CHAPTER 27

MATHEMATICAL MODELS OF DYNAMIC PHYSICAL SYSTEMS

K. Preston White, Jr.
Department of Systems Engineering
University of Virginia
Charlottesville, Virginia

27.1 RATIONALE

The design of modern control systems relies on the formulation and analysis of mathematical models of dynamic physical systems. This is simply because a model is more accessible to study than the physical system the model represents. Models typically are less costly and less time consuming to construct and test. Changes in the structure of a model are easier to implement, and changes in the behavior of a model are easier to isolate and understand. A model often can be used to achieve insight when the corresponding physical system cannot, because experimentation with the actual system is too dangerous or too demanding. Indeed, a model can be used to answer "what if" questions about a system that has not yet been realized or actually cannot be realized with current technologies.

Mechanical Engineers' Handbook, 2nd ed., Edited by Myer Kutz.
ISBN 0-471-13007-9 © 1998 John Wiley & Sons, Inc.

The type of model used by the control engineer depends upon the nature of the system the model represents, the objectives of the engineer in developing the model, and the tools which the engineer has at his or her disposal for developing and analyzing the model. A mathematical model is a description of a system in terms of equations. Because the physical systems of primary interest to the control engineer are dynamic in nature, the mathematical models used to represent these systems most often incorporate difference or differential equations. Such equations, based on physical laws and observations, are statements of the fundamental relationships among the important variables that describe the system. Difference and differential equation models are expressions of the way in which the current values assumed by the variables combine to determine the future values of these variables.

Mathematical models are particularly useful because of the large body of mathematical and computational theory that exists for the study and solution of equations. Based on this theory, a wide range of techniques has been developed specifically for the study of control systems. In recent years, computer programs have been written that implement virtually all of these techniques. Computer software packages are now widely available for both simulation and computational assistance in the analysis and design of control systems.

It is important to understand that a variety of models can be realized for any given physical system. The choice of a particular model always represents a tradeoff between the fidelity of the model and the effort required in model formulation and analysis. This tradeoff is reflected in the nature and extent of simplifying assumptions used to derive the model. In general, the more faithful the model is as a description of the physical system modeled, the more difficult it is to obtain general solutions. In the final analysis, the best engineering model is not necessarily the most accurate or precise. It is, instead, the simplest model that yields the information needed to support a decision. A classification of various types of models commonly encountered by control engineers is given in Section 27.8.

A large and complicated model is justified if the underlying physical system is itself complex, if the individual relationships among the system variables are well understood, if it is important to understand the system with a great deal of accuracy and precision, and if time and budget exist to support an extensive study. In this case, the assumptions necessary to formulate the model can be minimized. Such complex models cannot be solved analytically, however. The model itself must be studied experimentally, using the techniques of computer simulation. This approach to model analysis is treated in Section 27.7.

Simpler models frequently can be justified, particularly during the initial stages of a control system study. In particular, systems that can be described by linear difference or differential equations permit the use of powerful analysis and design techniques. These include the transform methods of classical control theory and the state-variable methods of modern control theory. Descriptions of these standard forms for linear systems analysis are presented in Sections 27.4, 27.5, and 27.6.

During the past several decades, a unified approach for developing lumped-parameter models of physical systems has emerged. This approach is based on the idea of idealized system elements, which store, dissipate, or transform energy. Ideal elements apply equally well to the many kinds of physical systems encountered by control engineers. Indeed, because control engineers most frequently deal with systems that are part mechanical, part electrical, part fluid, and/or part thermal, a unified approach to these various physical systems is especially useful and economic. The modeling of physical systems using ideal elements is discussed further in Sections 27.2, 27.3, and 27.4.

Frequently, more than one model is used in the course of a control system study. Simple models that can be solved analytically are used to gain insight into the behavior of the system and to suggest candidate designs for controllers. These designs are then verified and refined in more complex models, using computer simulation. If physical components are developed during the course of a study, it is often practical to incorporate these components directly into the simulation, replacing the corresponding model components. An iterative, evolutionary approach to control systems analysis and design is depicted in Fig. 27.1.

27.2 IDEAL ELEMENTS

Differential equations describing the dynamic behavior of a physical system are derived by applying the appropriate physical laws. These laws reflect the ways in which energy can be stored and transferred within the system. Because of the common physical basis provided by the concept of energy, a general approach to deriving differential equation models is possible. This approach applies equally well to mechanical, electrical, fluid, and thermal systems and is particularly useful for systems that are combinations of these physical types.

27.2.1 Physical Variables

An idealized *two-terminal* or *one-port* element is shown in Fig. 27.2. Two *primary physical variables* are associated with the element: a through variable $f(t)$ and an across variable $v(t)$. *Through variables* represent quantities that are transmitted through the element, such as the force transmitted through a spring, the current transmitted through a resistor, or the flow of fluid through a pipe. Through variables have the same value at both ends or terminals of the element. *Across variables* represent the difference

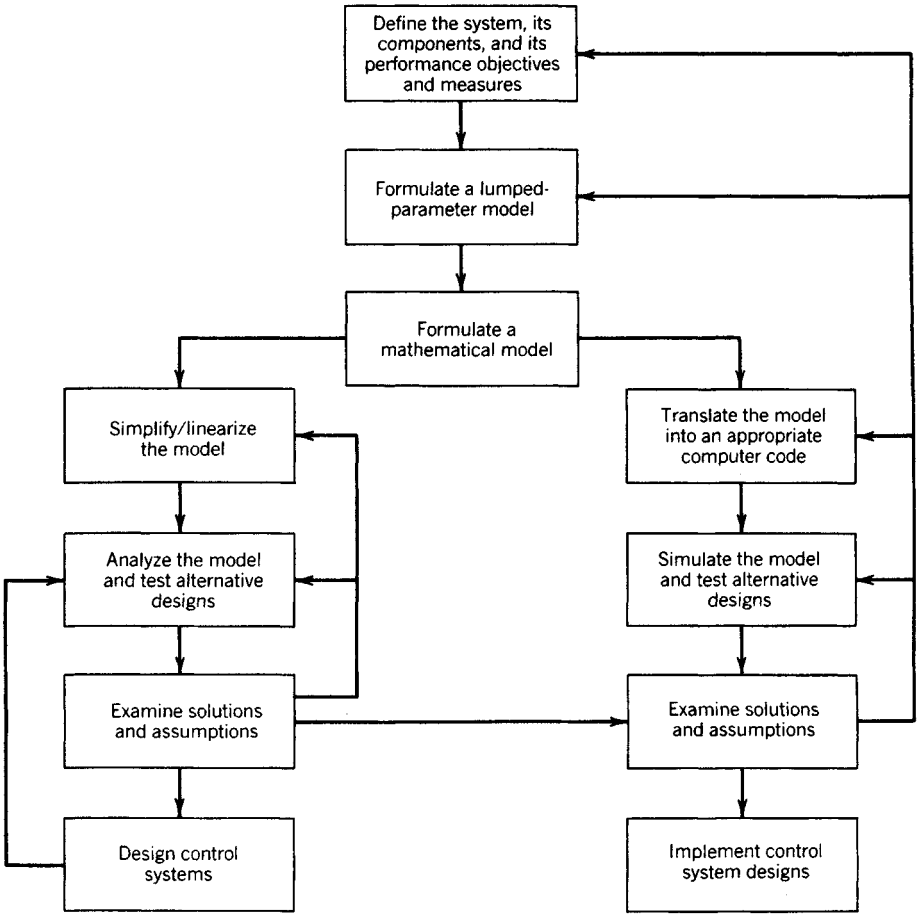


Fig. 27.1 An iterative approach to control system design, showing the use of mathematical analysis and computer simulation.

in state between the terminals of the element, such as the velocity difference across the ends of a spring, the voltage drop across a resistor, or the pressure drop across the ends of a pipe. *Secondary physical variables* are the integrated through variable $h(t)$ and the integrated across variable $x(t)$. These represent the accumulation of quantities within an element as a result of the integration of the associated through and across variables. For example, the momentum of a mass is an integrated through variable, representing the effect of forces on the mass integrated or accumulated over time. Table 27.1 defines the primary and secondary physical variables for various physical systems.

27.2.2 Power and Energy

The flow of *power* $P(t)$ into an element through the terminals 1 and 2 is the product of the through variable $f(t)$ and the difference between the across variables $v_2(t)$ and $v_1(t)$. Suppressing the notation for time dependence, this may be written as

$$P = f(v_2 - v_1) = fv_{21}$$

A negative value of power indicates that power flows out of the element. The *energy* $E(t_a, t_b)$ transferred to the element during the time interval from t_a to t_b is the integral of power, that is,

$$E = \int_{t_a}^{t_b} P\,dt = \int_{t_a}^{t_b} fv_{21}\,dt$$

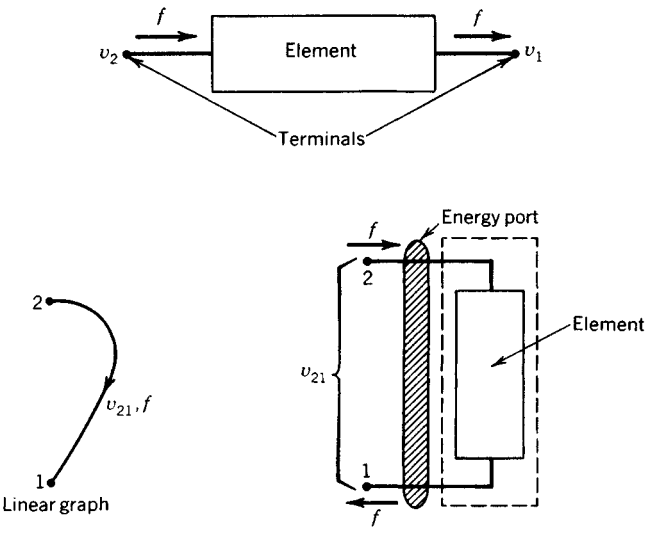


Fig. 27.2 A two-terminal or one-port element, showing through and across variables.[1]

A negative value of energy indicates a net transfer of energy out of the element during the corresponding time interval.

Thermal systems are an exception to these generalized energy relationships. For a thermal system, power is identically the through variable $q(t)$, heat flow. Energy is the integrated through variable $\mathcal{H}(t_a, t_b)$, the amount of heat transferred.

By the *first law of thermodynamics,* the net energy stored within a system at any given instant must equal the difference between all energy supplied to the system and all energy dissipated by the system. The generalized classification of elements given in the following sections is based on whether the element stores or dissipates energy within the system, supplies energy to the system, or transforms energy between parts of the system.

27.2.3 One-Port Element Laws

Physical devices are represented by idealized system elements, or by combinations of these elements. A physical device that exchanges energy with its environment through one pair of across and through variables is called a *one-port* or *two-terminal* element. The behavior of a one-port element expresses the relationship between the physical variables for that element. This behavior is defined mathematically by a *constitutive relationship.* Constitutive relationships are derived empirically, by experimentation, rather than from any more fundamental principles. The *element law,* derived from the corresponding constitutive relationship, describes the behavior of an element in terms of across and through variables and is the form most commonly used to derive mathematical models.

Table 27.1 Primary and Secondary Physical Variables for Various Systems[1]

System	Through Variable f	Integrated Through Variable h	Across Variable v	Integrated Across Variable x
Mechanical–translational	Force F	Translational momentum p	Velocity difference v_{21}	Displacement difference x_{21}
Mechanical–rotational	Torque T	Angular momentum h	Angular velocity difference Ω_{21}	Angular displacement difference Θ_{21}
Electrical	Current i	Charge q	Voltage difference v_{21}	Flux linkage λ_{21}
Fluid	Fluid flow Q	Volume V	Pressure difference P_{21}	Pressure–momentum Γ_{21}
Thermal	Heat flow q	Heat energy $\mathcal{H}$	Temperature difference θ_{21}	Not used in general

Table 27.2 summarizes the element laws and constitutive relationships for the one-port elements. Passive elements are classified into three types. *T-type* or *inductive storage* elements are defined by a single-valued constitutive relationship between the through variable $f(t)$ and the integrated across-variable difference $x_{21}(t)$. Differentiating the constitutive relationship yields the element law. For a linear (or ideal) *T*-type element, the element law states that the across-variable difference is proportional to the rate of change of the through variable. Pure translational and rotational compliance (springs), pure electrical inductance, and pure fluid inertance are examples of *T*-type storage elements. There is no corresponding thermal element.

A-type or *capacitive storage elements* are defined by a single-valued constitutive relationship between the across-variable difference $v_{21}(t)$ and the integrated through variable $h(t)$. These elements store energy by virtue of the across variable. Differentiating the constitutive relationship yields the element law. For a linear *A*-type element, the element law states that the through variable is proportional to the derivative of the across-variable difference. Pure translational and rotational inertia (masses), and pure electrical, fluid, and thermal capacitance are examples.

It is important to note that when a nonelectrical capacitance is represented by an *A*-type element, one terminal of the element must have a constant (reference) across variable, usually assumed to be zero. In a mechanical system, for example, this requirement expresses the fact that the velocity of a mass must be measured relative to a noninertial (nonaccelerating) reference frame. The constant velocity terminal of a pure mass may be thought of as being attached in this sense to the reference frame.

D-type or *resistive elements* are defined by a single-valued constitutive relationship between the across and the through variables. These elements dissipate energy, generally by converting energy into heat. For this reason, power always flows into a *D*-type element. The element law for a *D*-type energy dissipator is the same as the constitutive relationship. For a linear dissipator, the through variable is proportional to the across-variable difference. Pure translational and rotational friction (dampers or dashpots), and pure electrical, fluid, and thermal resistance are examples.

Energy-storage and energy-dissipating elements are called *passive* elements, because such elements do not supply outside energy to the system. The fourth set of one-port elements are *source elements,* which are examples of *active* or power-supplying elements. Ideal sources describe interactions between the system and its environment. A pure *A-type source* imposes an across-variable difference between its terminals, which is a prescribed function of time, regardless of the values assumed by the through variable. Similarly, a pure *T-type source* imposes a through-variable flow through the source element, which is a prescribed function of time, regardless of the corresponding across variable.

Pure system elements are used to represent physical devices. Such models are called *lumped-element models.* The derivation of lumped-element models typically requires some degree of approximation, since (1) there rarely is a one-to-one correspondence between a physical device and a set of pure elements and (2) there always is a desire to express an element law as simply as possible. For example, a coil spring has both mass and compliance. Depending on the context, the physical spring might be represented by a pure translational mass, or by a pure translational spring, or by some combination of pure springs and masses. In addition, the physical spring undoubtedly will have a nonlinear constitutive relationship over its full range of extension and compression. The compliance of the coil spring may well be represented by an ideal translational spring, however, if the physical spring is approximately linear over the range of extension and compression of concern.

27.2.4 Multiport Elements

A physical device that exchanges energy with its environment through two or more pairs of through and across variables is called a *multiport element.* The simplest of these, the idealized *four-terminal* or *two-port* element, is shown in Fig. 27.3. Two-port elements provide for transformations between the physical variables at different energy ports, while maintaining instantaneous continuity of power. In other words, net power flow into a two-port element is always identically zero:

$$P = f_a v_a + f_b v_b = 0$$

The particulars of the transformation between the variables define different categories of two-port elements.

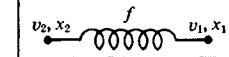
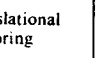
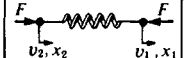
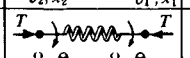
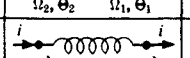
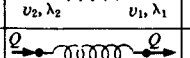
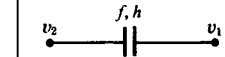
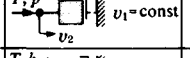
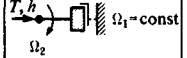
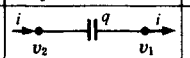
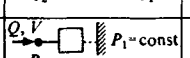
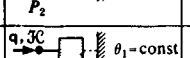
A *pure transformer* is defined by a single-valued constitutive relationship between the integrated across variables or between the integrated through variables at each port:

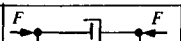
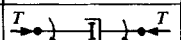
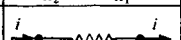
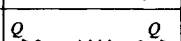
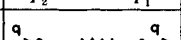
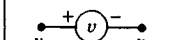
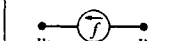
$$x_b = f(x_a) \qquad \text{or} \qquad h_b = f(h_a)$$

For a linear (or ideal) transformer, the relationship is proportional, implying the following relationships between the primary variables:

$$v_b = n v_a, \qquad f_b = -\frac{1}{n} f_a$$

Table 27.2 Element Laws and Constitutive Relationships for Various One-Port Elements[1]

Type of element	Physical element	Linear graph	Diagram	Constitutive relationship	Energy or power function	Ideal elemental equation	Ideal energy or power
T-type energy storage $\mathcal{E} \geqslant 0$  	Translational spring	 	 	$x_{21} = f(F)$	$\mathcal{E} = \int_0^F F\, dx_{21}$	$v_{21} = \frac{1}{k}\frac{dF}{dt}$	$\mathcal{E} = \frac{1}{2}\frac{F^2}{k}$
	Rotational spring	 	 	$\Theta_{21} = f(T)$	$\mathcal{E} = \int_0^T T\, d\Theta_{21}$	$\Omega_{21} = \frac{1}{K}\frac{dT}{dt}$	$\mathcal{E} = \frac{1}{2}\frac{T^2}{K}$
Pure: $x_{21} = f(f)$ $\mathcal{E} = \int_0^f f\, dx_{21}$ Ideal: $x_{21} = Lf$ $\mathcal{E} = \frac{1}{2}Lf^2$	Inductance	 	 	$\lambda_{21} = f(i)$	$\mathcal{E} = \int_0^i i\, d\lambda_{21}$	$v_{21} = L\frac{di}{dt}$	$\mathcal{E} = \frac{1}{2}Li^2$
	Fluid inertance	 	 	$\Gamma_{21} = f(Q)$	$\mathcal{E} = \int_0^Q Q\, d\Gamma_{21}$	$P_{21} = I\frac{dQ}{dt}$	$\mathcal{E} = \frac{1}{2}IQ^2$
A-type energy storage $\mathcal{E} \geqslant 0$  	Translational mass	 	 	$p = f(v_2)$	$\mathcal{E} = \int_0^{v_2} v_2\, dp$	$F = m\frac{dv_2}{dt}$	$\mathcal{E} = \frac{1}{2}mv_2^2$
	Inertia	 	 	$h = f(\Omega_2)$	$\mathcal{E} = \int_0^{\Omega_2} \Omega_2\, dh$	$T = J\frac{d\Omega_2}{dt}$	$\mathcal{E} = \frac{1}{2}J\Omega_2^2$
	Electrical capacitance	 	 	$q = f(v_{21})$	$\mathcal{E} = \int_0^{v_{21}} v_{21}\, dq$	$i = C\frac{dv_{21}}{dt}$	$\mathcal{E} = \frac{1}{2}Cv_{21}^2$
Pure: $h = f(v_{21})$ $\mathcal{E} = \int_0^{v_{21}} v_{21}\, dh$ Ideal: $h = Cv_{21}$ $\mathcal{E} = \frac{1}{2}Cv_{21}^2$	Fluid capacitance	 	 	$V = f(P_2)$	$\mathcal{E} = \int_0^{P_2} P_2\, dV$	$Q = C_f\frac{dP_2}{dt}$	$\mathcal{E} = \frac{1}{2}C_fP_2^2$
	Thermal capacitance	 	 	$\mathfrak{C} = f(\theta_2)$	$\mathcal{E} = \int_0^{\theta_2} \mathbf{q}\, dt = \mathfrak{C}$	$\mathbf{q} = C_t\frac{d\theta_2}{dt}$	$\mathcal{E} = C_t\theta_2$

D-type energy dissipators								
$\mathcal{P} \geqslant 0$		Translational damper			$F = \mathrm{f}(v_{21})$	$\mathcal{P} = Fv_{21}$	$F = bv_{21}$	$\mathcal{P} = bv_{21}^2$
		Rotational damper			$T = \mathrm{f}(\Omega_{21})$	$\mathcal{P} = T\Omega_{21}$	$T = B\Omega_{21}$	$\mathcal{P} = B\Omega_{21}^2$
		Electrical resistance			$i = \mathrm{f}(v_{21})$	$\mathcal{P} = iv_{21}$	$i = \frac{1}{R} v_{21}$	$\mathcal{P} = \frac{1}{R} v_{21}^2$
Pure	Ideal	Fluid resistance			$Q = \mathrm{f}(P_{21})$	$\mathcal{P} = QP_{21}$	$Q = \frac{1}{R_f} P_{21}$	$\mathcal{P} = \frac{1}{R_f} P_{21}^2$
$f = \mathrm{f}(v_{21})$	$f = \frac{1}{R} v_{21}$							
$\mathcal{P} = v_{21}\mathrm{f}(v_{21})$	$\mathcal{P} = \frac{1}{R} v_{21}^2 = Rf^2$	Thermal resistance			$\mathrm{q} = \mathrm{f}(\theta_{21})$	$\mathcal{P} = \mathrm{q}$	$\mathrm{q} = \frac{1}{R_t} \theta_{21}$	$\mathcal{P} = \frac{1}{R_t} \theta_{21}$
Energy sources		*A*-type across-variable source			$v_{21} = \mathrm{f}(t)$	$\mathcal{P} = fv_{21}$		
$\mathcal{P} \gtrless 0$								
$\mathcal{E} \gtrless 0$		*T*-type through-variable source			$f = \mathrm{f}(t)$	$\mathcal{P} = fv_{21}$		

Nomenclature

λ = energy, $\mathcal{P}$ = power

f = generalized through-variable, F = force, T = torque, i = current, Q = fluid flow rate, q = heat flow rate

h = generalized integrated through-variable, p = translational momentum, h = angular momentum, q = charge, l' = fluid volume displaced, $\mathcal{H}$ = heat

v = generalized across-variable, v = translational velocity, Ω = angular velocity, v = voltage, P = pressure, θ = temperature

x = generalized integrated across-variable, x = translational displacement, Θ = angular displacement, λ = flux linkage, Γ = pressure-momentum

L = generalized ideal inductance, $1/k$ = reciprocal translational stiffness, $1/K$ = reciprocal rotational stiffness, L = inductance, I = fluid inertance

C = generalized ideal capacitance, m = mass, J = moment of insertia, C = capacitance, C_j = fluid capacitance, C_t = thermal capacitance

R = generalized ideal resistance, $1/b$ = reciprocal translational damping, $1/B$ = reciprocal rotational damping, R = electrical resistance, R_j = fluid resistance, R_t = thermal resistance

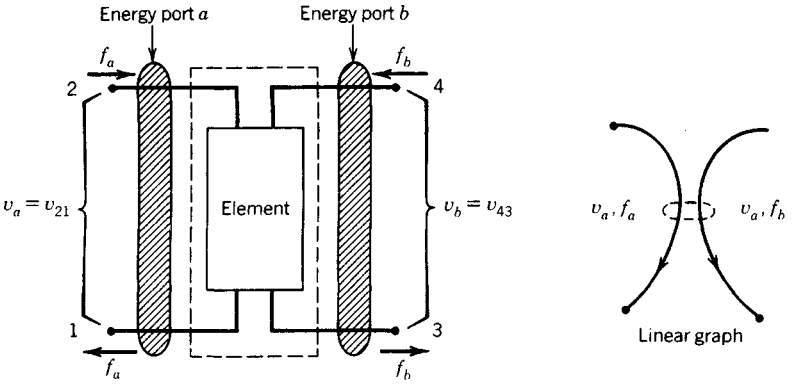


Fig. 27.3 A four-terminal or two-port element, showing through and across variables.

where the constant of proportionality n is called the *transformation ratio.* Levers, mechanical linkages, pulleys, gear trains, electrical transformers, and differential-area fluid pistons are examples of physical devices that typically can be approximated by pure or ideal transformers. Figure 27.4 depicts some examples. *Pure transmitters,* which serve to transmit energy over a distance, frequently can be thought of as transformers with $n = 1$.

A *pure gyrator* is defined by a single-valued constitutive relationship between the across variable at one energy port and the through variable at the other energy port. For a linear gyrator, the following relations apply:

$$v_b = rf_a, \qquad f_b = \frac{-1}{r} v_a$$

where the constant of proportionality is called the *gyration ratio* or *gyrational resistance.* Physical devices that perform pure gyration are not as common as those performing pure transformation. A mechanical gyroscope is one example of a system that might be modeled as a gyrator.

In the preceding discussion of two-port elements, it has been assumed that the type of energy is the same at both energy ports. A *pure transducer,* on the other hand, changes energy from one physical medium to another. This change may be accomplished either as a transformation or a gyration. Examples of *transforming transducers* are gears with racks (mechanical rotation to mechanical translation), and electric motors and electric generators (electrical to mechanical rotation and vice versa). Examples of *gyrating transducers* are the piston-and-cylinder (fluid to mechanical) and piezoelectric crystals (mechanical to electrical).

More complex systems may have a large number of energy ports. A common *six-terminal* or *three-port element* called a *modulator* is depicted in Fig. 27.5. The flow of energy between ports a and b is controlled by the energy input at the modulating port c. Such devices inherently dissipate energy, since

$$P_a + P_c \geq P_b$$

although most often the modulating power P_c is much smaller than the power input P_a or the power output P_b. When port a is connected to a pure source element, the combination of source and modulator is called a *pure dependent source.* When the modulating power P_c is considered the input and the modulated power P_b is considered the output, the modulator is called an *amplifier.* Physical devices that often can be modeled as modulators include clutches, fluid valves and couplings, switches, relays, transistors, and variable resistors.

27.3 SYSTEM STRUCTURE AND INTERCONNECTION LAWS

27.3.1 A Simple Example

Physical systems are represented by connecting the terminals of pure elements in patterns that approximate the relationships among the properties of component devices. As an example, consider the mechanical-translational system depicted in Fig. 27.6*a*, which might represent an idealized automobile suspension system. The inertial properties associated with the masses of the chassis, passenger compartment, engine, and so on, all have been lumped together as the pure mass m_1. The inertial prop-

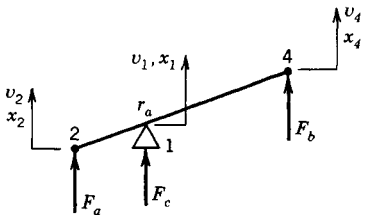
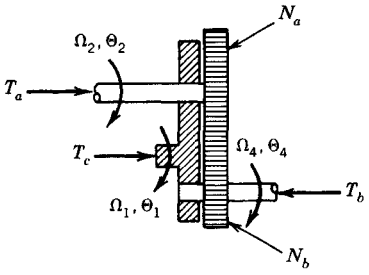
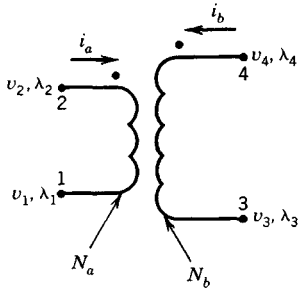
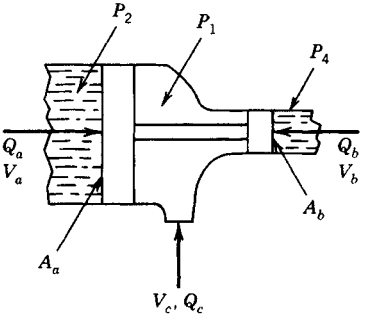
System	Symbol	Pure transformer	Ideal transformer	Transformation ratio
Mechanical translation (lever)		$x_{41} = f(x_{21})$	$v_{41} = nv_{21}$ $F_b = -\frac{1}{n} F_a$	$n = -\frac{r_b}{r_a}$ Lever ratio
Mechanical rotational (gears)		$\Theta_{41} = f(\Theta_2)$	$\Omega_{41} = n\Omega_{21}$ $T_b = -\frac{1}{n} T_a$	$n = -\frac{N_a}{N_b}$ Gear ratio
Electrical (magnetic)		$\lambda_{43} = f(\lambda_{21})$	$v_{43} = nv_{21}$ $i_b = -\frac{1}{n} i_a$	$n = \frac{N_b}{N_a}$ Turns ratio
Fluid (differential piston)		$V_b = f(V_a)$	$P_{41} = nP_{21}$ $Q_b = -\frac{1}{n} Q_a$	$n = \frac{A_a}{A_b}$ Area ratio

Fig. 27.4*a* Examples of transforms and transducers: pure transformers.[1]

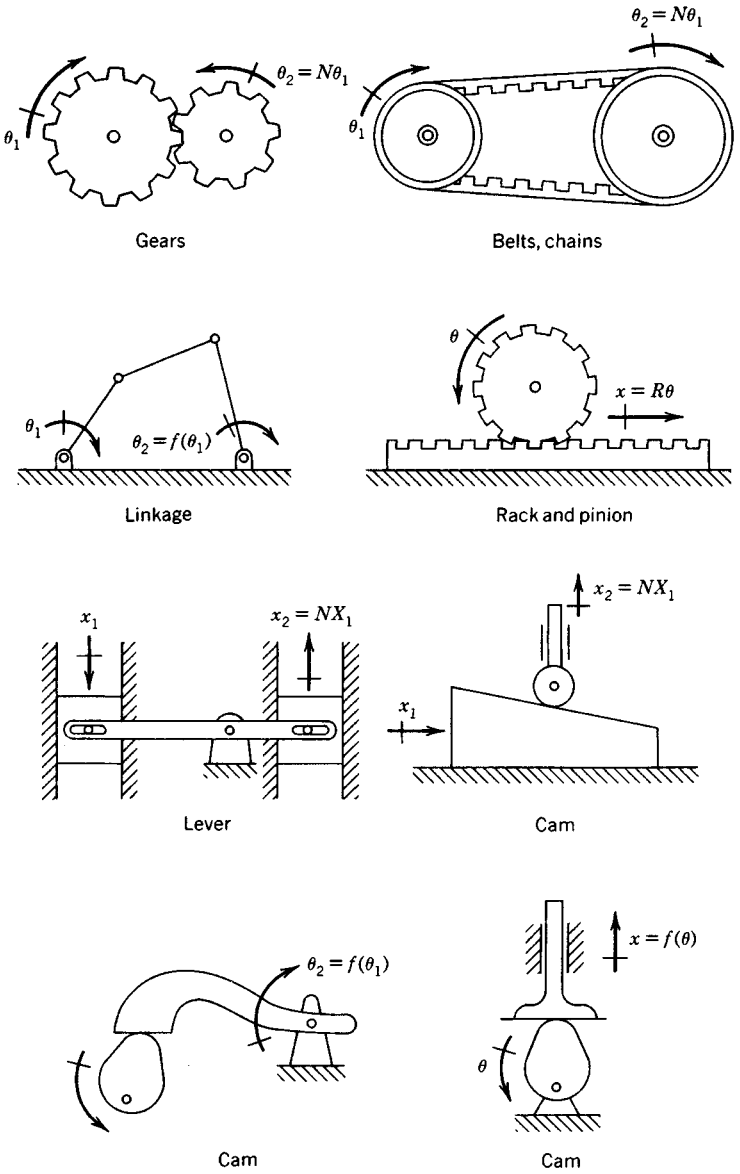


Fig. 27.4*b* Examples of transformers and transducers: pure mechanical transformers and transforming transducers.[2]

erties of the unsprung components (wheels, axles, etc.) have been lumped into the pure mass m_2. The compliance of the suspension is modeled as a pure spring with stiffness k_1 and the frictional effects (principally from the shock absorbers) as a pure damper with damping coefficient b. The road is represented as an input or source of vertical velocity, which is transmitted to the system through a spring of stiffness k_2, representing the compliance of the tires.

27.3.2 Structure and Graphs

The *pattern of interconnections* among elements is called the *structure* of the system. For a one-dimensional system, structure is conveniently represented by a *system graph.* The system graph for the idealized automobile suspension system of Fig. 27.6*a* is shown in Fig. 27.6*b*. Note that each distinct across variable (velocity) becomes a distinct *node* in the graph. Each distinct through variable

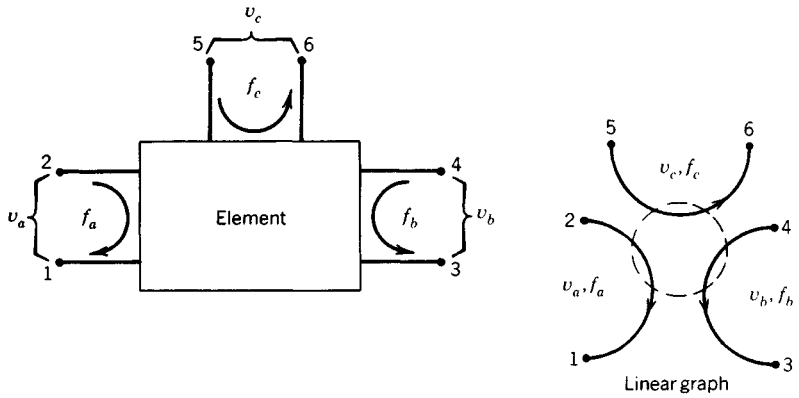
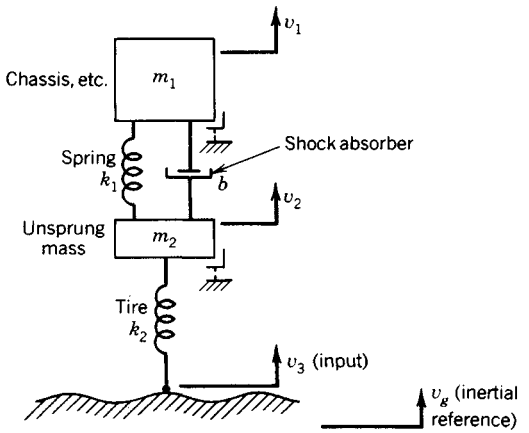
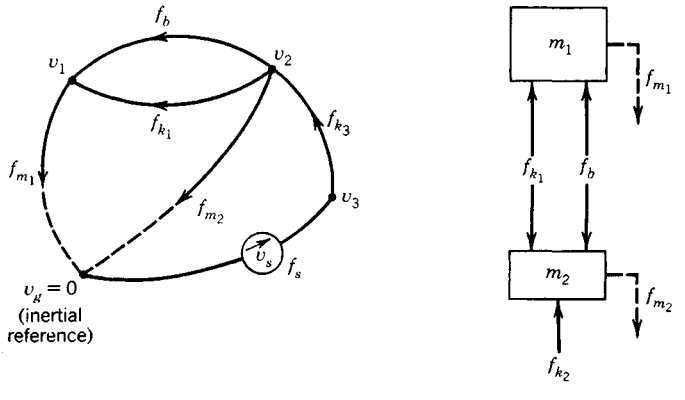


Fig. 27.5 A six-terminal or three-port element, showing through and across variables.



(a)



(b) *(c)*

Fig. 27.6 An idealized model of an automobile suspension system: (*a*) lumped-element model, (*b*) system graph, (*c*) free-body diagram.

(force) becomes a *branch* in the graph. Nodes coincide with the terminals of elements and branches coincide with the elements themselves. One node always represents *ground* (the constant velocity of the inertial reference frame v_g), and this is usually assumed to be zero for convenience. For non-electrical systems, all the A-type elements (masses) have one terminal connection to the reference node. Because the masses are not physically connected to ground, however, the convention is to represent the corresponding branches in the graph by dashed lines.

System graphs are oriented by placing arrows on the branches. The orientation is arbitrary and serves to assign reference directions for both the through-variable and the across-variable difference. For example, the branch representing the damper in Fig. 27.6*b* is directed from node 2 (tail) to node 1 (head). This assigns $v_b = v_{21} = v_2 - v_1$ as the across-variable difference to be used in writing the damper elemental equation

$$f_b = bv_b = bv_{21}$$

The reference direction for the through variable is determined by the convention that power flow $P_b = f_b v_b$ into an element is positive. Referring to Fig. 27.6*a*, when v_{21} is positive, the damper is in compression. Therefore, f_b must be positive for compressive forces in order to obey the sign convention for power. By similar reasoning, tensile forces will be negative.

27.3.3 System Relations

The structure of a system gives rise to two sets of *interconnection laws* or *system relations.* Continuity relations apply to through variables and compatibility relations apply to across variables. The interpretation of system relations for various physical systems is given in Table 27.3.

Continuity is a general expression of dynamic equilibrium. In terms of the system graph, continuity states that the algebraic sum of all through variables entering a given node must be zero. Continuity applies at each node in the graph. For a graph with n nodes, continuity gives rise to n continuity equations, $n - 1$ of which are independent. For node i, the continuity equation is

$$\sum_j f_{ij} = 0$$

where the sum is taken over all branches (i, j) incident on i.

For the system graph depicted in Fig. 27.6*b*, the four continuity equations are

$$\begin{aligned}
\text{node 1:} \quad & f_{k_1} + f_b - f_{m_1} = 0 \\
\text{node 2:} \quad & f_{k_2} - f_{k_1} - f_b - f_{m_2} = 0 \\
\text{node 3:} \quad & f_s - f_{k_2} = 0 \\
\text{node } g: \quad & f_{m_1} + f_{m_2} - f_s = 0
\end{aligned}$$

Only three of these four equations are independent. Note, also, that the equations for nodes 1 through 3 could have been obtained from the conventional *free-body diagrams* shown in Fig. 27.6*c*, where f_{m_1} and f_{m_2} are the *D'Alembert forces* associated with the pure masses. Continuity relations are also known as *vertex, node, flow,* and *equilibrium relations.*

Compatibility expresses the fact that the magnitudes of all across variables are scalar quantities. In terms of the system graph, compatibility states that the algebraic sum of the across-variable differences around any closed path in the graph must be zero. Compatibility applies to any closed path in the system. For convenience and to ensure the independence of the resulting equations, continuity is usually applied to the *meshes* or "windows" of the graph. A one-part graph with n nodes and b branches will have $b - n + 1$ meshes, each mesh yielding one independent compatibility equation. A planar graph with p separate parts (resulting from multiport elements) will have $b - n + p$ independent compatibility equations. For a closed path q, the compatibility equation is

Table 27.3 System Relations for Various Systems

System	Continuity	Compatibility
Mechanical	Newton's first and third laws (conservation of momentum)	Geometrical constraints (distance is a scalar)
Electrical	Kirchhoff's current law (conservation of charge)	Kirchhoff's voltage law (potential is a scalar)
Fluid	Conservation of matter	Pressure is a scalar
Thermal	Conservation of energy	Temperature is a scalar

$$\sum_q v_{ij} = 0$$

where the summation is taken over all branches (i, j) on the path.

For the system graph depicted in Fig. 27.6*b*, the three compatibility equations based on the meshes are

$$\begin{aligned} &\text{path } 1 \to 2 \to g \to 1: && -v_b + v_{m_2} - v_{m_1} = 0 \\ &\text{path } 1 \to 2 \to 1: && -v_{k_1} + v_b = 0 \\ &\text{path } 2 \to 3 \to g \to 2: && -v_{k_2} - v_s - v_{m_2} = 0 \end{aligned}$$

These equations are all mutually independent and express apparent geometric identities. The first equation, for example, states that the velocity difference between the ends of the damper is identically the difference between the velocities of the masses it connects. Compatibility relations are also known as *path, loop,* and *connectedness* relations.

27.3.4 Analogs and Duals

Taken together, the element laws and system relations are a complete mathematical model of a system. When expressed in terms of generalized through and across variables, the model applies not only to the physical system for which it was derived, but to any physical system with the same generalized system graph. Different physical systems with the same generalized model are called *analogs.* The mechanical rotational, electrical, and fluid analogs of the mechanical translational system of Fig. 27.6*a* are shown in Fig. 27.7. Note that because the original system contains an inductive storage element, there is no thermal analog.

Systems of the same physical type, but in which the roles of the through variables and the across variables have been interchanged, are called *duals.* The analog of a dual—or, equivalently, the dual of an analog—is sometimes called a *dualog.* The concepts of analogy and duality can be exploited in many different ways.

27.4 STANDARD FORMS FOR LINEAR MODELS

The element laws and system relations together constitute a complete mathematical description of a physical system. For a system graph with n nodes, b branches, and s sources, there will be $b - s$

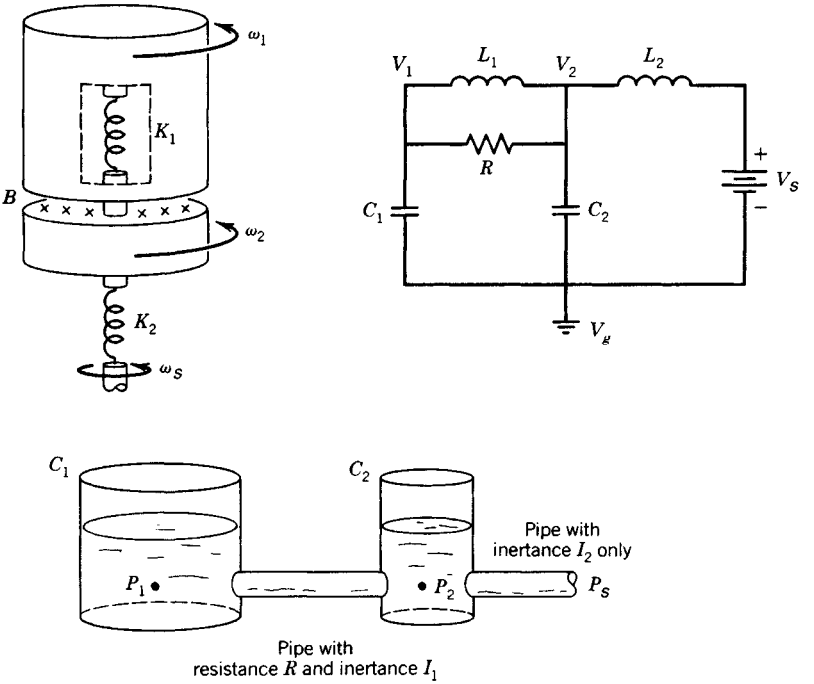


Fig. 27.7 Analogs of the idealized automobile suspension system depicted in Fig. 27.6.

element laws, $n - 1$ continuity equations, and $b - n + 1$ compatibility equations. This is a total of $2b - s$ differential and algebraic equations. For systems composed entirely of linear elements, it is always possible to reduce these $2b - s$ equations to either of two standard forms. The *input/output* or *I/O form* is the basis for *transform* or so-called *classical linear systems analysis.* The *state-variable form* is the basis for *state-variable* or so-called *modern linear systems analysis.*

27.4.1 I/O Form

The classical representation of a system is the "black box," depicted in Fig. 27.8. The system has a set of p *inputs* (also called *excitations* or *forcing functions*), $u_j(t)$, $j = 1, 2, \ldots, p$. The system also has a set of q *outputs* (also called *response variables*), $y_k(t)$, $k = 1, 2, \ldots, q$. Inputs correspond to sources and are assumed to be known functions of time. Outputs correspond to physical variables that are to be measured or calculated.

Linear systems represented in I/O form can be modeled mathematically by *I/O differential equations.* Denoting as $y_{kj}(t)$ that part of the kth output $y_k(t)$ that is attributable to the jth input $u_j(t)$, there are $(p \times q)$ I/O equations of the form

$$\frac{d^n y_{kj}}{dt^n} + a_{n-1}\frac{d^{n-1}y_{kj}}{dt^{n-1}} + \cdots + a_1\frac{dy_{kj}}{dt} + a_0 y_{kj}(t) = b_m\frac{d^m u_j}{dt^m} + b_{m-1}\frac{d^{m-1}u_j}{dt^{m-1}} + \cdots + b_1\frac{du_j}{dt} + b_0 u_j(t)$$

where $j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, q$. Each equation represents the dependence of one output and its derivatives on one input and its derivatives. By the *principle of superposition,* the kth output in response to all of the inputs acting simultaneously is

$$y_k(t) = \sum_{j=1}^{p} y_{kj}(t)$$

A system represented by nth-order I/O equations is called an nth-*order system.* In general, the order of a system is determined by the number of *independent* energy-storage elements within the system, that is, by the combined number of T-type and A-type elements for which the initial energy stored can be independently specified.

The coefficients $a_0, a_1, \ldots, a_{n-1}$ and $b_0, b_1, \ldots, b_m$ are parameter groups made up of algebraic combinations of the system physical parameters. For a system with constant parameters, therefore, these coefficients are also constant. Systems with constant parameters are called *time-invariant* systems and are the basis for classical analysis.

27.4.2 Deriving the I/O Form—An Example

I/O differential equations are obtained by combining element laws and continuity and compatibility equations in order to eliminate all variables except the input and the output. As an example, consider the mechanical system depicted in Fig. 27.9*a*, which might represent an idealized milling machine. A rotational motor is used to position the table of the machine tool through a rack and pinion. The motor is represented as a torque source T with inertia J and internal friction B. A flexible shaft, represented as a torsional spring K, is connected to a pinion gear of radius R. The pinion meshes with a rack, which is rigidly attached to the table of mass m. Damper b represents the friction opposing the motion of the table. The problem is to determine the I/O equation that expresses the relationship between the input torque T and the position of the table x.

The corresponding system graph is depicted in Fig. 27.9*b*. Applying continuity at nodes 1, 2, and 3 yields

$$\begin{aligned} \text{node 1:} \quad & T - T_J - T_B - T_K = 0 \\ \text{node 2:} \quad & T_K - T_p = 0 \\ \text{node 3:} \quad & -f_r - f_m - f_b = 0 \end{aligned}$$

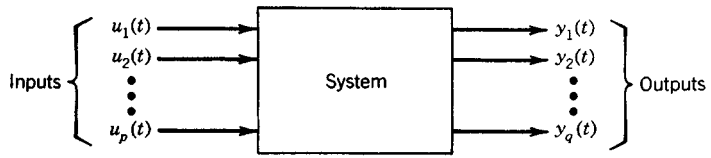


Fig. 27.8 Input/output (I/O) or "black box" representation of a dynamic system.

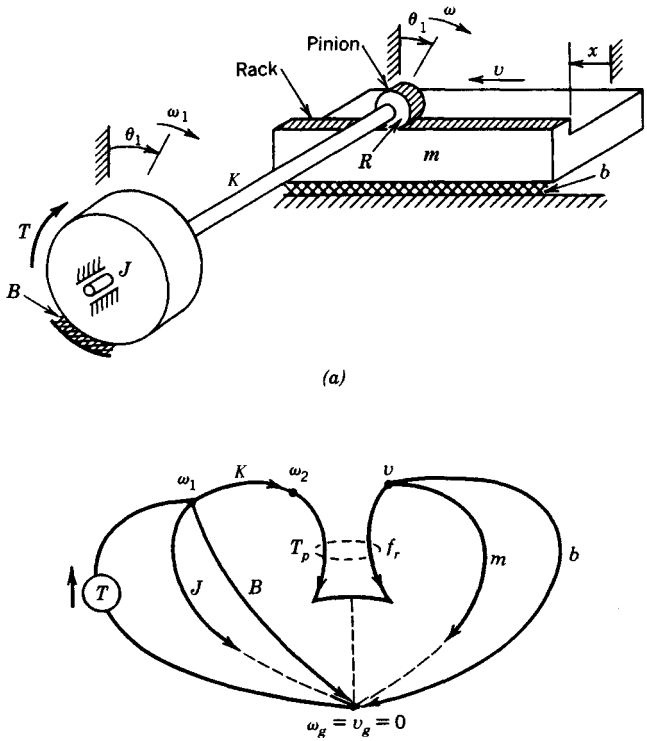


Fig. 27.9 An idealized model of a milling machine: (*a*) lumped-element model,[3] (*b*) system graph.

Substituting the elemental equation for each of the one-port elements into the continuity equations and assuming zero ground velocities yields

$$\begin{aligned}
\text{node 1:} \quad & T - J\dot{\omega}_1 - B\omega_1 - K\int(\omega_1 - \omega_2)dt = 0 \\
\text{node 2:} \quad & K\int(\omega_1 - \omega_2)dt - T_p = 0 \\
\text{node 3:} \quad & -f_r - m\dot{v} - bv = 0
\end{aligned}$$

Note that the definition of the across variables for each element in terms of the node variables, as above, guarantees that the compatibility equations are satisfied. With the addition of the constitutive relationships for the rack and pinion

$$\omega_2 = \frac{1}{R}v \qquad \text{and} \qquad T_p = -Rf_r$$

there are now five equations in the five unknowns ω_1, ω_2, v, T_p, and f_r. Combining these equations to eliminate all of the unknowns except v yields, after some manipulation,

$$a_3\frac{d^3v}{dt^3} + a_2\frac{d^2v}{dt^2} + a_1\frac{dv}{dt} + a_0v = b_1T$$

where

$$a_3 = Jm, \qquad a_1 = \frac{JK}{R^2} + Bb + mK, \qquad b_1 = \frac{K}{R}$$

$$a_2 = Jb + mB, \qquad a_0 = \frac{BK}{R^2} + Kb$$

Differentiating yields the desired I/O equation

$$a_3\frac{d^3x}{dt^3} + a_2\frac{d^2x}{dt^2} + a_1\frac{dx}{dt} + a_0x = b_1\frac{dT}{dt}$$

where the coefficients are unchanged.

For many systems, combining element laws and system relations can best be achieved by *ad hoc* procedures. For more complicated systems, formal methods are available for the orderly combination and reduction of equations. These are the so-called *loop method* and *node method* and correspond to procedures of the same names originally developed in connection with electrical networks. The interested reader should consult Ref. 1.

27.4.3 State-Variable Form

For systems with multiple inputs and outputs, the I/O model form can become unwieldy. In addition, important aspects of system behavior can be suppressed in deriving I/O equations. The "modern" representation of dynamic systems, called the *state-variable form,* largely eliminates these problems. A state-variable model is the maximum reduction of the original element laws and system relations that can be achieved without the loss of any information concerning the behavior of a system. State-variable models also provide a convenient representation for systems with multiple inputs and outputs and for systems analysis using computer simulation.

State variables are a set of variables $x_1(t), x_2(t), \ldots, x_n(t)$ internal to the system from which any set of outputs can be derived, as depicted schematically in Fig. 27.10. A set of state variables is the minimum number of independent variables such that by knowing the values of these variables at any time t_0 and by knowing the values of the inputs for all time $t \geq t_0$, the values of the state variables for all future time $t \geq t_0$ can be calculated. For a given system, the number n of state variables is unique and is equal to the order of the system. The definition of the state variables is not unique, however, and various combinations of one set of state variables can be used to generate alternative sets of state variables. For a physical system, the state variables summarize the *energy state* of the system at any given time.

A complete state-variable model consists of two sets of equations, the *state* or *plant equations* and the *output equations.* For the most general case, the state equations have the form

$$\begin{aligned}
\dot{x}_1(t) &= f_1[x_1(t),x_2(t), \ldots, x_n(t),u_1(t),u_2(t), \ldots, u_p(t)]\\
\dot{x}_2(t) &= f_2[x_1(t),x_2(t), \ldots, x_n(t),u_1(t),u_2(t), \ldots, u_p(t)]\\
&\vdots\\
\dot{x}_n(t) &= f_n[x_1(t),x_2(t), \ldots, x_n(t),u_1(t),u_2(t), \ldots, u_p(t)]
\end{aligned}$$

and the output equations have the form

$$\begin{aligned}
y_1(t) &= g_1[x_1(t),x_2(t), \ldots, x_n(t),u_1(t),u_2(t), \ldots, u_p(t)]\\
y_2(t) &= g_2[x_1(t),x_2(t), \ldots, x_n(t),u_1(t),u_2(t), \ldots, u_p(t)]\\
&\vdots\\
y_q(t) &= g_q[x_1(t),x_2(t), \ldots, x_n(t),u_1(t),u_2(t), \ldots, u_p(t)]
\end{aligned}$$

These equations are expressed more compactly as the two vector equations

$$\begin{aligned}
\dot{x}(t) &= f[x(t),u(t)]\\
y(t) &= g[x(t),u(t)]
\end{aligned}$$

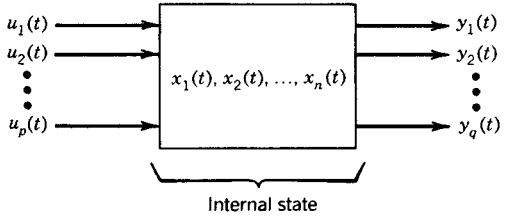


Fig. 27.10 State-variable representation of a dynamic system.

where

$$\dot{x}(t) = \text{the } (n \times 1) \textit{ state vector}$$
$$u(t) = \text{the } (p \times 1) \textit{ input or control vector}$$
$$y(t) = \text{the } (q \times 1) \textit{ output or response vector}$$

and f and g are vector-valued functions.

For linear systems, the state equations have the form

$$\begin{aligned}
\dot{x}_1(t) &= a_{11}(t)x_1(t) + \cdots + a_{1n}(t)x_n(t) + b_{11}(t)u_1(t) + \cdots + b_{1p}(t)u_p(t)\\
\dot{x}_2(t) &= a_{21}(t)x_1(t) + \cdots + a_{2n}(t)x_n(t) + b_{21}(t)u_1(t) + \cdots + b_{2p}(t)u_p(t)\\
&\vdots\\
\dot{x}_n(t) &= a_{n1}(t)x_1(t) + \cdots + a_{nn}(t)x_n(t) + b_{n1}(t)u_1(t) + \cdots + b_{np}(t)u_p(t)
\end{aligned}$$

and the output equations have the form

$$\begin{aligned}
y_1(t) &= c_{11}(t)x_1(t) + \cdots + c_{1n}(t)x_n(t) + d_{11}(t)u_1(t) + \cdots + d_{1p}(t)u_p(t)\\
y_2(t) &= c_{21}(t)x_1(t) + \cdots + c_{2n}(t)x_n(t) + d_{21}(t)u_1(t) + \cdots + d_{2p}(t)u_p(t)\\
&\vdots\\
y_n(t) &= c_{q1}(t)x_1(t) + \cdots + c_{qn}(t)x_n(t) + d_{q1}(t)u_1(t) + \cdots + d_{qp}(t)u_p(t)
\end{aligned}$$

where the coefficients are groups of parameters. The linear model is expressed more compactly as the two linear vector equations.

$$\begin{aligned}
\dot{x}(t) &= A(t)x(t) + B(t)u(t)\\
y(t) &= C(t)x(t) + D(t)u(t)
\end{aligned}$$

where the vectors x, u, and y are the same as the general case and the matrices are defined as

$A = [a_{ij}]$ is the $(n \times n)$ *system matrix*

$B = [b_{jk}]$ is the $(n \times p)$ *control, input,* or *distribution matrix*

$C = [c_{lj}]$ is the $(q \times n)$ *output matrix*

$D = [d_{lk}]$ is the $(q \times p)$ *output distribution matrix*

For a time-invariant linear system, all of these matrices are constant.

27.4.4 Deriving the "Natural" State Variables—A Procedure

Because the state variables for a system are not unique, there are an unlimited number of alternative (but equivalent) state-variable models for the system. Since energy is stored only in generalized system storage elements, however, a natural choice for the state variables is the set of through and across variables corresponding to the independent T-type and A-type elements, respectively. This definition is sometimes called the set of *natural state variables* for the system.

For linear systems, the following procedure can be used to reduce the set of element laws and system relations to the natural state-variable model.

Step 1. For each independent T-type storage, write the element law with the derivative of the through variable isolated on the left-hand side, that is, $\dot{f} = L^{-1}v$.

Step 2. For each independent A-type storage, write the element law with the derivative of the across variable isolated on the left-hand side, that is, $\dot{v} = C^{-1}f$.

Step 3. Solve the compatibility equations, together with the element laws for the appropriate D-type and multiport elements, to obtain each of the across variables of the independent T-type elements in terms of the natural state variables and specified sources.

Step 4. Solve the continuity equations, together with the element laws for the appropriate D-type and multiport elements, to obtain the through variables of the A-type elements in terms of the natural state variables and specified sources.

Step 5. Substitute the results of step 3 into the results of step 1; substitute the results of step 4 into the results of step 2.

Step 6. Collect terms on the right-hand side and write in vector form.

27.4.5 Deriving the "Natural" State Variables—An Example

The six-step process for deriving a natural state-variable representation, outlined in the preceding section, is demonstrated for the idealized automobile suspension depicted in Fig. 27.6:

Step 1

$$\dot{f}_{k_1} = k_1 v_{k_1}, \qquad \dot{f}_{k_2} = k_2 v_{k_2}$$

Step 2

$$\dot{v}_{m_1} = m_1^{-1} f_{m_1}, \qquad \dot{v}_{m_2} = m_2^{-1} f_{m_2}$$

Step 3

$$v_{k_1} = v_b = v_{m_2} - v_{m_1}, \qquad v_{k_2} = -v_{m_2} - v_s$$

Step 4

$$f_{m_1} = f_{k_1} + f_b = f_{k_1} + b^{-1}(v_{m_2} - v_{m_1})$$
$$f_{m_2} = f_{k_2} - f_{k_1} - f_b = f_{k_2} - f_{k_1} - b^{-1}(v_{m_2} - v_{m_1})$$

Step 5

$$\dot{f}_{k_1} = k_1(v_{m_2} - v_{m_1}), \qquad \dot{v}_{m_1} = m_1^{-1}[f_{k_1} + b^{-1}(v_{m_2} - v_{m_1})]$$
$$\dot{f}_{k_2} = k_2(-v_{m_2} - v_s), \qquad \dot{v}_{m_2} = m_2^{-1}[f_{k_2} - f_{k_1} - b^{-1}(v_{m_2} - v_{m_1})]$$

Step 6

$$\frac{d}{dt}\begin{bmatrix} f_{k_1} \\ f_{k_2} \\ v_{m_1} \\ v_{m_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -k_1 & k_1 \\ 0 & 0 & 0 & -k_2 \\ 1/m_1 & 0 & -1/m_1 b & 1/m_1 b \\ -1/m_2 & 1/m_2 & 1/m_2 b & -1/m_2 b \end{bmatrix}\begin{bmatrix} f_{k_1} \\ f_{k_2} \\ v_{m_1} \\ v_{m_2} \end{bmatrix} + \begin{bmatrix} 0 \\ -k_2 \\ 0 \\ 0 \end{bmatrix} v_s$$

27.4.6 Converting from I/O to "Phase-Variable" Form

Frequently, it is desired to determine a state-variable model for a dynamic system for which the I/O equation is already known. Although an unlimited number of such models is possible, the easiest to determine uses a special set of state variables called the *phase variables.* The phase variables are defined in terms of the output and its derivatives as follows:

$$x_1(t) = y(t)$$
$$x_2(t) = \dot{x}_1(t) = \frac{d}{dt}y(t)$$
$$x_3(t) = \dot{x}_2(t) = \frac{d^2}{dt^2}y(t)$$
$$\vdots$$
$$x_n(t) = \dot{x}_{n-1}(t) = \frac{d^{n-1}}{dt^{n-1}}y(t)$$

This definition of the phase variables, together with the I/O equation of Section 27.4.1, can be shown to result in a state equation of the form

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_{n-1}(t) \\ x_n(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_{n-1}(t) \\ x_n(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u(t)$$

and an output equation of the form

$$y(t) = [b_0 \quad b_1 \cdots b_m] \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

This special form of the system matrix, with ones along the upper off-diagonal and zeros elsewhere except for the bottom row, is called a *companion matrix.*

27.5 APPROACHES TO LINEAR SYSTEMS ANALYSIS

There are two fundamental approaches to the analysis of linear, time-invariant systems. *Transform methods* use rational functions obtained from the Laplace transformation of the system I/O equations. Transform methods provide a particularly convenient algebra for combining the component submodels of a system and form the basis of so-called *classical control theory. State-variable methods* use the vector state and output equations directly. State-variable methods permit the adaptation of important ideas from linear algebra and form the basis for so-called *modern control theory.* Despite the deceiving names of "classical" and "modern," the two approaches are complementary. Both approaches are widely used in current practice and the control engineer must be conversant with both.

27.5.1 Transform Methods

A *transformation* converts a given mathematical problem into an equivalent problem, according to some well-defined rule called a *transform.* Prudent selection of a transform frequently results in an equivalent problem that is easier to solve than the original. If the solution to the original problem can be recovered by an inverse transformation, the three-step process of (1) transformation, (2) solution in the *transform domain,* and (3) inverse transformation, may prove more attractive than direct solution of the problem in the original problem domain. This is true for fixed linear dynamic systems under the *Laplace transform,* which converts differential equations into equivalent algebraic equations.

Laplace Transforms: Definition

The one-sided Laplace transform is defined as

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

and the inverse transform as

$$f(t) = \mathcal{L}^{-1}[F(s)] = \frac{1}{2\pi j}\int_{\sigma-j\omega}^{\sigma+j\omega} F(s)e^{-st}\,ds$$

The Laplace transform converts the function $f(t)$ into the transformed function $F(s)$; the inverse transform recovers $f(t)$ from $F(s)$. The symbol $\mathcal{L}$ stands for the "Laplace transform of"; the symbol $\mathcal{L}^{-1}$ stands for "the inverse Laplace transform of."

The Laplace transform takes a problem given in the *time domain,* where all physical variables are functions of the *real variable t,* into the *complex-frequency domain,* where all physical variables are functions of the complex frequency $s = \sigma + j\omega$, where $j = \sqrt{-1}$ is the imaginary operator. Laplace transform pairs consist of the function $f(t)$ and its transform $F(s)$. Transform pairs can be calculated by substituting $f(t)$ into the defining equation and then evaluating the integral with s held constant. For a transform pair to exist, the corresponding integral must converge, that is,

$$\int_0^\infty |f(t)|e^{-\sigma^* t}\,dt < \infty$$

for some real $\sigma^* > 0$. Signals that are physically realizable always have a Laplace transform.

Tables of Transform Pairs and Transform Properties

Transform pairs for functions commonly encountered in the analysis of dynamic systems rarely need to be calculated. Instead, pairs are determined by reference to a *table of transforms* such as that given in Table 27.4. In addition, the Laplace transform has a number of properties that are useful in determining the transforms and inverse transforms of functions in terms of the tabulated pairs. The most important of these are given in a *table of transform properties* such as that given in Table 27.5.

Table 27.4 Laplace Transform Pairs

$F(s)$	$f(t),\ t \geq 0$
1. 1	$\delta(t)$, the unit impulse at $t = 0$
2. $\frac{1}{s}$	1, the unit step
3. $\frac{n!}{s^{n+1}}$	t^n
4. $\frac{1}{s + a}$	e^{-at}
5. $\frac{1}{(s + a)^n}$	$\frac{1}{(n - 1)!} t^{n-1} e^{-at}$
6. $\frac{a}{s(s + a)}$	$1 - e^{-at}$
7. $\frac{1}{(s + a)(s + b)}$	$\frac{1}{(b - a)} (e^{-at} - e^{-bt})$
8. $\frac{s + p}{(s + a)(s + b)}$	$\frac{1}{(b - a)} [(p - a)e^{-at} - (p - b)e^{-bt}]$
9. $\frac{1}{(s + a)(s + b)(s + c)}$	$\frac{e^{-at}}{(b - a)(c - a)} + \frac{e^{-bt}}{(c - b)(a - b)} + \frac{e^{-ct}}{(a - c)(b - c)}$
10. $\frac{s + p}{(s + a)(s + b)(s + c)}$	$\frac{(p - a)e^{-at}}{(b - a)(c - a)} + \frac{(p - b)e^{-bt}}{(c - b)(a - b)} + \frac{(p - c)e^{-ct}}{(a - c)(b - c)}$
11. $\frac{b}{s^2 + b^2}$	$\sin bt$
12. $\frac{s}{s^2 + b^2}$	$\cos bt$
13. $\frac{b}{(s + a)^2 + b^2}$	$e^{-at} \sin bt$
14. $\frac{s + a}{(s + a)^2 + b^2}$	$e^{-at} \cos bt$
15. $\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$	$\frac{\omega_n}{\sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1 - \zeta^2}\, t, \quad \zeta < 1$
16. $\frac{\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)}$	$1 + \frac{1}{\sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1 - \zeta^2}\, t + \phi)$ $\phi = \tan^{-1} \frac{\sqrt{1 - \zeta^2}}{\zeta} + \pi$ (third quadrant)

Poles and Zeros

The response of a dynamic system most often assumes the following form in the complex-frequency domain

$$F(s) = \frac{N(s)}{D(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0} \tag{27.1}$$

Functions of this form are called *rational functions,* because these are the ratio of two polynomials $N(s)$ and $D(s)$. If $n \geq m$, then $F(s)$ is a *proper rational function;* if $n > m$, then $F(s)$ is a *strictly proper rational function.*

In factored form, the rational function $F(s)$ can be written as

$$F(s) = \frac{N(s)}{D(s)} = \frac{b_m(s - z_1)(s - z_2) \cdots (s - z_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)} \tag{27.2}$$

Table 27.5 Laplace Transform Properties

$f(t)$	$F(s) = \int_0^\infty f(t)e^{-st}\,dt$
1. $af_1(t) + bf_2(t)$	$aF_1(s) + bF_2(s)$
2. $\dfrac{df}{dt}$	$sF(s) - f(0)$
3. $\dfrac{d^2f}{dt^2}$	$s^2F(s) - sf(0) - \left.\dfrac{df}{dt}\right\|_{t=0}$
4. $\dfrac{d^nf}{dt^n}$	$s^nF(s) - \sum_{k=1}^{n} s^{n-k}g_{k-1}$, $g_{k-1} = \left.\dfrac{d^{k-1}f}{dt^{k-1}}\right\|_{t=0}$
5. $\int_0^t f(t)\,dt$	$\dfrac{F(s)}{s} + \dfrac{h(0)}{s}$, $h(0) = \left.\int f(t)\,dt\right\|_{t=0}$
6. $\begin{cases} 0, & t < D \\ f(t-D), & t \geq D \end{cases}$	$e^{-sD}F(s)$
7. $e^{-at}f(t)$	$F(s + a)$
8. $f\left(\dfrac{t}{a}\right)$	$aF(as)$
9. $f(t) = \int_0^t x(t-\tau)y(\tau)\,d\tau = \int_0^t y(t-\tau)x(\tau)\,d\tau$	$F(s) = X(s)Y(s)$
10. $f(\infty) = \lim_{s\to 0} sF(s)$	
11. $f(0+) = \lim_{s\to\infty} sF(s)$	

The roots of the numerator polynomial $N(s)$ are denoted by z_j, $j = 1, 2, \ldots, m$. These numbers are called the *zeros* of $F(s)$, since $F(z_j) = 0$. The roots of the denominator polynomial are denoted by p_i, $1, 2, \ldots, n$. These numbers are called the *poles* of $F(s)$, since $\lim_{s\to p_i} F(s) = \pm\infty$.

Inversion by Partial-Fraction Expansion

The *partial-fraction expansion theorem* states that a strictly proper rational function $F(s)$ with *distinct (nonrepeated) poles* p_i, $i = 1, 2, \ldots, n$, can be written as the sum

$$F(s) = \frac{A_1}{s - p_1} + \frac{A_2}{s - p_2} + \cdots + \frac{A_n}{s - p_n} = \sum_{i=1}^{n} A_i \left(\frac{1}{s - p_i}\right) \tag{27.3}$$

where the A_i, $i = 1, 2, \ldots, n$, are constants called *residues*. The inverse transform of $F(s)$ has the simple form

$$f(t) = A_1e^{p_1t} + A_2e^{p_2t} + \cdots + A_ne^{p_nt} = \sum_{i=1}^{n} A_ie^{p_it}$$

The *Heaviside expansion theorem* gives the following expression for calculating the residue at the pole p_i,

$$A_i = (s - p_i)F(s)\big|_{s=p_i} \qquad \text{for } i = 1, 2, \ldots, n$$

These values can be checked by substituting into Eq. (27.3), combining the terms on the right-hand

side of Eq. (27.3), and showing the result yields the values for all the coefficients b_j, $j = 1, 2, \ldots, m$, originally specified in the form of Eq. (27.3).

Repeated Poles

When two or more poles of a strictly proper rational function are identical, the poles are said to be *repeated* or *nondistinct.* If a pole is repeated q times, that is, if $p_i = p_{i+1} = \cdots = p_{i+q-1}$, then the pole is said to be of *multiplicity* q. A strictly proper rational function with a pole of multiplicity q will contain q terms of the following form

$$\frac{A_{i1}}{(s - p_i)^q} + \frac{A_{i2}}{(s - p_i)^{q-1}} + \cdots + \frac{A_{iq}}{(s - p_i)}$$

in addition to the terms associated with the distinct poles. The corresponding terms in the inverse transform are

$$\left(\frac{1}{(q-1)!} A_{i1} t^{(q-1)} + \frac{1}{(q-2)!} A_{i2} t^{(q-2)} + \cdots + A_{iq}\right) e^{p_i t}$$

The corresponding residues are

$$A_{i1} = (s - p_i)^q F(s)\big|_{s=p_i}$$

$$A_{i2} = \left(\frac{d}{ds}\left[(s - p_i)^q F(s)\right]\right)\Bigg|_{s=p_i}$$

$$\vdots$$

$$A_{iq} = \frac{1}{(q-1)!}\left(\frac{d^{(q-1)}}{ds^{(q-1)}}\left[(s - p_i)^q F(s)\right]\right)\Bigg|_{s=p_i}$$

Complex Poles

A strictly proper rational function with complex conjugate poles can be inverted using partial-fraction expansion. Using a method called *completing the square,* however, is almost always easier. Consider the function

$$\begin{aligned} F(s) &= \frac{B_1 s + B_2}{(s + \sigma - j\omega)(s + \sigma + j\omega)} \\ &= \frac{B_1 s + B_2}{s^2 + 2\sigma s + \sigma^2 + \omega^2} \\ &= \frac{B_1 s + B_2}{(s + \sigma)^2 + \omega_2} \end{aligned}$$

From the transform tables the Laplace inverse is

$$\begin{aligned} f(t) &= e^{-\sigma t}[B_1 \cos \omega t + B_3 \sin \omega t] \\ &= Ke^{-\sigma t} \cos(\omega t + \phi) \end{aligned}$$

where $B_3 = (1/\omega)(B_2 - aB_1)$
$K = \sqrt{B_1^2 + B_3^2}$
$\phi = -\tan^{-1}(B_3/B_1)$

Proper and Improper Rational Functions

If $F(s)$ is not a strictly proper rational function, then $N(s)$ must be divided by $D(s)$ using *synthetic division.* The result is

$$F(s) = \frac{N(s)}{D(s)} = P(s) + \frac{N^*(s)}{D(s)}$$

where $P(s)$ is a polynomial of degree $m - n$ and $N^*(s)$ is a polynomial of degree $n - 1$. Each term of $P(s)$ may be inverted directly using the transform tables. $N^*(s)/D(s)$ is a strictly proper rational function and may be inverted using partial-fraction expansion.

Initial-Value and Final-Value Theorems

The limits of $f(t)$ as time approaches zero or infinity frequently can be determined directly from the transform $F(s)$ without inverting. The *initial-value theorem* states that

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

where the limit exists. If the limit does not exist (i.e., is infinite), the value of $f(0_+)$ is undefined. The *final-value theorem* states that

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

provided that (with the possible exception of a single pole at $s = 0$) $F(s)$ has no poles with nonnegative real parts.

Transfer Functions

The Laplace transform of system I/O equation may be written in terms of the transform $Y(s)$ of the system response $y(t)$ as

$$Y(s) = \frac{G(s)N(s) + F(s)D(s)}{P(s)D(s)} = \left(\frac{G(s)}{P(s)}\right)\left(\frac{N(s)}{D(s)}\right) + \frac{F(s)}{P(s)}$$

where (a) $P(s) = a_n s^n + a_{n-1} + \cdots + a_1 s + a_0$ is the *characteristic polynomial* of the system,
(b) $G(s) = b_m s^m + b_{m-1} s^{m-1} + \cdots + b_1 s + b_0$ represents the *numerator dynamics* of the system,
(c) $U(s) = N(s)/D(s)$ is the transform of the input to the system, $u(t)$, assumed to be a rational function, and

$$\text{(d) } F(s) = a_n y(0) s^{n-1} + \left(a_n \frac{dy}{dt}(0) + a_{n-1} y(0)\right) s^{n-2} + \cdots + \left(a_n \frac{d^{n-1}y}{dt^{n-1}}(0) + a_{n-1}\frac{d^{n-2}y}{dt}(0) + \cdots + a_1 y(0)\right)$$

reflects the initial system state [i.e., the initial conditions on $y(t)$ and its first $n - 1$ derivatives].

The transformed response can be thought of as the sum of two components,

$$Y(s) = Y_{zs}(s) + Y_{zi}(s)$$

where (e) $Y_{zs}(s) = [G(s)/P(s)][N(s)/D(s)] = H(s)U(s)$ is the transform of the *zero-state response,* that is, the response of the system to the input alone, and
(f) $Y_{zi}(s) = F(s)/P(s)$ is the transform of the *zero-input response,* that is, the response of the system to the initial state alone.

The rational function
(g) $H(s) = Y_{zs}(s)/U(s) = G(s)/P(s)$ is the *transfer function* of the system, defined as the Laplace transform of the ratio of the system response to the system input, assuming zero initial conditions.

The transfer function plays a crucial role in the analysis of fixed linear systems using transforms and can be written directly from knowledge of the system I/O equation as

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

Impulse Response

Since $U(s) = 1$ for a unit impulse function, the transform of the zero-state response to a unit impulse input is given by the relation (g) as

$$Y_{zs}(s) = H(s)$$

that is, the system transfer function. In the time domain, therefore, the unit *impulse response* is

$$h(t) = \begin{cases} 0 & \text{for } t \leq 0 \\ \mathcal{L}^{-1}[H(s)] & \text{for } t > 0 \end{cases}$$

This simple relationship is profound for several reasons. First, this provides for a direct characterization of time-domain response $h(t)$ in terms of the properties (poles and zeros) of the rational function $H(s)$ in the complex-frequency domain. Second, applying the convolution transform pair (Table 27.5) to relation (e) above yields

$$Y_{zs}(t) = \int_0^t h(\tau)u(t - \tau)\, d\tau$$

In words, the zero-state output corresponding to an arbitrary input $u(t)$ can be determined by convolution with the impulse response $h(t)$. In other words, the impulse response completely characterizes the system. The impulse response is also called the system *weighing function.*

Block Diagrams

Block diagrams are an important conceptual tool for the analysis and design of dynamic systems, because block diagrams provide a graphic means for depicting the relationships among system variables and components. A block diagram consists of unidirectional blocks representing specified system components or subsystems, interconnected by arrows representing system variables. Causality follows in the direction of the arrows, as in Fig. 27.11, indicating that the output is caused by the input acting on the system defined in the block.

Combining transform variables, transfer functions, and block diagrams provides a powerful graphical means for determining the overall transfer function of a system, when the transfer functions of its component subsystems are known. The basic blocks in such diagrams are given in Fig. 27.12. A block diagram comprising many blocks and summers can be reduced to a single transfer function block by using the diagram transformations given in Fig. 27.13.

27.5.2 Transient Analysis Using Transform Methods

Basic to the study of dynamic systems are the concepts and terminology used to characterize system behavior or performance. These ideas are aids in *defining* behavior, in order to consider for a given context those features of behavior which are desirable and undesirable; in *describing* behavior, in order to communicate concisely and unambiguously various behavioral attributes of a given system; and in *specifying* behavior, in order to formulate desired behavioral norms for system design. Characterization of dynamic behavior in terms of standard concepts also leads in many cases to analytical shortcuts, since key features of the system response frequently can be determined without actually solving the system model.

Parts of the Complete Response

A variety of names is used to identify terms in the response of a fixed linear system. The complete response of a system may be thought of alternatively as the sum of:

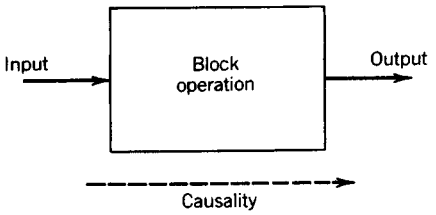


Fig. 27.11 Basic block diagram, showing assumed direction of causality or loading.

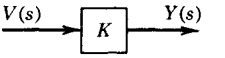
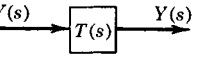
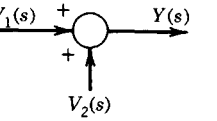
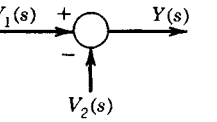
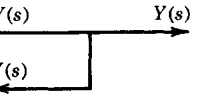
Type	Input–Output Relations: Time Domain	Input–Output Relations: Transform Domain	Symbol
(*a*) Multiplier	$y(t) = Kv(t)$	$Y(s) = KV(s)$	
(*b*) General transfer function	$y(t) = \mathcal{L}^{-1}[T(s)V(s)]$	$Y(s) = T(s)V(s)$	
(*c*) Summer	$y(t) = v_1(t) + v_2(t)$	$Y(s) = V_1(s) + V_2(s)$	
(*d*) Comparator	$y(t) = v_1(t) - v_2(t)$	$Y(s) = V_1(s) - V_2(s)$	
(*e*) Takeoff point	$y(t) = v(t)$	$Y(s) = V(s)$	

Fig. 27.12 Basic block diagram elements.[4]

1. The *free response* (or complementary or homogeneous solution) and the *forced response* (or particular solution). The free response represents the natural response of a system when inputs are removed and the system responds to some initial stored energy. The forced response of the system depends on the form of the input only.
2. The *transient response* and the *steady-state response.* The transient response is that part of the output that decays to zero as time progresses. The steady-state response is that part of the output that remains after all the transients disappear.
3. The *zero-state response* and the *zero-input response.* The zero-state response is the complete response (both free and forced responses) to the input when the initial state is zero. The zero-input response is the complete response of the system to the initial state when the input is zero.

Test Inputs or Singularity Functions

For a stable system, the response to a specific input signal will provide several measures of system performance. Since the actual inputs to a system are not usually known *a priori,* characterization of the system behavior is generally given in terms of the response to one of a standard set of *test input signals.* This approach provides a common basis for the comparison of different systems. In addition, many inputs actually encountered can be approximated by some combination of standard inputs. The most commonly used test inputs are members of the family of *singularity functions,* depicted in Fig. 27.14.

First-Order Transient Response

The standard form of the I/O equation for a first-order system is

$$\frac{dy}{dt} + \frac{1}{\tau}y(t) = \frac{1}{\tau}u(t)$$

where the parameter τ is called the system *time constant.* The response of this standard first-order system to three test inputs is depicted in Fig. 27.15, assuming zero initial conditions on the output $y(t)$. For all inputs, it is clear that the response approaches its steady state monotonically (i.e., without

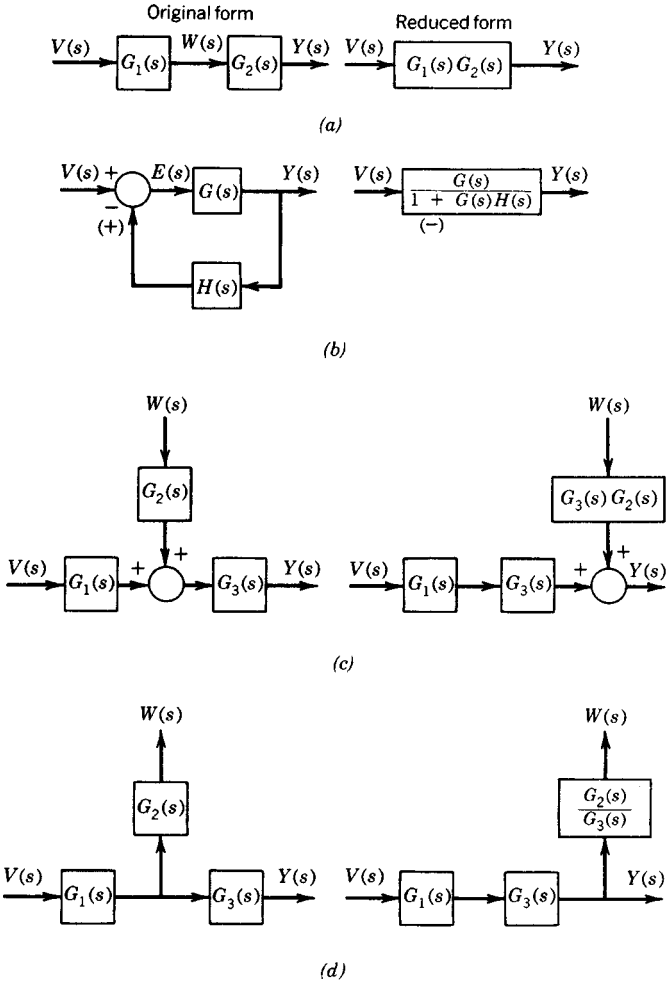


Fig. 27.13 Representative block diagram transformations: (*a*) series or cascaded elements, (*b*) feedback loop, (*c*) relocated summer, (*d*) relocated takeoff point.[4]

oscillations) and that the *speed of response* is completely characterized by the time constant τ. The transfer function of the system is

$$H(s) = \frac{Y(s)}{U(s)} = \frac{1/\tau}{s + 1/\tau}$$

and therefore $\tau = -p^{-1}$, where p is the system pole. As the absolute value of p increases, τ decreases and the response becomes faster.

The response of the standard first-order system to a step input of magnitude u for arbitrary initial condition $y(0) = y_0$ is

$$y(t) = y_{ss} - [y_{ss} - y_0]e^{-t/\tau}$$

where $y_{ss} = u$ is the steady-state response. Table 27.6 and Fig. 27.16 record the values of $y(t)$ and $\dot{y}(t)$ for $t = k\tau$, $k = 0, 1, \ldots, 6$. Note that over any time interval of duration τ, the response increases approximately 63% of the difference between the steady-state value and the value at the beginning of the time interval, that is,

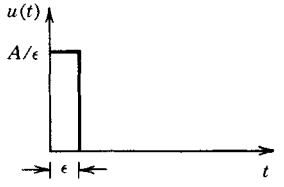
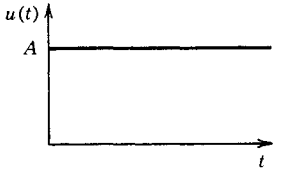
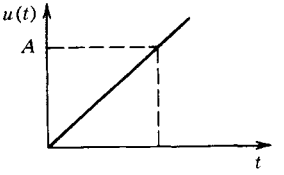
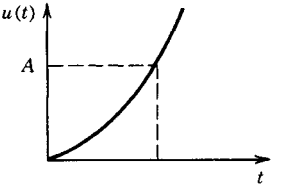
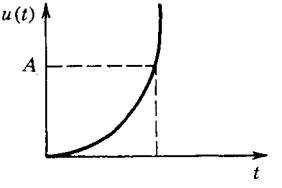
Function	Graph	$u(t)$	$U(s)$
Impulse		$u(t) = \begin{cases} \lim_{\epsilon \to 0} A/\epsilon, & 0_+ < t < \epsilon \\ 0, & \text{otherwise} \end{cases}$	$U(s) = A$
Step		$u(t) = \begin{cases} A, & 0_+ < t \\ 0, & \text{otherwise} \end{cases}$	$U(s) = \dfrac{A}{s}$
Ramp		$u(t) = \begin{cases} At, & 0_+ < t \\ 0, & \text{otherwise} \end{cases}$	$U(s) = \dfrac{A}{s^2}$
Parabolic		$u(t) = \begin{cases} At^2, & 0_+ < t \\ 0, & \text{otherwise} \end{cases}$	$U(s) = \dfrac{2A}{s^3}$
General		$u(t) = \begin{cases} At^n, & 0_+ < t \\ 0, & \text{otherwise} \end{cases}$	$U(s) = \dfrac{n!A}{s^{n+1}}$

Fig. 27.14 Family of singularity functions commonly used as test inputs.

$$y(t + \tau) - y(t) \approx 0.63212[y_{ss} - y(t)]$$

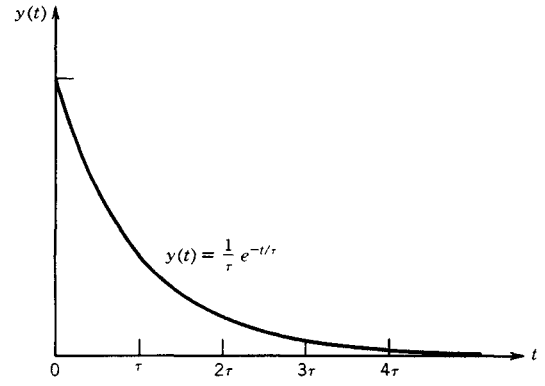
Note also that the slope of the response at the beginning of any time interval of duration τ intersects the steady-state value y_{ss} at the end of the interval, that is,

$$\frac{dy}{dt}(t) = \frac{y_{ss} - y(t)}{\tau}$$

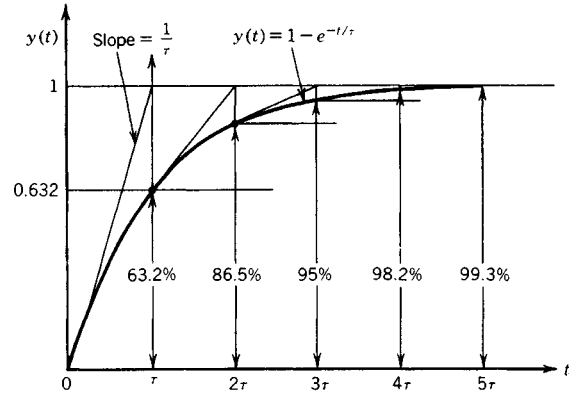
Finally, note that after an interval of four time constants, the response is within 98% of the steady-state value, that is,

$$y(4\tau) \approx 0.98168(y_{ss} - y_0)$$

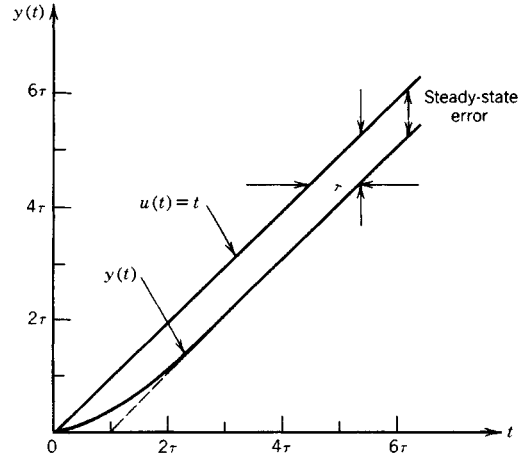
For this reason, $T_s = 4\tau$ is called the (2%) *setting time.*



(a)



(b)



(c)

Fig. 27.15 Response of a first-order system to (*a*) unit impulse, (*b*) unit step, and (*c*) unit ramp inputs.

Table 27.6 Tabulated Values of the Response of a First-Order System to a Unit Step Input

t	$y(t)$	$\dot{y}(t)$
0	0	τ^{-1}
τ	0.632	$0.368\tau^{-1}$
2τ	0.865	$0.135\tau^{-1}$
3τ	0.950	$0.050\tau^{-1}$
4τ	0.982	$0.018\tau^{-1}$
5τ	0.993	$0.007\tau^{-1}$
6τ	0.998	$0.002\tau^{-1}$

Second-Order Transient Response

The standard form of the I/O equation for a second-order system is

$$\frac{d^2y}{dt^2} + 2\zeta\omega_n \frac{dy}{dt} + \omega_n^2 y(t) = \omega_n^2 u(t)$$

with transfer function

$$H(s) = \frac{Y(s)}{U(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

The system poles are obtained by applying the quadratic formula to the characteristic equation as

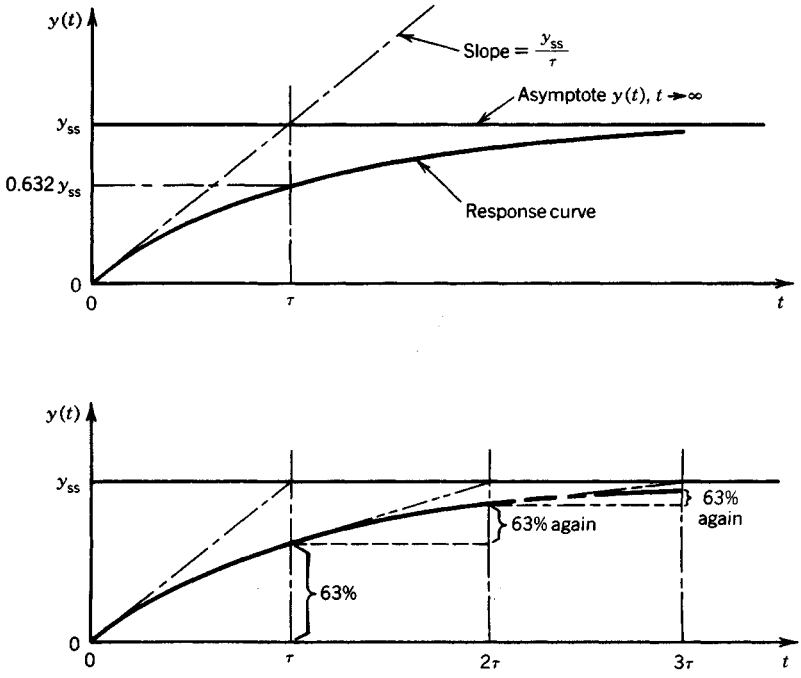


Fig. 27.16 Response of a first-order system to a unit step input, showing the relationship to the time constant.

$$p_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$$

where the following parameters are defined: ζ is the *damping ratio*, ω_n is the *natural frequency*, and $\omega_d = \omega_n\sqrt{1-\zeta^2}$ is the *damped natural frequency*.

The nature of the response of the standard second-order system to a step input depends on the value of the damping ratio, as depicted in Fig. 27.17. For a stable system, four classes of response are defined.

1. *Overdamped Response* ($\zeta > 1$). The system poles are real and distinct. The response of the second-order system can be decomposed into the response of two cascaded first-order systems, as shown in Fig. 27.18.
2. *Critically Damped Response* ($\zeta = 1$). The system poles are real and repeated. This is the limiting case of overdamped response, where the response is as fast as possible without overshoot.
3. *Underdamped Response* ($1 > \zeta > 0$). The system poles are complex conjugates. The response oscillates at the damped frequency ω_d. The magnitude of the oscillations and the speed with which the oscillations decay depend on the damping ratio ζ.
4. *Harmonic Oscillation* ($\zeta = 0$). The system poles are pure imaginary numbers. The response oscillates at the natural frequency ω_n and the oscillations are undamped (i.e., the oscillations are sustained and do not decay).

The Complex *s*-Plane

The location of the system poles (roots of the characteristic equation) in the *complex s-plane* reveals the nature of the system response to test inputs. Figure 27.19 shows the relationship between the location of the poles in the complex plane and the parameters of the standard second-order system. Figure 27.20 shows the unit impulse response of a second-order system corresponding to various pole locations in the complex plane.

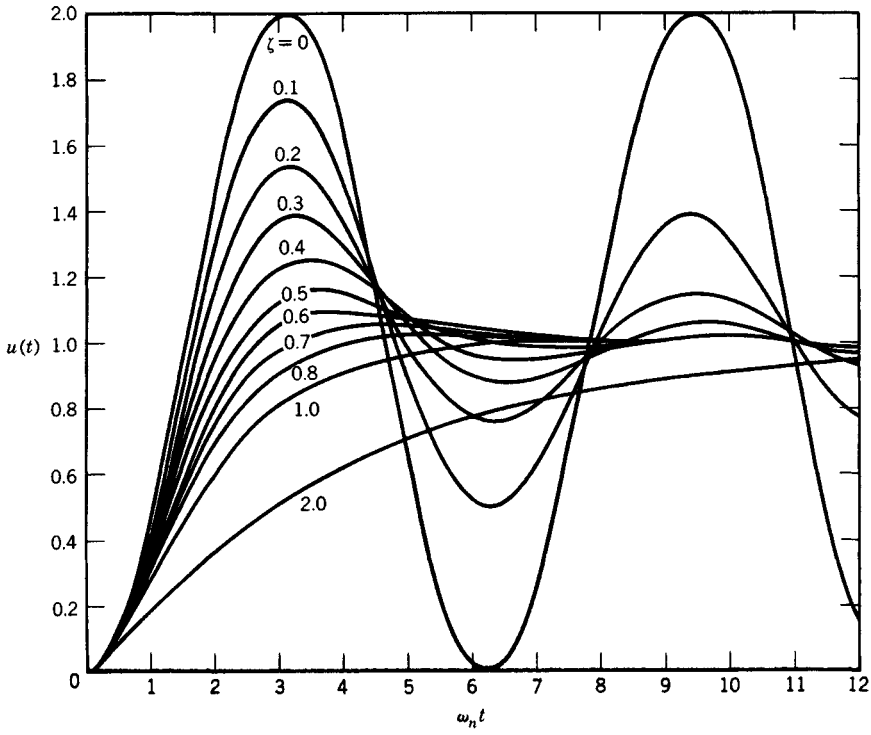


Fig. 27.17 Response of a second-order system to a unit step input for selected values of the damping ratio.

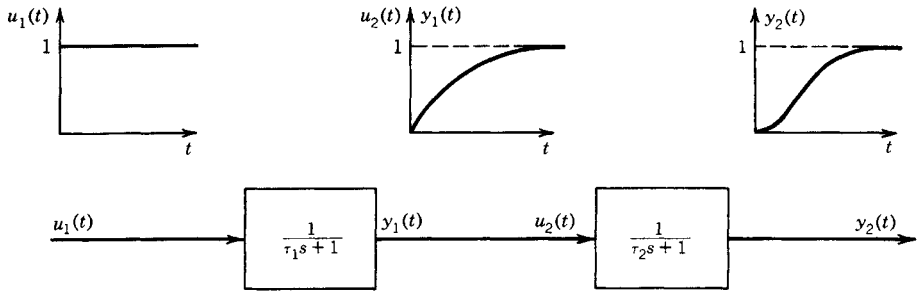


Fig. 27.18 Overdamped response of a second-order system decomposed into the responses of two first-order systems.

Transient Response of Higher-Order Systems

The response of third- and higher-order systems to test inputs is simply the sum of terms representing component first- and second-order responses. This is because the system poles must either be real, resulting in first-order terms, or complex, resulting in second-order underdamped terms. Furthermore, because the transients associated with those system poles having the largest real part decay the most slowly, these transients tend to dominate the output. The response of higher-order systems therefore tends to have the same form as the response to the *dominant poles,* with the response to the *subdominant poles* superimposed over it. Note that the larger the relative difference between the real parts of the dominant and subdominant poles, the more the output tends to resemble the dominant mode of response.

For example, consider a fixed linear third-order system. The system has three poles. The poles may either all be real, or one may be real while the other pair is complex conjugates. This leads to the three forms of step-response shown in Fig. 27.21, depending on the relative locations of the poles in the complex plane.

Transient Performance Measures

The transient response of a system is commonly described in terms of the measures defined in Table 27.7 and shown in Fig. 27.22. While these measures apply to any output, for a second-order system these can be calculated exactly in terms of the damping ratio and natural frequency, as shown in column three of the table. A common practice in control system design is to determine an initial design with dominant second-order poles that satisfy the performance specifications. Such a design can easily be calculated and then modified as necessary to achieve the desired performance.

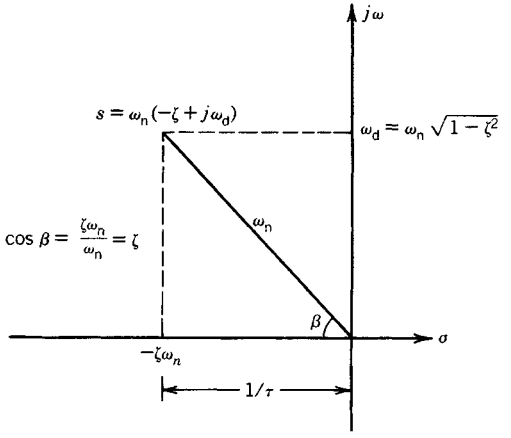


Fig. 27.19 Location of the upper complex pole in the s-plane in terms of the parameters of the standard second-order system.

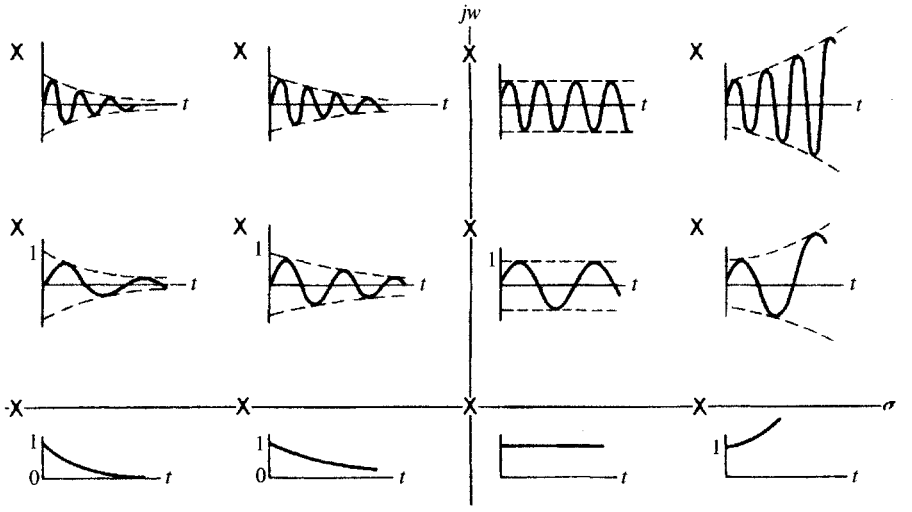


Fig. 27.20 Unit impulse response for selected upper complex pole locations in the *s*-plane.

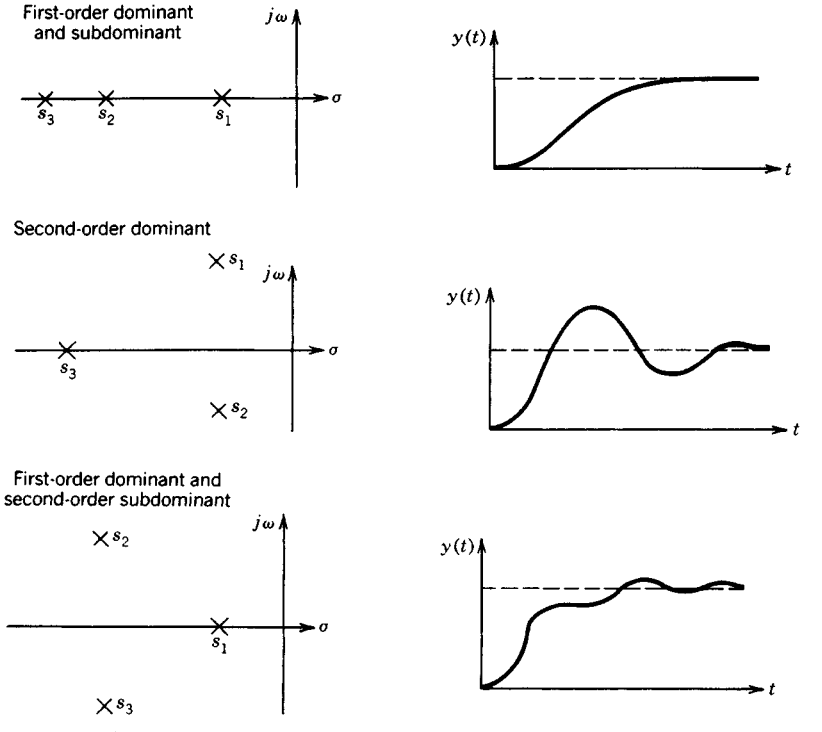


Fig. 27.21 Step response of a third-order system for alternative upper-complex pole locations in the *s*-plane.

Table 27.7 Transient Performance Measures Based Upon Step Response

Performance Measure	Definition	Formula for a Second-Order System
Delay time, t_d	Time required for the response to reach half the final value for the first time	
10–90% rise time, t_r	Time required for the response to rise from 10 to 90% of the final response (used for overdamped responses)	
0–100% rise time, t_r	Time required for the response to rise from 0 to 100% of the final response (used for underdamped responses)	$t_r = \frac{\pi - \beta}{\omega_d}$ where $\beta = \cos^{-1}\zeta$
Peak time, t_p	Time required for the response to reach the first peak of the overshoot	$t_p = \frac{\pi}{\omega_d}$
Maximum overshoot, M_p	The difference in the response between the first peak of the overshoot and the final response	$M_p = e^{-\zeta\pi/\sqrt{1-\zeta^2}}$
Percent overshoot, PO	The ratio of maximum overshoot to the final response expressed as a percentage	$PO = 100e^{-\zeta\pi/\sqrt{1-\zeta^2}}$
Setting time, t_s	The time required for the response to reach and stay within a specified band centered on the final response (usually a 2% or 5% of final response band)	$t_s = \frac{4}{\zeta\omega_n}$ (2% band) $t_s = \frac{3}{\zeta\omega_n}$ (5% band)

The Effect of Zeros on the Transient Response

Zeros arise in a system transfer function through the inclusion of one or more derivatives of $u(t)$ among the inputs to the system. By sensing the rate(s) of change of $u(t)$, the system in effect *anticipates* the future values of $u(t)$. This tends to increase the speed of response of the system relative to the input $u(t)$.

The effect of a zero is greatest on the modes of response associated with neighboring poles. For example, consider the second-order system represented by the transfer function

$$H(s) = K\frac{s - z}{(s - p_1)(s - p_2)}$$

If $z = p_1$, then the system responds as a first-order system with $\tau = -p_2^{-1}$; whereas if $z = p_2$, then the system responds as a first-order system with $\tau = -p_1^{-1}$. Such *pole-zero cancellation* can only be achieved mathematically, but it can be approximated in physical systems. Note that by diminishing the residue associated with the response mode having the larger time constant, the system responds more quickly to changes in the input, confirming our earlier observation.

27.5.3 Response to Periodic Inputs Using Transform Methods

The response of a dynamic system to periodic inputs can be a critical concern to the control engineer. An input $u(t)$ is *periodic* if $u(t + T) = u(t)$ for all time t, where T is a constant called the period. Periodic inputs are important because these are ubiquitous: rotating unbalanced machinery, reciprocating pumps and engines, ac electrical power, and a legion of noise and disturbance inputs can be approximated by periodic inputs. Sinusoids are the most important category of periodic inputs, because these are frequently occurring, easily analyzed, and form the basis for analysis of general periodic inputs.

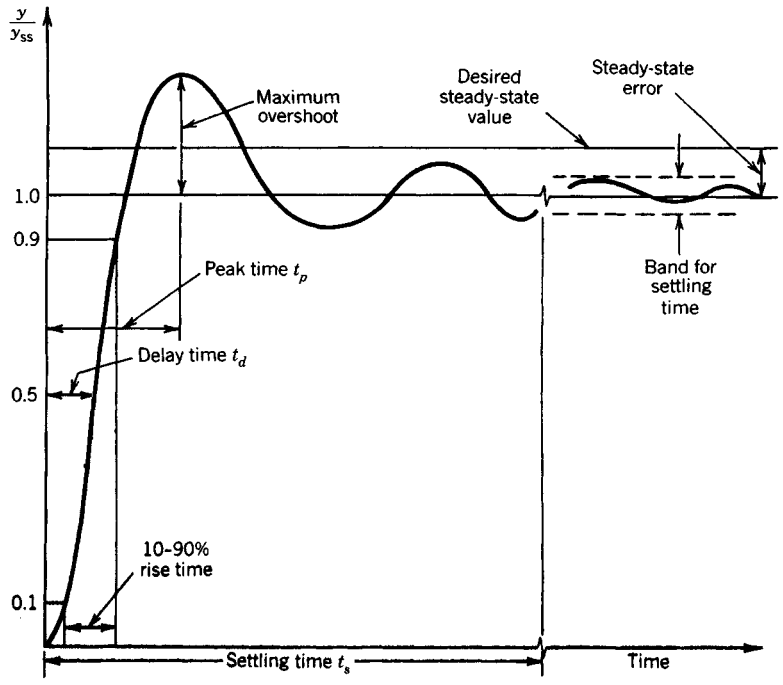


Fig. 27.22 Transient performance measures based on step response.

Frequency Response

The *frequency response* of a system is the steady-state response of the system to a sinusoidal input. For a linear system, the frequency response has the unique property that the response is a sinusoid of the same frequency as the input sinusoid, differing only in amplitude and phase. In addition, it is easy to show that the amplitude and phase of the response are functions of the input frequency, which are readily obtained from the system transfer function.

Consider a system defined by the transfer function $H(s)$. For an input

$$u(t) = A \sin \omega t$$

the corresponding steady-state output is

$$y_{ss}(t) = AM(\omega) \sin[\omega t + \phi(\omega)]$$

where $M(\omega) = |H(j\omega)|$ is called the *magnitude ratio*
$\phi(\omega) = \angle H(j\omega)$ is called the *phase angle*
$H(j\omega) = H(s)|_{s=j\omega}$ is called the *frequency transfer function*

The frequency transfer function is obtained by substituting $j\omega$ for s in the transfer function $H(s)$. If the complex quantity $H(j\omega)$ is written in terms of its real and imaginary parts as $H(j\omega) = \text{Re}(\omega) + j\text{Im}(\omega)$, then

$$M(\omega) = [\text{Re}(\omega)^2 + \text{Im}(\omega)^2]^{1/2}$$
$$\phi(\omega) = \tan^{-1}[\text{Im}(\omega)/\text{Re}(\omega)]$$

and in polar form

$$H(j\omega) = M(\omega)e^{j\phi(\omega)}$$

Frequency Response Plots

The frequency response of a fixed linear system is typically represented graphically, using one of three types of frequency response plots. A *polar plot* is simply a plot of the vector $H(j\omega)$ in the complex plane, where $\mathrm{Re}(\omega)$ is the abscissa and $\mathrm{Im}(\omega)$ is the ordinate. A *logarithmic plot* or *Bode diagram* consists of two displays: (1) the magnitude ratio in decibels $M_{db}(\omega)$ [where $M_{db}(\omega) = 20 \log M(\omega)$] versus $\log \omega$, and (2) the phase angle in degrees $\phi(\omega)$ versus $\log \omega$. Bode diagrams for normalized first- and second-order systems are given in Fig. 27.23. Bode diagrams for higher-order systems are obtained by adding these first- and second-order terms, appropriately scaled. A *Nichols diagram* can be obtained by cross plotting the Bode magnitude and phase diagrams, eliminating $\log \omega$. Polar plots and Bode and Nichols diagrams for common transfer functions are given in Table 27.8.

Frequency Response Performance Measures

Frequency response plots show that dynamic systems tend to behave like *filters*, "passing" or even amplifying certain ranges of input frequencies, while blocking or attenuating other frequency ranges. The range of frequencies for which the amplitude ratio is no less than 3 db of its maximum value is called the *bandwidth* of the system. The bandwidth is defined by upper and lower *cutoff frequencies* ω_c, or by $\omega = 0$ and an upper cutoff frequency if $M(0)$ is the maximum amplitude ratio. Although the choice of "down 3 db" used to define the cutoff frequencies is somewhat arbitrary, the bandwidth is usually taken to be a measure of the range of frequencies for which a significant portion of the input is felt in the system output. The bandwidth is also taken to be a measure of the system speed of response, since attenuation of inputs in the higher-frequency ranges generally results from the inability of the system to "follow" rapid changes in amplitude. Thus, a narrow bandwidth generally indicates a sluggish system response.

Response to General Periodic Inputs

The *Fourier series* provides a means for representing a general periodic input as the sum of a constant and terms containing sine and cosine. For this reason the *Fourier series,* together with the superposition principle for linear systems, extends the results of frequency response analysis to the general case of arbitrary periodic inputs. The Fourier series representation of a periodic function $f(t)$ with period $2T$ on the interval $t^* + 2T \geq t \geq t^*$ is

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi t}{T} + b_n \sin \frac{n\pi t}{T} \right)$$

where

$$a_n = \frac{1}{T} \int_{t^*}^{t^*+2T} f(t) \cos \frac{n\pi t}{T} \, dt$$

$$b_n = \frac{1}{T} \int_{t^*}^{t^*+2T} f(t) \sin \frac{n\pi t}{T} \, dt$$

If $f(t)$ is defined outside the specified interval by a periodic extension of period $2T$, and if $f(t)$ and its first derivative are piecewise continuous, then the series converges to $f(t)$ if t is a point of continuity, or to $\frac{1}{2} [f(t_+) + f(t_-)]$ if t is a point of discontinuity. Note that while the Fourier series in general is infinite, the notion of bandwidth can be used to reduce the number of terms required for a reasonable approximation.

27.6 STATE-VARIABLE METHODS

State-variable methods use the vector state and output equations introduced in Section 27.4 for analysis of dynamic systems directly in the time domain. These methods have several advantages over transform methods. First, state-variable methods are particularly advantageous for the study of multivariable (multiple input/multiple output) systems. Second, state-variable methods are more naturally extended for the study of linear time-varying and nonlinear systems. Finally, state-variable methods are readily adapted to computer simulation studies.

27.6.1 Solution of the State Equation

Consider the vector equation of state for a fixed linear system:

$$\dot{x}(t) = Ax(t) + Bu(t)$$

The solution to this system is