Signals, Systems, and Controls

B. P. Lathi

Signals, Systems, and Controls

B. P. Lathi
Universidade Estadual de Campinas—São Paulo

Intext Educational Publishers
New York and London

Library of Congress Cataloging in Publication Data

Lathi, Bhagwandas Pannalal.
Signal, systems, and controls.
(Intext series in circuits, systems, communications, and computers)
Includes bibliographical references.
1. System analysis. 2. Control theory.
I. Title.
QA402.L36 003 73-464
ISBN 0-7002-2431-9

Intext Educational Publishers
257 Park Avenue South
New York, New York 10010

Preface

In this book I have attempted to integrate some of the basic ideas of network and system theory, signal analysis and processing, and control systems and simulation in a fashion that will be accessible to undergraduate students taking a second course in networks or systems. The approach to time-domain and frequency-domain analysis is unified.

The two methods are presented not as different approaches, but as essentially the same approach employing different bases for input signal representation. Such a viewpoint is very effective in practical terms and has philosophical appeal. Thus, the frequency transforms (Fourier and Laplace) are introduced not as mechanical operators which aid in solving integrodifferential equations, but as tools for representing a signal as the sum of exponential signals with complex frequencies. The response of a linear system to any input signal is seen as the sum of the responses of the system to various exponential components of the input. This approach not only gives a deeper appreciation of interaction of signals with systems, but also allows one to integrate smoothly the basic concepts of signal analysis and processing with those of system analysis. It also unmasks frequency-domain analysis to reveal that it is in fact a time-domain analysis in disguise.

In the development of discrete-time systems, discrete-time signals are introduced first. The analysis of discrete-time systems then unfolds along lines similar to those in continuous-time systems, taking advantage of the parallel that exists between the two types of systems. Hybrid or sampled-data systems are then treated as special cases in which the techniques of discrete-time analysis can be applied conveniently. I believe that this approach greatly facilitates the learning of discrete-time systems as well as sampled-data systems.

The concept of the state of a system is introduced in the first chapter. Identification of the initial state with the initial conditions immediately dispels the veil of mystery surrounding the concept of the state and helps to clarify the meaning of the state to the student. The system response is then discussed in terms of zero-input and zero-state components.

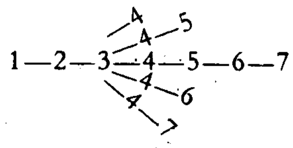
As in all my previous books I have used mathematics not so much to prove abstract axiomatic theory, as to enhance physical understanding. Logical motivation is provided for introducing new concepts. Whenever possible, theoretical results are interpreted heuristically, supported by carefully chosen examples and illustrations. Written primarily for juniors and seniors, the material included is, I believe, the absolute minimum that a prospective graduate of electrical engineering or systems engineering must acquire. The book is self-contained, requiring only a modest background in calculus and in the elements of network theory or dynamic systems. It can therefore be used effectively for self-study by practicing engineers.

Thanks are due to Professors J. B. Cruz, W. D. Thayer, M. E. Van Valkenburg, and Mr. J. M. Elfelt for several suggestions. Discussions with Professor W. H. Huggins were especially helpful. A. Alonso and D. Sousa assisted in proofreading. Thanks are also due to John Wiley & Sons for allowing me to reproduce certain material from my earlier works.

Notes to the Instructor

The entire contents of this book can be covered in about 90 classroom hours. By judicious choice of topics, it can be used for a course lasting anywhere from 30 to 90 hours. It is therefore suitable as a text for a course lasting for one to three quarters or one to two semesters.

Linear Systems: For a course on signals and systems (or systems theory), any one of the following combinations of chapters should be appropriate.



Chapter 4 gives two interpretations of the frequency domain: (1) the time-domain interpretation and (2) the conventional or transform interpretation. The instructor may omit the former interpretation (Secs. 4.1–4.8) without experiencing much discontinuity in the flow of the discussion. This, however, is not recommended unless there is no way of covering the desired material in a given time period.

Control Systems: The book can be used effectively as a text for a 30- to 45-hour course on control systems. For this purpose the following material is suggested. Chapters 1 and 2, part of Chapter 4 (Secs. 9–13, and 15–17 only), Chapter 5, and Appendixes A–F.

Contents

Chapter 4: Frequency-Domain Analysis Using Generalized Exponentials

Chapter 5: Feedback and Control

1

Introduction to Systems

1.1 INTRODUCTION

The dictionary gives several possible meanings of the word *system*. One of them is "A set of arrangement of things so related or connected as to form a unity or organic whole, as a solar system, irrigation system, or supply system." This rather broad definition includes all physical as well as nonphysical systems. Electrical, mechanical, electromechanical, hydraulic, acoustic, and thermal systems are examples of physical systems. Economic systems, political systems, traffic control systems, and industrial planning systems are examples of socioeconomic systems. The socioeconomic systems are made up of various components and their interactions. These are no more or less of systems than physical systems. As a result, the system theory applicable to physical and a class of nonphysical systems is emerging. Attempts are being made to apply system theory to such socioeconomic problems as education, transportation, public and private administration, and economic development. In this book we shall be concerned with physical systems in general, and with electrical, mechanical, and electromechanical systems in particular.

Each system performs some desired function (response or outputs) for a given set of driving functions (inputs). In electrical systems the driving functions are generally in the form of voltage and current sources, and the response will be voltages or currents at certain locations. For mechanical systems the inputs may be forces (or displacements), and the response may be displacement or velocity at some point. For a given system there may be several inputs (driving functions) acting simultaneously, and there may be several outputs (responses) of interest. Before defining any relationship between inputs (driving functions) and outputs (responses), let us consider a simple mechanical system consisting of a mass M which is acted upon by a force f as shown in Fig. 1.1.

In this case the driving function (input) is $f(t)$ and the output (response) is the velocity v. According to Newton's law, for a constant mass M the force f and the velocity v of the mass are related by

$$f = M\frac{dv}{dt}$$

or

$$v = \frac{1}{M}\int f(\tau)\,d\tau \qquad (1.1a)$$

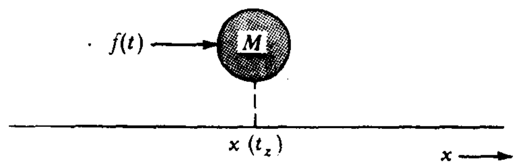


Figure 1.1

The velocity v is the response to the input force f. The velocity at any instant t is therefore the result of force f acting on M in the entire past. Hence the limits of integration in Eq. 1.1a are from $-\infty$ to t. Thus

$$v(t) = \frac{1}{M} \int_{-\infty}^{t} f(\tau)\, d\tau \tag{1.1b}$$

$$= \frac{1}{M} \int_{-\infty}^{0} f(\tau)\, d\tau + \frac{1}{M} \int_{0}^{t} f(\tau)\, d\tau \tag{1.1c}$$

From Eq. 1.1b (by letting $t = 0$) the first term on the left-hand side of Eq. 1.1c is seen to be $v(0)$. Hence

$$v(t) = v(0) + \frac{1}{M} \int_{0}^{t} f(\tau)\, d\tau \tag{1.1d}$$

Let us discuss the implications of Eq. 1.1d. From Eq. 1.1b, it is obvious that the velocity $v(t)$ of the mass at any instant t can be computed if we know the force that acted upon the mass in the entire past $(-\infty, t)$. In practice, however, it is impossible to keep record of force acting on a mass over the entire history of its existence. In such case the use of Eq. 1.1d proves very attractive. Suppose we know the force from some moment $t = 0$ onward we can still calculate $v(t)$ for $t \geq 0$ provided $v(0)$, the initial velocity (velocity at $t = 0$) was known. Thus $v(0)$ has all the relevant information about the entire past of the forces acting on M, that we need to calculate $v(t)$ for $t \geq 0$. The velocity $v(0)$ represents the value of the velocity at the initial moment $t = 0$ and is generally referred to as the *initial condition*. In the present case we arbitrarily chose the initial moment to be $t = 0$. We can, however, use $t = t_0$ as the initial instant.

Eq. 1.1d can be easily generalized as

$$v(t) = v(t_0) + \frac{1}{M} \int_{t_0}^{t} f(\tau)\, d\tau \tag{1.1e}$$

In the present situation we conclude that the response (velocity) for $t \geq t_0$ is a function of the initial condition $v(t_0)$ and the input $f(t)$ for $t \geq t_0$. This fact may be expressed as

$$v(t) = \phi[v(t_0), f(t)], \qquad t \geq t_0 \tag{1.2}$$

This result is true in general. A response of a system for $t \geq t_0$ is a function of the initial conditions at $t = t_0$ and the input(s) $f(t)$ for $t \geq t_0$.

In the present problem we needed only one initial condition. However, in general several initial conditions may be necessary. Consider again our problem of mass M acted upon by a force f. Let us determine the position x of the mass at some time t. We have

$$\frac{dx}{dt} = v$$

Hence

$$x(t) = \int_{-\infty}^{t} v(\tau)\, d\tau \tag{1.3a}$$

$$= \int_{-\infty}^{0} v(\tau)\, d\tau + \int_{0}^{t} v(\tau)\, d\tau \tag{1.3b}$$

From Eq. 1.3a (by letting $t = 0$), the first term on the right-hand side of Eq. 1.3b is seen to be $x(0)$. Hence

$$x(t) = x(0) + \int_{0}^{t} v(\zeta)\, d\zeta \tag{1.3c}$$

where ζ is the dummy variable of integration. Substituting Eq. 1.1d in Eq. 1.3c we obtain

$$x(t) = x(0) + \int_{0}^{t} \left[v(0) + \frac{1}{M} \int_{0}^{\zeta} f(\tau)\, d\tau \right] d\zeta$$

$$= x(0) + v(0)t + \frac{1}{M} \int_{0}^{t} \int_{0}^{\zeta} f(\tau)\, d\tau\, d\zeta \tag{1.3d}$$

It is obvious from Eq. 1.3d, that if the input force $f(t)$ is known from $t = 0$ onward, then to find the position $x(t)$ for $t \geq 0$, we need two initial conditions $x(0)$ and $v(0)$. Thus

$$x(t) = \phi[x(0), v(0), f(t)], \qquad t \geq 0$$

Note that the initial instant is arbitrarily chosen at $t = 0$. The results can be generalized for any value of the initial instant.

1.2 STATE OF A SYSTEM: THE VITAL KEY

The initial conditions at some $t = t_0$ collectively are called as the **state** of the system at $t = t_0$. Thus if a system has n initial conditions $x_1(t_0)$, $x_2(t_0)$, ..., $x_n(t_0)$, the state of the system at $t = t_0$ **(initial state)** is given by $x_1(t_0)$, $x_2(t_0)$, ..., $x_n(t_0)$. We may say that the state at some instant t_0 contains all the relevant information of the past history of the system that is needed to obtain the response for $t \geq t_0$ when the input is given for $t \geq t_0$. For the mass-force system in Fig. 1.1, the state of the system at $t = t_0$ is given by $x(t_0)$ and $v(t_0)$.

The state of a system at any time t_0 is the smallest set of numbers $x_1(t_0)$, $x_2(t_0)$, ..., $x_n(t_0)$ which is sufficient to determine the behavior of the system for all time $t \geq t_0$ when the input to the system is known for $t \geq t_0$.

In general there may be several inputs applied simultaneously at various points in a system and there may be several variables of interest which will be considered as response (output). For simplicity, we shall first consider the case of single-input, single-output system and then later extend the discussion to a general case of multiple-input, multiple-output system. A response $y(t)$ for $t \geq t_0$ of a system is a function of the state of the system at some initial instant $t = t_0$, and the input $f(t)$ for $t \geq t_0$. This can be expressed as

$$y(t) = \phi[x_1(t_0), x_2(t_0), \ldots, x_n(t_0), f(t)] \qquad t \geq t_0 \tag{1.4a}$$

For the sake of convenience the initial state at $t = t_0$ represented by numbers $x_1(t_0), x_2(t_0), \ldots, x_n(t_0)$, will be denoted by $\{x(t_0)\}$. Using this notation, Eq. 1.4a can be expressed as

$$y(t) = \phi[\{x(t_0)\}, f(t)], \qquad t \geq t_0 \tag{1.4b}$$

Figure 1.2 shows the block diagram representation of a system. A system is

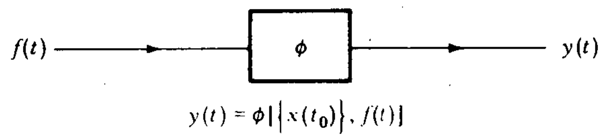


Figure 1.2

characterized by input(s), output(s) and the functional block diagram. The functional block diagram should be labeled by the input-output relationship (such as in Eq. 1.4b) for complete characterization of the system.

At this point we make an important observation. It was seen that the response $y(t)$ at any instant $t \geq t_0$ can be determined from the knowledge of initial state $\{x(t_0)\}$ and the input $f(t)$ over the interval (t_0, t). Let us consider the output y at $t = t_0$. From the above discussion it is evident that $y(t_0)$ can be determined from the knowledge of the initial state $\{x(t_0)\}$ and the input $f(t)$ over the interval (t_0, t_0). The latter is $f(t_0)$. Hence the response at any instant is determined completely from the knowledge of the state of the system at that instant (and the input at that instant). This result is also true for multiple-input, multiple-output systems. Every output (response) at any given instant t is completely determined by the state of the system (and the inputs) at that instant. Therefore the state of a system at some instant tells us everything about the system at that instant. It is evident that the state of a system is the single most important attribute of a system. It is the vital key to the system.

As an example, consider the electrical network shown in Fig. 1.3. We can easily show that if the capacitor voltage x_1 and the inductor current x_2 are known (along with the input) at any instant, then all voltage and currents in this network at that instant are known. From Fig. 1.3 it can be seen that

$$v_{R_1} = f(t) - x_1$$

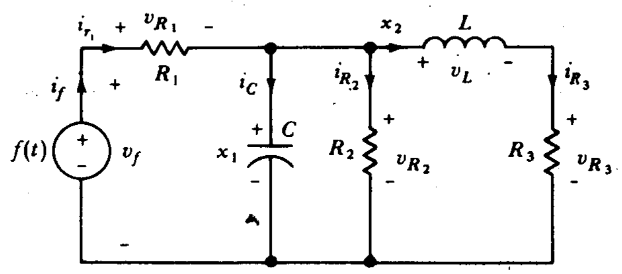


Figure 1.3

$$
\begin{aligned}
i_{R_1} &= \frac{1}{R_1} v_{R_1} = \frac{1}{R_1}[f(t) - x_1] \\
v_{R_2} &= x_1 \\
i_{R_2} &= \frac{1}{R_2} x_1 \\
i_C &= i_{R_1} - i_{R_2} - x_2 = \frac{1}{R_1}[f(t) - x_1] - \frac{1}{R_2} x_1 - x_2 \\
v_C &= x_1 \\
i_{R_3} &= x_2 \\
v_{R_3} &= R_3 x_2 \\
v_L &= v_{R_2} - v_{R_3} = x_1 - R_3 x_2 \\
i_L &= x_2 \\
i_f &= i_{R_1} = \frac{1}{R_1}[f(t) - x_1] \\
v_f &= f(t)
\end{aligned}
\tag{1.5}
$$

It can be seen that if we know the capacitor voltage x_1, the inductor current x_2, and the input $f(t)$ at any instant, all the voltages and the currents in the network at that instant are determined. Consequently the state of this network at any instant t_0 is given by $x_1(t_0)$, $x_2(t_0)$. For electrical networks, in general, it can be shown that all voltages, and the currents at any instant are determined by the values of all inductor currents and all capacitor voltages (and the inputs) at that instant. Hence in electrical network the state of a system at any instant is given by all inductor currents and all capacitor voltages. It will also be seen that in mechanical systems all forces, displacements, and velocities at any instant are determined from the knowledge of positions, and velocities of all junctions† (along with inputs) at that instant. Hence the positions and velocities of all junctions represent a state of a mechanical system.

The concept of a state is very important. As the term implies, state of a system represents its position or status.

†A junction is the point where two or more elements are connected.

STATE OF A SYSTEM IS NOT UNIQUELY SPECIFIED

At this point we shall note that the state of a system can be specified in several ways. In the force-mass system (Fig. 1.1), for example, we may specify the state of a system by $x(t_0)$, $v(t_0)$, or we may define new variables w_1 and w_2 as

$$\begin{aligned} w_1 &= a_{11}x + a_{12}v \\ w_2 &= a_{21}x + a_{22}v \end{aligned} \tag{1.6}$$

Solving these equations simultaneously, we can express x and v in terms of† w_1 and w_2. Therefore if w_1, and w_2 are known x and v are also known. Hence the response of the system can also be obtained from the knowledge of the input $f(t)$ and the initial conditions $w_1(t_0)$ and $w_2(t_0)$. Therefore, by definition $w_1(t_0)$, $w_2(t_0)$ also specify the state of the system.

In the electrical network shown in Fig. 1.3, one can easily show that the voltages and the currents in all the branches at any instant are determined from the knowledge of v_{R_1} and v_{R_3} (and the input) at that instant. Hence v_{R_1} and v_{R_3} also specify the state of the system. Other possible sets of variables which can specify the state of this network are (i_L, v_L), (i_C, v_C), (v_{R_1}, v_L), (i_C, v_L), and (i_C, v_{R_3}). It is left as an exercise for the reader to show that these sets can specify the state of the system in Fig. 1.3.

1.3 CLASSIFICATION OF SYSTEMS

The systems can be broadly classified into following categories:‡

1. Linear and nonlinear systems.
2. Constant-parameter and time-varying-parameter systems.
3. Instantaneous and dynamic systems.
4. Lumped-parameter and distributed-parameter systems.
5. Continuous-time and discrete-time systems.

We shall now discuss the nature of these classifications. To begin with, we shall consider the single-input, single-output system and then extend the results to a general case.

1. LINEAR AND NONLINEAR SYSTEMS

Before defining a linear system it is necessary to understand the important concept of linearity.

Linearity Concept

The reader is no doubt familiar with the rudimentary notions of linearity. Broadly speaking, the linearity property implies two important concepts (i) *homo-*

†This is true provided $a_{11}a_{22} - a_{12}a_{21} \neq 0$. This condition is implicit here.

‡There are few more classifications such as (1) discrete-state and continuous-state systems, and (2) deterministic and probabilistic systems. These classes, however, are beyond the scope of this book and will not be considered.

geneity and (ii) *superposition*. Homogeneity property implies that a k-fold increase in the input causes a k-fold increase in the output for any value of k (Fig. 1.4a). If $f(t)$ is the input and $y(t)$ is the corresponding response, then $ky(t)$ is the response when the input is $kf(t)$. This fact may be represented as

If

$$f(t) \rightarrow y(t)$$

then

$$kf(t) \rightarrow ky(t) \tag{1.7}$$

The superposition property states that if there are several inputs (or causes) acting on a system, then the total response of the system due to all the inputs (causes) can be determined on installment basis by considering only one input (cause) at a time and by assuming the remaining inputs (causes) to be zero. The total output is the sum of all such output components computed by considering one input (cause) at a time. This property may be expressed as follows. If $y_1(t)$ is the response of a system to the input (cause) $f_1(t)$ and $y_2(t)$ is the response of the same system to the input (cause) $f_2(t)$, then if both inputs are acting simultaneously—that is, when the input is $f_1(t) + f_2(t)$—the response is $y_1(t) + y_2(t)$. This property may be expressed as

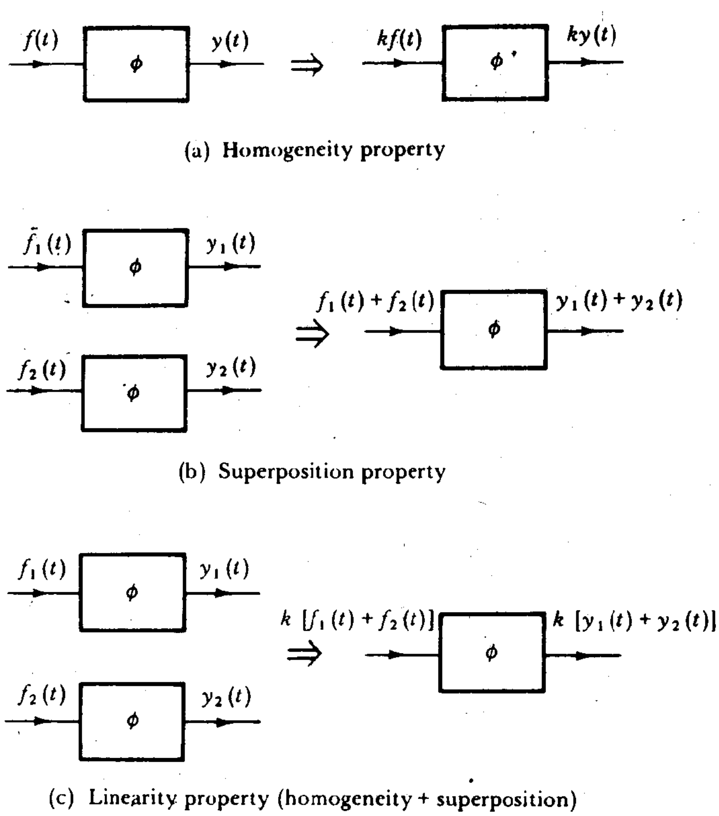


Figure 1.4

follows (see Fig. 1.4b):

If

$$f_1(t) \rightarrow y_1(t)$$
$$f_2(t) \rightarrow y_2(t)$$

then

$$f_1(t) + f_2(t) \rightarrow y_1(t) + y_2(t) \tag{1.8}$$

We can combine both the properties (Eqs. 1.7 and 1.8) into one equation as follows (see Fig. 1.4c):

If

$$f_1(t) \rightarrow y_1(t)$$
$$f_2(t) \rightarrow y_2(t)$$

then

$$k[f_1(t) + f_2(t)] \rightarrow k[y_1(t) + y_2(t)] \tag{1.9}$$

Note that Eq. 1.9 embodies the essence of both equations 1.7 and 1.8. Hence Eq. 1.9 represents the *linearity property* (homogeneity + superposition).

Definition of Linear and Nonlinear Systems

We shall now use the linearity concepts to define a linear system. As observed earlier, the output of a system depends upon not only the input $f(t)$ but also the initial state $\{x(t_0)\}$. We may view this as if the output (response) depends upon two different inputs or causes; $\{x(t_0)\}$ and $f(t)$. Consequently for a linear system we must demand that the output (response) should be given by a sum of two components, arising because of the two different causes. The component due to each cause is computed by assuming that only that cause is present and the other cause is zero. To be specific, the output of a linear system should be given by a sum of two components (i) the output of the system with the given initial state $\{x(t_0)\}$ and with zero input—that is, $f(t) = 0$ (*zero-input component*); and (ii) the output of the system with the given input $f(t)$ but with zero initial state—that is, $\{x(t_0)\} = 0$ (*zero-state component*). Thus the response $y(t)$ can be expressed as

$$\underbrace{y(t)}_{\text{total response}} = \underbrace{y_x(t)}_{\text{zero-input response}} + \underbrace{y_f(t)}_{\text{zero-state response}} \tag{1.10}$$

where $y_x(t)$ (the zero-input response) is a function of the initial state only and $y_f(t)$ is a function of the input $f(t)$ only. This property of a system which allows us to separate the components due to the initial state and the input is called the *decomposition property*. Thus the output of a linear system can be separated into two components. The first component (zero-input component) is obtained by letting the input be zero. This component of response is caused entirely by the initial conditions or the initial state. The second component (zero-state component) is obtained by letting the initial state be zero. This component of the response results due to the input alone. The decomposition property allows us to evaluate the two