop stability and the system type. It should also be noticed that having a very high 1oop gain as given by Eq. (89b) is very similar to having a free integrator in the open-loop transfer function. This notion is useful in regulation-type problems. In tracking systems the problem is compounded by the fact that the system type must be increased. Increasing the system type is usually accompanied by instabilities. To illustrate this, consider the following example.

Example 10 Synthesize a controller for the system shown in Fig. 33*a* so that the steady-state error due to a ramp input is zero.

The open-loop system is unstable. For tracking purposes transform the system to a closed-loop one as shown in Fig. 33*b*. To have zero steady-state error for a ramp, the system's open-loop transfer function must be at least type II and must be of a form that guarantees closed-loop stability. Therefore let

$$G_c(s) = \frac{G_1(s)}{s^2} \tag{94}$$

Table 4 Steady-State Error in Terms of Error Constants

Type	$cu_s(t)$	$ctu_s(t)$	$\frac{t^2}{2}\,u_s(t)$
0	$\frac{c}{1+K_p}$	∞	∞
1	0	$\frac{c}{K_v}$	∞
2	0	0	$\frac{c}{K_a}$
3	0	0	0

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

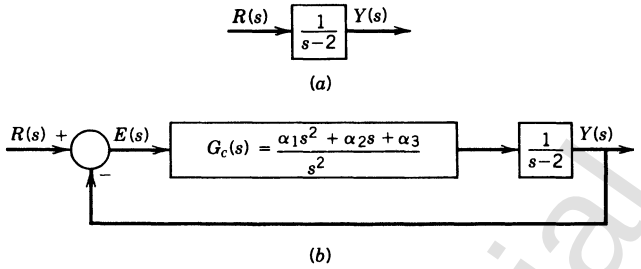


Figure 33 Tracking system: (*a*) open-loop system; (*b*) closed-loop system.

The CLCE is

$$1 + \frac{G_1(s)}{s^2}\frac{1}{s-2} = 0 \tag{95}$$

That is, $s^3 - 2s^2 + G_1(s) = 0$.

Now for closed-loop stability, $G_1(s)$ must be of the form

$$G_1(s) = \alpha_1 s^2 + \alpha_2 s + \alpha_3$$

yielding the CLCE

$$s^3 + (\alpha_1 - 2)s^2 + \alpha_2 s + \alpha_3 = 0 \tag{96}$$

From the Routh–Hurwitz stability criterion it can be shown that for asymptotic stability $\alpha_1 > 2$, $\alpha_3 > 0$, and $(\alpha_1 - 2)\alpha_2 - \alpha_3 > 0$. So pick $\alpha_1 = 4$, $\alpha_2 = 1$, $\alpha_3 = 1$, yielding the controller

$$G_c(s) = \frac{4s^2 + s + 1}{s^2} \tag{97}$$

In an actual design situation the choice of α_1, α_2, and α_3 will be dictated by the transient design specifications.

7.4 Indices of Performance

Sometimes the performance of control systems is given in terms of a performance index. A performance index is a number that indicates the ''goodness'' of system performance. Such performance indices are usually employed to get an optimal design in the sense of minimizing or maximizing the index of performance. The optimal parameter values depend directly upon the performance index chosen. Typical performance indices are $\int_0^\infty e^2(t)\,dt$, $\int_0^\infty te^2(t)\,dt$, $\int_0^\infty |e(t)|\,dt$, $\int_0^\infty t|e(t)|\,dt$. In control system design the task is to select the controller parameters to minimize the chosen index.

7.5 Integral-Square-Error (ISE) Criterion

According to the ISE criterion, the quality of system performance is evaluated by minimizing the integral

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

$$J = \int_0^\infty e^2(t)\, dt \tag{98}$$

A system designed by this criterion tends to show a rapid decrease in a large initial error. Hence the response is fast and oscillatory. Thus the system has poor relative stability. The ISE is of practical significance because $\int e^2(t)\, dt$ resembles power consumption for some systems.

7.6 Integral of Time-Multiplied Absolute-Error (ITAE) Criterion

According to the ITAE criterion, the optimum system is the one that minimizes the performance index:

$$J = \int_0^\infty t|e(t)|\, dt \tag{99}$$

This criterion weighs large initial errors lightly, and errors occurring late in the transient response are penalized heavily. A system designed by use of the ITAE has a characteristic that the overshoot in the transient response is small and oscillations are well damped.

7.7 Comparison of Various Error Criteria[2]

Figure 34 shows several error performance curves. The system considered is

$$\frac{C(s)}{R(s)} = \frac{1}{s^2 + 2\zeta s + 1} \tag{100}$$

The curves of Fig. 34 indicate that $\zeta = 0.7$ corresponds to a near-optimal value with respect to each of the performance indices used. At $\zeta = 0.7$ the system given by Eq. (100) results in rapid response to a step input with approximately 5% overshoot.

Table 5 summarizes the coefficients that will minimize the ITAE performance criterion for a step input to the closed-loop transfer function[5]

$$\frac{C(s)}{R(s)} = \frac{a_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{101}$$

Table 6[2] summarizes the coefficients that will minimize the ITAE performance criterion for a ramp input applied to the closed-loop transfer function:

$$\frac{C(s)}{R(s)} = \frac{b_1 s + b_0}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0} \tag{102}$$

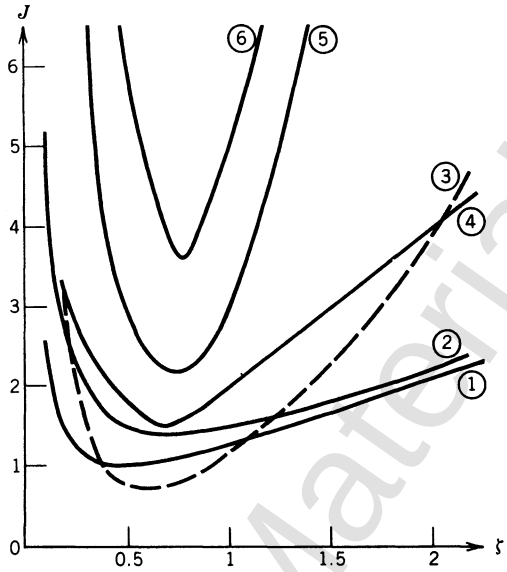
Figure 35 shows the response resulting from optimum coefficients for a step input applied to the normalized closed-loop transfer function given in Eq. (101), for ISE, IAE (integral absolute error), and ITAE.

8 SIMULATION FOR CONTROL SYSTEM ANALYSIS

Often in control system design the governing differential equations are reduced to a form that facilitates the shaping of a controller. The model order reductions involve things such as linearization, neglecting fast dynamics, parasitics, time delays, and so on. Once a controller is synthesized for a system, after making simplifying assumptions, it becomes nec-

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com



① $J = \int_0^\infty e^2\,dt$ ② $J = \int_0^\infty |e|\,dt$

③ $J = \int_0^\infty (e^2 + \dot{e}^2)\,dt$ ④ $J = \int_0^\infty t\,|e|\,dt$

⑤ $J = \int_0^\infty te^2\,dt$ ⑥ $J = \int_0^\infty t(|e| + |\dot{e}|)\,dt$

Figure 34 Error performance curves.

essary to try out the controller on the actual system. This involves experimentation. Computer simulation may also be viewed as experimentation where the synthesized controller is inserted into a very sophisticated model of the actual physical system and the system response computed. Extensive simulation studies can reduce the cost of experimentation and may even at times serve as the final controller to be implemented on the actual physical plant. So in simulation the burden of experiment is transformed into generating accurate models for all the components in a system without oversimplifying assumptions. Simulations can be performed on analog, digital, or hybrid computers.[7]

Table 5 Optimal Form of the Closed-Loop Transfer Function Based on the ITAE Criterion for Step Inputs

$\dfrac{C(s)}{R(s)} = \dfrac{a_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}, \; a_0 = \omega_n^n$
$s + \omega_n$
$s^2 + 1.4\omega_n s + \omega_n^2$
$s^3 + 1.75\omega_n s^2 + 2.15\omega_n^2 s + \omega_n^3$
$s^4 + 2.1\omega_n s^3 + 3.4\omega_n^2 s^2 + 2.7\omega_n^3 s + \omega_n^4$
$s^5 + 2.8\omega_n s^4 + 5.0\omega_n^2 s^3 + 5.5\omega_n^3 s^2 + 3.4\omega_n^4 s + \omega_n^5$
$s^6 + 3.25\omega_n s^5 + 6.60\omega_n^2 s^4 + 8.60\omega_n^3 s^3 + 7.45\omega_n^4 s^2 + 3.95\omega_n^5 s + \omega_n^6$

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

Table 6 Optimum Coefficients of $T(s)$ Based on the ITAE Criterion for a Ramp Input

$s^2 + 3.2\omega_n s + \omega_n^2$
$s^3 + 1.75\omega_n s^2 + 3.25\omega_n^2 s + \omega_n^3$
$s^4 + 2.41\omega_n s^3 + 4.93\omega_n^2 s^2 + 5.14\omega_n^3 s + \omega_n^4$
$s^5 + 2.19\omega_n s^4 + 6.50\omega_n^2 s^3 + 6.30\omega_n^3 s^2 + 5.24\omega_n^4 s + \omega_n^5$

8.1 Analog Computation[8]

An analog computer is a machine in which several physical components can be selected and interconnected in such a way that the equations describing the operation of the computer are analogous to that of the actual physical system to be studied. It is a continuous-time device operating in a real-time parallel mode, making it particularly suitable for the solution of differential equations and hence for the simulation of dynamic systems. The most commonly used analog computer is the electronic analog computer in which voltages at various places within the computer are proportional to the variable terms in the actual system. The ease of use and the direct interactive control over the running of such a computer allow full scope for engineering intuition and make it a valuable tool for the analysis of dynamic systems and the synthesis of any associated controllers. A facility frequently useful is the ability to slow down or speed up the problem solution. The accuracy of solution, since it is dependent on analog devices, is generally of the order of a few percent, but for the purposes of system analysis and design, higher accuracy is seldom necessary; also, this accuracy often matches the quality of the available input data.

The basic building block of the analog computer is the high-gain dc amplifier, represented schematically by Fig. 36. When the input voltage is $e_i(t)$, the output voltage is given by

$$e_o(t) = -Ae_i(t) \tag{103}$$

where A, the amplifier voltage gain, is a large constant value.

If the voltage to the input of an amplifier, commonly referred to as the summing junction, exceeds a few microvolts, then the amplifier becomes saturated or overloaded because the power supply cannot force the output voltage high enough to give the correct gain. Therefore, if an amplifier is to be operated correctly, its summing junction must be very close to ground potential and is usually treated as such in programming.

When this is used in conjunction with a resistance network as shown in Fig. 37, the resulting circuit can be used to add a number of voltages.

If

$$R_1 = R_2 = R_3$$

then

$$V_0 = -(V_1 + V_2 + V_3) \tag{104}$$

If R_1, R_2, R_3 are arbitrary, then

$$V_0 = -\left(\frac{R_f}{R_1} V_1 + \frac{R_f}{R_2} V_2 + \frac{R_f}{R_3} V_3\right) \tag{105}$$

If there is only one voltage input, then

Copyrighted material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

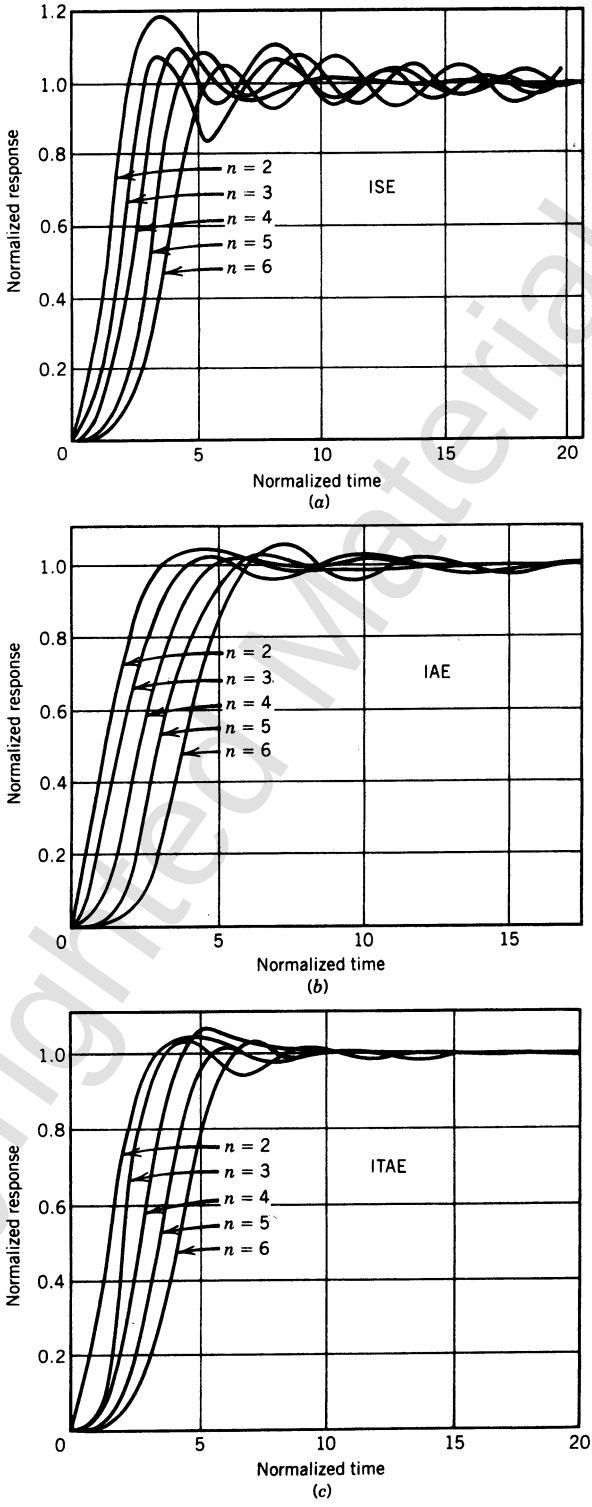


Figure 35 (*a*) Step response of a normalized transfer function using optimum coefficients of ISE, (*b*) IAE, and (*c*) ITAE.

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

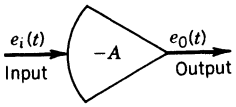


Figure 36 High-gain dc amplifier.

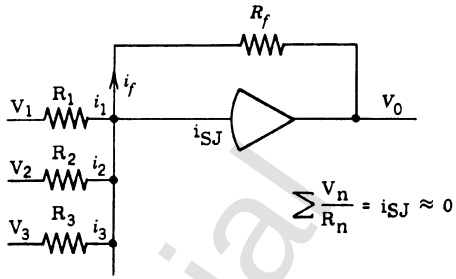


Figure 37 Current balance.

$$V_0 = \frac{R_f}{R_1} V_1 \cdots \text{multiplication by a constant} \tag{106}$$

It should be noted that in all cases there is a sign inversion. Usually the available ratios R_f/R_1 and so on are standardized to 1 and 10, the appropriate gain being selectable as required. The complete circuit comprising the high-gain amplifier, input resistors, and the feedback element is termed an *operational amplifier.* It is given symbolically in Fig. 38.

To multiply a voltage by a constant other than 10, use is made of a grounded potentiometer (usually a 10-turn helical potentiometer), as shown in Fig. 39. This permits multiplication by a constant in the range 0–1.

Since electrical circuits are very nearly linear, the generation of even the simplest nonlinearity, like a product, requires special consideration. One way this can be done is to use the feedback control concept shown in Fig. 40. Two or more potentiometers are mounted rigidly with a common shaft turning the sliders exactly together. A negative unit voltage, which commonly is 10–100 V, is applied across one of the potentiometers. The slider voltage is added to the voltage representing one of the factors. If they differ, then the motor turns the shaft in the direction that eliminates the error. The second factor voltage is applied across the other potentiometer. Its slider then reads the product. This device is called a servomultiplier. It is quite slow in operation because of the motor. Faster multiplication requires the use of special nonlinear networks using diodes. A diode allows current to flow in one direction only if polarities are in the indicated directions in the symbolic representations of Fig. 41. Combinations of diodes and resistors can be made such that the current–voltage curve is a series of straight-line segments. The circuit of Fig. 42 gives the current indicated there. Almost any shape of curve can be approximated by straight-line segments in this way; in particular a network can be made in which the current into the summing junction of an amplifier is approximately proportional to the square of the voltage applied.

The quarter-square multiplier uses two of these circuits, based on the identity

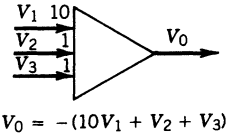


Figure 38 Symbol for summing amplifier.

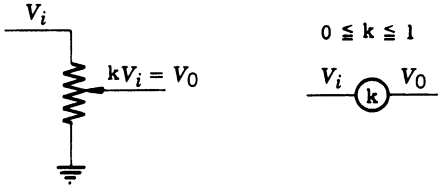


Figure 39 Potentiometer.

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

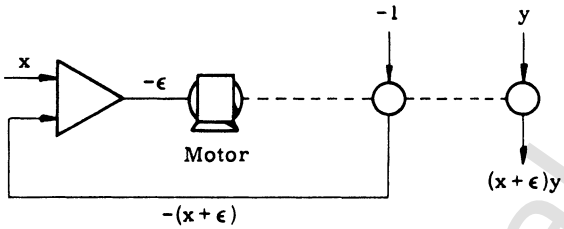


Figure 40 Servomultiplier.

$$xy = \tfrac{1}{4}\,[(x + y)^2 - (x - y)^2] \tag{107}$$

Its symbol is shown in Fig. 43. In most modern computers both polarities of input voltage must be provided. This has been indicated by two amplifiers on the inputs.

Division of voltages is accomplished by using a multiplier in the feedback path of an amplifier, as indicated in Fig. 44.

With the quarter-square multiplier, multiplication and division may be performed accurately at high frequencies. These complete the description of how the ordinary algebraic operations are performed on an analog computer.

Next, we turn to the operations of calculus, which enable differential equations to be solved. Consider the circuit of Fig. 45. The charge on a capacitor is the integral of the current; therefore the derivative of the charge is the current. The charge is the potential times the capacitance, which is fixed; thus current balance around the summing junction gives

$$\frac{V_1}{R} + C\dot{V}_2 = 0 \tag{108}$$

or

$$V_2(t) = V_0(0) - \frac{1}{RC}\int_0^t V_1(\tau)\,d\tau \tag{109}$$

Thus this circuit is an accurate integrator with a proportionality factor or time constant of RC. A common symbol is shown in Fig. 46. The initial condition can be set in with a special network furnished with most computers.

Differentiation can be performed with an input capacitor and a feedback resistor; however, it is not generally recommended because input signals are often contaminated by high-frequency noise, and the differentiator circuit amplifies this noise in direct proportion to its frequency. Since differential equations can always be written in terms of integrals, this causes no inconvenience.

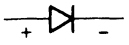


Figure 41 Diode.

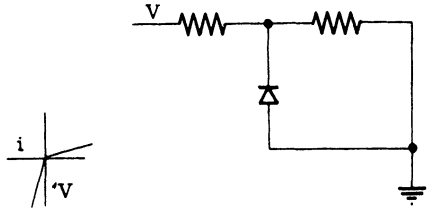


Figure 42 Diode network.

Copyrighted material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

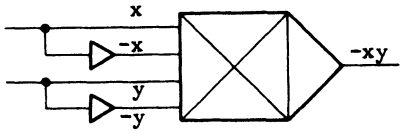


Figure 43 Quarter-square multiplier.

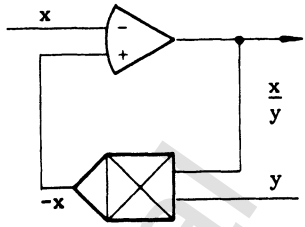


Figure 44 Division.

The circuit of Fig. 47 generates a simple time lag, as can be derived from its current balance:

$$\frac{V_1}{R_1} + \dot{V}_2 C + \frac{V_2}{R_2} = 0 \tag{110}$$

or

$$\dot{V}_2 + \frac{1}{R_2 C} V_2 = -\frac{V_1}{R_1 C} \tag{111}$$

The transfer function of this circuit is

$$G(s) = \frac{V_2}{V_1}(s) = \frac{-R_2/R_1}{R_2 C s + 1} = \frac{-R_2/R_1}{\tau s + 1}$$

Its time response to a pulse input is a simple exponential decay, with its time constant equal to R_2C.

The circuit corresponding to a damped sinusoid is shown in Fig. 48. Its differential equation is

$$\ddot{y} + 2\zeta\omega_n \dot{y} + \omega_n^2 y = \omega_n^2 f(t) \tag{112}$$

The transfer function is

$$G(s) = \frac{\omega_n^2}{s^2 + 2\omega_n \zeta s + \omega_n^2} = \frac{1}{s^2/\omega_n^2 + (2\zeta/\omega_n)\, s + 1} \tag{113}$$

A particular precaution that must be taken with analog computers is in scaling voltages to represent other physical variables. This must be done consistently for all variables and must be done in such a way that no amplifier is ever overloaded. Since amplifier outputs can never be fully predicted, some problems may require rescaling because of a bad initial

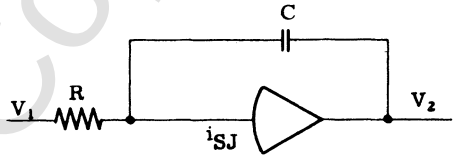


Figure 45 Current balance.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

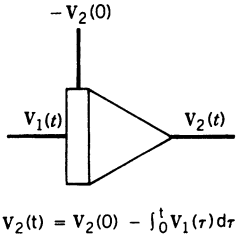


Figure 46 Integration.

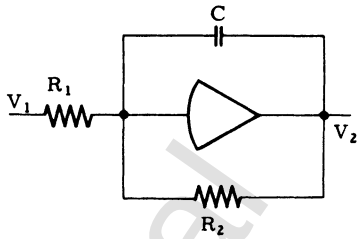


Figure 47 Time lag.

guess about the range of a dependent variable. This is particularly troublesome if nonlinear equipment is involved in the simulation. The necessity for proper prior scaling is one of the major disadvantages of analog computation.

With these basic circuit elements, the control engineer can proceed directly from a block diagram of a physical system to a wired program board without writing down differential equations because the circuits corresponding to the common transfer functions are combinations of those shown in Figs. 46–48. Special nonlinearities can be simulated by special networks. Simple ones such as multiplication, squaring, exponentiation, and generation of logarithms are often prewired directly to the program board.

Therefore, by suitable interconnection of components, even the most complicated system can be simulated in full dynamic character. Frequency response, stability limits, sensitivity to parameter changes, and effects of noise can all be tested and studied. Adjustments of the simulated controls for optimum rejection of noise can be done. Different types of control can be tried experimentally.

If care has been taken in formulating the system dynamics, then the analog computer can be made to behave precisely like the device it simulates.

Figures 49–53 show some special diode networks for simulating some of the special discontinuous nonlinearities that have been discussed previously.

Modem analog computers are generally equipped with additional types of special equipment, some of which are now listed and described below:

1. Comparator: Equivalent to the on–off control of Fig. 50.
2. Resolver: A circuit for rapidly and continuously converting from polar to Cartesian coordinates or the reverse.
3. Relays and switches: Operated by voltages generated in the simulation.
4. Memory amplifiers: An integrator into which an initial condition is set but to which no integrating input is patched will "remember" the initial condition.

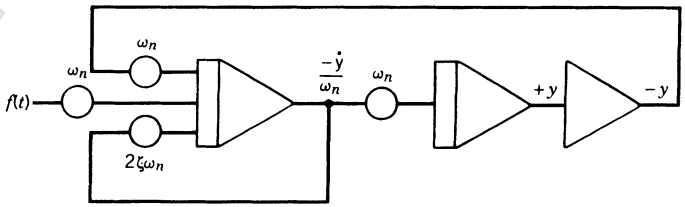


Figure 48 Sinusoid.

Copyrighted material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

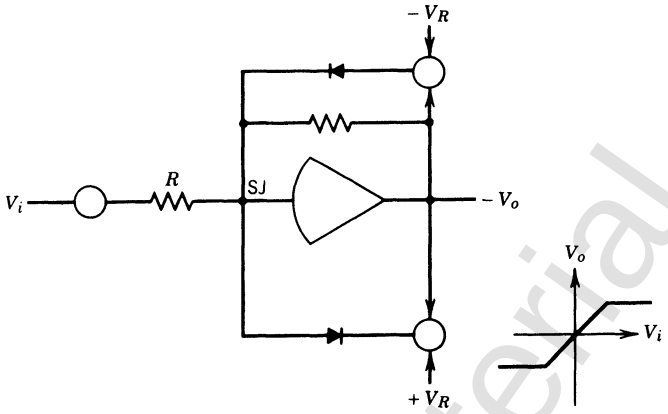


Figure 49 Saturation.

5. Repetitive operation at high speed: Solutions can be generated to repeat hundreds of times per second.
6. Outputs can be displayed on an oscilloscope or plotted on coordinate paper or on continuous recording equipment.

All these features make an analog computer a highly useful tool for design of control systems.

8.2 Digital Computation

An analog computer simultaneously solves all of the differential equations that model a physical system, and continuous voltage signals represent the variables of interest. This enables the machine to operate in real time. Significant disadvantages of analog simulation are the high cost of the computer due to the multiplicity of elements with demanding performance specifications, difficulties with scaling to avoid overloading of amplifiers, and rel-

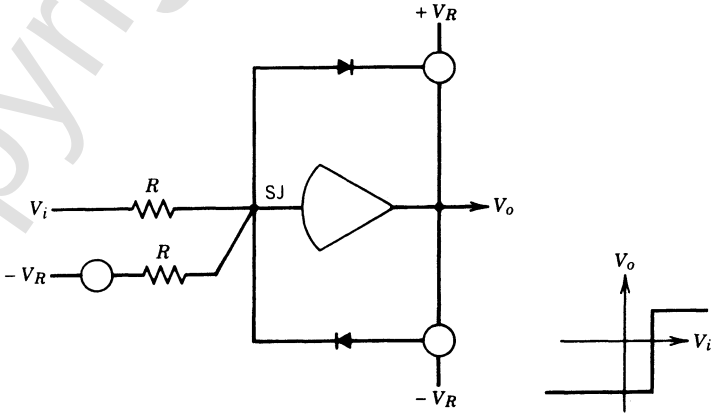


Figure 50 On–off controller.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

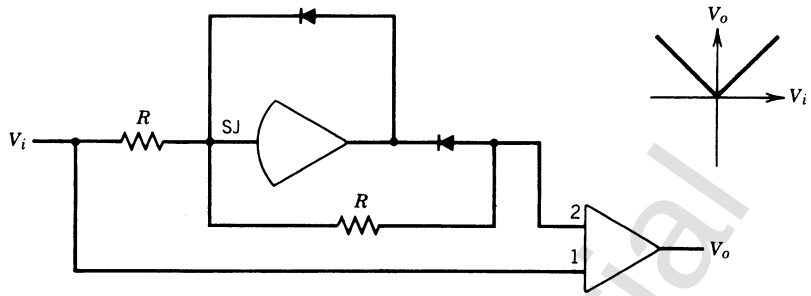


Figure 51 Rectifier.

atively limited accuracy and repeatability due in part to amplifier drift. As a consequence of the very rapid development of digital computer hardware and software giving ever greater capability and flexibility at reducing cost, system simulation is inevitably being carried out more and more on the digital computer. There is effectively no problem of overloading, enabling wide ranges of parameter variation to be accommodated.

The solution of a differential equation involves the process of integration, and for the digital computer analytical integration must be replaced by some numerical method that yields an approximation to the true solution. A number of special programming languages referred to as continuous system simulation languages (CSSLs) or simply as simulation languages are available as analytical tools to study the dynamic behavior of a wide range of systems without the need for a detailed knowledge of computing procedures. The languages are designed to be simple to understand and use, and they minimize programming difficulty by allowing the program to be written as a sequence of relatively self-descriptive statements. Numerous different languages with acronyms such as ACSL, CSMP, CSSL, DYNAMO, DARE, MIMIC, TELSIM, ENPORT, SCEPTRE, and SIMNON have been developed by computer manufacturers, software companies, universities, and others, some for specific families of machines and others for wider applications. Due to standardization, a number of the languages tend to be broadly similar. Symbolic names are used for the system variables, and the main body of the program is written as a series of simple statements based on the system state equations, block diagrams, or bond graphs. To these are added statements specifying initial parameter values and values of system constants and simple command statements controlling the running of the program and specifying the form in which the output is

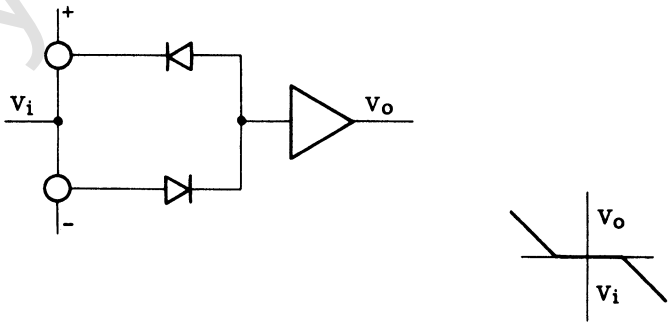


Figure 52 Dead zone.

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

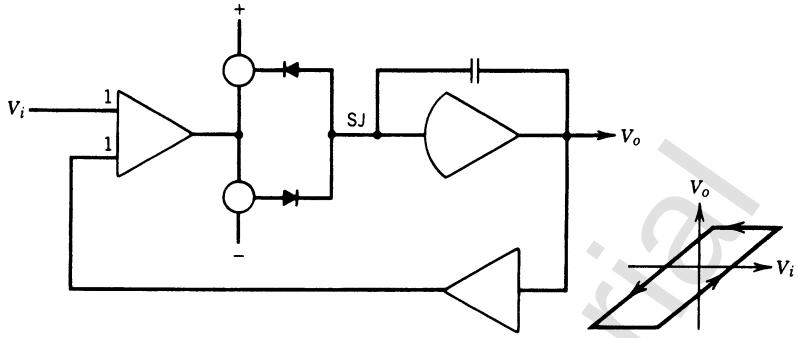


Figure 53 Hysteresis or backlash.

required. The short user-written program is then automatically translated into a FORTRAN (or other high-level language) program that is then compiled, loaded, and executed to produce a time history of the variables of interest in print or plot form. System constants and initial conditions can be altered and the program rerun without the need to retranslate and recompile. Many of the languages are designed to be run interactively from a graphics terminal and have the facility of displaying on the screen whichever output is of interest and, if the solution is not developing as desired, the capability of interrupting the program, changing parameters, and rerunning immediately.

The typical steps to be taken by a CSSL are shown in Fig. 54. The major process blocks include:

1. Reading the initial problem description, in which the circuit, bond graph, schematic, or block diagram information is communicated to the computer
2. Formulating the system and state equations
3. Performing the numerical integration of the state equations by a suitable method
4. Storing and displaying results

Most simulation languages differ with regard to use by the engineer in steps 1 and 2; for example, SCEPTRE for circuit descriptions, ENPORT for bond graph descriptions,

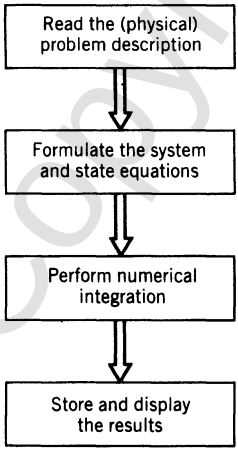


Figure 54 Typical simulation flowchart of the CSSL.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

CSMP for block diagrams, and DARE for state equations. The best guide to a program's characteristics is its user's manual.

In summary, the advantages of digital simulation are (i) simple program writing, (ii) accuracy and reproducibility, (iii) cost-effectiveness, (iv) where interactive facilities exist, keyboard entry of program and running, (v) inspection of plots on graphics display screen, and (vi) online modification and rerunning.

8.3 Hybrid Computation

Hybrid computation is a combination of analog and digital computations. In hybrid computations analog and digital machines are used in such a way that a problem can be programmed by exploiting the most efficient features of each. Much of the cost of such a machine arises from the rather complicated interface equipment required to make the necessary conversions between analog and digital signals and vice versa. The cost of an analog or hybrid computer is now justified only for certain large simulations where fast computing speed attained by parallel operation is important. Even this advantage may disappear as parallel computation is introduced into digital systems.

REFERENCES

1. E. O. Doebelin, *Control System Principles and Design,* Wiley, New York, 1985.

2. K. Ogata, *Modern Control Engineering,* Prentice-Hall, Englewood Cliffs, NJ, 1970.

3. M. Vidyasagar, *Nonlinear Systems Analysis,* Prentice-Hall, Englewood Cliffs, NJ, 1978.

4. Y. Takahashi, M. J. Rabins, and D. M. Auslander, *Control and Dynamic Systems,* Addison-Wesley, Reading, MA, 1972.

5. D. Graham and R. C. Lathrop, ''The Synthesis of Optimum Response: Criteria and Standard Forms,''AIEE *Transactions,* **72**, (Pt. II), 273–288 (1953).

6. R. C. Dorf, *Modern Control Systems,* Addison-Wesley, Reading, MA, 1986.

7. G. A. Korn and J. V. Wait, *Digital Continuous System Simulation,* Prentice-Hall, Englewood Cliffs, NJ, 1978.

8. O. W. Eshbach and M. Souders (eds.), *Handbook of Engineering Fundamentals,* 3rd ed., Wiley, New York, 1975.

BIBLIOGRAPHY

Bode, H. W., *Network Analysis and Feedback Amplifier Design,* Van Nostrand, New York, 1945.

Chestnut, H., and R. W. Mayer, *Servomechanisms and Regulating Systems Design,* 2nd ed., Vol. 1, Wiley, New York, 1959.

D'Azzo, J. J., and C. H. Houpis, *Linear Control System Analysis and Design,* McGraw-Hill, New York, 1988.

Distefano, J. J. III, A. R. Stubberud, and I. J. Williams, *Feedback and Control Systems* (Schaum's Outline Series), Shaum Publishing, New York, 1967.

Dransfield, P., *Engineering Systems and Automatic Control,* Prentice-Hall, Englewood Cliffs, NJ, 1968.

Elgerd, O. I., *Control Systems Theory,* McGraw-Hill, New York, 1967.

Evans, W. R., *Control-System Dynamics,* McGraw-Hill, New York, 1954.

Eveleigh, V. W., *Introduction to Control Systems Design,* McGraw-Hill, New York, 1972.

Franklin, G. F., J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems,* Addison-Wesley, Reading, MA, 1986.

Franklin, G. F., and J. D. Powell, *Digital Control of Dynamic Systems,* Addison-Wesley, Reading, MA, 1980.

Horowitz, I. M., *Synthesis of Feedback Systems,* Academic, New York, 1963.

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com

Houpis, C. H., and G. B. Lamont, *Digital Control Systems Theory, Hardware, Software,* McGraw-Hill, New York, 1985.

Kuo, B. C., *Digital Control Systems,* Holt, Rinehart and Winston, New York, 1980.

Kuo, B. C., *Automatic Control Systems,* Prentice-Hall, Englewood Cliffs, NJ, 1982.

Melsa, J. L., and D. G. Schultz, *Linear Control Systems,* McGraw-Hill, New York, 1969.

Nyquist, H., ''Regeneration Theory,'' Bell *System Technical Journal* **II,** 126–147 (1932).

Palm, W. J. III, *Modeling, Analysis and Control of Dynamic Systems,* Wiley, New York, 1983.

Phillips, C. L., and H. T. Nagle, Jr., *Digital Control System Analysis and Design,* Prentice-Hall, Englewood Cliffs, NJ, 1984.

Ragazzini, J. R., and G. F. Franklin, *Sampled Data Control Systems,* McGraw-Hill, New York, 1958.

Raven, F. H., *Automatic Control Engineering,* 4th ed., McGraw-Hill, New York, 1987.

Rosenberg, R. C., and D. C. Karnopp, *Introduction to Physical System Dynamics,* McGraw-Hill, New York, 1983.

Shinners, S. M., *Modern Control Systems Theory and Application,* Addison-Wesley, Reading, MA, 1972.

Truxal, J. G., *Automatic Feedback Control Synthesis,* McGraw-Hill, New York, 1955.

Copyrighted Material

Copyright © 2006 John Wiley & Sons Retrieved from: www.knovel.com