O. L. R. JACOBS

Introduction to control theory

CLARENDON PRESS · OXFORD

1974

Oxford University Press, Ely House, London W.1
GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
CAPE TOWN IBADAN NAIROBI DAR ES SALAAM LUSAKA ADDIS ABABA
DELHI BOMBAY CALCUTTA MADRAS KARACHI LAHORE DACCA
KUALA LUMPUR SINGAPORE HONG KONG TOKYO

ISBN 0 19 856108 3

PRINTED IN GREAT BRITAIN BY
WILLIAM CLOWES & SONS, LIMITED
LONDON, BECCLES AND COLCHESTER

Preface

CONTROL theory is a branch of applied mathematics devoted to analysis and design of control systems. Control systems are systems in which a controller interacts with a real process in order to influence its behaviour. A primary objective for most control systems is to make some real variable take a desired value, for example to regulate the temperature of an oven or to make the direction of a receiving aerial track a moving target. The objective is usually to be achieved by adjusting some other variable, such as heat input to the oven or force applied to the aerial, although the response to such adjustments in most real controlled processes is neither instantaneous nor certain. The non-instantaneous response is accounted for by regarding the controlling and controlled variables as input and output of a *dynamic* system described by differential or difference equations. The effect of uncertainties is reduced by using *feedback* to provide the controller with continuous indication of what adjustment is needed; for example, if the oven is too cold more heat must be supplied and if the aerial points to the left of its target it must be forced to turn to the right. Feedback is a characteristic feature of control systems: in addition to reducing the effects of uncertainties it modifies the dynamic behaviour of controlled processes and can cause instability. The main preoccupations of control theory are with analysis of system dynamics and with applications of probability theory to describe the behaviour of dynamic systems in the presence of uncertainty. Control theory is similar to other branches of applied mathematics in that the majority of solved theoretical problems are linear and the majority of real control systems non-linear. Techniques for applying the theoretical results to specific practical problems are beyond the scope of the theory.

Control theory originally developed as a branch of engineering science, but subsequently found applications elsewhere; for example in economics, in studies of social systems, and in biology.† Its development has passed through three stages.

† Most of the books about control in the Bibliography have an introductory chapter on engineering applications of control theory. Applications to economic systems are described in Tustin (1953), Allen (1968), IFAC IFORS (1973); to social systems in Forrester (1961, 1968, 1971) and to biological systems in Bayliss (1966), Milsum (1966), Milhorn (1966), Stark (1968), McFarland (1971).

(i) A classical stage originating with a study by Maxwell (1868) of speed-control systems and culminating during the second world war in designs of gunnery, radar, and other military control systems. In this stage classical analysis of ordinary linear differential equations was interpreted for and applied to control systems.

(ii) A more modern stage during the 1950s and 1960s when the attention of applied mathematicians was directed to aerospace and to complex industrial problems. In this stage multi-variable optimization methods were developed and applied under the influence of the contemporary development of digital computers.

(iii) The most recent stage which emphasizes the importance of uncertainty, almost completely neglected at earlier stages. In this stage control systems are regarded as stochastic systems and probability theory is applied.

The history of control theory is reflected in the existing range of books on the subject. There are many books about the classical theory of stage (i), often written by engineers for engineers and with emphasis on engineering applications. There are books about the more modern theory of stage (ii), often written with emphasis on mathematical rigour and on computational feasibility. There are a few books about the recent stochastic control theory of stage (iii), mainly restricted to linear systems. Books covering all three stages are still rare.

The purpose of this book is to provide an introduction to the main results from all three stages of control theory and to unify them in a single volume at a mathematical level suitable for final year undergraduates in engineering and for graduates.† The material is presented without much reference to specific applications and the book is not intended only for engineers. It is suitable for anyone whose mathematical background includes the classical solution of ordinary linear differential or difference equations and the elements of probability theory, which are reviewed in Chapters 1 and 12 respectively, complex numbers and matrix algebra, which are used without any special explanation, and a certain confidence in manipulating mathematical expressions.

The book is divided into three parts on a mathematical, rather than historical, basis.

Part I Deterministic linear systems described by equations having analytic solutions.

Part II Deterministic non-linear systems described by equations without analytic solutions.

† The book is in some of these respects similar to certain other recent books: Takahashi, Rabins, and Auslander (1970) covers a similar wide range of topics from all three stages but with less emphasis on theory; Anderson and Moore (1971) is written at a similar mathematical level and discusses modern control theory from stage (ii); Kwakernaak and Sivan (1972) unifies the theory of linear systems from all three stages.

Part III Systems with uncertainty, described with the help of probability theory.

This framework leads to several unconventional features in presentation.

(i) The uniform treatment of both continuous-time and discrete-time systems in the chapters of Part I about classical control theory. For example the z-transform precedes the Laplace transform in the presentation of Section 1.4 because the z-transform is mathematically simpler.

(ii) Chapter 9 on optimal control theory precedes the material on phase-plane analysis in Chapter 10 in order that it shall immediately follow the presentation of optimal linear control theory in Chapter 8 at the end of Part I.

(iii) Chapter 12 is an introduction to probability theory rather than to control theory. It is included as necessary background for Part III in the same way that Chapter I provides necessary background on ordinary dynamic equations for Part I.

Other features of the book are as follows.

(iv) Dynamic programming and Bayes's rule are emphasized as foundations for optimal and stochastic control theory.

(v) Chapter 13 on stochastic control theory precedes and motivates the introduction to estimation theory of Chapter 14. With this order of presentation the Kalman filter equations are used in Chapter 13 in advance of their derivation in Chapter 14.

(vi) The discussion in Chapter 15 of the structure of stochastic controllers has not previously appeared outside the research literature.

The book derives largely from lectures given to undergraduates in engineering science at the University of Edinburgh, the University of Oxford, and the University of California, San Diego, and should be suitable for similar course work elsewhere. Problems and their answers are provided for every chapter except Chapter 15 which presents research material. Selected chapters might be used for various courses, for example:

(i) *Classical control theory*
(Chapter 1 contains necessary background material.)
Chapters 2, 3, 4, 5.
(Chapter 10 and Section 11.3 optional.)
(ii) *Optimal control theory*
(Chapter 6 provides a link with classical control theory.)
Chapters 7, 8, 9; or Chapters 7, 8; or Chapters 8, 9; or Chapter 8.
(iii) *Non-linear control theory*
(Classical control theory may be a necessary prerequisite.)
Chapters 10, 11.

(iv) *Stochastic control theory*
(Chapter 12 contains necessary background material, Chapter 8 is a necessary prerequisite.)
Chapters 13, 14.
(Chapter 15 optional.)

Each chapter ends with a bibliography indicating further or alternative reading; the references are mainly to other books rather than to primary sources and are inevitably incomplete.

The book also is inevitably incomplete. The theory of observers mentioned at the end of Section 7.4, Popov's method mentioned in Section 11.2, and recent developments in the theory of linear multi-variable systems† might well have been included. Power spectral density functions were not used in Section 12.6, in spite of their intuitive appeal, because they are not needed for the derivation of any results. Practical applications, including computer realizations of control algorithms, were deliberately excluded in the belief that insight into the main results of control theory can be achieved by concentrating on the theory.

It is a pleasure to acknowledge the debts I have incurred in writing this book. The debt to other authors is evident from the Bibliography. The students who have attended my various courses and questioned what I told them may not know how much I learned from them. David Clarke, David Hughes, and David Witt read drafts of the book and made many helpful suggestions. Most of all I am grateful to my wife Sheila who encouraged me to persevere and who, together with our children, had to live with me while I did so.

† See MacFarlane (1973).

Contents

Part I
Deterministic linear systems

1. Ordinary linear systems

1.1. Dynamic equations

MOST control systems are dynamic systems characterized by variables that are functions of time. Such variables may be continuous-time variables, like the temperature at some particular place, or discrete-time variables, like a series of mid-day observations of temperature at the same place; Fig. 1.1 shows the difference between the two sorts of variable.

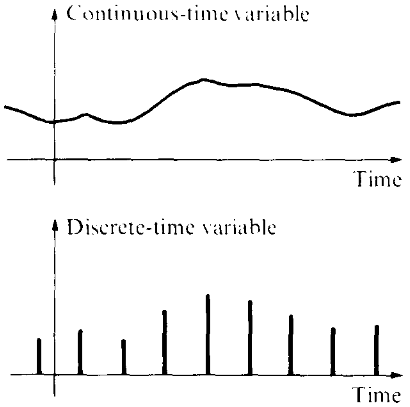

FIG. 1.1. Two sorts of function of time.

Dynamic systems are described for purposes of mathematical analysis by dynamic equations; continuous-time systems by differential equations and discrete-time systems by difference equations. A linear system is one that can be described by a linear dynamic equation. If a linear system is single-variabled, like that shown in Fig. 1.2, its dynamic equation would be either

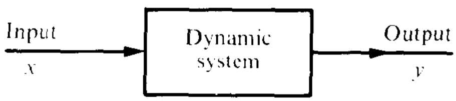

FIG. 1.2. Single-variabled dynamic system.

the continuous-time linear differential equation

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_1 \frac{dy}{dt} + a_0 y = b_m \frac{d^m x}{dt^m} + \cdots + b_0 x, \qquad (1.1)$$

where t represents time; or the discrete-time linear difference equation

$$a_n y(i + n) + \cdots + a_1 y(i + 1) + a_0 y(i) = b_m x(i + m) + \cdots + b_0 x(i), \quad (1.2)$$

where i is an integer counting the discrete-time instants. When these equations describe real physical systems the coefficients $a_n, \ldots, a_0, b_m, \ldots, b_0$ are all real and n, the order of the equation, is greater than or equal to m, the order of the forcing function. This condition $n \geqslant m$ reflects the fact that the system cannot respond to an input before the input has been applied. In addition the coefficients $a_n, \ldots, a_0, b_m, \ldots, b_0$ are often constant although the equations remain linear even when the coefficients vary with time.

Linear equations satisfy the principle of superposition, which states that

if input x_1 causes output y_1 and input x_2 causes output y_2,
then input $x_1 + x_2$ causes output $y_1 + y_2$.

Figure 1.3 illustrates the principle for a continuous-time equation.

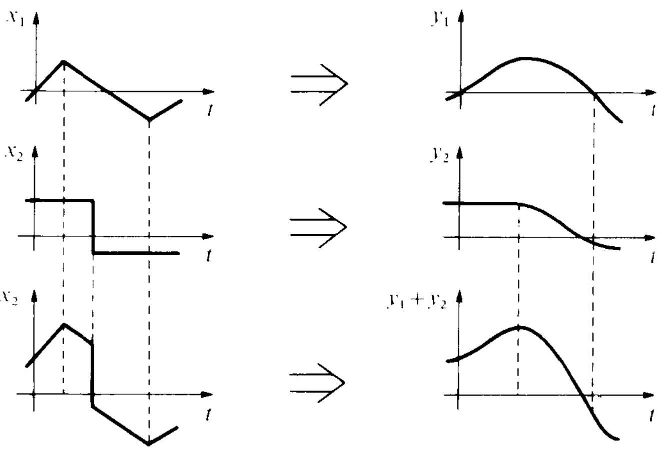

FIG. 1.3. Superposition.

It follows from the principle of superposition that the complete solution to a linear dynamic equation is the sum of two parts,

complete solution = particular integral + complementary function.

The particular integral, or 'driven response', is due to the input x, and the complementary function, or 'natural response', follows from the initial condition of the equation. In order to find the complete solution it is necessary to know the form of the input function x and to know n initial conditions.

The complementary function is found by setting the right-hand side of the equation to zero and seeking solutions to the resulting homogeneous equation in the general form

$$y_{CF} = \sum_{k=1}^{n} A_k \, e^{s_k t}$$

for continuous-time equations, or

$$y_{CF} = \sum_{k=1}^{n} A_k s_k^{\,i}$$

for discrete-time equations; in both cases the A_k are constants determined by the initial conditions and the s_k, which may be complex numbers, are the roots of the characteristic equation

$$a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0 = 0. \tag{1.3}$$

These roots of the characteristic equation determine the dynamic behaviour of the equation.

The most important aspect of the dynamic behaviour is the stability of the equation. A linear dynamic system is said to be 'stable' if its response to any input tends to a finite steady value after the input is removed; this implies that the complementary function must remain finite as time goes to infinity. It follows that the condition for stability of a continuous-time system is that the roots of the characteristic equation must all have negative real parts and that the condition for stability of a discrete-time system is that the roots of the characteristic equation must all have magnitude less than unity. Figure 1.4

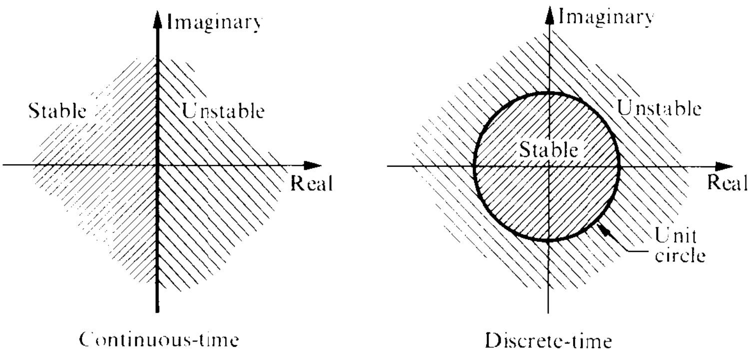

FIG. 1.4. Stability conditions on roots of the characteristic equation.

shows how these conditions can be interpreted on an Argand diagram, sometimes described as the 's-plane', as regions where the roots must, or must not, be for stability, or instability.

Stability is discussed more fully in Chapter 3. Chapter 1 is devoted to standard descriptions of the responses of ordinary linear dynamic equations.

Problem 1.1

1.2. Continuous-time responses

The typical, first-order differential equation

$$T\frac{dy}{dt} + y = x \tag{1.4}$$

has characteristic equation

$$Ts + 1 = 0$$

and complementary function

$$y_{CF} = A\,e^{-t/T}.$$

If the input x has the constant value $x = X$ the particular integral is

$$y_{PI} = X$$

and if the initial condition is $y(t = 0) = 0$, the general solution is

$$y = X(1 - e^{-t/T}). \tag{1.5}$$

This solution is sketched in Fig. 1.5 which shows how the constant T determines the time scale of the response. T is called the 'time constant' of the first-order equation. The time constant of an unknown first-order system can

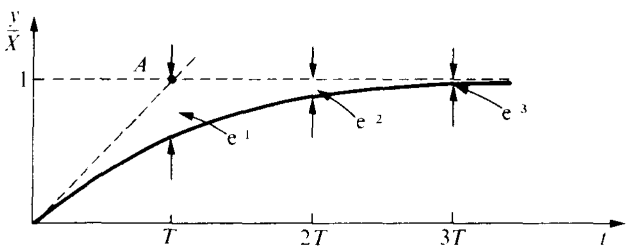

FIG. 1.5. Step response of first-order system.

easily be found from an experimental step response by noticing that the tangent to the initial response intersects the final value (point A in Fig. 1.5) after time T.

The typical, second-order differential equation is written

$$\frac{d^2y}{dt^2} + 2\zeta\omega_0\frac{dy}{dt} + \omega_0^2 y = \omega_0^2 x. \tag{1.6}$$

It has complementary function

$$y_{CF} = A_1\,e^{-\omega_0(\zeta+\sqrt{(\zeta^2-1)})t} + A_2\,e^{-\omega_0(\zeta-\sqrt{(\zeta^2-1)})t},$$

and when the input has the constant value $x = X$ and the initial conditions are $y(t = 0) = \mathrm{d}y/\mathrm{d}t\,(t = 0) = 0$ the general solution is as sketched in Fig. 1.6, which shows how the constant ω_0 determines the time scale and the

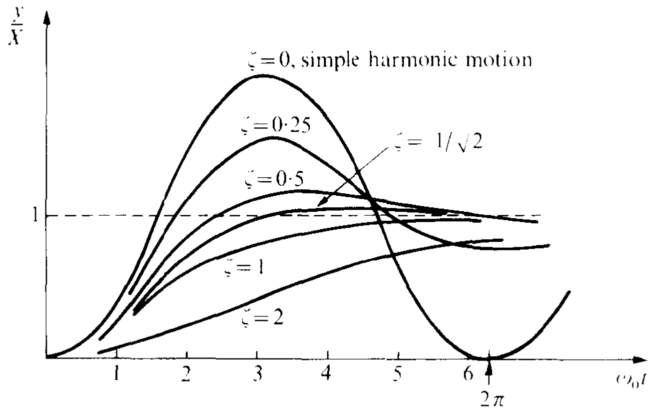

FIG. 1.6. Step responses of second-order systems.

constant ζ determines the shape of the response. ω_0 is called the 'natural frequency' and is expressed in radians per unit time; ζ is called the 'damping ratio' of the equation and is dimensionless. When ζ is less than unity the response is oscillatory and the equation is said to be 'under-damped', when ζ is greater than unity there is no overshoot and the equation is said to be 'over-damped', and when ζ is equal to unity the equation is said to be 'critically damped'.

The second-order equation can describe a position-control system where the angular position y of a rotating load of inertia J is controlled by a motor that applies torque u to the load. Figure 1.7 shows such a system in which the

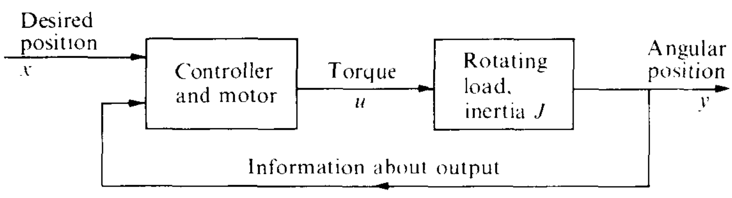

FIG. 1.7. Position-control system.

torque u is controlled by comparing information about the actual output y with information about the desired output x. The differential equation of the rotating load is given by Newton's equation of motion

$$J\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = u$$