



OpenCourseWare

Multiple-Input Describing Functions
and Nonlinear System Design

Arthur Gelb and Wallace Vander Velde

MULTIPLE-INPUT DESCRIBING FUNCTIONS AND NONLINEAR SYSTEM DESIGN

Arthur Gelb, Sc.D.

*President and Technical Director
The Analytic Sciences Corporation
Reading, Massachusetts*

Wallace E. Vander Velde, Sc.D.

*Professor of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts*

McGraw-Hill Book Company

*New York St. Louis
San Francisco Toronto
London Sydney*

**MULTIPLE-INPUT DESCRIBING FUNCTIONS
AND NONLINEAR SYSTEM DESIGN**

McGraw-Hill Electronic Sciences Series

Editorial Board

Ronald Bracewell Colin Cherry Willis W. Harman
Edward W. Herold John G. Linvill Simon Ramo John G. Truxal

ABRAMSON Information theory and coding
BATTIN Astronautical guidance
BLACHMAN Noise and its effect on communication
BREMER Superconductive devices
BROXMEYER Inertial navigation systems
GELB AND VANDER VELDE Multiple-input describing functions and nonlinear system design
GILL Introduction to the theory of finite-state machines
HANCOCK AND WINTZ Signal detection theory
HUELSMAN Circuits, matrices, and linear vector spaces
KELSO Radio ray propagation in the ionosphere
MERRIAM Optimization theory and the design of feedback control systems
MILSUM Biological control systems analysis
NEWCOMB Linear multiport synthesis
PAPOULIS The fourier integral and its applications
STEINBERG AND LEQUEUX (TRANSLATOR R. N. BRACEWELL) Radio astronomy
WEEKS Antenna engineering

PREFACE

The theory of automatic control has been advanced in important ways during recent years, particularly with respect to stability and optimal control. These are significant contributions which appeal to many workers, including the writers, because they answer important questions and are both theoretically elegant and practically useful. These theories do not, however, lay to rest all questions of importance to the control engineer. The designer of the attitude control system for a space vehicle booster which, for simplicity, utilizes a rate-switched engine gimbal drive, must know the characteristics of the limit cycle oscillation that the system will sustain and must have some idea of how the system will respond to attitude commands while continuing to limit-cycle. The designer of a chemical process control system must be able to predict the transient oscillations the process may experience during start-up due to the limited magnitudes of important variables in the system. The designer of a radar antenna pointing system with limited torque capability must be able to predict the rms pointing error due to random wind disturbances on the antenna, and must understand how these random disturbances will influence the behavior of the system in its response to command inputs. But more important than just being able to evaluate how a given system will behave in a postulated situation is the fact that these control engineers must *design* their

systems to meet specifications on important characteristics. Thus a complicated exact analytical tool, if one existed, would be of less value to the designer than an approximate tool which is simple enough in application to give insight into the trends in system behavior as a function of system parameter values or possible compensations, hence providing the basis for system design. As an analytical tool to answer questions such as these in a way which is useful to the system designer, the multiple-input describing function remains unexcelled.

This book is intended to provide a comprehensive documentation of describing function theory and application. It begins with a unified theory of quasi-linear approximation to nonlinear operators within which are embraced all the useful describing functions. It continues with the application of these describing functions to the study of a wide variety of characteristics of nonlinear-system operation under different input conditions. Emphasis is given to the *design* of these nonlinear systems to meet specified operating characteristics. The book concludes with a complete tabular and graphical presentation of the different describing functions discussed in the text, corresponding to a broad family of nonlinear functions. Dealing as it does with the single subject of describing functions, the book would seem to be very specialized in scope. And so it is. Yet the range of practical and important questions regarding the operation of nonlinear systems which this family of describing functions is capable of answering is so broad that the writers have had to set deliberate limits on the lengths of chapters to keep the book within reasonable size. Thus the subject is specialized to a single analytical tool which has exceedingly broad applicability.

This presentation is intended both for graduate students in control theory and for practicing control engineers. Describing function theory is applicable to problems other than the analysis and design of feedback control systems, and this is illustrated by some of the examples and problems in the book. But the principal application has been to control systems, and this has been the major focus of the book. The presentation is too comprehensive, and the subject too specialized, for the book to serve as the textbook in most graduate control courses, but it can serve very well as one of several reference books for such courses. In a graduate control course in the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology, the subject of this book is covered in a period of four or five weeks—twelve to fifteen lecture hours. The presentation of this book is not abbreviated primarily by omitting whole sections; rather, the principal ideas of almost every section are summarized briefly in class. A selection of these concepts is further developed through the problems. Such a presentation does not bring the student to the point of mastery of the subject, but it can give him a good understanding of the principal ideas underlying describing function theory and application. With this, the student can recognize the areas of useful applicability and can readily use the book as a reference to help him address the problems that arise in his professional experience.

The practicing control engineer should find the book valuable as a complete reference work in the subject area. If his background in mathematics is not sufficient to enable him to follow the theoretical development of Chapter 1 comfortably, he can omit that chapter and will still find a complete presentation in every chapter except Chapters 7 and 8, based on the physically motivated concept of

harmonic analysis of the nonlinearity output. Chapter 7, which includes random processes at the nonlinearity input, requires a statistical approach. But this too reduces to a rather simple matter in the very important class of problems involving static single-valued nonlinearities. Chapter 8 treats transient responses by related forms of quasi-linearization which are developed completely within that chapter. Thus it is hoped that every control engineer will find the principal ideas presented in a manner which is meaningful and appealing to him.

It is a pleasure for the writers to acknowledge the contributions of people who helped in different ways to see this project to completion. We express sincere appreciation to Hazel Leiblein for typing large portions of the manuscript; to Allan Dushman and Laurie J. Henrikson for a careful reading of several chapters; and to Martin V. Dixon, who volunteered to prepare the graphed data on the relay with dead zone which are presented in Appendix F. Special appreciation is due our understanding wives, Linda and Winni, who accepted long evenings over a period of several years without the company of their husbands.

Arthur Gelb
Wallace E. Vander Velde

CONTENTS

Preface v

Chapter 1 Nonlinear Systems and Describing Functions I

- 1.0 Introduction 1
- 1.1 Nonlinear-system Representation 3
- 1.2 Behavior of Nonlinear Systems 7
- 1.3 Methods of Nonlinear-system Study 9
- 1.4 The Describing Function Viewpoint 14
- 1.5 A Unified Theory of Describing Functions 18
- 1.6 About the Book 37

Chapter 2 Sinusoidal-input Describing Function (DF) 41

- 2.0 Introduction 41
- 2.1 Asymptotic Methods for the Study of Nonlinear Oscillations 43
- 2.2 Equivalent Linearization and the DF 47
- 2.3 DF Calculation for Frequency-independent Nonlinearities 55
- 2.4 DF Calculation for Frequency-dependent Nonlinearities 75
- 2.5 Synthesis of DFs 86

- 2.6 Techniques for Approximate Calculation of the DF 89
- 2.7 DF Inversion 97

Chapter 3 Steady-state Oscillations in Nonlinear Systems 110

- 3.0 Introduction 110
- 3.1 Determination of Limit Cycles 111
- 3.2 Limit Cycle Stability 120
- 3.3 Frequency Response of Non-limit-cycling Nonlinear Systems 125
- 3.4 Application to a Time-optimal Computer Control System 138
- 3.5 Linear and Nonlinear Loop Compensation Techniques 144
- 3.6 Treatment of Multiple Nonlinearities 154
- 3.7 Accuracy of the DF Approximation 162
- 3.8 Exact Methods for Relay Control Systems 185

Chapter 4 Transient Oscillations in Nonlinear Systems 212

- 4.0 Introduction 212
- 4.1 Analytic Description of Transient Oscillations 213
- 4.2 Relation to Other Work 220
- 4.3 Solution of the Equations Defining the Oscillation 223
- 4.4 Limit Cycle Dynamics 231
- 4.5 Limit Cycle Stability 241

Chapter 5 Two-sinusoid-input Describing Function (TSIDF) 250

- 5.0 Introduction 250
- 5.1 TSIDF Calculation 253
- 5.2 Subharmonic Oscillations 267
- 5.3 Frequency Response Counterexamples 271
- 5.4 Multiple Limit Cycles 274
- 5.5 Incremental-input Describing Function 276
- 5.6 Additional TSIDF Applications 286

Chapter 6 Dual-input Describing Function (DIDF) 297

- 6.0 Introduction 297
- 6.1 Mathematical Formulation of the DIDF 300
- 6.2 DIDF Calculation 305
- 6.3 Forced Response of Limit Cycling Nonlinear Systems 317
- 6.4 A Scheme for Parameter-adaptive Control 329
- 6.5 Application to an Adaptive Missile Roll Control System 332
- 6.6 Limit Cycles in Systems with an Asymmetric Nonlinearity 340
- 6.7 Artificial Dither and Signal Stabilization 345
- 6.8 TSIDF Calculation via the DF of a DIDF 357
- 6.9 Basis for Higher-order Approximations 358

Chapter 7 Random and Other Signals in Nonlinear Systems 365

- 7.0 Introduction 365
- 7.1 Statistical Linearization 366

7.2	Calculation of Random-input Describing Functions (RIDFs)	370
7.3	Feedback Systems with Random Signals and Noise	396
7.4	Feedback Systems with Random and Other Inputs	414
7.5	Alternative Approach for Nonlinearities with Memory	429

Chapter 8 Nonoscillatory Transients in Nonlinear Systems 438

8.0	Introduction	438
8.1	Transient-input Describing Function	439
8.2	Linear-system Approximations	443
8.3	Transient Response of Non-limit-cycling Nonlinear Systems	448
8.4	Design Procedure for a Specified Transient Response	452
8.5	Exponential-input Describing Function	454

Chapter 9 Oscillations in Nonlinear Sampled-data Systems 461

9.0	Introduction	461
9.1	Limit Cycles in Sampled Two-level Relay Systems	465
9.2	Limit Cycles in Other Sampled Nonlinear Systems	472
9.3	Stability of Limit Cycle Modes	492
9.4	Exact Verification of Limit Cycle Modes	497
9.5	Limit Cycles in Pulse-width-modulated Systems	505

Appendix A Amplitude-ratio-Decibel Conversion Table 515

Appendix B Table of Sinusoidal-input Describing Functions (DFs) 519

Appendix C Table of Dual-input Describing Functions (DIDFs) 539

Appendix D Table of Two-sinusoid-input Describing Functions (TSIDFs) 557

Appendix E Table of Random-input Describing Functions (RIDFs) 565

E.1	Gaussian-input RIDFs	565
E.2	Gaussian-plus-bias-input RIDFs	577
E.3	Gaussian-plus-bias-plus-sinusoid-input RIDFs	588

Appendix F Table of Sampled Describing Functions 611

Appendix G Analytical Justification for the Filter Hypothesis 625

Appendix H Introduction to Probability and Random Processes 629

Index 649

I NONLINEAR SYSTEMS AND DESCRIBING FUNCTIONS

I.0 INTRODUCTION

A system whose performance obeys the principle of superposition is defined as a *linear* system. This principle states that if the input $r_1(t)$ produces the response $c_1(t)$, and the input $r_2(t)$ yields the response $c_2(t)$, then for all a and b the response to the input $ar_1(t) + br_2(t)$ will be $ac_1(t) + bc_2(t)$; and this must be true for all inputs. A system is defined as *time-invariant* if the input $r(t + \tau)$ produces the response $c(t + \tau)$ for all input functions $r(t)$ and all choices for τ .

The simplest class of systems to deal with analytically is of course the class of *linear invariant* systems. For such systems the choice of time origin is of no consequence since any translation in time of the input simply translates the output through the same interval of time, and the responses to simple input forms can be superimposed to determine the responses to more complex input forms. This permits one in principle to *generalize from the response for any one input to the responses for all other inputs*. The elementary input function most commonly used as the basis for this generalization is the

2 NONLINEAR SYSTEMS AND DESCRIBING FUNCTIONS

unit-impulse function; the response to this input is often called the system *weighting function*. All possible modes of behavior of a linear invariant system are represented in its weighting function. Having once determined this function, which is just the response to one particular input, the performance of such a system can hold no surprises.

A *linear variable* system, although it still obeys the principle of superposition, is appreciably more difficult to deal with analytically. The response to a single input function in this case does not suffice to define the responses to all inputs; rather, a one-parameter infinity of such responses is needed. A single elementary form of input such as the unit-impulse function is adequate, but this input must be introduced, and the response determined, with all translations in time. Furthermore, the calculation of such responses very often cannot be done analytically. For invariant systems, the calculation of the weighting function requires the solution of a linear invariant differential equation, and there is a well-established procedure for finding the homogeneous solution to all such equations. There is no comparable general solution procedure for linear variable differential equations, and so the determination of the time-variable weighting function for linear variable systems usually eludes analytic attack.

Any system for which the superposition principle does not hold is defined to be *nonlinear*. In this case there is *no possibility of generalizing from the responses for any class of inputs to the response for any other input*. This constitutes a fundamental and important difficulty which necessarily requires any study of nonlinear systems to be quite specific. One can attempt to calculate the response for a specific case of initial conditions and input, but make very little inference based on this result regarding response characteristics in other cases.

In spite of the analytic difficulties, one has no choice but to attempt to deal in some way with nonlinear systems, because they occupy very important places in anyone's list of practical systems. In fact, linear systems can be thought of only as approximations to real systems. In some cases, the approximation is very good, but most physical variables, if allowed to take large values, will eventually run out of their range of reasonable linearity. *Limiting* is almost universally present in control systems since most instrumented signals can take values only in a bounded range. Many error detectors, such as a *resolver* or *synchro differential*, have a restricted range of linearity. Most drive systems, such as *electrical* and *hydraulic actuators*, can be thought of as linear over only small ranges, and others, such as *gas jets*, have no linear range at all. The use of *digital data processing* in control systems inevitably involves signal quantization. These are examples of nonlinear effects which the system designer would prefer to avoid, but cannot. There are good reasons why he might also choose to design some nonlinear effects into his system. The use of a *two- or three-level switch* as a controller,

switching the power supply directly into the actuator, often results in a considerable saving of space and weight, compared with a high-gain chain of linear amplification, ending with a power amplifier to drive the actuator. Also, controllers of sophisticated design, such as *optimal* or *final-value controllers*, often require nonlinear behavior.

As these few examples illustrate, the trend toward smaller and lighter-weight systems, the demand for higher-performance systems, and the increasing utility of digital operations in control systems, all conspire to broaden the place that nonlinear systems occupy. Thus the need for analytic tools which can deal with nonlinear systems in ways that are useful to the system designer continues to grow. This book treats a practical means of studying some of the performance characteristics of a broad class of nonlinear invariant systems. The techniques presented here can be, and have been, extended to some special cases of nonlinear variable systems, and the possibilities for doing so are relatively clear, once the basic analytic tool is well understood.

1.1 NONLINEAR-SYSTEM REPRESENTATION

Most systems can be considered an interconnection of components, or subsystems. In most cases, some of these subsystems are well characterized as linear, whereas others are more distinctly nonlinear. This results in a system configuration which is an interconnection of separable linear and nonlinear parts. The systems which are most commonly considered in this book are a further specialization of these: the class of systems which can be reduced to a *single-loop configuration with separable linear and nonlinear parts*. Some special cases of multiple-nonlinearity systems arranged in multiple loops which cannot be reduced are considered, but the configuration most commonly referred to is that of Fig. 1.1-1. This diagram could equally well represent a temperature control system, an inertial navigator platform gimbal servo, an aircraft stabilizer servo, a spacecraft telescope position control system, or a machine-tool positioning system. In each instance we might expect to

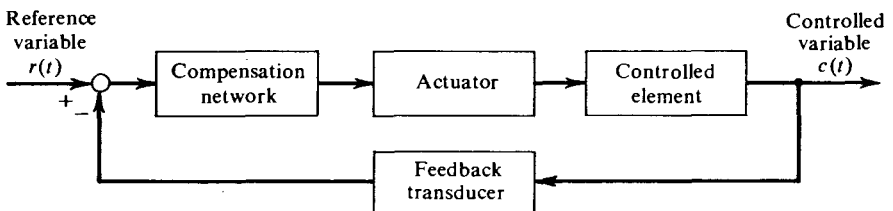


Figure 1.1-1 General control system block diagram.

find nonlinear effects in the actuator or feedback transducer, or even both, whereas the controlled element and loop compensation might well be linear.

To permit reference to certain classes of separable nonlinear elements, it is appropriate to classify them according to type in some sense. The broadest distinction to be made is between *explicit* and *implicit* nonlinearities. In the former case, the nonlinearity output is explicitly determined in terms of the required input variables, whereas in the latter case the output is defined only in an implicit manner, as through the solution of an algebraic or differential equation. Among the explicit nonlinearities, the next division is between *static* and *dynamic* nonlinearities. In the former case the nonlinearity output depends only on the input function, whereas in the latter, the output also depends on some derivatives of the input function. Among the static nonlinearities, a further distinction is drawn between *single-valued* and *multiple-valued* nonlinearities. In the case of a static, single-valued nonlinearity, the output is uniquely given in terms of the current value of the input, whereas more than one output may be possible for any given value of the input in the case of a static multiple-valued nonlinearity. The choice among the multiple values is made on the basis of the previous history of the input; thus such a nonlinearity is said to possess *memory*. One can imagine dynamic multiple-valued nonlinearities as well, but we shall not have occasion to refer to any such in this book. These are the major distinctions among nonlinearities from the point of view of the theory to be developed here. Other characteristics, such as *continuous* vs. *discontinuous*, are of little consequence here, but can be of supreme importance in other contexts.

An example of a static, single-valued, continuous, piecewise-linear nonlinearity is the *deadband gain*, or *threshold* characteristic (Fig. 1.1-2a). It could represent the acceleration input-voltage output relationship of a pendulous accelerometer, or the input-output characteristic of an analog angular position transducer. It is described by

$$y = \begin{cases} k(x - \delta) & \text{for } x \geq \delta \\ 0 & \text{for } -\delta \leq x < \delta \\ k(x + \delta) & \text{for } x < -\delta \end{cases} \quad (1.1-1)$$

where x and y denote the nonlinearity input and output, respectively. A static, multiple-valued, discontinuous, piecewise-linear nonlinearity is the *relay with deadband and hysteresis* (Fig. 1.1-2b). Arrows denote the direction in which this characteristic must be traversed in the determination of the output for a given input. The *history* of the input determines the value of the output in the multiple-valued regions. This characteristic is representative of the actuator switch in a temperature control system (in which case only the first quadrant portion applies) or the on-off gas jets in a spacecraft angular orientation system.

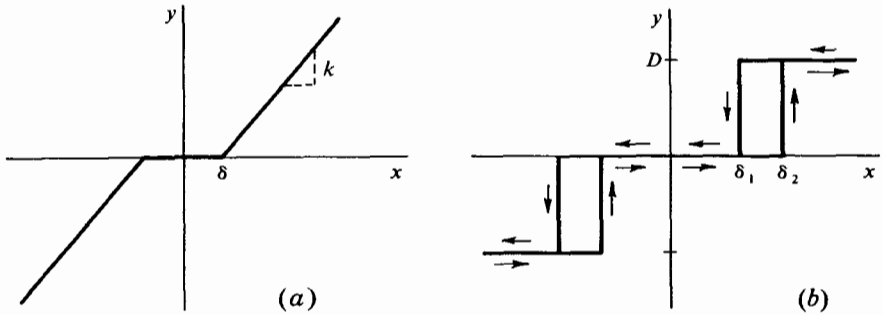


Figure 1.1-2 Examples of two static nonlinear characteristics. (a) Threshold; (b) relay with deadband and hysteresis.

Nonlinear differential equations illustrate *implicit dynamic* nonlinearities. For example, the equation

$$\dot{y}^{\frac{1}{3}} + 2y = x \quad (1.1-2)$$

represents such a nonlinearity, whereas

$$y = e^{\dot{x}+x} \quad (1.1-3)$$

portrays an *explicit dynamic* nonlinearity. It is to be noted that the process of converting implicit nonlinear differential relationships to explicit relationships is precisely the process of solving nonlinear differential equations—a process which is impossible for most equations of interest. For this reason, when they occur, we are usually forced to work directly with the implicit relationships themselves.

It is sometimes possible to recast a nonlinearity into a simpler form than that in which it is originally presented. An example is the implicit nonlinearity of Eq. (1.1-2). This differential relation can be represented by the feedback configuration of Fig. 1.1-3, just as if the equation were to be solved using

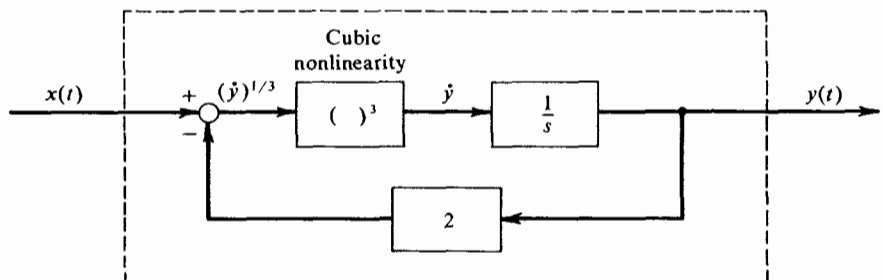


Figure 1.1-3 Closed-loop formulation of the implicit dynamic nonlinearity $\dot{y}^{\frac{1}{3}} + 2y = x$.

an analog computer. Only the explicit cubic nonlinearity appears in this diagram. Thus, if this nonlinear relation were part of a larger system, and the feedback path of Fig. 1.1-3 were absorbed into that system, we should have succeeded in trading an implicit dynamic nonlinearity for an explicit static single-valued nonlinearity. The static multiple-valued nonlinearity of Fig. 1.1-2b can be reduced to a static single-valued nonlinearity with a feedback path as shown in Fig. 1.1-4a. This, again, is an exact representation. An approximate representation of the hysteresis nonlinearity (multiple-valued) by the deadband gain nonlinearity (single-valued) in a feedback loop, together with an integrator and a high gain, is shown in Fig. 1.1-4b. The forward gain in this approximation must be chosen so that the bandwidth of the loop, when the deadband gain is operated in its linear range, is wide compared with the bandwidth of the system of which the hysteresis element is a part. In each case the feedback path of the transformed nonlinearity is then associated with the transfer of the rest of the system to separate the linear and nonlinear parts. The primary limitation on this technique of transforming a nonlinearity to simpler form is the fact that the feedback path of the transformed nonlinearity contains no filtering. This characteristic will be found undesirable for application of the theory developed in this book. The importance of this unfiltered feedback in any particular case depends on

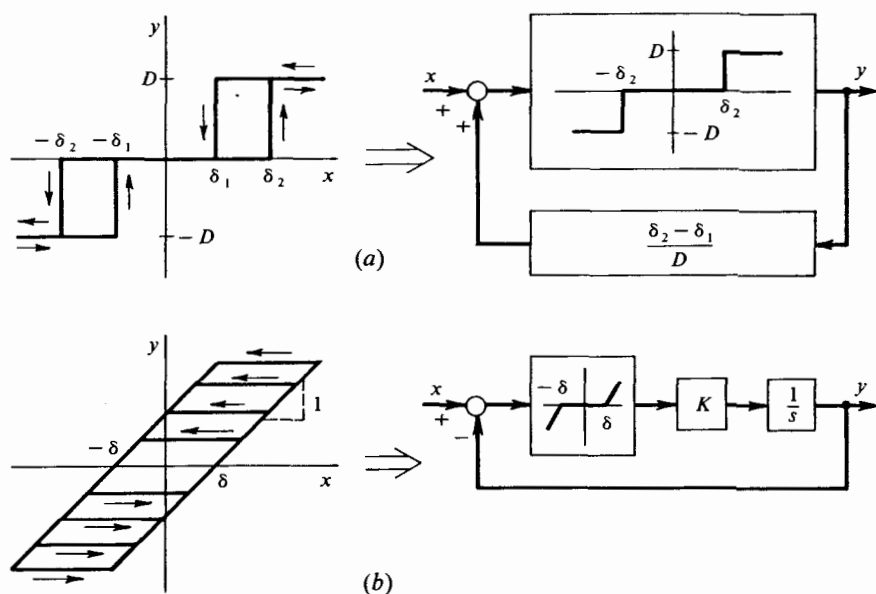


Figure 1.1-4 Transformation of multiple-valued nonlinearities into single-valued nonlinearities with feedback loops.

the relative amplitudes of the signals fed back to the nonlinearity through this path and through the linear part of the system.

If one wishes to study nonlinear differential equations which may arise out of any context whatsoever, it is often possible to collect the linear and nonlinear terms in the equation into separate linear and nonlinear parts, and then, using techniques similar to that of the preceding paragraph, arrange the resulting equation into a feedback configuration. The result is a closed loop having separated linear and nonlinear parts, which falls into the pattern of Fig. 1.1-1.

1.2 BEHAVIOR OF NONLINEAR SYSTEMS

The response of a linear invariant system to any forcing input can be expressed as a convolution of that input with the system weighting function. The response for any initial conditions of whatever magnitude can always be decomposed into the same set of normal modes which are properties of the system. The normal modes of all such systems are expressible as complex exponential functions of time, or as real exponentials and exponentially increasing or decreasing sinusoids. A special case of the latter is the undamped sinusoid, a singular situation. If a system displays an undamped sinusoidal normal mode, the amplitude of that mode in the response for a given set of initial conditions is, as for all other normal modes, dependent on the initial conditions.

The response characteristics of nonlinear systems, on the other hand, cannot be summarized in a way such as this. These systems display a most interesting variety of behavioral patterns which must be described and studied quite specifically. The most obvious departure from linear behavior is the *dependence of the response on the amplitude of the excitation*—either forcing input or initial conditions. Even in the absence of input, nonlinear systems have an important variety of response characteristics. A fairly common practical situation is that in which the system responds to small (in some sense) initial conditions by returning in a well-behaved manner to rest, whereas it responds to large initial conditions by diverging in an unstable manner. In other cases, the response to certain initial conditions may lead to a continuing oscillation, the characteristics of which are a property of the system, and not dependent on initial conditions. Such an oscillation may be viewed as a trajectory in the state space of the system which closes on itself and thus repeats; it is called a *limit cycle*. Nonlinear systems may have more than one limit cycle, and the one which is established will depend on the initial conditions, but the characteristics of each member of this discrete set of possible limit cycles are not dependent on initial conditions—they are properties of the system. This phenomenon is not possible in linear systems.

Since a limit cycle, once excited, will continue in the absence of further excitation, it constitutes a point of very practical concern for the system designer, and an analytic tool to study limit cycles is of evident importance to him.

A simple example of a limit cycling system is the Van der Pol oscillator, which obeys the equation

$$\ddot{x} - \eta(1 - x^2)\dot{x} + \omega_0^2 x = 0 \quad \eta > 0 \quad (1.2-1)$$

If $x(t)$ is always small compared with 1, this oscillator appears unstable since the effective damping is negative. Thus the small signals will tend to grow. If $x(t)$ were an oscillation with amplitude much greater than 1, it seems qualitatively clear that the average effective damping would appear positive, and the large signals would tend to decrease. At some amplitude of oscillation, where the average $|x(t)|$ is of the order of 1, the average effective damping would appear to be zero, and the oscillation would continue with that amplitude. This heuristic argument is given quantitative significance in the following chapters.

The response of nonlinear systems to forcing inputs presents an infinite variety of possibilities, just a few examples of which will be cited here. An input which plays a central role in linear invariant system theory is the sinusoid. Its importance is due primarily to the fact that the stability of a closed-loop linear system can be determined from the steady-state response of the open-loop system to the set of sinusoids of all frequencies. The amplitude of the sinusoids is of no consequence since only the ratio of the input to output is needed, and for linear systems this is independent of amplitude. The response of a nonlinear system to sinusoidal inputs is an important characteristic of the system too, because of the near-sinusoidal character of the actual inputs that many systems may see. But the nature of this response characteristic is much more complex in this case. The steady-state response of a nonlinear system to a sinusoidal input is dependent upon amplitude as well as frequency, in general. And even more interesting characteristics may appear. The response may be multiple-valued—a large-amplitude mode and a small-amplitude mode both possible for the same sinusoidal input. Or the predominant output response to an input sinusoid may even be at a different frequency, either a subharmonic or superharmonic of the input frequency. A system which limit-cycles in the absence of input may continue to do so in the presence of a sinusoidal input, both frequency components being apparent in the response, or the limit cycle may be quenched by the presence of the input. Or again, a system which does not limit-cycle in the absence of input may break out into a limit cycle in the presence of a sinusoidal input.

Most of these characteristics—the multiple-mode response, the possible quenching or exciting of limit cycles—may also be true of nonlinear systems responding to other periodic inputs or to random inputs. And so the story

could be continued; but the only point to be made here is that nonlinear systems display a most interesting variety of behavioral characteristics. The brief listing of some of them given here has favored those which can best be studied by the methods of this book.

1.3 METHODS OF NONLINEAR-SYSTEM STUDY

A number of possible means of studying nonlinear-system performance may be cited. All are of importance since different systems may be most amenable to analysis by different methods. Also, it was noted earlier that nonlinear-system study must be quite specific since generalization on performance characteristics seems impossible. It may be expected, then, that different analytic techniques will be best suited to the study of different performance characteristics. In effect, different techniques can be used to ask different questions about system performance.

The most common methods of nonlinear-system study are listed here, with brief comment for the purpose of viewing the describing function against a background of alternative methods. Of greatest interest is the *usefulness* of each of these methods in the *design* of practical nonlinear systems.

PROTOTYPE TEST.

The most certain means of studying a system is to build one and test it. The greatest advantage of this is that it *avoids* the necessity of choosing a *mathematical model* to represent the important characteristics of the system. The disadvantages are clear: the technique is of limited help in design because of the *time* required to construct a series of trial systems, the *cost* involved, and in many cases the *danger* involved in trying a system about which little is known.

COMPUTER SIMULATION

The capability of modern computers—analogue, digital, and hybrid—is such that very complete simulations of complex systems can be made and studied in a practical way. Dependence upon computer simulation, however, is not a very good first step in nonlinear-system design. The attempt to design through a sequence of blind trials on a computer is costly and unsatisfying. Each computer run answers the question, “What is the response to one particular set of initial conditions and one particular input?” One must always wonder if he has tried enough initial-condition and input combinations to have turned up all interesting response characteristics of the system. The

computer, used simply to make a direct simulation of the system, is thus a limited tool in the early phases of system design. However, since any other tool which is more useful for that purpose will almost surely involve approximations, computer simulation to check the design and verify system performance is an appropriate, if not essential, *final step* before building the real system.

CLOSED-FORM SOLUTIONS

There are a number of nonlinear differential equations, mostly of second order, for which exact solutions have been found or for which certain properties of the solutions have been tabulated. These constitute a very small number of *special cases*, and it is rare indeed when a control-system problem or other problem arising out of a significant physical situation can be made to fit one of these cases.

PHASE-PLANE SOLUTION

The dynamic properties of a system can be described in terms of the differential equations of state, and an attempt made to solve for the trajectories of the system in the state space. But this is just another way to solve nonlinear differential equations, and it is rarely possible to effect the solution. For the special cases of *first-* and *second-order systems*, however, this approach is useful because the two dimensions available on a flat piece of paper are adequate to display completely the state of these systems. Thus graphical techniques can be used to solve for the state, or phase, trajectories. This allows the response to be calculated for any set of initial conditions and for certain simple input functions. More important, however, is the fact that certain properties of the trajectories, such as their slopes, can be displayed over the whole phase plane. This information helps to alleviate concern over whether enough specific trajectories have been calculated to exhibit all interesting response characteristics. Thus, when such phase trajectories can be determined and their characteristics portrayed over the whole phase space, for certain inputs, one has a most valuable attack on the problem. But this can rarely be achieved for systems of greater than second order.

LYAPUNOV'S DIRECT METHOD

One of the most important properties of a system, *stability*, can in principle be evaluated without calculating the detailed responses of the system from given initial conditions with given inputs. All that is necessary is an indication of whether the state trajectories in the vicinity of an equilibrium point

tend to move generally toward or away from that point. This concept has most evident application to systems operating without command inputs; certain simple forms of input can also be considered in some cases. An analytic procedure for ignoring the detailed characteristics of the trajectories themselves, and just observing whether or not they tend in a generalized sense toward an equilibrium point, is given by the direct method of Lyapunov. A positive definite scalar function of the state variables which has certain required properties is defined. It is referred to as a Lyapunov function; we shall denote it as $V(x)$, where x is the vector of state variables. The time rate of change of this function, $\dot{V}(x)$, is calculated for motion of x along the system state trajectories. The sign of this derivative function in each region of the state space determines whether the state trajectories in that region tend generally toward or away from the origin of the space which is taken at an equilibrium point. Stability or instability of the system can be demonstrated by showing connected regions, including and surrounding the equilibrium point in which $\dot{V}(x)$ has consistently a negative or positive sign.

Failure of any number of choices for the form of the Lyapunov function to demonstrate stability or instability conclusively indicates nothing regarding system properties; it just means that the functions tried did not fit properly the characteristics of the system. Only for linear systems do we have well-defined procedures for choosing functions which give useful indications of stability. For nonlinear systems one can try different functional forms, but the search for a good one often goes unrewarded. To quote Popov, who has worked extensively with the method, "The study of stability by means of Lyapunov theorems is in principle universal, but in practice limited" (Ref. 11).

It is even possible to construct $V(x)$ functions whose time derivatives would indicate *bounds on a system limit cycle*. The concept is very appealing, but its implementation has so far failed to produce usefully tight quantitative bounds (Refs. 6, 12).

SERIES-EXPANSION SOLUTION

A whole family of techniques exists which develop the solutions of nonlinear differential equations or express the dynamic properties of nonlinear systems in expansions of various forms. These expansions may be a series of nonlinear-system operators, a power series in some small system parameter, a power series in the running variable—time in the case of dynamic systems—or of some other form. The central question related to these expansions is the speed with which the series *converge*. One can often solve nonlinear differential equations by simply assuming a series form for the solution, such as a power series in the running variable, and solving for the coefficients in the series which cause the solution to obey the differential equation. But the solution form chosen in this way is completely arbitrary, and one has no

reason to expect that it will fit the actual solution efficiently. For example, if a system actually has a solution of the form

$$y(t) = A \sin \omega t \quad (1.3-1)$$

the assumed solution form

$$y_a(t) = a + bt + ct^2 + \cdots \quad (1.3-2)$$

cannot generate the solution for an interval of time comparable even with one period of the oscillation with a reasonable number of terms in the series.

More rapidly convergent expansions can be made if one can solve for the approximate response of the system and develop the solution in a series of functions which fit this response efficiently. If such a solution to a nonlinear-system problem is to be achieved, the leading term in the expansion must be the solution to a simpler problem which we are able to solve, and each succeeding term must be derivable from this in some tractable manner. If we confess that the only problems we are really able to solve are linear problems (this statement is intended to be a bit overgeneral), we must expect that the leading term in most useful series solutions will be the solution of a linear problem, and subsequent terms in the expansion will attempt to account for the nonlinear characteristics of the system. Such expansions can then be expected to converge rapidly only if the system is "slightly nonlinear," that is, if the system properties are describable to a good approximation as properties of a linear system. But this is not true of some of the simplest and most commonplace of nonlinear systems, such as a relay-controlled servo. Thus series methods, although they will continue to hold an important place in nonlinear-system theory, are almost certain to be restricted in applicability.

LINEARIZATION

The problem of studying a nonlinear system can be avoided altogether by simply replacing each nonlinear operation by an approximating linear operation and studying the resulting linear system. This allows one to say a great deal about the performance of the approximating system, but the relation of this to the performance of the actual system depends on the validity of the linearizing approximations. Linearization of nonlinear operations ordinarily can be justified only for *small departures* of the variables from nominal operating values. This is pictured in Fig. 1.3-1. Any response which carries variables through a range which exceeds the limits of reasonable linear approximation cannot be described using this technique unless the system is repeatedly relinearized about new operating points, and the resulting solutions patched together. In addition, some commonplace nonlinearities, among them the two-level switch, have a discontinuity at the point which should be

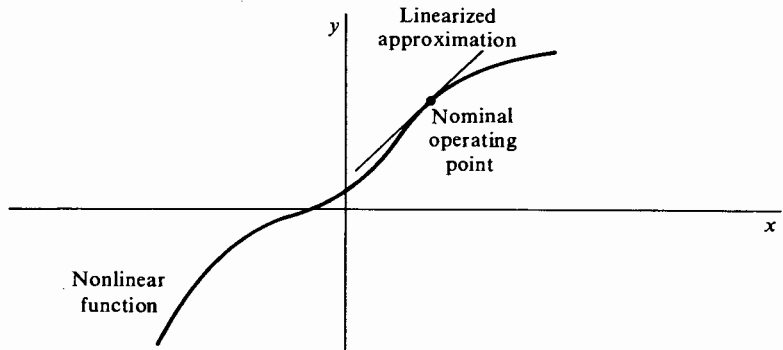


Figure 1.3-1 Illustration of true linearization.

chosen as the operating point. Linearization in the ordinary sense is not possible in these cases.

QUASI-LINEARIZATION

If the small-signal constraint of true linearization is to be relieved, but the advantages of a linear approximation retained, one must determine the operation performed by the nonlinear element on an *input signal of finite size* and approximate this in some way by a linear operation. This procedure results in different linear approximations for the same nonlinearity when driven by inputs of different forms, or even when driven by inputs of the same form but of different magnitude. The approximation of a nonlinear operation by a linear one which depends on some properties of the input is called *quasi-linearization*. It is a kind of linearization since it results in a linear description of the system, but it is not true linearization since the characteristics of the linear approximation change with certain properties of the signals circulating through the system.

This notion is illustrated in Fig. 1.3-2 for a general saturation-type nonlinearity. True or small-signal linearization about the origin would approximate the nonlinear function by a fixed gain which is the slope of the nonlinear function at the origin. However, if the signal at the input to this nonlinearity ranges into the saturated regions, it seems intuitively proper to say that the effective gain of the nonlinearity is lower than that for small signals around the origin. Such a gain, which depends on the magnitude of the nonlinearity input, is illustrated in the figure, and results from a quasi-linearization of the nonlinear function.

Quasi-linearization enjoys a very substantial advantage over true linearization in that there is *no limit to the range of signal magnitudes* which can be accommodated. Moreover, a completely linearized model can exhibit only

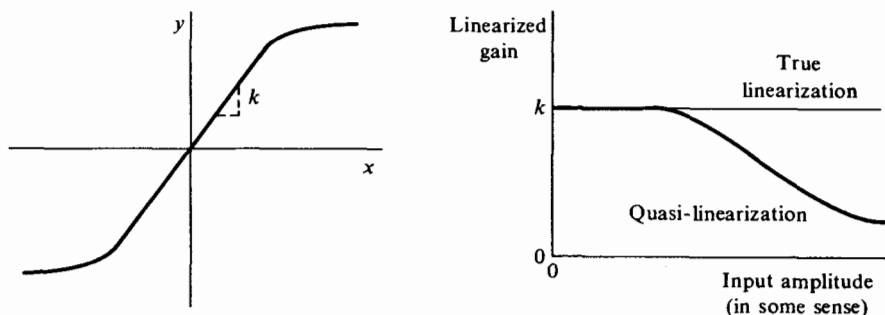


Figure 1.3-2 Illustration of quasi-linearization.

linear behavior, whereas a quasi-linearized model exhibits the basic characteristic of nonlinear behavior: *dependence of performance on signal amplitude*. On the other hand, a quasi-linearized model is more difficult to employ. A linearized model depends only on the system and the choice of nominal operating point. A quasi-linearized model depends on the system and certain properties of the signals circulating in that system. This gives rise to the inevitable requirement for simultaneous solution of two problems: (1) the quasi-linearized model is used to solve for the signals in the system, and (2) certain properties of these signals are used to define the quasi-linearized model. In spite of these difficulties, and more yet to be discussed, quasi-linearization stands as a most valuable tool in nonlinear-system study. A substantial number of interesting and important characteristics of nonlinear-system behavior can be studied better with this technique than with any other.

1.4 THE DESCRIBING FUNCTION VIEWPOINT

Quasi-linearization must be done for *specified input signal forms*. Any form of input to the nonlinear operator can be considered, the output calculated, and the result approximated by the result of a linear operation. However, for feedback-system configurations which are of primary interest to control engineers, the signal at the input to the nonlinearity depends both on the input to the system and the signal fed back within the system. The presence of the fed-back signal complicates considerably the determination of the form of signal which appears at the input to the nonlinearity. Practical solution of this problem for feedback-system configurations depends on avoiding the calculation of the signal form by *assuming* it to have a form which is guessed in advance. The forms which may reasonably be expected to appear at the nonlinearity input are those resulting from the filtering effect of the linear part

of the loop. This leads us to consider three basic signal forms with which to derive quasi-linear approximators for nonlinear operators:

1. *Bias*.¹ A constant component might be found in the signal at the nonlinearity input due to a bias in nonlinearity output which is propagated through the linear part and around the loop. Or, if the linear part contains one or more integrations, it can support a biased output even in the absence of a biased input.

2. *Sinusoid*. The limit of all periodic functions as the result of low-pass linear filtering is a sinusoid. Thus any periodic signal at the nonlinearity output would tend to look more like a sinusoid after propagation through the linear part and back to the nonlinearity input.

3. *Gaussian process*. Random processes with finite power density spectra tend toward gaussian processes as the result of low-pass linear filtering.² The restriction to finite power density spectra rules out bias and sinusoidal signals. But these have already been singled out for separate attention. Thus any random signal at the nonlinearity output may be expected to look more nearly gaussian after propagation through the linear part back to the nonlinearity input.

These three forms of signal, which we have some reason to expect to find at the input to the nonlinearity, are the principal bases for the calculation of approximators for nonlinear operators in this book. The quasi-linear approximating functions, which describe approximately the transfer characteristics of the nonlinearity, are termed *describing functions*. The major limitation on the use of these describing functions to describe system behavior is the requirement that the actual signal at the nonlinearity input approximate the form of signal used to derive the describing functions.

Within this requirement that the linear part of the system filter the output of the nonlinearity sufficiently, describing function theory provides answers to quite general questions about nonlinear-system operation. The response of systems to the whole class of inputs consisting of linear combinations of these limiting signal forms can be calculated. Even more general system inputs can be handled; the only requirement is that the input to the nonlinearity be of appropriate form. This includes, of course, the special case of zero input. The important problem of limit cycle determination is most expeditiously solved by describing function analysis. Situations involving certain combinations of limit cycle and forced response can also be treated: an example of a limit cycling system responding to a ramp command input and a random disturbance is discussed in the text. This must be considered

¹ Throughout the text, the term "bias" implies a constant, or dc, signal.

² This limiting behavior of many random processes is discussed briefly in Appendix H.

by any reasonable standard to be a very complicated problem, but it is susceptible to practical solution by the describing function technique. A significant further advantage is the fact that the solution procedure *does not break down*, in the sense of becoming very much more complicated, *for systems of higher order* than second or third, as is true of some analytic techniques. In the case of sinusoidal and constant signals, the order of the linear part of the system is of very little consequence. In the case of random signals, the analytic expressions for mean-squared signal levels become more complicated with increasing order of the system, but systems of order five, six, or seven are perfectly practical to deal with. And if graphical, rather than analytic, integration of the spectrum is used, systems of arbitrary order can be handled with nearly equal facility.

But the principal advantage of describing function theory is not that it permits the approximate calculation of the response of a given system to a given input or class of inputs; this can always be done by computer simulation. The real advantage, which justifies the development of an approximate theory such as this, is that it serves as a valuable aid to the *design* of nonlinear systems. There are certain situations in which describing functions permit a *direct synthesis* of nonlinear systems to optimize some performance index. In other cases, the *compensation* required to meet some performance specification becomes apparent upon application of describing function theory. In any case, the *trends in system performance characteristics* as functions of system parameters are probably more clearly displayed using describing function theory than with any other attack on nonlinear-system design. An analytic tool yielding this kind of information, even approximately, is of greater value to the system designer than an exact analytic tool which yields only specific information regarding the behavior of the system under specific circumstances.

The describing function technique has its limitations as well. The fundamental limitation is that the *form of the signal* at the input to the nonlinearity *must be guessed in advance*. For feedback configurations, this guess is usually taken to be one of the limiting signal forms discussed above for the reasons cited there. A less obvious limitation, which is probably true of every method of nonlinear-system study, is the fact that *the analysis answers only the specific questions asked of it*. If the designer does not ask about all important aspects of the behavior of a nonlinear system, describing function analysis will not disclose this behavior to him. For example, if one uses the two-sinusoid-input describing function to study subharmonic resonance, he would conclude—as many writers have—that a system with an odd static single-valued nonlinearity cannot support a subharmonic resonance of even order. Actually, the describing function is telling him that such a resonance cannot exist with just the two assumed sinusoids at the input to the nonlinearity. An even subharmonic resonance can indeed exist in such a

system, but it will be a biased asymmetric mode. Or again, use of the single-sinusoid-input describing function may indicate that a system has two stable limit cycles, and one might expect to see either one, depending on initial conditions. In some cases, however, use of the bias-plus-sinusoid-input describing function would show that in the presence of one of the limit cycles the system has an unstable small-signal mode. Thus the system is unable to sustain that limit cycle. We conclude that the analysis is tailored to the evaluation of particular response characteristics. The burden of deciding what characteristics should be inquired into rests with the system designer.

Another difficulty which the user of describing function theory must be alert to is the possibility of *multiple solutions*. Formulation of a problem using describing functions results in a simultaneous set of nonlinear algebraic relations to be solved. More than one solution may exist. These solutions represent different possible modes of response, some of which in some cases may be shown to be unstable. But the characteristics of these different solutions may be quite different, and the designer could be badly misled if he did not inquire into the possibility of other solutions. As an illustration of this, the gaussian-plus-sinusoid-input describing function can be used to determine how much random noise must be injected into a system to quench a limit cycle. The equations defining the behavior of the system may have a solution for a zero limit cycle amplitude and a certain rms value of the noise. However, one cannot conclude from this that the calculated rms value of noise will quench the limit cycle until he has assured himself that there is not also another solution for the same rms noise and a nonzero limit cycle amplitude.

A final limitation on the use of describing function theory is the fact that there is no satisfactory evaluation of the *accuracy* of the method. Research into this problem on the part of quite a few workers has resulted in some intuitively based criteria which are rather crude and some analytically based procedures which are impractical to use. All we have, then, is the fact that a great deal of experience with describing function techniques has shown that they work very well in a wide variety of problems. Furthermore, in those cases in which the technique does not work well, it is almost always obvious that the linear part of the system is providing very little low-pass filtering of the nonlinearity output. Finally, since the design of a nonlinear system must be based on the use of approximate analytic techniques, and these techniques will be inadequate to answer all questions regarding system behavior, the design must be checked—preferably by computer simulation—before it is approved. At that point in the design process one need not concern himself with checking the accuracy of the approximate analytic tools he has used in arriving at the design. Rather, his object is to check the design itself, to assure himself of its satisfactory performance in a variety of

simulated situations. The recommended procedure is, then, to use describing functions as an aid to system design, watching only for the obvious situations in which the theory might not have good application, and then to check the design by simulation.

1.5 A UNIFIED THEORY OF DESCRIBING FUNCTIONS

A quasi-linear approximator for a nonlinear operator is formed by noting the operation performed by the nonlinearity on an input of specified form, and approximating this operation in some way by a linear operation. We shall take the input to the nonlinearity to have a rather general form: consider the input $x(t)$ to be the sum of any number of signals, $x_i(t)$, each of an identifiable type. In later application, these input components $x_i(t)$ are taken to be constant signals, sinusoids, and gaussian processes, for the reasons discussed in the preceding section. For the purpose of the present theoretical development, however, no specialization is required. Corresponding to this form of input, the most general form of linear approximator for the nonlinearity is a parallel set of linear operators, one to pass each component of the input. Each input component to be considered is stationary, and if we restrict our attention to invariant nonlinearities, the linear operators which comprise the quasi-linear approximator can be taken as invariant at the outset. The resulting approximator for the nonlinearity is shown in Fig. 1.5-1. The $w_i(t)$ in this figure are the weighting functions for the filters which pass the different input components.

Having chosen this form for the quasi-linear approximator, it remains to decide on what basis to make the approximation, that is, what criterion to use in choosing the $w_i(t)$. The criterion used in the present development is minimum mean-squared error; the filters in the linear approximator are designed to minimize the mean-squared difference between the output of that approximator and the output of the nonlinearity. There are a number of reasons for this choice. As is always true in optimum linear theory—whether optimum filtering, optimum control, optimum estimation, or other special case—the quadratic error criterion, of all reasonable criteria, leads to the most tractable formulation of the optimum design problem. This suggests that the minimum mean-squared error criterion is advantageous, not only because it is analytically tractable, but also because the development based on this criterion runs a very close parallel to other optimum linear theory based on the same criterion. Thus, those who are familiar with, for example, Wiener's optimum filter theory for the separation of a signal from noise will have no difficulty following this theory. In addition, a criterion is desired which is universally applicable to all forms of input signal. No other such criterion has been demonstrated to give superior results over a broad range

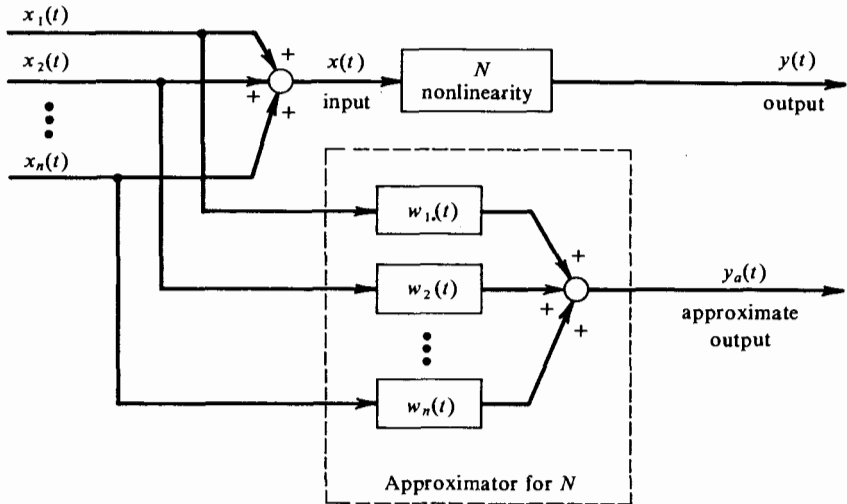


Figure 1.5-1 General linear approximator for a nonlinear operator.

of problems. This matter is discussed more fully in Chap. 2 with respect to sinusoidal inputs, and in Chap. 7 as it relates to random inputs.

This concept of forming a quasi-linear approximator to a nonlinear operator so as to minimize the mean-squared error was initiated by Booton (Ref. 3). He considered only a single input component—a random process with finite power density spectrum—and took the approximator to be a static gain. He showed separately that under the conditions he was considering, the optimum linear approximator, static or dynamic, was a static gain (Ref. 2). The concept was extended by Somerville and Atherton (Ref. 13) to treat an input of the type considered here: the sum of a number of components of identifiable form. They took the approximator to be a parallel set of static gains. In the present development we take the approximator to have the most general linear form: a parallel set of dynamic linear operators. Those cases in which the optimum linear operator is a static gain will appear as consequences of this theory.

The problem of determining the optimum linear approximator to a nonlinearity being driven by an input of specified form is treated here as a statistical problem. This is necessary to permit a unified attack. The repertory of input components to be considered must include random processes for which statistical techniques are essential. The deterministic signals to be considered as well can be formulated as simple forms of random processes; so a statistical approach embraces all forms of input. From this viewpoint, the mean-squared error which is to be minimized is seen as the expectation of the squared approximation error at any one time over all

members of the ensemble of possible inputs. The superscript bar used in this development indicates in every case an ensemble average. The reader who needs additional background in the statistics of random processes is referred to the brief presentation in Appendix H or the more complete discussions in Refs. 4 and 7 to 9.

THE OPTIMUM QUASI-LINEAR APPROXIMATOR

The linear approximator of the form shown in Fig. 1.5-1 which minimizes the mean-squared approximation error is now derived. The error in the approximation is

$$e(t) = y_a(t) - y(t) \quad (1.5-1)$$

and its mean-squared value

$$\overline{e(t)^2} = \overline{y_a(t)^2} - 2\overline{y_a(t)y(t)} + \overline{y(t)^2} \quad (1.5-2)$$

Now

$$y_a(t) = \sum_{i=1}^n \int_0^\infty w_i(\tau) x_i(t - \tau) d\tau \quad (1.5-3)$$

so

$$\begin{aligned} \overline{y_a(t)^2} &= \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 w_i(\tau_1) w_j(\tau_2) \overline{x_i(t - \tau_1) x_j(t - \tau_2)} \\ &= \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 w_i(\tau_1) w_j(\tau_2) \varphi_{ij}(\tau_1 - \tau_2) \end{aligned} \quad (1.5-4)$$

under the definition

$$\varphi_{ij}(\tau) = \overline{x_i(t) x_j(t + \tau)} \quad (1.5-5)$$

Also

$$\overline{y_a(t)y(t)} = \sum_{i=1}^n \int_0^\infty w_i(\tau) \overline{y(t) x_i(t - \tau)} d\tau \quad (1.5-6)$$

A necessary condition on the optimum set of weighting functions is derived from the observation that $\overline{e(t)^2}$ must be stationary with respect to variations in the $w_i(t)$ from the optimum set. To formulate an analytic statement of this requirement, we express each of the weighting functions as the optimum function plus a variation.

$$w_i(t) = w_{oi}(t) + \delta w_i(t) \quad (1.5-7)$$

The variations $\delta w_i(t)$ are arbitrary, except that they must represent physically realizable weighting functions. These expressions are used in Eqs. (1.5-4)

and (1.5-6) to give

$$\overline{y_a(t)^2} = \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 [w_{oi}(\tau_1) + \delta w_i(\tau_1)][w_{oj}(\tau_2) + \delta w_j(\tau_2)] \varphi_{ij}(\tau_1 - \tau_2) \quad (1.5-8)$$

$$\overline{y_a(t)y(t)} = \sum_{i=1}^n \int_0^\infty d\tau_1 [w_{oi}(\tau_1) + \delta w_i(\tau_1)] \overline{y(t)x_i(t - \tau_1)} \quad (1.5-9)$$

With these expressions for the terms appearing in Eq. (1.5-2), the mean-squared error can be written in expanded form. First, the terms which do not involve any of the variational functions constitute the stationary value of $\overline{e(t)^2}$; this value will be shown to be a minimum, and thus an optimum, value.

$$\begin{aligned} \overline{e(t)_o^2} &= \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 w_{oi}(\tau_1) w_{oj}(\tau_2) \varphi_{ij}(\tau_1 - \tau_2) \\ &\quad - 2 \sum_{i=1}^n \int_0^\infty d\tau_1 w_{oi}(\tau_1) \overline{y(t)x_i(t - \tau_1)} + \overline{y(t)^2} \end{aligned} \quad (1.5-10)$$

Next, the first-degree terms in variational functions constitute the first variation in $\overline{e(t)^2}$. It is this first variation which must vanish to define the stationary point.

$$\begin{aligned} \delta \overline{e(t)^2} &= \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 w_{oi}(\tau_1) \delta w_j(\tau_2) \varphi_{ij}(\tau_1 - \tau_2) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \delta w_i(\tau_1) w_{oj}(\tau_2) \varphi_{ij}(\tau_1 - \tau_2) \\ &\quad - 2 \sum_{i=1}^n \int_0^\infty d\tau_1 \delta w_i(\tau_1) \overline{y(t)x_i(t - \tau_1)} \\ &= 2 \sum_{i=1}^n \int_0^\infty d\tau_1 \delta w_i(\tau_1) \left[\sum_{j=1}^n \int_0^\infty d\tau_2 w_{oj}(\tau_2) \varphi_{ij}(\tau_1 - \tau_2) - \overline{y(t)x_i(t - \tau_1)} \right] \end{aligned} \quad (1.5-11)$$

In this reduction we have used the fact that $\varphi_{ji}(\tau_2 - \tau_1) = \varphi_{ij}(\tau_1 - \tau_2)$ for all i and j .

Finally, the second-degree term in variational functions is the second variation in $\overline{e(t)^2}$. The sign of this term determines the nature of the stationary point defined by the vanishing of the first variation.

$$\delta^2 \overline{e(t)^2} = \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \delta w_i(\tau_1) \delta w_j(\tau_2) \varphi_{ij}(\tau_1 - \tau_2) \quad (1.5-12)$$

This expression is identical in form with Eq. (1.5-4), the mean-squared output of the linear approximator. Thus the second variation may be interpreted as the mean-squared value of the sum of the outputs of a parallel set of filters with weighting functions $\delta w_i(t)$ subjected to the inputs $x_i(t)$. This must be positive; so the stationary point is shown to be a minimum.

The minimum mean-squared error is thus achieved by that set $w_{oj}(\tau_2)$ which causes Eq. (1.5-11) to be zero. But this must be true for arbitrary choices of the variations, $\delta w_i(\tau_1)$. We can be assured of the vanishing of the first variation, then, only if the bracketed term in Eq. (1.5-11) is zero for each value of i over the range of the τ_1 integration. Thus

$$\sum_{j=1}^n \int_0^\infty w_{oj}(\tau_2) \varphi_{ij}(\tau_1 - \tau_2) d\tau_2 = \overline{y(t)x_i(t - \tau_1)} \quad \tau_1 \geq 0, i = 1, 2, \dots, n \quad (1.5-13)$$

These are the conditions which define the optimum set of filter weighting functions. This result has a simple interpretation which is common to all optimum linear filter theory based on the mean-squared-error criterion. The left-hand member is $\overline{x_i(t)y_a(t + \tau_1)}$, the cross correlation between the i th input component and the approximate output. Because of the stationary character of the input components and nonlinearity output, the right-hand member can be written $\overline{x_i(t)y(t + \tau_1)}$, the cross correlation between the i th input component and the actual output of the nonlinearity. *The set of filters which minimizes the mean-squared approximation error is that set which equates, over the nonnegative range of their arguments, the input-output cross-correlation functions for the nonlinearity and its approximation. This cross-correlation equivalence is demanded for each component of the input according to Eq. (1.5-13). This property of cross-correlation equivalence is also true, for example, of the Wiener filter.*

Since the correlation between input and output is preserved in the approximate output, the error in the approximation must be uncorrelated with the input. This can readily be shown to be so; viz.,

$$\begin{aligned} \overline{x(t)e(t + \tau)} &= \sum_{i=1}^n \overline{x_i(t)[y_a(t + \tau) - y(t + \tau)]} \\ &= \sum_{i=1}^n \left[\sum_{j=1}^n \int_0^\infty w_{oj}(\tau_2) \overline{x_i(t)x_j(t + \tau - \tau_2)} d\tau_2 - \overline{y(t + \tau)x_i(t)} \right] \\ &= \sum_{i=1}^n \left[\sum_{j=1}^n \int_0^\infty w_{oj}(\tau_2) \varphi_{ij}(\tau - \tau_2) d\tau_2 - \overline{y(t)x_i(t - \tau)} \right] \\ &= 0 \quad \text{for } \tau \geq 0 \end{aligned} \quad (1.5-14)$$

according to Eq. (1.5-13). The correlation between actual and approximate

outputs is

$$\begin{aligned}
 \overline{y(t)y_a(t+\tau)} &= \sum_{i=1}^n \int_0^\infty w_{oi}(\tau_1) \overline{y(t)x_i(t+\tau-\tau_1)} d\tau_1 \\
 &= \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 w_{oi}(\tau_1) w_{oj}(\tau_2) \overline{\varphi_{ij}(\tau_1-\tau-\tau_2)} \\
 &= \overline{y_a(t)y_a(t-\tau)} \quad \tau \leq 0 \\
 &= \overline{y_a(t)y_a(t+\tau)} \quad \tau \geq 0
 \end{aligned} \tag{1.5-15}$$

So the cross correlation between actual and approximate outputs is equal to the autocorrelation of the approximate output over the indicated range of τ . The restrictions on the range of τ for which Eqs. (1.5-14) and (1.5-15) hold are due to the restricted range of τ_1 for which Eq. (1.5-13) is a required condition.

The statistics of the approximation error can now be written.

$$\begin{aligned}
 \overline{e(t)} &= \overline{y_a(t) - y(t)} \\
 &= \sum_{i=1}^n \int_0^\infty w_{oi}(\tau_1) \overline{x_i(t-\tau_1)} d\tau_1 - \overline{y(t)}
 \end{aligned} \tag{1.5-16}$$

If the input to the nonlinearity, $x(t) = \sum_{i=1}^n x_i(t)$, has a nonzero mean value, there is no prescribed way of associating this constant with the various components $x_i(t)$. The assignment of the mean to the $x_i(t)$ can be made arbitrarily with no loss of generality. The most convenient convention to employ is to assign all of the mean of $x(t)$ to one of the $x_i(t)$ which is just a constant, or bias, function. With this convention, all but one of the $x_i(t)$ are unbiased functions; the remaining $x_i(t)$ is a constant function equal to the mean of $x(t)$. Equation (1.5-16) then becomes

$$\overline{e(t)} = \overline{x(t)} \int_0^\infty w_{oi}(\tau_1) d\tau_1 - \overline{y(t)} \tag{1.5-17}$$

where $w_{oi}(\tau_1)$ is the weighting function of the filter which passes the bias input component. This weighting function has yet to be determined on the basis of minimizing the mean-squared approximation error; we shall find that the result also has the desirable property of reducing the mean error as expressed in Eq. (1.5-17) to zero.

The mean-squared approximation error, which is minimized by the set of filters that satisfy the conditions of Eq. (1.5-13), is

$$\begin{aligned}
 \overline{e(t)^2} &= \overline{y_a(t)^2} - 2\overline{y_a(t)y(t)} + \overline{y(t)^2} \\
 &= \overline{y_a(t)^2} - 2\overline{y_a(t)^2} + \overline{y(t)^2} \\
 &= \overline{y(t)^2} - \overline{y_a(t)^2}
 \end{aligned} \tag{1.5-18}$$

using Eq. (1.5-15) with $\tau = 0$. But this also gives

$$\overline{y_a(t)^2} = \overline{y(t)^2} - \overline{e(t)_o^2} \quad (1.5-19)$$

which demonstrates the fact that *linear filters which minimize the mean-squared approximation error always underestimate the mean-squared output of the nonlinearity*. This should not be interpreted necessarily as a failure of the approximator. The power in the output of the nonlinearity is spread over a wider band of frequencies than that of the approximate output, and thus suffers greater attenuation in passing through the linear part of the system. It is most important for the approximated system to yield nearly correct statistics for the input to the nonlinearity since the approximation itself depends on these quantities, and to achieve this we should expect that the mean-squared output of the approximator would have to be less than the mean-squared output of the nonlinearity.

The most general expression for the filter weighting functions is the solution of Eq. (1.5-13), a coupled set of simultaneous integral equations. However, the case of almost universal interest is that in which the various components of the input are *statistically independent*. Conditions under which this property may be expected to hold are discussed later, in connection with specific input forms. Independence of the input components, together with the convention of assigning any bias in $x(t)$ to one input component, thus leaving all but one component with zero mean, assures that the different input components are uncorrelated.

$$\begin{aligned} \varphi_{ij}(\tau) &= \overline{x_i(t)x_j(t+\tau)} \\ &= \overline{x_i(t)} \overline{x_j(t+\tau)} \\ &= 0 \quad i \neq j \end{aligned} \quad (1.5-20)$$

This serves to uncouple the components of Eq. (1.5-13) so that they become simply

$$\int_0^\infty w_{oi}(\tau_2) \varphi_{ii}(\tau_1 - \tau_2) d\tau_2 = \overline{y(t)x_i(t - \tau_1)} \quad \tau_1 \geq 0, i = 1, 2, \dots, n \quad (1.5-21)$$

This expression, which is hereafter used to define the filter weighting functions, defines a more specific form of cross-correlation equivalence than Eq. (1.5-13). For independent input components, the cross correlation between input and output of each filter is independently equated, over the nonnegative range of their arguments, to the cross correlation between that component of input and the output of the nonlinearity. It is still true, as was noted earlier, that the cross correlation between each component of input and the approximate output is being equated to the cross correlation between that component of input and the actual output of the nonlinearity; but in the case

of independent input components the entire cross correlation between $x_i(t)$ and $y_a(t)$ is due to the cross correlation between $x_i(t)$ and the output of the filter which passes that particular input component. So there is a kind of isolation among the parallel paths which make up the approximator to the nonlinearity; the signals passing through these paths are uncorrelated. However, each of the filter weighting functions depends not only on the form of the nonlinearity and the characteristics of the input component which that filter passes, but on the characteristics of all other input components as well, through the $y(t)$ term appearing in Eq. (1.5-21). Thus the basic nature of a nonlinear operator, the failure of the property of superposition, is evidenced in this approximator to the nonlinearity.

DESCRIBING FUNCTIONS CORRESPONDING TO SPECIFIC INPUT FORMS

We now specialize Eq. (1.5-21) to the cases of the limiting input signal forms under consideration, each in the presence of other independent input components.

Bias First determine the optimum filter to operate on the constant input component, the $x_i(t)$ which is the mean value of $x(t)$. To emphasize the nature of this input function we denote it by B , a bias, and write this component of the input explicitly.

$$x(t) = B + x_r(t) \quad (1.5-22)$$

Here the input to the nonlinearity is written as the sum of the bias component and the remainder, $x_r(t)$, which is the sum of all other components of $x(t)$. The remainder can be an arbitrary collection of functions provided that they are statistically independent of B . This is not to say that the *statistics* of the remainder are independent of B . When this nonlinearity is treated as part of a feedback system so that its input is related to its output, the statistics of the remainder, such as the standard deviation of a gaussian component, become functionally related to each other and to B . But the independence referred to here, which justifies the use of Eq. (1.5-21), simply requires that there be no correlation between the instantaneous value of the remainder at any one time and the value of the bias. This is always true since the bias is a deterministic quantity and the remainder is unbiased. Thus

$$\overline{Bx_r(t)} = \overline{Bx_r(t)} = 0 \quad (1.5-23)$$

The autocorrelation function for the bias input component, $x_i(t) = B$, is

$$\begin{aligned} \varphi_{ii}(\tau) &= \overline{x_i(t)x_i(t + \tau)} \\ &= \overline{B^2} \\ &= B^2 \end{aligned} \quad (1.5-24)$$

Equation (1.5-21) in this case becomes

$$\int_0^\infty w_B(\tau_2) B^2 d\tau_2 = \overline{y(t)} B \quad (1.5-25)$$

where the weighting function for the optimum filter to pass the bias input component has been written $w_B(\tau_2)$. Since we are considering only stationary statistics, the mean output of the nonlinearity is a constant, and the integral equation is clearly satisfied by

$$w_B(\tau_2) = \frac{1}{B} \overline{y(t)} \delta(\tau_2) \quad (1.5-26)$$

where $\delta(\tau_2)$ is the unit-impulse function of argument τ_2 . A filter whose weighting function, or response to a unit-impulse input, is a unit impulse scaled by K , is recognized to be just a static gain of magnitude K . This is not the only possible solution of Eq. (1.5-25), but any other solution would also represent a filter with the dc gain $(1/B)\overline{y(t)}$. The simplest solution is then a static gain of this magnitude. The generality involved in the original assumption of a dynamic filter to pass a bias component was not necessary since the only characteristic of that filter which is of consequence is its zero frequency gain. However, the analytic development was facilitated by the consistent assumption of a dynamic linear operator for each input component.

The linear filter operating on the bias input component which minimizes the mean-squared approximation error is thus found to be the static gain which equates the mean output of the approximator to that of the nonlinearity. This approximator may therefore be described as the unbiased minimum-variance quasi-linear approximator for the nonlinearity. This gain is called the *describing function* for the bias input component and is denoted by N_B .

$$\begin{aligned} N_B &= \frac{1}{B} \overline{y(t)} \\ &= \frac{1}{B} \overline{y(0)} \end{aligned} \quad (1.5-27)$$

This step utilizes the fact that only stationary inputs to the nonlinearity are considered. Since the nonlinearity is invariant, the output statistics are also stationary. Thus all expectations are independent of time, and can be calculated at any convenient time, such as $t = 0$. This is done repeatedly in what follows.

Sinusoid Now find the optimum filter to pass a sinusoidal component in the presence of other independent input components. The sinusoidal

component of the input is thus written explicitly.

$$x(t) = A \sin(\omega t + \theta) + x_r(t) \quad (1.5-28)$$

Note that the remainder is quite arbitrary provided that it is uncorrelated with the sinusoidal component we are singling out for attention; in particular, it may possibly include other sinusoidal components. The sinusoid considered here has a deterministic amplitude and frequency. If the phase angle of the sinusoid were known, there would remain no need to find a filter to approximate the transfer characteristics of the nonlinearity in passing the known sinusoid. One could then just replace the nonlinearity with a function generator which would produce the known output of the nonlinearity corresponding to the known sinusoidal input. This cannot in fact be done, since we do not have any a priori knowledge of the phase angle, and there is assumed to be no deterministic mechanism operating in the system which will fix the phase angle. The amplitude of a sinusoidal component at the input to a nonlinear device enclosed in a feedback loop is also unknown in advance, but the amplitude is determined by the nature of the system and its inputs. The phase angle, on the other hand, remains undetermined. It is measured against some arbitrary time reference. The basic performance of the system is independent of the choice of time reference, and thus of the phase angle of the sinusoidal component. There is no reason to anticipate that some phase angles will occur with greater likelihood than others; so the phase angle θ in Eq. (1.5-28) is taken as a random variable with uniform distribution over 2π radians. An important consequence of this particular distribution for θ , as noted in Appendix H, is that the resulting ensemble of all possible sinusoids of the form $A \sin(\omega t + \theta)$, with A and ω determined, is statistically stationary.

Since the phase angle θ is the only random variable associated with a sinusoidal input component, independence among several sinusoidal components means independence of their phase angles. A great many problem situations involving multiple sinusoids at the nonlinearity input do have this property of independence of phase. Some important situations in which this is not true are also discussed in Chap. 5. For the present, the describing function is determined under the assumption of independence.

For $x_i(t) = A \sin(\omega t + \theta)$, with A and ω determined, and θ uniformly distributed over 2π radians, the autocorrelation function is

$$\begin{aligned} \varphi_{ii}(\tau) &= \overline{x_i(t)x_i(t + \tau)} \\ &= \overline{x_i(0)x_i(\tau)} \\ &= \frac{1}{2\pi} \int_0^{2\pi} A^2 \sin \theta \sin(\omega\tau + \theta) d\theta \\ &= \frac{1}{2} A^2 \cos \omega\tau \end{aligned} \quad (1.5-29)$$

The left-hand member of Eq. (1.5-21) is in this case

$$\begin{aligned} & \int_0^\infty w_A(\tau_2) \frac{1}{2} A^2 \cos \omega(\tau_1 - \tau_2) d\tau_2 \\ &= \frac{1}{2} A^2 \cos \omega\tau_1 \int_0^\infty w_A(\tau_2) \cos \omega\tau_2 d\tau_2 + \frac{1}{2} A^2 \sin \omega\tau_1 \int_0^\infty w_A(\tau_2) \sin \omega\tau_2 d\tau_2 \end{aligned} \quad (1.5-30)$$

The right-hand member is

$$\begin{aligned} \overline{y(t)x_i(t - \tau_1)} &= \overline{y(0)x_i(-\tau_1)} \\ &= \overline{y(0)A \sin(\theta - \omega\tau_1)} \\ &= A \cos \omega\tau_1 \overline{y(0) \sin \theta} - A \sin \omega\tau_1 \overline{y(0) \cos \theta} \end{aligned} \quad (1.5-31)$$

Satisfaction of the equation for all nonnegative values of τ_1 requires

$$\frac{1}{2} A \int_0^\infty w_A(\tau_2) \cos \omega\tau_2 d\tau_2 = \overline{y(0) \sin \theta} \quad (1.5-32)$$

and

$$\frac{1}{2} A \int_0^\infty w_A(\tau_2) \sin \omega\tau_2 d\tau_2 = -\overline{y(0) \cos \theta} \quad (1.5-33)$$

The right-hand members of Eqs. (1.5-32) and (1.5-33) are statistics associated with the random variables $y(0)$ and θ ; they are constants. These equations are seen to be satisfied by

$$w_A(\tau_2) = \frac{2}{A} \overline{y(0) \sin \theta} \delta(\tau_2) + \frac{2}{A\omega} \overline{y(0) \cos \theta} \dot{\delta}(\tau_2) \quad (1.5-34)$$

where $\dot{\delta}(\tau_2)$ is the derivative of the unit-impulse function, the so-called "doublet." The integral of $f(t) \delta(t)$ across the point $t = 0$ is $f(0)$, if $f(t)$ is continuous at $t = 0$. Similarly, the integral of $f(t) \dot{\delta}(t)$ across $t = 0$ is $-f'(0)$, if $f(t)$ is continuous at $t = 0$, which can be verified by an integration by parts. The transfer function corresponding to this weighting function is the sum of a proportional plus derivative path.

$$W_A(s) = \frac{2}{A} \left[\overline{y(0) \sin \theta} + \left(\frac{1}{\omega} \overline{y(0) \cos \theta} \right) s \right] \quad (1.5-35)$$

This is not the only solution to Eqs. (1.5-32) and (1.5-33), but any other solution would represent a filter which has the same transfer at the frequency of the input sinusoid. As in the case of the bias input component, the assumption of an arbitrary dynamic linear operator to pass a sinusoidal input component is more general than necessary since only the transfer at the frequency of the input sinusoid is of consequence. At any one frequency, a linear filter is just a complex gain which modifies the amplitude and phase of the input. The real and imaginary parts of this complex gain may be thought of as in-phase and quadrature gains. Using notation appropriate

to this interpretation, the describing function for a sinusoidal input component in the presence of any other independent components is written

$$\begin{aligned} N_A &= n_p + jn_q \\ n_p &= \frac{2}{A} \overline{y(0) \sin \theta} \\ n_q &= \frac{2}{A} \overline{y(0) \cos \theta} \end{aligned} \quad (1.5-36)$$

For nonlinearities which are static and single-valued, so that $y(0)$ is given unambiguously in terms of $x(0)$, this quadrature gain is always zero. This can be seen from the following calculation. With $x(t)$ given by Eq. (1.5-28), we write

$$\begin{aligned} \overline{y(0) \cos \theta} &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_n f_n(p_1, \dots, p_n) y[A \sin \theta + x_r(0)] \cos \theta \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_n f_n(p_1, \dots, p_n) \int_0^{2\pi} y[A \sin \theta + x_r(0)] \cos \theta d\theta \end{aligned} \quad (1.5-37)$$

where $1/2\pi$ is the probability density function for θ ; $p_1, \dots, p_n$ are the random variables which define the remainder of the input, $x_r(0)$; and $f_n(p_1, \dots, p_n)$ is their joint probability density function. The θ integration in Eq. (1.5-37) is an integral over one period of a periodic function. The interval of integration is arbitrary provided only that it encompasses exactly one period.

$$\begin{aligned} &\int_0^{2\pi} y[A \sin \theta + x_r(0)] \cos \theta d\theta \\ &= \int_{-\pi/2}^{3\pi/2} y[A \sin \theta + x_r(0)] \cos \theta d\theta \\ &= \int_{-\pi/2}^{\pi/2} y[A \sin \theta + x_r(0)] \cos \theta d\theta + \int_{\pi/2}^{3\pi/2} y[A \sin \theta + x_r(0)] \cos \theta d\theta \end{aligned} \quad (1.5-38)$$

Change the variable of integration in the second term to $\theta' = \theta - \pi$.

$$\begin{aligned} \text{Second term} &= - \int_{-\pi/2}^{\pi/2} y[-A \sin \theta' + x_r(0)] \cos \theta' d\theta' \\ &= - \int_{-\pi/2}^{\pi/2} y[A \sin \theta'' + x_r(0)] \cos \theta'' d\theta'' \\ &= -\text{first term} \end{aligned} \quad (1.5-39)$$

The further change of variable $\theta'' = -\theta'$ was employed in this demonstration of the fact that Eq. (1.5-37), and thus the imaginary part of the describing function for a sinusoidal input to a static single-valued nonlinearity, is zero. For such a nonlinearity, the describing function for a sinusoidal input in the presence of any other uncorrelated inputs is a real static gain, the proportional gain of Eq. (1.5-36).

This gain is subject to an interesting interpretation.

$$N_A = \frac{1}{\pi A} \int_0^{2\pi} \left\{ \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_n f_n(p_1, \dots, p_n) y[A \sin \theta + x_r(0)] \right\} \sin \theta d\theta \quad (1.5-40)$$

The definitions of all variables and functions used here are identical with those of Eq. (1.5-37). The quantity in the braces—the integral with respect to $p_1, \dots, p_n$ —is the expectation of the output of the nonlinearity with θ fixed. This expectation is itself a nonlinear function of θ which appears only in the form of $A \sin \theta$. Call this new function $y'(A \sin \theta)$. In terms of this, the describing function for a sinusoidal input component in the presence of any other independent input components is written

$$N_A = \frac{1}{\pi A} \int_0^{2\pi} y'(A \sin \theta) \sin \theta d\theta \quad (1.5-41)$$

But this is just an ordinary harmonic analysis of the modified nonlinearity $y'(x)$. This interpretation is also true of N_A for more general nonlinearities than static single-valued. In that case both the sine and cosine components of the output of the modified nonlinearity must be calculated, and this calculation is considerably more difficult. If the remainder includes a random process, the determination of the expectation of the nonlinearity output with θ fixed is somewhat obscure when the present output depends not only on the present input, but also on the history of that input. However, if the input consists only of a bias and any number of sinusoids, the calculation of the modified nonlinearity is straightforward. In summary, then, *the describing function for a sinusoidal input component may be viewed as the amplitude and phase relationship between an input sinusoid and the fundamental harmonic component of the expectation of the output of the nonlinearity taken with respect to all statistical parameters except θ* . This definition of the gain of a nonlinearity to a sinusoid in the presence of other input components was employed by Vander Velde (Ref. 14), based on an intuitive argument. This property has since been utilized for computational purposes by a number of writers, among them Atherton (Ref. 1), Gusev (Ref. 5), Popov (Ref. 10), and Somerville and Atherton (Ref. 13).

Random signal The last of the signal forms which we are considering is the gaussian process.

$$x(t) = r(t) + x_r(t) \quad (1.5-42)$$

$r(t)$ is a member of a stationary ensemble, and the remainder, $x_r(t)$, is uncorrelated with $r(t)$. For $x_i(t) = r(t)$, the autocorrelation function is simply written as

$$\varphi_{ii}(\tau) = \varphi_{rr}(\tau) \quad (1.5-43)$$

and the corresponding form of Eq. (1.5-21) as

$$\begin{aligned} \int_0^\infty w_R(\tau_2) \varphi_{rr}(\tau_1 - \tau_2) d\tau_2 &= \overline{y(0)r(-\tau_1)} \\ &= \varphi_{ry}(\tau_1) \quad \tau_1 \geq 0 \end{aligned} \quad (1.5-44)$$

The right-hand member of this equation, defining the weighting function for the filter which passes a gaussian input component, is the cross correlation between the gaussian input component and the output of the nonlinearity. For a general nonlinearity, the output at time zero, $y(0)$, may depend not only on the current value of the input, but on certain properties of the past history of that input, or on certain derivatives of the input at time zero. The cross-correlation function in Eq. (1.5-44) is an average over all inputs to, and corresponding outputs from, the nonlinearity. The evaluation of this expectation requires the joint probability density function for all the random variables needed to define $y(0)$. Needless to say, this constitutes a formidable task even for dynamic nonlinearities of simple-appearing form.

Even if one is able to evaluate $\varphi_{ry}(\tau_1)$ in some cases, a substantial chore remains. The solution to the integral equation will not be obvious; the equation must be solved in the more general sense. This solution is not difficult if the transform of $\varphi_{ry}(\tau_1)$ can be taken and if the result is a rational function of the transform variable, or can be well approximated by a rational function. If so, the solution to the integral equation can be written down explicitly, since the equation is of the form of the Wiener-Hopf equation. The solution is derived in a number of books, including Refs. 4 and 7 to 9. If transform techniques cannot usefully be employed, the only practical alternative is likely to be numerical solution with computer help. The solution will be some general function for $w_R(\tau_2)$. Thus the optimum linear filter to approximate the effect of the general nonlinearity in passing a gaussian input component is not a static gain, but is indeed some dynamic linear filter, as one would surely expect.

Fortunately, this situation is simplified considerably in the very important case of a static single-valued nonlinearity. In this case the output of the nonlinearity depends only on the current value of the input, and the

right-hand member of Eq. (1.5-44) can be reduced to a convenient form if it is written out in detail.

$$\overline{y(0)r(-\tau_1)} = \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_n \int_{-\infty}^{\infty} dr_1 \int_{-\infty}^{\infty} dr_2 y[r_1 + x_r(0)]r_2 \\ \times f_n(p_1, \dots, p_n) f_2(r_1, 0; r_2, -\tau_1) \quad (1.5-45)$$

In this integral, the p_i are the statistical parameters involved in the remainder, $x_r(0)$, and $f_n(p_1, \dots, p_n)$ is the joint probability density function for these parameters. The remaining random variables are $r(0)$, whose general value is written as r_1 , and $r(-\tau_1)$, whose general value is written as r_2 . The joint probability density function for r_1 and r_2 is the bivariate normal distribution whose form is given in Eq. (H-70).

$$f_2(r_1, 0; r_2, -\tau_1) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2\sigma^2(1-\rho^2)} (r_1^2 - 2\rho r_1 r_2 + r_2^2) \right] \quad (1.5-46)$$

$$\text{where} \quad \rho = \frac{\varphi_{rr}(\tau_1)}{\sigma^2}$$

$$\text{and} \quad \sigma^2 = \varphi_{rr}(0)$$

Defining a new variable,

$$r_3 = \frac{r_2 - \rho r_1}{\sigma\sqrt{1-\rho^2}} \quad (1.5-47)$$

and writing Eq. (1.5-46) in terms of it,

$$f_2 = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2} \left(r_3^2 + \frac{r_1^2}{\sigma^2} \right) \right] \quad (1.5-48)$$

Now Eq. (1.5-45), with r_2 eliminated in favor of r_3 , is written

$$\overline{y(0)r(-\tau_1)} = \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_n \int_{-\infty}^{\infty} dr_1 \int_{-\infty}^{\infty} dr_3 y[r_1 + x_r(0)] f_n(p_1, \dots, p_n) \\ \times (\sigma\sqrt{1-\rho^2} r_3 + \rho r_1) \frac{1}{2\pi\sigma} \exp \left[-\frac{1}{2} \left(r_3^2 + \frac{r_1^2}{\sigma^2} \right) \right] \\ = \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_n \int_{-\infty}^{\infty} dr_1 y[r_1 + x_r(0)] f_n(p_1, \dots, p_n) \\ \times \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{r_1^2}{2\sigma^2} \right) \rho r_1 \\ = \frac{\varphi_{rr}(\tau_1)}{\sigma^2} \overline{y(0)r(0)} \quad (1.5-49)$$

The cross correlation between a gaussian input component and the output of a static single-valued nonlinearity is found to be proportional to the

autocorrelation of the gaussian signal. The preceding argument shows that this interesting property holds even if the input to the nonlinearity consists of the sum of the gaussian and other signals, provided that the other signals are not correlated with the gaussian. This is true only of an unbiased gaussian input, but that is the case of interest here under the convention of associating any bias in $x(t)$ with a single input component. For this case, the solution of Eq. (1.5-44) is obvious.

$$w_R(\tau_2) = \frac{1}{\sigma^2} \overline{y(0)r(0)} \delta(\tau_2) \quad (1.5-50)$$

So the optimum quasi-linear operator to pass a gaussian input component in the presence of any other uncorrelated components is just a static gain in the case of a static single-valued nonlinearity. This is a result which could not have been anticipated with certainty. This gain, written in describing function notation, is

$$N_R = \frac{1}{\sigma^2} \overline{y(0)r(0)} \quad (1.5-51)$$

These are the describing functions, based on minimum mean-squared approximation error, for the three signal forms which might be expected to appear at the input to the nonlinearity in a feedback system. These signals are chosen because they may be assumed on the basis of the filtering properties of the linear part of the system. It is possible to calculate the optimum quasi-linear approximator to operate on other signal forms as well, and this will be done in Chap. 8 following a different motivation. The point is that for a feedback system, if the forms of the signals at the nonlinearity input are not *assumed*, they must be *calculated*. And this is indeed a forbidding task in most cases.

MULTIPLE INPUTS OF THE SAME FORM

A property of these describing functions which will serve to define the most general form of nonlinearity input that need be considered should be noted here. The input has been taken to be any linear combination of independent signals of the following forms: bias, sinusoid, and gaussian random process. This includes the possibility of any number of signals of the same form. The potential complexity of this situation is considerably reduced if we note, first of all, that there is no reason to consider more than one bias component. Even if the constant signal at the input to the nonlinearity should arise from several different sources, there would be no way to distinguish them, and the sum of all the constant input components would be propagated by the nonlinearity as a single bias. Less obvious than this is the fact that gaussian signals enjoy the same property in the case of static single-valued nonlinearities. If the input to the nonlinearity includes several gaussian

components, say, $r_1(t), r_2(t), \dots, r_n(t)$, the describing function for the i th signal is given by Eq. (1.5-51) to be

$$N_{R_i} = \frac{1}{\sigma_i^2} \overline{y(0)r_i(0)} \quad (1.5-52)$$

which would appear to be different for each i . However, one can show by direct calculation that

$$\frac{1}{\sigma_i^2} \overline{y(0)r_i(0)} = \frac{1}{\sigma_R^2} \overline{y(0)R(0)} \quad (1.5-53)$$

where

$$R(t) = r_1(t) + r_2(t) + \dots + r_n(t)$$

and

$$\sigma_R^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$$

So the nonlinearity propagates each of the gaussian components with the same gain—that gain which corresponds to a single gaussian input component consisting of the sum of the individual gaussian contributors. This result may be interpreted in terms of the fact that the sum of a number of independent gaussian signals is a gaussian signal with variance equal to the sum of the variances of the contributors. Once summed, these contributors cannot be distinguished on the basis of the distribution of amplitudes of the total signal. They may still be distinguishable on the basis of their harmonic content, and the linear part of the system can be designed to separate a gaussian signal from a gaussian noise if their power spectral densities are distinct, but the static nonlinearity is insensitive to this property.

The remaining signal form, the sinusoid, does not have this additive property. Thus the most general input for which the describing functions need be calculated is the sum of a single bias, a single gaussian random process, and an arbitrary number of sinusoids.

THE DESCRIBING FUNCTIONS FOR SMALL SIGNALS

The describing functions can now be written in more specific form for the most general input of interest: the sum of a bias, a gaussian signal, and an arbitrary number of sinusoids.

$$x(t) = B + r(t) + A_1 \sin(\omega_1 t + \theta_1) + \dots + A_n \sin(\omega_n t + \theta_n) \quad (1.5-54)$$

These input components are considered statistically independent. Consider, for the purpose of this section, that the nonlinearity is static and single-valued, so that $y(0)$ depends only on $x(0)$. The random variables required to determine $x(0)$ are $r(0)$, which will be denoted simply as r , and the n phase angles θ_i . Being independent, these random variables have a joint probability density function which is the product of their individual density functions. The variable r has the normal distribution with zero mean and standard deviation σ , and each of the θ_i is uniformly distributed over

$(0, 2\pi)$. Thus the describing functions for the bias component, one of the sinusoidal components—say, the first one—and the gaussian component are written, according to Eqs. (1.5-27), (1.5-36), and (1.5-51), as

$$N_B = \frac{1}{B} \int_{-\infty}^{\infty} dr \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_n y(B + r + A_1 \sin \theta_1 + \cdots + A_n \sin \theta_n) \\ \times \frac{1}{(2\pi)^{n+\frac{1}{2}}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (1.5-55)$$

$$N_{A_1} = \frac{2}{A_1} \int_{-\infty}^{\infty} dr \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_n y(B + r + A_1 \sin \theta_1 + \cdots + A_n \sin \theta_n) \\ \times \sin \theta_1 \frac{1}{(2\pi)^{n+\frac{1}{2}}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (1.5-56)$$

$$N_R = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} dr \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_n y(B + r + A_1 \sin \theta_1 + \cdots + A_n \sin \theta_n) \\ \times r \frac{1}{(2\pi)^{n+\frac{1}{2}}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (1.5-57)$$

Given a specific form for the nonlinearity $y(x)$, these integrals can be evaluated to determine the general describing functions for a static single-valued nonlinearity.

For some purposes, an alternative form for these expressions is more useful. Making the change of variable,

$$x = B + r + A_1 \sin \theta_1 + \cdots + A_n \sin \theta_n \quad (1.5-58)$$

and eliminating r from the describing function expressions yields

$$N_B = \frac{1}{B} \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_n y(x) \frac{1}{(2\pi)^{n+\frac{1}{2}}\sigma} \\ \times \exp\left[-\frac{1}{2\sigma^2}(x - B - A_1 \sin \theta_1 - \cdots - A_n \sin \theta_n)^2\right] \quad (1.5-59)$$

$$N_{A_1} = \frac{2}{A_1} \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_n y(x) \sin \theta_1 \frac{1}{(2\pi)^{n+\frac{1}{2}}\sigma} \\ \times \exp\left[-\frac{1}{2\sigma^2}(x - B - A_1 \sin \theta_1 - \cdots - A_n \sin \theta_n)^2\right] \quad (1.5-60)$$

$$N_R = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_n y(x)(x - B - A_1 \sin \theta_1 - \cdots - A_n \sin \theta_n) \\ \times \frac{1}{(2\pi)^{n+\frac{1}{2}}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x - B - A_1 \sin \theta_1 - \cdots - A_n \sin \theta_n)^2\right] \quad (1.5-61)$$

With this form for the describing functions it is easy to determine the limiting values as the corresponding input components go to zero. If the nonlinearity is asymmetric, the mean output may remain nonzero as the mean input B goes to zero. In that case, N_B goes to infinity as B goes to zero, and it is preferable to work directly with the average output rather than with the effective gain of the nonlinearity to the average input. That average output is BN_B , with N_B given either by Eq. (1.5-55) or Eq. (1.5-59). But if the nonlinearity is odd, so that

$$y(-x) = -y(x) \quad (1.5-62)$$

then the average output of the nonlinearity goes to zero as the average input B does. In that case N_B remains finite as B goes to zero, and Eqs. (1.5-55) and (1.5-59) are indeterminate forms at $B = 0$. This indeterminate form can be evaluated for Eq. (1.5-59). Treat the integral as the numerator and B , from the $1/B$ coefficient, as the denominator. Differentiate both numerator and denominator with respect to B as required by L'Hospital's rule, and form the ratio of these derivatives. The result is exactly the right-hand member of Eq. (1.5-61), the describing function for the gaussian input component.

$$\lim_{B \rightarrow 0} N_B = N_R(B = 0) \quad (1.5-63)$$

Similarly, Eq. (1.5-60) is indeterminate for $A_1 = 0$. Differentiation of numerator and denominator with respect to A_1 gives

$$\begin{aligned} \lim_{A_1 \rightarrow 0} N_{A_1} &= 2 \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_n y(x) \sin^2 \theta_1 \\ &\times \frac{1}{(2\pi)^{n+1} \sigma^2} (x - B - A_2 \sin \theta_2 - \cdots - A_n \sin \theta_n) \\ &\times \exp \left[-\frac{1}{2\sigma^2} (x - B - A_2 \sin \theta_2 - \cdots - A_n \sin \theta_n)^2 \right] \end{aligned} \quad (1.5-64)$$

The θ_1 integration can now be done, with the result

$$\begin{aligned} \lim_{A_1 \rightarrow 0} N_{A_1} &= \int_{-\infty}^{\infty} dx \int_0^{2\pi} d\theta_2 \cdots \int_0^{2\pi} d\theta_n y(x) \\ &\times \frac{2\pi}{(2\pi)^{n+1} \sigma^3} (x - B - A_2 \sin \theta_2 - \cdots - A_n \sin \theta_n) \\ &\times \exp \left[-\frac{1}{2\sigma^2} (x - B - A_2 \sin \theta_2 - \cdots - A_n \sin \theta_n)^2 \right] \end{aligned} \quad (1.5-65)$$

But notice that if A_1 is set equal to zero in the expression for N_R [Eq. (1.5-61)], the θ_1 integration can be done, and the result is exactly the right-hand member of Eq. (1.5-65). So

$$\lim_{A_1 \rightarrow 0} N_{A_1} = N_R(A_1 = 0) \quad (1.5-66)$$

This demonstrates the very interesting fact that the effective gain of an odd memoryless nonlinearity to a small *bias* component in the presence of any other uncorrelated input components is the same as the effective gain of the nonlinearity to a small *sinusoid* of any frequency in the presence of the same additional input components. And this common gain is the same as the effective gain of that nonlinearity to a *gaussian* input component, again in the presence of the same additional input components. This suggests that the operation of such a nonlinearity on a small signal in the presence of other uncorrelated signals is *independent of the form of the small signal*. This property may seem even more remarkable when one notes that no assumption has been made as to the analyticity or differentiability of the nonlinear characteristic.

1.6 ABOUT THE BOOK

The argument developed in the preceding section provides a unified approach to all describing function theory. For the sake of the added insight it may provide, the more physically motivated historical development of describing functions, which is applicable to sinusoid and constant inputs, is also given in the following chapters. The original single-sinusoid-input describing function is derived in Chap. 2, and applied to the study of steady-state and transient oscillations in Chaps. 3 and 4. Multiple-input describing functions involving sinusoid and bias components at the input to the nonlinearity are derived and applied in Chaps. 5 and 6. The general input, including gaussian processes, sinusoids, and bias components, is treated in Chap. 7. Quasi-linearization based on other forms of nonlinearity input which are applicable to the study of transients in nonlinear systems is the subject of Chap. 8, and in Chap. 9, describing function theory is applied to the study of sampled-data nonlinear systems. Information which the reader may wish to refer to repeatedly has generally been presented in the Appendixes, where it can more conveniently be located.

Examples illustrating various applications of the theory are of simple engineering systems. Other applications, such as the solution of nonlinear differential equations arising in any other context, are treated in the same way, once the mathematical model is formulated. In each case, the system is presented as a mathematical model at the outset. This does not imply

that the techniques can usefully be applied only to such simple "mathematical" systems and have little applicability to practical systems. *Every topic considered in this book is a practically useful topic.* It was noted earlier that the describing function method does not suffer as badly as most other analytic methods upon increasing the order of the system to which it is applied. Illustrations could be given by starting with a description of a physical system, deriving a mathematical model of the system, and applying describing function theory to the model. However, this would lengthen the presentation considerably and would serve no better to illustrate the subject of this book, describing function theory and usage. The modeling of a physical system for analytic study is fundamental to engineering analyses by any technique, and is not within the scope of this book.

REFERENCES

1. Atherton, D. P.: The Evaluation of the Response of Single-valued Non-linearities to Several Inputs, *Proc. IEE*, vol. C-109, no. 15 (March, 1962), pp. 146-157.
2. Booton, R. C., Jr.: Nonlinear Control Systems with Random Inputs, *IRE Trans. Circuit Theory*, vol. CT-1, no. 1 (March, 1954), pp. 9-17.
3. Booton, R. C., Jr., M. V. Mathews, and W. W. Seifert: "Nonlinear Servomechanisms with Random Inputs," *M.I.T. DACL Rept. 70*, August, 1953.
4. Davenport, W. B., Jr., and W. L. Root: "Introduction to Random Signals and Noise," McGraw-Hill Book Company, New York, 1958.
5. Gusev, M. I.: Determining the Effect of Regular Signal Dynamics in the Statistical Linearization Method, *Automation and Remote Control*, vol. 21, no. 11 (May, 1961), pp. 1093-1100.
6. Jaffe, R. C.: Limit Cycles in Single and Multiple Loop On-Off Systems, Sc.D. thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Cambridge, Mass., June, 1963, pp. 215-236.
7. Laning, J. H., Jr., and R. H. Battin: "Random Processes in Automatic Control," McGraw-Hill Book Company, New York, 1956.
8. Lee, Y. W.: "Statistical Theory of Communication," John Wiley & Sons, Inc., New York, 1960.
9. Newton, G. C., L. A. Gould, and J. F. Kaiser: "Analytic Design of Linear Feedback Controls," John Wiley & Sons, Inc., New York, 1957.
10. Popov, E. P.: On Evaluating the Quality of Nonlinear Automatic Systems with Random Interference, *Automation and Remote Control*, vol. 20, no. 10 (June, 1960), pp. 1281-1287.
11. Popov, E. P.: "The Dynamics of Automatic Control Systems," Pergamon Press, New York, 1962, p. 550.
12. Rekasius, Z. V.: Lagrange Stability of Nonlinear Feedback Systems, *Joint Autom. Control Conf., Paper 2-3*, New York, 1962.
13. Somerville, M. J., and D. P. Atherton, Multi-gain Representation for a Single-valued Nonlinearity with Several Inputs and the Evaluation of Their Equivalent Gains by a Cursor Method, *Proc. IEE*, vol. C-105, no. 8 (July, 1958), pp. 537-549.
14. Vander Velde, W. E.: A Nonlinear Method of Automatic Control in the Presence of Random Noise, Sc.D. thesis, Massachusetts Institute of Technology, Department of Aeronautical Engineering, Cambridge, Mass., May, 1956.

PROBLEMS

- 1-1. The system of Fig. 1-1 has a single nonlinear part, N , and several linear parts, L_1 to L_4 . Rearrange this system into a single-loop configuration suitable for describing function analysis.

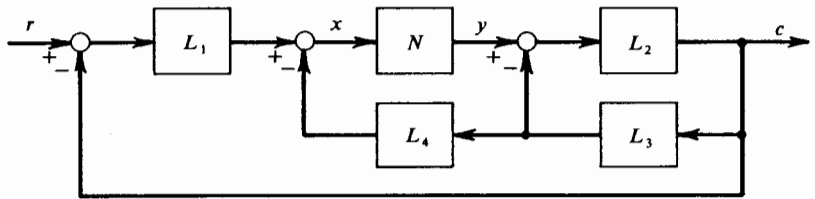


Figure 1-1

- 1-2. The differential equation

$$\frac{d^2y}{dt^2} + a \frac{dy}{dt} + f(y)y = 0$$

has a constant coefficient, a , and a nonlinear coefficient, $f(y)$. Draw the closed loop corresponding to this differential equation, and identify the linear and nonlinear parts.

- 1-3. Characterize Van der Pol's equation

$$\ddot{y} - \alpha(1 - \beta y^2)\dot{y} + y = x$$

in two different unity-feedback-system configurations, one of which has a single static nonlinear element. [Hint: Use the relationship $(d/dt)(y^3) = 3y^2\dot{y}$ to recast the original equation.]

- 1-4. Transform the rectangular hysteresis nonlinearity (see picture in Appendix B, nonlinearity 45) into a single-valued nonlinearity with a feedback path.
- 1-5. What would you expect the quasi-linearized gain of a relay with dead zone (see picture in Appendix B, nonlinearity 3) to look like as a function of input amplitude?
- 1-6. The theory of Sec. 1.5 demonstrates that the describing functions for a static single-valued nonlinearity are static gains. Take an approximator for the nonlinearity in the form of a parallel set of static gains, rather than arbitrary linear operators as in Fig. 1.5-1, and find the expressions for these gains which minimize the mean-squared approximation error. This leads easily to Eq. (1.5-27), n_r in Eq. (1.5-36), and Eq. (1.5-51).
- 1-7. The most general linear operator that need be considered to approximate the effect of a nonlinearity in passing a sinusoidal input component is the sum of a proportional plus derivative gain. This follows from the fact that the steady-state transfer of any linear filter at a prescribed frequency is just a complex gain which is equivalent to a proportional plus derivative path.
- Consider the input to a nonlinearity to be a sinusoid, $A \sin(\omega t + \theta)$, plus any uncorrelated remainder, $x_r(t)$. Use a proportional plus derivative operator to approximate the operation of the nonlinearity on the sinusoid, and derive expressions for the proportional and derivative gains which minimize the mean-squared approximation error. Thus derive Eq. (1.5-35).
- 1-8. Carry out the reduction of the expression for $\overline{\delta e(t)^2}$ indicated in Eq. (1.5-11).

- 1-9. What condition, used in deriving the describing function for a sinusoidal input [Eq. (1.5-36)], is violated in the case of two sinusoidal input components having rationally related frequencies?
- 1-10. The imaginary part of the describing function of a static single-valued nonlinearity for a sinusoidal input component was shown to be zero. Thus, in this case, Eq. (1.5-36) reads

$$N_A = \frac{2}{A} \overline{y(\theta) \sin \theta}$$

If only a sinusoid is present in the nonlinearity input, this is expressed

$$N_A = \frac{2}{A} \overline{y(A \sin \theta) \sin \theta}$$

where the expectation symbol implies an integration over the distribution of θ , which is uniform over the interval $(0, 2\pi)$.

$$N_A = \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \theta) \sin \theta d\theta$$

Alternatively, one could first compute the probability density function for $x = \sin \theta$ and then take the expectation by integration over the distribution of x . Carry out this alternative calculation, and derive the equivalent relation

$$N_A = \frac{2}{\pi A} \int_{-1}^1 \frac{y(Ax)x}{\sqrt{1-x^2}} dx$$

- 1-11. The calculation which leads to Eq. (1.5-53) is quite tedious in the general case of an arbitrary number of gaussian input components. Consider the special case of a nonlinearity input consisting of the sum of a bias, a sinusoid, and two gaussian signals, and verify that relation in the case of the describing functions for the two gaussian components.
- 1-12. Verify the limiting forms of the describing functions for small inputs [Eqs. (1.5-63) and (1.5-66)] in the cases of the three-input describing functions for which analytic expressions are given in Sec. E-3.

2 SINUSOIDAL-INPUT DESCRIBING FUNCTION (DF)

2.0 INTRODUCTION

Of the various describing functions discussed throughout this text, the *sinusoidal-input describing function* is by far the most widely known and used. In the following discussion the abbreviation DF is reserved for reference to this describing function.

As the name implies, the DF is a linearization of a nonlinear element subjected to a sinusoidal input. In the next two chapters we see that different interpretations regarding the origin of this sinusoidal input allow the study of totally unrelated modes of behavior occurring in nonlinear systems. However, in each such study the representation for the nonlinearity is its DF.

According to the theory of optimum quasi-linearization developed in the preceding chapter, the describing function for a nonlinearity driven by an input consisting of just a single sinusoidal component is given by a

specialization of Eqs. (1.5-36). These expressions are repeated here for convenient reference.

$$N_A = n_p + jn_q \quad (2.0-1a)$$

$$n_p = \frac{2}{A} \overline{y(0) \sin \theta} \quad (2.0-1b)$$

$$n_q = \frac{2}{A} \overline{y(0) \cos \theta} \quad (2.0-1c)$$

The nonlinearity input is in this case

$$x(t) = A \sin(\omega t + \theta) \quad (2.0-2)$$

where the amplitude and frequency, A and ω , are deterministic quantities, and θ is the random phase angle which is uniformly distributed over 2π radians.

$y(0)$ in Eqs. (2.0-1) is the output of the nonlinearity at an arbitrary time called zero. The output of a nonlinearity is a function of its input, and in some instances the relation is quite complex. To emphasize the fact that $y(t)$ depends on $x(t)$ in some way, we shall indicate in the notation a dependence on the current value of $x(t)$ and its first derivative.

$$y(t) = y[x(t), \dot{x}(t)] \quad (2.0-3)$$

This is not a restriction on the form of the nonlinearity, however. Any functional relation is acceptable. It is required only that the steady-state output history corresponding to a sinusoidal input be defined. In view of Eqs. (2.0-2) and (2.0-3),

$$y(0) = y(A \sin \theta, A\omega \cos \theta) \quad (2.0-4)$$

In general, the expectations indicated in Eqs. (2.0-1) are over all random parameters required to define $y(0)$. For the single-sinusoidal-input case, the only random parameter in the problem is the phase angle θ . The expectation is then a single integral over the range of θ , which can be taken as $0, 2\pi$. The probability density function for this uniformly distributed variable is $1/2\pi$ in that interval. Equations (2.0-1) are thus specialized in this case to

$$n_p = \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \theta, A\omega \cos \theta) \sin \theta \, d\theta \quad (2.0-5a)$$

$$n_q = \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \theta, A\omega \cos \theta) \cos \theta \, d\theta \quad (2.0-5b)$$

Historically, the basis for the DF was established by Krylov's and Bogoliubov's continuation in the field of nonlinear mechanics of an earlier work by Van der Pol. We shall briefly develop and study this basis to see

how it relates to the DF of current usage, derived from a quite different point of view. Once having established the significant relationships involved, the remainder of this chapter is then devoted to DF calculations for a variety of nonlinear elements. Static and dynamic, memoryless and with memory, implicit and explicit nonlinearities are treated. A compilation of the DFs for all the nonlinearities treated, as well as many others, is presented in Appendix B.

2.1 ASYMPTOTIC METHODS FOR THE STUDY OF NONLINEAR OSCILLATIONS

Starting with the well-known and extensively developed theory of linear systems, particularly of linear oscillators, a first step in the exploration of nonlinear oscillatory behavior is to study periodic regimes of the system characterized by

$$\ddot{x} + \omega_0^2 x + \mu f(x, \dot{x}) = 0 \quad (2.1-1)$$

where μ is a small parameter. In the limit as $\mu \rightarrow 0$ this system reduces to the classical *linear* oscillator.

By *asymptotic* methods we shall mean *approximating* schemes for the solution of Eq. (2.1-1), which also result in the *exact* solution in the limit as $\mu \rightarrow 0$. A great variety of such methods have been developed. Our ultimate objective is to develop methods which are also valid in the limit of large μ , as in the case of discontinuous nonlinearities. Nevertheless, the most intuitive point of view is born in the study of only slightly nonlinear systems.

Since the intent of this section is primarily to place in perspective the DF approach to be thoroughly studied in this and the following two chapters, we discuss only two different asymptotic methods. The two approaches are conventionally called the *perturbation method* and the *method of slowly varying amplitude and phase*.

PERTURBATION METHOD

Another name by which the general technique is known is the *method of small parameters*. This method was formulated by Poisson (ca. 1830) and studied throughout the nineteenth century by a number of astronomers, including Gylden and Lindstedt, culminating finally in the extensive work of Poincaré. Although the discussion following is without proof of convergence of the solution, a rigorous mathematical justification due to Poincaré (Ref. 42) is available in Stoker (Ref. 49).

For the system of Eq. (2.1-1) assume a solution of the form

$$\begin{aligned} x(t) &= x_0(t) + \mu x_1(t) + \mu^2 x_2(t) + \cdots \\ &= \sum_{i=0}^{\infty} \mu^i x_i(t) \end{aligned} \quad (2.1-2)$$

where μ is the small parameter to which the nonlinear term is proportional. Substituting $x(t)$ from Eq. (2.1-2) into (2.1-1) gives

$$\sum_{i=0}^{\infty} \mu^i \ddot{x}_i(t) + \omega_0^2 \sum_{i=0}^{\infty} \mu^i x_i(t) = -\mu f \left[\sum_{i=0}^{\infty} \mu^i x_i(t), \sum_{i=0}^{\infty} \mu^i \dot{x}_i(t) \right] \quad (2.1-3)$$

Consider the right-hand member of Eq. (2.1-3) as expanded in a Taylor series about $x(t) = x_0(t)$. This, of course, requires that $f(x, \dot{x})$ be an analytic function. We now argue that, for $\mu \rightarrow 0$, the leading term $x_0(t)$ in Eq. (2.1-2) must approach the corresponding exact solution $x(t)$ and that the terms involving $x_i(t)$ for $i \geq 1$ must be of increasingly smaller importance. Thus the initial conditions are distributed according to

$$x_0(0) = x(0) \quad \dot{x}_0(0) = \dot{x}(0) \quad (2.1-4a)$$

$$\text{and} \quad x_i(0) = 0 \quad \dot{x}_i(0) = 0 \quad \text{for } i \geq 1 \quad (2.1-4b)$$

and Eq. (2.1-3) is solved *recursively*, starting with low-ordered μ . Collating terms of like coefficient, μ^i , gives

$$\begin{aligned} \mu^0: \quad \ddot{x}_0 + \omega_0^2 x_0 &= 0 \\ \mu^1: \quad \ddot{x}_1 + \omega_0^2 x_1 &= -f(x_0, \dot{x}_0) \\ \mu^2: \quad \ddot{x}_2 + \omega_0^2 x_2 &= -\frac{\partial f}{\partial x}(x_0, \dot{x}_0)x_1 - \frac{\partial f}{\partial \dot{x}}(x_0, \dot{x}_0)\dot{x}_1 \\ &\vdots \end{aligned} \quad (2.1-5)$$

In this manner the postulated series expansion of $x(t)$ is determined.

A difficulty arises in the solution of this set of equations which, in one form or another, baffled the earlier nineteenth-century astronomers. This is due to the appearance in the solutions they obtained of apparently unbounded functions, called *secular terms*. It is easy to appreciate this difficulty by noting that the first few terms in the series expansion of a perfectly well-behaved bounded function can convey the picture of unboundedness. For example,

$$\sin(\omega_0 + \mu)t = \sin \omega_0 t + \mu t \cos \omega_0 t - \mu^2 \frac{t^2}{2} \sin \omega_0 t + \cdots \quad (2.1-6)$$

Many attempts to remove secular terms were made. One simple approach was to assume a dependence of solution frequency upon solution amplitude, should the physical picture support this observation, by taking

$$\omega^2 = \omega_0^2 + \mu \Psi_1(A) + \mu^2 \Psi_2(A) + \cdots \quad (2.1-7)$$

where ω is the actual fundamental frequency of the oscillation, and so choosing $\Psi_1(A), \Psi_2(A), \dots$ as to remove the secular terms as they arise (Ref. 6). Inserting both the expression for ω_0^2 from Eq. (2.1-7) and the series for $x(t)$ from Eq. (2.1-2) into Eq. (2.1-1) and once again collating terms of like coefficient μ^i give the recursive set of equations

$$\begin{aligned}\mu^0: \quad \ddot{x}_0 + \omega^2 x_0 &= 0 \\ \mu^1: \quad \ddot{x}_1 + \omega^2 x_1 &= x_0 \Psi_1(A) - f(x_0, \dot{x}_0) \\ \mu^2: \quad \ddot{x}_2 + \omega^2 x_2 &= x_0 \Psi_2(A) + x_1 \Psi_1(A) - \frac{\partial f}{\partial x}(x_0, \dot{x}_0)x_1 - \frac{\partial f}{\partial \dot{x}}(x_0, \dot{x}_0)\dot{x}_1 \\ &\dots\end{aligned}\tag{2.1-8}$$

One advantage of the perturbation method is that successive approximations of higher order can be formed in a systematic manner. It is quite useful for the study of both nonoscillatory and oscillatory systems. For the first type, though, a disadvantage is the requirement of an analytic functional description of the nonlinearity; and for the latter type an additional disadvantage is that a special mode of solution is required to eliminate occurrence of the secular terms. In what follows we present another procedure designed rather specifically for nonlinear oscillatory systems, which does not entail either of these disadvantages.

METHOD OF SLOWLY VARYING AMPLITUDE AND PHASE

For small μ , it is argued, the solution of Eq. (2.1-1) is fundamentally oscillatory in nature. Recognizing this behavior, Van der Pol (1924) proposed a solution of the form

$$x(t) = a(t) \sin \omega_0 t + b(t) \cos \omega_0 t \tag{2.1-9}$$

where $a(t)$ and $b(t)$ are to be determined by the method of *variation of parameters*. The principal feature in this solution is use of the *method of averaging* to reduce the resulting variational equations to autonomous form. Krylov and Bogoliubov (1937) and Bogoliubov and Mitropolsky (1958) follow the same original argument, but use a solution of the form

$$x(t) = A(t) \sin [\omega_0 t + \theta(t)] \tag{2.1-10}$$

As this formulation is fully equivalent, albeit more convenient to work with, than Van der Pol's, we develop the method of slowly varying amplitude and phase after Krylov and Bogoliubov (Ref. 29).

Let us assume the functional time dependence of A and θ without explicitly indicating it by an appended argument in t , and for convenience introduce the parameter ψ , defined by

$$\psi = \omega_0 t + \theta \tag{2.1-11}$$

We seek a solution of the form

$$x = A \sin \psi \quad (2.1-12a)$$

$$\dot{x} = A\omega_0 \cos \psi \quad (2.1-12b)$$

which requires the constraint

$$\dot{A} \sin \psi + A\dot{\theta} \cos \psi = 0 \quad (2.1-13)$$

determined by differentiation of Eq. (2.1-10) and comparison of the result with Eq. (2.1-12b). Forming the second derivative $\ddot{x}$ by differentiating $\dot{x}$ in Eq. (2.1-12b) gives

$$\ddot{x} = \dot{A}\omega_0 \cos \psi - A\omega_0(\omega_0 + \dot{\theta}) \sin \psi \quad (2.1-14)$$

Substituting for x and $\ddot{x}$ in Eq. (2.1-1) and repeating Eq. (2.1-13) gives the variational set

$$\dot{A}\omega_0 \cos \psi - A\omega_0\dot{\theta} \sin \psi = -\mu f(A \sin \psi, A\omega_0 \cos \psi) \quad (2.1-15a)$$

$$\dot{A} \sin \psi + A\dot{\theta} \cos \psi = 0 \quad (2.1-15b)$$

Solving Eqs. (2.1-15) for $\dot{A}$ and $\dot{\theta}$ yields

$$\dot{A} = -\frac{1}{\omega_0} \mu f(A \sin \psi, A\omega_0 \cos \psi) \cos \psi \quad (2.1-16a)$$

$$\dot{\theta} = \frac{1}{A\omega_0} \mu f(A \sin \psi, A\omega_0 \cos \psi) \sin \psi \quad (2.1-16b)$$

At this point we are at an apparent impasse, since this nonautonomous set usually cannot be solved exactly. From the requirement that μ be small, however, it may be inferred that in general $\dot{A}$ and $\dot{\theta}$ are small, and hence that A and θ are slowly varying functions of time. We are thus led to consider approximate solutions $\underline{A}$ and $\underline{\theta}$ obtained by an *averaging* of $\dot{A}$ and $\dot{\theta}$ over one complete cycle of the oscillatory solution. Equations (2.1-16) are thereby replaced by the averaged set

$$\dot{A} \approx -\frac{1}{2\pi\omega_0} \int_0^{2\pi} \mu f(A \sin \psi, A\omega_0 \cos \psi) \cos \psi \, d\psi \quad (2.1-17a)$$

$$\dot{\theta} \approx \frac{1}{2\pi\omega_0} \int_0^{2\pi} \frac{1}{A} \mu f(A \sin \psi, A\omega_0 \cos \psi) \sin \psi \, d\psi \quad (2.1-17b)$$

In effecting each of these integrals, A is held constant. Integrating the right-hand member of Eq. (2.1-17a) yields a first-order differential equation of the *separable-variables* type in A and t . Call the solution of this equation $\underline{A}(t)$. Then, integrating the right-hand member of Eq. (2.1-17b), in which A is first brought out from under the integral sign, yields another first-order differential equation. Replacing A by $\underline{A}$, one solves this differential equation

to obtain $\theta(t)$. The first-approximation solution to Eq. (2.1-1) then takes the form of Eq. (2.1-10); i.e., it is $x(t) = A(t) \sin [\omega_0 t + \theta(t)]$.

The resulting first-order solution, so obtained by Krylov and Bogoliubov (*op. cit.*), is simpler in form than is the solution obtained by the perturbation method. However, higher-order approximations do not follow as simply. Methods of obtaining higher-order approximations are deferred to Chap. 3. It is to be noted that $f(x, \dot{x})$ need not be analytically expressed before computation of the integrals in Eqs. (2.1-17) can be effected. Both discontinuous and continuous nonlinear functions $f(x, \dot{x})$ can be treated with equal ease.

2.2 EQUIVALENT LINEARIZATION AND THE DF

Having just studied the method of slowly varying amplitude and phase, we are in a position to interpret the solution obtained in terms of an equivalent linear representation for the nonlinearity $\mu f(x, \dot{x})$. For continuity with what shall become the central part of this and the two following chapters, we rename the nonlinear element

$$y(x, \dot{x}) = \mu f(x, \dot{x}) \quad (2.2-1)$$

and proceed to the concept of harmonic linearization.

AN INTERPRETATION OF PREVIOUS RESULTS

According to Krylov and Bogoliubov, the solution to

$$\ddot{x} + \omega_0^2 x + y(x, \dot{x}) = 0 \quad (2.2-2)$$

can be sought in the form

$$x = A \sin \psi \quad (2.2-3)$$

where, as shown in the previous section,

$$\begin{aligned} \dot{A} &\approx -\frac{1}{2\pi\omega_0} \int_0^{2\pi} y(A \sin \psi, A\omega_0 \cos \psi) \cos \psi \, d\psi \\ \dot{\theta} &\approx \frac{1}{2\pi A\omega_0} \int_0^{2\pi} y(A \sin \psi, A\omega_0 \cos \psi) \sin \psi \, d\psi \end{aligned} \quad (2.2-4)$$

Let us attempt to identify these results in terms of the solution to an already well-known system, the linear second-order damped oscillator. For this system the differential equation of motion is

$$\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x = 0 \quad (2.2-5)$$

where ζ is the damping ratio, and ω_n is the natural frequency. The exact solution of this equation can be written in the form

$$\begin{aligned} x &= A_1 e^{-\zeta \omega_n t} \sin(\omega_n \sqrt{1 - \zeta^2} t + \theta_1) \\ &= A \sin(\omega t + \theta_1) \end{aligned} \quad (2.2-6)$$

where A_1 and θ_1 are determined by the initial conditions. In any event, the solution is a sinusoid with *amplitude* A , where

$$A = A_1 e^{-\zeta \omega_n t} \quad (2.2-7)$$

from which we immediately determine the relationship

$$\frac{\dot{A}}{A} = -\zeta \omega_n \quad (2.2-8)$$

The frequency of the exact solution, squared, is

$$\omega^2 = \omega_n^2 (1 - \zeta^2) \quad (2.2-9)$$

Using the interpretations for $\zeta \omega_n$ and ω_n^2 afforded by the above, we may rewrite (2.2-5) in the form

$$\ddot{x} - 2 \frac{\dot{A}}{A} \dot{x} + \left[\omega^2 + \left(\frac{\dot{A}}{A} \right)^2 \right] x = 0 \quad (2.2-10)$$

Note that, as yet, there is no approximation incorporated in this equation.

The gap in this argument can now be closed, since here we observe that, in accord with the results of Krylov and Bogoliubov [Eqs. (2.2-2) to (2.2-4)], the coefficients of Eq. (2.2-10) can be expressed in the form

$$\ddot{x} + \frac{n_q(A, \omega_0)}{\omega_0} \dot{x} + [\omega_0^2 + n_p(A, \omega_0)] x \approx 0 \quad \omega = \omega_0 + \dot{\theta} \quad (2.2-11)$$

where the following definitions have been made:

$$n_p(A, \omega_0) = \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \psi, A \omega_0 \cos \psi) \sin \psi d\psi \quad (2.2-12a)$$

$$n_q(A, \omega_0) = \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \psi, A \omega_0 \cos \psi) \cos \psi d\psi \quad (2.2-12b)$$

and only terms up to first order in the implicit small parameter μ have been retained, viz.,

$$\left(\frac{\dot{A}}{A} \right)^2 \approx \left(\frac{\dot{\theta}}{\omega_0} \right)^2 \approx 0 \quad (2.2-13)$$

Comparing Eqs. (2.2-2) and (2.2-11), and calling $s = d/dt$, it is apparent that the linearized version of the original nonlinear term $y(x, \dot{x})$, valid to second

order in the implicit small parameter, is

$$y(x, \dot{x}) \approx \left[n_p(A, \omega_0) + \frac{n_q(A, \omega_0)}{\omega_0} s \right] x \quad (2.2-14)$$

The above linearization is fundamental to all that follows, and was, in fact, the first derivation of the DF. What has, in effect, been done is the replacement of a small nonlinearity by a linear *proportional plus derivative* network, whose coefficients are *not* constant, but rather are functions of the *amplitude* and *frequency* of the system oscillation. This procedure is called *harmonic linearization*, particularly in the Russian literature.

It is significant to note here the suggestion by Krylov and Bogoliubov that this linearization (derived in terms of a second-order system) also be used for nonlinearities in systems of higher order than second. It should also be noted that this linearization is exactly that which results from the general theory of Chap. 1. In particular, Eqs. (2.2-12) are identical with Eqs. (2.0-5).

PHYSICALLY MOTIVATED DF DEFINITION

During the years 1947 to 1950, in at least five different countries a new point of view came into being. This new point of view was essentially a physically motivated harmonic linearization, which, as we shall see, yields results equivalent to those derived by Krylov and Bogoliubov. Kochenburger (Ref. 27) is generally credited with its introduction in the United States, while Tustin (Ref. 52) of England, Oppelt (Ref. 41) of Germany, Goldfarb (Ref. 15) of Russia, and Dutilh (Ref. 10) of France played similar roles in their respective countries. Since the time of appearance of the papers cited above, a veritable wealth of material has been published relating to the DF and modifications thereof.

It is DF philosophy simply to replace a system nonlinearity by a linear gain, chosen in a fashion which renders similar the responses of the nonlinearity and its approximation, in some sense, to the same sinusoidal input. This is an attempt to extend the transfer function concept of linear-system studies to the nonlinear-system problem. In this vein we observe that, if a nonlinearity $y(x, \dot{x})$ is excited by a sinusoidal input ($\psi = \omega t$),

$$x = A \sin \psi \quad (2.2-15)$$

then the output is expressible by the Fourier series expansion¹

$$y(A \sin \psi, A\omega \cos \psi) = \sum_{n=1}^{\infty} A_n(A, \omega) \sin [n\omega t + \varphi_n(A, \omega)] \quad (2.2-16)$$

¹ No output dc term is present since in this chapter we restrict our attention to odd nonlinearities. Other nonlinearities are treated in Chap. 6.

and the sinusoidal-input describing function (DF), denoted $N(A, \omega)$, is by definition

$$\begin{aligned} N(A, \omega) &= \frac{\text{phasor representation of output component at frequency } \omega}{\text{phasor representation of input component at frequency } \omega} \\ &= \frac{A_1(A, \omega)}{A} e^{j\phi_1(A, \omega)} \end{aligned} \quad (2.2-17)$$

In other words, *the DF is the complex fundamental-harmonic gain of a nonlinearity in the presence of a driving sinusoid.*

The concepts of transfer magnitude and phase changes are embodied in this definition, a direct carry-over from familiar linear theory. In fact, the DF for a nonlinear element is *analogous* to the transfer function for a linear element, reducing *identically* to this transfer function in the purely linear case. We must note, however, that whereas for a linear device the response to an infinite spectrum of sinusoids *completely* defines the response to any other form of excitation, for a nonlinear device the response to an infinite spectrum of sinusoids is *not* definitive about the response to other inputs. *Linear superposition is not valid.* Nevertheless, the DF does lead to simple analysis and synthesis techniques for nonlinear systems which are reminiscent of the familiar frequency response methods exploited so successfully in linear-system analysis and synthesis.

A pictorial view of the DF, including all higher harmonics which are excluded in the DF formulation (collectively called the *residual*), is shown in Fig. 2.2-1.

An equation for the DF in terms of $y(x, \dot{x})$ is easily obtained. Multiplying both sides of Eq. (2.2-16) by either $\sin \psi$ or $\cos \psi$, and integrating to

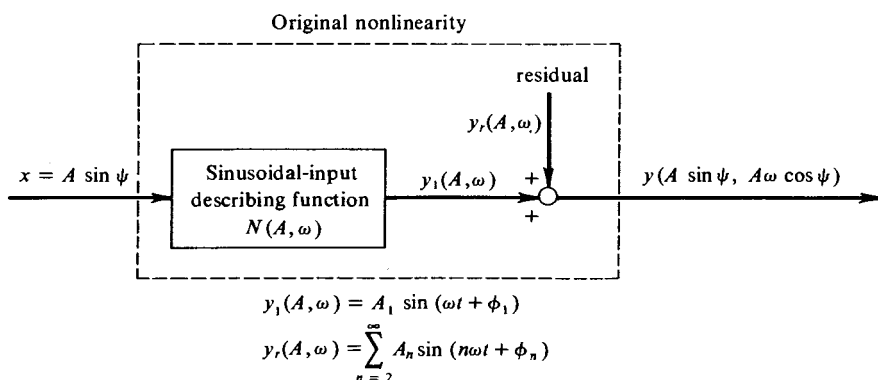


Figure 2.2-1 Definition of the DF.

determine the first Fourier coefficients, we find the relationships

$$\begin{aligned} A_1 \cos \varphi_1 &= \frac{1}{\pi} \int_0^{2\pi} y(A \sin \psi, A\omega \cos \psi) \sin \psi \, d\psi \\ A_1 \sin \varphi_1 &= \frac{1}{\pi} \int_0^{2\pi} y(A \sin \psi, A\omega \cos \psi) \cos \psi \, d\psi \end{aligned} \quad (2.2-18)$$

Now, multiplying the second of these equations by j , adding the two equations, and dividing both sides of the resultant equation by A , we get

$$\frac{A_1}{A} e^{j\varphi_1} = \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi, A\omega \cos \psi) e^{-j\psi} \, d\psi \quad (2.2-19)$$

where the relationship

$$e^{j\varphi} = \cos \varphi + j \sin \varphi$$

has been used. Comparing Eqs. (2.2-17) and (2.2-19), the equation for the DF in terms of the system nonlinearity becomes

$$N(A, \omega) = \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi, A\omega \cos \psi) e^{-j\psi} \, d\psi \quad (2.2-20)$$

Before attempting to tie in this result with the linearization of Krylov and Bogoliubov, we first observe the relation of this DF to mean-squared approximation error.

A PROPERTY OF THE DF

Historically, the DF as a linearizing complex gain was defined as opposed to having been derived. It was then noted that the equivalent linearization thus defined also minimized the mean-squared approximation error. This observation results from the following calculation.

Calling the magnitude and phase shift of the linearizing complex gain ρ_N and θ_N , respectively, we seek to minimize the following error measure:

$$\bar{e}^2 = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} e^2(t) \, dt \quad (2.2-21)$$

where

$$\begin{aligned} e &= y(x, \dot{x}) - y_{\text{approx}}(A, \omega) \\ &= y(A \sin \psi, A\omega \cos \psi) - A\rho_N \sin(\omega t + \theta_N) \end{aligned} \quad (2.2-22)$$

A stationary point of Eq. (2.2-21) exists when, simultaneously,

$$\frac{\partial \bar{e}^2}{\partial \rho_N} = \frac{\partial \bar{e}^2}{\partial \theta_N} = 0 \quad (2.2-23)$$

Differentiation with respect to ρ_N yields, after simplification,

$$\rho_N = \frac{\omega}{\pi A} \int_0^{2\pi/\omega} y(A \sin \psi, A\omega \cos \psi) \sin (\omega t + \theta_N) dt \quad (2.2-24)$$

Similarly, differentiation with respect to θ_N yields, after simplification,

$$\frac{\omega}{\pi A} \int_0^{2\pi/\omega} y(A \sin \psi, A\omega \cos \psi) \cos (\omega t + \theta_N) dt = 0 \quad (2.2-25)$$

Equations (2.2-24) and (2.2-25) can be combined in the form

$$\begin{aligned} \rho_N &= \frac{\omega}{\pi A} \int_0^{2\pi/\omega} y(A \sin \psi, A\omega \cos \psi) [\sin (\omega t + \theta_N) + j \cos (\omega t + \theta_N)] dt \\ &= j \frac{\omega}{\pi A} e^{-j\theta_N} \int_0^{2\pi/\omega} y(A \sin \psi, A\omega \cos \psi) e^{-j\omega t} dt \end{aligned} \quad (2.2-26)$$

which can in turn be rearranged to yield ($\psi = \omega t$)

$$\rho_N e^{j\theta_N} = \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi, A\omega \cos \psi) e^{-j\psi} d\psi \quad (2.2-27)$$

By comparison with Eq. (2.2-20) we conclude that the describing function defined as the first harmonic gain of a nonlinearity driven by a sinusoid also minimizes the mean-squared error in the linearizing approximation.

For the general theory of quasi-linear approximation developed in Chap. 1, minimum mean-squared approximation error was taken as the starting point. This defining condition permits the *derivation* of describing functions for nonlinearities driven by any form of input, whereas the first-harmonic-gain concept is applicable only to sinusoidal inputs. It was also shown there that the equivalent complex gain of a nonlinearity to a sinusoid input component in the presence of any other independent input components is the first harmonic gain of a pseudo-nonlinearity, defined by averaging over the effects of the other input components. If the input consists of a sinusoid only, this reduces to the first harmonic gain of the actual nonlinearity.

SUMMARY OF DF FORMS

The DF has been written both in the form of a dynamic proportional plus derivative element [Eq. (2.2-14)] and in the form of a static complex gain [Eq. (2.2-20)]. These seemingly different linearizations can be at once

reconciled by expanding the right-hand side of Eq. (2.2-20). Expansion yields

$$\begin{aligned} N(A, \omega) &= \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi, A \omega \cos \psi) e^{-j\psi} d\psi \\ &= \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi, A \omega \cos \psi) (\cos \psi - j \sin \psi) d\psi \\ &= n_p(A, \omega) + j n_q(A, \omega) \end{aligned} \quad (2.2-28)$$

where $n_p(A, \omega)$ and $n_q(A, \omega)$, defined earlier in Eqs. (2.2-12) and repeated here for convenience, are

$$n_p(A, \omega) = \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \psi, A \omega \cos \psi) \sin \psi d\psi \quad (2.2-29a)$$

$$n_q(A, \omega) = \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \psi, A \omega \cos \psi) \cos \psi d\psi \quad (2.2-29b)$$

Observing the complex-gain point of view, it follows from Eq. (2.2-28) that $n_p(A, \omega)$ and $n_q(A, \omega)$ are the *in-phase* and *quadrature* gains of the nonlinearity. That is, An_p is the in-phase (sin) component of the nonlinearity output, and An_q is the *quadrature* (cos) component of the nonlinearity output, referred to the input. Hence the mnemonic use of p and q .

Continuing, let us interpret the complex gain $N(A, \omega)$ in Eq. (2.2-28) as a linear operator which, acting upon a sinusoidal input $x = A \sin \omega t$, gives the appropriately phased sinusoidal output. Calling $s = d/dt$, this is equivalent to writing

$$N(A, \omega) = n_p(A, \omega) + \frac{n_q(A, \omega)}{\omega} s \quad (2.2-30)$$

which, of course, is precisely the result of Eq. (2.2-14).

Summarizing the above discussion, we have the following useful DF representations for a nonlinearity:

■ DF as a proportional plus derivative element

$$N(A, \omega) = n_p(A, \omega) + \frac{n_q(A, \omega)}{\omega} s \quad (2.2-31)$$

■ DF as a complex gain

$$N(A, \omega) = n_p(A, \omega) + j n_q(A, \omega) \quad (2.2-32)$$

or

$$N(A, \omega) = \rho_N(A, \omega) e^{j\theta_N(A, \omega)} \quad (2.2-33)$$

where the magnitude-phase angle representation is found according to

$$\begin{aligned}\rho_N(A, \omega) &= \sqrt{n_p^2(A, \omega) + n_q^2(A, \omega)} \\ \theta_N(A, \omega) &= \tan^{-1} \frac{n_q(A, \omega)}{n_p(A, \omega)}\end{aligned}\quad (2.2-34)$$

the corresponding set of inverse relationships being

$$\begin{aligned}n_p(A, \omega) &= \rho_N(A, \omega) \cos \theta_N(A, \omega) \\ n_q(A, \omega) &= \rho_N(A, \omega) \sin \theta_N(A, \omega)\end{aligned}\quad (2.2-35)$$

Throughout the development thus far we have assumed a nonlinearity output which, in general, is a function of the input as well as its derivative(s). Certain computational simplifications occur when the nonlinearity is static, and additional computational ease results when the static nonlinearity is odd. For example:

■ DF calculation for a general dynamic nonlinearity [$y = y(x, \dot{x})$]

$$N(A, \omega) = \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi, A\omega \cos \psi) e^{-j\psi} d\psi \quad (2.2-36)$$

■ DF calculation for a general static nonlinearity [$y = y(x)$]

$$N(A) = \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi) e^{-j\psi} d\psi \quad (2.2-37)$$

■ DF calculation for an odd static nonlinearity [$y(x) = -y(-x)$]

$$N(A) = \frac{2j}{\pi A} \int_0^{\pi} y(A \sin \psi) e^{-j\psi} d\psi \quad (2.2-38)$$

Single-valued characteristics are termed *memoryless*; multivalued characteristics are said to possess *memory*. For all memoryless nonlinearities, $n_q(A, \omega) = 0$. The ranges of integration can be further restricted. For example:

■ DF calculation for a general memoryless static nonlinearity

$$N(A) = \frac{2}{\pi A} \int_{-\pi/2}^{\pi/2} y(A \sin \psi) \sin \psi d\psi \quad (2.2-39)$$

■ DF calculation for an odd memoryless static nonlinearity

$$N(A) = \frac{4}{\pi A} \int_0^{\pi/2} y(A \sin \psi) \sin \psi d\psi \quad (2.2-40)$$

The remainder of this chapter is devoted to various aspects of the computation of DFs in the complex-gain representation. The questions of DF usage, DF accuracy, and DF validity are deferred to the following chapter.

2.3 DF CALCULATION FOR FREQUENCY-INDEPENDENT NONLINEARITIES

In this section we deal with nonlinearities whose input-output characteristics fall into the class defined by

$$y = y(x)$$

Such nonlinearities are *static*, displaying no dependence upon the input derivatives. At present, it is further required that the nonlinear characteristics possess *odd symmetry*. This requirement is expressed as

$$y(x) = -y(-x)$$

It is to be noted that DF determination for *asymmetric* nonlinearities is straightforward. DF application in such instances, however, is confounded by the presence of a nonlinearity output bias whose presence is not accounted for in the DF formulation. This difficulty may be overcome by artificially shifting the nonlinearity characteristic along its output axis to the point at which an input sinusoid results in an unbiased output. A less artificial approach to this problem is taken in Sec. 6.6, however, in which such nonlinearities are dealt with directly.

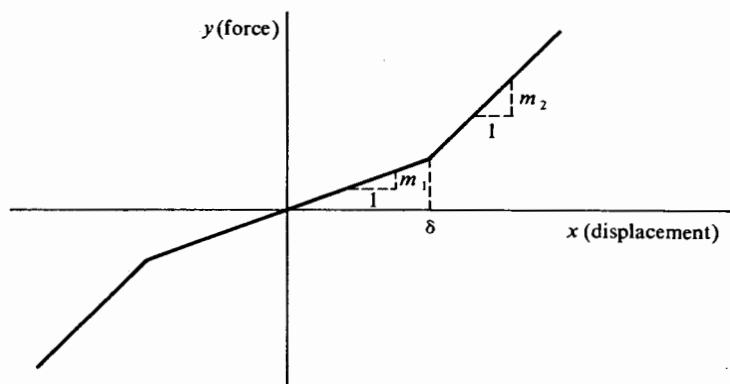
GAIN-CHANGING ELEMENT

The symmetrical gain-changing element with two slope discontinuities is shown in Fig. 2.3-1a. One common mechanical system giving rise to such a force-displacement characteristic is the dynamic vibration mount, for which a (massless) spring arrangement is depicted in Fig. 2.3-1b. In terms of the spring constants defined in the illustration, it is clear that $m_1 = 2k_1$ and $m_2 = 2(k_1 + k_2)$. For this nonlinearity the actual response to a driving sinusoid and the corresponding first harmonic are shown in Fig. 2.3-2. Since the response is not frequency-dependent, the independent variable of the input and output graphs has been chosen as the angle ψ (where $\psi = \omega t$). There are two regions of interest, those in which y assumes different dependencies upon x . Let us define the angle ψ_1 according to

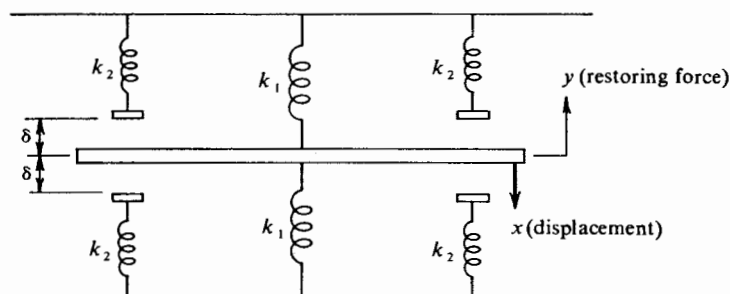
$$A \sin \psi_1 = \delta$$

or

$$\psi_1 = \sin^{-1} \frac{\delta}{A} \quad (2.3-1)$$



(a) Input-output characteristic



(b) Parallel translation spring assembly

Figure 2.3-1 Gain-changing nonlinear element.

Then the two regions of interest are defined in terms of ψ_1 as follows:

$$\begin{aligned} 0 < \psi \leq \psi_1 & \quad y = m_1 x \\ \psi_1 < \psi \leq \frac{\pi}{2} & \quad y = (m_1 - m_2)\delta + m_2 x \end{aligned} \quad (2.3-2)$$

We may proceed directly to the calculation of $N(A)$ after Eq. (2.2-40).

$$\begin{aligned} N(A) &= \frac{4}{\pi A} \int_0^{\pi/2} y(A \sin \psi) \sin \psi \, d\psi \\ &= \frac{4}{\pi A} \int_0^{\psi_1} m_1 A \sin^2 \psi \, d\psi + \frac{4}{\pi A} \int_{\psi_1}^{\pi/2} [(m_1 - m_2)\delta + m_2 A \sin \psi] \sin \psi \, d\psi \\ &= \frac{2(m_1 - m_2)}{\pi} \left(\psi_1 - \frac{1}{2} \sin 2\psi_1 + 2 \frac{\delta}{A} \cos \psi_1 \right) + m_2 \end{aligned} \quad (2.3-3)$$

Equations (2.3-1) and (2.3-3) define the DF for the gain-changing nonlinearity. These may be combined, with the result

$$N(A) = \frac{2(m_1 - m_2)}{\pi} \left[\sin^{-1} \frac{\delta}{A} + \frac{\delta}{A} \sqrt{1 - \left(\frac{\delta}{A} \right)^2} \right] + m_2 \quad (2.3-4)$$

Note that this result is valid only for $A \geq \delta$. For $A < \delta$, $N(A) = m_1$, which is the linear gain.

It is worthwhile to observe that in deriving the DF for this gain-changing nonlinearity we have simultaneously accomplished the DF derivation for some simpler nonlinearities. For example, as can be seen from Fig. 2.3-1a, setting $m_1 = 0$ results in the dead-zone characteristic; setting $m_2 = 0$ results in the saturation characteristic; setting $m_2 = 0$ and simultaneously $m_1 \delta = D$, $m_1 \rightarrow \infty$, results in the ideal-relay characteristic; and, of course, setting $m_1 = m_2$ results in the linear-gain characteristic. Figure 2.3-3 illustrates each of the above-mentioned nonlinearities, together with the appropriate DFs as derived from Eq. (2.3-4).

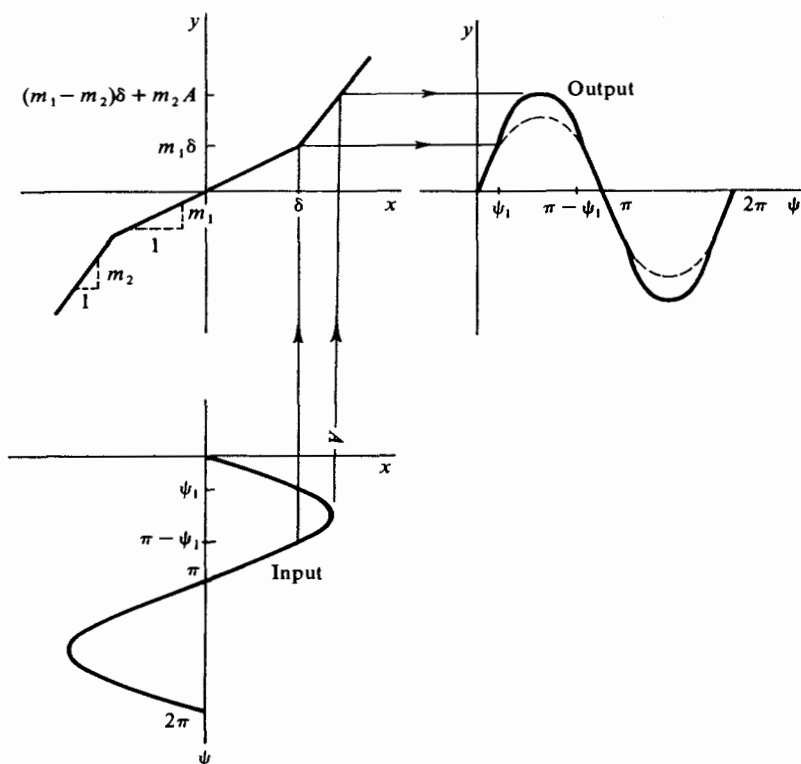
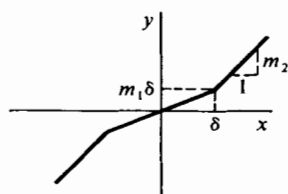


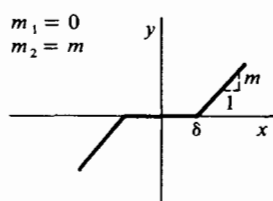
Figure 2.3-2 Input-output relationship for a gain-changing nonlinearity.



$$N(A) = m_1 \quad A \leq \delta$$

$$= \frac{2(m_1 - m_2)}{\pi} \left[\sin^{-1} \left(\frac{\delta}{A} \right) + \left(\frac{\delta}{A} \right) \sqrt{1 - \left(\frac{\delta}{A} \right)^2} \right] + m_2 \quad A > \delta$$

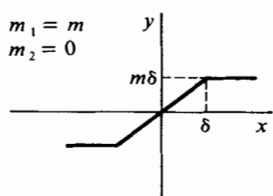
(a) Gain-changing element



$$N(A) = 0 \quad A \leq \delta$$

$$= \frac{m}{\pi} \left[\pi - 2 \sin^{-1} \left(\frac{\delta}{A} \right) - 2 \left(\frac{\delta}{A} \right) \sqrt{1 - \left(\frac{\delta}{A} \right)^2} \right] \quad A > \delta$$

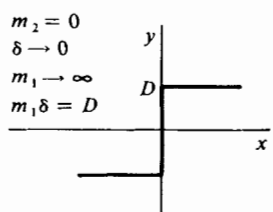
(b) Dead zone



$$N(A) = m \quad A \leq \delta$$

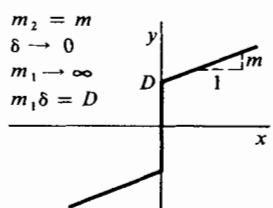
$$= \frac{2m}{\pi} \left[\sin^{-1} \left(\frac{\delta}{A} \right) + \left(\frac{\delta}{A} \right) \sqrt{1 - \left(\frac{\delta}{A} \right)^2} \right] \quad A > \delta$$

(c) Saturation



$$N(A) = \frac{4m_1\delta}{\pi A} = \frac{4D}{\pi A}$$

(d) Ideal relay



$$N(A) = \frac{4D}{\pi A} + m$$

(e) Preload

Figure 2.3-3 DFs for the gain-changing element and related nonlinearities.

Examination of the DFs for Fig. 2.3-3a to c reveals that a single functional form is recurrent. For this form we shall coin the name *saturation function* and the designation $f(\delta/A)$, where

$$f\left(\frac{\delta}{A}\right) = \begin{cases} \frac{2}{\pi} \left[\sin^{-1} \frac{\delta}{A} + \frac{\delta}{A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} \right] & \text{for } \frac{\delta}{A} \leq 1 \\ 1 & \text{for } \frac{\delta}{A} > 1 \end{cases} \quad (2.3-5)$$

A plot of $f(\delta/A)$ appears in Fig. 2.3-4. The ordinate scale is linear (rather than logarithmic, which form we shall find quite convenient in actual DF use) in order that the algebraic operations of addition and subtraction may be easily performed during DF calculation. Appendix A is an amplitude-ratio-decibel conversion table with which logarithmic DF representation

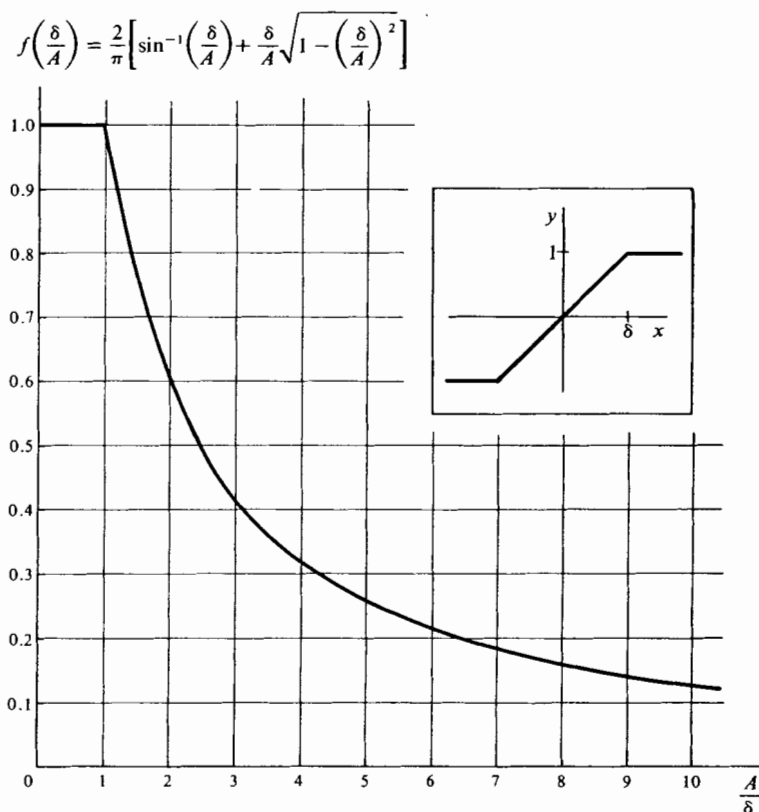


Figure 2.3-4 Saturation function.

may be accomplished conveniently, wherever desired. In terms of the saturation function $f(\delta/A)$, we have¹

Gain-changing element:

$$N(A) = (m_1 - m_2)f\left(\frac{\delta}{A}\right) + m_2 \quad (2.3-6)$$

Dead zone:

$$N(A) = m \left[1 - f\left(\frac{\delta}{A}\right) \right] \quad (2.3-7)$$

Saturation:

$$N(A) = m f\left(\frac{\delta}{A}\right) \quad (2.3-8)$$

Observe that Fig. 2.3-4 is a normalized plot of the saturation DF; hence the name saturation function.

GENERAL PIECEWISE-LINEAR ODD MEMORYLESS NONLINEARITY

We are now in a position to consider a fairly general case, namely, one where the nonlinear characteristic is any piecewise-linear odd memoryless and frequency-independent function.² Figure 2.3-5 illustrates a typical four-segment characteristic (first quadrant only, eight segments in all). All δ_i 's

¹ This form of DF presentation was motivated by the work of Magnus (Ref. 36), who used a slightly less compact notation. See Graham and McRuer (Ref. 16) for details of Magnus' presentation.

² A similar development is given by Gille et al. (Ref. 14).

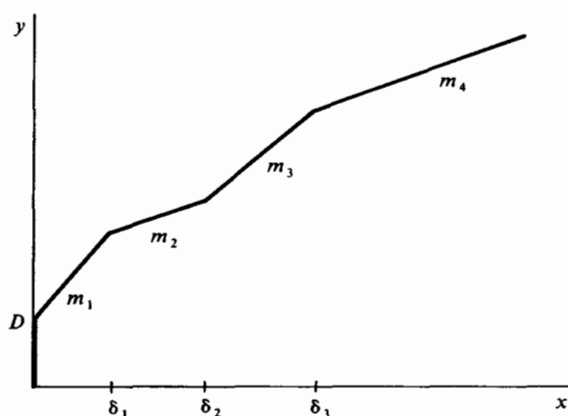


Figure 2.3-5 General piecewise-linear odd memoryless nonlinearity.

indicate abscissa breakpoints, and m_i indicates the slope occurring in the range $\delta_{i-1} < x \leq \delta_i$. The nonlinearity output y in terms of its input x can be written in the particular form

$$\begin{aligned} 0 < x \leq \delta_1 & \quad y = m_1 x + D \\ \delta_1 < x \leq \delta_2 & \quad y = (m_1 - m_2)\delta_1 + m_2 x + D \\ \delta_2 < x \leq \delta_3 & \quad y = (m_1 - m_2)\delta_1 + (m_2 - m_3)\delta_2 + m_3 x + D \\ \delta_3 < x & \quad y = (m_1 - m_2)\delta_1 + (m_2 - m_3)\delta_2 + (m_3 - m_4)\delta_3 \\ & \quad \quad \quad + m_4 x + D \end{aligned} \quad (2.3-9)$$

As before, the input is taken as a sinusoid

$$x = A \sin \psi$$

and we may directly compute the DF for various ranges of A . In the first range ($0 < A \leq \delta_1$), behavior is like that of a preload nonlinearity. The DF in this range is therefore as given in Fig. 2.3-3e. Continuing in other ranges, we note that the DF can very conveniently be described in terms of the saturation function, $f(\delta/A)$. It is readily verified that

$$\begin{aligned} N(A) = \frac{4D}{\pi A} + (m_1 - m_2)f\left(\frac{\delta_1}{A}\right) + (m_2 - m_3)f\left(\frac{\delta_2}{A}\right) \\ + (m_3 - m_4)f\left(\frac{\delta_3}{A}\right) + m_4 \end{aligned} \quad (2.3-10)$$

The reader may show that $N(A)$ is appropriately defined over *all* A . Clearly, this sequence can be directly extended by induction to an n -segment nonlinearity. Moreover, the actual magnitude of the composite nonlinearity DF can be obtained easily by repeated use of Fig. 2.3-4.

Example 2.3-1 Derive the DF for the piecewise-linear nonlinearity containing dead zone, a linear band, and saturation (Fig. 2.3-6).

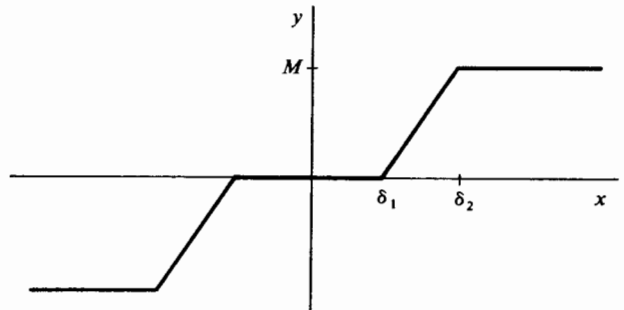


Figure 2.3-6 Nonlinear characteristic with dead zone, a linear band, and saturation.

By observing that the more general nonlinearity of Fig. 2.3-5 is reducible to the specific case at hand under the value assignments

$$D = 0, \quad m_2 = \frac{M}{\delta_2 - \delta_1}, \quad m_1 = m_3 = m_4 = 0$$

we obtain directly from Eq. (2.3-10)

$$N(A) = \frac{M}{\delta_2 - \delta_1} \left[f\left(\frac{\delta_2}{A}\right) - f\left(\frac{\delta_1}{A}\right) \right] \quad (2.3-11)$$

which takes on the following special values in the ranges indicated:

$$0 < A \leq \delta_1 \quad N(A) = 0$$

$$\delta_1 < A \leq \delta_2 \quad N(A) = \frac{M}{\delta_2 - \delta_1} \left[1 - f\left(\frac{\delta_1}{A}\right) \right]$$

$$\delta_2 < A \quad N(A) = \frac{M}{\delta_2 - \delta_1} \left[f\left(\frac{\delta_2}{A}\right) - f\left(\frac{\delta_1}{A}\right) \right]$$

QUANTIZER

A very useful DF in the study of systems containing a linear analog-digital converter is that for the staircase, or quantizer, nonlinearity, shown in Fig. 2.3-7. For all continuous inputs x , the output can assume only the discrete values $0, \pm D, \pm 2D$, etc.; hence the name suggesting quantum jumps.

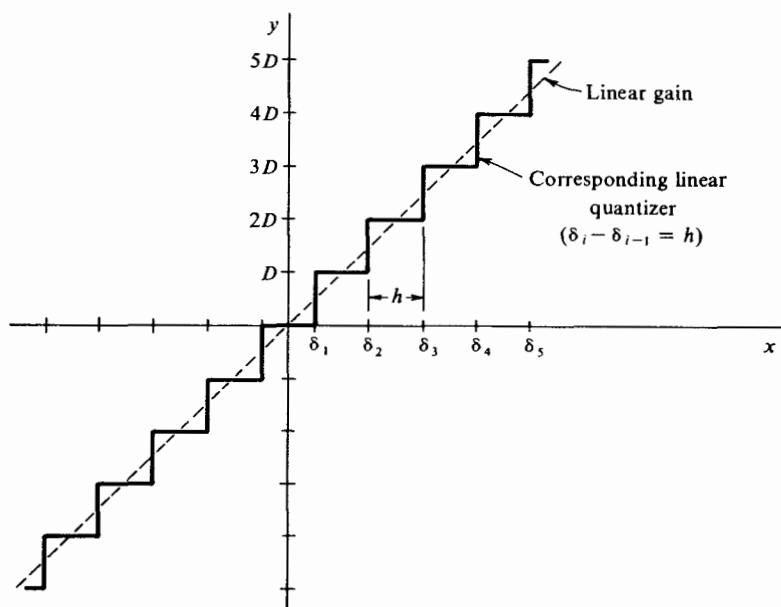


Figure 2.3-7 Linear quantizer characteristic.

nonlinearity form may be well related to a polynomial fit, either by sinusoidal testing or growing harmonic exponential testing (Ref. 35).

The input-output characteristic for an *odd* n th-order polynomial-type nonlinearity is of the form (n is a positive odd integer in this equation)

$$y = c_n x^n + c_{n-1} x^{n-2} |x| + \cdots + c_2 x |x| + c_1 x \quad (2.3-15)$$

The general term of Eq. (2.3-15) can be rewritten as (arbitrary n)

$$y = \begin{cases} c_n x^n & x \geq 0 \\ -c_n |x|^n & x < 0 \end{cases} \quad (2.3-16)$$

which is an odd function of x . The DF for the general odd polynomial term is therefore computed by integrating over the interval $0 < \psi < \pi/2$, as follows:

$$\begin{aligned} N(A) &= \frac{4}{\pi A} \int_0^{\pi/2} y(A \sin \psi) \sin \psi \, d\psi \\ &= \frac{4}{\pi A} \int_0^{\pi/2} c_n (A \sin \psi)^n \sin \psi \, d\psi \\ &= \frac{4}{\pi} c_n A^{n-1} \int_0^{\pi/2} \sin^{n+1} \psi \, d\psi \\ &= \frac{2c_n}{\sqrt{\pi}} A^{n-1} \frac{\Gamma[(n+2)/2]}{\Gamma[(n+3)/2]} \quad n > -2 \end{aligned} \quad (2.3-17)$$

where $\Gamma(\arg)$ is the gamma function of the indicated argument.¹ In the event that n is a positive odd integer, $N(A)$ can be rewritten as

$$N(A) = c_n \frac{n(n-2)(n-4) \cdots (3)}{(n+1)(n-1)(n-3) \cdots (4)} A^{n-1} \quad n \text{ is an odd integer} > 1 \quad (2.3-18)$$

If n is a positive even integer, we obtain the following expansion for $N(A)$:

$$N(A) = c_n \frac{4}{\pi} \frac{n(n-2)(n-4) \cdots (2)}{(n+1)(n-1)(n-3) \cdots (3)} A^{n-1} \quad n \text{ is an even integer} > 0 \quad (2.3-19)$$

¹ Those properties of the gamma function required for use in Eq. (2.3-17) are

$$\Gamma(k+1) = k! \quad \text{for integers } k \geq 0$$

$$\Gamma(k+1) = k\Gamma(k) \quad \text{for arbitrary } k > -1$$

and

$$\Gamma(\frac{1}{2}) = \sqrt{\pi} \quad \Gamma(1) = \Gamma(2) = 1$$

These results can be combined to form the DF for the particular odd polynomial nonlinearity described by

$$y = c_1x + c_2x|x| + c_3x^3 + c_4x^3|x| + c_5x^5 + \cdots \quad (2.3-20)$$

which is

$$N(A) = c_1 + \frac{8}{3\pi} c_2A + \frac{3}{4} c_3A^2 + \frac{32}{15\pi} c_4A^3 + \frac{5}{8} c_5A^4 + \cdots \quad (2.3-21)$$

A plot of the DF for a one-term odd polynomial nonlinearity appears in Appendix B. For values of n greater than unity, the nonlinear characteristic is of increasing slope with increasing input (*hard*). For values of n less than unity, the characteristic is of the saturating variety (*soft*). This accounts for the behavior of DF magnitude as a function of A . As n approaches zero, the odd polynomial characteristic approaches the ideal-relay characteristic; this bound is shown for comparison.

HARMONIC NONLINEARITY

Error-detecting synchros commonly used in ac servomechanisms are capable of continuous rotation as required to follow constant-angular-velocity command inputs. The gain characteristic of a synchro pair is a sinusoidal function of angular position following error. Thus, for the synchro pair, we may write

$$y = M \sin mx \quad (2.3-22)$$

where y is the ac output amplitude, and x is the input angular-position error. The DF for this harmonic nonlinearity is given by

$$\begin{aligned} N(A) &= \frac{4}{\pi A} \int_0^{\pi/2} y(A \sin \psi) \sin \psi \, d\psi \\ &= \frac{4M}{\pi A} \int_0^{\pi/2} \sin (mA \sin \psi) \sin \psi \, d\psi \\ &= \frac{2MJ_1(mA)}{A} \end{aligned} \quad (2.3-23)$$

where $J_1(mA)$ is the Bessel function of order one for real arguments. A plot of the DF for this nonlinearity in Appendix B indicates regions of -180° phase shift, as well as regions of 0° phase shift. Boundaries of each region are given by zero crossings of the Bessel function, $J_1(mA)$.

HYSTERESIS

Electromagnetic current-actuated relays generally have different pull-in and drop-out input-current values. The nonlinear input-output characteristics

for such relays are, as a consequence, multivalued. A typical characteristic is given in Fig. 2.3-8.

To compute the DF for this characteristic we employ the complex exponential form as follows:

$$\begin{aligned}
 N(A) &= \frac{2j}{\pi A} \int_0^\pi y(A \sin \psi) e^{-j\psi} d\psi \\
 &= \frac{2j}{\pi A} \left[\int_0^{\psi_1} (-D) e^{-j\psi} d\psi + \int_{\psi_1}^{\psi_2} 0 d\psi + \int_{\psi_2}^\pi D e^{-j\psi} d\psi \right] \\
 &= \frac{2D}{\pi A} (e^{-j\psi_1} + e^{-j\psi_2}) \quad (2.3-24)
 \end{aligned}$$

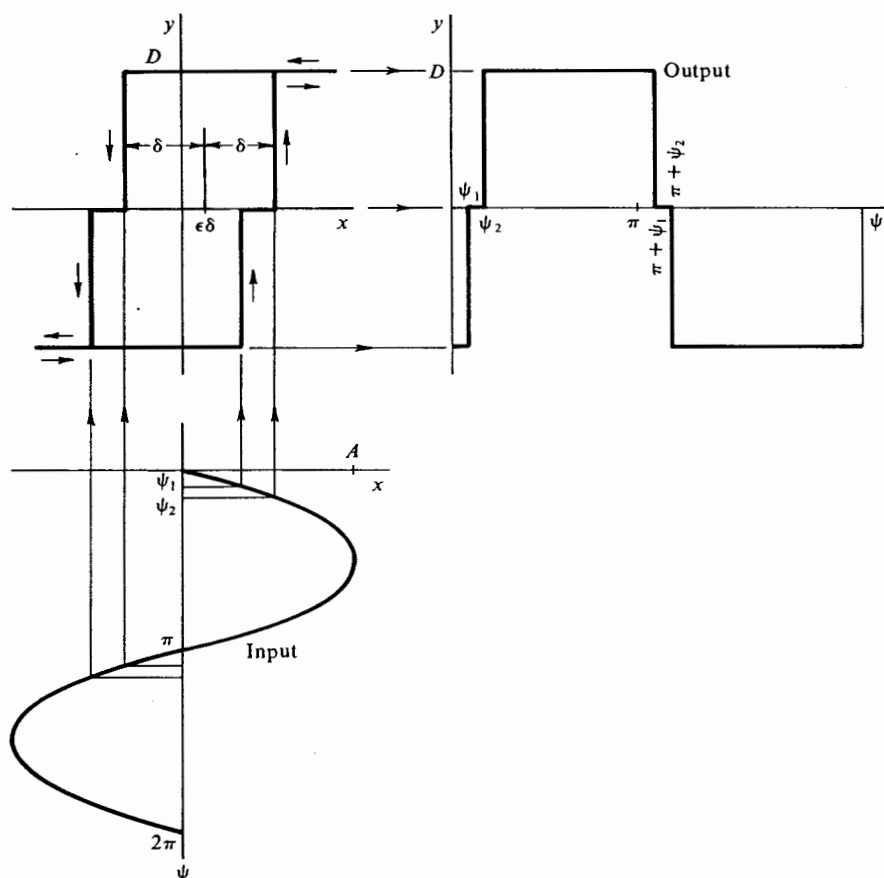


Figure 2.3-8 Hysteresis characteristic with input and output waveforms.

$$\begin{aligned} \text{where} \quad A \sin \psi_1 &= \delta(1 - \epsilon) \quad \text{or} \quad \psi_1 = \sin^{-1} \frac{\delta}{A} (1 - \epsilon) \\ A \sin \psi_2 &= \delta(1 + \epsilon) \quad \text{or} \quad \psi_2 = \sin^{-1} \frac{\delta}{A} (1 + \epsilon) \end{aligned}$$

The DF can be rewritten in terms of the hysteresis characteristic parameters as

$$N(A) = \frac{2D}{\pi A} \left[\sqrt{1 - \left(\frac{\delta}{A}\right)^2 (1 - \epsilon)^2} + \sqrt{1 - \left(\frac{\delta}{A}\right)^2 (1 + \epsilon)^2} \right] - j \frac{4D\delta}{\pi A^2} \quad (2.3-25)$$

Hence we see for the first time a nonlinearity giving rise to a DF which is complex. Both the magnitude and phase of the DF are functions of A . In general, multivalued characteristics will lead to complex DFs.¹ A normalized plot of DF magnitude and phase angle is presented in Appendix B for several values of ϵ .

A special case of the nonlinear characteristic of Fig. 2.3-8 is the *rectangular hysteresis characteristic*, derived by setting $\epsilon = 0$. It is frequently referred to as *toggle* because of its occurrence in mechanical spring-loaded toggle switches. The DF for rectangular hysteresis is given by either of the following forms, derived from Eq. (2.3-24) or (2.3-25) by setting ϵ to zero.

$$\begin{aligned} N(A) &= \frac{4D}{\pi A} e^{-j \sin^{-1}(\delta/A)} \\ &= \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} - j \frac{4D\delta}{\pi A^2} \end{aligned} \quad (2.3-26)$$

BACKLASH

Backlash in gearing can be defined as the amount by which a tooth space exceeds the thickness of a mating tooth. A linear force motor engaging its load through a linkage with backlash b is shown in Fig. 2.3-9. The effect of backlash, which is ever-present in geared systems, is almost always destabilizing. For this reason antibacklash (spring-loaded) gearing is often employed. Another approach to circumventing the destabilizing action of backlash in geared systems is to operate the motor with an output velocity bias. This approach is used in the turntable testing of high-quality gyroscopes, for example, where under other circumstances the backlash between motor and turntable could introduce sufficient measurement error to invalidate the testing.

¹ Although there are multivalued characteristics for which $n_g(A) = 0$. See Prob. 2-13.

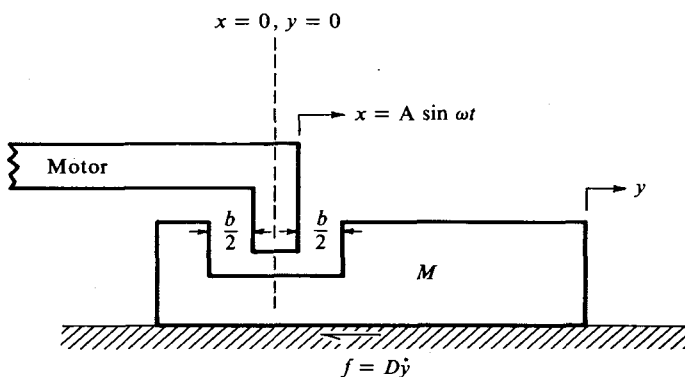


Figure 2.3-9 Linear motor driving a viscous friction plus inertia load through a linkage with backlash b .

In order to develop the DF for backlash in the system of Fig. 2.3-9, we consider two limiting cases. In the first of these the friction forces on the load are dominant; this is therefore referred to as *friction-controlled backlash*. In the second the load inertia forces are dominant; this is referred to as *inertia-controlled backlash*. In Sec. 2.4 the more general case, including both friction and inertia forces simultaneously, is treated. In that case a frequency-dependent DF results. In all cases we consider the motor an ideal drive in the sense that it provides a sinusoidal output displacement independent of load force requirements.

Friction-controlled backlash In this case we take $M = 0$. A plot of the motor input motion, load output motion, and equivalent backlash characteristic is shown in Fig. 2.3-10. Let us note here that it is not sufficient merely to indicate that a certain amount of backlash is present in a given

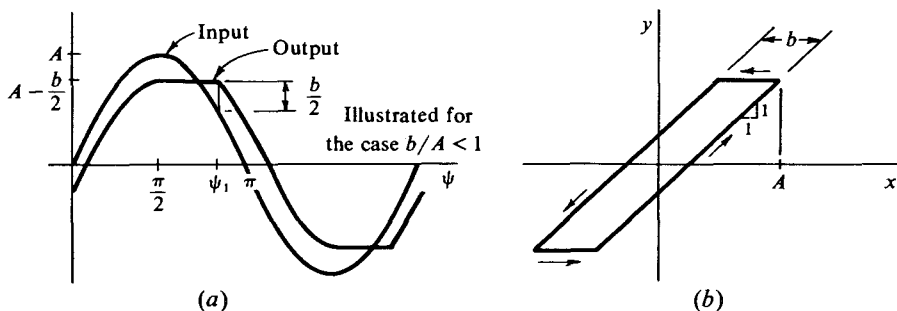


Figure 2.3-10 (a) Waveforms for friction-controlled backlash. (b) Equivalent backlash characteristic.

situation; one must further specify the load dynamics in order to derive the equivalent backlash input-output characteristic.

In this case the motor and load are in contact up to the point at which motor velocity reverses. Contact is not reestablished until the backlash is closed on the other side. Bouncing between motor and load is assumed negligible. The DF is computed according to

$$\begin{aligned}
 N(A) &= \frac{2j}{\pi A} \int_0^\pi y(A \sin \psi) e^{-j\psi} d\psi \\
 &= \frac{2j}{\pi A} \left[\int_0^{\pi/2} \left(A \sin \psi - \frac{b}{2} \right) e^{-j\psi} d\psi + \int_{\pi/2}^{\psi_1} \left(A - \frac{b}{2} \right) e^{-j\psi} d\psi \right. \\
 &\quad \left. + \int_{\psi_1}^\pi \left(A \sin \psi + \frac{b}{2} \right) e^{-j\psi} d\psi \right] \\
 &= \frac{1}{\pi} \left[\frac{3}{2} \pi - \psi_1 - 2 \left(1 - \frac{b}{A} \right) \cos \psi_1 + \sin \psi_1 \cos \psi_1 \right] \\
 &\quad - \frac{j}{\pi} \left[2 - 2 \left(1 - \frac{b}{A} \right) \sin \psi_1 - \cos^2 \psi_1 \right] \quad (2.3-27)
 \end{aligned}$$

The angle ψ_1 , which defines the point at which the backlash is closed during the negative-velocity part of the cycle, is given by

$$\psi_1 = \pi - \sin^{-1} \left(1 - \frac{b}{A} \right) \quad (2.3-28)$$

where $\pi/2 < \psi_1 \leq \pi$ for $0 < b/A \leq 1$; the arcsin is interpreted as an angle in the first quadrant. In terms of b/A the DF can thus be rewritten as

$$\begin{aligned}
 N(A) &= \frac{1}{\pi} \left[\frac{\pi}{2} + \sin^{-1} \left(1 - \frac{b}{A} \right) + \left(1 - \frac{b}{A} \right) \sqrt{\frac{2b}{A} - \left(\frac{b}{A} \right)^2} \right] - j \frac{1}{\pi} \left[\frac{2b}{A} - \left(\frac{b}{A} \right)^2 \right] \\
 &\quad (2.3-29)
 \end{aligned}$$

The real part of $N(A)$ can be further rewritten in terms of the saturation function [Eq. (2.3-5)], viz.,

$$\begin{aligned}
 n_p(A) &= \text{Re} [N(A)] \\
 &= \frac{1}{2} \left[1 + f \left(1 - \frac{b}{A} \right) \right] \quad (2.3-30)
 \end{aligned}$$

in which form its plotting is facilitated since use can now be made of Fig. 2.3-4. The same expression for $N(A)$ results for values $1 < b/A \leq 2$. In that case $\pi < \psi_1 \leq 3\pi/2$, and the arcsin is interpreted as an angle in the fourth quadrant. A plot of the magnitude and phase angle of $N(A)$, calibrated in b/A , is given in Fig. 2.3-12.

Inertia-controlled backlash In this case we set $D = 0$. Input-output waveforms and the resulting equivalent backlash characteristic are presented in Fig. 2.3-11. Here we see evidence that the specification of backlash is indeed incomplete unless the load description is also included.

At the values $\psi = n\pi$, $n = 0, 1, 2, \dots$, the motor imparts to the load its maximum velocity, $\dot{x}_{\max} = A\omega = \dot{y}$. Motor-load separation then occurs, and the load coasts with constant velocity until the backlash is closed on the other side. The angle ψ_1 at which the backlash is again closed is given by

$$\psi_1 = \sin \psi_1 + \frac{b}{A} \quad (2.3-31)$$

Again we assume bouncing between motor and load to be negligible. The DF is computed according to

$$\begin{aligned} N(A) &= \frac{2j}{\pi A} \int_0^\pi y(A \sin \psi) e^{-j\psi} d\psi \\ &= \frac{2j}{\pi A} \left[\int_0^{\psi_1} \left(A\psi - \frac{b}{2} \right) e^{-j\psi} d\psi + \int_{\psi_1}^\pi \left(A \sin \psi + \frac{b}{2} \right) e^{-j\psi} d\psi \right] \\ &= \frac{1}{\pi} (\pi + 2 \sin \psi_1 - \psi_1 - \sin \psi_1 \cos \psi_1) - j \frac{1}{\pi} (1 - \cos \psi_1)^2 \end{aligned} \quad (2.3-32)$$

See Fig. 2.3-12.

If, instead of taking either $M = 0$ or $D = 0$ in each of the above DF calculations, we were to allow M and D to be nonzero, it would still be true that in the limit of increasing input frequency ($\omega \rightarrow \infty$) inertia forces would

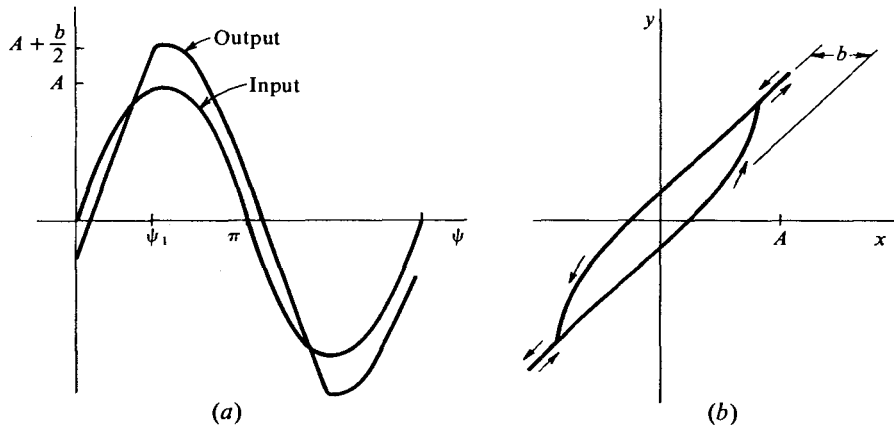


Figure 2.3-11 (a) Waveforms for inertia-controlled backlash. (b) Equivalent backlash characteristic.

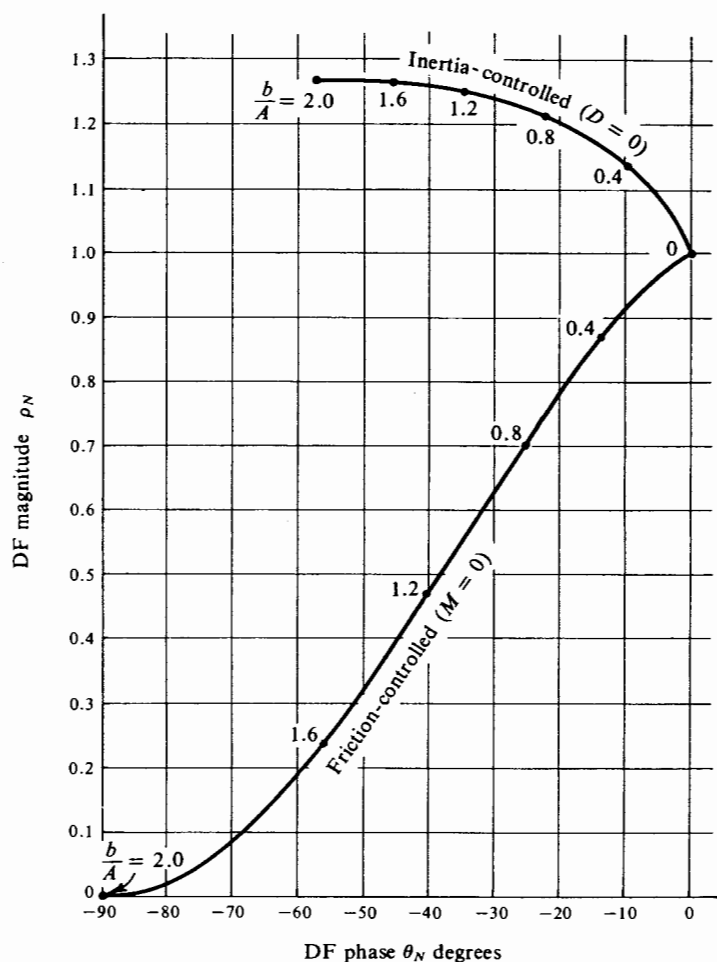


Figure 2.3-12 DFs for friction-controlled and inertia-controlled backlash.

predominate, and in the limit of decreasing input frequency ($\omega \rightarrow 0$) friction forces predominate. Hence we might expect the two curves of Fig. 2.3-12 to form the upper and lower bounds of the DF for backlash with both inertia and friction, as input frequency is varied. We indeed observe this behavior in the frequency-dependent-backlash DF calculation of Sec. 2.4.

HARMONIC GENERATION

The amount by which the output of a sinusoidally forced nonlinearity differs from its first harmonic has been called the *residual*, or *remnant*. Let us

consider the actual output harmonic content associated with several of the nonlinearities whose output first harmonics have been previously determined.

General piecewise-linear odd memoryless nonlinearity The ratio of output k th-harmonic amplitude to input amplitude is ($\psi_i = \sin^{-1} \delta_i/A$)

$$\begin{aligned}
 \frac{A_k}{A} &= \frac{4}{\pi A} \int_0^{\pi/2} y(A \sin \psi) \sin k\psi \, d\psi \\
 &= \frac{4}{\pi A} \left(\int_0^{\psi_1} y \sin k\psi \, d\psi + \int_{\psi_1}^{\psi_2} y \sin k\psi \, d\psi + \int_{\psi_2}^{\psi_3} y \sin k\psi \, d\psi \right. \\
 &\quad \left. + \int_{\psi_3}^{\pi/2} y \sin k\psi \, d\psi \right) \\
 &= \frac{2}{\pi k} \left[(m_1 - m_2) \left(\frac{\sin(k-1)\psi_1}{k-1} + \frac{\sin(k+1)\psi_1}{k+1} \right) \right. \\
 &\quad + (m_2 - m_3) \left(\frac{\sin(k-1)\psi_2}{k-1} + \frac{\sin(k+1)\psi_2}{k+1} \right) \\
 &\quad \left. + (m_3 - m_4) \left(\frac{\sin(k-1)\psi_3}{k-1} + \frac{\sin(k+1)\psi_3}{k+1} \right) + m_4 \frac{\pi}{2} \right] \quad (2.3-33)
 \end{aligned}$$

where the values of y for the different intervals of integration are given by Eq. (2.3-9). The result presented above is valid only for $A > \delta_3$, but its alteration to accommodate either larger or smaller regions is straightforward.

Saturation The ratio A_k/A for this nonlinear characteristic can be derived from the above result by choosing

$$m_1 = m \quad m_2 = m_3 = m_4 = 0 \quad (2.3-34)$$

from which it follows that

$$\frac{A_k}{A} = \frac{2m}{\pi k} \left[\frac{\sin[(k-1)\sin^{-1}(\delta/A)]}{k-1} + \frac{\sin[(k+1)\sin^{-1}(\delta/A)]}{k+1} \right] \quad (2.3-35)$$

Harmonic-amplitude ratios are plotted up to the seventh-harmonic term in Fig. 2.3-13.

Polynomial-type nonlinearities Consider a one-term n th-order polynomial, n odd. The harmonic-amplitude ratio is found according to

$$\begin{aligned}
 \frac{A_k}{A} &= \frac{4}{\pi A} \int_0^{\pi/2} y(A \sin \psi) \sin k\psi \, d\psi \\
 &= \frac{4}{\pi} c_n A^{n-1} \int_0^{\pi/2} \sin^n \psi \sin k\psi \, d\psi \\
 &= \frac{c_n A^{n-1} \sin(k\pi/2) \Gamma(n+1)}{2^{n-1} \Gamma[(n+k+2)/2] \Gamma[(n-k+2)/2]} \quad (2.3-36)
 \end{aligned}$$

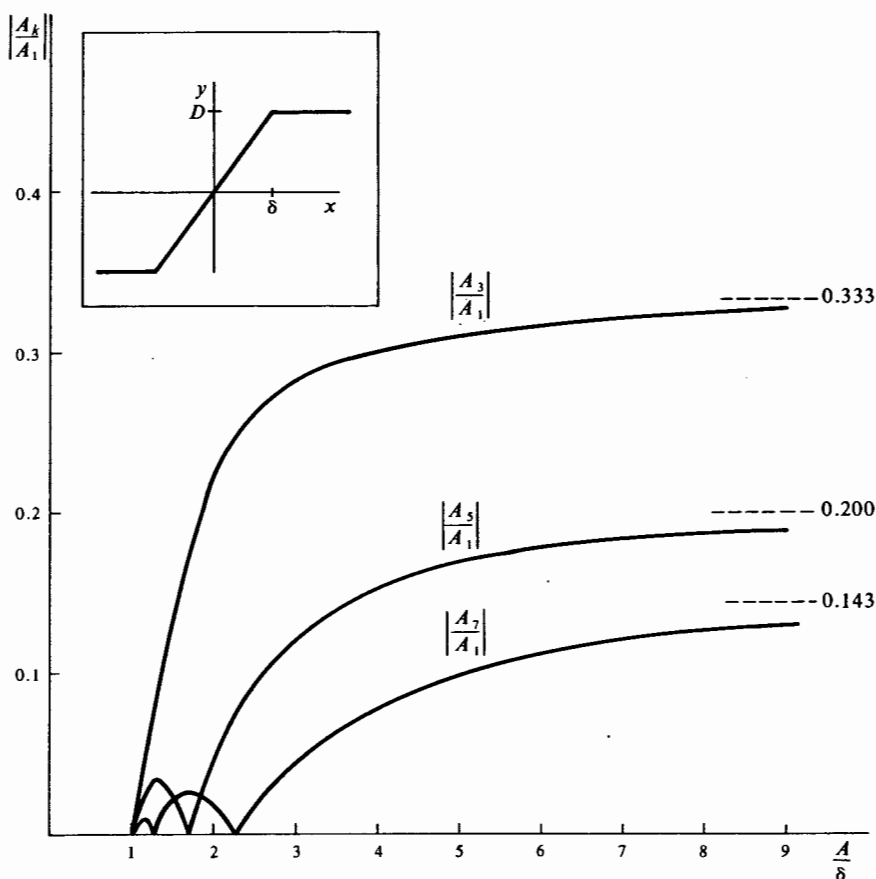


Figure 2.3-13 Output harmonic content for saturation.

Results of the related integration for an n th-order odd nonlinearity, n even, are identical. Using this expression, the total output harmonic content of many-term polynomial nonlinearities may be built up, one frequency at a time. As expected, $A_k = 0$ for k even. Only odd harmonics exist. When n is odd, we see that all harmonics of order greater than n are zero.

$$A_k = 0 \quad \text{for all } k > n, n \text{ odd} \quad (2.3-37)$$

This can be deduced from Eq. (2.3-36) with the aid of the fact that the value of the gamma function for all negative-integer arguments is infinite. For n even, however, all output odd harmonics exist.

Symmetrical square-law nonlinearity This characteristic is defined by

$$y = c_2 x |x| \quad (2.3-38)$$

Harmonic content is given by

$$\left| \frac{A_k}{A} \right| = \frac{c_2 A}{\Gamma[(4+k)/2] \Gamma[(4-k)/2]} \quad k \text{ odd} \quad (2.3-39)$$

whence we find

$$\left| \frac{A_3}{A_1} \right| = 0.200 \quad \left| \frac{A_5}{A_1} \right| = 0.029 \quad \left| \frac{A_7}{A_1} \right| = 0.010 \quad (2.3-40)$$

Cubic nonlinearity Proceeding as before,

$$y = c_3 x^3 \quad (2.3-41)$$

and we find

$$\left| \frac{A_3}{A_1} \right| = 0.33 \quad \left| \frac{A_5}{A_1} \right| = 0 \quad \left| \frac{A_7}{A_1} \right| = 0 \quad (2.3-42)$$

Harmonic nonlinearity This characteristic is given by

$$y = M \sin mx \quad (2.3-43)$$

Hence the output k th harmonic is

$$\begin{aligned} A_k &= \frac{4}{\pi} \int_0^{\pi/2} y(A \sin \psi) \sin k\psi \, d\psi \\ &= \frac{4M}{\pi} \int_0^{\pi/2} \sin(mA \sin \psi) \sin k\psi \, d\psi \\ &= M[1 - (-1)^k J_k(mA)] \end{aligned} \quad (2.3-44)$$

where $J_k(mA)$ is the Bessel function of order k . The harmonic-amplitude ratios of interest are

$$\left| \frac{A_3}{A_1} \right| = \left| \frac{J_3(mA)}{J_1(mA)} \right| \quad \left| \frac{A_5}{A_1} \right| = \left| \frac{J_5(mA)}{J_1(mA)} \right| \quad \left| \frac{A_7}{A_1} \right| = \left| \frac{J_7(mA)}{J_1(mA)} \right| \quad (2.3-45)$$

These functions are highly oscillatory and reach peaks of up to 70 in the interval $0 < mA \leq 10$. For our present purposes it is sufficient to study the behavior for large mA , in which case the Bessel function is well approximated by

$$J_k(mA) \approx \sqrt{\frac{2}{\pi mA}} \cos \left(mA - \frac{k\pi}{2} - \frac{\pi}{4} \right) \quad (2.3-46)$$

Applying this asymptotic representation to the harmonic ratios above, we find

$$\left| \frac{A_3}{A_1} \right| = \left| \frac{A_5}{A_1} \right| = \left| \frac{A_7}{A_1} \right| = 1 \quad (2.3-47)$$

which clearly indicates the presence of substantial harmonic content in the range $mA \gg 1$.

Rectangular hysteresis For this characteristic we easily obtain

$$\begin{aligned} |A_k| &= \left| \frac{2j}{\pi} \int_0^\pi y(A \sin \psi) e^{-jk\psi} d\psi \right| \\ &= \left| \frac{4D}{\pi k} e^{-jk \sin^{-1}(\delta/A)} \right| \\ &= \frac{4D}{\pi k} \quad k \text{ odd} \end{aligned} \quad (2.3-48)$$

from which the harmonic ratios are

$$\left| \frac{A_3}{A_1} \right| = \frac{1}{3} \quad \left| \frac{A_5}{A_1} \right| = \frac{1}{5} \quad \left| \frac{A_7}{A_1} \right| = \frac{1}{7} \quad (2.3-49)$$

Generalization For monotonically increasing nonlinear characteristics the k th-harmonic output amplitude ratio $|A_k/A_1|$ generally is on the order of $1/k$. Characteristics which are nonmonotonically increasing are apt to possess substantially higher harmonic-amplitude ratios. In most cases, calculation is readily executed.

In the final analysis it is the transfer function of the linear elements of a closed-loop system which emphasizes or deemphasizes loop harmonic content. For those cases in which the loop linear elements contain no resonance peaks at frequencies beyond the nonlinearity input fundamental frequency, DF linearization of a nonlinearity may generally be employed without excessive error due to the presence of unaccounted-for harmonics. This topic will be further pursued in the following chapter.

2.4 DF CALCULATION FOR FREQUENCY-DEPENDENT NONLINEARITIES

In this section we examine some methods for dealing with *dynamic* nonlinearities, those for which outputs depend upon inputs and their derivatives. As discussed previously, they are represented as

$$y = y(x, \dot{x}) \quad (2.4-1)$$

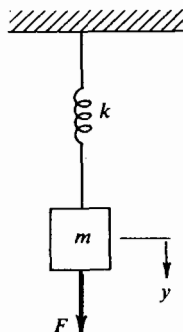


Figure 2.4-1 Simple linear mass-spring system.

At any single input frequency a DF magnitude vs. phase-angle plot for varying input amplitude may be constructed, with different such plots belonging to different input frequencies. Thus it will be observed that DFs for dynamic nonlinearities may be generally portrayed by a onefold infinity of graphs, with input frequency as a free parameter.

LINEAR MECHANICAL SYSTEM

As an example of a very simple system possessing a frequency-dependent DF, consider the single-degree-of-freedom mass-spring system of Fig. 2.4-1. For a sinusoidal force input

$$F = A \sin \omega t \quad (2.4-2)$$

the differential equation of motion of the mass-spring assembly is given by

$$m\ddot{y} + ky = A \sin \omega t \quad (2.4-3)$$

for which the exact *forced* solution is

$$y = \frac{A/m}{k/m - \omega^2} \sin \omega t \quad (2.4-4)$$

Consequently, the DF is given by

$$N(A, \omega) = \frac{1}{A} \frac{A/m}{\omega_n^2 - \omega^2} \quad (2.4-5)$$

where $\omega_n^2 = k/m$. The amplitude dependence cancels, with the result

$$N(\omega) = \frac{1/m}{\omega_n^2 - \omega^2} \quad (2.4-6)$$

which is a function of ω for fixed system parameters. In fact, the DF presented is *precisely* the linear transfer function of the system from force

the point of motor-load separation, which occurs when the motor *decelerates* faster than the load. Using the fact that the velocities of motor and load are identical at separation, one can show that ψ_s is given by

$$\psi_s = \tan^{-1} \frac{1}{\gamma} \quad (2.4-7)$$

where

$$\gamma = \frac{M}{D} \omega \quad (2.4-8)$$

Contact is reestablished when the backlash is again taken up. This occurs at $\psi = \psi_c$, where

$$y(\psi_c) = x(\psi_c) + \frac{b}{2} \quad (2.4-9)$$

which, after insertion of $x(\psi_c)$ and $y(\psi_c)$ in terms of system parameters, can be rewritten in the form

$$\sqrt{1 - \gamma^2} - \frac{b}{A} = \sin \psi_c + \frac{\gamma^2 e^{[1/\gamma][\tan^{-1} 1/\gamma - \psi_c]}}{\sqrt{1 + \gamma^2}} \quad (2.4-10)$$

The in-phase and quadrature components of the DF are found by the now-familiar integration schemes, yielding

$$\begin{aligned} n_p(A, \omega) &= \frac{2}{\pi A} \int_0^\pi y(A \sin \psi, A \omega \cos \psi) \sin \psi \, d\psi \\ &= \frac{2\gamma^2}{\pi(1 + \gamma^2)} \left[\frac{1 + \gamma^2}{2\gamma^2} (\psi_s - \psi_c + \pi + \sin \psi_c \cos \psi_c) + \frac{1}{2\gamma} \right. \\ &\quad \left. + \frac{1}{\gamma} \left(\frac{b}{A} - \sqrt{1 + \gamma^2} \right) \left(\frac{1}{\gamma} \cos \psi_c - \sin \psi_c \right) \right. \\ &\quad \left. - \frac{1}{\gamma} \sin^2 \psi_c - \sin \psi_c \cos \psi_c \right] \quad (2.4-11) \end{aligned}$$

and

$$\begin{aligned} n_q(A, \omega) &= \frac{2}{\pi A} \int_0^\pi y(A \sin \psi, A \omega \cos \psi) \cos \psi \, d\psi \\ &= \frac{-2\gamma^2}{\pi(1 + \gamma^2)} \left[1 + \frac{1}{2\gamma^2} + \left(\frac{1 - \gamma^2}{2\gamma^2} \right) \sin^2 \psi_c \right. \\ &\quad \left. + \frac{1}{\gamma} \left(\frac{b}{A} - \sqrt{1 + \gamma^2} \right) \left(\frac{1}{\gamma} \sin \psi_c + \cos \psi_c \right) \right. \\ &\quad \left. + \frac{1}{\gamma} \sin \psi_c \cos \psi_c \right] \quad (2.4-12) \end{aligned}$$

The magnitude and phase of the DF are both functions of A and ω ; viz.,

$$\begin{aligned}\rho_N(A, \omega) &= \sqrt{n_p^2(A, \omega) + n_q^2(A, \omega)} \\ \theta_N(A, \omega) &= \tan^{-1} \frac{n_q(A, \omega)}{n_p(A, \omega)}\end{aligned}\quad (2.4-13)$$

A plot of $\rho_N(A, \omega)$ versus $\theta_N(A, \omega)$ for various A and ω (the latter characterized by γ) is shown in Appendix B. The whole family of DF loci, as expected, are contained between the curves for the inertia- and friction-controlled backlash cases. Unlike the friction-controlled backlash nonlinearity, the response of inertia-controlled backlash need not be zero for all $A < b/2$. In fact, the DF can be defined for all $A > b/3.72$. Beyond this point, in the direction of decreasing A , both subharmonics and aperiodic responses occur. Correspondingly, all intermediate cases shown can be extended somewhat to values of b/A in excess of 2.0, but less than 3.72. Studies of this extension, as well as the cases of backlash with load inertia and coulomb friction, both with zero and nonzero input-velocity biases, are available in the literature (Refs. 11, 16, 44, 46, 52).

NONLINEAR CLEGG INTEGRATOR

The Clegg integrator (Refs. 5, 33) represents an attempt to synthesize a nonlinear circuit possessing the amplitude-frequency characteristic of a linear integrator while avoiding the 90° phase lag associated with the linear transfer function. Clearly, no linear circuit can accomplish this objective since the linear integrator is itself a minimum-phase network.

A functional diagram of the Clegg integrator, which switches on input zero crossings, is illustrated in Fig. 2.4-3a. Basically, operation consists of the input being gated through one of two integrators (the output of the other is simultaneously reset) in accordance with zero-crossing detector (ZCD) commands. Implementation of this integrator, including gates and ZCD, can be effected with four diodes, four RC networks, and two operational amplifiers (Ref. 5). Input and output waveforms are shown in Fig. 2.4-3b. In the interval $0 < \psi < \pi$, the output is ($\psi = \omega t$)

$$\begin{aligned}y(\psi) &= \int_0^{\psi/\omega} x d\left(\frac{\psi}{\omega}\right) \\ &= \frac{1}{\omega} \int_0^\psi A \sin \psi d\psi \\ &= \frac{A}{\omega} (1 - \cos \psi)\end{aligned}\quad (2.4-14)$$

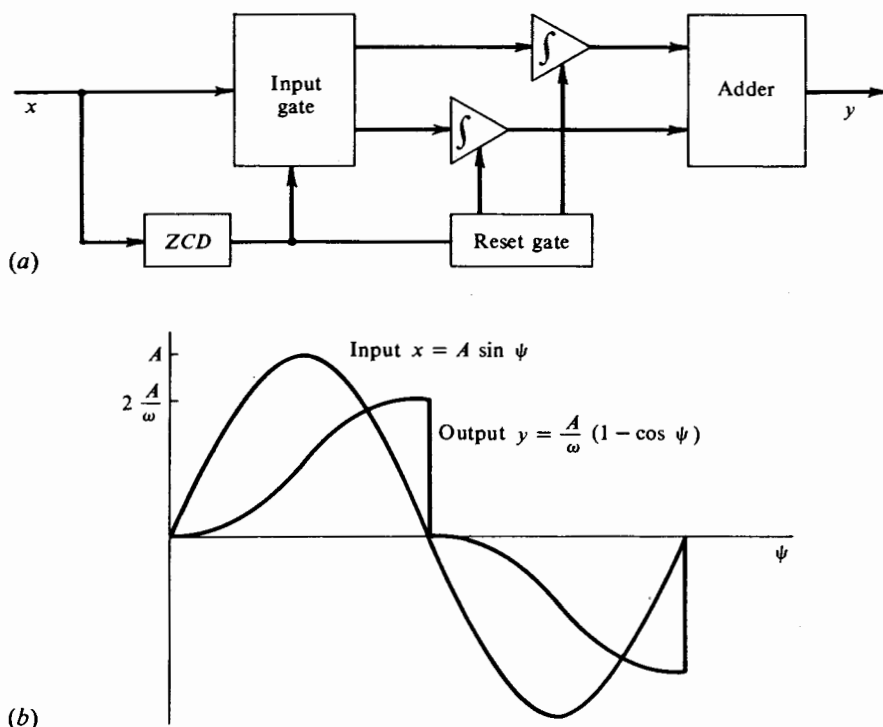


Figure 2.4-3 (a) Nonlinear Clegg integrator. (b) Associated input-output waveforms.

The DF is thus given by

$$\begin{aligned}
 N(A, \omega) &= \frac{2j}{\pi A} \int_0^\pi y(A \sin \psi, A\omega \cos \psi) e^{-j\psi} d\psi \\
 &= \frac{2j}{\pi A} \int_0^\pi \frac{A}{\omega} (1 - \cos \psi) e^{-j\psi} d\psi \\
 &= \frac{4}{\pi \omega} \left(1 - j \frac{\pi}{4} \right)
 \end{aligned} \tag{2.4-15}$$

and it is evident that the (in this case) undesired dependence of the DF upon input amplitude has been successfully avoided, with the result

$$\begin{aligned}
 N(\omega) &= \frac{1}{\omega} \sqrt{1 + \left(\frac{4}{\pi} \right)^2} e^{-j \tan^{-1}(\pi/4)} \\
 &= \frac{1.619}{\omega} e^{-j38.15^\circ}
 \end{aligned} \tag{2.4-16}$$

This DF has associated with it approximately 52° less phase lag than that for the linear integrator. Its merit as a system compensation network from the point of view of loop stability is thus apparent.

Harmonic content in the sinusoidally forced output of this nonlinear integrator follows directly from the observation that the output is the sum of a square wave of amplitude A/ω and a negative cosine wave of amplitude A/ω , both of equal period. The output k th harmonic is therefore due entirely to the square-wave portion

$$A_k = \frac{4A}{\pi\omega k} \quad k \text{ odd} \quad (2.4-17)$$

and the harmonic-amplitude ratios of interest are

$$\left| \frac{A_3}{A_1} \right| = \frac{1}{3} \quad \left| \frac{A_5}{A_1} \right| = \frac{1}{5} \quad \left| \frac{A_7}{A_1} \right| = \frac{1}{7} \quad (2.4-18)$$

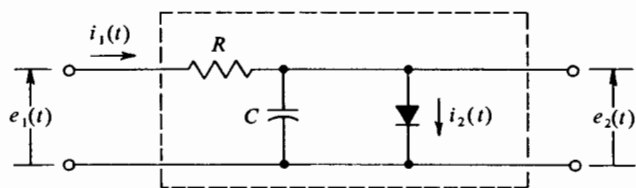
which are typical, as previously noted.

FREQUENCY-INDEPENDENT DFs FROM FREQUENCY-DEPENDENT NONLINEARITIES

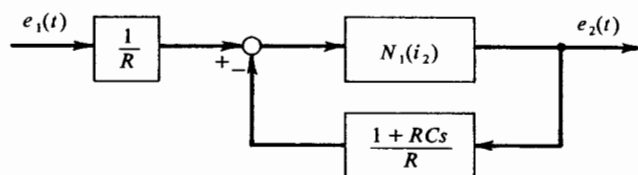
We have seen that the frequency-dependent DF can be treated by a simple extension of the basic DF concept. The result of this approach is a family of DF loci, perhaps with frequency as a convenient family parameter. Although it is true that analysis can now proceed within this framework, one additional approach is well worth consideration. The intent of this method is the divorce of linear frequency-dependent and nonlinear amplitude-dependent parts of an otherwise amplitude- and frequency-dependent nonlinear element. Although such a division cannot always be effected, when it can, the resultant nonlinear element will be far simpler to handle. In particular, the DF associated with the nonlinearity will be frequency-independent (Ref. 3).

Consider the four-terminal nonlinear RC network of Fig. 2.4-4a. For the purpose of demonstration we shall work under the condition that the nonlinear network must be treated as shown and that no physical reorganization within some larger system is possible. Given the voltage-current characteristic $e_2 = N_1(i_2)$ of the network diode, we may proceed to manipulate system variables to obtain the desired end result. Laplace transform notation is most convenient in this endeavor. In this form the equations governing system behavior are

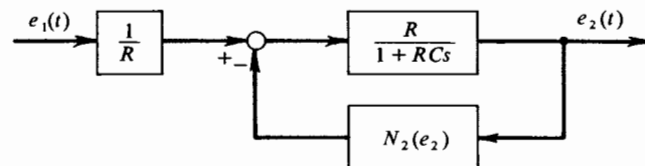
$$\begin{aligned} E_1(s) - E_2(s) &= RI_1(s) \\ I_1(s) - I_2(s) &= CsE_2(s) \\ e_2 &= N_1(i_2) \end{aligned} \quad (2.4-19)$$



(a) RC network with diode rectifier



(b) Block diagram of equivalent system



(c) Another block diagram of equivalent system

Figure 2.4-4 Frequency-dependent nonlinearity and two equivalent block-diagram representations.

which may easily be represented by block diagram, as in Fig. 2.4-4b. This block diagram contains only linear frequency-dependent elements and nonlinear frequency-independent elements. The desired separation has been accomplished since the isolated nonlinear part of the block diagram is frequency-independent. If the nonlinear RC network were part of a larger and otherwise linear system, the linear feedback branch of this network could be associated with the rest of the system to yield a final block diagram in which there would exist a single amplitude-dependent nonlinearity plus other purely linear elements. Analytic studies of this system would be in terms of a frequency-independent DF, considerably more convenient than equivalent studies using the corresponding frequency-dependent DF.

For the above example it is also possible to generate another useful block diagram. This new configuration contains a frequency-independent nonlinearity in the feedback path, which happens to be the inverse of the

nonlinearity representation given before, namely, $i_2 = N_2(e_2)$. Figure 2.4-4c depicts this arrangement. Stout (Ref. 50) has shown that any two-part system containing a single explicit nonlinearity can always be reduced to four topologically identical and mathematically equivalent block diagrams. In these diagrams the nonlinear element may appear as either a forward or feedback block, with its input and output in either a normal or reversed cause-effect relationship.

DF CALCULATION FOR IMPLICIT DYNAMIC NONLINEARITIES

Many nonlinearities are best described in terms of input-output differential equations. This is as opposed to some explicit dynamical description, for example. Under this circumstance the nonlinearity output waveform is not generally available directly in terms of its input, as has previously been assumed. The *implicit* dynamical relationship between input and output must therefore be dealt with by some special means. In order to demonstrate one useful method of approach, we compute the DF for the dynamic nonlinearity described by

$$\ddot{y} + 3y^2\dot{y} + y = x \quad (2.4-20)$$

where x and y are the nonlinearity input and output, respectively. To find $y(t)$ in response to the harmonic input

$$x = A \sin(\omega t + \theta) \quad (2.4-21)$$

one must possess the general solution of Eq. (2.4-20). Generally speaking, the solutions to nonlinear differential equations are unknown. Rather than obtain $y(t)$, and thus derive the DF by performing the usual Fourier expansion, we now are forced to seek an alternative approach. An artifice frequently worthwhile is to assume for the form of the nonlinearity output

$$y = Y \sin \omega t \quad (2.4-22)$$

and to solve for the sinusoidal input which results in this output. First-harmonic approximation allows execution of this method.

Inserting Eqs. (2.4-21) and (2.4-22) into (2.4-20), we get

$$-\omega^2 Y \sin \omega t + 3\omega Y^3 \sin^2 \omega t \cos \omega t + Y \sin \omega t = A \sin(\omega t + \theta) \quad (2.4-23)$$

The second term on the left-hand side may be expanded into first- and third-harmonic portions, viz.,

$$\sin^2 \omega t \cos \omega t = \frac{1}{4} \cos \omega t - \frac{1}{4} \cos 3\omega t \quad (2.4-24)$$

Dropping the third-harmonic term, Eq. (2.4-23) becomes

$$(1 - \omega^2) Y \sin \omega t + \frac{3}{4} \omega Y^3 \cos \omega t = A \sin(\omega t + \theta) \quad (2.4-25)$$

Collating coefficients of $\sin \omega t$ and $\cos \omega t$ yields the following equations in A , Y , θ :

$$\begin{aligned}(1 - \omega^2)Y &= A \cos \theta \\ \frac{3}{4}\omega Y^3 &= A \sin \theta\end{aligned}\tag{2.4-26}$$

Equations (2.4-26) may be simultaneously solved to yield $\theta(A, \omega)$, $Y(A, \omega)$.

$$(1 - \omega^2)^2 Y^2 + \left(\frac{3}{4}\omega Y^3\right)^2 = A^2\tag{2.4-27}$$

$$\theta = \tan^{-1} \left(\frac{3}{4} \frac{\omega Y^2}{1 - \omega^2} \right)\tag{2.4-28}$$

Equation (2.4-27) is an implicit relationship for $Y(A, \omega)$, which, once obtained, may be used to find $\theta(A, \omega)$ given explicitly by Eq. (2.4-28). The first-harmonic gain of the nonlinearity is given by

$$\begin{aligned}N(A, \omega) &= \frac{Y(A, \omega)}{A} e^{-j\theta(A, \omega)} \\ &= \rho_N(A, \omega) e^{j\theta_N(A, \omega)}\end{aligned}\tag{2.4-29}$$

Equations (2.4-27) and (2.4-28) may be rewritten in terms of ρ_N and θ_N .

$$\begin{aligned}\rho_N^6 + \left(\frac{4}{3} \frac{1 - \omega^2}{\omega A^2} \right)^2 \rho_N^2 - \left(\frac{4}{3\omega A^2} \right)^2 &= 0 \\ \theta_N &= -\tan^{-1} \left(\frac{3}{4} \frac{\omega \rho_N^2 A^2}{1 - \omega^2} \right)\end{aligned}\tag{2.4-30}$$

Thus the frequency-dependent DF has been determined. These results are identical with those presented elsewhere (Ref. 39), the same nonlinear equation in that instance treated by a method of Stoker (Ref. 49). Observe that the exact output first harmonic has not been calculated; rather, an approximation to it has been arrived at. Better approximations can be generated by assuming a more complete description of y in Eq. (2.4-22); however, the labor entailed rapidly increases.

EXTENSIONS OF THE DF CONCEPT

Another method for dealing with implicit dynamical nonlinearities has been developed by Klotter (Ref. 26), who replaces the Fourier harmonic concept of the DF with a corresponding "Hamilton harmonic" concept. In particular, he chooses the nonlinearity output amplitude and phase so as to minimize the integral whose Euler equation coincides with the given differential equation. This process is somewhat analogous to minimizing the mean-squared error in conventional DF formulation by selecting the Fourier series

representation of a nonlinearity output [Eq. (2.2-23)]. One significant difference between the two methods is that the Fourier coefficients are fixed, independent of the degree of approximation employed, whereas the Hamilton harmonics depend in some way upon rejected higher harmonics (the residual). DFs generated by both methods are reported to show first-harmonic amplitude differences of the order of 10 percent.

It is possible to propose any number of other linearization schemes based upon a sinusoidal input. For example, the equivalent gain could be chosen to minimize the average approximation error or absolute magnitude of the error rather than mean-squared error as in DF formulation. Although the above-mentioned alternatives do not appear to have any advantage over the DF, an "rms DF" proposed by Gibson and Prasanna-Kumar (Ref. 12) shows some promise in application to systems with odd single-valued nonlinearities. It is defined by

$$N_{\text{rms}}(A) = \left[\frac{\int_0^{2\pi} [y(A \sin \psi)]^2 d\psi}{\int_0^{2\pi} (A \sin \psi)^2 d\psi} \right]^{\frac{1}{2}} \quad (2.4-31)$$

That is, the equivalent sinusoidal output of the nonlinearity is chosen to have the same rms value as the actual output. Using the notation of Fig. 2.2-1, it is readily demonstrated that Eq. (2.4-31) can be written as (odd nonlinearity)

$$N_{\text{rms}}(A) = \frac{\sqrt{A_1^2 + A_3^2 + A_5^2 + A_7^2 + \cdots}}{A}$$

Rankine and D'Azzo have proposed a "corrected conventional DF" based upon a truncated version of the rms DF, as follows:

$$N_{\text{ccDF}}(A) = \frac{\sqrt{A_1^2 + A_3^2}}{A} \quad (2.4-32)$$

In application to the study of a wide variety of systems, the corrected conventional DF was found to be consistently more accurate than either the DF or rms DF, although all results were, in fact, quite good (Ref. 43).

Another mechanism for studying the limit cycle behavior of certain nonlinear systems is the "elliptic describing function" (Ref. 24). It is particularly interesting in that the limit cycle waveshape is determined *after* the amplitude and frequency have been found. This differs from other DF methods wherein the waveshape is specified *a priori*. However, the computation associated with the elliptic-describing-function method and its practical restriction to single-valued odd static nonlinearities appear to rule it out as a generally useful analytical tool.

For the remainder of this and the following two chapters we confine our consideration to the conventional DF.

2.5 SYNTHESIS OF DFs

The DF for a complex nonlinearity can be synthesized by vector addition of the DFs for a group of similar nonlinearities whose *parallel* combination has the identical input-output characteristic. The method relates to vector addition of sinusoids and is exact within its own framework. Series-connected nonlinearities do not support the same generalizations. These cases are treated in what follows.

Consider n nonlinearities $N_1, N_2, \dots, N_n$ in parallel such that their total output $y(x)$ is

$$\begin{aligned} y(x) &= y_1(x) + y_2(x) + \dots + y_n(x) \\ &= \sum_{i=1}^n y_i(x) \end{aligned} \quad (2.5-1)$$

It then follows that the DF for the composite nonlinearity is given by

$$\begin{aligned} N(A) &= \frac{j}{\pi A} \int_0^{2\pi} \left[\sum_{i=1}^n y_i(A \sin \psi) \right] e^{-j\psi} d\psi \\ &= \frac{j}{\pi A} \sum_{i=1}^n \int_0^{2\pi} y_i(A \sin \psi) e^{-j\psi} d\psi \\ &= \sum_{i=1}^n N_i(A) \end{aligned} \quad (2.5-2)$$

Thus the DF for the composite nonlinearity is the sum of DFs for the individual nonlinearities comprising the composite nonlinearity. Since, in general, the DF for the i th nonlinearity is a complex quantity, Eq. (2.5-2) represents a vector addition. In the case of frequency- and amplitude-dependent elemental nonlinearities, the DF for a parallel combination is given as

$$N(A, \omega) = \sum_{i=1}^n N_i(A, \omega) \quad (2.5-3)$$

by direct analogy with Eq. (2.5-2). Once again the sum implies a vector addition.

One may assess the harmonic content of the composite nonlinearity output merely by performing appropriate vector additions on the individual harmonic terms of like frequency, again in a manner completely analogous to that of Eq. (2.5-2).

An example of the synthesis of a complex nonlinearity from simpler forms is the construction of the nonrectangular hysteresis-type nonlinearity (Fig. 2.5-1a) from the rectangular-hysteresis and linear-gain functions (Fig.

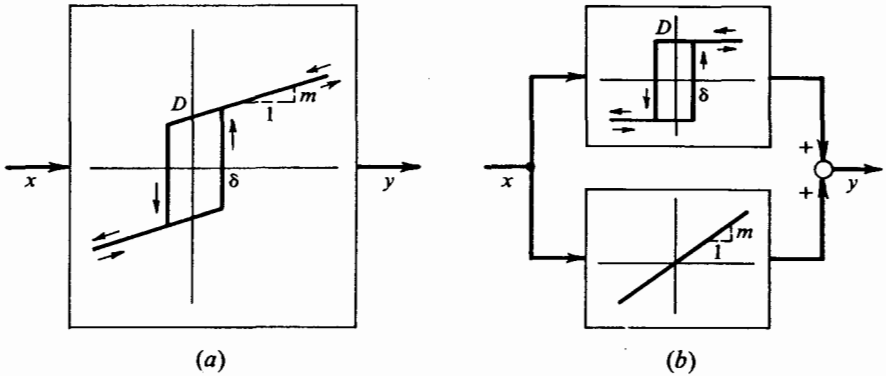


Figure 2.5-1 Synthesis of a simple hysteresis characteristic (a) from the elemental forms of rectangular hysteresis and linear gain (b).

2.5-1b). By executing two simple sketches the doubtful reader may convince himself of the validity of this maneuver, observing that the sinusoidally forced output of the simple nonrectangular-hysteresis element and the sum of sinusoidally forced outputs of the rectangular-hysteresis and linear-gain elements are identical. Accordingly, we have [Eq. (2.3-26)]

$$\begin{aligned}
 N(A) &= N_1(A) + N_2(A) \\
 &\quad \text{rectangular} \quad \text{linear} \\
 &\quad \text{hysteresis} \quad \text{gain} \\
 &= \frac{4D}{\pi A} e^{-j \sin^{-1}(\delta/A)} + m \quad (2.5-4)
 \end{aligned}$$

In terms of a real and imaginary part, or a magnitude and phase angle, the composite DF may be directly rewritten as follows:¹

$$\begin{aligned}
 N(A) &= \left[\frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} + m \right] - j \left(\frac{4D\delta}{\pi A^2} \right) \\
 &= \left[\left(\frac{4D}{\pi A} \right)^2 + m^2 + \frac{8Dm}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} \right]^{\frac{1}{2}} \\
 &\quad \times \exp \left\{ -j \tan^{-1} \left[\frac{4D(\delta/A)}{4D\sqrt{1 - (\delta/A)^2} + \pi A m} \right] \right\} \quad (2.5-5)
 \end{aligned}$$

¹ The expression giving the real and imaginary parts of $N(A)$ is least cumbersome of the two presented in Eq. (2.5-5). Since this is generally the case for complex nonlinearities with memory, Appendix B lists DFs in terms of their real and imaginary parts.

The harmonics generated by this composite nonlinearity are given by Eq. (2.3-49) and are due entirely to rectangular hysteresis, the linear gain generating no output harmonic content whatever.

Now consider the case of series-connected nonlinearities. To demonstrate the complexity of this situation, Fig. 2.5-2 illustrates the series connection of friction-controlled backlash followed by a limiter with dead zone. The two possible overall characteristics resulting for various ranges of A are also presented. It is at once evident that the decomposition, or synthesis procedure, is all but hopeless. No simple results exist comparable with the case of parallel-connected nonlinearities.

The question arises as to whether any simplified solution for the DF exists. In partial answer to this question, Gronner (Ref. 19) has shown that the exact DF for friction-controlled backlash followed by dead zone and an approximate DF computed directly by *multiplication* of the DFs of the individual nonlinearities compare quite well. Such a procedure implicitly assumes that the output of the first nonlinearity may in some sense be considered sinusoidal, in order to utilize the DF of the second nonlinearity in subsequent overall approximate DF calculation. This could be the case if, for example, the original nonlinearities were separated by a linear filter which, during analysis,

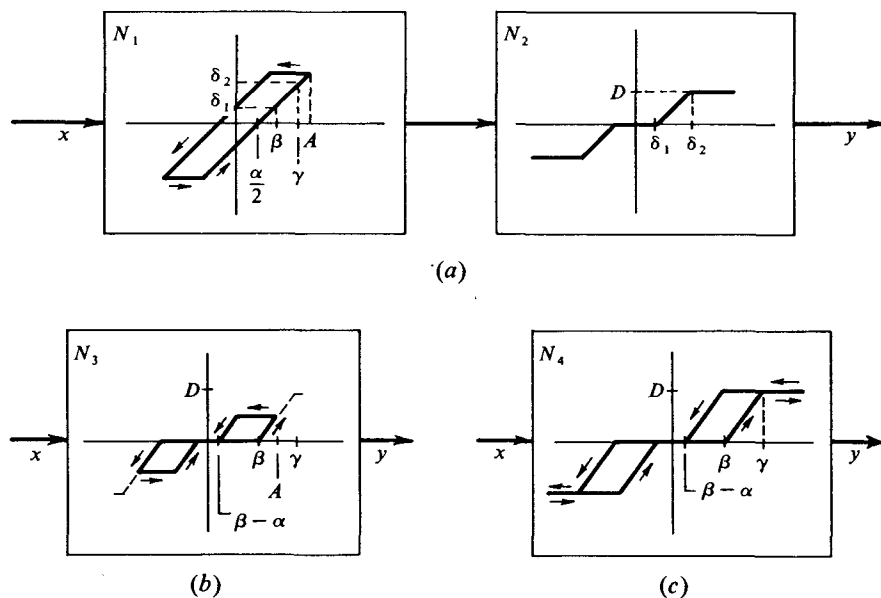


Figure 2.5-2 Series-connected nonlinearities. Elemental forms (a) and the corresponding overall characteristics for (b) $\beta < A < \gamma$ and (c) $A > \gamma$.

was associated with the first nonlinearity. Unfortunately, generalizations regarding expected accuracy of DF calculation using this procedure for a variety of nonlinearity combinations are not available.

2.6 TECHNIQUES FOR APPROXIMATE CALCULATION OF THE DF

Given any nonlinear characteristic, the corresponding DF can theoretically be evaluated using the techniques of previous sections. However, in practice, nonlinear characteristics are often known only by measurement of a physical system, precise analytic relationships remaining unknown. The DF can still be evaluated following an analytic curve fit of the experimentally derived characteristic, but time required to effect this curve fit and subsequent DF calculation may be unwarranted, considering that analysis following derivation of the DF is of an approximate nature. In fact, even given a nonlinear characteristic of precise analytical definition, an approximate DF calculation can often be justified on the grounds of approximate ultimate analysis. Furthermore, exact calculation can be tedious and lengthy.

To begin with, it may be possible to evaluate the DF experimentally. If a system nonlinearity can be isolated and excited with a sine wave of known amplitude and frequency, the application of a harmonic analyzer to the nonlinearity output directly yields all information necessary for DF specification. Instruments have been designed which automatically compute the frequency response of nonlinear systems based upon measurement of the evoked response to harmonic excitation (Refs. 4, 21, 57). Here we concern ourselves with approaches which can be executed with pencil and paper, starting at the point just after determination of the nonlinear characteristic. Thus we seek approximate methods for expediting hand calculation of the DF.

The most straightforward approach to graphic DF evaluation, starting with the nonlinear characteristic, is point-by-point derivation of the nonlinearity output waveform and harmonic analysis of it by direct area measurements (such as discussed in Ref. 8). Clearly, this is not the best manner of calculation. It would be more desirable, for example, to work directly with the nonlinear characteristic, without ever having to graph the actual output waveform. Several methods of computation having this virtue are considered. Let us restrict our attention to odd static symmetric nonlinearities.

PIECEWISE-LINEAR APPROXIMATION

The first approach deserving of mention simply prescribes an n -segment piecewise-linear fit to any given nonlinear characteristic. The DF for the

resultant nonlinear characteristic is either given by the general formula of Eq. (2.3-10), if that characteristic is memoryless, or can be derived by the methods of Sec. 2.3, if the characteristic possesses memory. Three or four segments per quadrant usually result in acceptable accuracy. The extension of this point of view to a piecewise-polynomial approximation is evident.

ANALYTIC SOLUTION OF THE DF INTEGRAL

A basis for the approximate analytic evaluation of the DF consists in the development of an approximate expansion of the exact DF integral formulation (Refs. 36, 51). Consider the case of a single-valued characteristic, for which the DF is given by

$$N(A) = \frac{2}{\pi A} \int_{-\pi/2}^{\pi/2} y(A \sin \psi) \sin \psi d\psi \quad (2.6-1)$$

Under the transformation

$$u = \sin \psi \quad du = \cos \psi d\psi \quad (2.6-2)$$

Eq. (2.6-1) can be rewritten as

$$N(A) = \frac{2}{\pi A} \int_{-1}^1 y(Au) \frac{u du}{\sqrt{1-u^2}} \quad (2.6-3)$$

The evaluation of a related integral is

$$\int_{-1}^1 \frac{g(u) du}{\sqrt{1-u^2}} \approx \frac{\pi}{6} [g(1) + g(-1) + 2g(0.5) + 2g(-0.5)] \quad (2.6-4)$$

This result can be demonstrated as exact for the case of $g(u)$ as the series expansion

$$g(u) = a_2 u^2 + a_3 u^3 + a_4 u^4 + a_5 u^5$$

which, by virtue of the required connecting identity

$$g(u) = uy(Au) \quad (2.6-5)$$

implies a nonlinear characteristic of the form

$$y(u) = b_1 u + b_2 u^2 + b_3 u^3 + b_4 u^4$$

From Eqs. (2.6-4) and (2.6-5) it follows that the DF is given by

$$N(A) \approx \frac{1}{3A} \left[y(A) - y(-A) + y\left(\frac{A}{2}\right) - y\left(-\frac{A}{2}\right) \right] \quad (2.6-6)$$

For characteristics which are odd, this reduces to

$$N(A) \approx \frac{2}{3A} \left[y(A) + y\left(\frac{A}{2}\right) \right] \quad (2.6-7)$$

The significance of these simple results can be demonstrated by example.

Example 2.6-1 (a) Find the DF for the odd polynomial characteristic

$$y(x) = c_1 x + c_2 x |x| + c_3 x^3 + c_4 x^3 |x| + c_5 x^5$$

Applying the formula obtained above, we get

$$N(A) \approx c_1 + \frac{5}{6} c_2 A + \frac{3}{4} c_3 A^2 + \frac{1}{2} c_4 A^3 + \frac{1}{10} c_5 A^4$$

The exact DF [Eq. (2.3-21)] is repeated for convenience:

$$N(A) = c_1 + \frac{8}{3\pi} c_2 A + \frac{3}{4} c_3 A^2 + \frac{32}{15\pi} c_4 A^3 + \frac{5}{8} c_5 A^4$$

In analyzing the result, first observe that the terms in c_1 and c_3 are exact. This is expected because of the formulation of $g(u)$. The terms in c_2 and c_4 are in error by less than 4 percent. Since $g(u)$ does not imply any odd terms of even power in the nonlinear characteristic, these results are encouraging. The term in c_5 is in error by 10 percent.

(b) Find the DF for an ideal-relay characteristic

$$y(x) = D \frac{x}{|x|}$$

By Eq. (2.6-7) we get the result

$$N(A) \approx \frac{4D}{3A}$$

The exact result is

$$N(A) = \frac{4D}{\pi A}$$

from which we observe that only a 4.5 percent error has been incurred in using the approximate DF calculation. This excellent result can be explained in part by noting that the same approximate DF formulation as in Eqs. (2.6-6) and (2.6-7) is obtained for characteristics given by

$$y(u) = \frac{b_{-1}}{u} + b_0 + b_1 u + b_2 u^2$$

where the leading term is seen to be discontinuous (more precisely, *undefined*) at the origin. Again the result is encouraging.

Another approximation to $N(A)$, derived in a manner similar to that presented above and yielding even better accuracies, is (odd nonlinearity, see Ref. 46)

$$N(A) \approx \frac{1}{3A} \left[y(A) + y\left(\frac{A}{2}\right) + \sqrt{3} y\left(\frac{\sqrt{3}A}{2}\right) \right] \quad (2.6-8)$$

By application to Example 2.6-1 it is verified that approximation errors using

this result are smaller than those previously obtained. Odd polynomials of odd exponent up to 9 may be treated with zero error. Calculation of the DF for saturation by either Eq. (2.6-7), the first approximation, or (2.6-8), the second approximation, yields errors of less than 5 percent, the latter with smaller error over most of the range considered. Approximately computed DFs for saturation, increasing gain, relay with dead zone, and harmonic nonlinearities are shown with their exact counterparts in Fig. 2.6-1.

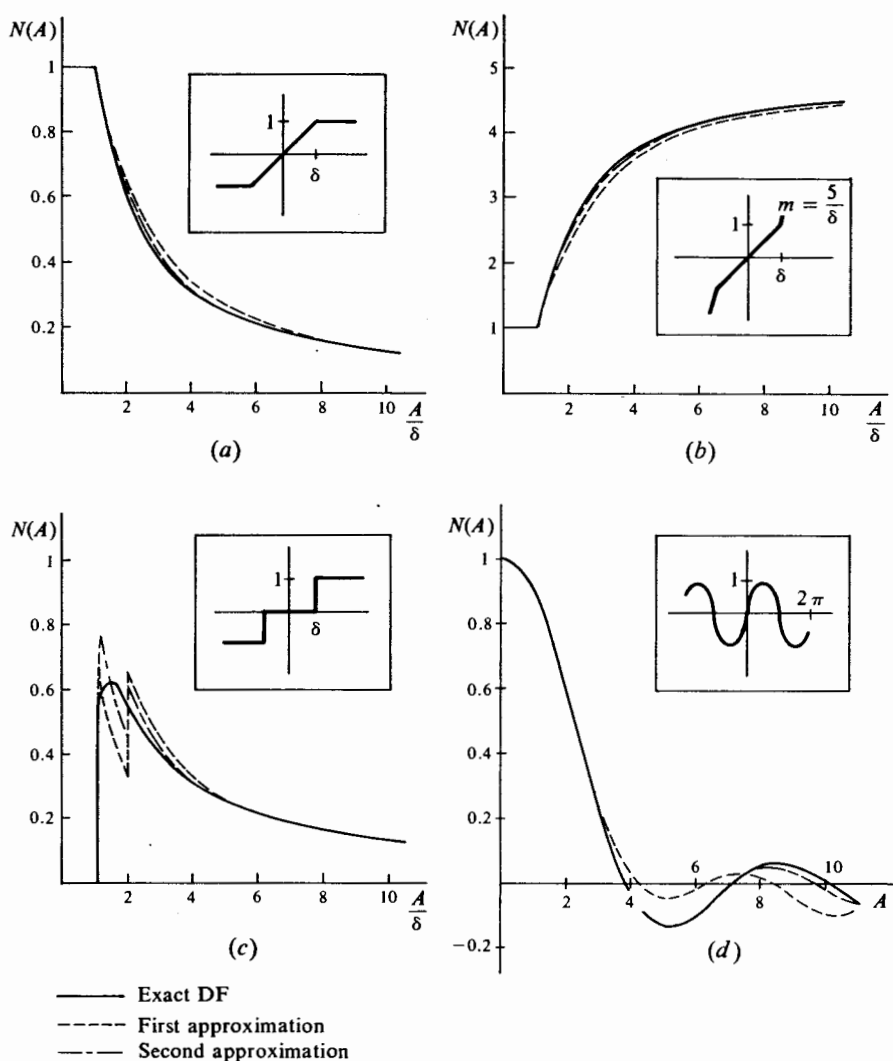


Figure 2.6-1 Approximately computed DFs for (a) saturation, (b) increasing gain, (c) relay with dead zone, and (d) harmonic nonlinearity.

The approximate DF solution presented in this section has been generalized to encompass arbitrary nonlinear characteristics with memory. Allowing the double-valued $y(x)$ to be separated into $y_1(x)$, valid when x is decreasing, and $y_2(x)$, valid when x is increasing, it has been shown (Ref. 38) that

$$n_p(A) \approx \frac{1}{6A} \left[y_1(A) + y_2(A) - y_1(-A) - y_2(-A) + y_1\left(\frac{A}{2}\right) + y_2\left(\frac{A}{2}\right) - y_1\left(-\frac{A}{2}\right) - y_2\left(-\frac{A}{2}\right) \right] \quad (2.6-9a)$$

and

$$n_q(A) \approx -\frac{1}{6\pi A} \left[y_1(A) - y_2(A) + 4y_1\left(\frac{A}{2}\right) - 4y_2\left(\frac{A}{2}\right) + 2y_1(0) - 2y_2(0) + y_1(-A) - y_2(-A) + 4y_1\left(-\frac{A}{2}\right) - 4y_2\left(-\frac{A}{2}\right) \right] \quad (2.6-9b)$$

For symmetric characteristics [$y_1(x) = -y_2(-x)$] these expressions become

$$n_p(A) \approx \frac{1}{3A} \left[y_1(A) + y_2(A) + y_1\left(\frac{A}{2}\right) + y_2\left(\frac{A}{2}\right) \right] \quad (2.6-10a)$$

and

$$n_q(A) \approx -\frac{1}{3\pi A} \left[y_1(A) - y_2(A) + 4y_1\left(\frac{A}{2}\right) - 4y_2\left(\frac{A}{2}\right) + 2y_1(0) \right] \quad (2.6-10b)$$

These results reduce to Eq. (2.6-7) in the case of memoryless characteristics.

NUMERICAL SOLUTION OF THE DF INTEGRAL

For simplicity, the following development is limited to *odd* nonlinear characteristics which may have memory. The DF can be written as

$$\begin{aligned} N(A) &= \frac{2}{\pi A} \int_0^\pi y(A \sin \psi) \sin \psi \, d\psi + j \frac{2}{\pi A} \int_0^\pi y(A \sin \psi) \cos \psi \, d\psi \\ &= \frac{2}{\pi A} \int_{\psi=0}^{\psi=\pi} y(A \sin \psi) d(-\cos \psi) + j \frac{2}{\pi A} \int_{\psi=0}^{\psi=\pi} y(A \sin \psi) d(\sin \psi) \end{aligned} \quad (2.6-11)$$

These integrals can be approximated by finite sums. In so doing, it is helpful to write separately the integrals over two ranges in ψ , from 0 to $\pi/2$ and from

$\pi/2$ to π . Thus

$$\begin{aligned}
 N(A) &= \frac{2}{\pi A} \left[\int_{\psi=0}^{\psi=\pi/2} y(A \sin \psi) d(-\cos \psi) + \int_{\psi=\pi/2}^{\psi=\pi} y(A \sin \psi) d(-\cos \psi) \right] \\
 &\quad + j \frac{2}{\pi A} \left[\int_{\psi=0}^{\psi=\pi/2} y(A \sin \psi) d(\sin \psi) + \int_{\psi=\pi/2}^{\psi=\pi} y(A \sin \psi) d(\sin \psi) \right] \\
 &\approx \frac{2}{\pi A} \left[\sum_{i=1}^{n/2} y(A \sin \psi_i) \delta(-\cos \psi_i) + \sum_{i=n/2+1}^n y(A \sin \psi_i) \delta(-\cos \psi_i) \right] \\
 &\quad + j \frac{2}{\pi A} \left[\sum_{i=1}^{n/2} y(A \sin \psi_i) \delta(\sin \psi_i) + \sum_{i=n/2+1}^n y(A \sin \psi_i) \delta(\sin \psi_i) \right]
 \end{aligned} \tag{2.6-12}$$

It has been assumed that an n -term summation is performed, with n even. Taking equal increments of $\delta(-\cos \psi_i)$ in the n_p integration, and equal increments of $\delta(\sin \psi_i)$ in the n_q integration, Eq. (2.6-12) can be rewritten as

$$\begin{aligned}
 N(A) &\approx \frac{2}{\pi A} |\delta(-\cos \psi)| \left[\sum_{i=1}^{n/2} y(A \sin \psi_i) + \sum_{i=n/2+1}^n y(A \sin \psi_i) \right] \\
 &\quad + j \frac{2}{\pi A} |\delta(\sin \psi)| \left[\sum_{i=1}^{n/2} y(A \sin \psi_i) - \sum_{i=n/2+1}^n y(A \sin \psi_i) \right]
 \end{aligned} \tag{2.6-13}$$

The minus sign in the second set of brackets results from the fact that the sign of the increment $\delta(\sin \psi_i)$ is negative for ψ increasing from $\pi/2$ to π (that is, i increasing from $n/2 + 1$ to n).

Both of the uniform increment magnitudes, $|\delta(-\cos \psi)|$ and $|\delta(\sin \psi)|$, can be seen to have the value $2/n$. Calling

$$u_i = \sin \psi_i$$

we can write the compact expression

$$N(A) \approx \frac{2}{\pi A} \frac{2}{n} \sum_{i=1}^{n/2} y(Au_i) + j \frac{2}{\pi A} \frac{2}{n} \sum_{i=1}^n S_i y(Au_i) \tag{2.6-14}$$

which is the result sought. The term S_i accounts for the sign change in the n_q integration; viz.,

$$S_i = \begin{cases} +1 & \text{for } 1 \leq i \leq \frac{n}{2} \\ -1 & \text{for } \frac{n}{2} + 1 \leq i \leq n \end{cases}$$

Observe that u_i goes from 0 to 1 and back to 0, as ψ varies from 0 to $\pi/2$ to π . The values of u_i give the points at which $y(Au)$ is evaluated. These

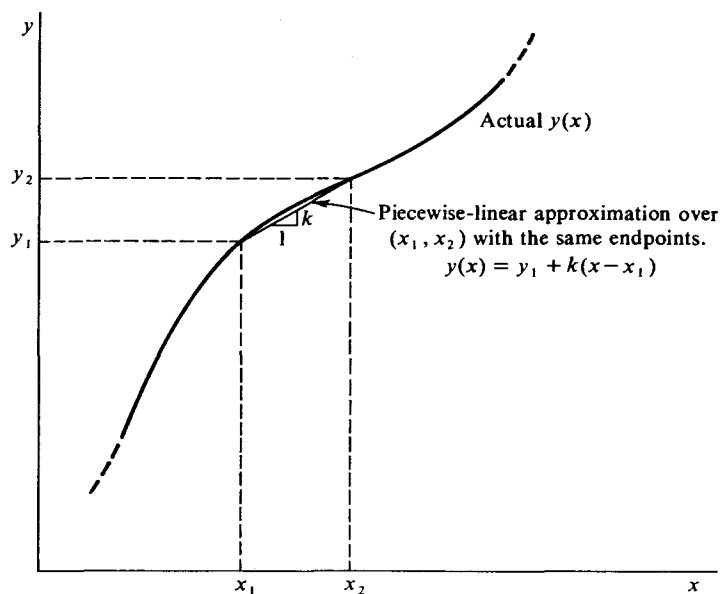


Figure 2.6-2 Approximation for numerical integration over one segment of the nonlinear characteristic.

points should be chosen so that the summations of Eq. (2.6-14) are good approximations to the corresponding integrals of Eq. (2.6-11). One reasonable way to choose the u_i is to require that the approximation to each integral be exact for an "average" function, perhaps a function which is linear over each summation interval and has the same endpoints (Fig. 2.6-2). The result of this choice for the u_i is just the well-known trapezoidal integration rule.

Consider the terms relevant to the n_q integration. From Eq. (2.6-11) we have, for a typical summation interval (u_1, u_2) ,

$$\begin{aligned} \int_{u_1}^{u_2} y(Au) du &\approx \int_{u_1}^{u_2} [y_1 + kA(u - u_1)] du \\ &= (u_2 - u_1)(y_1 - kAu_1) + \frac{kA}{2} (u_2^2 - u_1^2) \end{aligned}$$

and from Eq. (2.6-14) the corresponding term is

$$|\delta u| y(Au_i) = (u_2 - u_1)[y_1 + kA(u_i - u_1)]$$

Equating these quantities yields

$$u_i = \frac{u_1 + u_2}{2} \quad \text{for the } n_q \text{ integration}$$

A parallel development for the n_p integration results in the determination that¹

$$u_i = \frac{\sin^{-1} u_2 - \sin^{-1} u_1 - u_2 \sqrt{1 - u_2^2} + u_1 \sqrt{1 - u_1^2}}{2(\sqrt{1 - u_1^2} - \sqrt{1 - u_2^2})}$$

for the n_p integration

In each case u_i is the value of u , intermediate to u_1 and u_2 , at which $y(Au_i)$ is evaluated. The values of u_i for $n = 10$ and $n = 20$ are given in Table 2.6-1.

TABLE 2.6-1 VALUES OF u_i FOR DF DETERMINATION BY TRAPEZOIDAL INTEGRATION

n_p integration		n_q integration	
$n = 10$	$n = 20$	$n = 10$	$n = 20$
$u_i = 0.4088$	$u_i = 0.2936$	$u_i = 0.10$	$u_i = 0.05$
0.7095	0.5239	0.30	0.15
0.8634	0.6600	0.50	0.25
0.9520	0.7590	0.70	0.35
0.9933	0.8344	0.90	0.45
	0.8924		0.55
	0.9362		0.65
	0.9678		0.75
	0.9882		0.85
	0.9983		0.95

These integration formulas are particularly easy to use with a desk calculator since no multiplication is required to scale the values of $y(Au_i)$, which are read from a graph of $y(x)$ or computed. Using 10 increments per unity interval in u in each summation, one can expect DF computation accuracies of better than 5 percent.

PLACING BOUNDS ON THE DF

In any DF calculation a convenient independent check is certainly desirable. The general shape of the DF as a function of A can be easily determined by inspection of the nonlinear characteristic in question. Quantitative estimates, we shall see, are also feasible.

Consider an arbitrary nonlinear function $y(x)$ for which we seek DF information, and two associated curves $y_u(x)$ and $y_i(x)$. The latter curves are

¹ This is readily shown with the aid of the identities:

$$\int \frac{u \, du}{\sqrt{1 - u^2}} = -\sqrt{1 - u^2} \quad \int \frac{u^2 \, du}{\sqrt{1 - u^2}} = \frac{1}{2}(\sin^{-1} u - u\sqrt{1 - u^2})$$

required to bound $y(x)$ from above and below, respectively, but are otherwise arbitrary. Odd single-valued characteristics are assumed for convenience. The following inequality then holds for all x in the range of interest.

$$y_l(x) < y(x) < y_u(x) \quad \text{for } 0 < x < A_1 \quad (2.6-15)$$

where A_1 defines the highest value of x of interest. This implies the integral inequality

$$\int_0^{\pi/2} y_l(A \sin \psi) \sin \psi \, d\psi < \int_0^{\pi/2} y(A \sin \psi) \sin \psi \, d\psi < \int_0^{\pi/2} y_u(A \sin \psi) \sin \psi \, d\psi \quad (2.6-16)$$

which may be rewritten as the DF inequality

$$N_l(A) < N(A) < N_u(A) \quad (2.6-17)$$

where $N_l(A)$ and $N_u(A)$ are the DFs belonging to $y_l(x)$ and $y_u(x)$, respectively. Hence bounds may be placed on $N(A)$ provided only that $N_l(A)$ and $N_u(A)$ are known.

The most convenient functions $y_u(x)$ and $y_l(x)$ are simply straight lines passing through the origin. If the lines have slopes k_u and k_l , then we have, simply,

$$k_l < N(A) < k_u \quad (2.6-18)$$

Figure 2.6-3 illustrates the application of this simple bounding process to a typical nonlinear characteristic. Stronger bounds on $N(A)$ can be obtained by choosing better upper and lower nonlinearity bounding functions.

2.7 DF INVERSION

Given the DF data $n_p(A)$ and $n_q(A)$, one may be interested in a nonlinearity to which this corresponds. This inverse DF question is of interest in the evaluation of a physical device whose harmonic response has been experimentally obtained, and in the synthesis of a nonlinear compensatory network whose DF has been prescribed. It is intuitively clear that since the DF is an incomplete measure of a nonlinear characteristic, DF inversion is necessarily nonunique. For example, one can readily find nonlinear characteristics with memory which possess DFs identical with other characteristics with no memory (cf. Prob. 2-13). Below, we consider several cases where $n_q(A) = 0$ and the recovered memoryless nonlinearity is odd. The extension to asymmetric nonlinearities is straightforward, although of less interest. For a treatment of the more general case in which $n_q(A) \neq 0$, the reader is referred to Gibson et al. (Ref. 13), where an analytic DF inversion solution is developed.

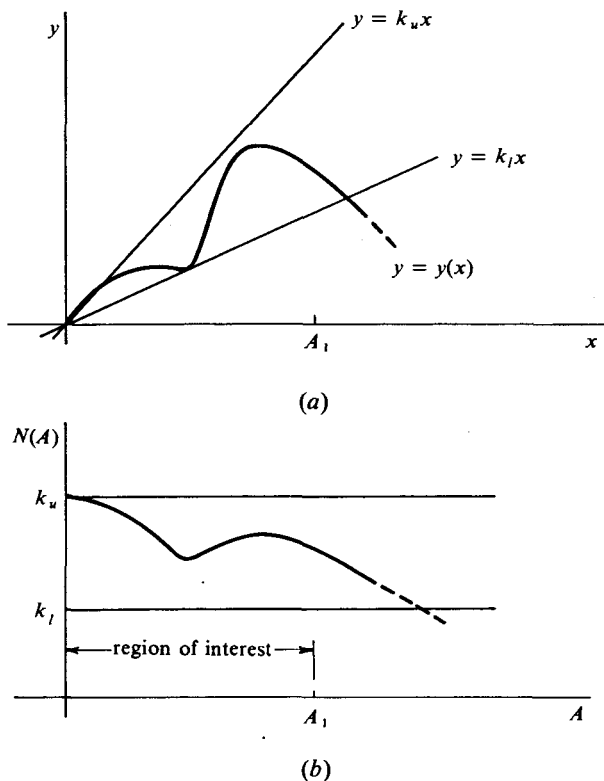


Figure 2.6-3 Bounding of the DF (b) by forming a straight-line envelope for the nonlinear characteristic (a).

POLYNOMIAL-BASED DF INVERSION

Assume that a polynomial curve fit to $N(A)$ has been accomplished. The result is

$$N(A) = \sum_{k=0}^n a_k A^k \quad (2.7-1)$$

In Sec. 2.3 it was demonstrated that for an odd-polynomial n th-order nonlinearity the DF is given by

$$N(A) = \frac{2}{\pi} \sum_{k=1}^n c_k A^{k-1} \frac{\Gamma[(k+2)/2]}{\Gamma[(k+3)/2]} \quad (2.7-2)$$

By comparing Eqs. (2.7-1) and (2.7-2) we immediately establish the result

$$c_k = \frac{\pi}{2} \frac{\Gamma[(k+3)/2]}{\Gamma[(k+2)/2]} a_{k-1} \quad (2.7-3)$$

which defines an odd-polynomial nonlinearity.

PIECEWISE-LINEAR-BASED DF INVERSION

In Sec. 2.3 it was shown that the DF for an odd n -segment, piecewise-linear memoryless nonlinearity can be written in the form

$$N(A) = \frac{4M}{\pi A} + \sum_{i=1}^{n-1} (m_i - m_{i+1}) f\left(\frac{\delta_i}{A}\right) + m_n \quad (2.7-4)$$

If $M \neq 0$, assume that it is removed from Eq. (2.7-4) by a curve fit in the vicinity of $A \approx 0$. The nonlinear characteristic to be determined thus has a discontinuity of value $2M$ at the origin. Now let the DF-data abscissa be divided into equal subdivisions of width δ . By computing $N(A)$ for $A = \delta/2, 3\delta/2, 5\delta/2$, etc., the following set of equations results from the repeated application of Eq. (2.7-4):

$$\begin{aligned} N\left(\frac{\delta}{2}\right) &= m_1 \\ N\left(\frac{3\delta}{2}\right) &= (m_1 - m_2)f\left(\frac{\delta}{3}\right) + m_2 \\ N\left(\frac{5\delta}{2}\right) &= (m_1 - m_2)f\left(\frac{\delta}{5}\right) + (m_2 - m_3)f\left(\frac{\delta}{5}\right) + m_3 \\ &\dots \end{aligned} \quad (2.7-5)$$

This set is easily solved by applying the solution for m_1 from the first equation to the second, then using the resulting solution for m_2 in the third, and so forth. For sufficiently small δ , the nonlinearity so determined will closely represent the required nonlinearity. If the nonlinearity for which the DF was determined actually was piecewise-linear, and if δ happens to be chosen at a value which is some integral fraction of each breakpoint, δ_i , of the actual nonlinearity, the piecewise-linear function determined by means of Eq. (2.7-5) will be precisely the actual nonlinearity.

APPROXIMATE DF INVERSION

By repeated application of the approximate DF formula [Eq. (2.6-7)] it is possible to secure an approximate DF inversion.

$$\begin{aligned} N(A) &\approx \frac{2}{3A} \left[y(A) + y\left(\frac{A}{2}\right) \right] \\ N(2A) &\approx \frac{1}{3A} [y(2A) + y(A)] \\ N(4A) &\approx \frac{1}{6A} [y(4A) + y(2A)] \\ &\dots \end{aligned} \quad (2.7-6)$$

Starting with the first equation above, and inserting in it the values for $y(A)$, $y(2A)$, $y(4A)$, etc., appearing in subsequent equations, we easily generate the result

$$\begin{aligned} y\left(\frac{A}{2}\right) &\approx \frac{3A}{2} N(A) - y(A) \\ &\approx \frac{3A}{2} N(A) - 3AN(2A) + y(2A) \\ &\approx \frac{3A}{2} N(A) - 3AN(2A) + 6AN(4A) - y(4A) \\ &\dots \end{aligned} \quad (2.7-7)$$

which is expressible as

$$y(A) \approx 3A \sum_{i=0}^{\infty} (-2)^i N(2^{i+1}A) \quad (2.7-8)$$

Had we instead employed the relationships

$$\begin{aligned} N(A) &\approx \frac{2}{3A} \left[y(A) + y\left(\frac{A}{2}\right) \right] \\ N\left(\frac{A}{2}\right) &\approx \frac{4}{3A} \left[y\left(\frac{A}{2}\right) + y\left(\frac{A}{4}\right) \right] \\ N\left(\frac{A}{4}\right) &\approx \frac{8}{3A} \left[y\left(\frac{A}{4}\right) + y\left(\frac{A}{8}\right) \right] \\ &\dots \end{aligned} \quad (2.7-9)$$

and proceeded as in Eq. (2.7-7), the corresponding result would be

$$y(A) \approx \frac{3A}{2} \sum_{i=0}^{\infty} \left(-\frac{1}{2}\right)^i N\left(\frac{A}{2^i}\right) \quad (2.7-10)$$

Equations (2.7-8) and (2.7-10) are the approximate DF inversion formulas sought. The former is applied to a given set of DF data when the values of $N(A)$ decrease with increasing values of A , thus leading to a rapidly converging series for $y(A)$. In the opposite case, Eq. (2.7-10) is employed. In either case the resulting function is $y(x)$.

Example 2.7-1 Given the DF data $N(A) = 4D/\pi A$, employ Eq. (2.7-10) to find a corresponding $y(x)$.

In this case one must proceed with some care to ensure that the series expansion for $y(A)$ converges. Consider the replacement of Eq. (2.7-10) by

$$y(A) \approx \frac{3A}{2} \lim_{\epsilon \rightarrow 1} \sum_{i=0}^{\infty} \left(-\frac{\epsilon}{2}\right)^i N\left(\frac{A}{2^i}\right) \quad (2.7-11)$$

Inserting the given DF data, we now get

$$\begin{aligned}
 y(A) &\approx \frac{3A}{2} \lim_{\epsilon \rightarrow 1} \sum_{i=0}^{\infty} \left(-\frac{\epsilon}{2}\right)^i \frac{4D}{\pi A} (2^i) \\
 &= \frac{6D}{\pi} \lim_{\epsilon \rightarrow 1} \sum_{i=0}^{\infty} (-\epsilon)^i \\
 &= \frac{6D}{\pi} \lim_{\epsilon \rightarrow 1} \frac{1}{1 + \epsilon} \\
 &= \frac{3D}{\pi}
 \end{aligned}$$

The reason for the particular form chosen for Eq. (2.7-11) is at once evident, since without it (and the understanding that ϵ approaches 1 from below), the resulting infinite series would have been the alternating series $1 - 1 + 1 - 1 + 1 - \dots$, etc.

Interpreting the result as an odd single-valued nonlinearity, it follows that the characteristic is described by

$$y(x) = \begin{cases} \frac{3D}{\pi} & \text{for } x \geq 0 \\ -\frac{3D}{\pi} & \text{for } x < 0 \end{cases}$$

which the reader from the outset knew to be an ideal-relay characteristic. The inversion is thus in error by 4.5 percent. Aside from cases where $N(A)$ is inversely proportional to A (such as in this example), no special precaution need be taken, and Eqs. (2.7-8) and (2.7-10) can be freely employed.

Up to this point we have developed the facility to calculate DFs for a very wide class of nonlinearities. In the following chapter we turn our attention to the use of these DFs in the study of nonlinear feedback systems.

REFERENCES

1. Andronow, A. A., and C. E. Chaikin: "Theory of Oscillations," Princeton University Press, Princeton, N.J., 1949.
2. Bhattacharyya, B. P.: Describing-function Expressions for Sine-type Functional Nonlinearity in Feedback Control Systems, *J. IEE*, vol. B-108, no. 41 (September, 1961), pp. 529-534.
3. Casserly, G., and J. G. Truxal: Measurement and Stabilization of Nonlinear Feedback Systems, *IRE Conv. Record*, pt. 4 (1956), pp. 52-60.
4. Choudhoury, A. K., M. S. Basu, and A. K. Mahalanabis: A Transfer Function Analyzer for Linear and Nonlinear Components, *Electron. Eng.*, London, vol. 33 (June, 1961), pp. 382-385.
5. Clegg, J. C.: A Nonlinear Integrator for Servomechanisms, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 77 (March, 1958), pp. 41-42.

6. Cunningham, W. J.: "Introduction to Nonlinear Analysis," McGraw-Hill Book Company, New York, 1958.
7. Deekshatulu, B. L.: A Graphical Method of Evaluating the Describing Function, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 81 (July, 1962), pp. 101-106.
8. Dotort, I. K.: Harmonic Analysis by Direct Area Measurements, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (March, 1956), pp. 16-19.
9. Douce, J. L.: A Note on the Evaluation and Response of a Non-linear Element to Sinusoidal and Random Signals, *J. IEE* (London), vol. C-105 (October, 1957), p. 88.
10. Dutilh, J.: Théorie des servomécanismes à relais, *Onde Élec.*, vol. 30 (1950), pp. 438-445.
11. Freeman, E. A.: The Effect of Speed-dependent Friction and Backlash on the Stability of Automatic Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 78 (January, 1959), pp. 680-691.
12. Gibson, J. E., and K. S. Prasanna-Kumar: A New RMS Describing Function for Single-valued Nonlinearities, *Proc. IRE*, vol. 49 (August, 1961), p. 1321.
13. Gibson, J. E., E. G. di Tada, J. C. Hill, and E. S. Ibrahim: Describing Function Inversion: Theory and Computational Techniques, *Purdue Univ., Control Inform. Systems Lab. TR-EE62-1D*, Lafayette, Ind., December, 1962.
14. Gilie, J. C., M. J. Pelegrin, and P. Decaulne: "Feedback Control Systems," McGraw-Hill Book Company, New York, 1959.
15. Goldfarb, L. C.: On Some Nonlinear Phenomena in Regulatory System, *Avtomatika i Telemekhanika*, vol. 8, no. 5 (1947), pp. 349-383 (in Russian). Translation: U. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956.
16. Graham, D., and D. McRuer: "Analysis of Nonlinear Control Systems," John Wiley & Sons, Inc., New York, 1961.
17. Grief, H. D.: Describing Function of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (September, 1953), pp. 243-248.
18. Grobner, W., and N. Hoffreiter: "Integraltafel," 3d ed., Springer-Verlag OHG, Berlin, 1961.
19. Gronner, A. D.: The Describing Function of Backlash Followed by a Dead Zone, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 77 (November, 1958), pp. 403-409.
20. Haas, V. B., Jr.: Coulomb Friction in Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (May, 1953), pp. 119-126.
21. Hamos, L. v.: Beitrag zur Frequenzanalyse von nichtlinearen Systemen, "Regelungstechnik: Moderne Theorien und ihre Verwendbarkeit," R. Oldenbourg KG, Munich, 1957, pp. 211-215 (in German).
22. Hartel, W.: Abschätzung der Oberwellen bei gleichstromvormagnetisierten Drosseln, *Arch. Elektrotech.*, vol. 36 (1942), p. 556.
23. Jahnke, E., and F. Emde: "Tables of Functions," 4th ed., Dover Publications, Inc., 1945.
24. Jopling, A. D., and R. A. Johnson: The Elliptic Describing Function, *Proc. Joint Autom. Control Conf.*, Stanford, Calif., June, 1964, pp. 175-184.
25. Kavanagh, R. J.: An Approximation to the Harmonic Response of Saturating Devices, *J. IEE* (London), vol. C-107, no. 11 (January, 1960), pp. 127-133.
26. Klotter, K.: An Extension of the Conventional Concept of the Describing Function, *Proc. Symp. Nonlinear Circuit Analysis*, Polytechnic Institute of Brooklyn, New York, April, 1956, pp. 151-162.
27. Kochenburger, R. J.: Analysis and Synthesis of Contactor Servomechanisms, Sc.D. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, Cambridge, Mass., 1949.

28. Kochenburger, R. J.: Limiting in Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (July, 1953), pp. 180-194.
29. Krylov, N., and N. Bogoliubov: "Introduction to Nonlinear Mechanics," Princeton University Press, Princeton, N.J., *Annals of Mathematics Studies*, vol. 11, 1947.
30. Ku, Y. H., and C. F. Chen: A New Approach to the Evaluation and Interpretation of Describing Functions, *AIEE Paper* CP61-1084, September, 1961.
31. Lakshmi-Bai, C.: Transformation Method for the Study of Non-linear Components, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 79 (September, 1960), pp. 249-254.
32. Lauber, R.: A New Method to Derive the Describing Function of Certain Nonlinear Transfer Systems, *Proc. IFAC*, Basel, Switzerland, August, 1963, pp. 188/1-188/5.
33. Levinson, E.: Non-linear Feedback Control Systems, pt. 3, *Electro-Technology*, vol. 72 (September, 1962), pp. 136-146.
34. Levinson, E.: Gain-Phase Relations of Nonlinear Circuits, *IRE Conv. Record*, pt. IV, 1958, pp. 141-159.
35. Lory, H. L., D. C. Lai, and W. H. Huggins: On the Use of Growing Harmonic Exponentials to Identify Static Nonlinear Operators, *IRE Trans. Autom. Control*, vol. AC-4, no. 2 (November, 1959), pp. 91-99.
36. Magnus, K.: "Über ein Verfahren zur Untersuchung nicht-linearer Schwingungs- und Regelungs-Systeme," VDI-Verlag GmbH, Düsseldorf, Germany, 1955. Translated into English by the Langley Research Center, NASA, Hampton, Va., October, 1958.
37. Makow, D. M.: Harmonic Analysis and Describing Function of Nonlinear Systems, *IRE Intern. Conv. Record*, pt. 4, 1961, pp. 69-76.
38. McBride, L. E., Jr.: Approximate Describing Function for Double-valued Nonlinearities, *Electronics Letters*, London, vol. 1, no. 3 (May, 1965).
39. Mishkin, E., and L. Braun, Jr.: "Adaptive Control Systems," McGraw-Hill Book Company, New York, 1961.
40. Neilsen, K. L.: "Methods in Numerical Analysis," The Macmillan Company, New York, 1956.
41. Oppelt, W.: Locus Curve Method for Regulators with Friction, *Z. Deut. Ingr.*, Berlin, vol. 90 (1948), pp. 179-183. Translated in *Report* 1691, National Bureau of Standards, Washington, 1952.
42. Poincaré, H.: "Les méthodes nouvelles de la mécanique céleste," vol. 1, Gauthier-Villars, Paris, 1892; reprinted by Dover Publications, Inc., New York, 1957.
43. Rankine, R. R., Jr., and J. J. D'Azzo: A Corrected-conventional Describing Function for Intermediate-order Nonlinear Control Systems, *Proc. Northeast Electronics Res. Eng. Meeting*, November, 1964, pp. 38-39.
44. Satyendra, K. N.: Describing Functions Representing the Effect of Inertia, Backlash and Coulomb Friction on the Stability of an Automatic Control System, I, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (September, 1956), pp. 243-249.
45. Semmelhack, H. P., and P. M. DeRusso: A Nonsinusoidal Describing Function for Stiction and Coulomb Friction, *IEEE Paper* CP63-921.
46. Silberberg, M. Y.: A Note on the Describing Function of an Element with Coulomb, Static, and Viscous Friction, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 76 (January, 1957), pp. 423-425.
47. Sridhar, R.: A General Method for Deriving the Describing Functions for a Certain Class of Nonlinearities, *IRE Trans. Autom. Control*, vol. AC-5, no. 2 (June, 1960), pp. 135-141.
48. Stein, W. A., and G. J. Thaler: Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-sensitive Describing Function, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 77 (May, 1958), pp. 91-96.
49. Stoker, J. J.: "Non-linear Vibrations in Mechanical and Electrical Systems," Interscience Publishers, Inc., New York, 1950.

50. Stout, T. M.: Block Diagram Transformations for Systems with One Nonlinear Element, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (July, 1956), pp. 130-141.
51. Tsyppin, Ya. Z.: Discussion of a paper by J. C. West, in "Regelungstechnik: Moderne Theorien und ihre Verwendbarkeit," R. Oldenbourg KG, Munich, 1957, pp. 163-164 (in German).
52. Tustin, A.: The Effects of Backlash and of Speed Dependent Friction on the Stability of Closed-cycle Control Systems, *J. IEE* (London), vol. 94, pt. II (May, 1947), pp. 143-151.
53. von Wyss, M. R.: The Describing Function for Saturated Amplification in Series with Saturated Integration, *SAAB Tech. Note* 46, Saab Aircraft Company, Linköping, Sweden, 1960.
54. Watson, G. N.: "A Treatise on the Theory of Bessel Functions," Cambridge University Press, New York, 1922, chap. 10.
55. West, J. C.: The Describing Function Technique Applied to Non-linear Systems, Part I, *Process Control and Automation*, vol. 7 (September, 1960), pp. 421-426.
56. Zaborsky, J., and H. J. Harrington: Describing Function for Hydraulic Valves, *Trans. AIEE*, pt. I, vol. 76 (1957), pp. 183-198.
57. Zakharov, K. V., and V. K. Svyatodukh: An Instrument for Determining the Frequency Characteristics of Nonlinear Systems, *Automation and Remote Control*, vol. 21 (June, 1960), pp. 1630-1637.

PROBLEMS

- 2-1. Show by use of the perturbation method that the first-order corrected fundamental component of the solution to the problem of a mass on a nonlinear spring

$$\ddot{x} + \omega_0^2 x + \mu x^3 = 0$$

subject to the initial conditions

$$x(0) = A_0 \quad \dot{x}(0) = 0$$

is given by

$$x = A_0 \left(1 - \frac{\mu A_0^2}{32\omega_0^2} \right) \cos \omega t$$

$$\omega^2 = \omega_0^2 + \frac{3\mu A_0^2}{4}$$

- 2-2. Show by use of the method of slowly varying amplitude and phase that the solution to Prob. 2-1 is

$$x = A_0 \cos \omega t$$

$$\omega = \omega_0 + \frac{3\mu A_0^2}{8\omega_0}$$

- 2-3. By use of the method of Krylov and Bogoliubov, show that the approximate transient solution to Van der Pol's equation

$$\ddot{x} - \epsilon(1 - x^2)\dot{x} + x = 0$$

is given by

$$x = \frac{A_0 e^{(\epsilon/2)t}}{\sqrt{1 + (A_0^2/4)(e^{\epsilon t} - 1)}} \sin(t + \theta_0)$$

where A_0 and θ_0 define the initial state of the system. Discuss the relationship between the system steady state and its initial conditions.

- 2-4. Derive the DF for an odd quantizer nonlinearity with arbitrarily spaced breakpoints δ_i and arbitrary output levels D_i .
- 2-5. Derive the overall DF for the following chain of elements: ideal relay-integrator-ideal relay. The integrator gain is K , and the relay drive levels are $\pm D$.
- 2-6. By noting that (rotary) coulomb friction is a torque-velocity phenomenon, show that the DF relating input rotation *angle* to output torque is

$$N(A, \omega) = \frac{4T}{\pi A} e^{(j\pi/2)}$$

where T is the coulomb-friction magnitude, and A is the input-angle amplitude.

Using this result, find the DF (angle input-angle output) for spring-coupled coulomb friction (massless case), and compare it with that for spring-coupled viscous friction.

What is the DF for inertia with coulomb friction (torque input-angle output)?

- 2-7. A useful technique for obtaining DFs for functionally characterized nonlinearities is to generate the power series of these functions in order to apply DF results already developed for polynomials. Using this technique, find DFs for the following nonlinear characteristics, and discuss the range of validity of each:

(a) $y = \tanh^{-1} x$

(b) $y = 1 - e^{-|x|}$

(c) $y = \frac{\sin x}{|x|}$

(d) $y = x \cos x$

- 2-8. Obtain the DF and third-harmonic amplitude ratio for the nonlinear characteristic of Fig. 2-1, such as might correspond to an optical-tracker error signal as a function of tracker misalignment in a satellite attitude control system. (Hint: Approximations may prove expedient.)

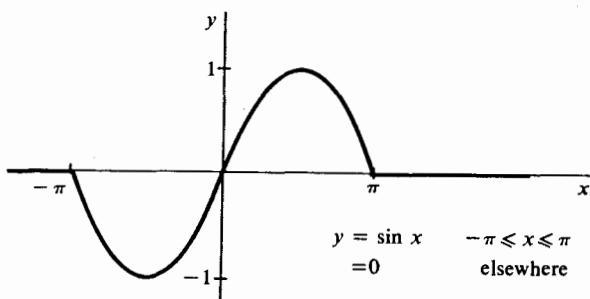


Figure 2-1 Optical-tracker error characteristic.

2-9. In determining the DF for a *dynamic* element of the form

$$y = f(\dot{x})$$

is it legitimate to simply observe the block-diagram equivalence of Fig. 2-2, and hence conclude the relationship

$$N_{\dot{x}}(A, \omega) = sN_x(A)$$

where $N_x(A)$ is the DF for the *static* nonlinearity $f(x)$?

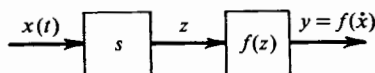


Figure 2-2 Series dynamic nonlinearity.

2-10. Find the DF for the parallel-connected dynamic nonlinearity system of Fig. 2-3. How would you recommend testing this two-port system to experimentally determine parameters δ_1 , D_1 , δ_2 , and D_2 , knowing a priori the form of the nonlinearity?

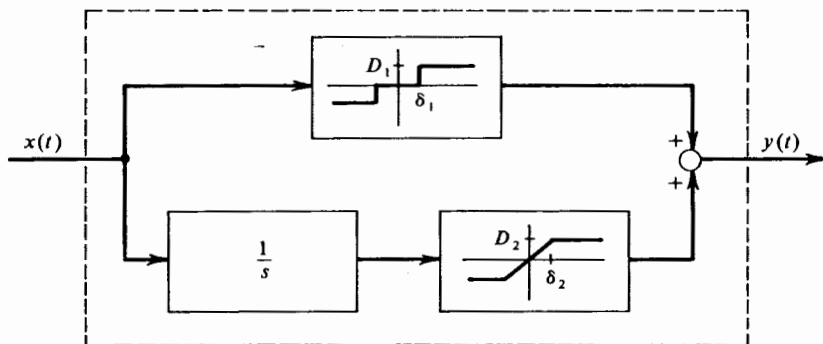


Figure 2-3 Parallel dynamic nonlinearity.

2-11. The nonlinear element described by

$$\ddot{y} + \dot{y}^3 = x$$

is imbedded in a larger system to be analyzed by DF methods. Find the DF for this nonlinear element.

2-12. Find and graph the DF for the nonlinear element of Fig. 2-4. By also plotting DFs for the two limiters (drawn dotted), show that the original DF is bounded above and below.

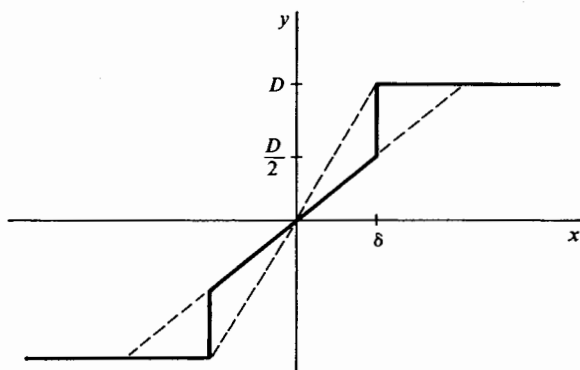


Figure 2-4 Nonlinearity with linear zone and abrupt limiting.

- 2-13. (a) Derive the DF for the nonlinear characteristic with memory of Fig. 2-5, and hence demonstrate that it is non-phase-shifting.
 (b) Synthesize the same DF by parallel addition of several memoryless nonlinearities, and hence demonstrate that DF inversion is indeed not unique.

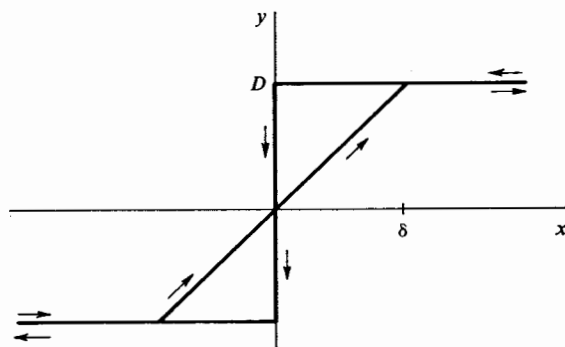


Figure 2-5 Nonlinear characteristic with memory.

- 2-14. In a certain torque feedback servo loop, the torquer input current is derived from a current supply with a piecewise-linear limiter characteristic (input breakpoints $\pm\delta$ and output saturation levels $\pm D$). The torque generator has an odd square-law characteristic from input current to output torque. Show that the DF from current-supply input command to torque-generator output torque is

$$N(A) = \frac{8}{3\pi} \left(\frac{D}{\delta} \right)^2 A \left\{ 1 - \left[1 - \left(\frac{\delta}{A} \right)^2 \right]^{\frac{3}{2}} \right\}$$

2-15. Show that the exact DF for the nonlinear chain illustrated in Fig. 2-6 is given by

$$N(A, \omega) = \begin{cases} \frac{4D_1KD_2}{\pi\delta\omega A} \sin\left(\frac{\delta\omega}{D_1K}\right) e^{-j\pi/2} & \text{for } \frac{\delta\omega}{D_1K} \leq \frac{\pi}{2} \\ \frac{4D_1KD_2}{\pi\delta\omega A} e^{-j\pi/2} & \text{for } \frac{\delta\omega}{D_1K} > \frac{\pi}{2} \end{cases}$$

and compare this with the result obtained by multiplying the DFs of the individual elements of the chain. Under what conditions does DF multiplication yield an acceptable solution?

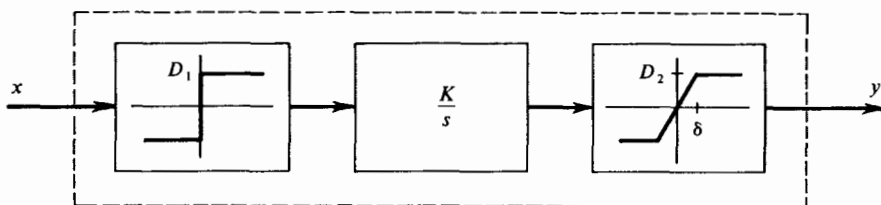


Figure 2-6 Dynamic multiple-nonlinearity chain.

2-16. One form of slowly saturating nonlinear characteristic is governed by Froelich's equation

$$y = \frac{x}{1 + c|x|}$$

Plot the appropriately normalized DF for this characteristic, and compare it with the DF for a limiter with the same slope at the origin and the same peak output level. Discuss the shape of characteristic and DF for the nonlinearity

$$y = \left(\frac{x}{1 + c|x|} \right)^n$$

for values $n < 1$ and $n > 1$.

2-17. The Chebyshev polynomial of order k is given by

$$T_k(x) = \cos(k \cos^{-1} x)$$

(a) Show that the k th-harmonic amplitude in the output of a sinusoidally forced static single-valued nonlinearity $y(x)$ can be written in the form ($k \geq 1$)

$$A_k = \frac{2}{\pi} \int_{-A}^A y(x) T_k\left(\frac{x}{A}\right) \frac{dx}{\sqrt{A^2 - x^2}}$$

where A is the input amplitude.

(b) Next, develop a method¹ for recovering the nonlinear characteristic from its

¹ Described by D. P. Atherton: Determination of Nonlinearity from the Harmonic Response, *Electronics Letters*, London, vol. 2, no. 4 (April, 1966), p. 152.

harmonic response by expanding the nonlinearity in the form

$$y(x) = \sum_{l=0}^{\infty} \alpha_l T_l\left(\frac{x}{A}\right)$$

and using the orthogonality relationship

$$\frac{2}{\pi} \int_{-A}^A T_l\left(\frac{x}{A}\right) T_k\left(\frac{x}{A}\right) \frac{dx}{\sqrt{A^2 - x^2}} = \begin{cases} 1 & \text{for } k = l \\ 0 & \text{for } k \neq l \end{cases}$$

2-18. Prove the relationship

$$\text{Im} [N(A)] = -\frac{S}{\pi A^2}$$

where S is the area enclosed by $y(x)$ as x varies from zero to A to $-A$ and back to zero, and thereby establish from this point of view that memoryless nonlinearities possess only real DFs. (*Hint*: A contour integral may be helpful.)

2-19. Show that the rms DF and corrected conventional DF for a relay with dead-zone nonlinearity are given by ($A > \delta$)

$$N_{\text{rms}}(A) = \frac{D}{A} \sqrt{\frac{2}{\pi} \left(\pi - 2 \sin^{-1} \frac{\delta}{A} \right)}$$

$$N_{\text{ccDF}}(A) = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A} \right)^2} \sqrt{1 + \frac{1}{9} \left[1 - \left(\frac{2\delta}{A} \right)^2 \right]^2}$$

2-20. Show that the DF for a flip-flop with first-order dynamic response is given by (see Fig. 2-7):

$$N(A, \omega) = \frac{4D}{\pi A} \frac{1}{1 + j\omega\tau}$$

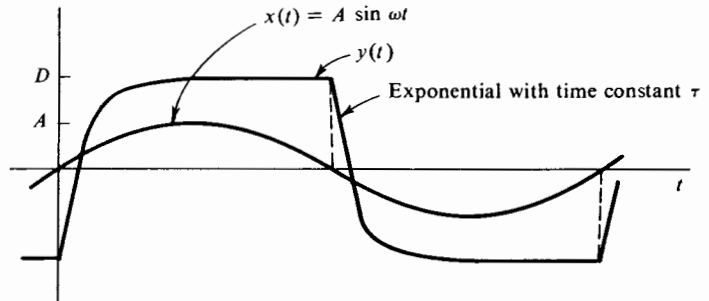


Figure 2-7 Flip-flop response to input sinusoid.

3 STEADY-STATE OSCILLATIONS IN NONLINEAR SYSTEMS

3.0 INTRODUCTION

The preceding chapter introduced the notion of a sinusoidal-input describing function (DF). Some of the implications of this type of *linearization* are discussed there. Here we apply the DF to the study of steady-state oscillations. For DF utilization to be meaningful, certain conditions must be fulfilled by the nonlinearity, and also by the system in which the nonlinearity is present:

1. The nonlinear element is time-invariant.¹
2. No subharmonics are generated by the nonlinearity in response to a sinusoidal input.
3. The system filters nonlinearity output harmonics to the extent that only a trivial quantity is fed back.

Condition 3 is the so-called *filter hypothesis*. We shall have a great deal to say about it in this chapter, for this heuristically motivated requirement is certainly a basic factor underlying DF success or failure.

Certain periodically varying nonlinearities are also admissible.

Three types of steady-state oscillations are of interest:

1. Forced oscillations
2. Conservative free oscillations
3. Limit cycles

For our present purpose a *forced* oscillation is taken to be a systematic response whose frequency is precisely the forcing signal frequency and whose amplitude depends on the forcing signal amplitude. Forced oscillations are encountered in the study of frequency response, in which connection we find a useful application for the DF. The next two oscillation types are behavioral modes of unforced systems. A *conservative free oscillation* is an *initial condition-dependent* periodic mode associated with nondissipative (conservative) systems. A *continuous* range of conservative-free-oscillation amplitudes and frequencies may be possible in a given system. *Limit cycle* denotes an *initial condition-independent* periodic mode occurring in dissipative (nonconservative) systems. Only a *discrete* set of limit cycle amplitudes and frequencies may exist in a given system. We shall find the DF a most powerful tool for the study of both types of unforced behavior of nonlinear systems.

3.1 DETERMINATION OF LIMIT CYCLES

The limit cycle phenomenon is deserving of special attention since it is apt to occur in any physical nonlinear system. A limit cycle can be desirable, for example, by providing the vibration (dither) which minimizes frictional effects in mechanical systems. In fact, it can be absolutely necessary for proper system performance, as we shall see when studying a certain adaptive control scheme in Chap. 6. On the other hand, a limit cycle can cause mechanical failure of a control system (destructive limit cycle) or operator discomfort and other undesirable effects, as in an aircraft autopilot.

Although precise knowledge of the waveform of a limit cycle is usually not mandatory, knowledge of the *existence* of a limit cycle, as well as its approximate amplitude and frequency, is a prerequisite to good system design.

THE DF LINEARIZED SYSTEM

Consider a physical nonlinear system as represented by Fig. 3.1-1a. To effect a limit cycle study, the nonlinear element is characterized by its DF, and the linear element by its frequency response function $L(j\omega)$. $N(A, \omega)$ is used to describe the effect of the nonlinearity on the closed-loop system as we are about to seek the presence of an oscillation which is sinusoidal at

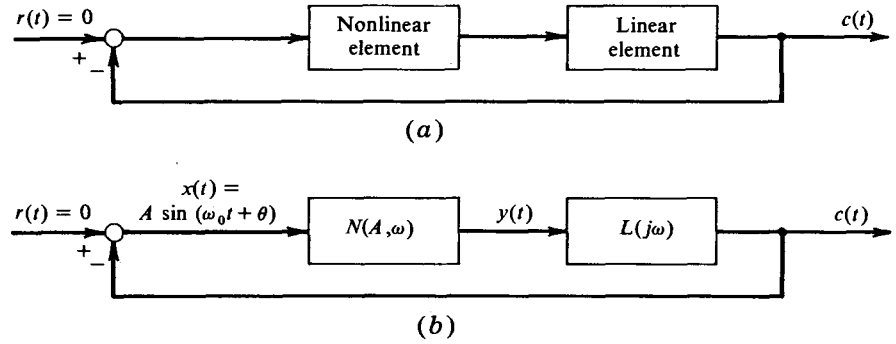


Figure 3.1-1 (a) Single-nonlinearity closed-loop system. (b) Corresponding linearized system for DF limit cycle study.

station x due to the filtering qualities of the loop linear part. This yields the quasi-linearized system of Fig. 3.1-1b, which, for constant A , is a constant-coefficient linear system.

Linear theory is now applied to the quasi-linearized system, and points of neutral stability are sought. Any undamped oscillations so arrived at are interpreted as limit cycles in the original nonlinear system.

The equation defining undamped oscillations of the DF linearized system is readily obtained. Since the input is taken to be zero, the following loop relationship is apparent:

$$X(j\omega) = -L(j\omega)Y(j\omega) \quad (3.1-1)$$

Treating the DF as a linearized gain between x and y gives

$$Y(j\omega) = N(A, \omega)X(j\omega) \quad (3.1-2)$$

Writing these two equations in matrix notation results in

$$\begin{bmatrix} N(A, \omega) & -1 \\ 1 & L(j\omega) \end{bmatrix} \begin{bmatrix} X(j\omega) \\ Y(j\omega) \end{bmatrix} = 0$$

which possesses a nontrivial solution only if the determinant of the square matrix is zero. Thus

$$\begin{vmatrix} N(A, \omega) & -1 \\ 1 & L(j\omega) \end{vmatrix} = 0$$

or

$$1 + N(A, \omega)L(j\omega) = 0 \quad (3.1-3)$$

Solutions of this equation yield the amplitudes and frequencies of the loop limit cycles.

Before proceeding to both algebraic and graphical solutions of Eq. (3.1-3), we digress to present a different method, which centers attention directly on the characteristic differential equation of the DF linearized system.

ANALYTIC ROUTH-HURWITZ TEST

The Routh-Hurwitz stability criterion for linear systems is well known (cf. Refs. 21, 29, 104). It is a means for determining conditions under which the linear system is neutrally stable. Equation (3.1-3), written as a polynomial in descending powers of s , is the characteristic equation to which this criterion is presently applied. In the case of a third-order system we have, for example,

$$s^3 + a_2s^2 + a_1s + a_0 = 0 \quad (3.1-4)$$

The stability boundary is given by

$$a_1a_2 - a_0 = 0 \quad (3.1-5)$$

provided that a_0 , a_1 , and a_2 are positive. We now have one relationship in the unknowns A and ω_0 . The remaining relationship is a result of the fact that the characteristic equation must contain an undamped second-order factor, the limit cycle. Hence we have

$$(s^2 + \omega_0^2)(s + b) = 0 \quad (3.1-6)$$

Expansion yields

$$s^3 + bs^2 + \omega_0^2s + b\omega_0^2 = 0$$

whence we observe by comparison with Eq. (3.1-4) that

$$a_1 = \omega_0^2 \quad (3.1-7)$$

This is the second of the required relationships. The limit cycle parameters can now be determined algebraically.

For non-phase-shifting nonlinearities Eq. (3.1-5) is a function of A alone. Once having solved for A , ω_0 follows immediately from Eq. (3.1-7). In order to use this method with phase-shifting nonlinearities, the proportional plus derivative DF formulation [Eq. (2.2-30)] must be employed. Solution of the resulting equations in this case, however, is apt to require graphical assistance since Eq. (3.1-5) is no longer an equation in the single variable A . Simultaneous nonlinear algebraic equations must be dealt with. Table 3.1-1 summarizes the results of this technique as applied to second-, third-, and fourth-order characteristic equations. All a_i must be positive.

The application of this technique for limit cycle determination is demonstrated by the following example.

**TABLE 3.1-1 ROUTH-HURWITZ
LIMIT CYCLE DETERMINATION**

Characteristic equation	Limit cycle equations
$s^2 + a_1s + a_0 = 0$	$a_1 = 0$ $\omega_0^2 = a_0$
$s^3 + a_2s^2 + a_1s + a_0 = 0$	$a_1a_2 - a_0 = 0$ $\omega_0^2 = a_1$
$s^4 + a_3s^3 + a_2s^2 + a_1s + a_0 = 0$	$a_1a_2a_3 - a_1^2 - a_0a_3^2 = 0$ $\omega_0^2 = \frac{a_1}{a_3}$

Example 3.1-1 Use the Routh-Hurwitz test to determine all possible limit cycles of the relay control system of Fig. 3.1-2.

The DF for a relay with dead zone, given by Eq. (2.3-14), is repeated below for convenience ($A > \delta$).

$$n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} \quad n_q = 0$$

Thus the characteristic equation of interest, $1 + N(A)L(s) = 0$, is representable as

$$s^3 + 2\zeta\omega_n s^2 + \omega_n^2 s + \frac{4\omega_n^2 KD}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} = 0$$

The coefficients a_i are given by

$$a_2 = 2\zeta\omega_n \quad a_1 = \omega_n^2 \quad a_0 = \frac{4\omega_n^2 KD}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2}$$

It follows from the second entry in Table 3.1-1 that the limit cycle frequency is exactly the undamped natural frequency of the second-order part of $L(s)$,

$$\omega_0 = \omega_n$$

and, additionally, that

$$(2\zeta\omega_n)(\omega_n^2) - \frac{4\omega_n^2 KD}{\pi A_0} \sqrt{1 - \left(\frac{\delta}{A_0}\right)^2} = 0$$

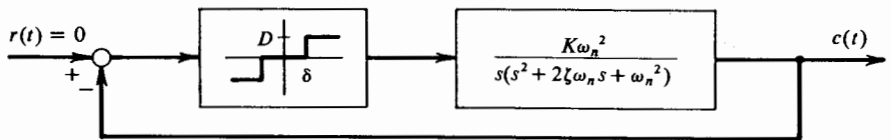


Figure 3.1-2 Example relay control system.

The last of these equations may be solved for the possible limit cycle amplitudes A_0 .

$$A_0 = \frac{\sqrt{2KD}}{\pi\zeta\omega_n} \sqrt{1 \pm \sqrt{1 - \left(\frac{\pi\zeta\omega_n\delta}{KD}\right)^2}}$$

It is worthwhile to make several observations regarding these results. First, since the limit cycle amplitude must be a real quantity, we see that no limit cycle is possible for δ in the range

$$\delta > \frac{KD}{\pi\zeta\omega_n} \quad \text{condition for no limit cycle}$$

Second, when δ is not in the above range, *two* limit cycles are possible, corresponding to the two choices of sign under the radical in the solution found for A_0 . We shall see in Sec. 3.2 that of these two limit cycle states, only one is stable.

ALGEBRAIC SOLUTION OF THE LIMIT CYCLE EQUATION

An algebraic solution of Eq. (3.1-3) can be effected directly by expressing $N(A, \omega)$ in its complex form.

$$1 + [n_p(A_0, \omega_0) + jn_q(A_0, \omega_0)]\{\text{Re}[L(j\omega_0)] + j\text{Im}[L(j\omega_0)]\} = 0$$

This relationship is satisfied when the real and imaginary parts on both sides are equated. The following pair of equations result.

$$\text{Real:} \quad n_p(A_0, \omega_0) \text{Re}[L(j\omega_0)] - n_q(A_0, \omega_0) \text{Im}[L(j\omega_0)] = -1 \quad (3.1-8a)$$

$$\text{Imaginary:} \quad n_p(A_0, \omega_0) \text{Im}[L(j\omega_0)] + n_q(A_0, \omega_0) \text{Re}[L(j\omega_0)] = 0 \quad (3.1-8b)$$

These two equations in the two unknowns defining the limit cycle are solved simultaneously. When the linear elements are of low order and the DF expressions are relatively simple, analytic solutions for A_0 and ω_0 are readily achieved.

Example 3.1-2 Find the limit cycles displayed by the relay control system of Fig. 3.1-2 by analytically employing Eqs. (3.1-8a) and (3.1-8b).

The DF for a relay with dead-zone nonlinearity, given by Eq. (2.3-14), is repeated here for convenience ($A > \delta$):

$$n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} \quad n_q = 0$$

The real and imaginary parts of $L(j\omega)$ are

$$\text{Re}[L(j\omega)] = \frac{-2K\zeta\omega_n^3}{(\omega_n^2 - \omega^2)^2 + (2\zeta\omega\omega_n)^2} \quad \text{Im}[L(j\omega)] = \frac{-K\omega_n^2(\omega_n^2 - \omega^2)}{\omega[(\omega_n^2 - \omega^2)^2 + (2\zeta\omega\omega_n)^2]}$$

From Eq. (3.1-8b) we conclude that

$$\text{Im}[L(j\omega_0)] = 0$$

This implies that the limit cycle frequency is equal to the natural frequency of the second-order part of $L(j\omega)$.

$$\omega_0 = \omega_n$$

Inserting this value into Eq. (3.1-8a) gives

$$\frac{K}{2\zeta\omega_n} \frac{4D}{\pi A_0} \sqrt{1 - \left(\frac{\delta}{A_0}\right)^2} = 1$$

which has two solutions:

$$A_0 = \frac{\sqrt{2KD}}{\pi\zeta\omega_n} \sqrt{1 \pm \sqrt{1 - \left(\frac{\pi\zeta\omega_n\delta}{KD}\right)^2}}$$

These results are identical with those derived in Example 3.1-1.

An alternative formulation of Eq. (3.1-3) in terms of magnitudes and phase angles rather than real and imaginary parts is

$$\text{Magnitude:} \quad |N(A_0, \omega_0)| |L(j\omega_0)| = 1 \quad (3.1-9a)$$

$$\text{Angle:} \quad \angle N(A_0, \omega_0) + \angle L(j\omega_0) = -\pi - 2n\pi \quad (3.1-9b)$$

for $n = 0, 1, 2, \dots$. This formulation is clearly equivalent to that of Eqs. (3.1-8a) and (3.1-8b), although one or the other may be more convenient to apply to a specific system.

Example 3.1-3 Determine the limit cycle amplitude and frequency for the nonlinear time-lag differential equation

$$\dot{x}(t) + \delta x(t - T) = \epsilon x^3(t - T)$$

which occurs in the study of cycles in the shipbuilding industry (Ref. 78), nonlinear oscillators with time delay, and certain process control applications.

The block diagram for this system is shown in Fig. 3.1-3. Note that the time-lag behavior is simply represented by the factor e^{-sT} , the Laplace transform of an ideal delay. The equation for a limit cycle is

$$\frac{1}{j\omega_0} e^{-j\omega_0 T} (\delta - \epsilon \frac{3}{4} A_0^2) = -1$$

where the DF for a cubing operator, $N(A) = \frac{3}{4} A^2$, has been employed. Writing angle and magnitude conditions, we get

$$-\frac{\pi}{2} - \omega_0 T = -\pi - 2n\pi$$

which results in the determination that

$$\omega_0 = (2n + \frac{1}{2}) \frac{\pi}{T} \quad \text{for } n = 0, 1, 2, \dots$$

and

$$\frac{1}{\omega_0} (1)(\delta - \epsilon \frac{3}{4} A_0^2) = 1$$

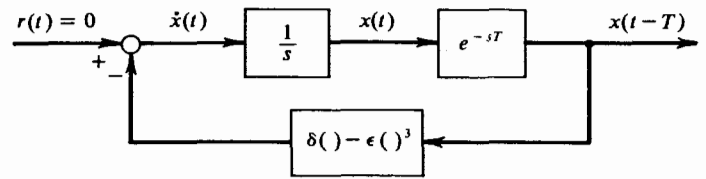


Figure 3.1-3 Nonlinear time-lag system.

which implies

$$A_0 = \sqrt{\frac{4}{3\epsilon} \left[\delta - (2n + \frac{1}{2}) \frac{\pi}{T} \right]}$$

Hence, for values of δ in excess of $\pi/2T$, a limit cycle is predicted. Experimental results substantiate this lower bound. We shall encounter this example once again in Chap. 4. At that point a more detailed discussion of limit cycle behavior is presented.

For systems with linear elements of high order or with relatively complex DFs, and particularly when various nonlinearities are to be studied, the graphical procedure described next is usually found the most expedient means to limit cycle determination.

GRAPHICAL LIMIT CYCLE DETERMINATION

The closed-loop characteristic equation for linear systems with a variable-gain parameter K is usually written in the form

$$KL(j\omega) = -1 \quad (3.1-10)$$

where the *fixed* point $(-1, 0)$ takes on the significance of the stability boundary demarcation. Polar plots, Bode plots, gain-phase plots, and (although more indirectly) root-locus plots of familiar linear theory are all graphical mechanisms to focus attention on the relative positions of the system linear elements' locus and the stability point. Equation (3.1-10) may be written, alternatively, as

$$L(j\omega) = -\frac{1}{K} \quad (3.1-11)$$

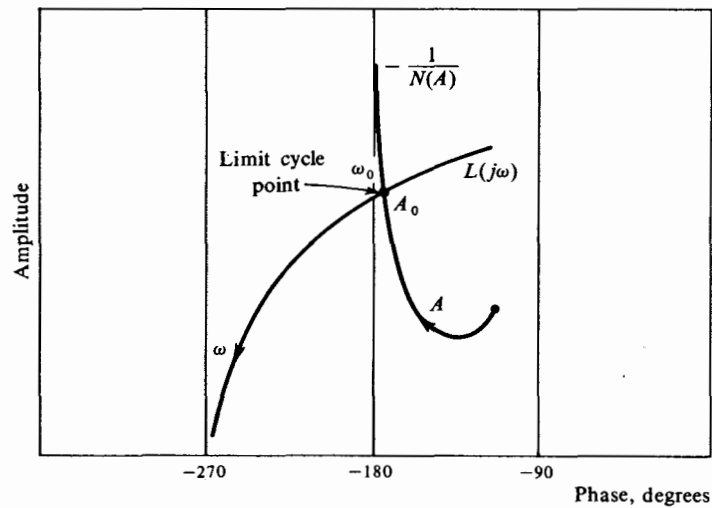
in which case the stability point is $(-1/K, 0)$, a *variable* quantity depending upon the gain K . In this form the effect of the parameter K on the system is singled out. Similarly, a particularly useful form for the characteristic equation of a nonlinear system [Eq. (3.1-3)] is¹

$$L(j\omega) = -\frac{1}{N(A, \omega)} \quad (3.1-12)$$

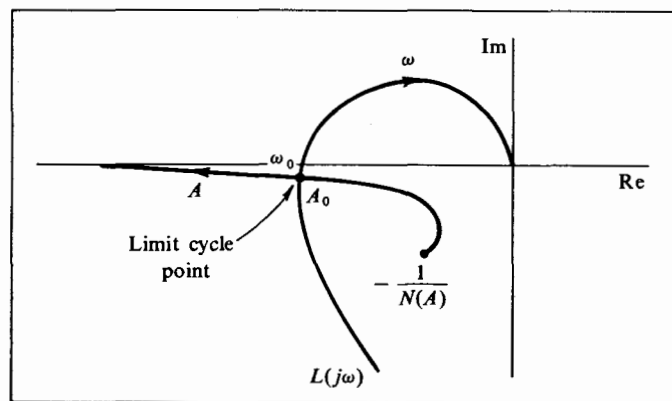
¹ As pointed out by Gibson (Ref. 29, p. 348), this analogy is not to be viewed as an extension of the Nyquist criterion, it being based on a contour integration around the right half-plane. Equation (3.1-12) has been arrived at by other reasoning.

In concept, $N(A, \omega)$ replaces the linear gain K , although it may be both amplitude- and frequency-dependent.

Whereas satisfaction of the equality of Eq. (3.1-11) implies an undamped second-order oscillation, corresponding satisfaction of Eq. (3.1-12) indicates a limit cycle. By plotting $L(j\omega)$ and $-1/N(A, \omega)$ loci on the same set of



(a)



(b)

Figure 3.1-4 Graphical limit cycle determination. $L(j\omega)$ is third-order, and $N(A)$ is a phase-shifting function (corresponding to a relay with hysteresis). Amplitude-phase plot (a) and polar plot (b) representations. A_0 and ω_0 are the limit cycle amplitude and frequency, respectively.

coordinates, possible limit cycle solutions are made to occur as the intersections of two curves. Since one curve represents the loop linear elements and the other represents the nonlinearity, the variation of limit cycle amplitude and frequency with parameter changes of either the linear or nonlinear elements is available with a minimum amount of work by simply altering the appropriate curve.

Limit cycle behavior can be studied in either amplitude-phase (gain-phase), polar, or root-locus coordinates. No limit cycle information exists in any one of these coordinate systems which is not also present in the other two. Since the root-locus construction for phase-shifting nonlinearities is more cumbersome than either of the corresponding gain-phase or polar-plot constructions, it is abandoned here as a working tool.

Typical limit cycle determinations involve constructions such as are shown in Fig. 3.1-4, where third-order linear elements and a phase-shifting DF have been assumed. Arrows indicate the direction of increasing A on the $-1/N(A)$ locus, and increasing ω on the linear elements' frequency locus. The frequency of the indicated limit cycle is read from the calibrated $L(j\omega)$ locus and its amplitude is read from the calibrated $-1/N(A)$ locus.

Whether the polar or amplitude-phase plot is used for this construction is a matter of individual preference. It is the authors' preference to use the amplitude-phase-plane representation, in part because of the general availability of templates for use in linear compensation, and at least in some part because it seems to be a more appealing data presentation.

Example 3.1-4 Graphically identify the two limit cycles found as the solutions to Example 3.1-1.

The nonlinearity is non-phase-shifting; hence $-1/N(A)$ lies along the -180° line on the amplitude-phase plot. From Fig. B.1 we observe that $N(A)$ is not a monotonic function of A ; it peaks at exactly $A = \sqrt{2}\delta$. For this reason the amplitude of the $-1/N(A)$ locus starts at $+\infty$, comes down to a minimum at $A = \sqrt{2}\delta$, and returns to $+\infty$, as A increases from zero. In order to illustrate this behavior, the plot of $-1/N(A)$ has been distorted, as shown in Fig. 3.1-5.

Both predicted limit cycles are evident as the intersections of $L(j\omega)$ and $-1/N(A)$. The possibility of no limit cycle is also apparent from this illustration. All that is required is a choice of parameters such that the curves do not intersect at all. As an exercise the reader may formulate this condition from the geometry of Fig. 3.1-5 and compare the result with that calculated in Example 3.1-1.

Limit cycle determination in systems with amplitude- and frequency-dependent DFs is, in principle, quite simple. In practice, the graphical procedure can be tedious. Basically, what one does is to plot several representative members of the family $N(A, \omega)$, each at a *fixed* value of ω . Intersections between $L(j\omega)$ and the various $-1/N(A, \omega)$ loci are examined to determine those for which the frequency *label* of $-1/N(A, \omega)$ and the frequency *calibration* of $L(j\omega)$ match. Such intersections denote the limit

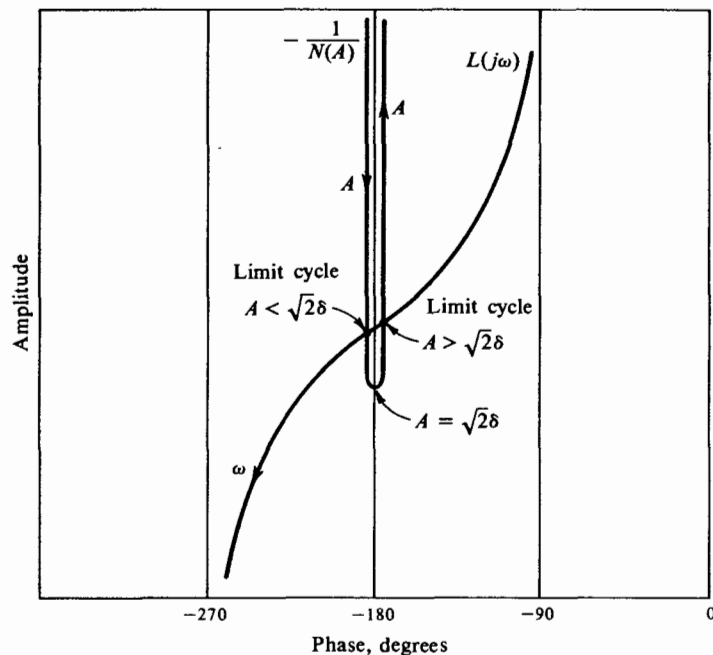


Figure 3.1-5 Determination of the limit cycles of an example relay control system.

cycles. Figure 3.1-6 illustrates limit cycle determination in the case where $L(j\omega)$ is third-order and the nonlinear characteristic is a hypothetical relay with frequency-sensitive rectangular hysteresis. At $\omega = \omega_3$ the frequency markings on both curves coincide; hence that is the limit cycle frequency.

We can now determine all possible limit cycles according to DF theory. As pointed out in the text, some of these are stable, others are not.

3.2 LIMIT CYCLE STABILITY

The question of limit cycle stability is posed in terms of the behavior of a theoretical limit cycle state following amplitude and/or frequency perturbations. If the limit cycle returns to its original equilibrium state it will be called *stable*, whereas if either its amplitude or frequency grows or decays until another equilibrium state is reached, it will be called *unstable*. The possibility of theoretically unbounded amplitude or frequency growth is included in the definition of instability by considering the state at infinity to be an equilibrium state.

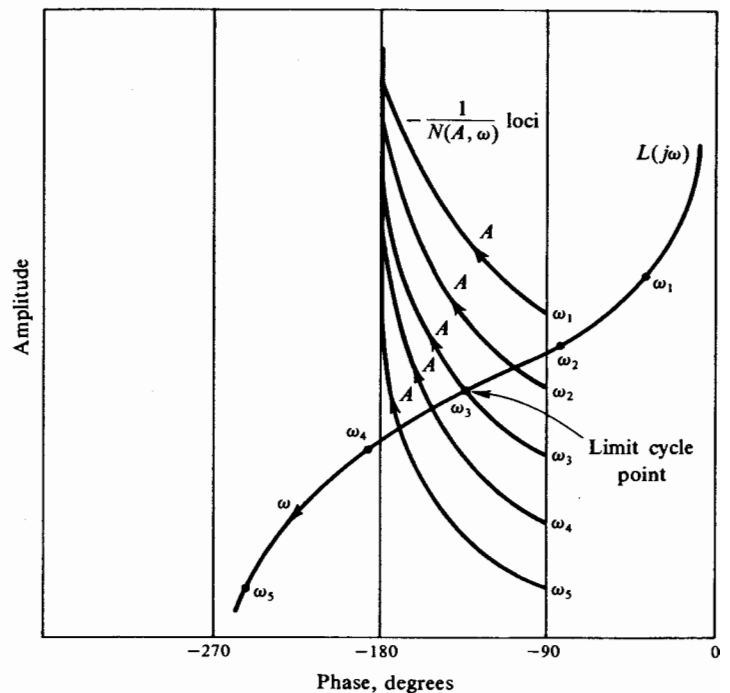


Figure 3.1-6 Graphical limit cycle determination for a frequency-sensitive nonlinearity. $L(j\omega)$ is third-order, and $N(A, \omega)$ is phase-shifting. Limit cycle frequency is ω_3 radians/time.

In terms of the behavior of small perturbations, the question of limit cycle stability is a question about a transient oscillation. In that sense it is most properly treated in Chap. 4. Nevertheless, by making what we shall call “quasi-static” arguments,¹ it will be possible to account for limit cycle stability in the great majority of cases. Analytical and graphical quasi-static stability tests proceed easily within the framework already established.

ANALYTICAL STABILITY DETERMINATION

The stability of a limit cycle under small perturbations about an equilibrium state is readily studied in terms of the DF linearized system characterized by Eq. (3.1-3). We follow Loeb’s discussion (Ref. 61) of a method originally due to Cahen (Ref. 13). Identical results have been obtained from a geometrical argument by Popov (Ref. 88, p. 586).

¹ That is, arguments ignoring terms involving rate of change of frequency and rate of change of damping. See Chap. 4.

Calling A_0 and ω_0 the amplitude and frequency of an equilibrium limit cycle state, respectively, we have seen that the following identity holds:

$$1 + N(A_0, \omega_0)L(j\omega_0) = 0 \quad (3.2-1)$$

The explicit complex form of this equation can be obtained by expressing the quantities $N(A_0, \omega_0)$ and $L(j\omega_0)$ in terms of their real and imaginary parts, resulting in

$$U(A_0, \omega_0) + jV(A_0, \omega_0) = 0 \quad (3.2-2)$$

where the definitions of U and V are evident. Let us now allow small perturbations in the limit cycle amplitude, rate of change of amplitude, and frequency by introducing the following changes in Eq. (3.2-2):

$$\begin{aligned} A_0 &\rightarrow A_0 + \Delta A \\ \omega_0 &\rightarrow \omega_0 + \Delta\omega + j\Delta\sigma \end{aligned} \quad (3.2-3)$$

The perturbation in the rate of change of amplitude has been associated with the frequency term, a device which becomes clear upon thinking of the limit cycle in the form $A_0 \exp(j\omega_0 t)$. That is, $\Delta\sigma = -\dot{A}/A$. Hence we have

$$U(A_0 + \Delta A, \omega_0 + \Delta\omega + j\Delta\sigma) + jV(A_0 + \Delta A, \omega_0 + \Delta\omega + j\Delta\sigma) = 0 \quad (3.2-4)$$

By definition, ΔA , $\Delta\omega$, and $\Delta\sigma$ are small quantities. The Taylor series expansion of Eq. (3.2-4) about the equilibrium state, valid to first-order terms, becomes, after removal of the quiescent terms [Eq. (3.2-2)],

$$\frac{\partial U}{\partial A} \Delta A + \frac{\partial U}{\partial \omega} (\Delta\omega + j\Delta\sigma) + j \frac{\partial V}{\partial A} \Delta A + j \frac{\partial V}{\partial \omega} (\Delta\omega + j\Delta\sigma) = 0 \quad (3.2-5)$$

Satisfaction of this equation requires that its real and imaginary parts separately vanish.

$$\begin{aligned} \frac{\partial U}{\partial A} \Delta A - \frac{\partial V}{\partial \omega} \Delta\sigma + \frac{\partial U}{\partial \omega} \Delta\omega &= 0 \\ \frac{\partial V}{\partial A} \Delta A + \frac{\partial U}{\partial \omega} \Delta\sigma + \frac{\partial V}{\partial \omega} \Delta\omega &= 0 \end{aligned} \quad (3.2-6)$$

Eliminating $\Delta\omega$ from this set of equations yields a single relationship which can be put in the form

$$\left(\frac{\partial U}{\partial A} \frac{\partial V}{\partial \omega} - \frac{\partial U}{\partial \omega} \frac{\partial V}{\partial A} \right) \Delta A = \left[\left(\frac{\partial U}{\partial \omega} \right)^2 + \left(\frac{\partial V}{\partial \omega} \right)^2 \right] \Delta\sigma \quad (3.2-7)$$

For a limit cycle to be stable, a positive increment ΔA must lead to a positive $\Delta\sigma$, and similarly, a negative ΔA must cause a negative $\Delta\sigma$. In other words,

the sign of $\Delta\sigma/\Delta A$ must always be positive for a stable limit cycle condition to exist. Hence, for the proposed limit cycle equilibrium condition to be stable, it is necessary that

$$\frac{\partial U}{\partial A} \frac{\partial V}{\partial \omega} - \frac{\partial U}{\partial \omega} \frac{\partial V}{\partial A} > 0 \quad (3.2-8)$$

Example 3.2-1 Determine the stability of each of the limit cycles found in Example 3.1-1.

The equation to which we shall apply the analytic stability test is, after simplification ($s = j\omega$),

$$-2\zeta\omega_n\omega^2 + \frac{4DK\omega_n^2}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} + j(\omega\omega_n^2 - \omega^3) = 0$$

the real and imaginary parts of which are U and V , respectively. Taking the required four partial derivatives and evaluating them at (A_0, ω_0) results in

$$\begin{aligned} \left. \frac{\partial U}{\partial A} \right|_{A_0, \omega_0} &= -\frac{4DK\omega_n^2}{\pi A_0^3} \left[\frac{1 - 2(\delta/A_0)^2}{\sqrt{1 - (\delta/A_0)^2}} \right] & \left. \frac{\partial V}{\partial A} \right|_{A_0, \omega_0} &= 0 \\ \left. \frac{\partial U}{\partial \omega} \right|_{A_0, \omega_0} &= -4\zeta\omega_n^2 & \left. \frac{\partial V}{\partial \omega} \right|_{A_0, \omega_0} &= -2\omega_n^2 \end{aligned}$$

Inserting these values into Eq. (3.2-8) immediately yields, as the requirement for a stable oscillation, the condition that

$$1 - 2\left(\frac{\delta}{A_0}\right)^2 > 0$$

or equivalently,

$$A_0 > \sqrt{2}\delta$$

We have already ascertained in Example 3.1-4 that, of the two limit cycles which are predicted, one corresponds to $A_0 > \sqrt{2}\delta$ and one to $A_0 < \sqrt{2}\delta$. From the above analysis it is clear that only the larger-amplitude limit cycle is stable.

GRAPHICAL STABILITY DETERMINATION

The stability of a limit cycle state may be assessed graphically, under the same qualifications which apply to the analytic tests. The argument is borrowed from linear-system stability theory, namely, if the limit cycle amplitude perturbation is positive, we require a stable system configuration in which energy is dissipated until the amplitude decays to its unperturbed value; whereas if the amplitude perturbation is negative, we require an unstable system configuration in which energy is absorbed until the amplitude grows to its unperturbed value. This behavior guarantees a locally stable limit cycle state.

Figure 3.2-1 illustrates amplitude-phase, polar, and root-locus plots for a nonlinear system with a single-valued saturating nonlinearity [$N(A)$ is real and decreases with increasing A]. In the root-locus plot the entire quantity

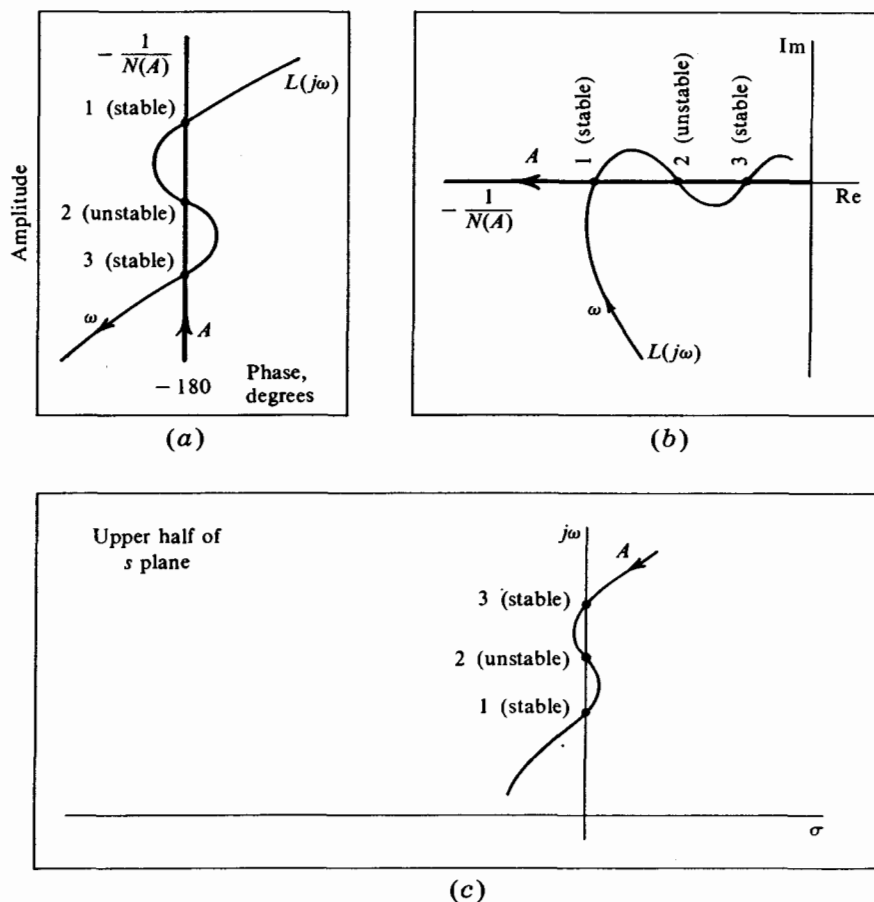


Figure 3.2-1 Amplitude-phase plot (a), polar plot (b), and root-locus plot (c) of a nonlinear system displaying two stable and one unstable limit cycle modes.

$N(A)L(s)$ is dealt with [corresponding to $KL(s)$ in the totally linear case]; the arrow on the root-locus branch illustrates pole migration for increasing A . Limit cycles in this construction are determined by intersections of the locus of roots and the $j\omega$ axis. Now, by the argument described above, we find that limit cycle states 1 and 3 are stable (convergent), whereas limit cycle state 2 is unstable (divergent). Consider the high-frequency intersection, state 3, for example. A positive perturbation in limit cycle amplitude ($A_3 \rightarrow A_3 + \Delta A$) places the stability point $[-1/N(A_3 + \Delta A), 0]$ to the left of the equilibrium state in the polar-plane representation and up in the amplitude-phase-plane representation, and puts a closed-loop root in the left half

of the s plane, each of which would correspond to a stable configuration in the totally linear case. Conversely, a negative perturbation in A_3 leads to an unstable system configuration. Hence, for small perturbations, the resultant limit cycle converges to its original state with increasing time; it is locally *stable* (convergent).

Example 3.2-2 Determine the stability of each of the limit cycles found in Example 3.1-1.

Inspecting pertinent Fig. 3.1-5, it is clear from the argument above that only the larger-amplitude limit cycle oscillation is stable.

Returning to the system of Fig. 3.2-1, we note that if this system is turned on with *zero* initial condition at $c(t)$, the amplitude will grow until state 3 is reached, at which time a stable limit cycle will exist. For positive amplitude perturbations of sufficient magnitude so that state 2 is reached and exceeded, however, the amplitude will eventually converge to state 1 rather than return to state 3. This is a result of the divergent nature of equilibrium state 2.

Physically, limit cycle amplitude perturbations can result from command or disturbance input transients as well as system parameter perturbations. Thus a large transient input to the above system in limit cycle state 3 can cause it to fall into limit cycle state 1. Similarly, a transient decrease in the linear elements leading phase contribution¹ in the frequency band from ω_2 to ω_3 , sufficiently large to temporarily cause only *one* intersection near the frequency ω_1 , will also result in the limit cycle falling from state 3 into state 1, provided that the parameter transient is of sufficient duration. This is an *irreversible* situation, since termination of the parameter transient with the system in limit cycle state 1 will not return it to state 3. Clearly, other situations can be proposed which are *reversible*; that is, removal of the parameter transient restores the original system state. A rigorous study of such situations can be found in Bautin (Ref. 6).

3.3 FREQUENCY RESPONSE OF NON-LIMIT-CYCLING NONLINEAR SYSTEMS

Although the frequency response data of a nonlinear system cannot be inverted as in the linear case to obtain the exact time response to any given input signal (linear superposition is invalid), for weakly nonlinear systems this process will yield useful approximate results. Further, despite the fact that control systems do not generally receive purely sinusoidal inputs, for certain specific cases sinusoidal response data may be quite significant. These

¹ Caused by either an increase of lagging phase or a decrease of leading phase in the constituents of $L(s)$.

include a periodic, although nonsinusoidal, input and a nonlinear subsystem within the control loop of a larger system which is either lightly damped (i.e., whose transient response is highly oscillatory) or which displays a limit cycle. Thus, for one reason or another, the sinusoidal response description of a nonlinear system is desirable. It will be shown that the DF may be used to represent nonlinearity behavior in a harmonically excited system, leading to a frequency response solution.

Frequency response as calculated with the DF has significance only if the system under consideration is not in a limit cycle state. It turns out that certain non-limit-cycling systems break into a limit cycle oscillation when forced by a sinusoidal input; limit cycles in certain other systems are quenched by the introduction of a sinusoidal input. It is therefore necessary to examine the state of a sinusoidally forced system to determine whether a limit cycle can exist. This requires the use of a two-sinusoid-input describing function, as we shall see in Chap. 5. Similarly, testing the stability of a DF-derived frequency response solution requires methods discussed in Chap. 5. For the

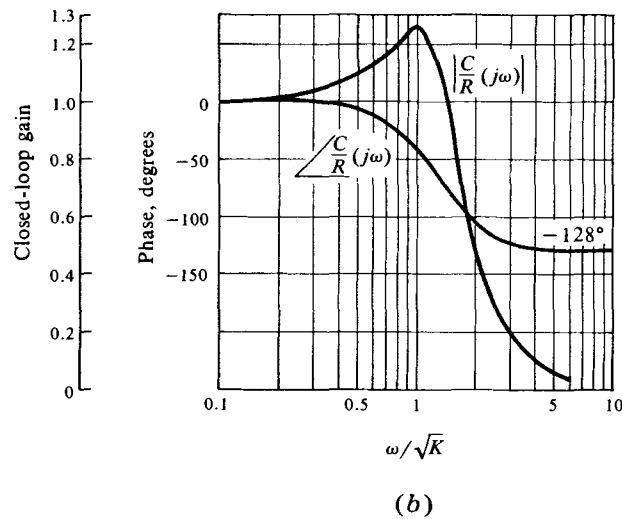
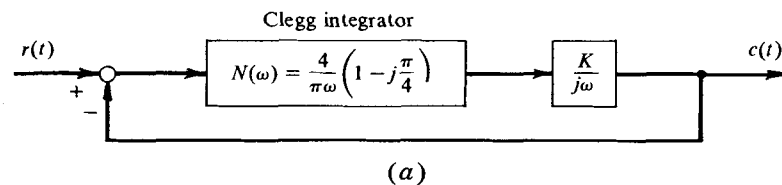


Figure 3.3-1 (a) System containing a Clegg integrator. (b) Its closed-loop frequency response. (Levinson, Ref. 60, pt. 4, p. 132.)

moment, however, we assume DF study of frequency response to be valid wherever applied.

Example 3.3-1 Determine the frequency response of the nonlinear system containing a Clegg integrator, shown in Fig. 3.3-1a.

Since the DF for the Clegg integrator depends on frequency alone [cf. Eq. (2.4-15)], one proceeds as in the linear case to arrive directly at a closed-form solution for $C/R(j\omega)$. Thus

$$\begin{aligned} \frac{C}{R}(j\omega) &= \frac{N(\omega)K/j\omega}{1 + N(\omega)K/j\omega} \\ &= \frac{1 + j(4/\pi)}{1 - (\omega^2/K) + j(4/\pi)} \\ &= \frac{\sqrt{1 + (4/\pi)^2}}{\sqrt{(1 - (\omega^2/K))^2 + (4/\pi)^2}} \exp \left[-j \tan^{-1} \left(\frac{(4\omega^2/\pi K)}{1 - (\omega^2/K) + (4/\pi)^2} \right) \right] \end{aligned}$$

The magnitude and phase angle determined above are plotted in Fig. 3.3-1b. Note the lower closed-loop phase than would be associated with the corresponding purely linear system.

In general, the DF is a function of both A and ω . It is the dependence upon A which forces the use of more complicated means for determination of frequency response. Referring to Fig. 3.3-2, let $r(t)$ represent the simple harmonic excitation

$$r(t) = M_r \sin \omega t \quad (3.3-1)$$

Assuming $x(t)$ to be a sinusoid of amplitude A and frequency ω , the transfer function to $c(t)$ becomes

$$\frac{C}{R}(j\omega, A) = \frac{L_1(j\omega)N(A, \omega)L_2(j\omega)}{1 + L_1(j\omega)N(A, \omega)L_2(j\omega)L_3(j\omega)} \quad (3.3-2)$$

where the nonlinearity is represented by its DF. This is the basis for a frequency response derivation. In this notation, applicable to all system variables, capitals are used to designate frequency-domain-transformed quantities of otherwise lowercase-designated time functions. Continuing, we observe that the transfer function relating C to R is dependent upon A

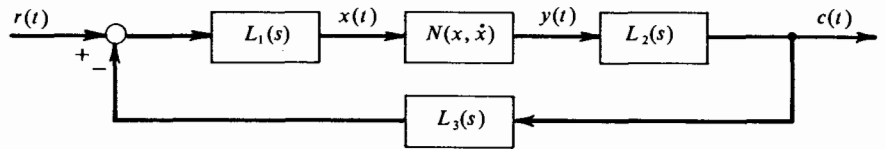


Figure 3.3-2 General nonlinear control system considered.

and ω , the descriptors of an as yet unknown X . One may negotiate this circular difficulty by working with the related transfer function

$$\frac{X}{R}(j\omega, A) = \frac{L_1(j\omega)}{1 + L_1(j\omega)N(A, \omega)L_2(j\omega)L_3(j\omega)} \quad (3.3-3)$$

and considering A and ω as the independent variables (Ref. 68). At any given input frequency ω , let us assume a value of A . This procedure, which specifies all of X but its phase, also defines the ratio X/R by virtue of Eq. (3.3-3). Working with the magnitude of X/R , the value of M_r corresponding to A is immediately determined as

$$M_r = \frac{A}{\left| \frac{X}{R}(j\omega, A) \right|} \quad (3.3-4)$$

The value of A corresponding to a particular combination of M_r and ω can be found routinely by inspecting accumulated data derived as above. At this point the transfer function C/R [Eq. (3.3-2)] may be determined (for a particular M_r), thus defining the frequency response of the system. It is clear that all frequency response plots such as either $|C/R|$ versus ω or $\angle C/R$ versus ω must carry the value of M_r for which the plots were derived.

In order to generate a complete set of data relating C to R for the range of expected input amplitudes and frequencies, one must use all the data that relate $N(A, \omega)$ to R in the repeated solutions of Eq. (3.3-3). A systematized procedure making substantial use of available graphical aids for transfer function evaluation renders this approach worthwhile (Ref. 44).

A CONSTRUCTION FOR FREQUENCY RESPONSE

For circumstances under which it is desirable to study directly the effects of a variety of nonlinearities on the closed-loop frequency response, linear elements remaining fixed, a different solution to the frequency response problem is available. Let us make the following definitions:

$$\begin{aligned} L_1 &= \rho_1 e^{j\theta_1} \\ L_1 L_2 L_3 &= \rho_2 e^{j\theta_2} \\ N &= \rho_N e^{j\theta_N} \end{aligned} \quad (3.3-5)$$

Functional dependence of ρ_1 , ρ_2 , θ_1 , and θ_2 upon ω , and of ρ_N and θ_N upon both ω and A , are omitted for convenience. Considering the absolute magnitude of each side of Eq. (3.3-3), we may write

$$\begin{aligned} \frac{A}{M_r} &= \left| \frac{\rho_1 \exp(j\theta_1)}{1 + \rho_2 \rho_N \exp[j(\theta_2 + \theta_N)]} \right| \\ &= \frac{\rho_1}{\sqrt{1 + 2\rho_2 \rho_N \cos(\theta_2 + \theta_N) + (\rho_2 \rho_N)^2}} \end{aligned} \quad (3.3-6)$$

which can be put in an alternative form,

$$A\rho_N = -\frac{A \cos(\theta_2 + \theta_N)}{\rho_2} \pm \frac{1}{\rho_2} \sqrt{\rho_1^2 M_r^2 - A^2 \sin^2(\theta_2 + \theta_N)} \quad (3.3-7)$$

This equation may be used to get a graphical solution for the frequency response. Both sides can be separately plotted as functions of A , ω remaining fixed. Intersections of the plotted curves represent solutions to Eq. (3.3-3).

For systems with nonlinearities whose DFs are non-phase-shifting, $\theta_N = 0$, and Eq. (3.3-7) reduces to

$$A\rho_N = -\frac{A \cos \theta_2}{\rho_2} \pm \frac{1}{\rho_2} \sqrt{\rho_1^2 M_r^2 - A^2 \sin^2 \theta_2} \quad (3.3-8)$$

which is relatively simple to investigate.¹ To begin with, the right-hand side of this equation is not a function of the DF. Hence changing the nonlinearity does not alter this part of the construction. For a fixed value of frequency, A is the only independent variable of this equation. It is readily established that a plot of the right-hand side versus A is an ellipse.

To facilitate the graphical procedure, the key points defining this ellipse are shown in Fig. 3.3-3. Ordinate and abscissa intercepts are indicated. Coordinates of the maximum ordinate and abscissa excursions serve to locate approximately the tip of the ellipse. Consider, for a moment, that in the DF study of frequency response, the system involved is required not to limit-cycle. Where single-valued nonlinearities are present, this restriction implies that θ_2 lies in the range $-\pi < \theta_2 < 0$ (cf. Sec. 3.1). Since $\csc \theta_2$ is negative throughout this range, the association of $-\csc \theta_2$ with coordinates of the tip of the ellipse in Fig. 3.3-3 is explained. Because the $\cot \theta_2$ term also appears, we see that for the semimajor axis of the ellipse to actually fall in the first quadrant (as shown), it is required that θ_2 lie in the range $-\pi < \theta_2 < -\pi/2$.

Example 3.3-2 Qualitatively determine the character of the harmonically forced response of the system illustrated in Fig. 3.3-4a.

First we observe that the DF for the system nonlinearity is non-phase-shifting. For this system we have

$$L_1(s) = L_3(s) = 1 \quad \text{and} \quad L_2(s) = \frac{K}{s(s+1)}$$

which imply

$$\begin{aligned} \rho_1 &= 1 & \theta_1 &= 0 \\ \rho_2 &= \frac{K}{\omega \sqrt{1 + \omega^2}} & \theta_2 &= -\frac{\pi}{2} - \tan^{-1} \omega \end{aligned}$$

¹ We follow Levinson's original work (Ref. 58) in this presentation.

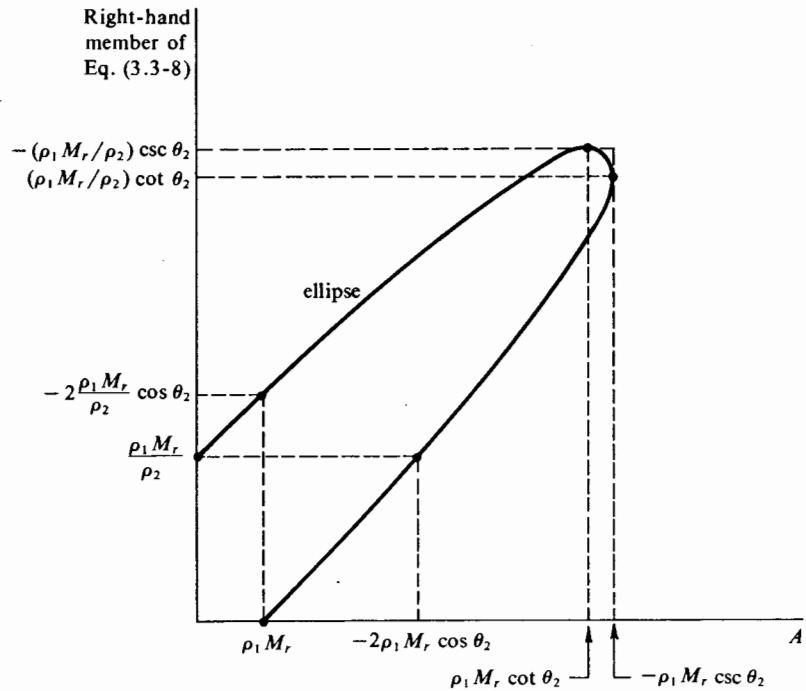


Figure 3.3-3 A frequency response construction aid, M_r and ω constant.

Utilizing Fig. 3.3-3, a group of ellipses corresponding to various input frequencies can be drawn. The result of this construction for a single value of M_r is shown in Fig. 3.3-4b. Since $\rho_1 = 1$, all ellipses pass through the point $(M_r, 0)$.

The quantity $A\rho_N$ is simply the amplitude of the nonlinearity output fundamental. In terms of increasing A , the plot of $A\rho_N$ is like a gain-changing function. This can be shown exactly by referring to the DF calculation for the gain-changing nonlinearity available in Chap. 2, or by noting the limiting results

$$A\rho_N \rightarrow m_1 A \quad \text{as } A \rightarrow 0$$

$$A\rho_N \rightarrow m_2 A \quad \text{as } A \rightarrow \infty$$

from which a rough plot of $A\rho_N(A)$ can be made. Generally speaking, the shape of $A\rho_N$ closely resembles the nonlinear characteristic itself, with all sharp corners rounded off.

The magnitude of ρ_N can now be found on a point-by-point basis. For example, at $\omega = \omega_s$, it follows that $A = A_s$, and hence that $\rho_N(A_s)$ is known. For certain values of ω , this point-by-point process is unique. Other values of ω result in multiple intersections in Fig. 3.3-4b, indicating a multivalued solution. Utilizing all values of ρ_N found graphically, amplitude and phase plots of $C/R(j\omega)$ can gradually be built up. Figure 3.3-5 illustrates such a solution for the particular M_r of Fig. 3.3-4b.

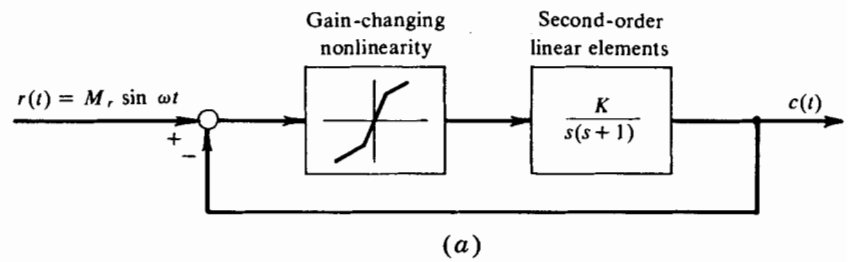


Figure 3.3-4a Example system for frequency response study.

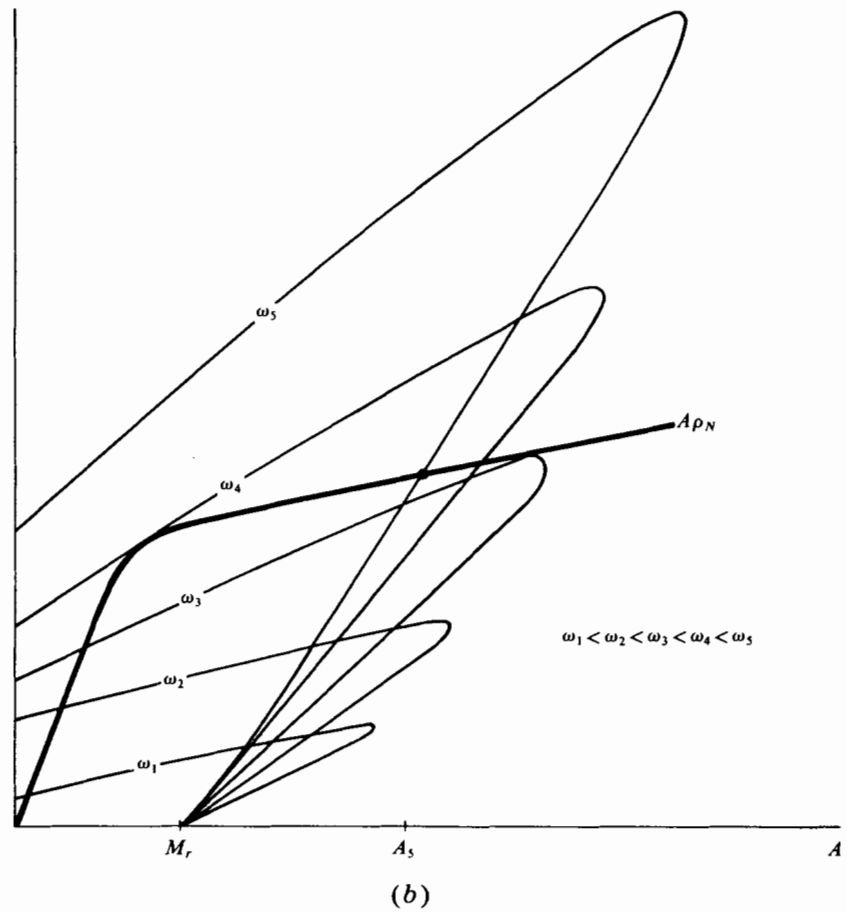


Figure 3.3-4b Frequency response construction for example system.

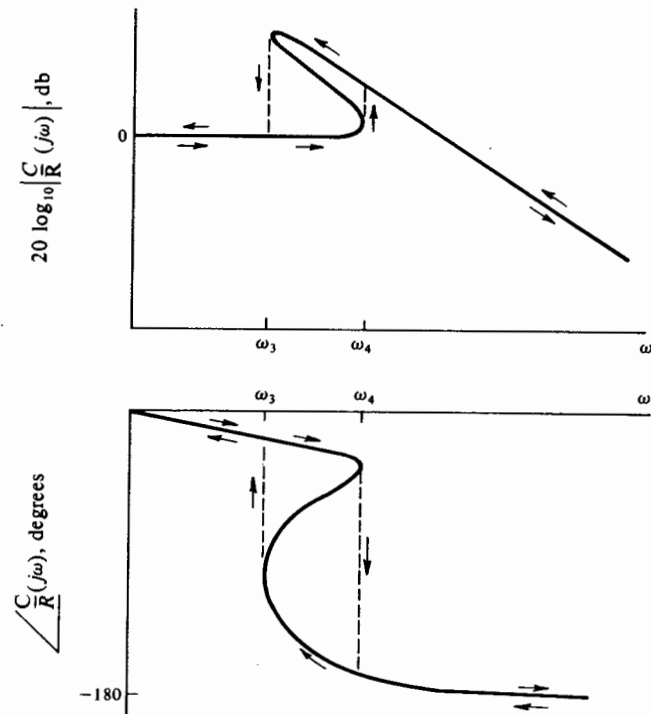


Figure 3.3-5 Closed-loop frequency response curves for a particular M_r . Amplitude and phase jumps occur.

The closed-loop frequency response curves of the example system appear to exhibit a hysteretic behavior. They indicate discontinuities in the amplitude and phase response for an input frequency either increasing and surpassing ω_4 or decreasing and falling below ω_3 . This form of discontinuous closed-loop frequency response has been appropriately labeled *jump resonance*. It is unlike any response provided by a linear system, and is due to the gain-changing character of a system nonlinearity. As a result of the jumps for increasing and decreasing frequency, a portion of the triple-valued amplitude and phase response curves is apparently never traversed. This portion of the curves may therefore be considered as representing unstable steady-state solutions. In fact, this portion cannot be observed experimentally. Proof of such an instability has been demonstrated by Stoker (Ref. 99) and Bonenn (Ref. 10).

Generally speaking, the amplitude resonance peak bends to the left for saturating nonlinear characteristics and to the right for increasing gain characteristics. This is illustrated in Fig. 3.3-6. As implied before, arrows indicate permissible motion along the locus in question.

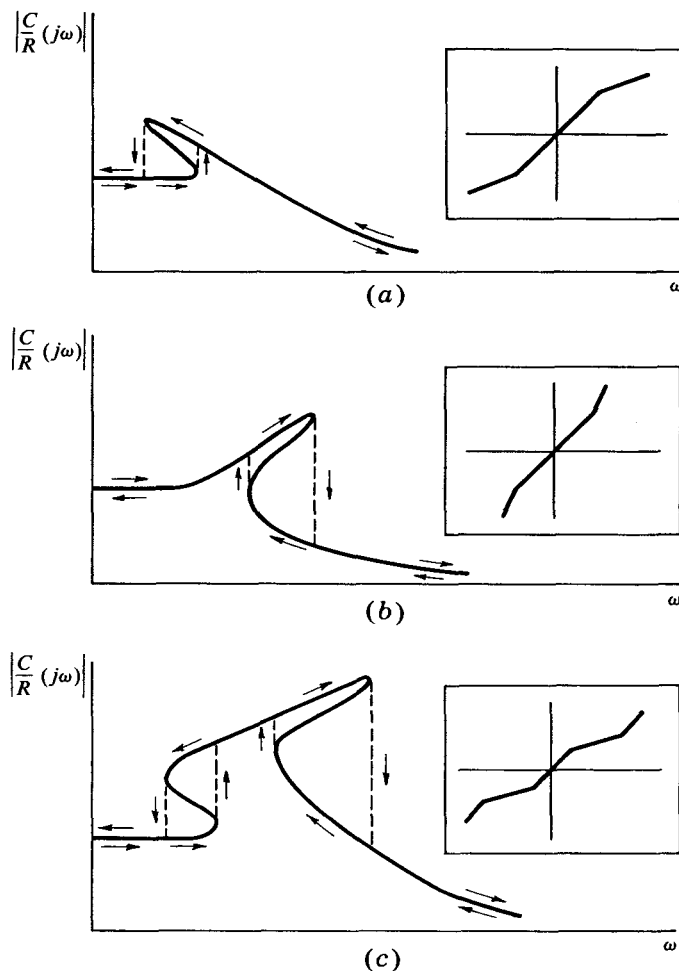


Figure 3.3-6 Decreasing (a), increasing (b), and combination (c) non-linear gain characteristics shown with typical system amplitude jump responses.

Note that for a linear device $A\rho_N$ is directly proportional to A ; it follows that a construction similar to that of Fig. 3.3-4b will yield only single intersections with each ellipse. The amplitude resonance peak in this instance is neither to the left nor to the right; *jump resonance cannot occur*.

CRITICAL JUMP RESONANCE CURVE

A curve which encloses a region of the complex plane such that jump resonance can occur if the open-loop frequency response locus of the system linear

part passes into this region is called a *critical jump resonance curve*. It is derived in the following discussion after a method due to Hatanaka (Ref. 43). The nonlinearity is restricted to be a single-valued odd static operator.¹

Recall that at the jump resonance point $|C/R|$ is discontinuous. It follows that $|X/R|$ is also discontinuous. That is, considering the input amplitude M_r to be a function of the error amplitude A for a fixed angular frequency ω , the same value of M_r corresponds to different values of A . If the nonlinearity DF is differentiable with respect to A and possesses a continuous first derivative, the condition under which jump resonance occurs is thus

$$\left(\frac{\partial M_r}{\partial A}\right)_{\omega=\text{const}} = 0 \quad \text{at the jump resonance point} \quad (3.3-9)$$

This condition prevails at points 1 and 2 of Fig. 3.3-7, which is a typical plot of A versus M_r (at fixed ω) for the system illustrated. Further, corresponding to the region between points 1 and 2, the requirement for unstable forced responses is that the slope of the curve be negative (Ref. 10).

$$\left(\frac{\partial M_r}{\partial A}\right)_{\omega=\text{const}} < 0 \quad \text{for region in which jump resonance occurs} \quad (3.3-10)$$

Using the above results, we can proceed to the derivation of interest. First, note that by use of the definition

$$L(j\omega) = U(\omega) + jV(\omega)$$

one can put the *magnitude* of the system error transfer function

$$\frac{X}{R}(j\omega, A) = \frac{1}{1 + N(A)L(j\omega)}$$

in the form

$$M_r = AN(A) \sqrt{\left[U(\omega) + \frac{1}{N(A)}\right]^2 + [V(\omega)]^2} \quad (3.3-11)$$

Taking the partial derivative of M_r in accordance with Eq. (3.3-9) yields

$$\left[U(\omega) + \frac{1}{N(A)}\right] \left[U(\omega) + \frac{1}{N^*(A)}\right] + V(\omega)^2 = 0 \quad (3.3-12)$$

where N^* is defined by

$$N^*(A) = N(A) + AN'(A) \quad (3.3-13)$$

and the prime indicates differentiation with respect to A . From Eq. (3.3-12) we see that contours in the polar plane for constant values of A at the jump

¹ Another treatment of this problem, including results for the case of a nonlinearity with memory, can be found in Ref. 25.

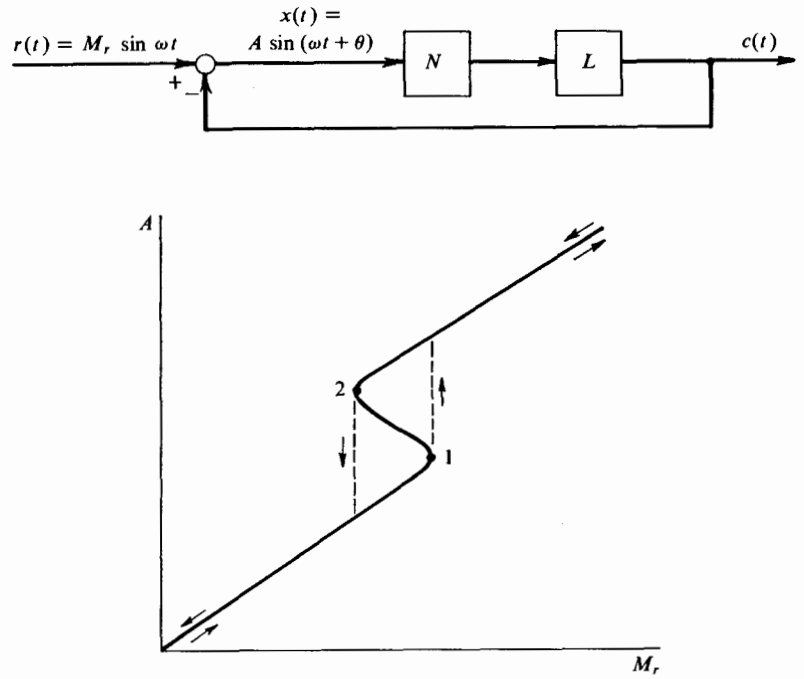


Figure 3.3-7 Steady-state error amplitude vs. forcing amplitude in a system exhibiting jump resonance.

resonance point are a family of circles. Denote this family of contours $f(U, V, A) = 0$.

We now seek the *envelope* of the family of contours $f(U, V, A) = 0$. It is proved in calculus that such an envelope satisfies the relationships

$$f(U, V, A) = 0 \quad \text{and} \quad \frac{\partial f(U, V, A)}{\partial A} = 0 \quad (3.3-14)$$

Operating on Eq. (3.3-12) in accordance with Eqs. (3.3-14) results in

$$U = -\frac{N'(A)N^*(A) + N(A)N'^*(A)}{N'(A)[N^*(A)]^2 + N'^*(A)[N(A)]^2}$$

$$V = \pm \frac{\sqrt{N'(A)N'^*(A)} [N(A) - N^*(A)]}{N'(A)[N^*(A)]^2 + N'^*(A)[N(A)]^2} \quad (3.3-15)$$

which is the desired critical jump resonance curve specified parametrically in A .

Example 3.3-3 Find the critical jump resonance curve for the nonlinearity $y = x^3$.
From Eqs. (3.3-15) we directly obtain

$$U = - \frac{(\frac{3}{2}A)(\frac{9}{4}A^2) + (\frac{3}{4}A^2)(\frac{9}{2}A)}{(\frac{3}{2}A)(\frac{81}{16}A^4) + (\frac{9}{2}A)(\frac{9}{16}A^4)} = - \frac{2}{3} \frac{1}{A^2}$$

$$V = \pm \frac{\sqrt{(\frac{3}{2}A)(\frac{9}{2}A)(\frac{3}{4}A^2 - \frac{9}{4}A^2)}}{(\frac{3}{2}A)(\frac{81}{16}A^4) + (\frac{9}{2}A)(\frac{9}{16}A^4)} = \pm \frac{2}{3\sqrt{3}} \frac{1}{A^2}$$

Eliminating A yields

$$V = \pm \frac{1}{\sqrt{3}} U$$

where it is understood that U takes only negative values. This locus is plotted in Fig. 3.3-8. Crosshatching indicates the interior of the region in which jump resonance occurs since everywhere within this region Eq.(3.3-10) is satisfied.

The critical jump resonance curves for a gain-changing nonlinearity, derived as above, are shown in Fig. 3.3-9. We shall have occasion to recall these results in Chap. 5, where they are again derived, albeit from quite a different point of view.

The methods discussed for obtaining the frequency response of a nonlinear system can be applied as well to nonlinear subsystems such as were treated in

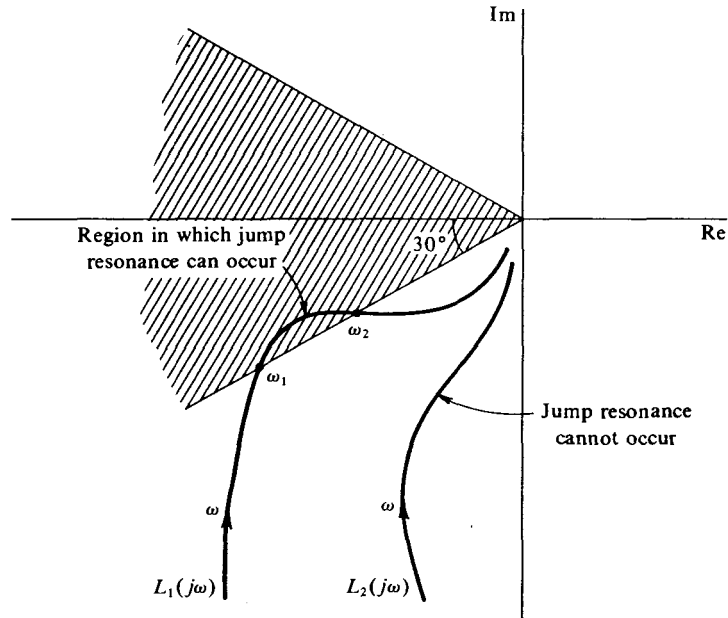


Figure 3.3-8 Critical jump resonance curve for cubic nonlinearity.

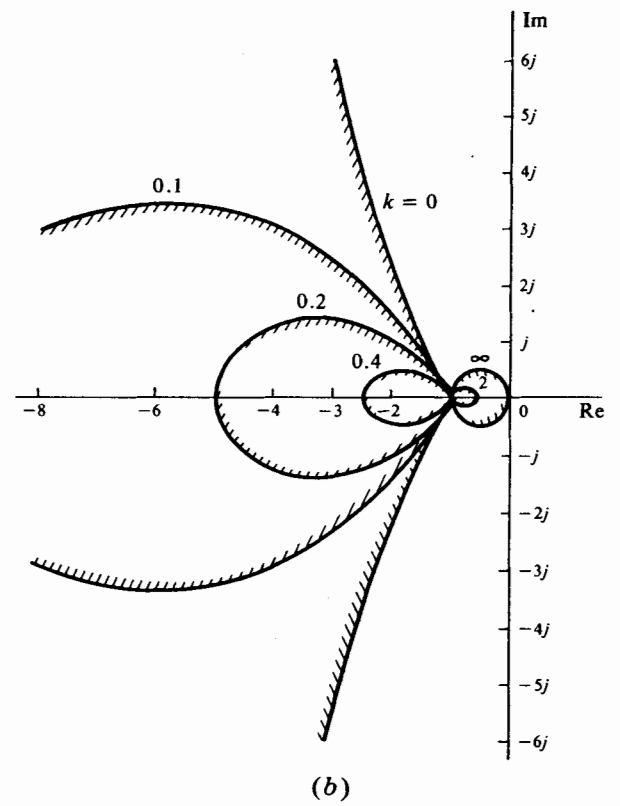
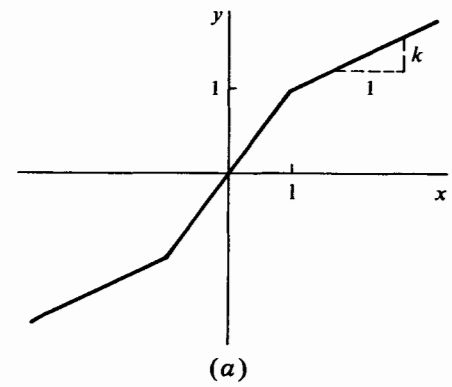


Figure 3.3-9 Critical jump resonance curve (b) for gain-changing nonlinearity (a). (Hatanaka, Ref. 43.)

Sec. 2.4. In particular, we now have additional means for DF derivation in the case of implicit dynamical nonlinearities. Before leaving the topic of nonlinear-system frequency response, one additional system is well worth considering, the time-optimal computer control system.

3.4 APPLICATION TO A TIME-OPTIMAL COMPUTER CONTROL SYSTEM

A time-optimal control system is one in which the error and its $n - 1$ derivatives (where n is the order of the controlled process) are simultaneously returned to zero in the shortest possible time. For a $1/s^n$ plant, the system input is characterized by an n -term Taylor series. To demonstrate DF utilization for frequency response investigation, we shall choose a simple $1/s^2$ controlled process in the system configuration of Fig. 3.4-1. An ideal relay is the controlled-process driver, as in all time-optimal systems. This follows from the desire to achieve either *maximum* load acceleration or *maximum* load deceleration in a correspondingly maximum control effort operation. The computer receives all system inputs and outputs, operating on them in such a way as to switch the ideal relay according to the requirements for time-optimal control. Examples of the controlled-process and relay-driver combination depicted include an interplanetary vehicle driven for attitude control about each of two axes by a two-position continuously operating microrocket (Ref. 25), an acceleration-switching hydraulic servovalve (Ref. 38), etc. Sparacino (Ref. 96) has treated this and other second-order processes. The method to be described closely follows his presentation.

The *control law* implemented by the computer must be derived as a prelude to DF evaluation. In order to do this we utilize the fact that for an n th-order system with real poles the error and its $n - 1$ derivatives may be returned to zero time optimally, with at most $n - 1$ switchings. We thus require but one switching. The system error trajectory will therefore be comprised of two parts, the second part necessarily going through the origin of the phase plane (the coordinates of which are error rate versus error).

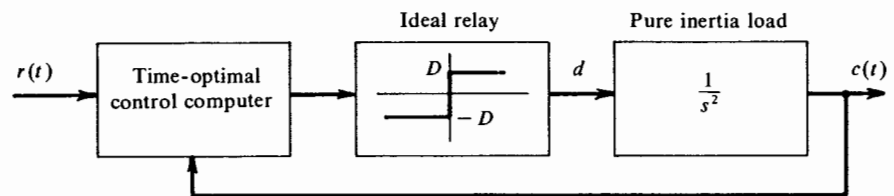


Figure 3.4-1 Time-optimal control scheme for a second-order controlled process.

Assuming that the system input may be represented by a two-term Taylor series, the equation of motion for the second-order process at hand during the *last* (second) trajectory may be written as

$$e = e_s + \dot{e}_s(t - t_s) - \frac{d}{2}(t - t_s)^2 \quad \text{for } t \geq t_s \quad (3.4-1)$$

where d is the ideal-relay output ($+D$ or $-D$),

$$e = r - c \quad (3.4-2)$$

and e_s is the final value of e for the *initial* trajectory. We now impose the requirements of time-optimal control, namely,

$$e(T) = \dot{e}(T) = 0 \quad (3.4-3)$$

where T is the shortest time in which Eq. (3.4-3) can be achieved. Figure 3.4-2 illustrates a typical input, output, and corresponding error trajectory. Solving Eqs. (3.4-1) and (3.4-3) by eliminating time gives

$$e_s = -\frac{\dot{e}_s^2}{2d} \quad (3.4-4)$$

We must therefore switch when the error satisfies Eq. (3.4-4), for in that case the last trajectory will be perfectly executed, having begun with the proper initial conditions. The control computer therefore chooses the sign of d according to ($\text{sgn} = \text{signum} = \text{"sign of"}\text{"}$)

$$\text{sgn } d = \text{sgn} \left(e + \frac{\dot{e}^2}{2D} \text{sgn } \dot{e} \right) \quad (3.4-5)$$

We are now in a position to derive the DF for the computer and relay, in combination. Since the computer acts upon the quantity $r - c$, the system of Fig. 3.4-1 may be redrawn as in Fig. 3.4-3. $N(e, \dot{e})$ represents the dynamics implied by Eq. (3.4-5). In particular, the output of N is a square wave of amplitude D (first harmonic $4D/\pi$) and phase shift θ_N (see Fig. 3.4-4). Assuming a sinusoidal input

$$e = A \sin \omega t \quad (3.4-6)$$

it follows from Eq. (3.4-4) that the switching criterion is

$$A \sin \omega t = -\frac{(A\omega)^2 \cos^2 \omega t}{2d} \quad (3.4-7)$$

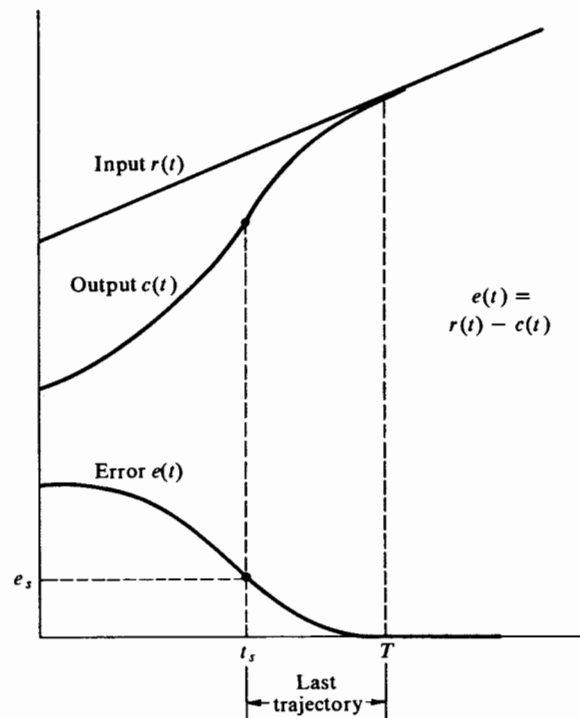


Figure 3.4-2 Typical time history of the input, output, and error in a time-optimal control system.

The substitutions $\omega t = -\theta_N$ and $d = D$, indicated by Fig. 3.4-4, yield θ_N in the form

$$\frac{\sin \theta_N}{\cos^2 \theta_N} = \frac{A\omega^2}{2D} \quad (3.4-8)$$

or equivalently,

$$\theta_N = \sin^{-1} \left[-\frac{D}{A\omega^2} + \sqrt{\left(\frac{D}{A\omega^2}\right)^2 + 1} \right] \quad (3.4-9)$$

The DF for the computer-and-relay combination is therefore given by

$$N(A, \omega) = \frac{4D}{\pi A} \exp \left[j \sin^{-1} \left(-\frac{D}{A\omega^2} + \sqrt{\left(\frac{D}{A\omega^2}\right)^2 + 1} \right) \right] \quad (3.4-10)$$

which is amplitude- and frequency-dependent. Nevertheless, the frequency response of this time-optimal control system can now be obtained by application of the methods of Sec. 3.3.

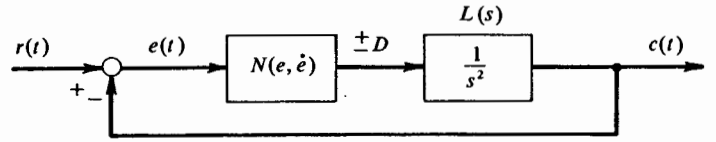


Figure 3.4-3 Equivalent system for the frequency response analysis.

Another avenue to the frequency response of this system was found by Sparacino (*op. cit.*), who observed that the open-loop transfer function of this system, $N(A, \omega)L(j\omega)$, could be written in terms of a *single* parameter, which we shall call F ; namely,

$$\begin{aligned} N(A, \omega)L(j\omega) &= \frac{4D}{\pi A \omega^2} \exp \left(j \left\{ -\pi + \sin^{-1} \left[-\frac{D}{A \omega^2} + \sqrt{\left(\frac{D}{A \omega^2} \right)^2 + 1} \right] \right\} \right) \\ &= \frac{4}{\pi F} \exp \left(j \left\{ -\pi + \sin^{-1} \left[-\frac{1}{F} + \sqrt{\left(\frac{1}{F} \right)^2 + 1} \right] \right\} \right) \end{aligned} \quad (3.4-11)$$

$$\text{where} \quad F = \frac{A \omega^2}{D} \quad (3.4-12)$$

This definition for F obviates the need for construction of a family of graphs in determination of the system frequency response, a quite desirable aspect. Such a manipulation is clearly not always possible.

We may now adopt the viewpoint that $N(A, \omega)L(j\omega)$, or, as it were, $NL(F)$, is a new nonlinearity N' in a closed-loop system wherein all linear

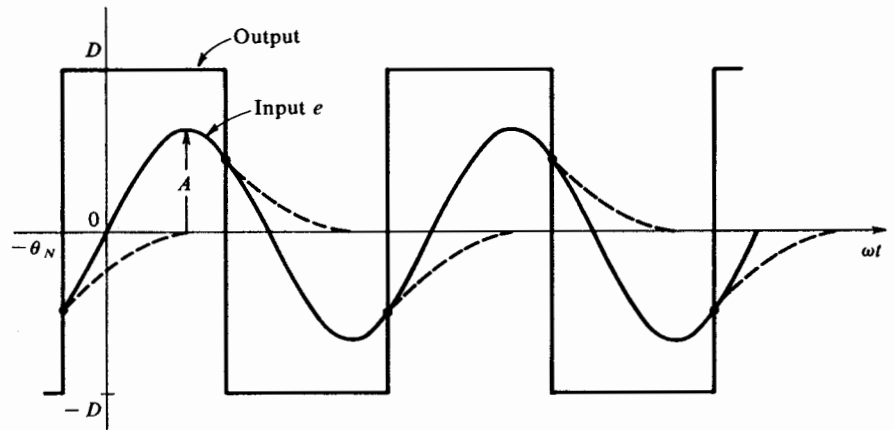


Figure 3.4-4 Sinusoidal response of N for the time-optimal control of a pure inertia load.

elements have unity amplitude transfer and zero phase shift. Thus, from Eqs. (3.3-5),

$$\begin{aligned} \rho_1 = \rho_2 = 1 \quad \theta_1 = \theta_2 = 0 \\ \rho_{N'} = \frac{4}{\pi F} \quad \theta_{N'} = -\pi + \sin^{-1} \left[-\frac{1}{F} + \sqrt{\left(\frac{1}{F}\right)^2 + 1} \right] \end{aligned} \quad (3.4-13)$$

Equation (3.3-6), governing the appropriate graphical construction, now becomes

$$\frac{A}{M_r} = \frac{1}{\sqrt{1 + 2\rho_{N'} \cos \theta_{N'} + (\rho_{N'})^2}} \quad (3.4-14)$$

where, as before, M_r is the magnitude of the system input sinusoid. Multiplying and dividing the left-hand member of Eq. (3.4-14) by ω^2/D ,

$$\frac{A}{M_r} = \frac{F}{M_r \omega^2 / D} \quad (3.4-15)$$

A plot of the left- and right-hand members of Eq. (3.4-14) yields solution points at observed intersections. The complicated right-hand member as a

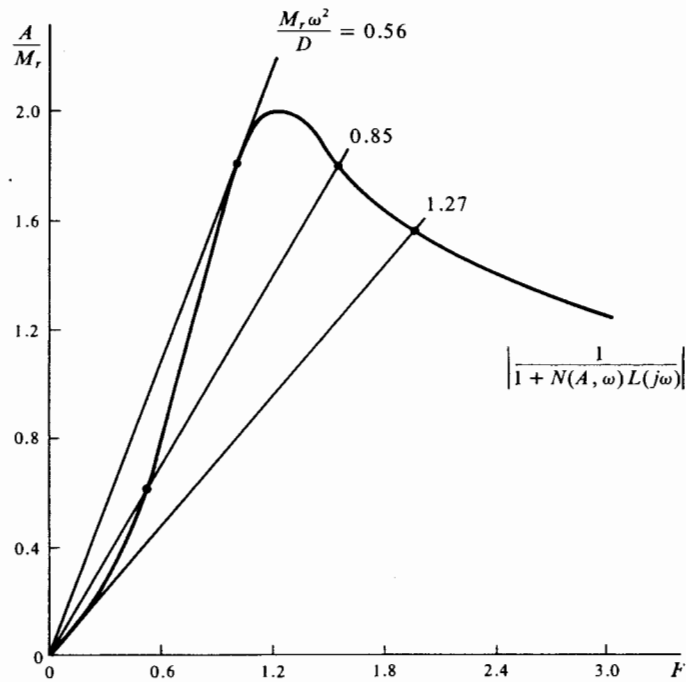


Figure 3.4-5 Graphical construction yielding A/M_r for a time-optimally-controlled pure inertia load. (Adapted from Sparacino, Ref. 96.)

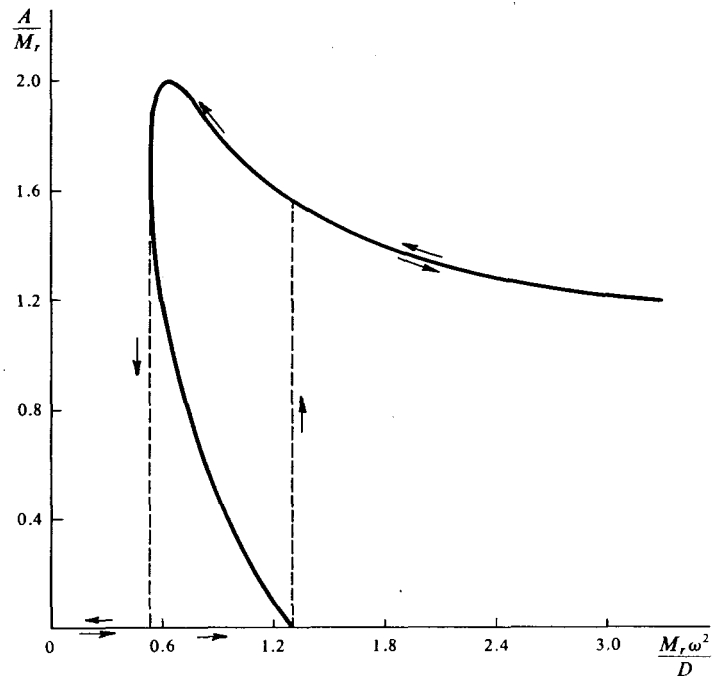


Figure 3.4-6 Frequency-amplitude response of a time-optimally-controlled pure inertia load. (Adapted from Sparacino, Ref. 96.)

function of F need be drawn but once, with the left-hand straight-line member drawn for a variety of values of $M_r \omega^2 / D$, its reciprocal slope. Figure 3.4-5 illustrates these constructions. Note that $M_r \omega^2$ is the peak input acceleration and that, for the $1/s^2$ linear elements, D represents the peak output acceleration. Thus $M_r \omega^2 / D$ is the ratio of peak commanded acceleration to peak output acceleration capability, a physically significant quantity. Changing the slope of the left-hand member of Eq. (3.4-14) therefore directly corresponds to sweeping out all the input dynamics-system dynamics possibilities. From another point of view, for fixed system dynamics the act of changing $M_r \omega^2 / D$ directly corresponds to changing the command input. With either point of view the entire system harmonic response is immediately evident from the plot.

For values of $M_r \omega^2 / D$ less than 0.56, one intersection occurs at $F = 0$. This implies that $A = 0$, or that the system follows the sinusoidal input with zero steady-state following error. For values of $M_r \omega^2 / D$ between 0.56 and 1.27, three intersections occur. An abrupt transition from the previous state is indicated, with the system now displaying finite steady-state following error. For $M_r \omega^2 / D > 1.27$, two intersections exist, one of which is nonzero. Figure 3.4-6 summarizes these results by displaying the error response vs. the

quantity $M_r\omega^2/D$. Since it is both the frequency response (input amplitude constant) and amplitude response (input frequency constant), it has been so labeled. Arrows indicate possible motion of the solution along the frequency-amplitude response locus. The trivial solution for $M_r\omega^2/D > 1.27$ has been removed from expected possibilities, based on the physical reasoning that for $M_r\omega^2/D > 1$ a following error would have to exist.

Experimental results reported (Sparacino, *op. cit.*) confirm the presence of jump resonance as well as the entire frequency-amplitude response of Fig. 3.4-6. The critical values of $M_r\omega^2/D$ so determined were approximately 0.60 and 1.15 as compared with 0.56 and 1.27 determined by the DF method. We may therefore conclude that the DF study predicted all the salient harmonic response behavioral patterns for the time-optimal controlled pure inertia load, with less than 10 percent error in determination of all critical parameter values.

3.5 LINEAR AND NONLINEAR LOOP COMPENSATION TECHNIQUES

The conventional compensation problem is to choose a network which, when inserted in a control loop, causes the overall loop (including fixed elements) to meet specifications. One fixed element is usually the loop final element; that is, its output is the controlled variable $c(t)$. Others, of course, are possible, any of which can be either linear or nonlinear. An inherent loop nonlinearity is referred to as unintentional (or incidental); examples are output member backlash, power element saturation, and feedback element threshold. Intentional nonlinearities include nonlinear compensation networks such as those discussed later on in this section.

Steady-state requirements for loop compensation commonly are (1) to rid an otherwise limit cycling system of its limit cycle (as in the case of a position servo with motor-load backlash), (2) to decrease the limit cycle amplitude to some acceptable value where its complete elimination is impossible (as in an on-off aircraft autopilot system), or (3) to permit control of the limit cycle amplitude and/or frequency (as in a vibration controller). Compensation selection for steady-state requirements is readily accomplished via DF usage.

Transient response specifications such as percent overshoot and settling time are difficult to deal with in non-limit-cycling systems. In totally linear systems it is well known that feedforward lead compensation tends to shorten the transient response time, whereas feedforward lag compensation increases transient response time. Moreover, feedback lag compensation has an effect somewhat equivalent to feedforward lead, and vice versa for feedback lead compensation. In non-limit-cycling nonlinear systems it is true that there

is often a certain qualitative agreement between the effect of linear compensation on the actual transient response and that described above for the totally linear case. Because of the inapplicability of linear superposition to these nonlinear systems, however, such *equivalences* are weak, and should at most be used as a rough guide in design. Chapters 4 and 8 are directed toward the study of transient phenomena, and further inferences regarding compensation for transient response can be drawn from material presented there.

Transient response characteristics of limit cycling systems are readily compensated, as we shall find in Chap. 6. With the aid of a dual-input describing function, it will be shown there that the above-mentioned equivalences are no longer weak. The reader is asked to temporarily accept this statement without proof.

Many theories are available to assist in the choice of loop compensation for totally linear systems. Such is not the case for nonlinear systems, especially of third and higher order. In this section we briefly examine several means of compensating these nonlinear systems, using DF theory as a guide to compensation selection.

LINEAR COMPENSATION

Here we presume the existence of a nonlinear fixed element, unintentional or otherwise, and inquire of the use of linear loop compensation.

The steady-state effects of both lead and lag networks on a linear system are conveniently displayed on the amplitude-phase plane, where linear compensation is easily carried out. Let us assume that a system with a third-order transfer function also contains an odd memoryless nonlinearity, whose DF lies on the -180° line in the amplitude-phase plane (see Fig. 3.5-1). The compensation network has the transfer function

$$L_1(s) = \frac{\alpha\tau s + 1}{\tau s + 1} \quad \text{lead network for } \alpha > 1, \text{ lag network for } \alpha < 1 \quad (3.5-1)$$

for which the insert in Fig. 3.5-1 illustrates the frequency locus in log magnitude vs. phase coordinates. We recall that in the process of linear compensation of linear systems the lag network is used for its *attenuation* characteristic at high frequencies (relative to $1/\alpha\tau$), whereas the lead network is used for its *phase-lead* characteristic at low frequencies (relative to $1/\tau$). These portions of the amplitude-phase loci of $L_1(j\omega)$ have been darkened in the insert.

The limit cycle resulting from the use of $L_1(s)$ as a *lead* network has a *higher frequency* and *smaller amplitude* than the limit cycle in the uncompensated system (A_0, ω_0). The lead network has been chosen to supply maximum phase lead in the vicinity of ω_0 . When $L_1(s)$ is a *lag* network, a (slightly) *lower-frequency* and *smaller-amplitude* limit cycle is seen to result. The lag

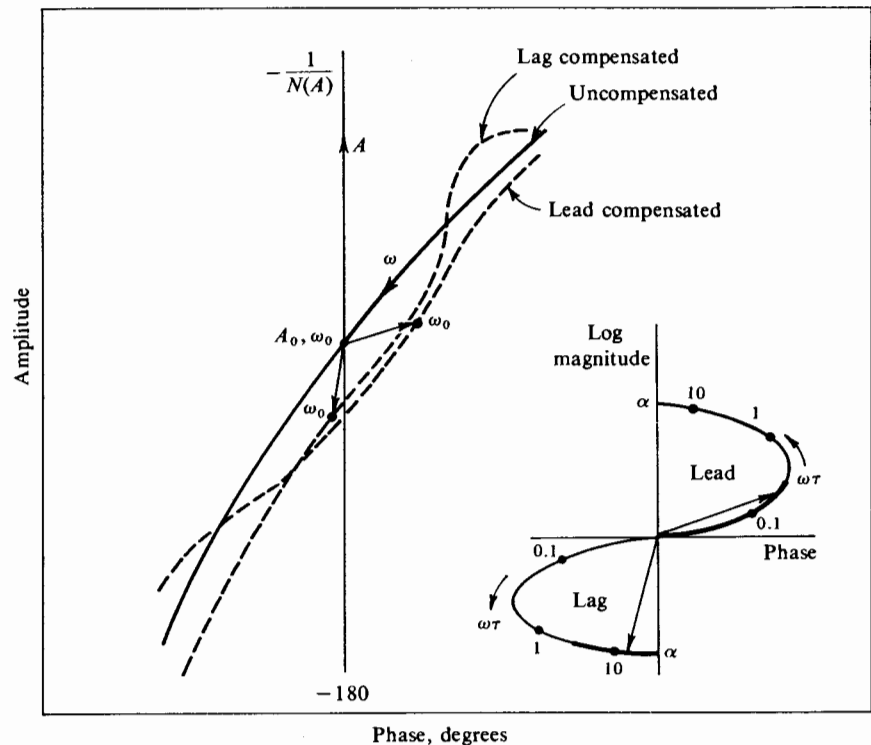


Figure 3.5-1 Effect of linear lead and lag compensation networks viewed on the amplitude-phase plane.

network chosen has its upper breakpoint well below ω_0 , so that in the vicinity of ω_0 , only its attenuation, and not its phase lag, predominates.

Quite commonly the system designer is constrained to place a compensation network at locations between the system output and the nonlinearity input. In a control loop for a pulse-rebalanced accelerometer, for example, compensation *must* be placed in the feedback path since the feedforward path is physically inaccessible. Referring to Fig. 3.5-1, if we assume $L_1(s)$ to be located between the system output and the nonlinearity input, it follows that the limit cycle amplitude *at the system output* is smaller when lead compensation is used than when lag compensation is used.

To see this, recall first that the limit cycle amplitude solved for graphically is that which appears at the nonlinearity *input*. The amplitude at other stations around the loop must be found by working forward or back from this point. Also, the lead network presents a *higher-than-unity gain* to the limit cycle, whereas the lag network presents a *lower-than-unity gain*. Thus, for roughly equal limit cycle amplitudes at the nonlinearity input (as indicated

by Fig. 3.5-1), it must follow that the output limit cycle amplitude is smaller in a lead-compensated system.

The effect of a linear gain K is simply to raise or lower a frequency response locus on the amplitude-phase plane. This operation, which does not affect limit cycle frequency in the case of non-phase-shifting DFs, is shown in Fig. 3.5-2.

Between the use of linear lead, lag, and pure gain networks, limit cycle amplitude and frequency can theoretically be altered as desired to meet specifications. The prevention of limit cycles (where possible) without unnecessarily dropping the loop gain is usually a matter of supplying sufficient phase lead to avoid the crossing of the $-1/N(A, \omega)$ locus by the $L(j\omega)$ locus (see Fig. 3.5-3). Because of familiar noise and saturation considerations, however, a practical upper limit on α of the lead compensation network is about 10.

An additional linear compensation technique occasionally used to achieve a specified limit cycle amplitude at a particular point in a control loop is to

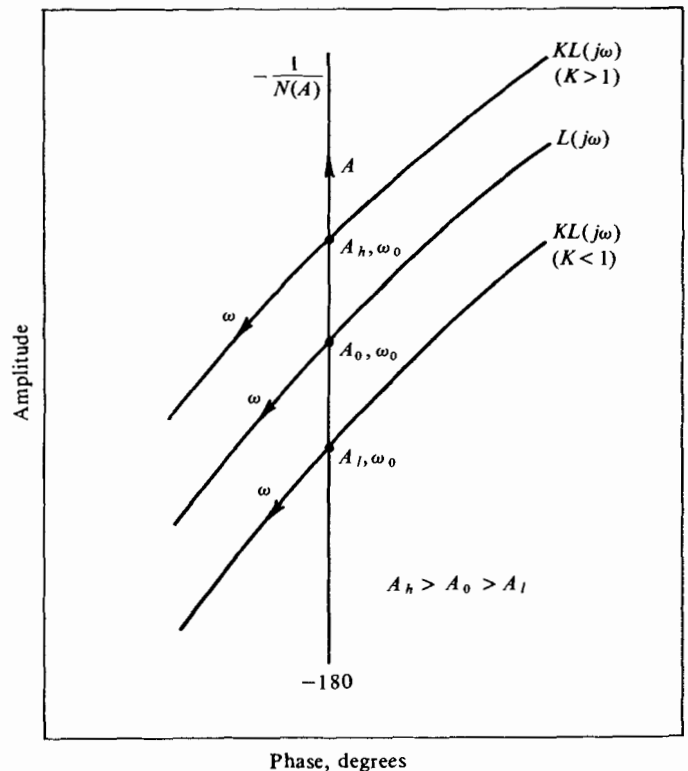


Figure 3.5-2 Effect of linear gain.

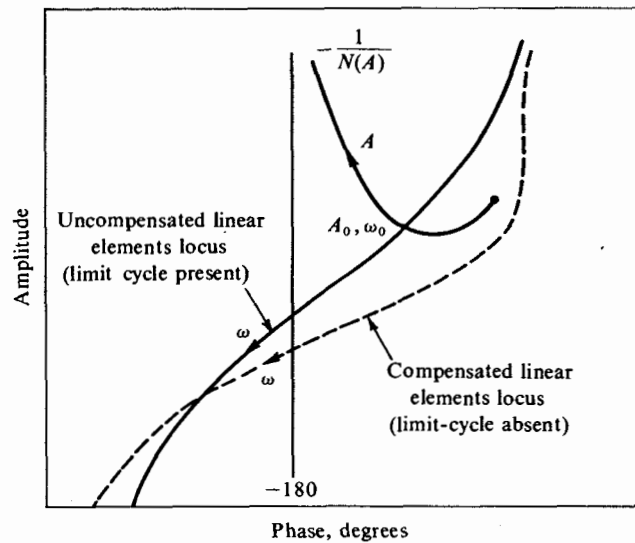


Figure 3.5-3 Lead compensation to remove a limit cycle in a high-order system containing a relay with dead zone and hysteresis.

insert a lead (lag) network, followed elsewhere by the reciprocal lag (lead) network. The net effect of this reciprocal network pair is to alter the limit cycle amplitude at all stations between the networks, but not its frequency. If one network is in the forward path and its reciprocal is in the feedback path, loop command response characteristics are clearly affected and must be accounted for.

NONLINEAR MODEL FEEDBACK

One of the interesting procedures available for synthesizing a nonlinear compensation network is based on what has been called complementary model feedback (Ref. 73). In its basic form, nonlinear model feedback serves to linearize a nonlinear feedback system by providing a signal, which, when added to the nonlinearity output, renders its net output a linear function of its net input. The addition of signals does not actually take place at the nonlinearity output, however, since that would require a nonlinear compensation network operating at high power levels in all cases in which the original nonlinearity is a power element. The signal addition, as a result, conceptually takes place in the feedback path (Fig. 3.5-4b), although it can be implemented in the feedforward path (Fig. 3.5-4c)

Consider the case of *exact* model feedback, $N^* = N$, $L_2^* = L_2$. The complementary nonlinearity $1 - N^*$ becomes $1 - N$. In the case where N

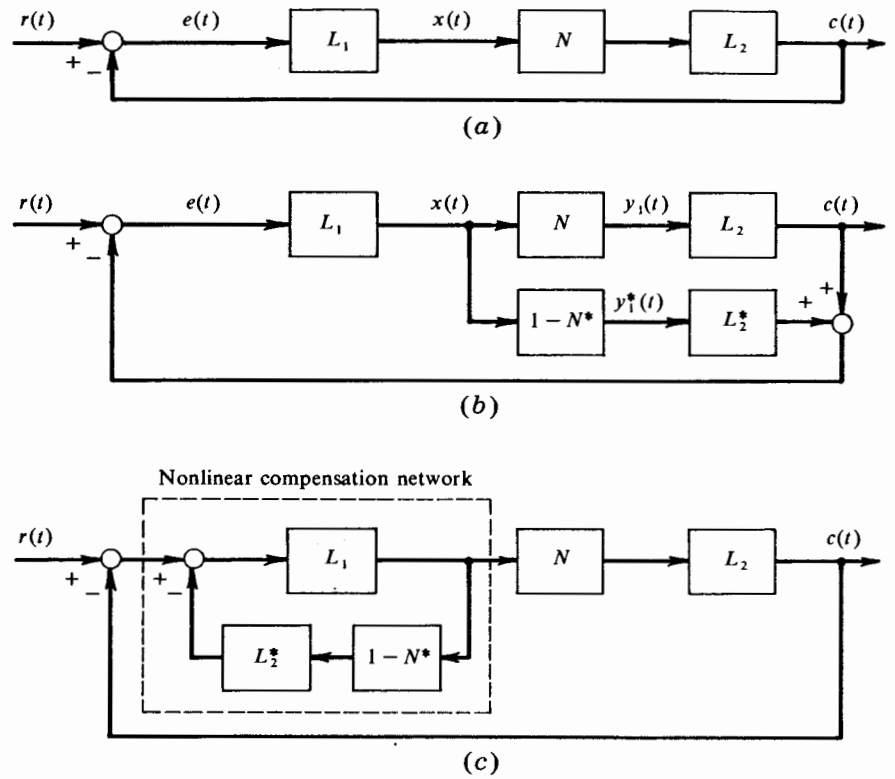


Figure 3.5-4 Nonlinear model feedback. (a) Original system; (b) system with model feedback; (c) equivalent nonlinear compensator.

is an abrupt saturation nonlinearity, for example, $1 - N$ is a piecewise-linear dead zone, and vice versa. It is clear that

$$Y_1(s) + Y_1^*(s) = X(s) \quad (3.5-2)$$

which implies that the transfer function from $X(s)$ to $E(s)$ is independent of N , viz.,

$$\begin{aligned} \frac{E}{X}(s) &= - \frac{Y_1(s)L_2(s) + Y_1^*(s)L_2^*(s)}{X(s)} \\ &= - \frac{[Y_1(s) + Y_1^*(s)]L_2(s)}{X(s)} \\ &= -L_2(s) \end{aligned} \quad (3.5-3)$$

Stability of this totally *linear* loop is therefore governed by the equation

$$1 + L_1(s)L_2(s) = 0 \quad (3.5-4)$$

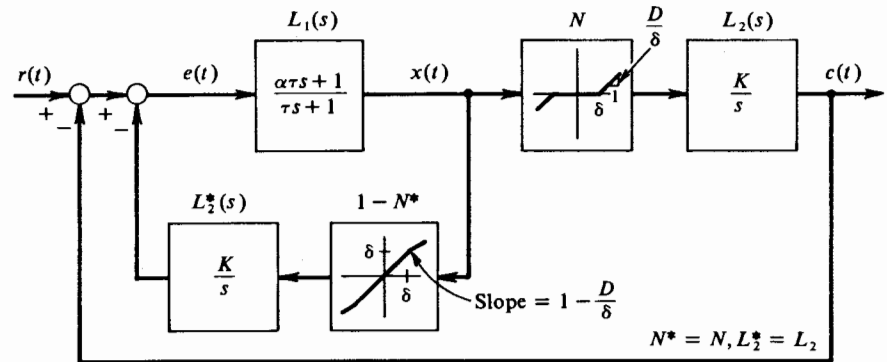


Figure 3.5-5 System requiring the use of inexact model compensation. (Adapted from Mishkin and Braun, Ref. 72, p. 209.)

independent of the loop nonlinearities. DF approximations have not, as yet, entered our discussion. The exact output in response to any input is obtained by solving for $x(t)$ on a purely linear basis,

$$x(t) = \mathcal{L}^{-1}[X(s)] = \mathcal{L}^{-1}\left[\frac{L_1(s)R(s)}{1 + L_1(s)L_2(s)}\right] \quad (3.5-5)$$

operating on the resulting time function by N , and passing the output of N through $L_2(s)$.

It is possible that either or both of N^* and L_2^* will be different from N and L_2 , respectively. This can be due either to an *inaccuracy* in the model elements or to the use of an *intentionally* inexact model (Ref. 72). As an example of a situation where an inexact model should be intentionally employed, consider the system of Fig. 3.5-5, subject to a ramp input $r(t) = Mt$. As a consequence of the integration in $L_2(s)$ [hence in $L_2^*(s)$], the output signal $c(t)$ will remain *at rest* if the following relationship holds:

$$K\delta \geq M \quad (3.5-6)$$

for in this instance a *constant* value of x less than δ results, which never exceeds the threshold of N . The difficulty, of course, is due to the integration in $L_2^*(s)$. System response degradation is perhaps at an all-time high; the use of an inexact model is certainly indicated. A first choice for $L_2^*(s)$ might be an *approximate* integration, $L_2^*(s) = K/(\tau s + 1)$. Certainly the steady-state ramp response problem would be avoided. Note that this example points out the need for calculation of the system input-output response properties, independent of the fact that the loop transfer may be linear and possess suitable stability.

To investigate loop stability under any of the circumstances for which $N^* \neq N$ and/or $L_2^* \neq L_2$, we employ the DF representation for N and

$1 - N$, which we denote, as usual, by $N(A, \omega)$ and $1 - N(A, \omega)$, respectively. Representing the errors and/or inaccuracies in L_2^* and $1 - N^*$ by additive terms (denoted by Δ) as follows:

$$\begin{aligned} L_2^*(s) &= L_2(s) + \Delta L_2(s) \\ 1 - N^*(A, \omega) &= 1 - N(A, \omega) - \Delta N(A, \omega) \end{aligned} \quad (3.5-7)$$

it is easily demonstrated with reference to Fig. 3.5-4b that loop stability is now governed by the equation

$$1 + L_1(s)L_2(s) \left\{ 1 - \Delta N(A, \omega) + \frac{\Delta L_2}{L_2}(s) [1 - N(A, \omega) - \Delta N(A, \omega)] \right\} = 0 \quad (3.5-8)$$

The long term in braces can be treated as some new amplitude- and frequency-dependent nonlinearity, with DF analysis proceeding as described in previous sections of this chapter.

Generally speaking, the design with exact models is unnecessary, perhaps difficult, and can lead to severely degraded system performance unless care is taken to avoid situations such as the one described above. The use of intentionally inexact models, however, represents an attractive area for nonlinear compensation.

OTHER NONLINEAR NETWORKS

Linear minimum-phase networks are such that input-output gain and phase relationships are themselves related. Therefore choice of one constrains the form of the other. Linear non-minimum-phase networks display more phase lag than the corresponding minimum-phase networks (i.e., for the same attenuation characteristics), and thus they are of little use in system compensation. Nonlinear networks can be synthesized, however, where DF gain and phase relationships are chosen separately to suit the designer (Ref. 59). An example of one such circuit is the nonlinear Clegg integrator discussed in Sec. 2.4. It was shown to have a DF with an amplitude vs. frequency characteristic identical with that of a linear (minimum-phase) integrator and an associated DF phase of -38° as opposed to -90° for the linear integrator. This phase characteristic is certainly a desirable feature for loop compensation purposes, in which excessive phase lag can cause degraded or even unstable system behavior.

Systems with saturation in a tachometer feedback path may be unstable at large-feedback-signal levels. A useful compensation network would reduce the loop gain to large signals. Large-amplitude oscillations can occur in a conditionally stable system (where a gain reduction leads to instability). This behavior can be compensated by the use of a nonlinear network which provides phase lead at large-signal levels. Systems with

backlash all too often display low-amplitude limit cycles. A network providing phase lead at low-amplitude signals can be used for stabilization in these instances. Networks providing phase lead at low amplitudes are also useful in the stabilization of systems with spring-coupled coulomb friction (Ref. 60). Mechanizations for the three nonlinear networks indicated above are illustrated in Fig. 3.5-6. Note that the mechanizations indicated are not unique; many other useful forms can be developed.

The first of the networks illustrated is simply a gain-changing element drawn in a feedback mechanization. As such, the DF for this network can be derived directly as shown in Sec. 2.3. We shall present an approximate DF derivation, however, which will find real use when applied to the nonlinear networks of Fig. 3.5-6b and c, for which the exact DF cannot be

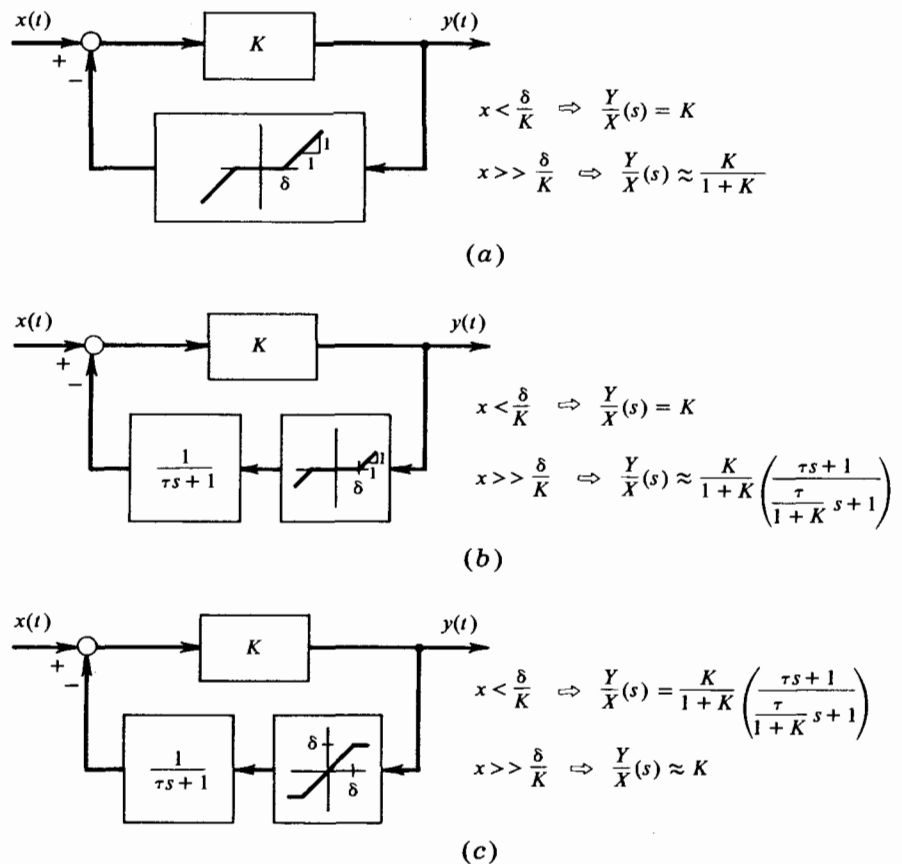


Figure 3.5-6 Nonlinear networks providing (a) gain reduction at large-signal levels; (b) phase lead at large-signal levels; (c) phase lead at small-signal levels.

found in any convenient analytic way. Let $N^\circ(A)$ represent the DF for the feedback dead-zone element in Fig. 3.5-6a. Now assume a sinusoidal *output* (we have already used this artifice in the study of implicit dynamic nonlinearities in Sec. 2.4)

$$y = A \sin \omega t \quad (3.5-9)$$

Since for this network there is no internal phase shift, a first-harmonic loop balance (ignoring fed-back harmonics) is achieved when

$$[A_{in} \sin \omega t - AN^\circ(A) \sin \omega t]K = A \sin \omega t \quad (3.5-10)$$

from which the frequency-invariant input-output DF is found to be

$$\begin{aligned} N(A) &= \frac{A}{A_{in}} \\ &= \frac{1}{N^\circ(A) + 1/K} \end{aligned} \quad (3.5-11)$$

Several observations are in order. First, the DF is a function of A (at the output) as opposed to A_{in} . This presents no particular difficulty in its use, of course; one need only recognize that a given magnitude of the DF corresponds to a particular value of A . Second, the formulation, although approximate, does yield exact results in the limiting cases $A_{out} \rightarrow 0$ [where $N(A_{out}) = K$] and $A_{out} \rightarrow \infty$ [where $N(A_{out}) = K/(1 + K)$], with small errors elsewhere. For example, in the case where $K = 1$ and $\delta = 1$, the maximum error in the approximate DF referred to the exact DF is less than 2 percent. Third, and most important, in application of this DF the filter hypothesis must be reversed. That is, for perfect DF results we *require* certain harmonics at x (the residual is now associated with the input), which must be generated by the passage of a pure sinusoid through a linear system (the loop linear elements). Obviously, this is a physically unrealizable requirement. Hence *perfect* results cannot be achieved using this DF model except at signal levels causing only linear operation. Nonetheless, *useful* results can be achieved. In this connection, Smith (Ref. 95, p. 461) points out that reversal of a cause-effect relationship is sometimes physically indicated. For example, in the analysis of magnetic amplifiers with parallel ac windings or with a zero-impedance bias source, the flux is essentially sinusoidal. Therefore the DF should be formulated in terms of flux input and magnetomotive force output, although the physical phenomenon is usually described as magnetomotive force for cause and flux for effect.

Perhaps the best means of determining $N(A, \omega)$ in the case of frequency-variant nonlinear networks is actual laboratory testing of a piece of hardware or a computer simulation of the network. Alternatively, an approximate means of calculation such as demonstrated in this section can be employed.

Following the approximate DF calculation procedure, it is easily shown that the approximate DF for the network of Fig. 3.5-6b is given by

$$N(A, \omega) = \frac{1}{\sqrt{\left(\frac{1}{K}\right)^2 + \frac{2N^\circ(A)}{K(1 + \tau^2\omega^2)} + N^\circ(A)^2}} e^{j \tan^{-1} \left[\frac{KN^\circ(A)\tau\omega}{1 + \tau^2\omega^2 + KN^\circ(A)} \right]} \quad (3.5-12)$$

Similarly, the approximate DF for the network of Fig. 3.5-6c is identical with the expression above, with the replacement of $N^\circ(A)$ for dead zone by $N^\circ(A)$ for saturation. Note the *phase lead* of each of these DFs, as required at either large- or small-signal levels.

3.6 TREATMENT OF MULTIPLE NONLINEARITIES

It is quite possible that nonlinearities will be present at more than one station around a control loop. In a typical active satellite attitude control system, for example, both the stabilizing (on-off) gas jets and the error transducer (perhaps a sun sensor) are quite decidedly nonlinear. In a synchro-controlled heavy-gun positioning system, both synchro error transducer and (saturating) hydraulic motor are again nonlinear, although in this case both nonlinearities are apparent only at large-signal levels. In fact, the particular combination of nonlinear error transducer and nonlinear power element are common occurrences in control-system applications. Thus it is fitting that we examine the possibility of using DF techniques in the study of multiple nonlinearity systems.

GENERAL MULTIPLE NONLINEARITY SYSTEMS

By a general two-nonlinearity system is meant one such as is illustrated in Fig. 3.6-1, where $N_1(x_1, \dot{x}_1)$ and $N_2(x_2, \dot{x}_2)$, as well as $L_1(s)$ and $L_2(s)$, are different. Let us replace the two nonlinearities by their respective DFs, $N_1(A_1, \omega)$ and $N_2(A_2, \omega)$. The frequency response of this system can be determined by solution of the equation

$$\frac{X_1}{R}(j\omega, A) = \frac{1}{1 + N_1(A_1, \omega)L_1(j\omega)N_2(A_2, \omega)L_2(j\omega)} \quad (3.6-1)$$

provided that the use of the DF for linearization is permissible. Clearly, one requirement for this to be true is that $L_1(j\omega)$ and $L_2(j\omega)$ are *each* low-pass

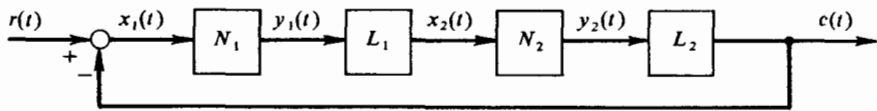


Figure 3.6-1 General two-nonlinearity system.

filters, so that the periodic inputs to both of the nonlinearities are nearly sinusoidal.

Equation (3.6-1) can be solved in a manner very similar to the first solution discussed in the one nonlinearity case, Sec. 3.3. By *choosing* values of A_1 and ω , $N_1(A_1, \omega)$ is determined and A_2 is, as a consequence, also determined.

$$A_2 = A_1 |N_1(A_1, \omega)| |L_1(j\omega)| \quad (3.6-2)$$

Thus the corresponding value of $N_2(A_2, \omega)$ is established, and Eq. (3.6-1) is soluble for X_1/R . This approach is in essence the same as considering the quantity $N_1(A_1, \omega)L_1(j\omega)N_2(A_2, \omega)$ equivalent to a new nonlinearity, $N(A_1, \omega)$, and proceeding as in earlier discussions of the single nonlinearity case.

Limit cycle study for this system corresponds to the determination of nontrivial solutions to the set of four relationships

$$\begin{bmatrix} N_1 & -1 & 0 & 0 \\ 0 & L_1 & -1 & 0 \\ 0 & 0 & N_2 & -1 \\ 1 & 0 & 0 & L_2 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \end{bmatrix} = 0$$

or equivalently,

$$\begin{vmatrix} N_1 & -1 & 0 & 0 \\ 0 & L_1 & -1 & 0 \\ 0 & 0 & N_2 & -1 \\ 1 & 0 & 0 & L_2 \end{vmatrix} = 0 \quad (3.6-3)$$

which reduces to the single equation

$$1 + N_1(A_1, \omega)L_1(j\omega)N_2(A_2, \omega)L_2(j\omega) = 0 \quad (3.6-4)$$

Of course, we come to the same conclusion by setting the denominator of Eq. (3.6-1) to zero, corresponding to writing the characteristic equation for the purely linear case. Equation (3.6-4) can be solved by any of the techniques for limit cycle determination discussed in Sec. 3.1 by treating the quantity $N_1(A_1, \omega)L_1(j\omega)N_2(A_2, \omega)$ as an equivalent single nonlinearity $N(A_1, \omega)$. When $L_1(j\omega)$ is a low-pass function, this simplified approach will be in small error. On the other hand, should $L_1(j\omega)$ turn out to be a lead-lag or similar non-low-pass compensation network, the simplified approach could lead to large error.

A more accurate solution to both the frequency response and limit cycle problems can be had by computing the DF for the complete chain $N_1(x_1, \dot{x}_1)L_1(j\omega)N_2(x_2, \dot{x}_2)$. This is accomplished by determining the actual first harmonic in y_2 when x_1 is a pure sinusoid. It is, of course, a more difficult approach. Regarding this approach, observe that even if N_1 and

N_2 are only amplitude-dependent, N will be both amplitude- and frequency-dependent because of the presumed frequency dependence of L_1 . Thus, in general, the treatment of multiple nonlinearity systems is bound to be somewhat more laborious than the treatment of single nonlinearity systems.

Example 3.6-1 Calculate the limit cycle frequency for the two-nonlinearity system of Fig. 3.6-2.

As N_2 is multiple-valued, it is convenient to seek a solution of Eq. (3.6-4) in the form

$$N_1(A_1)L_1(j\omega)L_2(j\omega) = -\frac{1}{N_2(A_2)}$$

The phase-shifting elements in this equation are $L_1(j\omega)L_2(j\omega)$ and $N_2(A_2)$. Thus, given values of ω must imply specific values of A_2 such that an appropriate loop phase shift occurs. Noting this phase requirement, we plot $L_1(j\omega)L_2(j\omega)$ and $-1/N_2(A_2)$ on amplitude-phase coordinates, as illustrated in Fig. 3.6-3. Any vertical line on this plot intercepts appropriate pairs of values (A_2, ω) .

A_1 is now expressed as

$$A_1 = A_2 |N_2(A_2)| |L_2(j\omega)|$$

and is calculated for several pairs of the values (A_2, ω) . This allows the calculation of several values of $|N_1(A_1)|$. Adjoining the resulting magnitudes (expressed in decibels, since this allows for graphical addition) to $L_1(j\omega)L_2(j\omega)$ at corresponding frequencies gives the *constructed locus*, whose intersections with $-1/N_2(A_2)$ satisfy both steady-state oscillation amplitude and phase-shift requirements, and hence yield the limit cycle solutions of interest. Figure 3.6-3 details the determination of one point on the constructed locus corresponding to the values $A_2 = 10$, $\omega = 3.9$.

By this process we observe as the limit cycle solution,

$$A_2 = 6.2 \quad \omega = 3.55$$

This compares extremely well with the analog computer solution

$$A_2(\text{exper}) = 6.2 \quad \omega(\text{exper}) = 3.49$$

particularly in view of the simplifying approximations employed.

LIMIT CYCLES IN SYMMETRIC MULTIPLE NONLINEARITY SYSTEMS

This section presents a method for transformation of a symmetric multiple nonlinearity system to a *single* nonlinearity system for the study of single-frequency limit cycle behavior. It is conjectured that there need be no

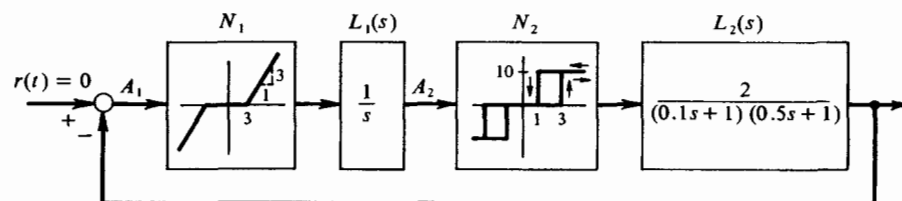


Figure 3.6-2 Example two-nonlinearity system.

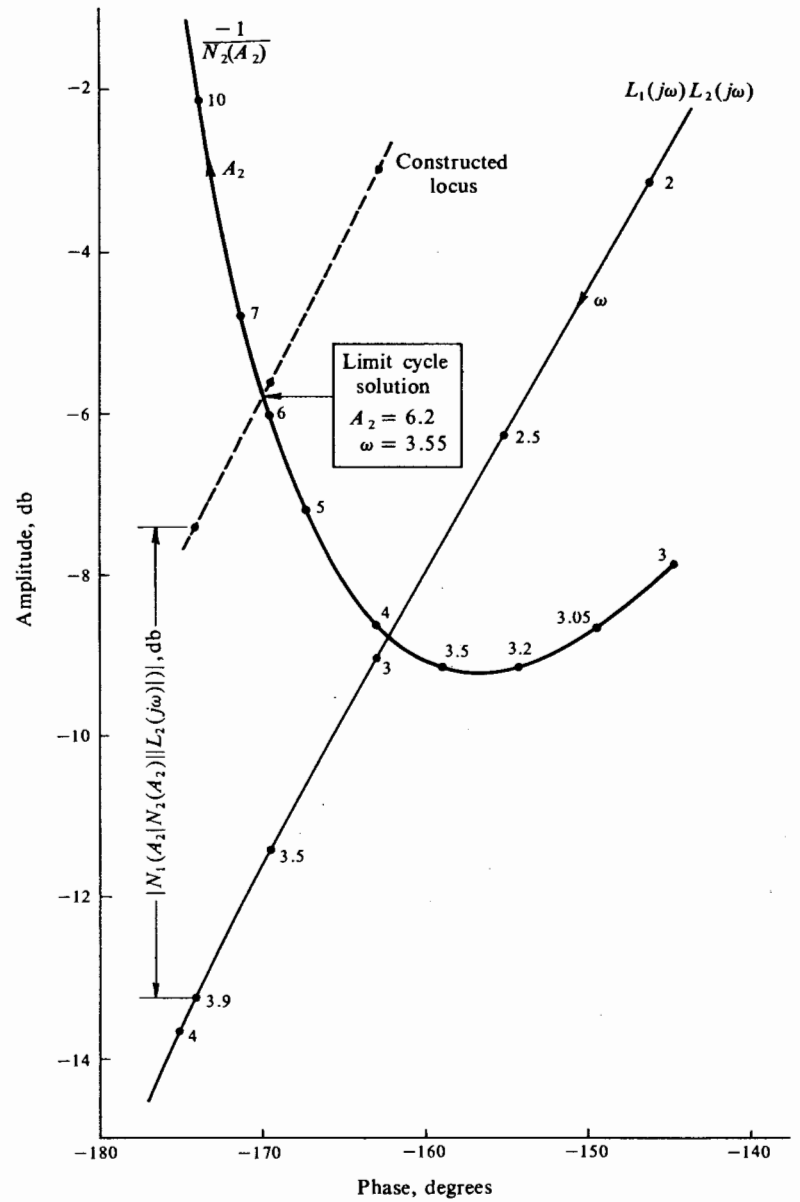


Figure 3.6-3 Graphical limit cycle solution in example two-nonlinearity system.

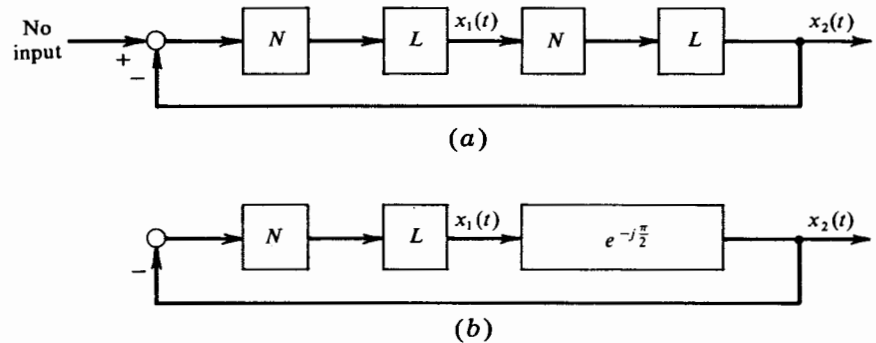


Figure 3.6-4 (a) Symmetric two-nonlinearity single-loop system. (b) Single-nonlinearity equivalent for limit cycle study.

restriction on the form of the linear elements, but that the nonlinearity is required to be odd. The resulting transformation, which apparently can be used for multiple- as well as single-loop systems, is exact. We first consider single-loop systems.

The essence of the matter is simply to observe that in a single-loop symmetric multiple nonlinearity system, any limit cycle which propagates must preserve its waveform at certain stations around the loop. Hence nonlinear operations on the limit cycle have a net effect *identical* with linear time delay (Ref. 27).

Consider the closed loop of Fig. 3.6-4. There is no input; we wish to study unforced oscillations. As the illustration implies, we have assumed a symmetric system in the sense that the two nonlinearities and the two linear elements are identical, respectively. It is plausible to argue that if the loop

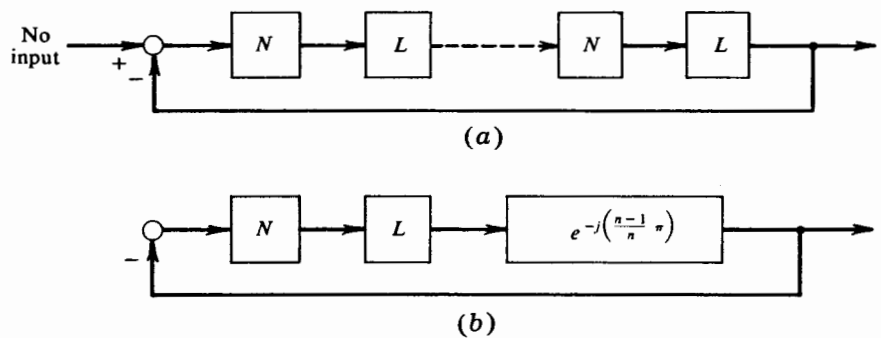


Figure 3.6-5 (a) Symmetric n -nonlinearity single-loop system. (b) Single-nonlinearity equivalent for limit cycle study.

does support a limit cycle, then, except for phase, the waveforms $x_1(t)$ and $x_2(t)$ must also be identical. In order to sustain a limit cycle we must have

$$x_1(t) = x_2\left(t + \frac{T_0}{4}\right) \quad (3.6-5)$$

where T_0 is the limit cycle period. This follows from the requirement for a loop phase shift of -360° in order to sustain an oscillation and the fact that the summing junction provides -180° of phase. Thus each of the pairs NL effectively provide a phase shift of one-quarter period, or -90° . Hence it is argued that an equivalent representation for the system of Fig. 3.6-4a is as shown in Fig. 3.6-4b. Further, it is argued that this representation must be exact for limit cycles of the assumed form.

By a continuation of the same argument, we conclude the results depicted in Fig. 3.6-5 for a symmetric n -nonlinearity single-loop system.

Example 3.6-2 Determine the functions ω_0 versus ζ ($\omega_n = 1$) and ω_0 versus ω_n ($\zeta = 0.5$) for a three-nonlinearity single-loop system with identical linear elements given by

$$L(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

and identical ideal-relay nonlinearities.

In the single nonlinearity equivalent for the three-nonlinearity case, the loop linear element is $L(s) \exp[-j(2\pi/3)]$. Hence $L(j\omega_0)$ must supply 60° of phase lag, as follows:

$$\angle L(j\omega_0) = -\tan^{-1} \frac{2\zeta\omega_n\omega_0}{\omega_n^2 - \omega_0^2} = -\frac{\pi}{3}$$

Solving for ω_0 yields

$$\omega_0 = \frac{\zeta\omega_n}{\sqrt{3}} \left[\sqrt{1 + \left(\frac{3}{\zeta^2}\right)} - 1 \right]$$

Theoretical and experimental results are plotted in Fig. 3.6-6.

As can be seen, the theoretical results are quite good. By observing simulated outputs of the linear elements, the theoretical phase shift of -60° per block (NL constituting a block) is verified. The slight discrepancy between theoretical and experimental results for high ζ is a consequence only of the approximations in describing function theory (i.e., waveforms in this region are highly nonsinusoidal). Exact analyses of the equivalent single-nonlinearity systems must lead to the exact limit cycle predictions. During simulation it was noted that all results were initial-condition-independent, as expected.

It is possible to extend these results to the cases of coupled multiple-loop systems. Linearly coupled nonlinear loops and nonlinearly coupled linear loops can both be studied, although only the latter case leads to unique results. Before discussing the possibility of nonuniqueness, let us first obtain the equivalent single-nonlinearity system.

Consider the system of Fig. 3.6-7a, where L_1 and L_2 are linear networks and N is a nonlinear operator. The symmetry in this system is apparent.

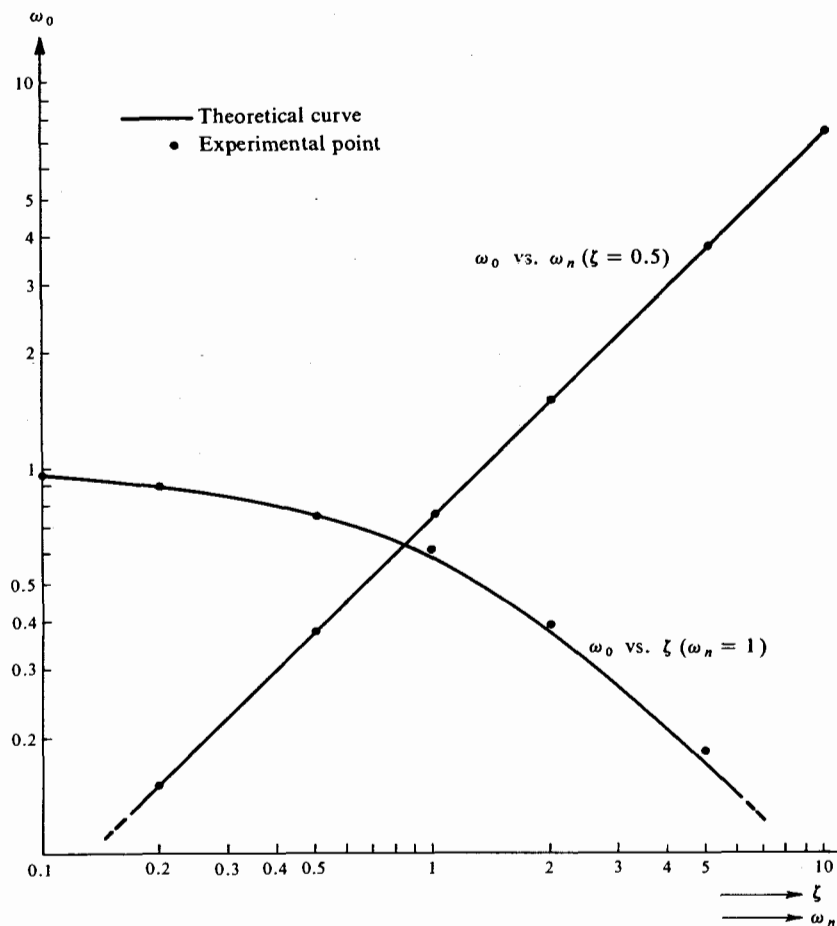


Figure 3.6-6 Experimental results for a symmetric three-nonlinearity single-loop system. ω_0 = limit cycle frequency.

Following the argument presented earlier, we should assert that $x_1(t)$ and $x_2(t)$ are identical in waveshape, but possibly of different phase. Since the phase shift from x_1 to x_2 and back to x_1 must be -360° or a multiple thereof, and since the phase shift from x_1 to x_2 is identical with that from x_2 to x_1 , we should argue that in order for the coupled system to oscillate at one frequency with a repeated waveshape, the phase shift from x_1 to x_2 must be a multiple of -180° . Hence we obtain the equivalent system of Fig. 3.6-7b. It is of the single-loop variety, with an equivalent linear element $L_1(s) \pm L_2(s)$.

Other cases can be treated in a similar manner. Nonlinearly coupled

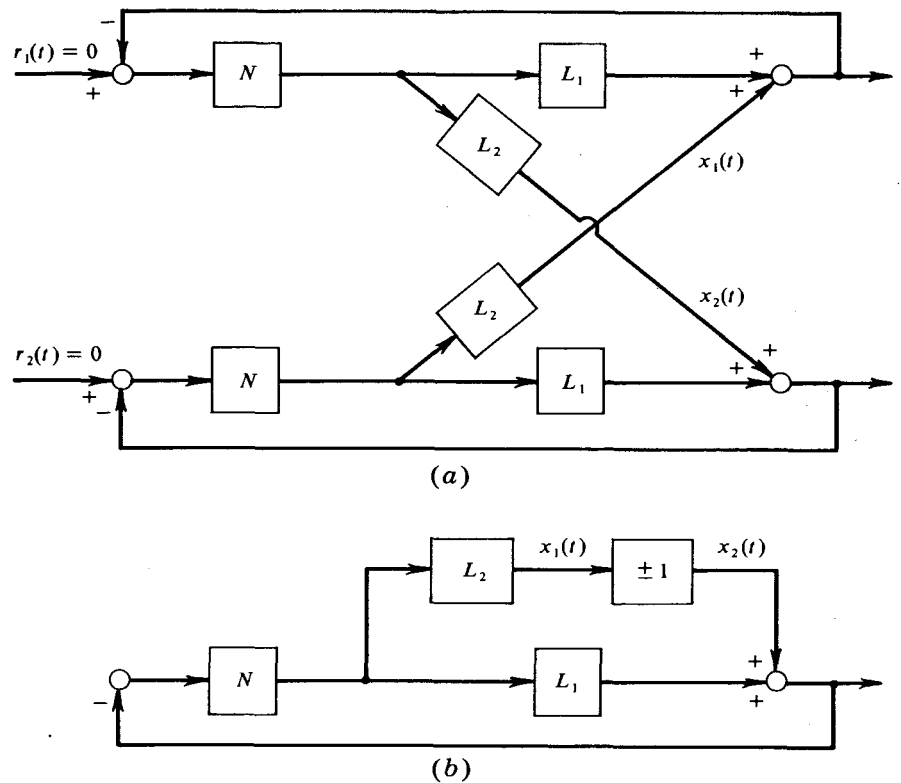


Figure 3.6-7 (a) Symmetric two-nonlinearity multiple-loop system. (b) Single-nonlinearity equivalent for limit cycle study.

linear loops, for example, degenerate to the single-loop case presented in the first part of this discussion, provided that one of the coupling elements contains a sign-changing term. In this situation any limit cycles which occur *must* couple both linear loops, and the results so obtained are unique.

As an example of a possibility for which this model is incomplete, consider the case where N and $L_1(s)$ are of such forms that each of the two loops present when $L_2(s) = 0$ can sustain two or more stable limit cycles. Now let us take the situation where both loops are supporting different limit cycles and $L_2(s)$ is small. The equivalent representation of Fig. 3.6-7b fails to account uniquely for this situation, indicating only that two stable limit cycles can indeed occur for sufficiently small values of $L_2(s)$. The equivalent representation does allow study of limit cycle behavior, but it excludes some of the possible limit cycle cases.

3.7 ACCURACY OF THE DF APPROXIMATION¹

As is so often the case in work with various types of describing functions, accuracy estimators are not part of the theoretical bundle from which the describing function itself is formulated. Additional means must be sought out, with which to provide accuracy estimates. These schemes must inevitably take cognizance of the *residual*, that part of the nonlinearity output ignored in the actual describing function development. In the case of the sinusoidal-input describing function, the residual consists of the entire sinusoidally forced nonlinearity output, minus the fundamental component. These harmonics, along with the linear elements, determine whether, and to what degree, DF solution of a problem will be successful.

In a number of text examples thus far, we have seen that results obtained through use of the DF approximation have been quite successful from a qualitative point of view, and that quantitative results have as well been generally excellent. It is also possible to demonstrate that both qualitative and quantitative conclusions can be in gross error, as shall be done later in this section. By examination of a few selected examples, we shall see that satisfaction of the *filter hypothesis* requirement is indeed the keystone of DF success. Moreover, we shall be in a position to estimate the accuracy of DF

¹ The text discussion throughout this chapter is based on the *filter hypothesis*, namely, that the loop linear elements attenuate nonlinearity output harmonics to the point that the nonlinearity input waveform is very nearly sinusoidal. Aizerman (Ref. 1) points out that for nonlinearities characterized by $y = mx + \mu\varphi(x)$, where μ is a small parameter, if for $\mu = 0$ the resulting closed-loop system has a complex pair of very lightly damped poles, the nonlinearity input will also be very nearly sinusoidal. He calls these the conditions of the *autoresonance hypothesis*.

In the case of autoresonance we can postulate beforehand that the limit cycle frequency is very nearly equal to the natural frequency of the lightly damped pole pair. Call this frequency ω^* . Then a small correction to ω^* , called $\delta\omega$, is found from the expansion of

$$L(j\omega^* + j\delta\omega) = -\frac{1}{N(A)}$$

in the form

$$L(j\omega^*) + \left. \frac{dL(j\omega)}{d\omega} \right|_{\omega=\omega^*} \delta\omega = -\frac{1}{N(A)}$$

Graphical solution of this equation is readily implemented.

Aizerman (*op. cit.*) argues that when $N(A)$ is real, $\delta\omega = 0$, and both the *autoresonance hypothesis* and the *filter hypothesis* lead to identical results. However, when $N(A)$ is a complex function of A , the employment of the *autoresonance hypothesis* in the study of systems which satisfy the *filter hypothesis* can lead to considerable error. On the other hand, use of the *filter hypothesis* in situations where the *autoresonance hypothesis* is satisfied does not lead to contradictory results, and in fact gives a more accurate solution. For these reasons and the additional fact that in problems of automatic control the *autoresonance hypothesis* is rarely satisfied, we shall pursue this topic no further.

results a priori, an essential feature of DF theory, based upon casual inspection of system linear elements and nonlinearities.

An independent discussion of the filter hypothesis is presented in Appendix G.

IMPORTANCE OF HARMONICS

This topic is profitably pursued first by example, and later by discussion. We shall examine five different systems in some detail, emphasizing the role played by harmonics in DF success or failure.

The first three examples are studies of a conservative second-order system with various nonlinearities: piecewise-linear *preload*, *harmonic*, and *odd polynomial* characteristics, in that order. The first example could arise physically from an imperfect electronic amplifier simulation (finite output impedance) of an ideal relay controlling a motor of negligible time constant with a pure inertia load. The second example is a study of large-amplitude oscillations of an ideal pendulum. The last of the first three examples could have its physical origin in the behavior of a mass on a nonlinear polynomial spring. Exact solutions are derived, along with DF solutions for these examples, in order that percent error in DF utilization can be established. The fourth example is a brief study of a pulse-excited damped clock pendulum, with the exact solution also given. The last example concerns a temperature control loop and the apparent reasons for which DF analysis fails to predict behavior of this system appropriately.

Example 3.7-1 Determine the free-oscillation relationships which exist in the conservative second-order nonlinear system of Fig. 3.7-1. $N(x)$ is the *preload* nonlinearity.

The differential equation governing the behavior of this system is

$$\ddot{x} + y(x) = 0$$

By defining a new quantity $z = \dot{x}$ and writing $\ddot{x}$ in the form $z dz/dx$, we can separate variables to obtain

$$z dz = -y(x) dx$$

Integrating this result from A_0 , the peak value of x at which z is zero, to the literal value x , we get

$$\frac{z^2}{2} = - \int_{A_0}^x y(x) dx \quad \text{or} \quad z = \left[2 \int_x^{A_0} y(x) dx \right]^{\frac{1}{2}}$$

If the nonlinearity is odd, the free-oscillation waveform will be symmetric with respect to a quarter cycle. Using the above solution for z , we can establish T_0 , the oscillation period, as four times the interval during which x grows from 0 to A_0 .

$$T_{0(\text{exact})} = 4 \int_0^{A_0} \left[2 \int_x^{A_0} y(x) dx \right]^{-\frac{1}{2}} dx$$

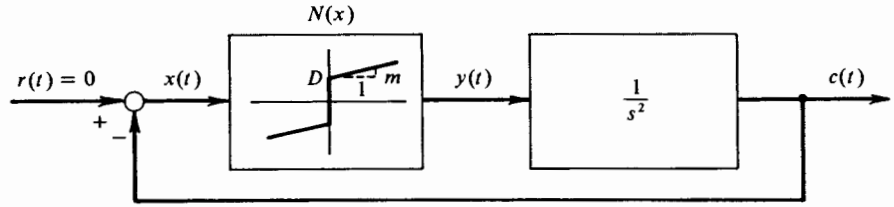


Figure 3.7-1 Nonlinear conservative second-order system.

The free-oscillation frequency is therefore determined exactly as

$$\omega_{0(\text{exact})} = \frac{2\pi}{T_{0(\text{exact})}} = \frac{\pi}{\sqrt{2}} \left\{ \int_0^{A_0} \left[\int_x^{A_0} y(x) dx \right]^{-\frac{1}{2}} dx \right\}^{-1} \quad (3.7-1)$$

We can now specialize this result to the case of a preload nonlinearity. Proceeding as required,

$$\begin{aligned} \int_x^{A_0} y(x) dx &= \int_x^{A_0} \left(mx + D \frac{x}{|x|} \right) dx \\ &= \frac{m}{2} A_0^2 + D |A_0| - \frac{m}{2} x^2 - D |x| \end{aligned}$$

which yields, as the solution for ω_0

$$\begin{aligned} \omega_{0(\text{exact})} &= \frac{\pi}{\sqrt{2}} \left[\int_0^{A_0} \frac{dx}{\sqrt{(m/2)A_0^2 + DA_0 - (m/2)x^2 - Dx}} \right]^{-1} \\ &= \frac{\pi}{\sqrt{2}} \left[\frac{\sqrt{2}}{\sqrt{m}} \sin^{-1} \left(\frac{mx + D}{mA_0 + D} \right) \Big|_0^{A_0} \right]^{-1} \\ &= \frac{\sqrt{m}}{1 - \frac{2}{\pi} \sin^{-1} \left(\frac{1}{1 + (m/D)A_0} \right)} \end{aligned} \quad (3.7-2a)$$

Observe that this system does not *limit-cycle*; in contrast, it supports a *conservative free oscillation*, the difference between the two being that whereas the former corresponds to a single *discrete* equilibrium state of the system over some finite range of initial conditions, the latter corresponds to a continuous *spectrum* of equilibrium states dependent directly on the initial conditions.

For comparison let us obtain the frequency of oscillation by DF usage. In DF application the differential equation of motion of the system is written in the linear form

$$\ddot{x} + N(A)x = 0$$

from which it is immediately evident that the frequency of oscillation is

$$\begin{aligned} \omega_{0(\text{DF})} &= \sqrt{N(A_0)} \\ &= \sqrt{\frac{4D}{\pi A_0} + m} \\ &= \sqrt{m} \sqrt{1 + \frac{4}{\pi} \frac{D}{mA_0}} \end{aligned} \quad (3.7-2b)$$

where $N(A)$ is as calculated in Sec. 2.3.

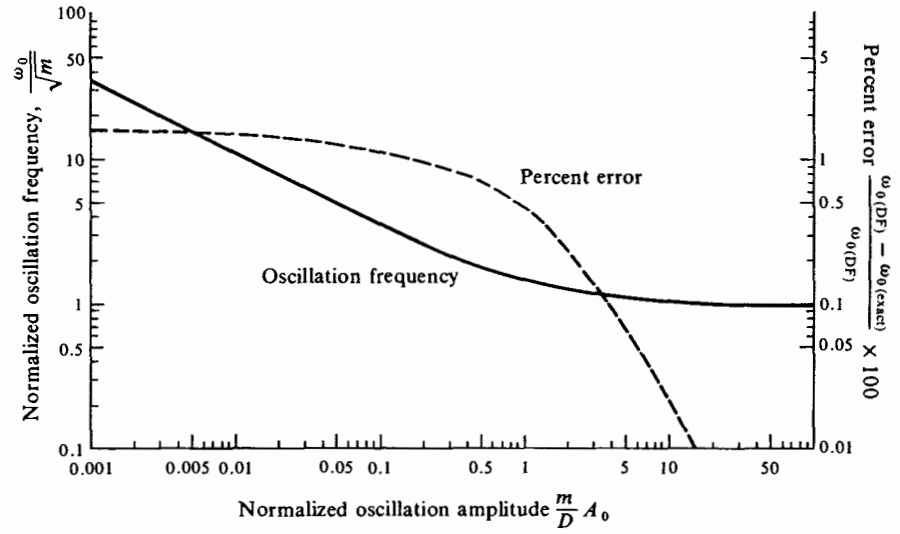


Figure 3.7-2 Conservative-free-oscillation frequency vs. amplitude in a doubly integrating second-order system with a preload nonlinearity.

Figure 3.7-2 is a plot of $\omega_0/\sqrt{m}$ for both the DF and exact oscillation frequency calculations. The maximum error in DF approximation over the entire range of mA_0/D is 1.6 percent, occurring in the limit as $mA_0/D \rightarrow 0$. In this region the preload nonlinearity is indistinguishable from an ideal relay. As $mA_0/D \rightarrow \infty$, the purely linear case is approached, and the error goes to zero. The free-oscillation frequency curves are indistinguishable.

Example 3.7-2 Determine the oscillation frequency of a simple pendulum of length l in a constant gravity field g .

The equation with which we are concerned is well known, namely,

$$\ddot{x} + \frac{g}{l} \sin x = 0$$

With the aid of the identity

$$\cos x = 1 - 2 \sin^2 \frac{x}{2}$$

and the definition for φ ,

$$\sin \varphi = \csc \frac{A_0}{2} \sin \frac{x}{2} \quad \begin{cases} x = A_0 \Rightarrow \varphi = \frac{\pi}{2} \\ x = 0 \Rightarrow \varphi = 0 \end{cases}$$

the exact solution can be obtained from Eq. (3.7-1), in the form

$$\omega_{0(\text{exact})} = \frac{\pi}{2} \sqrt{\frac{g}{l}} \left[K \left(\sin \frac{A_0}{2} \right) \right]^{-1} \quad (3.7-3a)$$

where $K(\arg)$ is the complete elliptic integral of the first kind. For $k < 1$,

$$K(k) = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}}$$

Another form of this integral appears in Eq. (2.3-29).

For comparison, the DF solution is obtained directly by use of Eq. (2.3-23).

$$\begin{aligned} \omega_{0(\text{DF})} &= \sqrt{\frac{g}{l}} \sqrt{N(A_0)} \\ &= \sqrt{\frac{g}{l}} \sqrt{\frac{2J_1(A_0)}{A_0}} \end{aligned} \quad (3.7-3b)$$

where $J_1(A)$ is the Bessel function of order 1 and argument A .

The normalized frequency of conservative free oscillations obtained by each of these methods is plotted versus A_0 in Fig. 3.7-3. The percent error in approximation is also shown. From this curve we see that for oscillation amplitudes less than 75° the error in approximation is 0.1 percent. By way of contrast, the error at 75° obtained by linearizing the sine function for small angles ($\sin x \sim x$) is 10 percent. At an amplitude of 130° , the error in DF approximation is 2 percent, and the linear approximation is in error by 32 percent.

For completeness we mention that the DF-predicted oscillation amplitudes for which the linearly predicted oscillation frequencies are in error as indicated above are themselves in error by 1 percent (at an exact amplitude of 75°) and 4 percent (at an exact amplitude of 130°).

Example 3.7-3 Compare the DF and exact free-oscillation frequency solutions for a conservative second-order system with a one-term odd polynomial nonlinearity.

The odd nonlinearity is defined by

$$y = \begin{cases} x^n & n \text{ odd} \\ x^{n-1} |x| & n \text{ even} \end{cases}$$

Proceeding as in the previous examples, the exact solution is found in the form

$$\begin{aligned} \omega_{0(\text{exact})} &= \frac{\pi}{\sqrt{2}} \left(\int_0^{A_0} \frac{dx}{\sqrt{[1/(n+1)](A_0^{n+1} - x^{n+1})}} \right)^{-1} \\ &= \frac{\pi A_0^{(n-1)/2}}{\sqrt{2}\sqrt{n+1}} \left(\int_0^1 \frac{du}{\sqrt{1 - u^{n+1}}} \right)^{-1} \end{aligned} \quad (3.7-4a)$$

For values of n less than 3, the definite integral above can either be evaluated directly ($n = 1$) or expressed in terms of standard elliptic integrals ($n = 2, 3$). Values of n in excess of 3 can be accommodated by numerical integration. The DF solution is, by virtue of Eq. (2.3-20),

$$\begin{aligned} \omega_{0(\text{DF})} &= \sqrt{N(A_0)} \\ &= \sqrt{\frac{2}{\sqrt{\pi}} \frac{\Gamma[(n+2)/2]}{\Gamma[(n+3)/2]}} A_0^{(n-1)/2} \end{aligned} \quad (3.7-4b)$$

As both solutions for oscillation frequency display identical dependences upon A_0 , the percent error attendant on DF utilization can be assessed with a single calculation. Table 3.7-1 lists the result of this calculation for various nonlinearity powers from 1 ($y = x^1$,

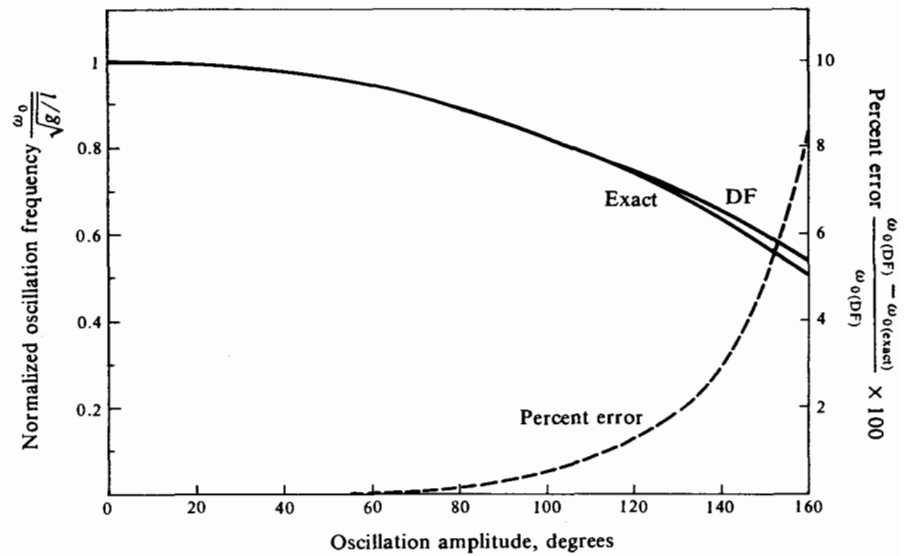


Figure 3.7-3 Conservative-free-oscillation parameters for a simple pendulum.

linear case) to 9 ($y = x^3$). Also tabulated is the ratio $|A_3/A_1|_x$ at the input to the nonlinearity. This ratio is computed by taking the harmonic amplitude ratio $|A_3/A_1|_y$ at the nonlinearity output (having assumed a sinusoidal input) and multiplying by $\frac{1}{3}$, the amount by which the linear element attenuates the third harmonic relative to the first harmonic.

A very significant, albeit quite expected, conclusion can be drawn from this tabulation, namely, that the DF solution accuracy improves with decreasing third-harmonic magnitudes

TABLE 3.7-1 FREE-OSCILLATION RESULTS IN A CONSERVATIVE SECOND-ORDER SYSTEM WITH AN ODD POLYNOMIAL NONLINEARITY

n	Describing function result	Exact result	Percent error	Harmonic amplitude ratio at x
	$A_0^{(1-n)/2} \times \omega_{0(DF)}$	$A_0^{(1-n)/2} \times \omega_{0(exact)}$	$\frac{\omega_{0(DF)} - \omega_{0(exact)}}{\omega_{0(DF)}} \times 100\%$	$\left \frac{A_3}{A_1} \right _x \times 100\%$
1	1	1	0	0
2	0.921	0.914	0.76	2.2
3	0.866	0.847	2.2	3.7
4	0.824	0.795	3.5	4.8
5	0.791	0.748	5.3	5.6
6	0.763	0.711	6.7	6.2
7	0.740	0.678	8.4	6.7
8	0.719	0.648	10.0	7.1
9	0.702	0.622	11.3	7.4

at the nonlinearity input. For values of the ratio $|A_3/A_1|_z$ less than 5 percent, it is seen that DF results are accurate to better than 5 percent. This result is quite representative of DF accuracy in second- and higher-order systems.

Example 3.7-4¹ Determine the limit cycle amplitude of a pulse-excited clock pendulum with damping ratio ζ . The velocity variation resulting from impact is a constant, V .

The equation we are dealing with is

$$\ddot{x} + 2\zeta\dot{x} + x = f(x, \dot{x})$$

where the nonlinear characteristic is as shown in Fig. 3.7-4. By assuming a sinusoidal input, $x = A \sin \omega t$, and computing the DF, we find

$$n_p(A, \omega) = 0 \quad n_q(A, \omega) = \frac{4D\delta}{\pi A^2}$$

Since $f(x, \dot{x})$ is a narrow pulse, it follows by integration of the differential equation of motion over the pulse duration that the change in velocity is equal to the pulse area. That is (τ = pulse-time duration),

$$V = \tau D$$

Continuing, the velocity of the assumed sinusoidal oscillation in the vicinity of $t = 0$ [$\dot{x}(0) = A_0\omega_0$] times the pulse-time duration must equal 2δ , the nonlinearity "on" width.

$$2\delta = A_0\omega_0\tau$$

Combining these equations, we get for the linearized nonlinearity the DF representation [Eq. (2.2-14)]

$$f(x, \dot{x}) = n_p(A_0, \omega_0)x + \frac{n_q(A_0, \omega_0)}{\omega_0} \dot{x} = \frac{2V}{\pi A_0} \dot{x}$$

Inserting this into the original equation of motion yields

$$\ddot{x} + 2\left(\zeta - \frac{V}{\pi A_0}\right)\dot{x} + x = 0$$

from which it is clear that we get a steady-state oscillation only if

$$A_{0(DF)} = \frac{V}{\pi\zeta} \quad (3.7-5a)$$

The subscript DF is used to distinguish this result from the exact limit cycle amplitude found by Magnus (*op. cit.*),

$$A_{0(exact)} = V \frac{\exp [-(\zeta/\sqrt{1-\zeta^2}) \tan^{-1}(\sqrt{1-\zeta^2}/\zeta)]}{1 - \exp [-(\pi\zeta/\sqrt{1-\zeta^2})]} \quad (3.7-5b)$$

Figure 3.7-5 is a plot of exact and approximate solutions as well as percent error in DF solution versus ζ . For low ζ the pendulum is nearly undamped and oscillates almost sinusoidally. The error in this region is small. The DF solution percent error is seen to

¹ This is one of many examples investigated by Magnus (Ref. 65) in an excellent paper on DF theory.

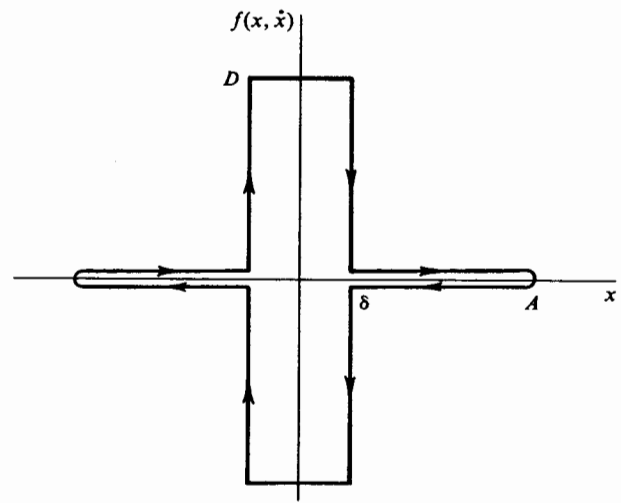


Figure 3.7-4 Nonlinear characteristic of clock-pendulum pulse exciter.

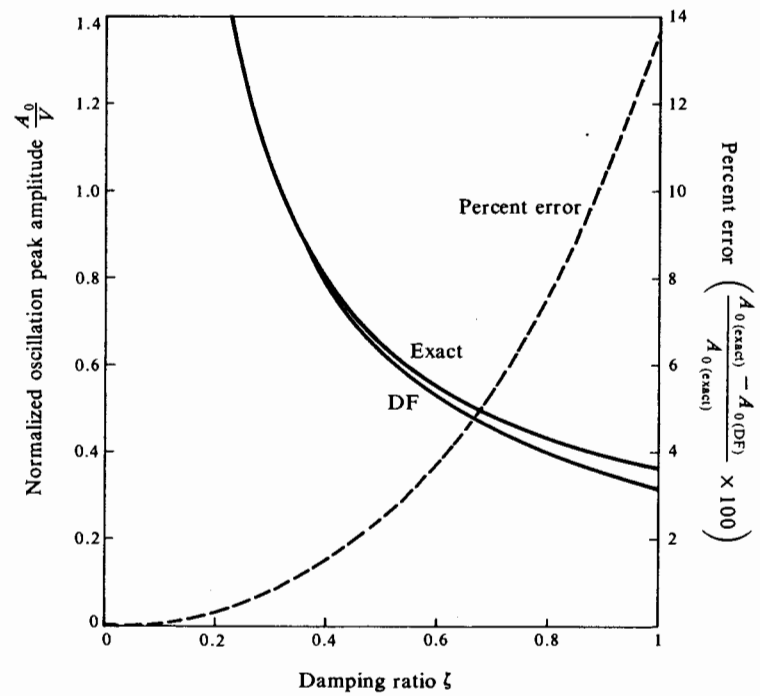


Figure 3.7-5 Solution for limit cycle amplitude of a pulse-excited clock pendulum.

increase with increasing ζ . We note here that with increasing ζ the oscillation becomes more nonsinusoidal. This is due to failure of the loop to attenuate higher harmonics sufficiently, relative to the fundamental.

Example 3.7-5 Investigate the limit cycle behavior of a temperature control system for which the block diagram is as shown in Fig. 3.7-6a.

Using DF methods, we immediately conclude that *no* oscillation is predicted to take place. The graphical construction pertinent to this conclusion is shown in Fig. 3.7-6b. The DF result is obviously in complete error, since a limit cycle must occur in this on-off system if $DK > \delta$. Why, then, we may ask, does the DF formulation yield such poor results?

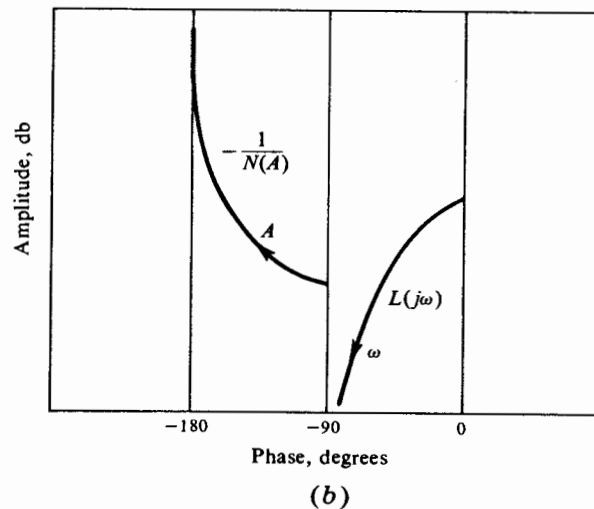
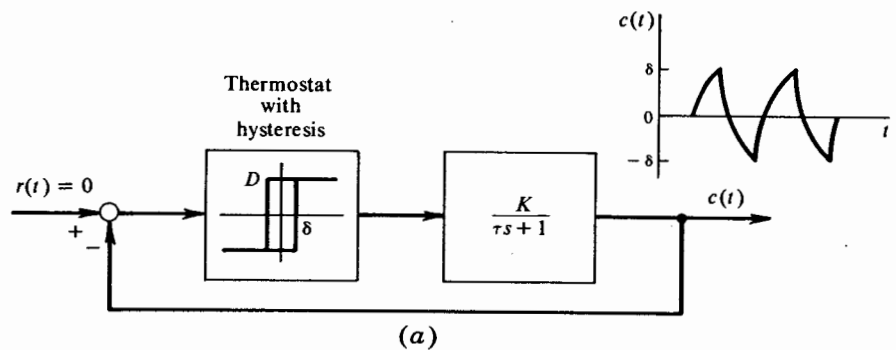


Figure 3.7-6 (a) Temperature-regulating system. (b) DF amplitude-phase-plot limit cycle construction.

The answer to this enigma becomes evident if the exact limit cycle waveform is considered. This has been illustrated in Fig. 3.7-6a, and is seen to consist of matched single-exponential segments. The waveform suffers slope discontinuities twice during each limit cycle period, and is, generally speaking, hardly sinusoidal. *It is waveform harmonic content which is responsible for the DF failure.*

Without knowledge of the exact waveform, however, these poor results can be and should have been anticipated. A simple calculation demonstrates this point. The third- to first-harmonic amplitude ratio has been shown to be $\frac{1}{3}$ at the nonlinearity output. The linear elements attenuate each of the harmonics by the factor $K/\sqrt{1 + n^2\omega_0^2\tau^2}$, where n is the harmonic number. Thus, at the input to the nonlinearity, we have

$$\left| \frac{A_3}{A_1} \right|_x = \frac{1}{3} \sqrt{\frac{1 + \omega_0^2\tau^2}{1 + 9\omega_0^2\tau^2}} \quad (3.7-6)$$

Without specifying the value of $\omega_0\tau$ it is clear that the following limiting inequality holds:

$$0.111 < \left| \frac{A_3}{A_1} \right|_x < 0.333 \quad (3.7-7)$$

That is, at the input to the nonlinearity, the third-harmonic amplitude is somewhere between 11.1 percent and 33.3 percent of the fundamental. As a result of this large amplitude at x , the unaccounted for phase shift of the third harmonic is sufficient to upset the original input zero-crossing assumptions (in this *on-off relay* problem only the input zero crossings, *not* its detailed waveshape, are influential on limit cycle behavior), and thus negate the DF calculation.

That the DF approach fails in the last example is not of the slightest consequence; there are better ways to study this low-order system.¹ What is important is the fact that clear indication exists which shows the basic assumptions of DF theory to be unsatisfied. We shall pursue this point further.

IMPORTANCE OF HARMONICS (CONTINUED)

Reviewing each of the foregoing examples, we can make one consistent observation, namely, that DF solution accuracy was always degraded by an increasing harmonic content at the input to the nonlinearity.

It is not sufficient merely to know the harmonic content at the nonlinearity output; the filtering influence of the linear elements must also be assessed. From what we have seen so far, a doubly integrating linear filter provides sufficient attenuation of higher harmonics to yield very reasonable DF results

¹ Exact solution by matching exponential waveforms at the switching boundaries, for example. Another method of study, the phase plane (see Chap. 1), is of general use for the efficient and complete study of first- and second-order nonlinear systems. Still another method, Tsytkin's exact method, is of use in the study of limit cycles in relay systems and is examined in some detail later in this chapter. At that time Example 3.7-5 is treated in a most satisfactory manner.

for a wide variety of nonlinearities, whereas a first-order linear lag filter insufficiently attenuates these harmonics and can lead to incorrect DF results. Many other examples have been studied along the same lines. The results indicate clearly that as the order¹ of the linear portion of the system *increases*, the error in DF application is *reduced*. This is an interesting situation since the phase plane and other related techniques are well suited to the study of second- (and occasionally third-) order systems. In this light the DF ought to be considered as an adjunct to these methods, for use in studying third- (and occasionally second-) as well as higher-order systems. That higher-order systems are more amenable to simple DF analysis is indeed a happy circumstance, being quite the opposite of a trend one normally anticipates in system work. The reason for this is simply that repeated integrations of any periodic waveform of frequency ω eventually reduce that waveform to a sinusoid of frequency ω . Consider an arbitrary periodic function $y(t)$ expressed in its complex Fourier series representation.

$$y(t) = \sum_{k=-\infty}^{\infty} A_k e^{jk\omega t} \quad (3.7-8)$$

Let us integrate this expression m times and dispose of zero-frequency terms.

$$\begin{aligned} y^{(m)}(t) &= \sum_{k=-\infty}^{\infty} \frac{A_k}{(jk\omega)^m} e^{jk\omega t} \\ &= \sum_{k=-\infty}^{\infty} A_k^{(m)} e^{jk\omega t} \end{aligned} \quad (3.7-9)$$

The superscript denotes the number of integrations performed. If we form the ratio of k th harmonic to first harmonic and let $m \rightarrow \infty$, we see indeed that only the first harmonic remains.

$$\frac{A_k^{(m)}}{A_1^{(m)}} = \frac{[A_k/(jk\omega)^m]}{[A_1/(j\omega)^m]} = \frac{1}{k^m} \frac{A_k}{A_1} \rightarrow 0 \quad \text{as } m \rightarrow \infty \quad (3.7-10)$$

The study of repeated integrations on a square wave points out dramatically how the fundamental square-wave frequency is eventually singled out. Figure 3.7-7 is a demonstration of this process, with each successive integral normalized in amplitude and freed of the bias resulting from integration. After the second integration the (parabolic) waveform already looks quite sinusoidal. After an infinite number of integrations, a pure sinusoid is converged upon. We hasten to point out that this exercise is meant only to lend physical credence to the accuracy development. Replacement of the pure integrations by first-order-lag filters would tend to slow down the convergence to a sinusoid, etc.

¹ More specifically, the excess of poles over zeros in the frequency region up to $5\omega_0$. This presumes no lightly damped pole pairs at the frequency of nonlinearity output harmonics.

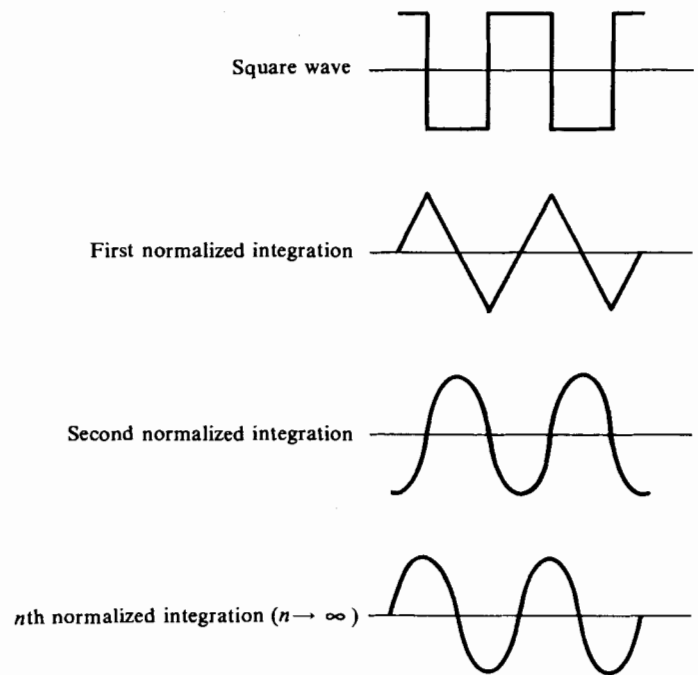


Figure 3.7-7 Effect of repeated integration on a square wave (normalized).

The use of asymptotic Bode plots is very convenient for determining the attenuation of all harmonic frequencies by the loop linear elements. Of course, the same information also exists on amplitude-phase and polar plots, but these plots are parametrized by ω , and hence must be carefully examined to ascertain actual attenuation ratios. Figure 3.7-8 shows the amplitude-phase plots for two different linear open-loop systems. Were the loops to be closed by the insertion of a non-phase-shifting nonlinearity, one ought to feel secure in his estimation that DF analysis would yield superior results in case *a* as opposed to case *b*, simply because of the observation that third- and higher-harmonic filtering is more pronounced in this case.

A word of caution with regard to the accuracy expected in graphical interpretation of DF solutions is in order. When the loci $L(j\omega)$ and $-1/N(A,\omega)$ indicated on the amplitude-phase plot are nearly orthogonal at the limit cycle intersection, results are ordinarily of 5 percent to 10 percent accuracy. Nearly parallel loci at intersection are apt to yield substantially poorer results. Low-frequency limit cycles often exist when the two loci approach each other at low frequencies but indicate no intersection. Such limit cycles can occasionally be determined by making a second-order

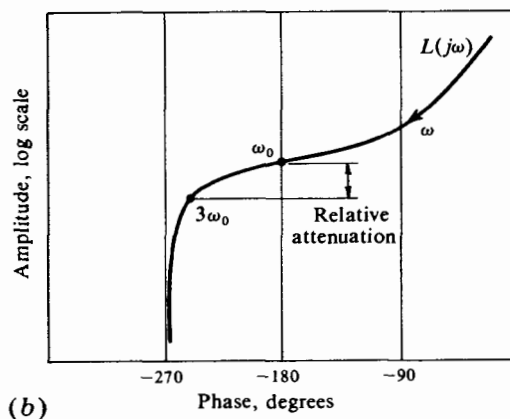
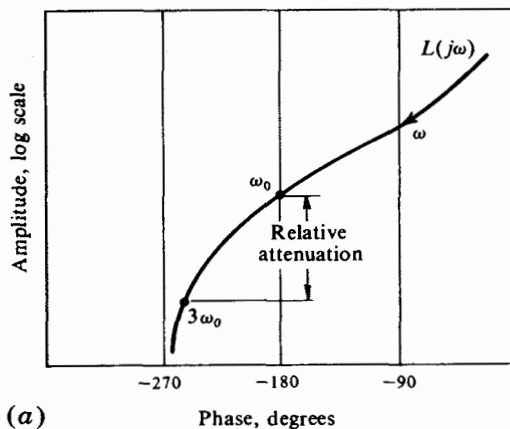


Figure 3.7-8 Estimation of the importance of harmonics by use of the amplitude-phase plane. (a) Less important; (b) more important.

approximation to the linear elements and studying the resultant system by phase-plane techniques. At any rate, in the latter instance it is desirable to determine a higher-order DF approximation as a check on DF accuracy. One such approximation is presented next. Another can be found in Popov (Ref. 88).

REFINED DF APPROXIMATION

Refining the DF formulation implies accounting for some of the higher harmonics comprising the residual, heretofore neglected. The means for achieving such DF refinement must be uncomplicated, for without this

constraint all the benefit of DF usage will be mitigated. We shall proceed with this point of view.

What we are after is essentially a new DF which describes the fundamental gain of the nonlinearity in the presence of the fundamental and *simultaneously* some of the higher harmonics. In Chap. 5 we shall see that *exact* calculations for one such DF are indeed possible, but quite complicated. Even then only the third harmonic is considered. Although such calculations definitely have their own domain of utility (study of subharmonic responses, for example), we should still like to seek out a DF second approximation, whose primary function is to serve as an accuracy check in areas of questionable DF results.

A consideration of the significance of feedback in linear systems provides us with the direction to follow, in order to find a scheme for DF refinement. Consider a unity feedback linear system with forward-path elements $L(j\omega)$ and input $R(j\omega)$. The output $C(j\omega)$ can be constructed by tracing the input around the loop an infinity of times. Thus, when the input is first applied, the output is, for an instant, $R(j\omega)L(j\omega)$. Presently, some feedback occurs, and the output is reduced by the amount $-R(j\omega)L^2(j\omega)$. Additional feedback occurs, and the following series can be constructed ($|L(j\omega)| < 1$):

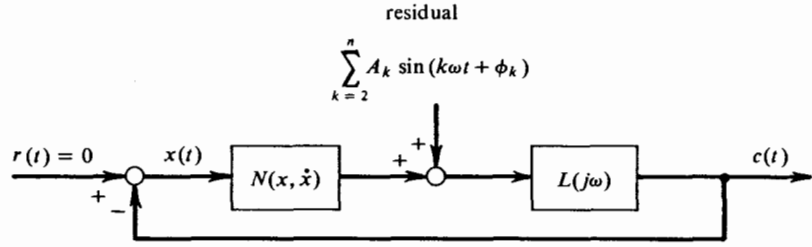
$$\begin{aligned} C(j\omega) &= R(j\omega)L(j\omega) - R(j\omega)L^2(j\omega) + R(j\omega)L^3(j\omega) - \cdots \\ &= R(j\omega)L(j\omega)[1 - L(j\omega) + L^2(j\omega) - \cdots] \\ &= R(j\omega) \frac{L(j\omega)}{1 + L(j\omega)} \end{aligned} \quad (3.7-11)$$

This, of course, is the familiar linear-system closed-loop transfer function result. We now make the observation that if $L(j\omega)$ is a low-pass function, higher powers of $L(j\omega)$ tend to zero in the frequency range of interest. Under this condition the transfer function can be *approximated* by taking only the first few terms of the series, namely,

$$\frac{C(j\omega)}{R(j\omega)} \approx L(j\omega)[1 - L(j\omega)] \quad (3.7-12)$$

This process is tantamount to considering only the first portion, as it were, of an infinite series of feedbacks. The same approach has been adopted for synthesizing a refined DF approximation (Ref. 34).

Suppose that by means of DF theory the nonlinear system of Fig. 3.7-9 has been found to oscillate at a frequency ω . In this first DF approximation, the input to the nonlinearity, $x_{(1)}(t)$, has been taken as a pure sinusoid. All higher harmonics were ignored. As a second approximation we now take the input to the nonlinearity to be the pure sinusoid calculated above plus the feedback quantity comprised of the filtered residual. We now compute the output of the nonlinearity in response to the second-approximation



First approximation: $x_{(1)}(t) = A \sin \omega t$

Second approximation: $x_{(2)}(t) = A \sin \omega t - \sum_{k=2}^n A_k |L(jk\omega)| \sin(k\omega t + \phi_k + \angle L(jk\omega))$

Figure 3.7-9 Formulation of a refined DF approximation.

input $x_{(2)}(t)$, and define the second-approximation DF in the conventional manner as the complex ratio of output first-harmonic term to input first-harmonic term. The effect of higher harmonics is implicit in this formulation, although they have by no means been accounted for exactly. To obtain corrected results, one then uses the refined DF and performs otherwise ordinary DF calculations. The technique is demonstrated in detail by the following example.

Example 3.7-6 Determine the refined DF solution for the limit cycle displayed by the third-order ideal-relay control system of Fig. 3.7-10. Relay drive levels are $\pm D$.

The first DF solution for the limit cycle frequency and amplitude is easily found to be

$$\omega_0 = \omega_n \quad A_0 = \frac{2DK}{\pi\zeta\omega_n}$$

Hence the first-approximation nonlinearity input is

$$x_{(1)}(t) = \frac{2DK}{\pi\zeta\omega_n} \sin \omega_n t$$

The residual consists of the odd harmonics associated with a square wave of amplitude D , as shown in Fig. 3.7-10. By passing this residual through $L(j\omega)$, we find as the second approximation to the nonlinearity input

$$x_{(2)}(t) = \frac{2DK}{\pi\zeta\omega_n} \sin \omega_n t - \sum_{k=3,5,\dots}^{\infty} \frac{4D}{\pi k} |L(jk\omega_n)| \sin[k\omega_n t + \angle L(jk\omega_n)]$$

where the amplitude and phase transfer of $L(jk\omega_n)$ are ($k > 1$)

$$\begin{aligned} |L(jk\omega_n)| &= \frac{K\omega_n^2}{k\omega_n \sqrt{(\omega_n^2 - k^2\omega_n^2)^2 + (2k\zeta\omega_n^2)^2}} \\ &= \frac{K}{k\omega_n \sqrt{(k^2 - 1)^2 + (2k\zeta)^2}} \end{aligned}$$

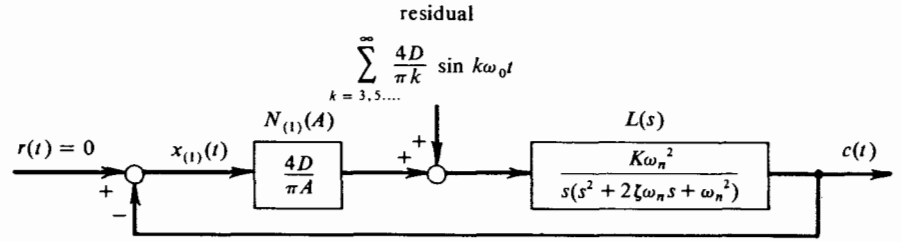


Figure 3.7-10 Third-order ideal-relay control system, first DF approximation.

and

$$\begin{aligned} \angle L(jk\omega_n) &= -\frac{\pi}{2} - \tan^{-1} \frac{2k\zeta\omega_n^2}{\omega_n^2 - k^2\omega_n^2} \\ &= -\frac{3\pi}{2} + \tan^{-1} \frac{2k\zeta}{k^2 - 1} \end{aligned}$$

Let us now restrict our attention to the third harmonic only. In principle, additional harmonics can be carried along, but for demonstration they serve only to clutter the issue. Thus $k = 3$, and $x_{(2)}(t)$ can be simplified to the following form:

$$x_{(2)}(t) = \frac{2DK}{\pi\zeta\omega_n} \left[\sin \omega_n t - \frac{\zeta}{36\sqrt{1 + (3\zeta/4)^2}} \cos \left(3\omega_n t + \tan^{-1} \frac{3\zeta}{4} \right) \right]$$

With this as the new input to the nonlinearity, the output will again be a square wave with fundamental-component amplitude $4D/\pi$. For this reason the *magnitude* of the refined DF is unchanged, remaining $4D/\pi A$. The *phase shift* of the refined DF is different from zero degrees, however, because of the presence of the third harmonic at the relay input. Presuming that the first DF approximation was not in great error, the sign of $x_{(2)}(t)$ will change at some value of t which is small ($t \approx 0$), allowing for an expansion of $x_{(2)}(t)$ in the vicinity of $t = 0$. Performing this expansion and setting $x_{(2)}(t)$ to zero yields

$$\omega_n t - \frac{\zeta}{36\sqrt{1 + (3\zeta/4)^2}} \left[\cos \left(\tan^{-1} \frac{3\zeta}{4} \right) - 3\omega_n t \sin \left(\tan^{-1} \frac{3\zeta}{4} \right) \right] = 0$$

Solving for $\omega_n t$, we find

$$\omega_n t = \frac{2\zeta}{9(8 + 5\zeta^2)}$$

Since the solution indicates a *positive* value for t , the phase shift associated with the refined DF is *lagging*. Thus the refined DF, designated $N_{(2)}(A)$, is

$$N_{(2)}(A) = \frac{4D}{\pi A} \exp \left[\frac{-j2\zeta}{9(8 + 5\zeta^2)} \right]$$

The refined approximate system for limit cycle study is illustrated in Fig. 3.7-11. It is impractical to observe the effect of $N_{(2)}(A)$ on the loop graphically, since the maximum phase shift involved is only 1.1° (at $\zeta = \sqrt{2}$). The condition for loop oscillation can be

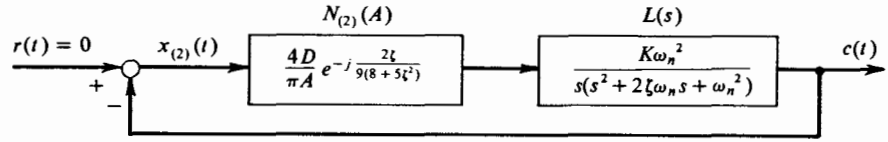


Figure 3.7-11 Third-order ideal-relay control system, refined DF approximation.

studied analytically, however. We require an open-loop phase shift of -180° , at the new limit cycle frequency ω_0 ; thus

$$-\frac{\pi}{2} - \tan^{-1} \left(\frac{2\zeta\omega_0\omega_n}{\omega_n^2 - \omega_0^2} \right) - \frac{2\zeta}{9(8+5\zeta^2)} = -\pi$$

or

$$\tan^{-1} \frac{2\zeta(\omega_0/\omega_n)}{1 - (\omega_0/\omega_n)^2} = \frac{\pi}{2} - \frac{2\zeta}{9(8+5\zeta^2)}$$

Continuing,

$$\begin{aligned} \frac{2\zeta(\omega_0/\omega_n)}{1 - (\omega_0/\omega_n)^2} &= \tan \left[\frac{\pi}{2} - \frac{2\zeta}{9(8+5\zeta^2)} \right] \\ &= \cot \frac{2\zeta}{9(8+5\zeta^2)} \\ &\approx \frac{9(8+5\zeta^2)}{2\zeta} \end{aligned}$$

Solving for ω_0 , we finally obtain

$$\omega_0 \approx \omega_n \left[1 - \frac{2\zeta^2}{9(8+5\zeta^2)} \right]$$

To a second approximation we find that the limit cycle frequency does indeed depend upon ζ . In the region of high ζ , the frequency correction term is largest. Such an outcome is reasonable, since it is precisely in this region that the limit cycle waveform harmonic content is greatest.

The first-order DF solution for ω_0 , as well as the second-order and exact solutions, are illustrated in Fig. 3.7-12.¹ The exact solution for the limit cycle frequency is the *implicit* relationship

$$\frac{\omega_n}{\omega_0} = \frac{2}{\pi\sqrt{1-\zeta^2}} \left[\frac{\sin \left(2 \tan^{-1} \frac{\sqrt{1-\zeta^2}}{\zeta} \right) \sinh \left(\frac{\pi\zeta\omega_n}{\omega_0} \right) - \cos \left(2 \tan^{-1} \frac{\sqrt{1-\zeta^2}}{\zeta} \right) \sin \left(\frac{\pi\sqrt{1-\zeta^2}\omega_n}{\omega_0} \right)}{\cosh \left(\frac{\pi\zeta\omega_n}{\omega_0} \right) + \cos \left(\frac{\pi\sqrt{1-\zeta^2}\omega_n}{\omega_0} \right)} \right]$$

which can be obtained by the method of piecewise solution of differential equations and boundary condition matching or Tsytkin's method (Sec. 3.8), among others. As can be

¹ Note the expanded ordinate scale in this figure.

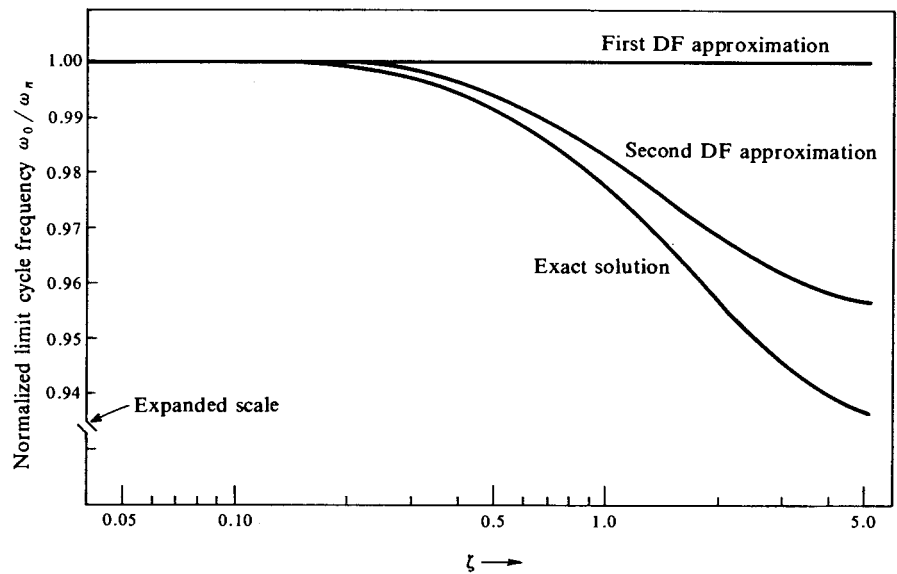


Figure 3.7-12 DF and exact limit cycle frequency solutions for example third-order relay system.

seen, it is a rather awkward expression, and is most reasonably dealt with by digital computer.

In the range $0 < \zeta < 1$, the first DF approximation has a maximum error of 2.3 percent, whereas the second DF approximation error is less than 0.6 percent. Over the semi-infinite range in ζ , the first DF approximation is in maximum error of 10.3 percent ($\zeta \rightarrow \infty$), whereas the second DF approximation has a maximum error of 4.9 percent ($\zeta \rightarrow \infty$). From these data it is clear that the second approximation does yield qualitative results missed by the first approximation, namely, the ζ dependence of ω_0 .

Note that in the region of high ζ , an amplitude-phase plot would show the $-1/N(A)$ and $L(j\omega)$ loci to be nearly parallel at intersection. It was pointed out earlier that under this circumstance the accuracy of DF results is expected to be degraded somewhat. The validity of this rule of thumb is redetermined consistently, as in the above example.

DF ACCURACY AND VALIDITY STUDIES

Up to this point we have attempted to both justify the validity of DF usage and assess DF accuracy by way of examples. To a certain extent this procedure has led us to develop various rules of thumb which permit exploitation of the DF as a device for nonlinear system study. However, as must certainly be the case, the rules of thumb are primarily stated *qualitatively*. There exist, on the other hand, several fairly general approaches to the accuracy and validity questions. These, we shall see, lead to *quantitative* statements concerning accuracy and validity.

The systems to which we are likely to apply DF methods are commonly quite complex. Therefore it will probably be no surprise to learn that the quantitative statements which we shall be able to make in literal terms will be cumbersome and of little value in actual analysis and design. Nevertheless, such statements, and the theories from which they are derived, do form a concrete mathematical basis for DF usage. It is for this reason alone that they are important in the context of our work and worthy of mention.

In what follows, because of space limitations, we present only the rudiments of several important accuracy and validity studies (along with source references). The interested reader is encouraged to follow up this presentation, especially since each of these studies is of some interest in its own right.

Bogoliubov and Mitropolsky (Ref. 8) extended the mathematical foundations upon which the earlier work of Krylov and Bogoliubov (Sec. 2.1) was based. One important result of their work was a recursive method for obtaining an asymptotic series solution of the autonomous equation (μ is a small parameter)

$$\ddot{x} + \omega_0^2 x = \mu f(x, \dot{x}) \quad (3.7-13)$$

The series solution is asymptotic in the sense that the error of the n th approximation is proportional to the $(n + 1)$ st power of the small parameter μ .

The form of the series solution is

$$x = A \cos \psi + \mu x^{(1)}(A, \psi) + \mu^2 x^{(2)}(A, \psi) + \cdots \quad (3.7-14)$$

where $\psi = \omega_0 t + \theta$, and $x^{(i)}$ are periodic functions of ψ , with period 2π . Further, A and ψ are required to satisfy the following differential equations:

$$\dot{A} = \mu A^{(1)}(A) + \mu^2 A^{(2)}(A) + \cdots \quad (3.7-15a)$$

$$\dot{\psi} = \omega_0 + \mu \Psi^{(1)}(A) + \mu^2 \Psi^{(2)}(A) + \cdots \quad (3.7-15b)$$

This formulation excludes the appearance of secular terms in the solution, which arise in the usual method of expansion in powers of a small parameter. Using Eqs. (3.7-14) and (3.7-15), one seeks a solution of Eq. (3.7-13). That is, one wants to find $x^{(i)}(A, \psi)$, $A^{(i)}(A)$, and $\Psi^{(i)}(A)$, which, simultaneously, lead to a solution of Eq. (3.7-13) to prescribed accuracy. The procedure to be followed is analogous to that used in the perturbation method (Sec. 2.1).

Bogoliubov and Mitropolsky (*op. cit.*) point out that only the first two or three terms of Eq. (3.7-14) can be computed in practice, because of the complexity of the formulas involved. Hence the practical applicability of the method is governed by the asymptotic properties of a few-term solution as $\mu \rightarrow 0$, rather than the actual convergence of Eq. (3.7-14). It is readily shown that the first series term in this recursive solution is the same as that originally developed by Krylov and Bogoliubov.

Among many examples considered by Bogoliubov and Mitropolsky is the simple pendulum, for which we have already obtained DF and exact solutions (Example 3.7-2). The DF solution, of course, is identical with the first-approximation solution of Bogoliubov and Mitropolsky. This follows since we have already indicated the equivalence of their solution to that of Krylov and Bogoliubov, and have earlier shown (Sec. 2.2) the relationship of the Krylov-Bogoliubov method to the DF method. For an oscillation amplitude of 160° we have seen that the first approximation to the oscillation frequency is in error by about 8 percent (Fig. 3.7-3). The second approximation, including the term $x^{(1)}(A, \psi)$, is reported (Ref. 8, p. 64) to be in error by only 3 percent. Considering the degree of nonlinearity present at such a large oscillation amplitude, we should conclude that these results are excellent and, in particular, that the second-order asymptotic solution can indeed represent a substantial improvement over the first-order solution. Unfortunately, we note that these results are applicable only to second-order systems.

Johnson (Ref. 48) studied the problem of DF accuracy directly. His work is based on earlier studies of Bulgakov (Ref. 12), and is addressed to the determination of the second term in a series for which DF analysis provides the first term. This additional term, once found, is used to correct DF results and to indicate DF solution accuracy. As we have already noted, general calculations of this sort are complicated and seldom of use in practice. Since this turns out to be the case here, we indicate Johnson's principal results without discussing his method in any detail.

Consider a single nonlinearity system to be in an oscillatory state. For a related hypothetical system define the input amplitude to the nonlinearity x and the oscillation frequency ω in terms of the following series:

$$x = x_0 + \mu x_1 + \frac{\mu^2}{2!} x_2 + \frac{\mu^3}{3!} x_3 + \cdots \quad (3.7-16a)$$

$$\omega = \omega_0 \left(1 + \mu \lambda_1 + \frac{\mu^2}{2!} \lambda_2 + \frac{\mu^3}{3!} \lambda_3 + \cdots \right) \quad (3.7-16b)$$

where μ is an artificially introduced parameter multiplying the nonlinearity, for which the value $\mu = 1$ reduces the hypothetical system to the original system. In Eqs. (3.7-16), called the *generating solution*, x_0 and ω_0 are the DF solutions to the original system; $x_i (i \geq 1)$ and $\lambda_i (i \geq 1)$ constitute the DF correction terms.

One of Johnson's main results was the determination that $\lambda_1 = 0$; i.e., the first frequency correction is zero. He also provided formulas for the first amplitude correction and the second frequency correction, although the accuracy estimates they provide depend upon the (unknown) radius of convergence of his power series. Bass (Ref. 3) states, appropriately, that

"Johnson's real accomplishment was to produce a new heuristic motivation for the [DF] method; one seeks for the linearized differential equation a linear equation possessing a 'generating solution' whose first frequency correction vanishes when it is perturbed in the 'direction' of the given nonlinear equation."

In his paper (*op. cit.*), Johnson studied a second-order system with friction-controlled backlash, for which the DF method results in limit cycle amplitude and frequency errors of 2.50 and 0.97 percent, respectively. Applying his second-frequency and first-amplitude corrections, he was able to reduce these errors to 0.28 and 0.23 percent, respectively. Similar order-of-magnitude reductions in DF errors were obtained in second-order ideal-relay- and ideal-velocity-limited servos. These results are outstanding, particularly in view of the fact that the linear elements' transfer function is only of second order. They not only tend to bear out the validity of Johnson's work but also provide further evidence that DF accuracy is on the order of 5 to 10 percent for a rather wide class of nonlinear systems.

Bass (Ref. 5) attempts to give a rigorous treatment of the mathematical validity of the describing function method by use of topological arguments. He starts with a physical system whose behavior is governed by a nonlinear vector differential equation of the form

$$\ddot{x} = f(x, \dot{x}) \quad \text{where } f(-x, -\dot{x}) = -f(x, \dot{x}) \quad (3.7-17)$$

Then, with the definition $\theta = \nu t$ ($\nu = 2\pi/T$, T being a period, if one exists; he is essentially normalizing the problem), Eq. (3.7-17) becomes

$$\nu^2 \frac{d^2x}{d\theta^2} = f\left(x, \nu \frac{dx}{d\theta}\right)$$

which is then broken into its linear and nonlinear parts to give the final form of the physical-system equation.

$$\nu^2 \frac{d^2x}{d\theta^2} + \nu C_1 \frac{dx}{d\theta} + C_2 x = \Phi\left(x, \nu \frac{dx}{d\theta}\right) \quad (3.7-18)$$

Proceeding along typical DF technique lines, he defines the vector $x_0(\theta)$, which is just the fundamental harmonic term in the Fourier series of $x(\theta)$, and obtains what he terms the "hypothetical tractable system" equation

$$\nu^2 \frac{d^2x}{d\theta^2} + \nu C_1 \frac{dx}{d\theta} + C_2 x = \Phi\left(x_0, \nu \frac{dx_0}{d\theta}\right) \quad (3.7-19)$$

The "describing function problem" is then to determine the circumstances under which, if Eq. (3.7-19) has precisely one periodic solution, the physical system of Eq. (3.7-18) will also have at least one periodic solution.

What Bass does show is that if $\Phi(x, \dot{x})$ is *smooth*, then under appropriate conditions a periodic solution to Eq. (3.7-19) implies a periodic solution to Eq. (3.7-18). If $\Phi(x, \dot{x})$ is piecewise-smooth, he is able only to show the conditions under which there are periodic solutions to Eq. (3.7-19), and not what this implies in respect to the physical system, Eq. (3.7-18). For $\Phi(x, \dot{x})$ piecewise-smooth, he shows that Eq. (3.7-19) has the appropriate periodic solution if and only if a $2n$ -dimensional (x is an n vector) system of implicit equations is satisfied. The necessity of this condition is derived from consideration of the Fourier coefficients of $x(\theta)$ quite easily, though the sufficiency involves more work.

It might be remarked that Bass does not expect an engineer to verify theoretically all the requirements he states for his results to hold; that would be a very large and difficult undertaking. He is attempting to put the technique on a mathematically justifiable basis, not to give methods for calculations. Unfortunately, the requirement that $\Phi(x, \dot{x})$ [and consequently $f(x, \dot{x})$] be smooth leaves out a large class of important nonlinear control problems; however, he does supply these rigorous results for a "smooth forcing function" system. An important result of this work is a rigorous verification of the *filter hypothesis* of conventional DF usage, at least for smooth systems.

Sandberg (Ref. 91) considers the operator equation

$$x = LN(x + r) \quad (3.7-20)$$

where x = a vector function,

L = a linear operator,

N = a nonlinear operator,

r = a periodic forcing function,

and the equivalent linearization approximation

$$x_0 = \tilde{L}N(x_0 + r) \quad (3.7-21)$$

in which $\tilde{L}$ extracts the fundamental component of the Fourier series expansion of $LN(x + r)$ and x_0 is a periodic solution. Sandberg's analysis is carried out in the space of periodic functions square integrable over a period. His method is to determine conditions that guarantee that an operator derived from LN is a contraction mapping in the whole space. He presents conditions under which there exists a unique periodic response to an arbitrary periodic input with the same period, as well as an upper bound on the mean-squared error in using equivalent linearization. He also gives conditions under which subharmonics and self-sustained oscillations cannot occur.

Holtzman (Ref. 45) treats essentially the same problem as Sandberg (*op. cit.*). His approach differs from Sandberg's in that Holtzman tries to obtain

a contraction mapping only in a neighborhood of x_0 , and thus avoids global Lipschitzian type of requirements which often limit the applicability of the contraction mapping theorem. However, his analysis is not applicable to piecewise-linear nonlinearities as a result of requiring differentiability. In his work there are applications to time-lag systems and subharmonic oscillators.

In a different study, Holtzman compares the jump resonance criteria given by the DF method and the *circle criterion*, the latter being a mathematically rigorous sufficient condition for the absence of jump phenomena in nonlinear feedback systems. The result of that study, stated briefly, is that the circle criterion does not contradict the DF criterion. This can be construed as providing another favorable piece of evidence relative to validity of the DF method.

V. M. Popov (Ref. 89) studied the stability of a class of closed-loop nonlinear systems by using direct analytical methods. He succeeded in establishing certain general results which can be interpreted in terms of the system linear elements' frequency response and the loosely defined shape of the nonlinear characteristic. It is of considerable interest to examine DF analysis in view of these results.

The class of systems considered by Popov contains a single static memoryless nonlinearity which is sufficiently smooth to ensure the existence of a unique solution of the governing differential equations, and a linear part with only lagging phase. In one special case his results can be interpreted as follows: For nonlinear characteristics lying entirely within the first and third quadrants (shaded area of Fig. 3.7-13a), the system is stable provided only that the linear-part phase shift is always lagging and takes values between 0 and -180° .

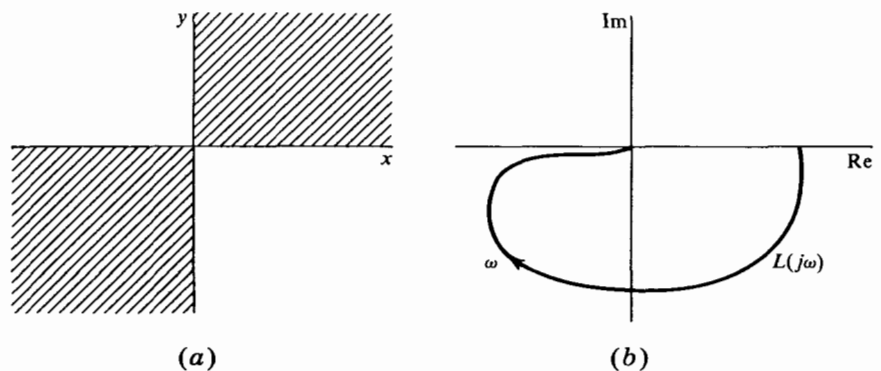


Figure 3.7-13 Region of allowable nonlinear characteristics (a) and typical linear-part polar plot (b) for stability, according to V. M. Popov's criterion.

Consider a DF analysis of this class of systems. As the nonlinear characteristic is single-valued and bounded by the lines $x = 0$ and $y = 0$, it follows (cf. Sec. 2.6) that $N(A)$ is real and of magnitude $0 \leq N(A) < \infty$. The polar plot of $-1/N(A)$ is thus the negative real axis. We immediately conclude that for all linear elements with phase lag never exceeding -180° the nonlinear systems are stable, that is, no limit cycle takes place. This is in complete agreement with Popov's result, although, according to DF analysis, no restriction on allowable phase *lead* is required (except that it must be less than 180°).

A generalization of Popov's criterion for lagging phase shift has been accomplished in the determination of stability conditions when the nonlinear characteristic lies in the sector between the lines $x = 0$ and $x = Ky$ (Ref. 90), and this work has been extended to certain cases, including leading phase shift, by further restricting the shape of the nonlinear characteristic (Ref. 115). Since there presently exists no neat polar-plane interpretations of these results, they are not further pursued here. One does observe from such work, however, the hint that DF analysis may be in error when applied to systems in which large phase lead occurs at low frequencies. Experimental results¹ appear to verify this point.

It is further worth noting that, because of their generality, results based upon extensions of Popov's criterion are always more conservative than the corresponding DF statements.

3.8 EXACT METHODS FOR RELAY CONTROL SYSTEMS

A control system in which the power amplifier is a relay device is called a relay (alternatively, on-off, contactor, bang-bang, bistable, etc.) control system. The relay amplifier is desirable because it can be simple, rugged, compact, and relatively cheap, while meeting high load-power requirements. Examples of symmetrical characteristics of four relay amplifiers are shown in Fig. 3.8-1, where δ_h and δ_d indicate hysteresis and dead-zone switching levels, respectively, and the associated drive levels are $\pm D$. The relay as a control-system component is enjoying an increasing popularity in both old and new control problems. Home temperature control systems, aircraft and missile adaptive control systems, and space-vehicle attitude control systems constitute typical applications. We have already seen the use of a relay for time-optimal control (Sec. 3.4). The range of variety in relay controller utilization is, in fact, enormous.

The actual design of relay control systems is a matter elsewhere discussed

¹ Brought to the authors' attention by R. W. Brockett, Massachusetts Institute of Technology. See Prob. 5-20.

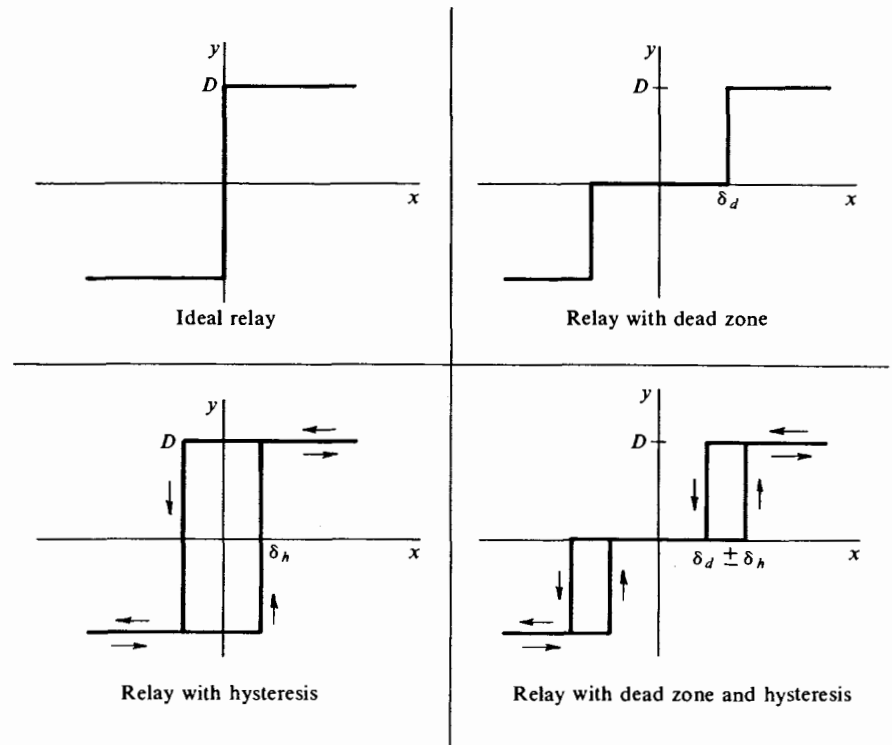


Figure 3.8-1 Common relay characteristics.

in this and other texts. Of importance here is the fact that relay control systems frequently either sustain limit cycles, are required to operate in the non-limit-cycle mode near a region of possible limit cycle behavior, or (less frequently) are subjected to near-harmonic forcing. Precise knowledge of steady oscillatory behavior is therefore desirable. It turns out that for systems incorporating relay-type devices, exact methods for steady-oscillation determination are available. Tsytkin's method is presented in some detail, because of its utility as a rapid check on DF results. Inasmuch as this method is of limited value in actual control-system design, it is treated as a DF validity test. As such, some discussion here is devoted to its analytical application, as opposed to its conventional graphical application. The emphasis in presentation is on application to low-order systems since we are perhaps at this time convinced of the reliability of DF results as applied to high-order systems. Although the development to follow deals only with symmetric two- or three-output-level devices, it is readily extended to asymmetric and other multiple-output-level devices.

PERIODICITY CONDITIONS

Consider a steady-state situation in which the closed-loop system of Fig. 3.8-2 sustains an oscillation. Sufficient conditions for the existence of this oscillation can be deduced, quite generally, from an observation of the periodic requirements on $x(t)$. For convenience we set a time origin at the instant of a positive switching (y becomes $+D$) and call T_0 the oscillation period ($T_0 = 2\pi/\omega_0$).

In the absence of any input $r(t)$, the oscillation will be symmetric about the origin for all symmetric relay characteristics. Thus conditions relating to periodicity need be written only over the half-cycle.

1. *Ideal relay.* The conditions sufficient for a periodic waveform are, in this case,¹

$$x\left(\frac{T_0^-}{2}\right) = 0 \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \quad (3.8-1)$$

since switching takes place at the zero crossing point for x . Equation (3.8-1), geometrically motivated, merely states that the value of x at the half-period equals the value of x at the period outset and that the slope is reversed at these instants. These conditions follow from the assumed "mirror-image" symmetry, familiar from Fourier series study.

2. *Hysteretic relay.* As in the case of the ideal relay, there are only two switchings per cycle. Sufficient conditions for periodicity are

$$x\left(\frac{T_0^-}{2}\right) = -\delta_h \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \quad (3.8-2)$$

which can be argued in a manner similar to that presented above.

3. *Relay with dead zone.* For this nonlinearity there are four switchings per limit cycle period, or two over the half-period. The first switching occurs some time before the half-period, denoted $\rho T_0/2$, with $0 < \rho < 1$. The second switching occurs at the half-period. Both sets of conditions, which must simultaneously be met for possible oscillatory behavior, are

$$x\left(\frac{\rho T_0^-}{2}\right) = \delta_d \quad \text{and} \quad \frac{dx}{dt}\left(\frac{\rho T_0^-}{2}\right) < 0 \quad (3.8-3a)$$

also

$$x\left(\frac{T_0^-}{2}\right) = -\delta_d \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \quad (3.8-3b)$$

¹ $\frac{T_0^-}{2} = \lim_{\epsilon \rightarrow 0} \left(\frac{T_0}{2} - \epsilon \right)$

4. *Hysteretic relay with dead zone.* Again there are two switchings per half-period. The sufficient conditions for oscillatory behavior are

$$x\left(\frac{\rho T_0^-}{2}\right) = \delta_a - \delta_h \quad \text{and} \quad \frac{dx}{dt}\left(\frac{\rho T_0^-}{2}\right) < 0 \quad (3.8-4a)$$

also

$$x\left(\frac{T_0^-}{2}\right) = -\delta_a - \delta_h \quad \text{and} \quad \frac{dx}{dt}\left(\frac{T_0^-}{2}\right) < 0 \quad (3.8-4b)$$

where, as before, $0 < \rho < 1$.

Having noted the geometric prerequisites for oscillation sustenance, it is now possible to establish conditions on the linear loop elements $L(s)$, so as to satisfy these prerequisites. One method of accomplishing this task, introduced by Tsytkin, is discussed next.

TSYTKIN'S METHOD

Assume the system of Fig. 3.8-2 to be in a steady oscillatory mode. For the sake of discussion, assume the relay to have drive levels $\pm D$ and switching levels $\pm \delta_h$. $L(s)$ is assumed rational in s , having all its poles in the left half-plane, with at least as many poles as zeros. When oscillating, the relay is assumed to flip every half-cycle. Then, for any oscillation waveform, the nonlinearity output $y(t)$ will be a square wave.¹

$$y(t) = Du(t) - 2Du(t - \tfrac{1}{2}T_0) + 2Du(t - T_0) - 2Du(t - \tfrac{3}{2}T_0) + \cdots \quad (3.8-5)$$

The Fourier series belonging to this periodic function is, for $t > 0$,

$$\begin{aligned} y(t) &= \frac{4D}{\pi} \left(\sin \omega_0 t + \frac{1}{3} \sin 3\omega_0 t + \cdots + \frac{1}{k} \sin k\omega_0 t + \cdots \right) \\ &= \frac{4D}{\pi} \sum_{k \text{ odd}} \frac{\sin k\omega_0 t}{k} \\ &= \frac{4D}{\pi} \sum_{k \text{ odd}} \frac{1}{k} \operatorname{Im} (e^{jk\omega_0 t}) \end{aligned} \quad (3.8-6)$$

Following Bergen (Ref. 7), we note that since $c(t)$ and $\dot{c}(t)$ are not necessarily continuous, it is convenient to define a function $L_1(s)$ by

$$L(s) = L_1(s) + L(\infty) \quad (3.8-7)$$

¹ $u(t - \tau)$ is the displaced unit-step function of value 1 for $t \geq \tau$, 0 for $t < \tau$.

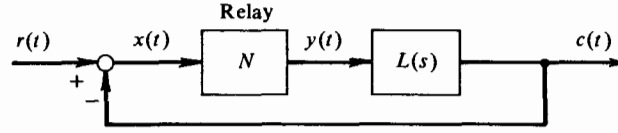


Figure 3.8-2 Closed-loop nonlinear system.

where $L_1(s)$ necessarily has more poles than zeros, and $L(\infty)$ is a constant. Since $L(s)$ has all its poles in the left half-plane, it follows from Eqs. (3.8-6) and (3.8-7) that

$$c(t) = \frac{4D}{\pi} \sum_{k \text{ odd}} \frac{1}{k} \text{Im} [L_1(jk\omega_0) e^{jk\omega_0 t}] + L(\infty) \dot{y}(t) \quad (3.8-8)$$

With $L_1(s)$ low-pass, the discontinuities in $c(t)$ appear only in the second term.

Limit cycles Since there is no loop input, $x(t) = -c(t)$, and the position periodicity condition [first part of Eq. (3.8-2)] at $t = T_0/2$ leads to

$$\sum_{k \text{ odd}} \frac{1}{k} \text{Im} [L_1(jk\omega_0)] = \frac{\pi}{4} \left[L(\infty) - \frac{\delta_h}{D} \right] \quad (3.8-9)$$

For the velocity periodicity condition we require $\dot{c}(t)$, which is derived by differentiating Eq. (3.8-8) with respect to time.

$$\dot{c}(t) = \frac{4D\omega_0}{\pi} \sum_{k \text{ odd}} \text{Re} [L_1(jk\omega_0) e^{jk\omega_0 t}] + L(\infty) \ddot{y}(t) \quad (3.8-10)$$

In evaluating $\dot{c}(t)$ care must be taken to account for possible discontinuities contributed by $L_1(s)$. [The delta function due to $\dot{y}(t)$ occurs after the relay action and may be discarded.] Discontinuities will appear if

$$\lim_{s \rightarrow \infty} sL_1(s)$$

is nonzero, and the magnitude of each discontinuity is

$$\dot{c}\left(\frac{T_0^+}{2}\right) - \dot{c}\left(\frac{T_0^-}{2}\right) = -2D \lim_{s \rightarrow \infty} sL_1(s) \quad (3.8-11)$$

The Fourier series of a piecewise-continuous function converges to the midpoint of each finite discontinuity; hence

$$\dot{c}\left(\frac{T_0}{2}\right) = -\frac{4D\omega_0}{\pi} \sum_{k \text{ odd}} \text{Re} [L_1(jk\omega_0)] + D \lim_{s \rightarrow \infty} sL_1(s) \quad (3.8-12)$$

Noting that $\dot{x}(t) = -\dot{c}(t)$ in the absence of an input, the velocity periodicity condition [second part of Eq. (3.8-2)] becomes

$$\sum_{k \text{ odd}}^{\infty} \operatorname{Re} [L_1(jk\omega_0)] < \frac{\pi}{4\omega_0} \lim_{s \rightarrow \infty} [sL_1(s)] \quad (3.8-13)$$

The periodicity equations derived above may be summarized by defining the so-called Tsyarkin locus, $T(j\omega)$, as

$$T(j\omega) = \sum_{k \text{ odd}}^{\infty} \operatorname{Re} [L_1(jk\omega)] + j \sum_{k \text{ odd}}^{\infty} \frac{1}{k} \operatorname{Im} [L_1(jk\omega)] \quad (3.8-14)$$

The Tsyarkin locus is a plot of $T(j\omega)$ as ω is varied over the range zero to infinity. We may restate the necessary conditions for a symmetric limit cycle oscillation at some frequency ω_0 in terms of the Tsyarkin locus, as follows:

$$\operatorname{Im} [T(j\omega_0)] = \frac{\pi}{4} \left[L(\infty) - \frac{\delta_h}{D} \right] \quad (3.8-15)$$

$$\operatorname{Re} [T(j\omega_0)] < \frac{\pi}{4\omega_0} \lim_{s \rightarrow \infty} [sL_1(s)]$$

In the case for which $L(s)$ has at least two more poles than zeros, Eqs. (3.8-15) reduce to the more common form of the Tsyarkin conditions

$$\begin{aligned} \operatorname{Im} [T(j\omega_0)] &= -\frac{\pi}{4} \frac{\delta_h}{D} \\ \operatorname{Re} [T(j\omega_0)] &< 0 \end{aligned} \quad (3.8-16)$$

Forced oscillations Let us assume an input of the form

$$r(t) = R \sin (\omega_r t + \varphi)$$

and a corresponding loop *forced* oscillation at the frequency ω_r . Then $x(t) = r(t) - c(t)$, and the position and velocity periodicity conditions, derived in a manner completely analogous to that presented above, are

$$\begin{aligned} -R \sin \varphi + \sum_{k \text{ odd}}^{\infty} \frac{1}{k} \operatorname{Im} [L_1(jk\omega_r)] &= \frac{\pi}{4} \left[L(\infty) - \frac{\delta_h}{D} \right] \\ -R \cos \varphi + \sum_{k \text{ odd}}^{\infty} \operatorname{Re} [L_1(jk\omega_r)] &< \frac{\pi}{4\omega_r} \lim_{s \rightarrow \infty} [sL_1(s)] \end{aligned} \quad (3.8-17)$$

These conditions can be rewritten in terms of the Tsyarkin locus $T(j\omega)$ as follows:

$$\begin{aligned} -R \sin \varphi + \operatorname{Im} [T(j\omega_r)] &= \frac{\pi}{4} \left[L(\infty) - \frac{\delta_h}{D} \right] \\ -R \cos \varphi + \operatorname{Re} [T(j\omega_r)] &< \frac{\pi}{4\omega_r} \lim_{s \rightarrow \infty} [sL_1(s)] \end{aligned} \quad (3.8-18)$$

Stability of periodic oscillations We have thus far been successful in determining whether a relay control system is *mathematically* capable of sustaining a periodic oscillation. The question now relates to whether in the presence of a small perturbation any given oscillating system will return to its original oscillatory state. The answer to this question will provide us with knowledge of the local asymptotic stability of an oscillatory mode in a *physical* relay system. Tsypkin (Ref. 106) describes a simple stability test based on the theory of sampled-data systems, the results of which we restate here:¹

A limit cycle mode in a relay control system is stable if the quantity $d\{\text{Im}[T(j\omega_0)]\}/d\omega_0$ is positive, and conversely, unstable if that quantity is negative.

The sign can be obtained either directly by inspection of the Tsypkin locus or analytically.

A forced oscillatory mode in a relay control system is stable if the operating point lies to the left of $T(j\omega_r)$, and conversely, unstable if it lies to the right.

Stability in this case is determined graphically.

GRAPHICAL APPLICATION OF TSYPKIN'S METHOD

In order to apply this method graphically, one must construct the Tsypkin locus $T(j\omega)$, defined by Eq. (3.8-14). This construction can proceed either from the amplitude-phase-plane plot of $L_1(j\omega)$ with the aid of an overlay (Ref. 80) or directly from the polar plot of $L_1(j\omega)$. The latter alternative is rather easily executed, making the polar plot a particularly convenient representation for this method (see Fig. 3.8-3).

As a result of the low-pass nature of $L_1(j\omega)$, the Tsypkin locus coincides with the polar plot of $L_1(j\omega)$ at high frequencies. This decreasing importance of higher harmonics is the basis of first-harmonic linearization, and leads to the generalization that *DF limit cycle prediction tends to be more reliable in the higher limit cycle frequency region*. Just what constitutes a *higher* frequency region of course depends upon the particular linear elements in question. We might also observe that although any linear elements' polar plot and its corresponding Tsypkin locus may tend toward the same trajectory in the polar plane, the *frequency calibration* for each will necessarily be different (Prob. 3-19).

Graphical construction for study of limit cycles follows from Eqs. (3.8-15). The study of forced oscillations is also readily accomplished graphically. Recalling that Eqs. (3.8-18) must be simultaneously satisfied in order for a

¹ The derivation may be found in Gille et al. (Ref. 30, Chap. 26).

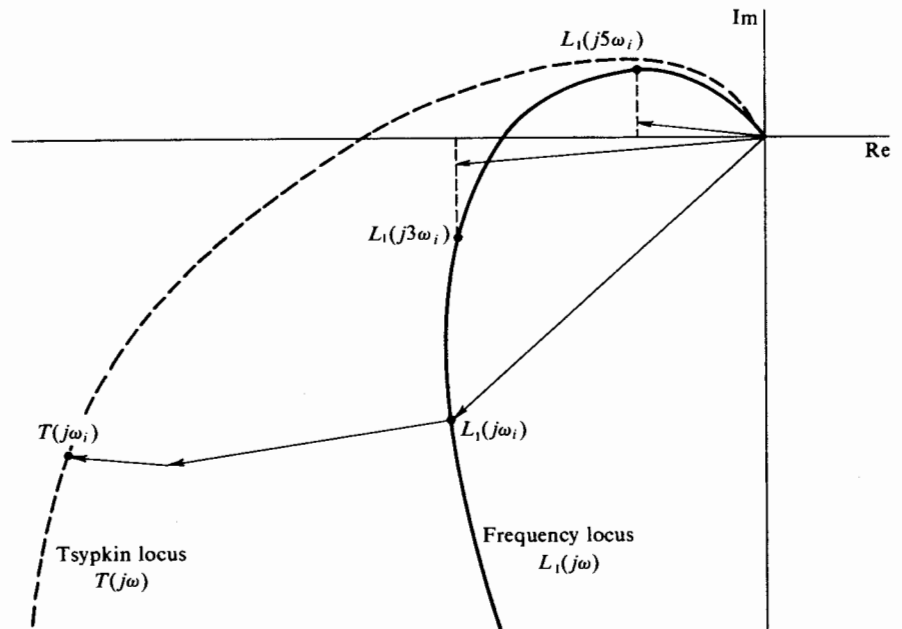


Figure 3.8-3 Obtaining the Tsypkin locus from the transfer locus $L_1(j\omega)$. $T(j\omega_i)$ is computed from $L_1(j\omega_i)$, $L_1(j3\omega_i)$, $L_1(j5\omega_i)$ according to Eq. (3.8-14). (Adapted from Gille, Pélegrin, and Decaulne, Ref. 30.)

forced oscillation to be possible, one is led to the construction of Fig. 3.8-4. By drawing a circle of radius R centered at the point on the Tsypkin locus given by $\omega = \omega_r$, all mathematical solutions are determined. In the illustration we see that two solutions of different phase appear possible. Based upon the stability test, however, we see that the solution labeled r_1 (to the left of ω_r) is the only stable solution, and that r_2 is in fact physically unrealizable. One may present some justification for the above argument at this point by noting that as $R \rightarrow \infty$, $\varphi_1 \rightarrow 0$, whereas $\varphi_2 \rightarrow \pi$. Certainly, we should expect the output to be in synchronism with the input in this limiting case, as the first solution indicates.

This graphical construction points out an interesting forced response pattern common to nonlinear systems, namely, *regional synchronous behavior*. As is evident from Fig. 3.8-4b, at a given input frequency ω_r for all input amplitudes below a certain value $R_{\min}$, no solution of Eqs. (3.8-18) is possible. The system actually sustains neither a limit cycle nor an oscillation at the frequency ω_r . Rather, a complex combination of the two occurs. For all values of R in the region $R > R_{\min}$, however, the system does indeed lock onto the input frequency, thus displaying regional synchronous behavior.

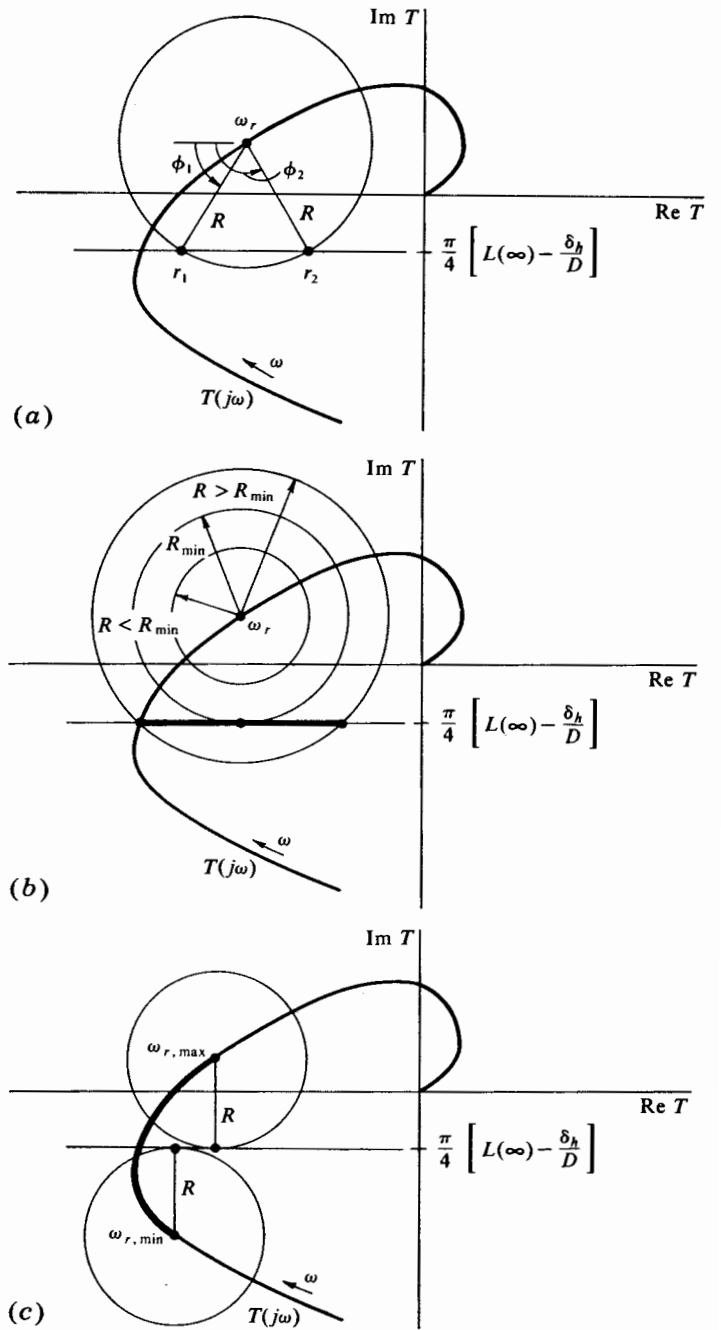


Figure 3.8-4 Graphical solution of the forced response equations (a). Construction of region of synchronism for fixed input frequency (b), for fixed input amplitude (c).

Viewed differently, for a fixed input amplitude R and varying input frequency ω_r , there will be some frequency region $\omega_{r,\max} > \omega_r > \omega_{r,\min}$ in which synchronous action can occur. Outside of this region, however, no such input-output synchronism is possible (Fig. 3.8-4c).

ANALYTICAL APPLICATION OF TSYPKIN'S METHOD

The problem we now face essentially reduces to the requirement for summing series of the form

$$\sum_{k=1,3,5,\dots}^{\infty} \operatorname{Re} [L_1(jk\omega_0)]$$

and

$$\sum_{k=1,3,5,\dots}^{\infty} \frac{1}{k} \operatorname{Im} [L_1(jk\omega_0)]$$

as encountered in Eq. (3.8-14), in terms of known functions. It happens that a summation procedure is available which solves this problem, having its basis in the theory of functions of a complex variable. By replacing a series of the form $\sum_{-\infty}^{\infty} f(n)$ with an appropriate contour integral, it is possible

TABLE 3.8-1 SUMMATION OF CERTAIN INFINITE SERIES

$L_1(s)$	$\sum_{k=1,3,5,\dots}^{\infty} \operatorname{Re} [L_1(jk\omega_0)]$	$\sum_{k=1,3,5,\dots}^{\infty} \frac{1}{k} \operatorname{Im} [L_1(jk\omega_0)]$
$\frac{1}{s+a}$	$\frac{\pi}{4\omega_0} \tanh \frac{\pi a}{2\omega_0}$	$-\frac{\pi}{4a} \tanh \frac{\pi a}{2\omega_0}$
$\frac{s+b}{(s+a)^2}$	$\frac{\pi}{8\omega_0 \cosh^2(\pi a/2\omega_0)} \times \left[\sinh \left(\frac{\pi a}{\omega_0} \right) - \frac{\pi}{\omega_0} (b-a) \right]$	$-\frac{\pi}{8a^2 \cosh^2(\pi a/2\omega_0)} \times \left[b \sinh \left(\frac{\pi a}{\omega_0} \right) - \frac{\pi a}{\omega_0} (b-a) \right]$
$\frac{(s+b_1)(s+b_2)}{(s+a)^3}$	$\frac{\pi}{16\omega_0 \cosh^2(\pi a/2\omega_0)} \times \left[2 \sinh \left(\frac{\pi a}{\omega_0} \right) - \frac{\pi^2}{\omega_0^2} (a-b_1) \right. \\ \left. \times (a-b_2) \tanh \left(\frac{\pi a}{2\omega_0} \right) + \frac{2\pi}{\omega_0} (2a-b_1-b_2) \right]$	$-\frac{\pi}{16a^3 \cosh^2(\pi a/2\omega_0)} \times \left[2b_1b_2 \sinh \left(\frac{\pi a}{\omega_0} \right) - \pi^2(a-b_1) \right. \\ \left. \times (a-b_2) \tanh \left(\frac{\pi a}{2\omega_0} \right) + 2\pi \frac{a}{\omega_0} (a^2 - b_1b_2) \right]$

Note: A useful relationship in applications of this table is

$$\frac{1}{(s+a)^2} = \lim_{b \rightarrow \infty} \left[\frac{1}{b} \frac{s+b}{(s+a)^2} \right]$$

to evaluate the series by application of the Cauchy residue formula. This manipulation, described elsewhere (Ref. 26), results in the formation of Table 3.8-1. In order to use this table, a given $L_1(s)$ is represented by its partial fraction expansion, and the terms are summed individually. Single-, double-, and triple-order poles can be accommodated by the three entries of the table.

Let us now consider the application of Tsytkin's method to a first-order relay control system which cannot be studied by the DF method (cf. Example 3.7-5).

Example 3.8-1 Investigate the possibility of limit cycle oscillations in the system of Fig. 3.8-5a. Consider time dimensioned in seconds.

For this system $L(\infty) = 0$; hence $L_1(s) = L(s)$. The first entry in Table 3.8-1 yields

$$\operatorname{Re} [T(j\omega_0)] = \frac{2\pi}{\omega_0} \tanh \frac{2\pi}{\omega_0}$$

and

$$\operatorname{Im} [T(j\omega_0)] = -\frac{\pi}{2} \tanh \frac{2\pi}{\omega_0}$$

From Eqs. (3.8-15) the conditions for sustained oscillation are

$$\frac{2\pi}{\omega_0} \tanh \frac{2\pi}{\omega_0} < \frac{2\pi}{\omega_0} \quad \text{or} \quad \tanh \frac{2\pi}{\omega_0} < 1$$

and

$$-\frac{\pi}{2} \tanh \frac{2\pi}{\omega_0} = -\frac{\pi}{5} \quad \text{or} \quad \tanh \frac{2\pi}{\omega_0} = 0.4$$

whence we observe that a limit cycle is indeed possible, and that its period is

$$T_0 = \frac{2\pi}{\omega_0} = \tanh^{-1} 0.4 = 0.424 \text{ sec}$$

The graphical construction is shown in Fig. 3.8-5b, from which we see that $d(\operatorname{Im} T)/d\omega > 0$ at $\omega = \omega_0$. Hence the limit cycle is stable.

Tsytkin's method can be applied to the study of systems displaying time-delay (transport-lag) phenomena. In such applications we deal with factors of the form e^{-sT} , which do not appear in Table 3.8-1. A possible and frequently instructive approach with which to circumvent this problem is to replace the time-delay factors by suitable Padé approximants (Ref. 104), which are given as ratios of polynomials in s . Then Table 3.8-1 can be employed, and system oscillations can be studied. In the interest of retaining an exact version of Tsytkin's method wherever possible, we digress to discuss an alternative summation procedure.

Following Guillemin (Ref. 37), one can generate the sums of Table 3.8-2 in closed form by an elegant sequence of elementary Fourier series manipulations. We shall demonstrate here that this table can be simply extended to

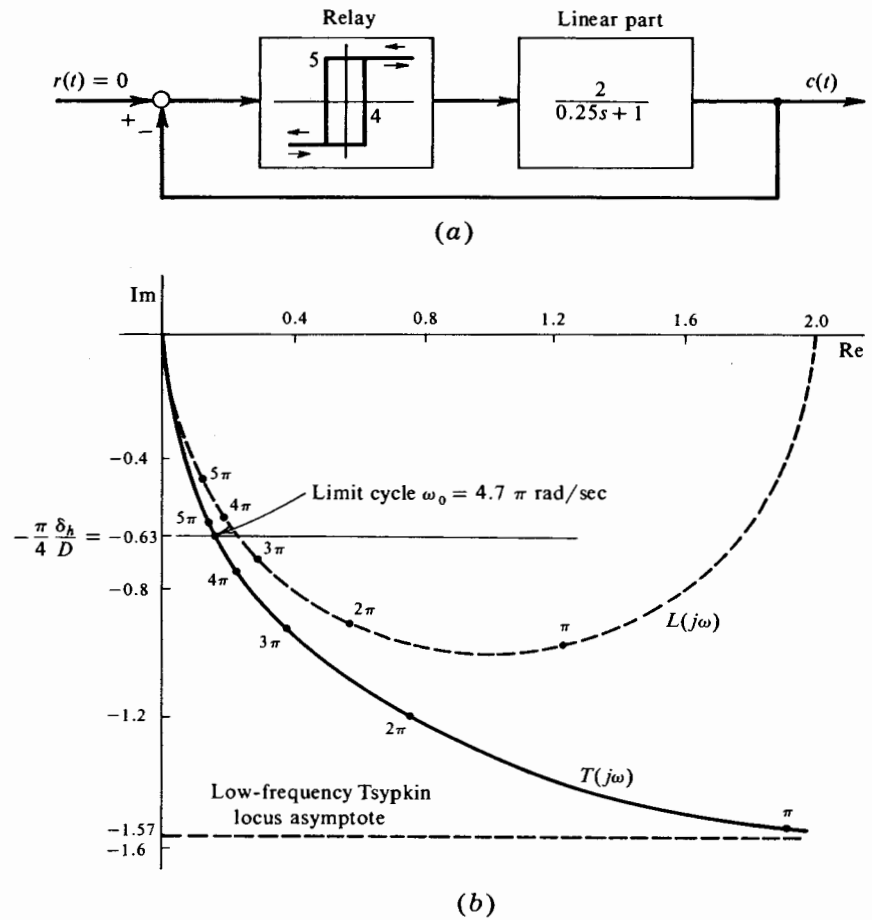


Figure 3.8-5 (a) Limit cycling system. (b) The graphical limit cycle solution.

include all cases of interest not listed. Observe that entry 3 can be derived from entry 2 merely by forming the function $f(x, a_1) - f(x, a_2)$ and performing the indicated algebra thereafter. Similarly, it is an algebraic procedure to extend these results to any order denominator comprised of terms $(k^2 + a_i^2)$. All series involving $\cos kx$ are derived from the related series involving $\sin kx$ by differentiation with respect to x . A summed series containing one second-order-denominator root $(k^2 + a_i^2)^2$ can be obtained by differentiation with respect to a_i of the same series with that root to the first power. Similarly, any-order-denominator root can be obtained by repeated differentiation. Finally, note that any order numerator can be

TABLE 3.8-2 SUMMATION OF CERTAIN INFINITE SERIES ($0 < x < \pi$)

$$\begin{aligned}
 (1) \quad & \sum_{k \text{ odd}}^{\infty} \frac{\sin kx}{k} = \frac{\pi}{4} \\
 (2) \quad & \sum_{k \text{ odd}}^{\infty} \frac{\sin kx}{k(k^2 + a^2)} = f(x, a) \\
 (3) \quad & \sum_{k \text{ odd}}^{\infty} \frac{\sin kx}{k(k^2 + a_1^2)(k^2 + a_2^2)} = \frac{f(x, a_1) - f(x, a_2)}{a_2^2 - a_1^2} \\
 (4) \quad & \sum_{k \text{ odd}}^{\infty} \frac{\cos kx}{k^2 + a^2} = f_z(x, a) \\
 (5) \quad & \sum_{k \text{ odd}}^{\infty} \frac{\cos kx}{(k^2 + a_1^2)(k^2 + a_2^2)} = \frac{f_z(x, a_1) - f_z(x, a_2)}{a_2^2 - a_1^2}
 \end{aligned}$$

where $f(x, a) = \frac{\pi}{4} \frac{\cosh(\pi a/2) - \cosh(\pi a/2 - ax)}{a^2 \cosh(\pi a/2)}$
 $f_z(x, a) = \frac{\partial f}{\partial x}(x, a) = \frac{\pi}{4} \frac{\sinh(\pi a/2 - ax)}{a \cosh(\pi a/2)}$

developed by forming terms such as

$$bf(x, a_1) + cf(x, a_2) \quad \text{or} \quad bf_z(x, a_1) + cf_z(x, a_2)$$

from entries 2 and 4 of the table, respectively.

Example 3.8-2 Determine the amplitude and frequency of the limit cycle displayed by the system of Fig. 3.8-6.

For this system, since $L(\infty) = 0$, we have

$$L_1(s) = \frac{Ke^{-s\tau}}{s}$$

The Tsytkin locus is computed according to

$$\operatorname{Re} [T(j\omega_0)] = -\frac{K}{\omega_0} \sum_{k=1,3,5,\dots}^{\infty} \frac{\sin k\omega_0\tau}{k} = -\frac{K}{\omega_0} \frac{\pi}{4}$$

$$\text{and} \quad \operatorname{Im} [T(j\omega_0)] = -\frac{K}{\omega_0} \sum_{k=1,3,5,\dots}^{\infty} \frac{\cos k\omega_0\tau}{k^2} = -\frac{K}{\omega_0} \frac{\pi}{4} \left(\frac{\pi}{2} - \omega_0\tau \right)$$

where we have used entries 1 and 4 of Table 3.8-2, the latter in the limit as $a \rightarrow 0$. The conditions for sustained oscillation are therefore

$$\begin{aligned}
 & -\frac{K}{\omega_0} \frac{\pi}{4} < 0 \\
 & -\frac{K}{\omega_0} \frac{\pi}{4} \left(\frac{\pi}{2} - \omega_0\tau \right) = -\frac{\pi}{4} \frac{\delta_h}{D}
 \end{aligned}$$

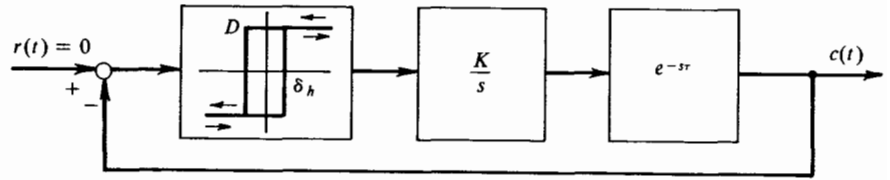


Figure 3.8-6 Simple relay control system.

the second of which yields, for $K > 0$,

$$T_0 = \frac{2\pi}{\omega_0} = 4 \left(\tau + \frac{\delta_h}{DK} \right)$$

By physical argument the symmetrical limit cycle waveform is established as triangular with zero average value. The peak amplitude A_0 is then the integrated nonlinearity output over one-quarter cycle.

$$A_0 = K \int_0^{T_0/4} D dt = \delta_h + DK\tau$$

METHODS OF HAMEL, BOHN, AND LUR'E

Additional methods for the exact solution for oscillations in relay control systems are now discussed briefly.

Hamel (Ref. 42) solves the problem of satisfying the required periodicity conditions by matching system trajectories in the phase plane. The existence of a periodic oscillation is then defined by an intersection in the phase plane of the so-called Hamel locus (the locus of points at which a switch must occur in order to sustain an oscillation) with the switch line $x = -\delta_h$ (for a hysteretic relay). It turns out that the Hamel and Tsytkin loci are fully equivalent in all systems with at least two more poles than zeros. In fact, one can construct the Hamel locus directly by an abscissa scale change and a coordinate rotation of the Tsytkin locus. Details of Hamel's method and the above-mentioned equivalence are available elsewhere (Ref. 30).

Bohn (Ref. 9) formulated the relay problem entirely in the time domain in order to overcome an apparent inadequacy in the Hamel and Tsytkin methods when applied to conservative systems. Although Tsytkin (Ref. 107) has subsequently pointed out that this inadequacy does, in fact, not exist, Bohn's formulation is nevertheless of interest. The result of this formulation is, for the hysteretic relay,

$$c(t) \Big|_{t=0} = -\delta_h \quad \text{and} \quad \frac{dc}{dt} \Big|_{t=T_0/2} > 0 \quad (3.8-19)$$

where

$$c(t) = \frac{D}{\pi j} \oint_c \frac{L(s)}{s} \frac{\exp(st)}{1 + \exp(sT_0/2)} ds \quad (3.8-20)$$

and the contour C encloses all left half-plane poles of $L(s)$, as well as the pole at the origin, but excludes all imaginary-axis pole pairs due to the zeros of $1 + \exp(sT_0/2)$. Proof of the validity of this formulation is left to the interested reader, who, upon closing the contour C in the right half-plane will also observe the re-creation of the usual Tsytkin conditions.

Lur'e (Ref. 63) treats the problem of steady oscillations directly in terms of the linear n th-order differential equation involved. By assuming a square-wave forcing of this differential equation, he formulates the nonlinear problem on a piecewise-linear basis. Then, by matching linear solutions at the switching instants, conditions for steady-state oscillations are determined. These conditions are presented in a compact canonic form in the reference cited, and are fully equivalent to the conditions derived by Bohn. Aizerman (Ref. 1) describes a method for the determination of self-oscillation modes in systems with an arbitrary piecewise-linear nonlinearity. This represents a degree of generalization beyond Lur'e's approach; however, in both instances the method of determining the periodic states reduces to a system of transcendental equations for which a general solution is not possible. These are in fact the same as the analytically summed Tsytkin conditions.

REFERENCES

1. Aizerman, M. A.: "Theory of Automatic Control," Pergamon Press, New York, 1963, chap. 5.
2. Andrievskii, V. R.: The Estimation of Self-oscillation Parameters in Nonlinear Automatic Control Systems, *Automation and Remote Control*, vol. 22 (September, 1961), pp. 146-150.
3. Bass, R. W.: Discussion of a paper by S. K. Chao, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (September, 1956), p. 231.
4. Bass, R. W.: Equivalent Linearization, Nonlinear Circuit Synthesis and the Stabilization and Optimization of Control Systems, *Proc. Symp. Nonlinear Circuit Analysis*, MRI Series, Polytechnic Institute of Brooklyn, New York, 1956, pp. 163-198.
5. Bass, R. W.: Mathematical Legitimacy of Equivalent Linearization by Describing Functions, *Proc. IFAC*, Moscow, 1960, Butterworth Scientific Publications, London, pp. 2074-2083.
6. Bautin, N.: "Povedenie dinamicheskikh sistem vblizi granits ustojchivosti," Gostekhizdat, Moscow, 1949.
7. Bergen, A. R.: A Note on Tsytkin's Locus, *IRE Trans. Autom. Control*, vol. AC-7, no. 3 (April, 1962), pp. 78-80.
8. Bogoliubov, N. N., and Y. A. Mitropolsky: "Asymptotic Methods in the Theory of Nonlinear Oscillations" (translated from Russian), Gordon and Breach, Science Publishers, Inc., New York, 1961.
9. Bohn, E. V.: Stability Margins and Steady-state Oscillations in On-Off Feedback Systems, *IRE Trans. Circuit Theory*, vol. CT-8, no. 2 (June, 1961), pp. 127-130.
10. Bonenn, Ze'ev: Stability of Forced Oscillations in Nonlinear Feedback Systems, *IRE Trans. Autom. Control*, vol. AC-6 (December, 1958), pp. 109-111.
11. Bonenn, Ze'ev: Connection between Various Methods of Investigating Absolute Stability, *IRE Trans. Autom. Control*, vol. AC-5, no. 4 (September, 1960), pp. 330-331.

12. Bulgakov, B. V.: Periodic Processes in Free Pseudo-linear Oscillatory Systems, *J. Franklin Inst.* (Philadelphia, Pa.), vol. 235 (June, 1943), pp. 591-616.
13. Cahen, G.: Perturbations des oscillateurs filtres, *Comptes Rend. Acad. Sci.*, vol. 235 (December, 1952), pp. 1614-1617.
14. Casserly, G., and J. G. Truxal: Measurement and Stabilization of Nonlinear Feedback Systems, *IRE Conv. Record*, pt. 4, 1956, pp. 52-60.
15. Chao, S. K.: Design of a Contactor Servo Using Describing Function Theory, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (September, 1956), pp. 223-233.
16. Chestnut, H.: Approximate Frequency-response Methods for Representing Saturation and Dead Band, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 269-287.
17. Choksy, N. H.: A New Analytic Stability Criterion for Use with Describing Functions, *JACC Proc.*, June, 1963, pp. 481-484.
18. Churchill, R. V.: "Complex Variables and Applications," 2d ed., McGraw-Hill Book Company, New York, 1960, chaps. 5 and 7.
19. Clauser, F. H.: A Theory of Nearly Linear Systems, *Proc. IFAC*, Moscow, Butterworth Scientific Publications, London, 1960, pp. 1557-1570.
20. Copson, E. T.: "Theory of Functions of a Complex Variable," Oxford University Press, New York, 1935.
21. Del Toro, V., and S. R. Parker: "Principles of Control System Engineering," McGraw-Hill Book Company, New York, 1960.
22. Dolgolenko, Yu. V.: The Approximate Derivation of Partial "Sliding" Periodic Modes in Relay Systems, *Automation and Remote Control*, vol. 18, no. 1 (April, 1958), pp. 1-27.
23. Emblar, J. N., and C. H. Weaver: A New Method for Compensating Contactor Servomechanisms, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 76 (September, 1957), pp. 165-170.
24. Freeman, E. A.: The Effect of Speed-dependent Friction and Backlash on the Stability of Automatic Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 78 (January, 1959), pp. 680-691.
25. Fukuma, A., and M. Matsubara: Jump Resonance Criteria of Nonlinear Systems, *IEEE Trans. Autom. Control*, vol. AC-11, no. 4 (October, 1966), pp. 699-706.
26. Gelb, A., and L. J. Henrikson: On the Analytic Application of Tsytkin's Method, *IRE Trans. Autom. Control*, vol. AC-8 (January, 1963), pp. 69-70.
27. Gelb, A.: Limit Cycles in Symmetric Multiple Nonlinearity Systems, *IRE Trans. Autom. Control*, vol. AC-8, no. 2 (April, 1963), pp. 177-178.
28. Gibson, J. E., and F. B. Tuteur: The Response of Relay Amplifiers with Feedback, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 76 (November, 1957), pp. 303-307.
29. Gibson, J. E.: "Nonlinear Automatic Control," McGraw-Hill Book Company, New York, p. 426.
30. Gille, J. C., M. J. Pélegrin, and P. Decaulne: "Feedback Control Systems," McGraw-Hill Book Company, New York, 1959, chap. 26.
31. Glatenok, I. V.: Evaluation of the Region of the True Periodic Solution Found Approximately by the Harmonic Balance Method, *Automation and Remote Control*, vol. 18, no. 11 (September, 1958), pp. 1175-1178.
32. Goldfarb, L. C.: On Some Nonlinear Phenomena in Regulatory Systems, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 239-259.
33. Grabbe, E. M., S. Ramo, and E. E. Wooldridge: "Handbook of Automation, Computation and Control," John Wiley & Sons, Inc., vol. 1, chap. 25.
34. Graham, D., and D. McRuer: "Analysis of Nonlinear Control Systems," John Wiley & Sons, Inc., New York, 1961.

35. Greif, H. D.: Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (September, 1953), pp. 243-248.
36. Grensted, P. E. W.: The Frequency-response Analysis of Nonlinear Systems, *Proc. IEE*, vol. C-102 (April, 1955), pp. 244-253.
37. Guillemin, E. A.: "The Mathematics of Circuit Analysis," John Wiley & Sons, Inc., New York, 1958, pp. 479-482.
38. Gurski, R. J.: Bang-bang Control of an Acceleration Switching Hydraulic Servovalve, 1961 *NEREM Record*, pp. 150-151.
39. Haas, V. B., Jr.: Coulomb Friction in Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (1953), pp. 119-123.
40. Hale, J. K.: "Oscillations in Nonlinear Systems," McGraw-Hill Book Company, New York, 1963.
41. Halstenberg, R. V.: Combined Hysteresis and Nonlinear Gains in Complex Control Systems, *IRE Trans. Autom. Control*, vol. AC-3 (July, 1958), pp. 51-58.
42. Hamel, B.: Contribution a l'étude mathématique des systèmes de réglage par tout-ou-rien, *CEMV*, no. 17, Service Technique Aeronautique, 1949.
43. Hatanaka, H.: The Frequency Responses and Jump-resonance Phenomena of Nonlinear Feedback Control Systems, *Trans. ASME, J. Basic Eng.*, vol. D-85 (June, 1963), pp. 236-242.
44. Hill, J. C.: Closed-loop Response of Nonlinear Systems by a Functional Transformation Approach, *IRE Trans. Autom. Control*, vol. AC-7, no. 4 (July, 1962), pp. 39-44.
45. Holtzman, J. M.: Contraction Maps and Equivalent Linearization, *Bell System Tech. J.*, vol. 46, no. 10 (December, 1967), pp. 2405-2435.
46. Hopkin, A. M., and K. Ogata: An Analytic Frequency-response Solution for a Higher-order Servomechanism with a Nonlinear Control Element, *Trans. ASME, J. Basic Eng.*, vol. D-81 (March, 1959), pp. 41-45.
47. Jaffe, R. C.: Limit Cycles in Single and Multiple Loop On-Off Systems, *M.I.T. Instrumentation Lab. Rept. T-358*, 1963.
48. Johnson, E. C.: Sinusoidal Analysis of Feedback-control Systems Containing Nonlinear Elements, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 71 (July, 1952), pp. 169-181.
49. Kislyakov, V. S.: Application of the Krylov-Bogoliubov Method of Constructing Asymptotic Approximations to the Investigation of Systems with Lag, *Automation and Remote Control*, vol. 21 (November, 1960), pp. 307-315.
50. Kislyakov, V. S.: Foundations for the Application of the Harmonic Linearization Method to an Investigation of Periodic Oscillations in Systems with Lag, *Automation and Remote Control*, vol. 21 (May, 1961), pp. 1051-1056.
51. Korolev, N. A.: Pulse Stabilization of Automatic Control Relay Systems, *Automation and Remote Control*, vol. 18, no. 5 (April, 1958), pp. 435-446.
52. Korolev, N. A.: Approximate Determination of the Auto-oscillation Parameters in Relay Systems with Lagged Feedback, *Automation and Remote Control*, vol. 20, no. 11 (June, 1960), pp. 1430-1434.
53. Kochenburger, R. J.: Limiting in Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (July, 1953), pp. 180-194.
54. Kochenburger, R. J.: A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, *Trans. AIEE*, pt. II, vol. 69 (1950), pp. 270-284.
55. Krasovskii, A. A.: On the Theory of Two-channel Follow-up Systems with a Relay Component in the AC Circuit, *Automation and Remote Control*, vol. 22 (April, 1961), pp. 911-919.
56. Landa, P. S., and S. P. Strelkov: On the Stability of an Aileron Control System Disturbed by Turbulence, *Automation and Remote Control*, vol. 22 (April, 1961), pp. 953-961.

57. Leonhard, A.: Extension of the Describing Function Method to the Investigation of Parametric Oscillations, *Proc. IFAC*, Basel, Switzerland, August, 1963, pp. 187/1-187/7.
58. Levinson, E.: Some Saturation Phenomena in Servomechanisms with Emphasis on the Tachometer Stabilized System, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 72 (March, 1953), pp. 1-9.
59. Levinson, E.: Gain-Phase Relations of Non-linear Circuits, *IRE Conv. Record*, pt. IV, 1958, pp. 141-159.
60. Levinson, E.: Non-linear Feedback Control Systems, Parts 3-5, *Electro-Technology*, vol. 72 (September, 1962), pp. 136-146; (October, 1962), pp. 131-140; (November, 1962), pp. 71-82.
61. Loeb, J. M.: Recent Advances in Non-linear Servo Theory, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 260-268.
62. Lozier, J. C.: A Steady-state Approach to the Theory of Saturable Servo Systems, *IRE Trans. Autom. Control*, vol. PGAC-1 (May, 1956), pp. 19-39.
63. Lur'e, A. I.: "Some Non-linear Problems in the Theory of Automatic Control," H.M. Stationery Office, London, 1957.
64. Magnus, K.: Approximate Criteria for Stability and Critical Point in Nonlinear Control Circuits, Metropolitan-Vickers translation 2257 of article in *Fachtagung Regelungstech., Contrib. 71*, Heidelberg, Germany, 1956, pp. 1-3.
65. Magnus, K.: On a Method of Investigating Nonlinear Systems of Oscillations and of Servomechanisms, NASA translation of original article in German (1955), OTS 59-16776, 1958.
66. Mahalanabis, A. K.: On Stabilization of Feedback Systems Affected by Hysteresis Nonlinearities, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 80 (November, 1961), pp. 277-285.
67. Manabe, S.: Analyze Nonlinear Controls with Saturation Functions, *Control Eng.*, vol. 8 (February, 1961), pp. 83-86.
68. McAllister, A. S.: A Graphical Method for Finding the Frequency Response of Nonlinear Closed-loop Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 80 (November, 1961), pp. 268-277.
69. Meyfarth, Philip F.: Bang-bang versus Linear Control of a Second-order Rate-type Servomotor, *Trans. ASME, J. Basic Eng.*, vol. D-82 (March, 1960), pp. 66-72.
70. Mikhail, S. L., and G. H. Fett: Frequency Response of Nonlinear Closed-loop Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 77 (November, 1958), pp. 436-438.
71. Minorsky, N.: "Nonlinear Oscillations," D. Van Nostrand Company, Inc., Princeton, N.J., 1962, chap. 14.
72. Mishkin, E., and L. Braun, Jr.: "Adaptive Control Systems," McGraw-Hill Book Company, New York, 1961, chap. 6.
73. Mishkin, E., and J. G. Truxal: Nonlinear Compensation Networks for Feedback Systems, *IRE Conv. Record*, pt. 4, 1957, pp. 3-7.
74. Mitropolsky, Y. A., and P. M. Senyk: Construction of an Asymptotic Solution for an Autonomic System with Strong Nonlinearity, OTS Publ. 62-15674, January, 1962.
75. Morosanov, I. S.: Methods of Extremum Control, *Automation and Remote Control*, vol. 18, no. 7 (June, 1958), pp. 1077-1092.
76. Morse, P. M., and H. Feshbach: "Methods of Theoretical Physics," McGraw-Hill Book Company, New York, 1953, pp. 413-414.
77. Moruzzi, R. L., and F. B. Tuteur: Nonlinear Servomechanism of Limited Dynamic Range, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 79 (November, 1960), pp. 314-320.
78. Mufti, I. H.: A Note on the Application of Harmonic Linearization, *IEEE Trans. Autom. Control*, vol. AC-8, no. 2 (April, 1963), pp. 175-177.

79. Naslin, P.: A Simplified Theory of Feedback Control Systems, Part II, *Process Control and Automation*, vol. 6 (May, 1959), pp. 206-209.
80. Naslin, P.: Summing Transmittances, *Process Control and Automation*, September, 1961, pp. 384-390.
81. Neimark, G.: O periodicheskikh rezhimakh i ustojchivosti relejnykh sistem, *Avtomatika i Telemekhanika*, vol. 14, no. 5 (1953), pp. 556-569.
82. Neimark, Yu. I., and L. P. Shil'nikov: On the Symmetric Periodic Motions of Multi-stage Relay Systems, *Automation and Remote Control*, vol. 20, no. 11 (June, 1960), pp. 1422-1429.
83. Nichols, N. B.: Backlash in a Velocity Lag Servomechanism, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 73 (January, 1954), pp. 462-467.
84. Numakura, T., T. Mivra, and A. Kamoi: Analyses of Servomechanisms by Describing Function Method on a Plane with Parameters as Coordinates, *AIEE Paper CP-57-1035*, 1957.
85. Ogata, K.: An Analytic Method for Finding the Closed-loop Frequency Response of Nonlinear Feedback Control Systems, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 76 (November, 1957), pp. 277-285.
86. Pestel, E.: Anwendung der Ljapunovschen Methode und des Verfahrens von Krylov-Bogoliubov auf ein technisches Beispiel, in "Nichtlineare Regelungsvorgänge," R. Oldenbourg KG, Munich, 1956.
87. Petelin, D. P.: Approximate Determination of the Auto-oscillations in the Automatic Control System for a Synchronous Motor, *Automation and Remote Control*, vol. 20 (July, 1959), pp. 13-19.
88. Popov, E. P.: "The Dynamics of Automatic Control Systems," Addison-Wesley Publishing Company, Reading, Mass., 1962.
89. Popov, V. M.: Absolute Stability of Nonlinear Systems of Automatic Control, *Automation and Remote Control*, vol. 22 (February, 1961), pp. 857-875.
90. Popov, V. M., and A. Halanay: On the Stability of Nonlinear Automatic Control Systems with Lagging Argument, *Automation and Remote Control*, vol. 23 (February, 1963), pp. 783-786.
91. Sandberg, I. W.: On the Response of Nonlinear Control Systems to Periodic Input Signals, *Bell System Tech. J.*, vol. 43 (May, 1964), pp. 911-926.
92. Shulkind, D.: Accuracy Requirements for Nonlinear Compensation of Backlash, *IRE Trans. Autom. Control*, vol. AC-5, no. 2 (June, 1960), pp. 79-84.
93. Silberberg, M. Y., and P. M. Schultheiss: Stability Criteria for Some Nonlinear Servomechanisms, *AIEE Paper DP 62-976*, 1962.
94. Slemon, G. R.: A Method of Approximate Steady-state Analysis for Nonlinear Networks, *J. IEE*, London, pt. I, vol. 100 (September, 1953), pp. 275-287.
95. Smith, O. J. M.: "Feedback Control Systems," McGraw-Hill Book Company, New York, 1958.
96. Sparacino, R. R.: A Synthesis Technique for Quasi-optimal Relay Control, Sc.D. thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Cambridge, Mass., September, 1961.
97. Starikova, M. V.: An Approximate Study of Self-oscillations in a Nonlinear Two-channel Automatic Control System, Wright-Patterson Air Force Base, Ohio, Foreign Technology Division, *Tech. Transl.* 62-833, June, 1962.
98. Stein, W. A., and G. J. Thaler: Obtaining the Frequency Response Characteristics of a Nonlinear Servomechanism from an Amplitude- and Frequency-sensitive Describing Function, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 77 (May, 1958), pp. 91-96.
99. Stoker, J. J.: "Non-linear Vibrations in Mechanical and Electrical Systems," Interscience Publishers, Inc., New York, 1950.

100. Syvi-Yan, T., and T. Lvi-Vy: Auto-oscillations in a Single-loop Automatic Control System Containing Two Symmetric Relays, *Automation and Remote Control*, vol. 20 (July, 1959), pp. 83-88.
101. Taft, V. A.: Stability of Periodic Conditions in Automatic Control Systems Found Approximately on the Basis of a Filter Hypothesis, *Automation and Remote Control*, vol. 19, no. 6 (March, 1959), pp. 550-555.
102. Thaler, G. J., and M. P. Pastel: "Analysis and Design of Nonlinear Feedback Control Systems," McGraw-Hill Book Company, New York, 1962.
103. Thomas, C. H.: Stability Characteristics of Closed-loop Systems with Dead-band, in R. Oldenburger (ed.), "Frequency Response," The Macmillan Company, New York, 1956, pp. 288-305.
104. Truxal, J. G.: "Automatic Feedback Control System Synthesis," McGraw-Hill Book Company, New York, 1955, pp. 548-550.
105. Tsypkin, Ya. Z.: Ob ustojchivosti periodicheskikh rezhimov v relejnykh sistemakh avtomaticheskovo regulirovaniya, *Avtomatika i Telemekhanika*, vol. 14, no. 5 (1953).
106. Tsypkin, Ya. Z.: "Teoriya relejnykh sistem avtomaticheskovo regulirovaniya," Gostekhizdat, Moscow, 1955, chaps. 8 and 9. Translation: "Sampling Systems Theory and Its Application," vols. 1 and 2, Pergamon Press and The Macmillan Company, New York, 1964.
107. Tsypkin, Ya. Z.: On the Determination of Steady-state Oscillations of On-Off Feedback Systems, *IRE Trans. Circuit Theory*, vol. CT-9, no. 3 (September, 1962).
108. Tustin, A.: The Effects of Backlash and of Speed-dependent Friction on the Stability of Closed-cycle Control Systems, *J. IEE*, London, pt. II, vol. 94 (May, 1947), pp. 143-151.
109. Urabe, M.: Galerkin's Procedure for Nonlinear Periodic Systems, *Arch. Rational Mech. Anal.*, vol. 20 (1965), pp. 120-152.
110. Van Horn, I. H.: Nonlinear Techniques Applied to the Analysis of Pilot Induced Oscillations, *IRE Conv. Record*, pt. 4, 1957, pp. 27-32.
111. Vavilov, A. A.: Investigation of Symmetric Self-oscillations in Relay Systems with Respect to Logarithmic Characteristics, *Automation and Remote Control*, vol. 24 (February, 1963), pp. 978-989.
112. West, J. C.: The Use of Frequency Response Analysis in Nonlinear Control Systems, in "Regelungstechnik, Moderne Theorien und ihre Verwendbarkeit," 1957 (in German).
113. West, J. C.: The Describing Function Technique Applied to Non-linear Systems, *Process Control and Automation*, vol. 7, pt. I (September, 1960), pp. 421-426; pt. II (October, 1960), pp. 469-473.
114. West, J. C., and P. N. Nikiforuk: The Frequency Response of a Servomechanism Designed for Optimum Transient Response, *Trans. AIEE*, pt. II, *Appl. Ind.*, vol. 75 (September, 1956), pp. 234-239.
115. Willems, J.: "On the Stability of Systems Containing a Single Nonlinearity," M.S. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, Cambridge, Mass., July, 1964.

PROBLEMS

- 3-1. A phase-shift oscillator has a cascaded RC-lag feedback network with transfer function

$$L(s) = \frac{K}{(\tau s + 1)^3}$$

The feedforward (amplifier) portion of the oscillator has a DF derived from consideration of both large and small oscillation amplitudes, given approximately by

$$N(A) \approx \begin{cases} \frac{5}{2+A} \angle -0.4A \text{ deg} & \text{for } A \leq 3 \\ \frac{3}{A} \angle -\frac{3.6}{A} \text{ deg} & \text{for } A > 3 \end{cases}$$

Plot the limit cycle amplitude and frequency as functions of τ (for $K = 4$) and as functions of K (for $\tau = 1$). What account need be taken of the amplifier phase contribution to the closed loop?

- 3-2. Using both polar and amplitude-phase plots, determine and study the stability of all limit cycles predicted by DF application to an ideal relay servo with linear elements given by

$$L(s) = \frac{Ke^{-sT}}{s}$$

Compute the exact amplitude and frequency of the single stable limit cycle, and thus show that the DF approximation leads to a perfect answer for both the waveform frequency and its fundamental amplitude component.

- 3-3. Investigate the existence and stability of limit cycle oscillations present in a servo with a limiter (abrupt saturation with drive levels $\pm D$ and input breakpoints $\pm \delta$) and linear elements

$$L(s) = \frac{K(\tau s + 1)^3}{s^3}$$

Use both analytical and graphical methods.

- 3-4. Investigate the possible appearance of steady-state oscillations in the equation

$$\ddot{\theta}(t) + a\dot{\theta}(t) + b\dot{\theta}(t - \tau) + \omega_0^2\theta(t) + g\theta^3(t) = 0$$

which arises in the study of a mechanism for minimizing the roll of a ship (W. J. Cunningham, "Introduction to Nonlinear Analysis," McGraw-Hill Book Company, New York, 1958, p. 333).

- 3-5. In a certain pneumatic control system the dominant system characteristics are known to be a nonlinear square-root-law downstream valve [$y(x) = K\sqrt{x}$] and three cascaded isolated accumulators. Each of the accumulators is describable by a first-order lag of unity gain and time constant $\tau = 5$. What value of valve gain K do you recommend for a fast-responding but limit-cycle-free system?
- 3-6. In a closed-loop system, an amplifier, a motor, and a gear train are used in the forward loop, and a simple angular sensor mounted on the load is used to provide the feedback signal. The amplifier has an adjustable gain K . The motor has a $\dot{\theta}$ /volt performance function $= 100/(1 + 2s)$ deg/sec/volt. The gear has a 10:1 reduction and a backlash of 5° at the input end. A pure viscous load is driven by the output shaft of the gear. The sensor has a sensitivity of 0.1 volt/deg. Study the limit cycle characteristic of this system as a function of K .
- 3-7. Find that value of gain K for which the closed-loop response of the system of Example 3.3-1 is down 3 db at a frequency of 10 radians/min. What closed-loop phase shift occurs at this frequency?

If the *nonlinear* integrator were replaced by an equivalent *linear* integrator, what would be the resultant closed-loop gain and phase shift at $\omega = 10$ radians/min?

- 3-8. Discuss the frequency response of the simple servomechanism of Fig. 3-1 when N is
- A linear gain
 - A limiter (abrupt saturation)
 - A cubing nonlinearity

Can jump resonance occur in any of the cases above? If not, devise a nonlinearity which will lead to jump resonance and discuss the resulting frequency response characteristics.

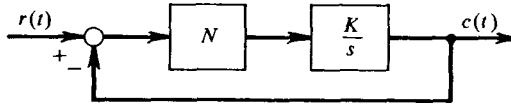


Figure 3-1 Simple servomechanism for frequency response study.

- 3-9. Sketch the closed-loop frequency response curves arising for the system of text Example 3.3-2 for the case in which the nonlinearity is a relay with drive levels $\pm D$ and dead zone 2δ . Consider the possibilities of both $M_r < \delta$ and $M_r > \delta$. What feedback compensation network can be employed to eliminate all jump resonances?
- 3-10. Demonstrate by manipulating the variables of Fig. 3-2 that the steady-state equation describing frequency response can be written as

$$\frac{M_r}{A} e^{-j\theta} - N(A) = L^{-1}(j\omega)$$

(Hint: Use of the notation $r = M_r e^{j\omega t}$, $x = A e^{j(\omega t + \theta)}$ may prove expedient.) (Aizerman, *op. cit.*)

Describe a graphical procedure for determining A and θ as functions of ω , and illustrate a situation which depicts jump resonance possibilities.

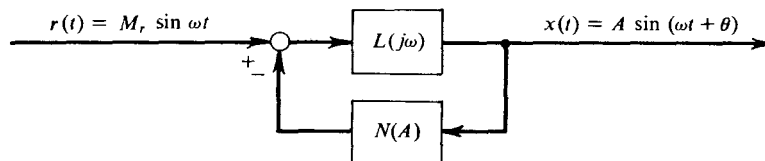


Figure 3-2 Nonlinear system for general frequency response study.

- 3-11. Using the critical frequency locus show that the system of Fig. 3-3 just fails to exhibit jump resonance phenomena if the following relationship holds:

$$\frac{\sqrt{ab}}{b-a} = \frac{1}{2\sqrt{3}}$$

[Hint: Work with the function $\varphi = \tan^{-1}(V/U)$, where $L = U + jV$.]

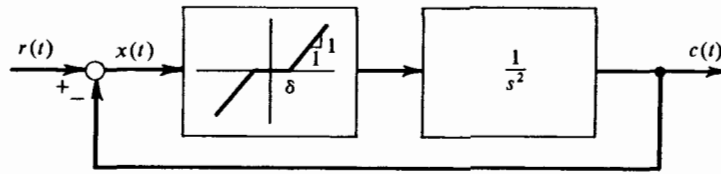


Figure 3-5 Conservative second-order system with dead zone.

3-17. For the following form of Rayleigh's equation,

$$\ddot{x} - \lambda(1 - \dot{x}^2)\dot{x} + x = 0$$

find the solution for limit cycle oscillations by the DF first approximation. By noting that the residual consists of a single term, construct the second-approximation DF solution, and thus show that the limit cycle frequency correct to second-order terms in λ is

$$\omega_0 \approx 1 + \frac{3\lambda^2 A^2}{64}$$

3-18. Using the DF method, compute first- and second-order approximation solutions for the limit cycle exhibited by the relay control system of Fig. 3-6.

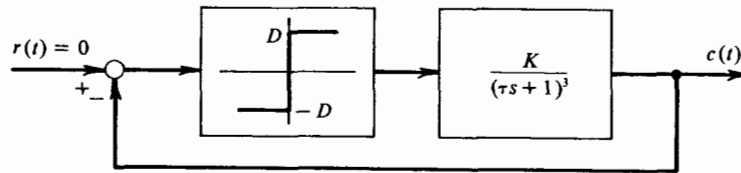


Figure 3-6 Relay control system.

- 3-19. Show that for all first- and second-order functions $L_1(s)$ the frequency calibration of the (coincident) linear elements' polar plot and corresponding Tsytkin locus in the high-frequency region tends toward a constant difference factor. [Hint: Consider the limit in increasing frequency of the function $|L_1(j\omega)/T(j\omega)|$.]
- 3-20. Discuss the nature of the limit cycle predicted by DF theory for the second-order ideal-relay system with $L(s) = K\omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$, and solve for the limit cycle exactly by use of Tsytkin's method. Repeat for $L(s) = K/(s + 1)^2$.
- 3-21. A satellite attitude control system with on-off thrust control and proportional plus rate feedback is shown in Fig. 3-7. Determine the limit cycle period for this system, and indicate probable accuracy of the result by tracing the relay-output third harmonic around the loop and comparing to the first-harmonic amplitude at the relay input.

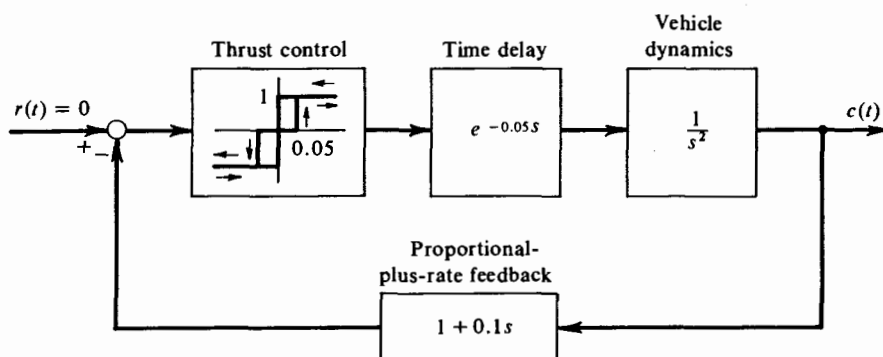


Figure 3-7 Satellite attitude control system.

3-22. Design the system of Fig. 3-8 to meet the specifications:

- No limit cycle in the absence of input
- Dead zone referred to the error $e(t) \leq 1$ unit

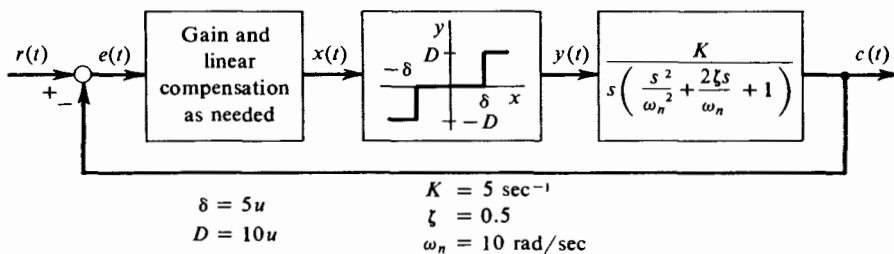


Figure 3-8

 3-23. Show that the range of values of K for which the system of Fig. 3-9 will *not* limit-cycle is given by

$$0 \leq K < \frac{\pi \zeta \omega_n \delta}{D}$$

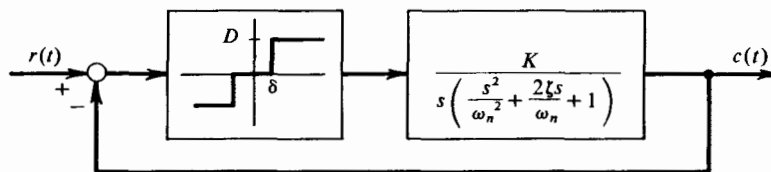
 If K is twice the minimum value which results in a limit cycle, show that the resulting limit cycle amplitude and frequency are $A \approx 3.9\delta$, $\omega_0 = \omega_n$.


Figure 3-9

- 3-24. The relay-controlled system of Fig. 3-10 must meet the following specifications:
- (a) It must have enough drive capability to follow input rates up to 50 units/sec.
 - (b) The limit cycle amplitude at $c(t)$ is to be no greater than 2 units.
- Design appropriate compensation. You may place it either just before or just after the switch.

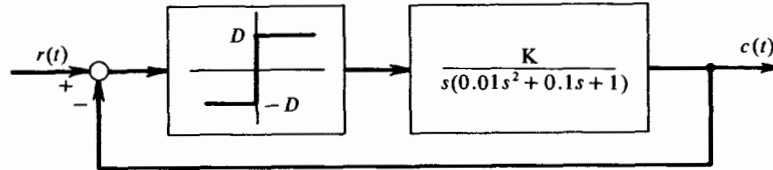


Figure 3-10

- 3-25. Sketch the input-output frequency response curves for amplitude and phase for the system of Fig. 3-11 with an input sinusoid of amplitude 15 units. You need not calculate the entire curve accurately. Just get some of the interesting points.

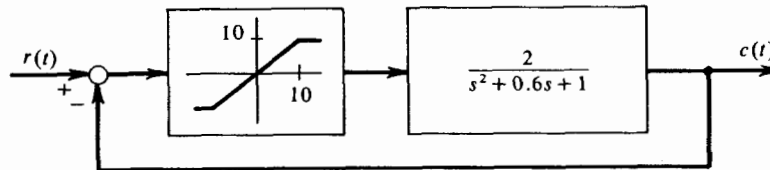


Figure 3-11

- 3-26. Would you expect single-sinusoid-input describing function theory to be useful in the applications of Fig. 3-12? If not, why not?

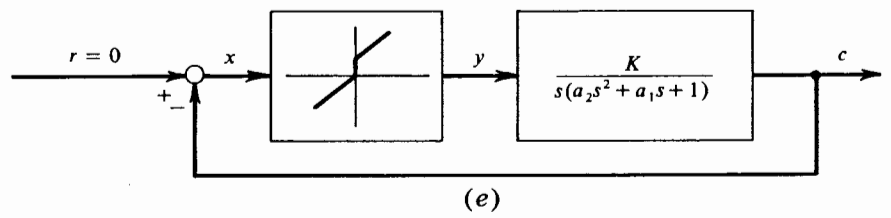
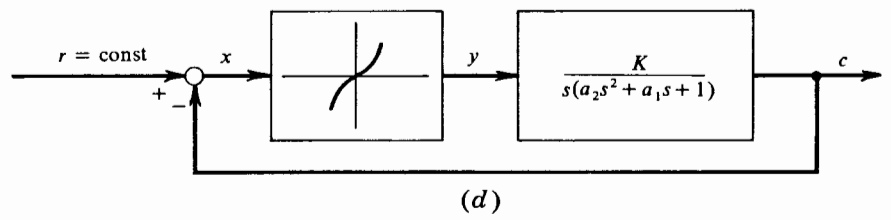
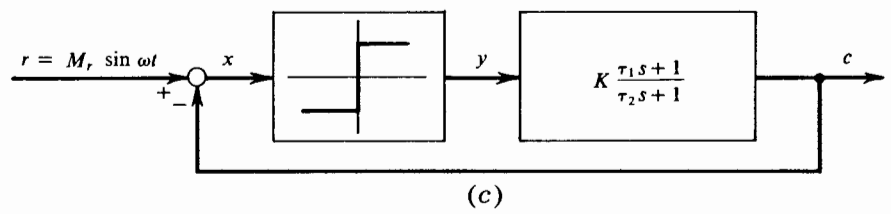
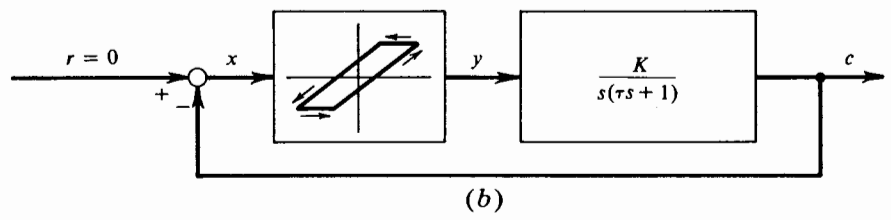
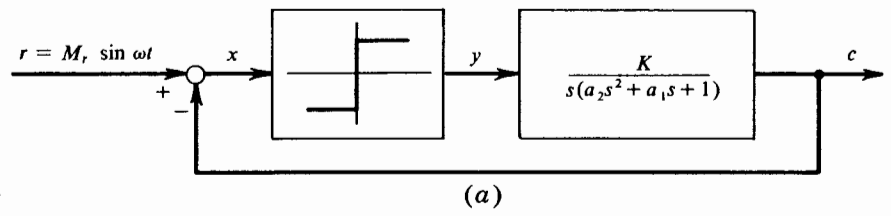


Figure 3-12

4 TRANSIENT OSCILLATIONS IN NONLINEAR SYSTEMS

4.0 INTRODUCTION

We turn now from consideration of steady-state oscillations to the more complicated study of transient oscillations in nonlinear systems. The term *transient oscillation* is intended to imply a function of time which is oscillatory in nature and which can be described as a sinusoid with slowly changing amplitude and frequency. Any function could be described as a sinusoid with changing amplitude and frequency if fast enough, and perhaps discontinuous, changes were permitted. For the purpose of this chapter, however, the restriction to slowly changing amplitude and frequency is necessary to permit the use of the DF to characterize the performance of the nonlinear part of the system.

The analytic description of transient oscillations in nonlinear systems is a matter of very real practical importance. A common illustration of this is a feedback servo with integral compensation to permit an increase in low-frequency gain. Such a system frequently has a substantial chain of amplification in the forward path, amplifying the error signal to get the desired gain.

If one of these amplifiers in the chain saturates at a small value of system error, the resulting loss of effective system gain results in a decrease in relative stability due to the phase lag at low frequencies introduced by the integral compensation. The consequence of this in high-performance servos, such as those which position the gimbals of an inertial guidance stable platform, is a violent oscillation when the servo is first turned on and the error is not within the linear "notch." This oscillation is damped, and the servo eventually settles into linear operation within the notch, but the nonlinear transient oscillation is an important characteristic of the servo, and an analytic description of this characteristic is of practical importance to the designer.

Another illustration is the design of a feedback loop around a limit cycling system which regulates the amplitude of the limit cycle at some station. The design of the amplitude-regulating loop can be pursued in a rational manner only if one has a description of the dynamics relating a change in a system parameter, such as a forward gain, to the resulting change in the amplitude of the limit cycle. A transfer function which represents this dynamic effect will be derived in the following sections as a special case of the general study of transient oscillations in nonlinear systems.

4.1 ANALYTIC DESCRIPTION OF TRANSIENT OSCILLATIONS

Consider the oscillatory performance of nonlinear systems which may be cast in the form of a single loop system with separable linear and nonlinear parts, as shown in Fig. 4.1-1. This is the same form as that treated in the preceding chapter on steady-state oscillations as shown in Fig. 3.3-2. The further restriction that $r = 0$ must be made to permit practical solution of this problem. This does not rule out consideration of all forced responses, because many cases of common interest, such as a step-function response, can be given an equivalent description in terms of zero input with appropriate initial conditions on system variables. It is the response due to initial conditions which is calculated here. In the case of steady-state oscillations it is possible without undue labor to consider an input of the same form as that of the system output, namely, a steady-state oscillation. In this case, the inclusion of an input in the form of a transient oscillation is much more laborious, and would seem to be of little practical consequence.

The linear parts of the system of Fig. 4.1-1 are time-invariant operators, and are considered to be given originally in terms of their transfer functions. The nonlinear part is characterized in this analysis by its DF. It is this approximation which limits the changing amplitude and frequency of the transient oscillation to slow changes; we shall later return to the question of

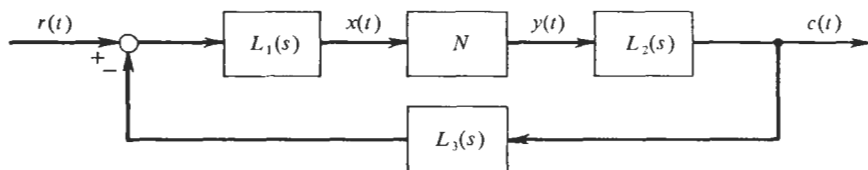


Figure 4.1-1 Form of system considered in study of transient oscillations.

what is meant by “slow.” The use of the DF also requires the conditions on the nonlinearity and the rest of the system stated in Sec. 3.0.

DIFFERENTIAL EQUATION FOR THE NONLINEARITY INPUT

Of the variables identified in Fig. 4.1-1, the only one that can conveniently be solved for initially is x because the DF for the nonlinearity part may depend on the instantaneous amplitude and frequency of the transient oscillation at x . Having found $x(t)$, it will be possible to solve for the variables at other stations, such as $c(t)$, but this will not be a simple matter. It might be noted at the outset that *in dealing with transient oscillations, not only the amplitude but the instantaneous frequency as well differ at different points around the loop.* Toward the solution for $x(t)$, we first write down the differential equation which it must satisfy. Considering the transfer functions to be rational functions of the Laplace transform variable s , we have

$$\begin{aligned} X(s) &= -L_1(s)L_2(s)L_3(s)Y(s) \\ &= -\frac{b_0 + b_1s + \cdots + b_ms^m}{a_0 + a_1s + \cdots + a_ns^n} Y(s) \end{aligned}$$

$$\text{or} \quad (a_ns^n + \cdots + a_0)X(s) = -(b_0 + \cdots + b_ms^m)Y(s) \quad (4.1-1)$$

This implies the differential equation

$$a_n \frac{d^n x(t)}{dt^n} + \cdots + a_0 x(t) = -\left[b_0 y(t) + \cdots + b_m \frac{d^m y(t)}{dt^m} \right] \quad (4.1-2)$$

If the solution to this differential equation is in the form of an oscillation with slowly changing amplitude and frequency, the nonlinear relationship between $x(t)$ and $y(t)$ can profitably be approximated as the DF for the nonlinear operation $y = y(x, \dot{x})$. This approximation is employed throughout this chapter, and for notational simplicity, the relation between $x(t)$ and $y(t)$, using the DF, is written as an equality.

$$y(t) = N(t)x(t) \quad (4.1-3)$$

Here $N(t)$ is the DF for the nonlinear part of the system. It is indicated as a function of time because it is explicitly a function of $A(t)$ and $\omega(t)$, the instantaneous amplitude and frequency of $x(t)$, both of which are functions of time.

$$a_n \frac{d^n x(t)}{dt^n} + \cdots + a_0 x(t) = - \left[b_0 N(t) x(t) + \cdots + b_m \frac{d^m N(t) x(t)}{dt^m} \right] \quad (4.1-4)$$

We now look for the solution to this equation in the form

$$x(t) = A(t) \exp [j\psi(t)] \quad (4.1-5)$$

subject to specified initial conditions $A(0)$, $\psi(0)$. Since the assumed form of the solution involves only these two constants of integration, only two initial conditions on x and its derivatives can be satisfied; $A(0)$ and $\psi(0)$ should be chosen to give $x(0)$ and $\dot{x}(0)$ correctly. The interpretation of the complex exponential form for $x(t)$ is, as always, that the physical variable $x(t)$ is either the real or imaginary part of this complex function.

To put this solution form into Eq. (4.1-4), we require the derivatives of $x(t)$ and of $N(t)x(t)$. For the terms on the left-hand side,

$$\begin{aligned} x &= A \exp(j\psi) \\ \frac{dx}{dt} &= \dot{A} \exp(j\psi) + j\dot{\psi} A \exp(j\psi) \\ &= \left(\frac{\dot{A}}{A} + j\dot{\psi} \right) A \exp(j\psi) \end{aligned} \quad (4.1-6)$$

By analogy with the standard description of exponentially decaying (or diverging) sinusoids, we define the relative rate of change of amplitude to be σ , an instantaneous exponential decay factor, and the instantaneous rate of change of phase angle to be ω .

$$\frac{\dot{A}}{A} = \sigma \quad (4.1-7)$$

$$\dot{\psi} = \omega \quad (4.1-8)$$

In this case, though, both σ and ω vary with time. Continuing the analogy further, we define a variable s as

$$s = \sigma + j\omega \quad (4.1-9)$$

In this case s has no interpretation in terms of an integral transform; it is simply a convenient variable, defined by the sequence of equations given above. In terms of this variable, Eq. (4.1-6) becomes simply

$$\frac{dx}{dt} = sx \quad (4.1-10)$$

Unfortunately, this pattern does not continue quite so simply to the higher derivatives.

$$\begin{aligned}
 \frac{d^2x}{dt^2} &= s\dot{x} + \dot{s}x \\
 &= s^2x + \dot{s}x \\
 &= (s^2 + \dot{s})x
 \end{aligned} \tag{4.1-11}$$

Similarly,

$$\begin{aligned}
 \frac{d^3x}{dt^3} &= s^2\dot{x} + 2s\dot{s}x + \dot{s}\dot{x} + \ddot{s}x \\
 &= (s^3 + 3s\dot{s} + \ddot{s})x
 \end{aligned} \tag{4.1-12}$$

Higher derivatives can be calculated as necessary. In every case, the result is expressible as x multiplied by a function of s and its derivatives.

For the terms in the right-hand member of Eq. (4.1-4),

$$\begin{aligned}
 \frac{d(Nx)}{dt} &= N\dot{x} + \dot{N}x \\
 &= (Ns + \dot{N})x
 \end{aligned} \tag{4.1-13}$$

It may be well to repeat this calculation in longer form to make clear the significance of the variables. Using the polar form for the DF,

$$\begin{aligned}
 N(t) &= \rho_N(t) \exp [j\theta_N(t)] \\
 Nx &= \rho_N A \exp [j(\psi + \theta_N)] \\
 \frac{d(Nx)}{dt} &= \dot{\rho}_N A \exp [j(\psi + \theta_N)] + \rho_N \dot{A} \exp [j(\psi + \theta_N)] \\
 &\quad + j(\dot{\psi} + \dot{\theta}_N) \rho_N A \exp [j(\psi + \theta_N)] \\
 &= \left[\left(\frac{\dot{A}}{A} + j\dot{\psi} \right) \rho_N \exp (j\theta_N) + \dot{\rho}_N \exp (j\theta_N) \right. \\
 &\quad \left. + j\dot{\theta}_N \rho_N \exp (j\theta_N) \right] A \exp (j\psi) \\
 &= (Ns + \dot{N})x
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 \frac{d^2(Nx)}{dt^2} &= Ns\dot{x} + N\dot{s}x + \dot{N}s x + \dot{N}\dot{x} + \ddot{N}x \\
 &= [N(s^2 + \dot{s}) + 2\dot{N}s + \ddot{N}]x
 \end{aligned} \tag{4.1-14}$$

and higher derivatives can be calculated as required.

Now these expressions for $x(t)$ and its derivatives and $N(t)x(t)$ and its derivatives can be put into Eq. (4.1-4), and the factorable $x(t)$ divided out of every term; that factor represents the trivial solution to the equation. The result is an equation among s and its derivatives and N and its derivatives. Since s and N are functions of A , $\dot{A}$, and $\omega = \dot{\psi}$, the complex equation, if written out in terms of its real and imaginary parts, becomes two nonlinear differential equations in A , ψ , and their derivatives. The solution to these equations defines $A(t)$ and $\psi(t)$, and thus the system variable $x(t)$, through Eq. (4.1-5). These differential equations, being nonlinear, generally do not permit an explicit solution. However, the purpose in deriving differential equations for A and ψ using an assumed solution form was to take advantage of the slowly changing nature of the parameters A and ω to effect useful approximate solutions.

REPLACEMENT RULES FOR TRANSIENT OSCILLATIONS

We shall return to the solution of the system equation, but let us first review the procedure involved in setting up that equation. Having once calculated the derivatives of the important variables, it is possible to go directly from the transfer function for the linear part of the system to the differential equation for A and ψ . From the block diagram of the system and the open-loop transfer function, one can immediately write an equation of the form of Eq. (4.1-1).

$$(a_n s^n + \cdots + a_1 s^1 + a_0 s^0)X(s) = -(b_0 s^0 + b_1 s^1 + \cdots + b_m s^m)Y(s) \quad (4.1-15)$$

The equation defining the transient oscillation is then derived from this by a few steps which can be formalized into simple rules:

1. Cross out $X(s)$ on the left, and $Y(s)$ on the right.
2. On the left-hand side, replace

s^0 by 1

s^1 by s

s^2 by $s^2 + \dot{s}$

s^3 by $s^3 + 3s\dot{s} + \ddot{s}$

s^4 by $s^4 + 6s^2\dot{s} + 3\dot{s}^2 + 4s\ddot{s} + \ddot{\dot{s}}$

s^5 by $s^5 + 10s^3\dot{s} + 15s\dot{s}^2 + 10s\ddot{s} + 10s^2\ddot{\dot{s}} + 5s\ddot{\dot{\dot{s}}} + \ddot{\dot{\dot{\dot{s}}}}$

.....

3. On the right-hand side, replace

s^0 by N
 s^1 by $Ns + \dot{N}$
 s^2 by $N(s^2 + \dot{s}) + 2\dot{N}s + \ddot{N}$
 s^3 by $N(s^3 + 3s\dot{s} + \ddot{s}) + 3\dot{N}(s^2 + \dot{s}) + 3\ddot{N}s + \ddot{\ddot{N}}$
.....

Replacement rules 2 and 3 can be expressed in compact form in terms of recursion formulas.¹ Let rule 2 read: “On the left-hand side, replace s^n by f_n ,” and rule 3: “On the right-hand side, replace s^n by g_n .” These replacement functions are given by

$f_{n+1} = \sum_{k=0}^n \binom{n}{k} (p^{n-k}s)f_k$ $f_0 = 1$ (4.1-16)

$g_{n+1} = \sum_{k=0}^n \binom{n}{k} [(p^{n-k}s)g_k + (p^{n+1-k}N)f_k]$ $g_0 = N$ (4.1-17)

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

and p^n represents d^n/dt^n .

The result of the application of these rules is a differential equation involving complex terms. One then writes two equations involving only real terms by equating real and imaginary parts of the original equation. This involves the expansion of terms like

$s^2 = \sigma^2 - \omega^2 + j2\sigma\omega$
 $s^3 = \sigma^3 - 3\sigma\omega^2 + j(3\sigma^2\omega - \omega^3)$

and the writing of derivatives like

$\dot{s} = \dot{\sigma} + j\dot{\omega}$
 $\ddot{s} = \ddot{\sigma} + j\ddot{\omega}$

The derivatives of N are more complicated. Using the real and imaginary form for $N(A,\omega)$, we have

$$N(A,\omega) = n_p(A,\omega) + jn_q(A,\omega)$$

$$\dot{N} = \frac{\partial n_p}{\partial A} \dot{A} + \frac{\partial n_p}{\partial \omega} \dot{\omega} + j \frac{\partial n_q}{\partial A} \dot{A} + j \frac{\partial n_q}{\partial \omega} \dot{\omega}$$

$$= A \frac{\partial n_p}{\partial A} \sigma + \frac{\partial n_p}{\partial \omega} \dot{\omega} + j \left(A \frac{\partial n_q}{\partial A} \sigma + \frac{\partial n_q}{\partial \omega} \dot{\omega} \right)$$

¹ Pointed out by B. C. Sherman.

The higher derivatives of N involve higher-ordered partial derivatives of n_p and n_q with respect to A and ω , as well as higher time derivatives of σ and ω . We note that in the quasi-static situation in which derivatives of s and N are ignored, the substitution rules reduce to identities; i.e., application of the rules reproduces the original system transfer function.

Example 4.1-1 As an example of the application of these rules, consider the system of Fig. 4.1-2. This simple system has an ideal relay driving a second-order plant in a closed-loop configuration. For this system, the equation corresponding to Eq. (4.1-15) is

$$(s^2 + bs)X(s) = -KY(s) \quad (4.1-18)$$

Application of the rules 1 to 3 gives

$$s^2 + \dot{s} + bs = -KN \quad (4.1-19)$$

as the equation defining the transient oscillation. Real and imaginary parts of this equation are

$$\text{Real:} \quad \sigma^2 - \omega^2 + \dot{\sigma} + b\sigma = -Kn_p \quad (4.1-20a)$$

$$\text{Imaginary:} \quad 2\sigma\omega + \dot{\omega} + b\omega = -Kn_q \quad (4.1-20b)$$

Using the DF for the ideal relay, these become

$$\text{Real:} \quad \sigma^2 - \omega^2 + \dot{\sigma} + b\sigma + \frac{4KD}{\pi A} = 0 \quad (4.1-21a)$$

$$\text{Imaginary:} \quad 2\sigma\omega + \dot{\omega} + b\omega = 0 \quad (4.1-21b)$$

These equations will be solved in Sec. 4.3.

SOLUTION FOR OTHER VARIABLES

In most cases, the solution for $x(t)$ serves adequately to describe the dynamic character of an oscillatory system. In some situations, however, one might wish to determine the response at some other station, such as $c(t)$ as shown in Fig. 4.1-1. From that figure one can write two relations which determine $c(t)$ in terms of $x(t)$.

$$C(s) = L_2(s)NX(s) \quad (4.1-22)$$

$$C(s) = -[L_1(s)L_3(s)]^{-1}X(s) \quad (4.1-23)$$

Each of these implies a differential equation of the form

$$g_n \frac{d^n c(t)}{dt^n} + \cdots + g_0 c(t) = f(t) \quad (4.1-24)$$

where the right-hand member is a known function of time composed of a linear combination of $N(t)x(t)$ and its derivatives if Eq. (4.1-22) is used, or a linear combination of $x(t)$ and its derivatives if Eq. (4.1-23) is used. In

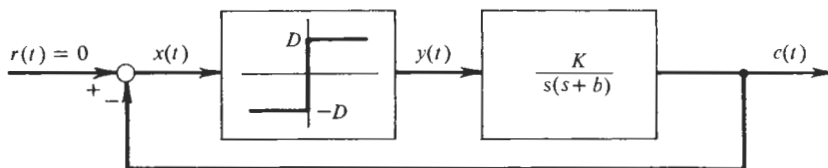


Figure 4.1-2 System of Example 4.1-1.

either case, $x(t)$, and consequently $N(t)$, are determined, together with their derivatives, by the solutions to the equations corresponding to Eq. (4.1-21); so $f(t)$ in Eq. (4.1-24) is determined. The choice between use of Eq. (4.1-22) or (4.1-23) is purely one of convenience, and must be decided in every individual case. For example, if $L_1(s)$ and $L_3(s)$ have no zeros, Eq. (4.1-24) takes the particularly simple form

$$c(t) = \frac{1}{g_0} f(t) \quad (4.1-25)$$

which can be handled readily. In more complicated cases this is still no more than the problem of calculating the output of a linear filter due to a specified time function input. Both exact and approximate methods of treating this problem are discussed in most books on control theory. (See, for example, Ref. 9, secs. 1.6, 6.6.)

The historical development of the analytical approach presented in this section is discussed in the next; then, in Sec. 4.3, we consider the solution of the equations defining a transient oscillation.

4.2 RELATION TO OTHER WORK

The solution of differential equations by means of fitting a sinusoid with slowly varying amplitude and phase angle is associated with a number of people, but foremost among them are Krylov and Bogoliubov, as the result of their extensive work in the period 1930 to 1940. Much of their attention was directed toward the discovery of stationary solutions corresponding to steady-state oscillations, and the stability of these solutions, but their approach was by way of the more general transient oscillation.

This work was cited as a beginning point in Chap. 2. The pertinent results presented there are that, for the differential equation

$$\ddot{x} + \omega_0^2 x + \mu f(x, \dot{x}) = 0 \quad (4.2-1)$$

Krylov and Bogoliubov find as their solution of first approximation

$$x(t) = A(t) \sin [\omega_0 t + \theta(t)] \quad (4.2-2)$$

where $A(t)$ and $\theta(t)$ satisfy the differential relations

$$\frac{dA}{dt} = -\frac{\mu}{2\pi\omega_0} \int_0^{2\pi} f(A \sin \psi, A\omega_0 \cos \psi) \cos \psi d\psi \quad (4.2-3)$$

and

$$\frac{d\theta}{dt} = \frac{\mu}{2\pi A\omega_0} \int_0^{2\pi} f(A \sin \psi, A\omega_0 \cos \psi) \sin \psi d\psi \quad (4.2-4)$$

The total rate of change of phase, ω , is

$$\omega = \omega_0 + \frac{\mu}{2\pi A\omega_0} \int_0^{2\pi} f(A \sin \psi, A\omega_0 \cos \psi) \sin \psi d\psi \quad (4.2-5)$$

In the notation of the preceding section, these results are

$$\sigma = \frac{\dot{A}}{A} = -\frac{n_q(A, \omega_0)}{2\omega_0} \quad (4.2-6)$$

$$\omega = \omega_0 + \frac{n_p(A, \omega_0)}{2\omega_0} \quad (4.2-7)$$

Application of the rules of the preceding section to Eq. (4.2-1) gives

$$s^2 + \dot{s} + \omega_0^2 + N(A, \omega) = 0 \quad (4.2-8)$$

which has real and imaginary parts

$$\text{Real:} \quad \sigma^2 - \omega^2 + \dot{\sigma} + \omega_0^2 + n_p(A, \omega) = 0 \quad (4.2-9a)$$

$$\text{Imaginary:} \quad 2\sigma\omega + \dot{\omega} + n_q(A, \omega) = 0 \quad (4.2-9b)$$

The Krylov and Bogoliubov solution may be recognized as an approximate solution to these equations. In the equation of imaginaries, drop the $\dot{\omega}$ term and approximate ω by ω_0 . This gives the solution for σ [Eq. (4.2-6)]. In the equation of reals, drop the σ^2 and $\dot{\sigma}$ terms and approximate ω by ω_0 in the evaluation of n_p . Then the neglect of n_p/ω_0^2 raised to powers greater than 1 gives the solution for ω [Eq. (4.2-7)].

This would seem to be a rather crude solution to Eqs. (4.2-9); in particular, it will be shown in examples that the omission of $\dot{\omega}$ in the equation of imaginaries is often serious. Moreover, we should like to include, in our consideration, systems for which $\omega_0 = 0$. In that case there is no evident approximation to be used for ω in evaluating n_p and n_q , and n_p/ω_0^2 cannot be considered a small quantity. We also wish to deal with systems of arbitrary order. Krylov and Bogoliubov had to consider a second-order equation of such a form that it permitted an asymptotic solution involving expansions in ascending powers of the small parameter μ to retain a degree of mathematical rigor. Under some conditions, for example, they are able to calculate bounds on the error in the solution due to the harmonic linearization of the nonlinearity.

The point of view taken in the preceding section was quite different. We began with a system of arbitrary order and, without stated motivation, employed the DF to characterize the nonlinearity, making no assumption about smallness of the nonlinear effect. This approach is actually motivated by the fact that it is the only practicable means of attacking an important class of problems; it is to some degree justified by the excellent results it yields in test problems. It is of some comfort to note, however, that after approaching the study of oscillatory nonlinear systems from a rather different point of view, Krylov and Bogoliubov were led to characterize the nonlinearity by the use of "harmonic linearization," which in the first approximation is exactly the describing function.

More recently, other writers have discussed the subject of oscillatory transients in nonlinear systems. Grensted (Ref. 5) and Voronov (Ref. 10), the latter following Popov (Ref. 7), give essentially the same argument as that of the preceding section. Neither of them, however, recognized the convenience of the variable $s = \dot{A}/A + j\dot{\omega}$ in writing out the derivatives of the oscillatory function. Clauser (Ref. 1) gives derivative formulas in the same form as those derived here, but he speaks of the result as an operational calculus for nonlinear oscillations. In this chapter we simply employ s as a convenient working variable and make no interpretation of it as a differential operator.

A number of writers have taken a quasi-static approach to the study of transient oscillations, among them Gelb and Vander Velde (Refs. 3 and 4), Lubbock and Barker (Ref. 6), and several others whose results are summarized by Thaler and Pastel (Ref. 8, sec. 4.9). In this approach one solves the standard characteristic equation of the system, using the DF to represent the nonlinearity, to determine the normal modes corresponding to prescribed values of A and ω . If the oscillatory mode, which is the one of interest, has a damping factor different from zero, this specifies the instantaneous relative rate of change of amplitude. The determination of this rate of change of amplitude as a function of the amplitude and corresponding frequency permits solution for the amplitude and frequency transients. If this solution is further approximated by considering the performance of the nonlinearity to be essentially fixed over each half-cycle of the oscillation, the transient is pieced together from half-cycles of sinusoids having a positive or negative damping factor which corresponds to the amplitude and frequency of the oscillation at the beginning of each half-cycle. This last procedure is appropriately referred to as *piecewise-sinusoidal approximation*.

There are many situations in which the quasi-static solution has a useful degree of accuracy. This is exactly the solution one obtains from the equations developed by the rules of the preceding section if all derivatives of σ and ω are dropped. Having the complete equations, including the derivative terms, one could first obtain the quasi-static solution, and from it

calculate the σ and ω derivatives to check their magnitudes. It will be found in the examples of the following sections that quasi-static solutions may be appreciably in error; usually, in these cases, it is the $\dot{\omega}$ contribution which is important.

4.3 SOLUTION OF THE EQUATIONS DEFINING THE OSCILLATION

We now address the question of how to solve the nonlinear differential equations for A and ω derived in Sec. 4.1. Only in rare cases would one expect to find an exact solution; so our purpose is to find useful approximate solutions. The approximations to be made are indicated by the fact that we have formulated the problem in terms of variables A and ω , which we expect to be slowly varying. Specifically, if the DF is to approximate the performance of the nonlinearity,¹ the input to the nonlinearity, $x(t)$, must approximate a sinusoid. That is, the relative changes in amplitude and frequency of $x(t)$ over one period of the oscillation must be small.

$$\frac{|\Delta A \text{ in one period}|}{A} = \frac{|A|}{A} \frac{2\pi}{\omega} = 2\pi \frac{|\sigma|}{\omega} \ll 1 \quad (4.3-1)$$

$$\frac{|\Delta \omega \text{ in one period}|}{\omega} = \frac{|\dot{\omega}|}{\omega} \frac{2\pi}{\omega} = 2\pi \frac{|\dot{\omega}|}{\omega^2} \ll 1 \quad (4.3-2)$$

With respect to ω (or ω^2), then, we should expect σ and $\dot{\omega}$ to be small, and any higher derivatives or higher powers of these quantities to be still smaller.

With this order for the importance of variables in mind, recall that the Krylov and Bogoliubov solution could be derived from our equations for A and ω by solving the equation of reals for ω and the equation of imaginaries for σ , after having dropped certain terms. Additional motivation for calculating ω from the equation of reals and σ from the equation of imaginaries is provided by considering the describing function for the nonlinear part as the sum of a proportional plus a derivative gain. The real part is the proportional gain, which acts as a "spring constant" in the oscillatory system, thus affecting the frequency of the oscillation. The imaginary part is the derivative gain, which influences the damping of the oscillation.

Following this pattern, and keeping as many terms as can conveniently be retained, we formulate a procedure for effecting a solution which will be termed the *small- σ solution*.

1. In the equation of reals, drop derivatives of σ and ω . Powers of σ greater than 2 can usually be dropped as well. Solve for $\omega(\sigma, A)$.

¹ For an interesting though considerably more complicated alternative to the DF as the descriptor for the nonlinearity, see Ref. 2.

2. Differentiate the resulting expression for ω to get $\dot{\omega}(\sigma, A)$ in which derivatives of σ are dropped.
3. In the equation of imaginaries, drop terms involving derivatives of σ and derivatives higher than the first of ω . Using $\omega(\sigma, A)$ and $\dot{\omega}(\sigma, A)$ from steps 1 and 2, solve for $\sigma(A)$. In this solution there is often little to be gained by retaining powers of σ greater than 1; powers greater than 2 should always be negligible.
4. Calculate time from the expression

$$t(A) = \int_{A(0)}^A \frac{1}{A\sigma(A)} dA$$

From this one can calculate directly $A(t)$, $\sigma(t)$, $\omega(t)$, $\psi(t)$, and thus $x(t)$.

The *quasi-static solution* results from dropping all derivative terms in σ , ω , and N from both equations. The two equations are then just algebraic equations for σ and ω , with A as a parameter, and can be solved by numerical iteration, graphing, or any other method. The solution is exactly that which one obtains by root-locus construction, using the original transfer function for the linear part of the system and treating A as a parameter for the purpose of evaluating the DF for the nonlinear part. Having this quasi-static solution for $\sigma(A)$ and $\omega(A)$, step 4 can be completed, as in the case of the small- σ solution. The quasi-static solution is mentioned primarily for the purpose of comparison.

If all derivative terms in σ , ω , and N are dropped from the differential equations for A and ω , as in the quasi-static case, and in addition σ is set equal to zero, the resulting equations are static relations in A and ω . If the solution, which we refer to as the *static*, or *steady-state*, *solution*, is nontrivial, the static values of A and ω define a steady-state oscillation which is a system limit cycle.

Example 4.3-1 We now illustrate these solution procedures in the case of the example system shown in Fig. 4.1-2. This solution can be generalized through nondimensionalization. If the nondimensional unit of time bt is employed, together with the nondimensional frequency variable

$$p = \frac{s}{b} = \frac{\sigma}{b} + j\frac{\omega}{b} = \mu + j\lambda \quad (4.3-3)$$

and the output level of the switch is associated with the gain of the linear part, the system diagram of Fig. 4.3-1 results in which there remains only a single parameter to characterize the system. That parameter, KD/b^2 , is the steady-state slewing velocity of the system in terms of nondimensional time, and for simplicity will be designated V .

$$V = \frac{KD}{b^2} \quad (4.3-4)$$

The system equation, corresponding to Eq. (4.1-15), is then

$$(p^2 + p)X(p) = -VY(p) \quad (4.3-5)$$

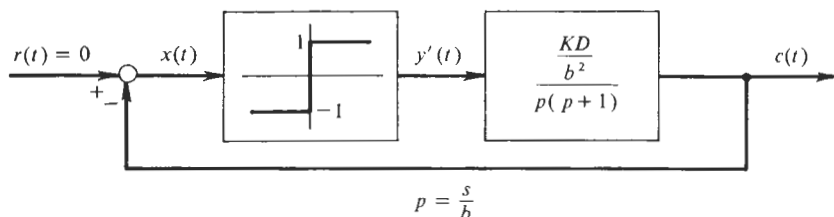


Figure 4.3-1 Nondimensionalized system of Example 4.3-1.

Application of the replacement rules gives

$$p^2 + \dot{p} + p = -VN \quad (4.3-6)$$

which has real and imaginary parts

$$\text{Real:} \quad \mu^2 - \lambda^2 + \dot{\mu} + \mu + \frac{4V}{\pi A} = 0 \quad (4.3-7a)$$

$$\text{Imaginary:} \quad 2\mu\lambda + \dot{\lambda} + \lambda = 0 \quad (4.3-7b)$$

To obtain the small- σ solution, the equation of reals is written

$$\mu^2 - \lambda^2 + \mu + \frac{4V}{\pi A} = 0 \quad (4.3-8)$$

$$\text{from which} \quad \lambda = \sqrt{\frac{4V}{\pi A} + \mu + \mu^2} \quad (4.3-9)$$

$$\text{Thus} \quad \dot{\lambda} = \frac{-(4V/\pi A^2)\dot{A} + \dot{\mu} + 2\mu\dot{\mu}}{2\lambda} \quad (4.3-10)$$

where the dot indicates differentiation with respect to bt . According to step 2, we drop the $\dot{\mu}$ terms. Also noting that $\dot{A}/A = \mu$, we get

$$\begin{aligned} \dot{\lambda} &= -\frac{(4V/\pi A)\mu}{2\lambda} \\ &= -\frac{1}{2\lambda} (\mu\lambda^2 - \mu^2 - \mu^3) \end{aligned} \quad (4.3-11)$$

The equation of imaginaries is then written according to step 3.

$$\begin{aligned} 2\mu\lambda + \dot{\lambda} + \lambda &= 0 \\ 2\mu + \frac{\dot{\lambda}}{\lambda} + 1 &= 0 \\ 2\mu - \frac{1}{2}\mu + \frac{1}{2}\left(\frac{\mu}{\lambda}\right)^2(1 + \mu) + 1 &= 0 \end{aligned} \quad (4.3-12)$$

At this point it would be possible to solve Eqs. (4.3-12) and (4.3-9) simultaneously for μ and λ as functions of A , but a simpler solution is adequate. According to Eq. (4.3-1),

μ/λ must be a small quantity if the DF is to represent the nonlinearity adequately, and there is no need to retain the second-degree term in this small quantity. Thus

$$\mu = -\frac{2}{3} \quad (4.3-13)$$

and

$$\lambda = \sqrt{\frac{4V}{\pi A} - \frac{2}{9}} \quad (4.3-14)$$

from Eq. (4.3-9).

For a system with unity feedback, $c(t) = -x(t)$; so this solution for $x(t)$ can be interpreted immediately in terms of $c(t)$ just by using initial conditions corresponding to $c(t)$. We shall work with the real parts of the complex functions, and use $\tau = bt$ as the non-dimensional time variable.

$$\begin{aligned} c(\tau) &= \text{Re} \{A(\tau) \exp [j\psi(\tau)]\} \\ &= A(\tau) \cos \psi(\tau) \end{aligned} \quad (4.3-15)$$

$$\dot{c}(\tau) = \mu(\tau)A(\tau) \cos \psi(\tau) - \lambda(\tau)A(\tau) \sin \psi(\tau) \quad (4.3-16)$$

The initial conditions are determined from these expressions.

$$c(0) = A(0) \cos \psi(0) \quad (4.3-17)$$

$$\dot{c}(0) = \mu(0)A(0) \cos \psi(0) - \lambda(0)A(0) \sin \psi(0) \quad (4.3-18)$$

For the case $c(0)/V = 1$, $\dot{c}(0) = 0$, we find

$$\frac{A(0)}{V} = 1.247 \quad \psi(0) = -36.68^\circ \quad (4.3-19)$$

Equations (4.3-19) and (4.3-13) imply

$$\frac{A(\tau)}{V} = 1.247 \exp \left(-\frac{2}{3}\tau \right) \quad (4.3-20)$$

and Eq. (4.3-14) becomes

$$\lambda(\tau) = \sqrt{\frac{4}{1.247\pi} \exp \left[\frac{2}{3}\tau \right] - \frac{2}{9}} \quad (4.3-21)$$

Finally,

$$\psi(\tau) = \psi(0) + \int_0^\tau \lambda(\tau_1) d\tau_1 \quad (4.3-22)$$

This can be integrated using tabulated forms to the result

$$\begin{aligned} \psi(\tau) = -36.68^\circ + \sqrt{2} \left\{ \frac{3}{\sqrt{2}} \lambda(\tau) - \tan^{-1} \frac{3}{\sqrt{2}} \lambda(\tau) \right. \\ \left. - \left[\frac{3}{\sqrt{2}} \lambda(0) - \tan^{-1} \frac{3}{\sqrt{2}} \lambda(0) \right] \right\} \end{aligned} \quad (4.3-23)$$

where $\lambda(\tau)$ is given by Eq. (4.3-21). The term $(3/\sqrt{2})\lambda(\tau)$ is in radian measure. Equations (4.3-15), (4.3-20), and (4.3-23) define the function $c(\tau)/V$, following the stated initial conditions. This function is plotted in Fig. 4.3-2, together with the exact solution. The agreement is seen to be excellent in spite of the substantial change in successive peaks in the early part of the response. The conditions of Eqs. (4.3-1) and (4.3-2), which should

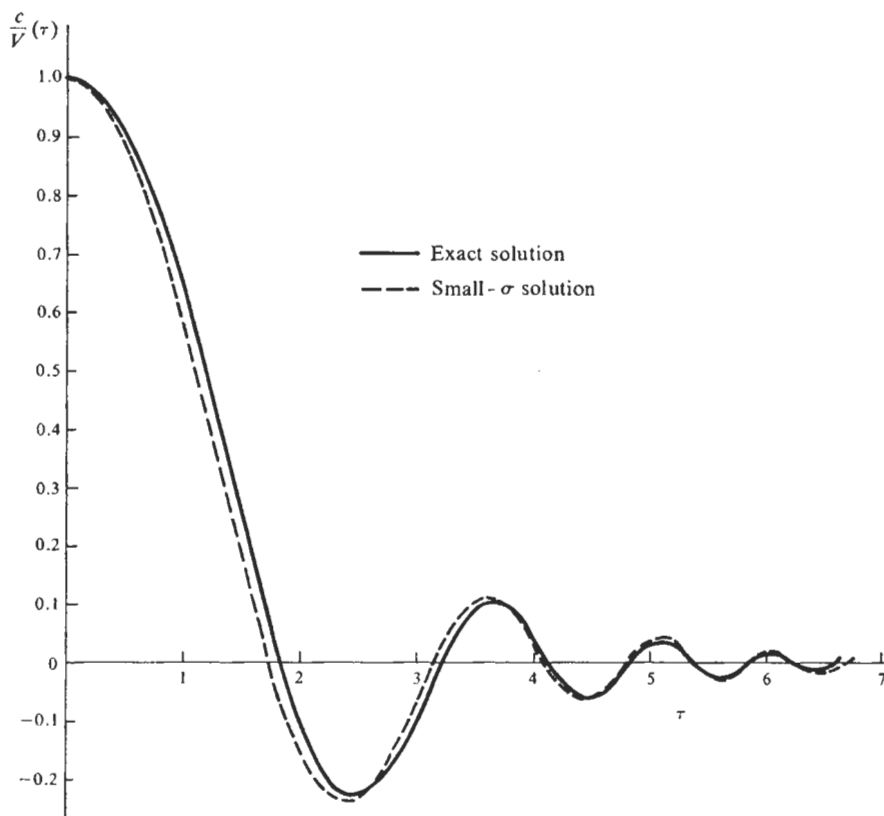


Figure 4.3-2 Solution of Example 4.3-1.

hold if one is to have confidence in characterizing the nonlinearity by its DF, are found to be badly violated in the early part of the response, when λ is small.

$$2\pi \frac{|\mu|}{\lambda} (0) = 4.67$$

$$2\pi \frac{|\dot{\lambda}|}{\lambda^2} (0) = 2.99$$

In spite of this, and the fact that the system linear part provides only second-order filtering of the harmonics in the nonlinearity output, the accuracy of the solution is excellent.

The quasi-static solution to this problem is readily found to be

$$\mu = -\frac{1}{2} \quad (4.3-24)$$

$$\lambda = \sqrt{\frac{4V}{\pi A} - \frac{1}{4}} \quad (4.3-25)$$

These values, obtained by solving Eqs. (4.3-7) after dropping the derivative terms, simply define the location of the closed-loop poles of this system in the p plane, considering the

amplitude of the oscillation to be quasi-stationary. The damping factor μ in this solution is in error by 25 percent, which is just the contribution of $\dot{\lambda}$ in the equation of imaginaries.

In many cases of practical importance, the complete solution to $x(t)$ or $c(t)$ is not required; it is often sufficient just to know how the amplitude of the oscillation behaves. The solution for $A(t)$ is appreciably simpler than the solution for $x(t)$; we illustrate this situation in the following example.

Example 4.3-2 Consider the feedback system consisting of a relay-controlled third-order plant as shown in Fig. 4.3-3. We require the history of the amplitude of the limit cycle as it builds up from zero to its steady-state value. Application of the rules of Sec. 4.1 gives as the equations of reals and imaginaries

$$\text{Real: } \frac{1}{\omega_n^2} (\sigma^3 - 3\sigma\omega^2 + 3\sigma\dot{\sigma} - 3\omega\dot{\omega} + \ddot{\sigma}) + \frac{2\zeta}{\omega_n} (\sigma^2 - \omega^2 + \dot{\sigma}) + \sigma + \frac{4KD}{\pi A} = 0 \quad (4.3-26a)$$

$$\text{Imaginary: } \frac{1}{\omega_n^2} (3\sigma^2\omega - \omega^3 + 3\sigma\dot{\omega} + 3\omega\dot{\sigma} + \ddot{\omega}) + \frac{2\zeta}{\omega_n} (2\sigma\omega + \dot{\omega}) + \omega = 0 \quad (4.3-26b)$$

For the small- σ solution the equation of reals is written

$$\frac{1}{\omega_n^2} (-3\sigma\omega^2) + \frac{2\zeta}{\omega_n} (\sigma^2 - \omega^2) + \sigma + \frac{4KD}{\pi A} = 0 \quad (4.3-27)$$

which gives

$$\omega = \sqrt{\frac{4KD/\pi A + \sigma + (2\zeta/\omega_n)\sigma^2}{2\zeta/\omega_n + 3\sigma/\omega_n^2}} \quad (4.3-28)$$

Before taking the derivative of ω , we note that in writing the equation of imaginaries, it will be convenient to divide through by ω . The derivative of ω will then appear in the form

$$\frac{\dot{\omega}}{\omega} = \frac{d(\ln \omega)}{dt} = \frac{1}{2} \frac{-(4KD/\pi A)\sigma}{4KD/\pi A + \sigma + (2\zeta/\omega_n)\sigma^2} \quad (4.3-29)$$

ignoring the $\dot{\sigma}$ terms. Then dropping $\dot{\sigma}$ and $\ddot{\omega}$ according to the stated procedure for the small- σ solution, the equation of imaginaries is written

$$\begin{aligned} \frac{1}{\omega_n^2} \left[3\sigma^2 - \frac{4KD/\pi A + \sigma + (2\zeta/\omega_n)\sigma^2}{2\zeta/\omega_n + 3\sigma/\omega_n^2} - \frac{3}{2} \frac{(4KD/\pi A)\sigma^2}{4KD/\pi A + \sigma + (2\zeta/\omega_n)\sigma^2} \right] \\ + \frac{2\zeta}{\omega_n} \left[2\sigma - \frac{1}{2} \frac{(4KD/\pi A)\sigma}{4KD/\pi A + \sigma + (2\zeta/\omega_n)\sigma^2} \right] + 1 = 0 \end{aligned} \quad (4.3-30)$$

which is expanded to

$$\left(2 + 12\zeta^2 + 16\zeta^2 \frac{1}{a} \right) \sigma^2 + 2\zeta\omega_n \left[1 + (1 + 6\zeta^2) \frac{1}{a} \right] \sigma + (2\zeta\omega_n)^2 \left(1 - \frac{1}{a} \right) \frac{1}{a} = 0 \quad (4.3-31)$$

keeping only terms to σ^2 . In this expression, the definition

$$a = \frac{\pi \zeta \omega_n}{2 KD} A \quad (4.3-32)$$

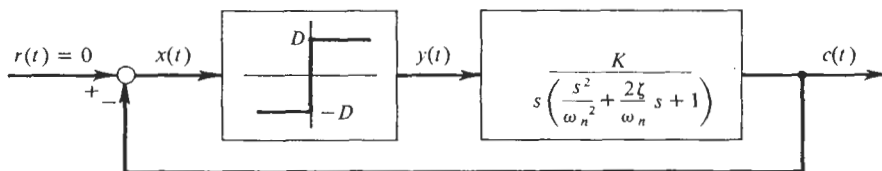


Figure 4.3-3 System of Example 4.3-2.

has been used. Note that, for this system, the steady-state limit cycle has $a = 1$, $\omega = \omega_n$, as can be derived from these equations by noting that $a = 1$ satisfies Eq. (4.3-31), with $\sigma = 0$, then evaluating ω from Eq. (4.3-28), with $a = 1$ and $\sigma = 0$.

Equation (4.3-31) is a nonlinear relation between a and $\sigma = \dot{a}/a$. A plot of $a\sigma$ versus a as defined by Eq. (4.3-31) is a phase-plane trajectory for the amplitude of the transient oscillation; this plot is shown in Fig. 4.3-4 for values of a ranging from 0.1 to 1.0. From this plot one can evaluate $a(t)$, using any of the standard methods of determining the time response from a phase-plane trajectory. (See, for example, Ref. 9, sec. 11.2.)

In the application to transient oscillations, the curves of $\dot{A} = A\sigma$ versus A for systems having a steady-state limit cycle generally have a near-linear characteristic for a substantial range of A around the limit cycle amplitude. In these cases it is convenient to fit the curve with one or more straight-line segments and treat each segment in the following manner. If

$$\dot{A} = m + nA \quad (4.3-33)$$

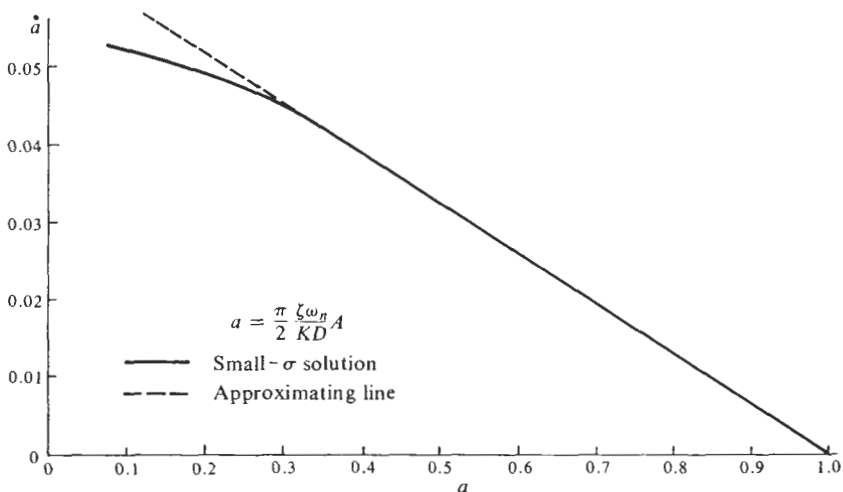


Figure 4.3-4 Phase trajectory for the amplitude of the transient oscillation. Example 4.3-2, $\zeta = 5$, $\omega_n = 1$ rad/sec.

over some interval of A , then the differential equation

$$\dot{A} - nA = m \quad (4.3-34)$$

having the solution

$$A(t) = -\frac{m}{n} + \left[\frac{m}{n} + A(0) \right] e^{nt} \quad (4.3-35)$$

holds over that interval. Exponential segments of $A(t)$ corresponding to linear segments of $\dot{A}(A)$ can then be pieced together, using a new time origin for each segment. Voronov (Ref. 10) has noted the usefulness of this procedure.

Example 4.3-2 (continued) For our present example, Eq. (4.3-31) as plotted in Fig. 4.3-4 can be fit quite well over the important part of its range by the single dashed line shown in the figure. The error in this fit at the small values of a is of little concern for two reasons: First, $\dot{a}$ is large in that region, so the value of a does not remain in that region long; and second, the original curve is not accurate at these small values of a because terms in σ^3 , $\dot{\sigma}$, and $\ddot{\omega}$ are not negligible there. Use of just the single approximating line gives

$$\dot{a} = 0.065 - 0.065a \quad (4.3-36)$$

and if this line is considered valid for the full interval of a ranging from 0 to 1, we have

$$a(t) = 1 - e^{-0.065t} \quad (4.3-37)$$

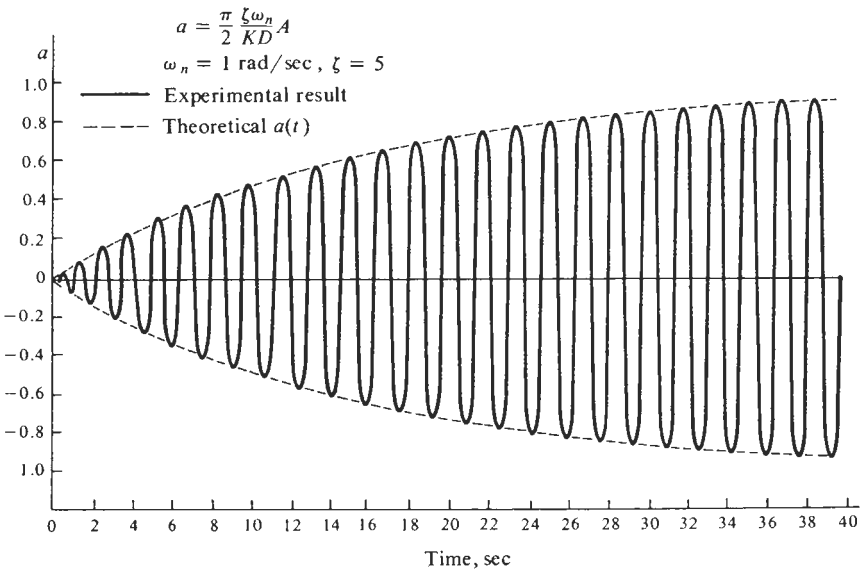


Figure 4.3-5 Solution of Example 4.3-2 (limit cycle build-up).

starting from the initial condition $a(0) = 0$. Actually, there is no analytic justification for presuming this solution to be valid for a less than 0.1; the solution effected here should be expected to define the limit cycle amplitude build-up over the range $a > 0.1$. However, it is found, when that solution is plotted with an experimentally observed build-up, that a continuation of the same exponential back to $a = 0$ continues to fit the observed data very well. The solution equation (4.3-37) is plotted in Fig. 4.3-5, together with a record of the limit cycle build-up in an analog simulation of this example system. The agreement is seen to be excellent.

4.4 LIMIT CYCLE DYNAMICS

A very important special case in the general study of transient oscillations is the subject of this section: the study of the dynamic characteristics of variations in limit cycles. This topic has particular relevance to the design of amplitude-regulating loops for limit cycling systems. The value of an automatic means of regulating the amplitude of a limit cycle in the presence of changing plant characteristics seems self-evident; the major considerations are the avoidance of large-amplitude limit cycles which would be costly, or uncomfortable, or perhaps dangerous, and the avoidance of loss of control capability, which usually accompanies very small amplitude limit cycles.

The general configuration of an amplitude-regulating loop is given in Fig. 4.4-1. The closed-loop limit cycling system is an operating control system. The amplitude of the limit cycle at some station in this system is sensed by an amplitude indicator, compared with the reference amplitude, and the error operated on to cause the adjustment of some parameter in the control system. The adjustment operator usually consists of an integrator plus any compensation that may be necessary to achieve the specified performance of the amplitude-regulating loop. The design of the regulating loop, including possible compensation, can be carried out in a direct manner only if the

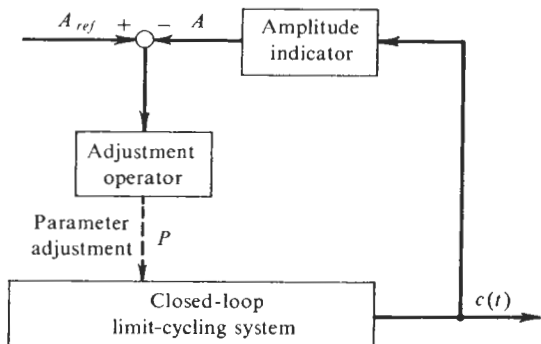


Figure 4.4-1 Limit cycle amplitude-regulating system.

dynamic response characteristics of each element in the loop are known and presented in a readily usable form. The adjustment operator is a standard linear device which presents no new problem. The amplitude indicator can be characterized to a good approximation by a readily derived transfer function (see, for example, Ref. 3, app. E). However, the dynamics of the change in limit cycle amplitude in response to a change in a parameter of the system is a more complicated matter. In this section we derive a transfer function to represent this dynamic characteristic.

The static sensitivity of the transfer from a parameter change, ΔP , to the resulting limit cycle amplitude change, ΔA , can be determined by the methods of Chap. 3. The steady-state limit cycle amplitude can be determined for several values of the parameter in the range of interest and the slope of the curve of A versus P taken as the static gain. But there is a dynamic effect as well, since the amplitude does not generally change instantaneously from one value to another when P is changed; there is usually a transient such as that pictured in Fig. 4.4-2. This transient is a transient oscillation, and can be analyzed by the methods of the preceding sections. However, it is of little value to the designer of an amplitude-regulating loop to know that he can plot the phase trajectory $\dot{A}$ versus A and thus deduce $A(t)$ for any particular situation. He needs a more readily usable description of the amplitude transient.

Such a description is indicated by the solution to the second example problem given in the preceding section. It was noted there that there was a substantial range of A around the limit cycle amplitude for which the curve of $\dot{A}$ versus A was very nearly linear, and that $A(t)$ was an exponential

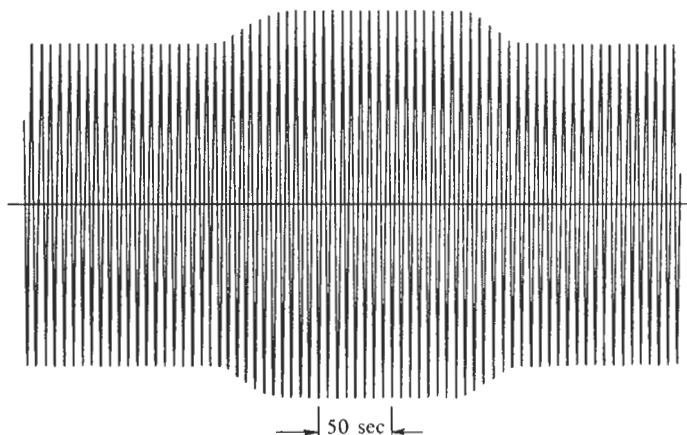


Figure 4.4-2 A limit cycle amplitude transient. System of Fig. 4.3-3, $\omega_n = 1 \text{ rad/sec}$, $\zeta = 5$.

function over this range. Furthermore, in an amplitude-regulating loop which is operating continuously, one would not expect $\dot{A}$ to depart very far from zero; it would just take small positive and negative values as necessary to adjust the amplitude. It seems quite appropriate for this purpose, then, to approximate $\dot{A}$ as a linear function of A , taking the value zero at the limit cycle amplitude corresponding to the selected operating point and having the slope of the curve of $\dot{A}$ versus A at $\dot{A} = 0$.

$$\dot{A} = \left. \frac{d\dot{A}}{dA} \right|_{\dot{A}=0} (A - A_0) \quad (4.4-1)$$

A_0 is the steady-state limit cycle amplitude. This first-order differential equation is the characteristic equation of a first-order lag with time constant

$$\tau = - \left[\left. \frac{d\dot{A}}{dA} \right|_{\dot{A}=0} \right]^{-1} \quad (4.4-2)$$

The desired transfer function which characterizes the changes in limit cycle amplitude in response to changes in a system parameter P is then

$$\frac{\Delta A}{\Delta P} = \frac{(\Delta A / \Delta P)_{ss}}{\tau s + 1} \quad (4.4-3)$$

where the gain is the steady-state change in A for a given change in P , which can be evaluated by static methods as described before, and τ is given by Eq. (4.4-2).

A more convenient expression for the time constant is obtained by noting that

$$\begin{aligned} \left. \frac{d\dot{A}}{dA} \right|_{\dot{A}=0} &= \left. \frac{d(A\sigma)}{dA} \right|_{\sigma=0} \\ &= \left[A \frac{d\sigma}{dA} + \sigma \right]_{\sigma=0} \\ &= A_0 \left. \frac{d\sigma}{dA} \right|_{\sigma=0} \end{aligned}$$

Thus

$$\tau = - \left[A_0 \left. \frac{d\sigma}{dA} \right|_{\sigma=0} \right]^{-1} \quad (4.4-4)$$

INDIRECT SOLUTION FOR THE TIME CONSTANT

One way of determining this time constant is to follow the procedure for the small- σ solution given in Sec. 4.3 to the point in step 3 where one has an equation relating σ to A . For the second example problem this is Eq. (4.3-31). For the present purpose there is no need to plot the complete

curve of $\dot{A}$ versus A ; one can simply take the first derivative of each term in this equation and thus evaluate $d\sigma/dA$ at the point $\sigma = 0$, $A = A_0$. It is clear that terms in σ of higher degree than the first make no contribution to the result; so it is convenient to drop the higher-degree terms in σ at the outset, write the linear solution $\sigma(A)$, and take the first derivative.

For Eq. (4.3-31) this gives

$$\sigma = -\frac{2\zeta\omega_n(1 - 1/a)(1/a)}{1 + (1 + 6\zeta^2)(1/a)} \quad (4.4-5)$$

$$\left. \frac{d\sigma}{da} \right|_{a=1} = -\frac{2\zeta\omega_n}{2 + 6\zeta^2} \quad (4.4-6)$$

The time constant for the limit cycle amplitude transfer function for this system shown in Fig. 4.3-3 is

$$\tau = \frac{1 + 3\zeta^2}{\zeta\omega_n} \quad (4.4-7)$$

and when nondimensionalized with respect to the steady-state limit cycle period, $T_0 = 2\pi/\omega_n$, this becomes

$$\frac{\tau}{T_0} = \frac{1 + 3\zeta^2}{2\pi\zeta} \quad (4.4-8)$$

This relation has been checked experimentally by analog simulation for values of ζ ranging from 0.01 to 10. The limit cycle transient was excited by a step change in the forward gain of the system; the changing limit cycle was observed at the output of the linear part. The amplitude response in every case approximated an exponential very closely. Figure 4.4-2 shows this response for the case $\zeta = 5$. The time constant of the response was read as the time at which the amplitude transient completed 63 percent of its total excursion. These experimental results are plotted in Fig. 4.4-3, together with the theoretical curve of Eq. (4.4-8). The agreement is seen to be excellent over the full 1,000:1 range of damping ratio. It may be noted that the minimum time constant occurs for $\zeta = 1/\sqrt{3} = 0.577$, at which value $\tau/T_0 = 0.552$. In the mid-range of damping ratio, the amplitude transient is essentially completed within one period of the limit cycle, whereas systems with very high or very low damping ratios have amplitude transients lasting many periods.

The time constant for the amplitude transient based on the quasi-static solution of Eqs. (4.3-26) is easily derived (Ref. 4); the result is

$$\frac{\tau}{T_0} = \frac{1 + 4\zeta^2}{2\pi\zeta} \quad (4.4-9)$$

This expression is also plotted in Fig. 4.4-3. The distinction between the small- σ solution and the quasi-static solution in this case appears for the

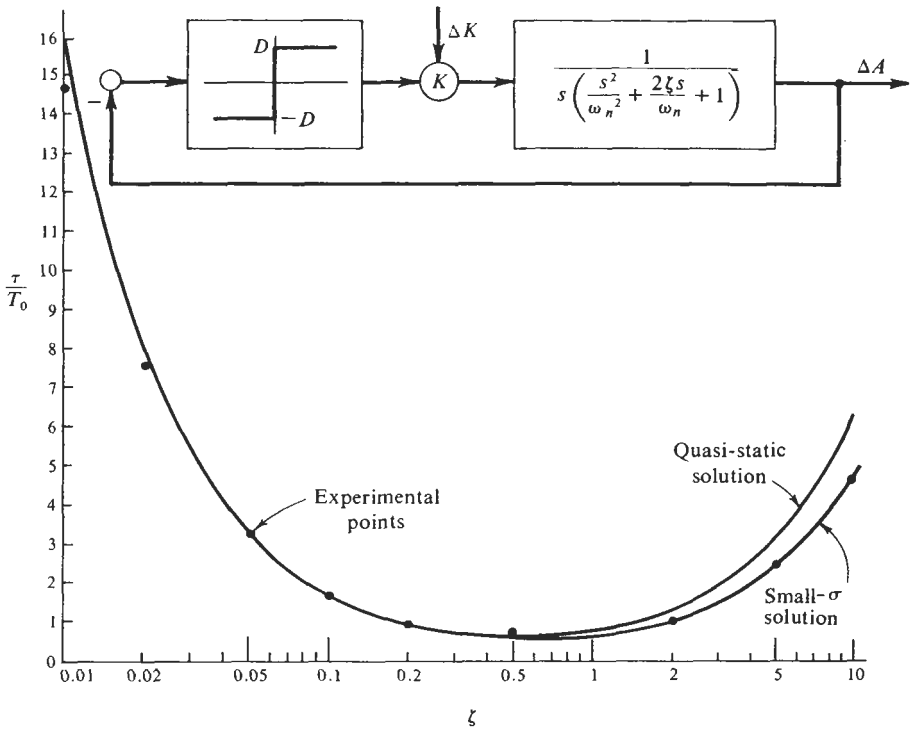


Figure 4.4-3 Time constant of limit cycle amplitude dynamics. System of Fig. 4.3-3.

large values of ξ only; the error in the quasi-static solution increases to 37 percent at $\xi = 10$.

DIRECT SOLUTION FOR THE TIME CONSTANT

The solution given above for the time constant of the limit cycle amplitude transfer function was pursued primarily to emphasize the relationship of limit cycle dynamics to more general transient oscillations. For the determination of the amplitude transfer function time constant in practice, a more direct procedure is preferable. It is not necessary to solve the equation of reals for ω and put explicit expressions for ω and $\dot{\omega}$ into the equation of imaginaries; rather, one can write first differentials for the equations of reals and imaginaries directly and, with suitable assumptions, solve for $d\sigma/dA$, evaluated at the steady-state limit cycle condition.

If the equations of reals and imaginaries are written dropping derivatives of σ and derivatives of ω of higher order than one, they are of the form

$$R(\sigma, \omega, A, \dot{\omega}) = 0 \quad (4.4-10)$$

$$I(\sigma, \omega, A, \dot{\omega}) = 0 \quad (4.4-11)$$

Terms involving powers of σ greater than 1 can be dropped since they cannot influence the end result, which is the first derivative of σ evaluated at $\sigma = 0$. All powers of ω should be retained. In Eqs. (4.4-10) and (4.4-11), terms in σ , ω , and $\dot{\omega}$ enter directly, A and ω enter through n_p and n_q , and σ and $\dot{\omega}$ enter through $\dot{n}_p$ and $\dot{n}_q$. The relationship among first differentials is

$$\frac{\partial R}{\partial \sigma} d\sigma + \frac{\partial R}{\partial \omega} d\omega + \frac{\partial R}{\partial A} dA + \frac{\partial R}{\partial \dot{\omega}} d\dot{\omega} = 0 \quad (4.4-12)$$

$$\frac{\partial I}{\partial \sigma} d\sigma + \frac{\partial I}{\partial \omega} d\omega + \frac{\partial I}{\partial A} dA + \frac{\partial I}{\partial \dot{\omega}} d\dot{\omega} = 0 \quad (4.4-13)$$

According to the procedure for the small- σ solution, the equation of reals is solved for ω ; in this case we solve for $d\omega$. It is no more difficult in this case to retain the effect of $d\dot{\omega}$.

$$d\omega = -\left(\frac{\partial R}{\partial \omega}\right)^{-1} \left(\frac{\partial R}{\partial \sigma} d\sigma + \frac{\partial R}{\partial A} dA + \frac{\partial R}{\partial \dot{\omega}} d\dot{\omega} \right) \quad (4.4-14)$$

From this we can write $\dot{\omega}$ as

$$\begin{aligned} \dot{\omega} &= \frac{d\omega}{dt} = -\left(\frac{\partial R}{\partial \omega}\right)^{-1} \left(\frac{\partial R}{\partial \sigma} \frac{d\sigma}{dt} + \frac{\partial R}{\partial A} \frac{dA}{dt} + \frac{\partial R}{\partial \dot{\omega}} \frac{d\dot{\omega}}{dt} \right) \\ &\approx -\left(\frac{\partial R}{\partial \omega}\right)^{-1} \left(\frac{\partial R}{\partial A} A_0 \dot{\sigma} \right) \end{aligned} \quad (4.4-15)$$

since $\dot{\sigma}$ and $\dot{\omega}$ terms are not being retained in the equations.

$$d\dot{\omega} = -\left(\frac{\partial R}{\partial \omega}\right)^{-1} \frac{\partial R}{\partial A} A_0 d\sigma \quad (4.4-16)$$

at the steady-state limit cycle point, where $\sigma = 0$. Now these expressions for $d\omega$ and $d\dot{\omega}$ can be used in the equation of imaginaries [Eq. (4.4-13)]:

$$\begin{aligned} \frac{\partial I}{\partial \sigma} d\sigma + \frac{\partial I}{\partial \omega} \left(-\frac{\partial R}{\partial \omega} \right)^{-1} \left[\frac{\partial R}{\partial \sigma} d\sigma + \frac{\partial R}{\partial A} dA + \frac{\partial R}{\partial \dot{\omega}} \left(-\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A} A_0 d\sigma \right] \\ + \frac{\partial I}{\partial A} dA + \frac{\partial I}{\partial \dot{\omega}} \left(-\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A} A_0 d\sigma = 0 \end{aligned}$$

This reduces to

$$\begin{aligned} \frac{d\sigma}{dA} = - \frac{\frac{\partial I}{\partial A} - \frac{\partial I}{\partial \omega} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A}}{\frac{\partial I}{\partial \sigma} - \frac{\partial I}{\partial \dot{\omega}} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A} A_0 - \frac{\partial I}{\partial \omega} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \left[\frac{\partial R}{\partial \sigma} - \frac{\partial R}{\partial \dot{\omega}} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A} A_0 \right]} \end{aligned} \quad (4.4-17)$$

Using Eq. (4.4-4), we have, finally, as the time constant of the transfer function for amplitude variations,

$$\tau = \frac{\frac{1}{A_0} \frac{\partial I}{\partial \sigma} - \frac{\partial I}{\partial \dot{\omega}} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A} - \frac{\partial I}{\partial \omega} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \left[\frac{1}{A_0} \frac{\partial R}{\partial \sigma} - \frac{\partial R}{\partial \dot{\omega}} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A} \right]}{\frac{\partial I}{\partial A} - \frac{\partial I}{\partial \omega} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A}} \quad (4.4-18)$$

This is an explicit expression for the time constant in terms of the partial derivatives of the real and imaginary functions evaluated at the steady-state limit cycle point: $\sigma = 0$, $\omega = \omega_0$, $A = A_0$, $\dot{\omega} = 0$.

The steady-state gain of the amplitude transfer function can be determined in a similar manner. The steady-state limit cycle is given by the solution of the equations of reals and imaginaries for the steady-state situation in which σ and all derivatives of σ and ω are zero. These equations are then just functions of A and ω , and of course the various parameters of the system. One of these parameters will be used to control the limit cycle amplitude; call that parameter P , and show explicitly the dependence of the real and imaginary functions on P . These functions, similar to those given in Eqs. (4.4-10) and (4.4-11), but including only the steady-state terms, are

$$R(A, \omega, P) = 0 \quad (4.4-19)$$

$$I(A, \omega, P) = 0 \quad (4.4-20)$$

We require the differential sensitivity relating changes in P to changes in A .

$$\frac{\partial R}{\partial A} dA + \frac{\partial R}{\partial \omega} d\omega + \frac{\partial R}{\partial P} dP = 0 \quad (4.4-21)$$

$$\frac{\partial I}{\partial A} dA + \frac{\partial I}{\partial \omega} d\omega + \frac{\partial I}{\partial P} dP = 0 \quad (4.4-22)$$

Elimination of the differential frequency change between these equations gives the desired result.

$$\frac{dA}{dP} = - \frac{\frac{\partial I}{\partial P} - \frac{\partial I}{\partial \omega} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial P}}{\frac{\partial I}{\partial A} - \frac{\partial I}{\partial \omega} \left(\frac{\partial R}{\partial \omega} \right)^{-1} \frac{\partial R}{\partial A}} \quad (4.4-23)$$

This is the steady-state gain of the amplitude transfer function between the parameter P and the amplitude of the limit cycle at the input to the nonlinear part of the system. If the amplitude is to be monitored at some other point in the system, the transfer between that point and $x(t)$, to which the amplitude above applies, is a linear one. This yields a simple relation for the change in amplitude based on the change in amplitude at $x(t)$, and the change in frequency, which can also be determined from Eqs. (4.4-21) and (4.4-22).

Equations (4.4-23) and (4.4-18) give the values of the parameters for the amplitude transfer function of first-order form as shown in Eq. (4.4-3). The partial derivatives in Eq. (4.4-23) are derivatives of the steady-state functions [Eqs. (4.4-19) and (4.4-20)]. But these are identical with the corresponding derivatives of the more complicated functions [Eqs. (4.4-10) and (4.4-11)] when evaluated at the steady-state limit cycle point as required for use in Eq. (4.4-18). Thus, in practice, it is not necessary to write the steady-state equations and take derivatives; the same derivatives required in the evaluation of the time constant may be used to evaluate the gain.

These expressions for the time constant and static gain of the amplitude transfer function may seem somewhat formidable, but in practice the calculation and evaluation of the required partial derivatives is carried out more rapidly than alternative procedures requiring graphing of functions and the determination of slopes.

This more direct procedure for the determination of the amplitude transfer function is illustrated by application to an example system which is a bit more complicated than those considered previously. In this case, the nonlinearity has both real and imaginary parts; the limit cycle frequency as well as amplitude varies with the controlling parameter; the linear part of the system includes a lead term; and the controlling parameter is a time constant rather than the more common forward gain.

Example 4.4-1 This system is shown in Fig. 4.4-4. The variable used to adjust the limit cycle is the time constant of the lead term τ ; thus the parameter P in Eq. (4.4-23) is in this case τ . The amplitude of the limit cycle is observed at the output $c(t)$, which is just the negative of $x(t)$, the input to the nonlinear part of the system. We proceed to derive the equations of the transient oscillation following the steps of Sec. 4.1. The system equation corresponding to Eq. (4.1-15) is

$$\left(\frac{1}{\omega_n^2} s^3 + \frac{2\zeta}{\omega_n} s^2 + s \right) X(s) = -(K + K\tau s) Y(s) \quad (4.4-24)$$

Application of the replacement rules gives

$$\frac{1}{\omega_n^2} (s^3 + 3s\dot{s} + \ddot{s}) + \frac{2\zeta}{\omega_n} (s^2 + \dot{s}) + s = -[KN + K\tau(Ns + \dot{N})] \quad (4.4-25)$$

which has real and imaginary parts

$$\begin{aligned} \text{Real: } \frac{1}{\omega_n^2} (\sigma^3 - 3\sigma\omega^2 + 3\sigma\dot{\sigma} - 3\omega\dot{\omega} + \ddot{\sigma}) + \frac{2\zeta}{\omega_n} (\sigma^2 - \omega^2 + \dot{\sigma}) \\ + \sigma + Kn_p + K\tau(n_p\sigma - n_\omega\omega + \dot{n}_p) = 0 \end{aligned} \quad (4.4-26a)$$

$$\begin{aligned} \text{Imaginary: } \frac{1}{\omega_n^2} (3\sigma^2\omega - \omega^3 + 3\sigma\dot{\omega} + 3\omega\dot{\sigma} + \ddot{\omega}) + \frac{2\zeta}{\omega_n} (2\sigma\omega + \dot{\omega}) \\ + \omega + Kn_q + K\tau(n_p\omega + n_\omega\sigma + \dot{n}_q) = 0 \end{aligned} \quad (4.4-26b)$$

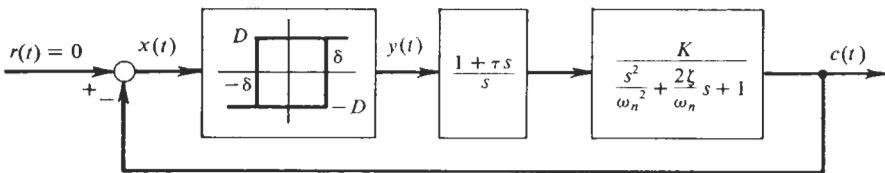


Figure 4.4-4 System of Example 4.4-1.

For the purpose of describing limit cycle dynamics, powers of σ greater than one are dropped, all derivatives of σ are dropped, and derivatives of ω of order greater than one are dropped.

$$R = \frac{1}{\omega_n^2} (-3\sigma\omega^2 - 3\omega\dot{\omega}) + \frac{2\zeta}{\omega_n} (-\omega^2) + \sigma + Kn_p + K\tau(n_p\sigma - n_q\omega + \dot{n}_p) \quad (4.4-27a)$$

$$I = \frac{1}{\omega_n^2} (-\omega^3 + 3\sigma\omega) + \frac{2\zeta}{\omega_n} (2\sigma\omega + \dot{\omega}) + \omega + Kn_q + K\tau(n_p\omega + n_q\sigma + \dot{n}_q) \quad (4.4-27b)$$

Partial derivatives of these functions with respect to σ , ω , A , $\dot{\omega}$, and τ are to be written. The contributions of the DF for the nonlinear part of the system to these derivatives are seen from the explicit statements of functional dependence:

$$n_p = n_p(A, \omega) \quad (4.4-28)$$

$$n_q = n_q(A, \omega) \quad (4.4-29)$$

$$\dot{n}_p = \frac{\partial n_p}{\partial A} A\sigma + \frac{\partial n_p}{\partial \omega} \dot{\omega} \quad (4.4-30)$$

$$\dot{n}_q = \frac{\partial n_q}{\partial A} A\sigma + \frac{\partial n_q}{\partial \omega} \dot{\omega} \quad (4.4-31)$$

The derivatives are

$$\frac{\partial R}{\partial \sigma} = -\frac{3}{\omega_n^2} \omega^2 + 1 + K\tau n_p + K\tau \frac{\partial n_p}{\partial A} A$$

$$\begin{aligned} \frac{\partial R}{\partial \omega} = & -\frac{6}{\omega_n^2} \sigma\omega - \frac{3}{\omega_n^2} \dot{\omega} - \frac{4\zeta}{\omega_n} \omega + K \frac{\partial n_p}{\partial \omega} + K\tau\sigma \frac{\partial n_p}{\partial \omega} - K\tau n_q - K\tau\omega \frac{\partial n_q}{\partial \omega} \\ & + K\tau \frac{\partial^2 n_p}{\partial A \partial \omega} A\sigma + K\tau \frac{\partial^2 n_p}{\partial \omega^2} \dot{\omega} \end{aligned}$$

$$\frac{\partial R}{\partial A} = K \frac{\partial n_p}{\partial A} + K\tau\sigma \frac{\partial n_p}{\partial A} - K\tau\omega \frac{\partial n_q}{\partial A} + K\tau \frac{\partial n_p}{\partial A} \sigma + K\tau \frac{\partial^2 n_p}{\partial A^2} A\sigma + K\tau \frac{\partial^2 n_p}{\partial \omega \partial A} \dot{\omega}$$

$$\frac{\partial R}{\partial \dot{\omega}} = -\frac{3\omega}{\omega_n^2} + K\tau \frac{\partial n_p}{\partial \omega}$$

$$\frac{\partial R}{\partial \tau} = Kn_p\sigma - Kn_q\omega + K\dot{n}_p$$

$$\frac{\partial I}{\partial \sigma} = \frac{3}{\omega_n^2} \dot{\omega} + \frac{4\zeta}{\omega_n} \omega + K\tau n_q + K\tau \frac{\partial n_q}{\partial A} A$$

$$\begin{aligned} \frac{\partial I}{\partial \omega} = & -\frac{3}{\omega_n^2} \omega^2 + \frac{4\zeta}{\omega_n} \sigma + 1 + K \frac{\partial n_q}{\partial \omega} + K\tau n_p + K\tau\omega \frac{\partial n_p}{\partial \omega} + K\tau\sigma \frac{\partial n_q}{\partial \omega} \\ & + K\tau \frac{\partial^2 n_q}{\partial A \partial \omega} A\sigma + K\tau \frac{\partial^2 n_q}{\partial \omega^2} \dot{\omega} \end{aligned}$$

$$\frac{\partial I}{\partial A} = K \frac{\partial n_q}{\partial A} + K\tau\omega \frac{\partial n_p}{\partial A} + K\tau\sigma \frac{\partial n_q}{\partial A} + K\tau \frac{\partial n_q}{\partial A} \sigma + K\tau \frac{\partial^2 n_q}{\partial A^2} A\sigma + K\tau \frac{\partial^2 n_q}{\partial \omega \partial A} \dot{\omega}$$

$$\frac{\partial I}{\partial \dot{\omega}} = \frac{3}{\omega_n^2} \sigma + \frac{2\zeta}{\omega_n} + K\tau \frac{\partial n_q}{\partial \omega}$$

$$\frac{\partial I}{\partial \tau} = Kn_p\omega + Kn_q\sigma + K\dot{n}_q$$

These derivatives are to be evaluated at the steady-state limit cycle for the values of system parameters chosen as the operating point. The limit cycle is defined by the static equations

$$\frac{2\zeta}{\omega_n}(-\omega^2) + Kn_p - K\tau n_q \omega = 0 \quad (4.4-32a)$$

$$-\frac{1}{\omega_n^2}\omega^3 + \omega + Kn_q + K\tau n_p \omega = 0 \quad (4.4-32b)$$

The nonlinearity in this case is rectangular hysteresis, for which the DF was calculated in Chap. 2. Equation (2.3-26) gives

$$n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} \quad (4.4-33a)$$

$$n_q = -\frac{4D\delta}{\pi A^2} \quad (4.4-33b)$$

The values selected for the nominal operating point are $\zeta = 2$, $\omega_n = 1$ radian/sec, $KD = 346$ volts/sec, $\delta = 10$ volts, $\tau = 0.22$ sec. A graphical solution of Eqs. (4.4-32) and (4.4-33) using these parameter values yields the results $A = 68.1$ volts, $\omega = 1.29$ radians/sec.

The steady-state limit cycle is then characterized by $A_0 = 68.1$ volts, $\omega_0 = 1.29$ radians/sec, $\sigma = \dot{\omega} = 0$. Using these values, the specified system parameters listed above, and Eqs. (4.4-33), the partial derivatives are found to be

$$\frac{\partial R}{\partial \sigma} = -3.96 \quad \frac{\partial I}{\partial \sigma} = 10.53$$

$$\frac{\partial R}{\partial \omega} = -10.11 \quad \frac{\partial I}{\partial \omega} = -2.58$$

$$\frac{\partial R}{\partial A} = -0.100 \quad \frac{\partial I}{\partial A} = 0.0012$$

$$\frac{\partial R}{\partial \dot{\omega}} = -3.87 \quad \frac{\partial I}{\partial \dot{\omega}} = 4.00$$

$$\frac{\partial R}{\partial \tau} = 1.228 \quad \frac{\partial I}{\partial \tau} = 8.26$$

Use of these numbers in Eqs. (4.4-23) and (4.4-18) yields the theoretical transfer function relating changes in the lead-term time constant to changes in the amplitude of the limit cycle at the output of the system.

$$\frac{\Delta A}{\Delta \tau} = \frac{-298}{4.51s + 1} \quad (4.4-34)$$

An analog simulation of this system was tested by making step changes in the lead-term time constant between the values 0.20 and 0.24, which is a 20 percent change centered around the operating-point value used in the calculation above. As in the previous example, the time constant of the resulting amplitude response was taken to be the time in which the amplitude executed 63 percent of its change. The observed time constant was 4.4 sec, just 2.5 percent different from the predicted value. This small difference is within the measurement error for the experiment. The observed steady-state sensitivity

based on the 20 percent change in τ was -263 volts/sec, which is 13 percent lower than the predicted value. The separate solution of the steady-state equations for $\tau = 0.20$ and 0.24 predicts the limit cycle amplitude in each case within 3 percent; so the 13 percent error above is attributable to the nonlinearity of the relation between τ and steady-state amplitude.

The description of limit cycle dynamics developed in this section has applied to the limit cycle as viewed at the input to the nonlinear part of the system, that is, at the point in the system which enjoys the full filtering effect of the linear part. If, however, the limit cycle is to be monitored at a point farther forward in the linear part, one would expect a somewhat different dynamic relation between the controlling parameter and the observed limit cycle amplitude. This may well be true, but experiment shows that as one observes at points farther forward, closer both to the nonlinearity and to the controlling parameter, the most dramatic change is that the whole nature of the amplitude response takes such a form as to defy description by a linear operator. The limit cycle amplitude response to a step change in forward gain may, for example, become quite sensitive to the phase angle in the limit cycle at which the step occurs, totally different responses resulting from a gain change at the zero crossing and at the peak of the cycle. From the point of view of the designer of an amplitude-regulating loop, this complication is of little consequence. The only change anticipated from the results derived here is the possibility of *less lag* in the amplitude dynamics for variables other than the input to the nonlinear part of the system. The results of this section then give a conservative estimate of the lag due to limit cycle dynamics, and may still be used to permit an analytical approach to the design of the system. That complicated phenomena such as those described above should occur in nonlinear systems is no surprise; the wonder is that the simple picture developed in this section does so well in characterizing the complex process of variations in the limit cycles of nonlinear systems.

We close this section with the observation that if the frequency, as well as the amplitude, of a limit cycle is to be monitored and perhaps controlled, the dynamics of limit cycle frequency changes are identical with the dynamics of amplitude changes, since the amplitude and frequency of a changing limit cycle are inextricably bound together as parts of a single transient oscillation.

4.5 LIMIT CYCLE STABILITY

We now return briefly to the matter of the stability of limit cycles, which was discussed earlier, in Sec. 3.2. At that point we had developed analytic tools for the description of steady-state oscillations only, and thus the stability of limit cycles was necessarily treated from a quasi-static point of view. That is to say, the truly dynamic properties of limit cycle transients which involve

derivatives of σ and ω were not included in the description. It is clear that the "right" way to determine the stability of an indicated steady-state oscillation is to determine the stability of the *transient* oscillation which results when the steady-state oscillation is perturbed; to do so requires the analytic machinery of this chapter. However, in each of the examples of transient oscillations given in the preceding sections, the quasi-static description of the transient was found to be in error by 25 or 30 percent at most, and one should expect it to be a rare case in which the quasi-static description of a transient is so grossly in error as to give a false indication of stability. It is fortunately true that the techniques of Sec. 3.2 almost always give a correct indication of stability; one has to search to find examples in which they fail.

A more complete description of the stability of limit cycles has already been developed in the characterization of limit cycle dynamics given in the preceding section. There a transfer function of first-order form was derived which described the dynamic properties of the response of a limit cycle to a change in a parameter of the system. The time constant of this first-order transfer function was thought of as measuring the speed of response of the transient. As a by-product, it also indicates the stability of the limit cycle transient, and thus of the limit cycle. If the sign of τ given by Eq. (4.4-18) is positive, the limit cycle is stable; if τ is negative, the limit cycle is unstable. This is the familiar stability criterion for an ordinary first-order linear system; it applies equally well to the first-order linear approximation which was developed for limit cycle dynamics. From another point of view, note that a positive τ corresponds to a case in which the slope of $\dot{A}$ versus A is negative in the vicinity of the steady-state limit cycle where $\dot{A} = 0$. Thus if A is greater than A_0 , $\dot{A}$ is negative and A decreases toward the steady-state value. Similarly, if A is less than A_0 , $\dot{A}$ is positive and A increases toward A_0 . On the other hand, if τ were negative, the slope of $\dot{A}$ versus A would be positive in the vicinity of $A = A_0$, $\dot{A} = 0$, and the same reasoning indicates a divergence of A away from A_0 following any disturbance.

It should be emphasized that although the sign of τ is a more complete indicator of limit cycle stability than any of the quasi-static methods, it still involves certain approximations. The original differential equations for A and ω derived in Sec. 4.1 incorporate the approximation of the nonlinear characteristic by its DF; in other respects these equations are exact. The small- σ solution of these equations is an approximate solution which ignores derivatives of σ and derivatives higher than the first of ω . This solution is further approximated in the expression for τ [Eq. (4.4-18)] by linearizing around the steady-state limit cycle point. The net difference between this stability criterion and any of the quasi-static methods given in Sec. 3.2 is then the inclusion of the effect of $\dot{\omega}$ on the dynamic properties of the system. The importance of this term to the *accuracy* of the description of transient oscillations was demonstrated in the examples given in the preceding

sections. That the effect of ω on the determination of the *stability* of a limit cycle can also be important is illustrated in the following example.

Example 4.5-1 Take as an example the system of Fig. 3.1-3. The open-loop linear part of this system is

$$L(s) = \frac{e^{-sT}}{s} \quad (4.5-1)$$

and the nonlinear part is

$$y = \delta x - \epsilon x^3 \quad (4.5-2)$$

or which the DF is

$$N(A) = \delta - \frac{3}{4}\epsilon A^2 \quad (4.5-3)$$

The form of these functions on a polar plot is shown in Fig. 4.5-1. The negative reciprocal DF for the nonlinear part of the system starts, for $A = 0$, at the point $-1/\delta$ and proceeds to the left for $A > 0$. It approaches minus infinity and changes sign at $A = \sqrt{4\delta/3\epsilon}$.

From Fig. 4.5-1 we observe a multiplicity of possible limit cycles, depending on the magnitude of δ (δ is shown as positive). The limit cycle at the point numbered 1 on the figure is possible if $-1/\delta$ lies inside the point where $L(j\omega)$ first crosses the negative real axis; that is,

$$\text{Limit cycle at 1 possible if } \delta > \frac{\pi}{2T} \quad (4.5-4)$$

Similarly,

$$\text{Limit cycle at 2 possible if } \delta > 2.5 \frac{\pi}{T} \quad (4.5-5)$$

and so on for higher-frequency limit cycles.

The stability of these limit cycles as determined by any quasi-static method can readily be seen from the graphic approach. If the point $-1/[N(A)]$ is considered an equivalent -1 point for the purpose of the Nyquist stability criterion, we see that a limit cycle at the point numbered 1, if perturbed by increasing its amplitude, indicates a stable system; so

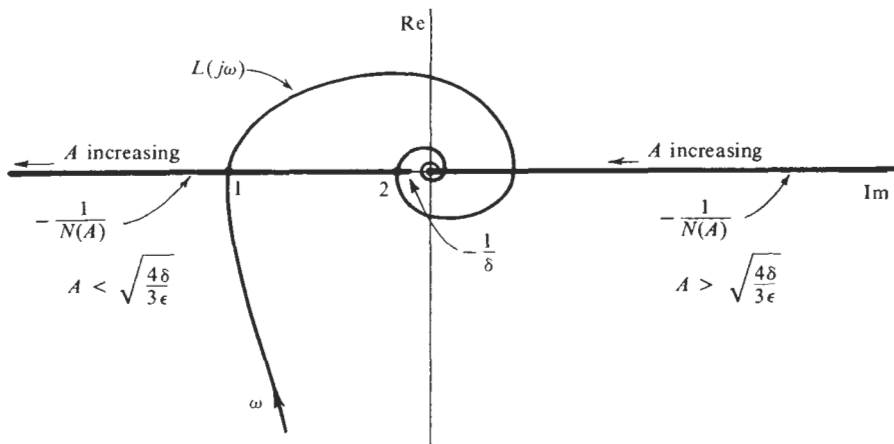


Figure 4.5-1 Polar plot of system of Example 4.5-1.

A would tend to decrease again. Similarly, if it is perturbed by decreasing its amplitude, it indicates an unstable system; so A would tend to increase again. These are the conditions for a stable limit cycle. Thus a quasi-static view of stability predicts that if δ is large enough to support the limit cycle mode numbered 1, that limit cycle will be stable no matter how large δ is. Experimental results show, however, that there is only a rather narrow band of values for δ for which this mode is possible and stable; this mode is actually unstable for all values of δ greater than some limit value—a behavioral pattern which a quasi-static description of stability fails to predict. The same argument shows the limit cycle mode numbered 2 and all higher-frequency limit cycles which may occur to be unstable.

We now test the stability of these limit cycles using the criterion of this section, the sign of the time constant of the transfer function for limit cycle dynamics. Defining $x(t)$, as usual, to be the input to the nonlinear part of the system, the differential equation which this closed-loop system obeys (see Fig. 3.1-3) is

$$\dot{x}(t + T) = -\delta x(t) + \epsilon x^3(t) \quad (4.5-6)$$

The substitution rules of Sec. 4.1 cannot be employed directly because of the time delay element in this system; we must address separately the matter of expressing $\dot{x}(t + T)$ as some function times $x(t)$, so that $x(t)$ can be divided out of each term of the equation as before. The form of the solution assumed is

$$x(t) = A(t)e^{j\psi(t)} \quad (4.5-7)$$

so

$$\dot{x}(t) = \dot{A}(t)e^{j\psi(t)} + j\omega(t)A(t)e^{j\psi(t)} \quad (4.5-8)$$

The purpose in using this form is that it expresses the solution in terms of $A(t)$ and $\omega(t)$, which we expect to be slowly varying functions. One may then expect that Taylor series expansions for $A(t + T)$ and $\omega(t + T)$ about the time t will be useful for the expression of $\dot{x}(t + T)$ in terms of variables evaluated at the time t . In fact, in the small- σ solution to be used, all derivatives of σ are ignored (corresponding to derivatives of A greater than the first), and derivatives of ω greater than the first are dropped. To be consistent with this, we shall use Taylor series expansions for $A(t + T)$ and $\omega(t + T)$ which retain only the first-derivative terms. Note that these expansions are used only for the purpose of expressing $\dot{x}(t + T)$ in a convenient form.

$$\dot{A}(t + T) \cong \dot{A}(t) \quad (4.5-9)$$

$$A(t + T) \cong A(t) + \dot{A}(t)T \quad (4.5-10)$$

$$\omega(t + T) \cong \omega(t) + \dot{\omega}(t)T \quad (4.5-11)$$

$$\psi(t + T) \cong \psi(t) + \omega(t)T + \frac{1}{2}\dot{\omega}(t)T^2 \quad (4.5-12)$$

Thus

$$\dot{x}(t + T) \cong [\dot{A} + j(\omega + \dot{\omega}T)(A + \dot{A}T)]e^{j(\psi + \omega T + \frac{1}{2}\dot{\omega}T^2)} \quad (4.5-13)$$

Each of the variables in the right-hand member is evaluated at the time t , and the functional dependence on t is omitted from the notation for the sake of brevity. Expanding and factoring out $Ae^{j\psi} = x(t)$ gives

$$\dot{x}(t + T) \cong [\sigma + j(\omega + \dot{\omega}T + \sigma\omega T + \sigma\dot{\omega}T^2)]x(t)e^{j(\omega T + \frac{1}{2}\dot{\omega}T^2)} \quad (4.5-14)$$

Using this expression and the DF for the x^3 nonlinear term [$N(A) = \frac{3}{4}A^2$] in Eq. (4.5-6) and dividing $x(t)$ out of each term gives

$$[\sigma + j(\omega + \dot{\omega}T + \sigma\omega T + \sigma\dot{\omega}T^2)]e^{j(\omega T + \frac{1}{2}\dot{\omega}T^2)} + \delta - \frac{3}{4}\epsilon A^2 = 0 \quad (4.5-15)$$

which has real and imaginary parts

$$R = \sigma \cos(\omega T + \frac{1}{2}\dot{\omega}T^2) - (\omega + \dot{\omega}T + \sigma\omega T + \sigma\dot{\omega}T^2) \sin(\omega T + \frac{1}{2}\dot{\omega}T^2) + \delta - \frac{3}{2}\epsilon A^2 = 0 \quad (4.5-16a)$$

$$I = \sigma \sin(\omega T + \frac{1}{2}\dot{\omega}T^2) + (\omega + \dot{\omega}T + \sigma\omega T + \sigma\dot{\omega}T^2) \cos(\omega T + \frac{1}{2}\dot{\omega}T^2) = 0 \quad (4.5-16b)$$

The limit cycle modes are identified as the static or steady-state solutions of these equations. Using the static conditions $\sigma = \dot{\omega} = 0$, the solution is

$$\omega_0 = (m + \frac{1}{2}) \frac{\pi}{T} \quad (4.5-17)$$

$$A_0 = \sqrt{\frac{4}{3\epsilon} \left[\delta - (-1)^m (m + \frac{1}{2}) \frac{\pi}{T} \right]} \quad (4.5-18)$$

for m any nonnegative integer. The condition for a real oscillation is

$$\delta > (-1)^m (m + \frac{1}{2}) \frac{\pi}{T} \quad (4.5-19)$$

All these steady-state results are of course identical with those derived earlier; the case $m = 0$ in Eq. (4.5-19) corresponds to Eq. (4.5-4), and $m = 2$ to Eq. (4.5-5).

The stability of these limit cycles is determined by the sign of the time constant of the limit cycle dynamics transfer function [Eq. (4.4-18)]. Toward the evaluation of this time constant, the required partial derivatives of the real and imaginary functions [Eqs. (4.5-16)] are calculated and evaluated at the limit cycle conditions $\sigma = \dot{\omega} = 0$, with ω given by Eq. (4.5-17) and A by Eq. (4.5-18). These derivatives are

$$\begin{aligned} \frac{\partial R}{\partial A} &= -\frac{3}{2}\epsilon A_0 & \frac{\partial I}{\partial A} &= 0 \\ \frac{\partial R}{\partial \sigma} &= -(-1)^m (m + \frac{1}{2})\pi & \frac{\partial I}{\partial \sigma} &= (-1)^m \\ \frac{\partial R}{\partial \omega} &= -(-1)^m & \frac{\partial I}{\partial \omega} &= -(-1)^m (m + \frac{1}{2})\pi \\ \frac{\partial R}{\partial \dot{\omega}} &= -(-1)^m T & \frac{\partial I}{\partial \dot{\omega}} &= -(-1)^m \frac{1}{2} (m + \frac{1}{2})\pi T \end{aligned}$$

which, when used in Eq. (4.4-18), determine the time constant to be

$$\tau = \frac{(-1)^m \{ 1/[(m + \frac{1}{2})\pi] + 2(m + \frac{1}{2})\pi \} - T\delta}{\frac{3}{2}\epsilon A_0^2} \quad (4.5-20)$$

This time constant is positive, and thus the limit cycle is indicated to be stable, if

$$\delta < \frac{(-1)^m}{T} \left[2(m + \frac{1}{2})\pi + \frac{1}{(m + \frac{1}{2})\pi} \right] \quad (4.5-21)$$

The condition for a real oscillation to exist was found to be

$$\delta > \frac{(-1)^m}{T} (m + \frac{1}{2})\pi \quad (4.5-22)$$

We are considering δ to be a positive quantity and are only interested in the positive frequency solutions; thus m is restricted to nonnegative even-integer values. For every such value of m there is a bounded range of δ for which a limit cycle is possible and stable, according to Eqs. (4.5-21) and (4.5-22):

$$\text{For } m = 0: \quad \frac{1.57}{T} < \delta < \frac{3.78}{T} \quad (4.5-23)$$

$$\text{For } m = 2: \quad \frac{7.85}{T} < \delta < \frac{15.83}{T} \quad (4.5-24)$$

$$\text{For } m = 4: \quad \frac{14.13}{T} < \delta < \frac{28.33}{T} \quad (4.5-25)$$

and so on for the higher-frequency limit cycles.

In this particular example, then, this dynamic stability criterion indicates an entirely different pattern of behavior for the system than do the quasi-static criteria. Computer experimentation with this system has demonstrated the behavior predicted above for the $m = 0$ limit cycle. The lower limit of δ for a limit cycle to exist is quite accurately predicted; the upper limit for the cycle to be stable was not determined accurately, but was found to lie between 4 and 5. The system is sensitive in that the initial amplitude must be taken rather close to the steady-state value if the system is to fall into the limit cycle mode. The higher-frequency limit cycle modes are not observed in the actual system. Multiple-input describing function theory shows clearly the reason for this. Although higher-frequency limit cycle modes are stable for limited ranges of δ , the system in the presence of these modes is unstable for any arbitrary small perturbation in addition to the limit cycle. Thus any additional perturbation diverges and prevents the continuation of the limit cycle. The analysis of this phenomenon requires the dual-input describing function of Chap. 6.

In this section we have presented a criterion for limit cycle stability which is more complete than those of Sec. 3.2 in that one more aspect of the dynamic character of a limit cycle transient, the effect of $\dot{\omega}$, is included. The fact that significantly new information is incorporated in this formulation is demonstrated by the example. It is not surprising that the more complete stability criterion is more laborious to apply than the simpler ones. Fortunately, the quasi-static indicators of stability are accurate in the vast majority of cases which present themselves in practice.

REFERENCES

1. Clauser, F. H.: The Transient Behavior of Non-linear Systems, *IRE Trans. Circuit Theory*, vol. CT-7, no. 4 (December, 1960), pp. 446-458.
2. Freeman, E. A.: Characterization of Nonlinearity for Transient Processes: Its Evaluation and Application, *IEEE Trans. Autom. Control*, vol. AC-12, no. 5 (October, 1967), pp. 491-501.
3. Gelb, A.: The Analysis and Design of Limit Cycling Adaptive Automatic Control Systems, Sc.D. thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Cambridge, Mass., 1961.

4. Gelb, A., and W. E. Vander Velde: On Limit Cycling Control Systems, *IEEE Trans. Autom. Control*, vol. AC-8, no. 2 (April, 1963), pp. 142-157.
5. Grensted, P. E. W.: Analysis of the Transient Response of Nonlinear Control Systems, *Trans. ASME*, vol. 80, no. 2 (February, 1958), pp. 427-432.
6. Lubbock, J. K., and H. A. Barker: Complex Frequency Response Diagrams and Their Use in the Design of Feedback Control Systems, Part 12, *Process Control and Automation*, vol. 9, no. 12 (December, 1962), pp. 563-570.
7. Popov, E. P.: Approximate Study of Transients in Nonlinear Automatic Control Systems by Harmonic Linearization, *Bull. Acad. Sci. U.S.S.R., Tech. Sci. Sec.* (1956), no. 9, pp. 3-23; no. 12, pp. 30-47.
8. Thaler, G. J., and M. P. Pastel: "Analysis and Design of Nonlinear Feedback Control Systems," McGraw-Hill Book Company, New York, 1962.
9. Truxal, J. G.: "Automatic Feedback Control System Synthesis," McGraw-Hill Book Company, New York, 1955.
10. Voronov, A. A.: An Approximate Determination of the Self-oscillation Startup Process in Certain Nonlinear Automatic Control Systems, *Automation and Remote Control*, vol. 18, no. 7 (June, 1958), pp. 683-694.

PROBLEMS

- 4-1. Derive the replacement formulas, Eqs. (4.1-16) and (4.1-17), using the relation ($p = d/dt$)

$$p^n(xy) = \sum_{k=0}^n \binom{n}{k} (p^{n-k}x)(p^ky)$$

Start with

$$f_{n+1} = \frac{1}{x} (p^{n+1}x)$$

$$= \frac{1}{x} p^n(sx)$$

and

$$g_{n+1} = \frac{1}{x} p^{n+1}(Nx)$$

$$= \frac{1}{x} \{p^n[(pN)x + s(Nx)]\}$$

- 4-2. Suppose the solution for $x(t) = A(t) \exp [j\psi(t)]$ has been determined and we wish to observe $c(t)$. Using Eq. (4.1-23), with

$$L_1(s)L_2(s) = \frac{g_0}{d_0 + d_1s}$$

write $c(t)$ in the form of a transient oscillation

$$c(t) = B(t) \exp [j\theta(t)]$$

and express $B(t)$ and $\theta(t)$ in terms of known functions. Note that the instantaneous amplitude and frequency of the transient oscillation at $c(t)$ are both different from those at $x(t)$.

- 4-3. Use the piecewise-sinusoidal-approximation technique described in Sec. 4.2 to calculate approximately the amplitude history for the limit cycle build-up of the system of Fig. 4.3-3. A root-locus plot is most convenient for determining the exponential damping factor corresponding to any amplitude of oscillation. Compare the result with Fig. 4.3-5.
- 4-4. The system of Fig. 4-1 starts from the initial condition $c(0)/KD = 100$, $\dot{c}(0) = 0$. Plot $A(t)/KD$ and $\omega(t)$, where $A(t)$ is the amplitude and $\omega(t)$ the frequency of the transient oscillation which $c(t)$ experiences.

Do both the small- σ and the quasi-static solutions. Discuss the range of applicability of these solutions. Can you improve either of them?

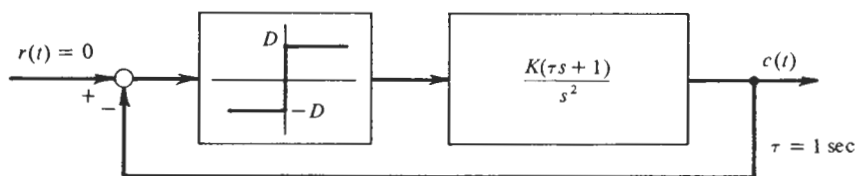


Figure 4-1

- 4-5. Find the small- σ solution to the differential equation

$$\frac{d^2 y}{dt^2} + \alpha y^2 \frac{dy}{dt} + \omega_0^2 y = 0$$

You should be able to write out explicit expressions for $A(t)$ and $\omega(t)$ in terms of arbitrary initial conditions.

- 4-6. The system of Fig. 4-2 is turned on with an initial error of 50 units. Find the time history of the amplitude of the resulting transient oscillation as it decays from the initial 50 units to 5 units, after which the system operates in the linear region. What is the initial frequency of this oscillation and the frequency when the amplitude has decayed to 5 units? For simplicity in the solution for $\sigma(A)$, drop powers of σ greater than 1.

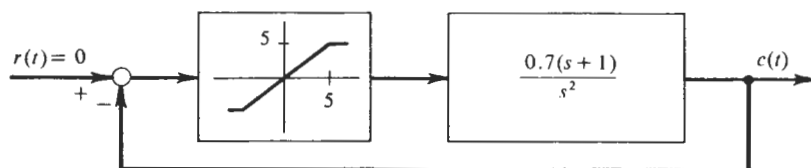


Figure 4-2

- 4-7. Nonspinning vehicles entering the earth's atmosphere at high speed exhibit oscillations in which the angle-of-attack history obeys approximately the following differential equation:

$$\ddot{\alpha} + M(\alpha)e^{at} = 0$$

$M(\alpha)$ is the nonlinear relation between angle of attack and restoring aerodynamic moment for a given dynamic pressure. The exponential term represents the increasing dynamic pressure as the vehicle descends into denser atmosphere.

$$M(\alpha) = 0.1\alpha - 0.02\alpha^3 \quad \alpha = 0.4 \text{ sec}^{-1}$$

Suppose the angle of attack at $t = 0$ is 1 radian. Assume the $\alpha(t)$ history to be in the form of a transient oscillation, and solve approximately for the amplitude and frequency of the oscillation. How long does it take for the amplitude to converge to 0.2 radian? What is the initial frequency and the frequency when the amplitude is 0.2 radian?

- 4-8. Derive the equations (the equations of reals and imaginaries) which the parameters of a transient oscillation in the system of Fig. 4-3 must obey. You may relate any delayed time function which appears in the problem to the present time function by assuming $\dot{A}$ and $\dot{\omega}$ are constant over the delay interval.

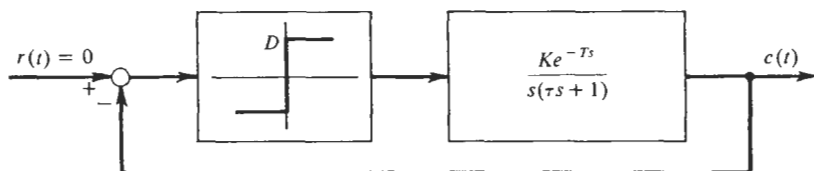


Figure 4-3

- 4-9. The system of Fig. 4-4 is in a steady-state limit cycle with $KD = 10$ units/sec. The forward gain is suddenly increased to $KD = 20$ units/sec. Plot the history of the amplitude and frequency of the transient in the limit cycle at $c(t)$.

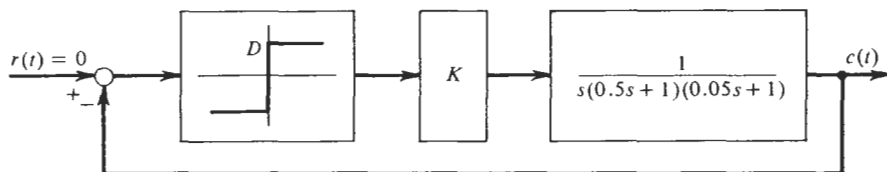


Figure 4-4

- 4-10. Use the direct-solution procedure [Eq. (4.4-18)] to find the time constant for limit cycle amplitude transients for the system of Fig. 4.3-3.

5 TWO-SINUSOID-INPUT DESCRIBING FUNCTION (TSIDF)

5.0 INTRODUCTION

Circumstances which lead to periodic but highly nonsinusoidal nonlinearity inputs render the DF approach invalid. In a number of such cases the nonlinearity input is well described as being composed of *two* additive sinusoids. These sinusoids result, in most cases of interest, from a sinusoidal input to the system, from system limit cycles, or both. The input to the nonlinearity is assumed, for the purpose of describing function calculation, to have the form

$$x(t) = A \sin(\omega_A t + \theta_A) + B \sin(\omega_B t + \theta_B) \quad (5.0-1)$$

in which the amplitudes A and B and frequencies ω_A and ω_B will be determined by the nature of the system and its inputs. From the point of view of the general describing function theory developed in Chap. 1, the random variables which characterize this input are the phase angles θ_A and θ_B .

When two or more sinusoidal components are assumed at the nonlinearity input, the statistical independence of these components, which if true

simplifies describing function calculation considerably, must be established with some care. Since each sinusoidal component is characterized by an amplitude and frequency which are deterministically fixed, and by a phase angle, the independence of the input components is established if the phase angles can properly be described as independent random variables. There are a number of important situations in which these phase angles are *not* independent. If one sinusoid were harmonically related to the other, the periods of the sinusoids would be commensurate, and a consistent phase relation would exist between them. The nature of the nonlinearity output would depend on this relative phase, and the quasi-linear approximator should reflect that dependence. In that case, with two sinusoidal inputs, only one phase angle could be treated as a random variable with a uniform distribution over one cycle; the other phase angle would be deterministically related to the first. With the statistical approach of Chap. 1, the coupled set of integral equations indicated in Eq. (1.5-13) would have to be solved to determine the optimum quasi-linear approximator for the nonlinearity. Alternatively, the interpretation of these describing functions as the amplitude and phase relation between each input component and the harmonic component of the same frequency in the nonlinearity output can be employed for describing function calculation. This latter point of view is expressed in the statements

$$N_A = \frac{\text{phasor representation of output component of frequency } \omega_A}{\text{phasor representation of input component of frequency } \omega_A} \quad (5.0-2)$$

$$N_B = \frac{\text{phasor representation of output component of frequency } \omega_B}{\text{phasor representation of input component of frequency } \omega_B} \quad (5.0-3)$$

where the subscript on N denotes the input component to which the describing function applies. These TSIDFs for the case of harmonically related inputs are in general functions of the amplitudes of both input components, the relative phase angle of these sinusoids, and the frequencies of each for a dynamic nonlinearity, the frequency ratio in the case of a static nonlinearity.

When one is studying the effect of a sub- or superharmonic of a sinusoidal signal, the frequencies of the two sinusoids under consideration are locked together into a harmonic relationship. In other cases, the frequencies of the two sinusoids are not locked into such a relation, but are determined, rather, by the nature of the system or its inputs. An example of this would be a limit cycling system responding to a sinusoidal input. One sinusoidal component at the nonlinearity input would have the frequency of the input to the system; the other would have the frequency of the system limit cycle. When the frequencies are determined by two unrelated mechanisms such as these, the periods of the two sinusoids may be thought of as incommensurate,

or the frequency ratio irrational, since the set of rational numbers is a set of zero measure. This means that if the frequencies are determined by unrelated causes, the event that they should be found to be rationally related is an event of zero probability. For sinusoids with incommensurate periods, the concept of a relative phase angle has no meaning, and one may properly think of the phase angles as given independently at random.

Situations in which the frequencies of the sinusoidal components at the nonlinearity input are not related harmonically thus satisfy the assumption of independence which led to the uncoupled expression for the optimum quasi-linear approximator [Eq. (1.5-21)] and its specialization to a sinusoidal input component in the presence of any other uncorrelated inputs [Eq. (1.5-36)], which is repeated here for convenience.

$$N_A = n_p + jn_q \quad (5.0-4a)$$

$$n_p = \frac{2}{A} \overline{y(0) \sin \theta_A} \quad (5.0-4b)$$

$$n_q = \frac{2}{A} \overline{y(0) \cos \theta_A} \quad (5.0-4c)$$

This is the describing function for the transfer of the sinusoidal component with amplitude A . The corresponding describing function for the transfer of the component with amplitude B is given by the same expression, with B replacing A . The indicated expectations in this case represent integrations over the distributions of the two random phase angles, θ_A and θ_B . Thus, for example, if the nonlinearity were static and single-valued, so $y(0)$ were given unambiguously in terms of $x(0)$, n_q would be zero, as was shown in Sec. 1.5, and the describing functions for the two input components would be written

$$N_A = \frac{1}{2\pi^2 A} \int_0^{2\pi} d\theta_A \int_0^{2\pi} d\theta_B y(A \sin \theta_A + B \sin \theta_B) \sin \theta_A \quad (5.0-5)$$

$$N_B = \frac{1}{2\pi^2 B} \int_0^{2\pi} d\theta_A \int_0^{2\pi} d\theta_B y(A \sin \theta_A + B \sin \theta_B) \sin \theta_B \quad (5.0-6)$$

The probability density functions for θ_A and θ_B have both been written as $1/2\pi$ over the interval $(0, 2\pi)$. The integrations could equally well be taken over any other interval which spans one cycle, such as $(-\pi, \pi)$. These TSIDFs can be seen to depend only on the amplitudes of the input components.

An important special case is that in which one of the sinusoids at the

input to the nonlinearity has an infinitesimal amplitude. The quasi-linear gain of the nonlinearity to such an infinitesimal sinusoid in the presence of another is called the *incremental-input describing function*. It is possible to express this describing function in terms of the DF for the nonlinearity.

In this chapter applications of the TSIDF and incremental-input describing function are made to various situations in which DF theory is inadequate. These include the two-sinusoid response of nonlinear systems, subharmonic oscillations, frequency response counterexamples, and multiple limit cycles; other applications are indicated. Jump resonance phenomena and transient oscillations are also briefly treated, despite the fact that we have already developed adequate means for their study via the DF. In the case of jump resonance phenomena a different and possibly more appealing point of view than that presented in Sec. 3.3 is shown to lead to precisely the same result obtained earlier. The study of transient oscillations is directed toward the determination of relative stability of steady-state oscillations. The procedure developed is simple to execute, and is comparable in accuracy with the quasi-static solution of Chap. 4. We begin with a treatment of TSIDF calculation.

5.1 TSIDF CALCULATION

Many different methods for TSIDF calculation are available. In general, they are all more difficult to execute than the various schemes for DF calculation treated in Chap. 2. Certainly, this is to be expected; the problem now under consideration is far more general. Nevertheless, once derived, application of the TSIDF proceeds in a remarkably simple way.

TSIDF CALCULATION BY DIRECT EXPANSION

This method of TSIDF calculation is applicable when representation of the nonlinearity is in terms of a low-order polynomial. The approach is therefore somewhat limited in scope. Nevertheless, this approach leads directly to several interesting results, some of which will be encountered in subsequent sections. Motivated by the work of West, Douce, and Livesley (Ref. 22), we proceed by example.

Example 5.1-1 Consider the cubic characteristic

$$y = x^3 \quad (5.1-1)$$

For an input of the form

$$x = A \cos \omega t + B \cos (\gamma \omega t + \theta) \quad (5.1-2)$$

one may determine output harmonic content by direct expansion, as follows:

$$\begin{aligned}
 y &= [A \cos \omega t + B \cos (\gamma \omega t + \theta)]^3 \\
 &= A^3 \cos^3 \omega t + 3A^2B \cos^2 \omega t \cos (\gamma \omega t + \theta) \\
 &\quad + 3AB^2 \cos \omega t \cos^2 (\gamma \omega t + \theta) + B^3 \cos^3 (\gamma \omega t + \theta) \\
 &= \left(\frac{3A^3}{4} + \frac{3AB^2}{2} \right) \cos \omega t + \frac{A^3}{4} \cos 3\omega t \\
 &\quad + \left(\frac{3A^2B}{2} + \frac{3B^3}{4} \right) \cos (\gamma \omega t + \theta) \\
 &\quad + \frac{3A^2B}{4} \{ \cos [(\gamma + 2)\omega t + \theta] + \cos [(\gamma - 2)\omega t + \theta] \} \\
 &\quad + \frac{3AB^2}{4} \{ \cos [(2\gamma + 1)\omega t + 2\theta] + \cos [(2\gamma - 1)\omega t + 2\theta] \} \\
 &\quad + \frac{B^3}{4} \cos (3\gamma \omega t + 3\theta)
 \end{aligned} \tag{5.1-3}$$

The last fully expanded equation is written in terms of only single-frequency components. Several distinct cases corresponding to different values of γ are of interest.

Case 1. Non-harmonically-related input sinusoids¹ In this event the only output terms of the same frequency as the input sinusoids are the first and third terms of Eq. (5.1-3). Terming $N_A(A, B)$ and $N_B(A, B)$ the TSIDFs for input frequencies ω and $\gamma\omega$, respectively, gives

$$\begin{aligned}
 N_A(A, B) &= \frac{3A^3/4 + 3AB^2/2}{A} \\
 &= \frac{3}{4}A^2 + \frac{3}{2}B^2
 \end{aligned} \tag{5.1-4}$$

and

$$\begin{aligned}
 N_B(A, B) &= \frac{3A^2B/2 + 3B^3/4}{B} \\
 &= \frac{3}{2}A^2 + \frac{3}{4}B^2
 \end{aligned} \tag{5.1-5}$$

The symmetry between these expressions, both of which are *real, independent of γ , and independent of θ* , is apparent.

Case 2. Harmonically related input sinusoids ($\gamma = \frac{1}{3}$) The output terms of interest in Eq. (5.1-3) are repeated for convenience:

$$\begin{aligned}
 y &= \left(\frac{3A^3}{4} + \frac{3AB^2}{2} \right) \cos \omega t + \left(\frac{3A^2B}{2} + \frac{3B^3}{4} \right) \cos \left(\frac{1}{3}\omega t + \theta \right) \\
 &\quad + \frac{3AB^2}{4} \cos \left(-\frac{1}{3}\omega t + 2\theta \right) + \frac{B^3}{4} \cos (\omega t + 3\theta) + \text{terms at other frequencies}
 \end{aligned} \tag{5.1-6}$$

¹ Cases where γ is irrational (e.g., not expressible as the ratio of two integers). Results obtained for the cubic nonlinearity also hold for harmonic frequency ratios (e.g., rational γ), provided $\gamma \neq \frac{1}{3}, 1, 3$.

As varied phase angles are associated with the terms to be considered, the phasor representation called for by Eqs. (5.0-2) and (5.0-3) is conveniently utilized, with the result that

$$\begin{aligned} N_A(A, B, \tfrac{1}{3}, \theta) &= \frac{(\tfrac{3}{4}A^3 + \tfrac{3}{2}AB^2)e^{j\omega t} + (\tfrac{1}{4}B^3)e^{j(\omega t + 3\theta)}}{Ae^{j\omega t}} \\ &= \tfrac{3}{4}A^2 + \tfrac{3}{2}B^2 + \tfrac{1}{4}(B^3/A)e^{j3\theta} \end{aligned} \quad (5.1-7a)$$

and

$$\begin{aligned} N_B(A, B, \tfrac{1}{3}, \theta) &= \frac{(\tfrac{3}{2}A^2B + \tfrac{3}{4}B^3)e^{j[(\omega/3)t + \theta]} + \tfrac{3}{4}AB^2e^{j[(\omega/3)t - 2\theta]}}{Be^{j[(\omega/3)t + \theta]}} \\ &= \tfrac{3}{2}A^2 + \tfrac{3}{4}B^2 + \tfrac{3}{4}ABe^{-j3\theta} \end{aligned} \quad (5.1-7b)$$

The TSIDFs are both *complex* and *dependent* upon θ .

Case 3. Harmonically related input sinusoids ($\gamma = 3$) This case is, essentially, identical with Case 2, with the roles of the higher- and the lower-frequency input terms reversed. It warrants no further discussion.

TSIDF CALCULATION BY DOUBLE FOURIER SERIES EXPANSION

This calculation is based on a procedure described by Bennett (Ref. 2), and later by Kalb and Bennett (Ref. 14). Corresponding to the input

$$x(t) = A \sin \omega t + B \sin (\gamma \omega t + \theta) \quad (5.1-8)$$

the nonlinearity output is sought in the form

$$\begin{aligned} y(t) &= y(A \sin \psi_1 + B \sin \psi_2) \\ &= \sum_{m=0}^{\infty} \sum_{\substack{n=-\infty (m \neq 0) \\ n=0 (m=0)}}^{\infty} [P_{mn} \sin (m\psi_1 + n\psi_2) + Q_{mn} \cos (m\psi_1 + n\psi_2)] \end{aligned} \quad (5.1-9)$$

where $\psi_1 = \omega t$ and $\psi_2 = \gamma \omega t + \theta$

The Fourier coefficients P_{mn} and Q_{mn} are determined in the usual manner by multiplying both sides of Eq. (5.1-9) by the multiplier of the coefficient which is to be found, and integrating the result throughout the square bounded by $\psi_1 = \pm\pi$, $\psi_2 = \pm\pi$. This yields

$$P_{mn} = \frac{1}{2\pi^2} \iint_{-\pi}^{\pi} y(A \sin \psi_1 + B \sin \psi_2) \sin (m\psi_1 + n\psi_2) d\psi_1 d\psi_2 \quad (5.1-10)$$

and

$$Q_{mn} = \frac{1}{2\pi^2} \iint_{-\pi}^{\pi} y(A \sin \psi_1 + B \sin \psi_2) \cos (m\psi_1 + n\psi_2) d\psi_1 d\psi_2 \quad (5.1-11)$$

In the case of static single-valued nonlinearities, an important property of Q_{mn} can be demonstrated by rewriting Eq. (5.1-11) as four integrals covering the ranges $-\pi/2 < \psi_1 < 3\pi/2$ and $-\pi/2 < \psi_2 < 3\pi/2$.

$$\begin{aligned}
 2\pi^2 Q_{mn} &= \int_{-\pi/2}^{\pi/2} d\psi_2 \int_{-\pi/2}^{\pi/2} d\psi_1 [y(A \sin \psi_1 + B \sin \psi_2) \cos(m\psi_1 + n\psi_2)] \\
 &\quad + \int_{-\pi/2}^{\pi/2} d\psi_2 \int_{\pi/2}^{3\pi/2} d\psi_1 [] \\
 &\quad + \int_{\pi/2}^{3\pi/2} d\psi_2 \int_{-\pi/2}^{\pi/2} d\psi_1 [] + \int_{\pi/2}^{3\pi/2} d\psi_2 \int_{\pi/2}^{3\pi/2} d\psi_1 [] \\
 &= I_1 + I_2 + I_3 + I_4
 \end{aligned} \tag{5.1-12}$$

The integrand, unchanged from one integral to the next, has been denoted by brackets. In I_3 and I_4 , the last two terms in Eq. (5.1-12), we make the substitutions

$$\psi_1^* = \pi - \psi_1 \quad \text{and} \quad \psi_2^* = \pi - \psi_2$$

which result in the determination that

$$I_3 + I_4 = \cos(m+n)\pi (I_1 + I_2)$$

Hence Q_{mn} can be written as

$$Q_{mn} = \frac{1 + \cos(m+n)\pi}{2\pi^2} (I_1 + I_2)$$

whence it follows that

$$Q_{mn} = 0 \quad \text{for } m+n \text{ odd} \tag{5.1-13}$$

Beginning with Eq. (5.1-10) and pursuing a similarly constructed argument, it is also readily demonstrated that

$$P_{mn} = 0 \quad \text{for } m+n \text{ even}$$

If, in addition, the nonlinearity is odd, it can be further proved that *all* $Q_{mn} = 0$.

Let us return to Eq. (5.1-9) and note that, for γ irrational, the only output term of frequency ω occurs for $m = 1, n = 0$. Correspondingly, the only output term of frequency $\gamma\omega$ occurs for $m = 0, n = 1$. In each of these instances it follows from previous arguments that $Q_{mn} = 0$. Hence we arrive at the TSIDF expressions

$$\begin{aligned}
 N_A(A, B) &= \frac{P_{10}}{A} \\
 &= \frac{1}{2\pi^2 A} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} y(A \sin \psi_1 + B \sin \psi_2) \sin \psi_1 d\psi_1 d\psi_2 \tag{5.1-14}
 \end{aligned}$$

$$\text{and } N_B(A, B) = \frac{P_{01}}{B}$$

$$= \frac{1}{2\pi^2 B} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} y(A \sin \psi_1 + B \sin \psi_2) \sin \psi_2 d\psi_1 d\psi_2 \quad (5.1-15)$$

These TSIDFs are real numbers. They are independent of θ and γ , and dependent only upon A and B . These expressions are exactly those of Eqs. (5.0-5) and (5.0-6), which result from the unified theory of Chap. 1.

If γ is rational, more than one term of Eq. (5.1-9) yields the frequency of interest. Calculating $N_A(A, B)$, for example, we look for all terms for which

$$m\omega + n\gamma\omega = \omega$$

$$\text{or } n = \frac{1}{\gamma} (1 - m) \quad (5.1-16)$$

Thus, when $\gamma = \frac{1}{3}$, the pairs of values (m, n) which satisfy Eq. (5.1-16) are $(0, 3)$, $(1, 0)$, $(2, -3)$, $(3, -6)$, etc.; and for $\gamma = 3$, the appropriate pairs are $(1, 0)$, $(4, -1)$, $(7, -2)$, $(10, -3)$, etc. In each of these terms $Q_{mn} = 0$, and the component of $y(t)$ at frequency ω is computed from Eq. (5.1-9) as

$$\begin{aligned} y(t) &= \sum_m P_{mn} \sin \left\{ m\omega t + \left[\frac{1}{\gamma} (1 - m) \right] \gamma\omega t + \left[\frac{1}{\gamma} (1 - m) \right] \theta \right\} \\ &= \sum_m P_{mn} \sin \left[\omega t + \frac{1}{\gamma} (1 - m) \theta \right] \\ &= P_N \sin (\omega t + \theta_N) \end{aligned}$$

The TSIDF is therefore given by

$$N_A(A, B) = \frac{P_N(A, B, \gamma, \theta)}{A} e^{j\theta_N(A, B, \gamma, \theta)}$$

which is complex and dependent upon γ and θ , in addition to A and B .

Example 5.1-2 Consider the ideal-relay characteristic defined by

$$y(A \sin \psi_1 \pm B \sin \psi_2) = \begin{cases} D & \text{for } \sin \psi_1 \pm k \sin \psi_2 \geq 0 \\ -D & \text{for } \sin \psi_1 \pm k \sin \psi_2 < 0 \end{cases}$$

where $k = B/A$, and for the present development we adopt the convention that A corresponds to the larger sinusoid. This ensures a solution to the switching condition $\sin \psi_1 \pm k \sin \psi_2 = 0$. In order to calculate $N_A(A, B)$, the double-integral formulation of Eq. (5.1-14) is written as follows:

$$\begin{aligned} N_A(A, B) = \frac{1}{2\pi^2 A} & \left(\int_{-\pi}^0 d\psi_2 \int_{-\pi}^0 d\psi_1 [] + \int_{-\pi}^0 d\psi_2 \int_0^{\pi} d\psi_1 [] \right. \\ & \left. + \int_0^{\pi} d\psi_2 \int_{-\pi}^0 d\psi_1 [] + \int_0^{\pi} d\psi_2 \int_0^{\pi} d\psi_1 [] \right) \end{aligned}$$

where, as before, the integrand is denoted simply by brackets. Changes of variable are now effected in the first three integrals, to convert all the limits of integration to the range 0 to π . Using the property of odd characteristics, $y(x) = -y(-x)$, the result of this manipulation is expressible as

$$N_A(A, B) = \frac{1}{\pi^2 A} \int_0^\pi d\psi_2 \int_0^\pi d\psi_1 \sin \psi_1 [y(A \sin \psi_1 + B \sin \psi_2) + y(A \sin \psi_1 - B \sin \psi_2)]$$

The calculation attendant on this integral formulation can be somewhat simplified by recasting it into four double integrals, each of $\pi/2$ -radian range in the independent variables ψ_1 and ψ_2 . Operating on these integrals by appropriate variable substitutions so as to achieve a range of integration 0 to $\pi/2$ in ψ_1 and ψ_2 , and collecting the results, yields

$$N_A(A, B) = \frac{4}{\pi^2 A} \int_0^{\pi/2} d\psi_2 \int_0^{\pi/2} d\psi_1 \sin \psi_1 [y(A \sin \psi_1 + B \sin \psi_2) + y(A \sin \psi_1 - B \sin \psi_2)]$$

As this formulation has not yet been specialized other than to odd memoryless characteristics, observe that it is generally applicable. We now proceed to the specific ideal-relay nonlinearity.

Over the entire range 0 to $\pi/2$ in each of the angles ψ_1 and ψ_2 , $y(A \sin \psi_1 + B \sin \psi_2) = D$. However, $y(A \sin \psi_1 - B \sin \psi_2)$ assumes both of the values $\pm D$, depending upon the sign of its argument. The angle which satisfies the switching condition $\sin \psi_1 - k \sin \psi_2 = 0$ is $\psi_1 = \sin^{-1}(k \sin \psi_2)$. Therefore we can write

$$\begin{aligned} N_A(A, B) &= \frac{4D}{\pi^2 A} \int_0^{\pi/2} d\psi_2 \left(\int_0^{\pi/2} \sin \psi_1 d\psi_1 \right. \\ &\quad \left. - \int_0^{\sin^{-1}(k \sin \psi_2)} \sin \psi_1 d\psi_1 + \int_{\sin^{-1}(k \sin \psi_2)}^{\pi/2} \sin \psi_1 d\psi_1 \right) \\ &= \frac{8D}{\pi^2 A} \int_0^{\pi/2} \cos [\sin^{-1}(k \sin \psi_2)] d\psi_2 \\ &= \frac{8D}{\pi^2 A} \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 \psi_2} d\psi_2 \\ &= \frac{8D}{\pi^2 A} E(k) \end{aligned} \quad (5.1-17a)$$

where $E(k)$ is the complete elliptic integral of the first kind. Denoting by $K(k)$ the complete elliptic integral of the second kind, it is readily proved that

$$N_B(A, B) = \frac{8D}{\pi^2 k^2 A} [E(k) - (1 - k^2)K(k)] \quad (5.1-17b)$$

$$\text{where} \quad E(k) = \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 \psi} d\psi \quad K(k) = \int_0^{\pi/2} \frac{d\psi}{\sqrt{1 - k^2 \sin^2 \psi}} \quad (5.1-18)$$

It is to be observed that N_A and N_B are indeed both real and independent of frequency. Expansion of these functions in powers of k yields

$$N_A(A, B) = \frac{4D}{\pi A} \left(1 - \frac{1}{4} k^2 - \frac{3}{64} k^4 - \frac{5}{256} k^6 - \dots \right) \quad (5.1-19a)$$

$$N_B(A, B) = \frac{2D}{\pi A} \left(1 + \frac{1}{8} k^2 + \frac{3}{64} k^4 + \frac{25}{1,024} k^6 + \dots \right) \quad (5.1-19b)$$

whence we deduce the useful relationship

$$\frac{N_B(A,B)}{N_A(A,B)} \approx \frac{1}{2} \quad \text{for } k < 1 \quad (5.1-20)$$

Calculation of the TSIDFs for harmonically related input sinusoids can be pursued as outlined earlier in the text or by single Fourier series expansion techniques. Rather than present details of the calculation here, we simply observe that the TSIDFs are now complex and dependent upon A , B , γ , and θ . It is of interest to note the *maximum* phase shift associated with $N_B(A,B,\gamma,\theta)$. The absolute value of this phase shift appears in Table 5.1-1 for various values of γ . It can be seen that, for a given k , the largest values occur at $\gamma = \frac{1}{3}$. For $k \leq 0.5$ and $\gamma \leq \frac{1}{3}$, this phase shift is less than 0.5° .

TABLE 5.1-1 MAXIMUM PHASE SHIFT IN $N_B(A,B,\gamma,\theta)$ FOR AN IDEAL RELAY†

γ	Maximum phase shift
$\frac{1}{3}$	$\tan^{-1} \frac{k}{4}$
$\frac{1}{2}$	$\tan^{-1} \frac{k}{8}$
$\frac{1}{5}$	$\tan^{-1} \frac{k}{64}$

† Amsler and Gorozdos, Ref. 1.

In the single-sinusoid-input case the nonlinearity output harmonics and the filtering properties of the loop linear elements determine whether DF analysis is valid. The familiar desired filtering property of the linear elements is amplitude attenuation with increasing frequency. In the two-sinusoid-input case a slightly different situation arises. This is a consequence of the fact that nonlinearity output terms can occur at frequencies *below* and *between*, as well as *above*, the input frequencies. In other words, because of the modulation process taking place, both sum and difference frequency terms are created. The sum frequency terms are necessarily at higher frequencies than either input sinusoid; hence for these terms the usual low-pass-filter requirement is necessary to validate TSIDF analysis. The difference frequency terms can be additionally troublesome, however, particularly if the loop linear elements have a resonance peak at one or more of these frequencies. It is certainly obvious at this point that to determine all these terms would be a monumental chore. In theory, all that is required is the evaluation of each term P_{mn} (and Q_{mn}) of interest. In practice, this determination is usually intractable.

Often the difference frequency terms are of quite small amplitude. This is particularly true when the frequency ratio γ is either very large or very

TABLE 5.1-2 OUTPUT OF AN IDEAL RELAY IN RESPONSE TO A TEST INPUT

Individual sinusoidal output components

Frequency	Amplitude
1	0.128
2	0.000
3	0.000
4	0.000
5	0.000
6	0.000
7	0.000
8	0.006
9	0.000
10	1.260
11	0.000
12	0.006

small. In support of this statement there is listed in Table 5.1-2 the output of an ideal relay (drive levels ± 1) in response to an input

$$x = 0.2 \sin t + \sin 10t$$

For all practical purposes there is *no* output harmonic distortion. At the frequencies $\omega = 1$ and $\omega = 10$, however, significant amplitudes do occur. These are very well predicted by Eqs. (5.1-19), the solutions derived for non-harmonically-related input sinusoids. That they also apply to the case of harmonically related input sinusoids will be explained in subsequent discussions of TSIDF approximation.

TSIDF CALCULATION BY INTEGRAL REPRESENTATION

This method has been employed by Gibson and Sridhar (Ref. 10) for the case of non-harmonically-related input sinusoids, using techniques of random-process theory. Below, we develop the TSIDFs for non-harmonically-related input sinusoids with the aid of completely deterministic arguments.

The output of an instantaneous nonlinear element can be written as a function of the input using the integral representation

$$y(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(ju) \exp(jux) du \quad (5.1-21)$$

where $Y(ju)$ is the Fourier transform of $y(x)$,

$$Y(ju) = \int_{-\infty}^{\infty} y(x) \exp(-jux) dx \quad (5.1-22)$$

The convergence of this integral can usually be enforced artificially, as is subsequently demonstrated by example. If $y(x)$ is an odd function of x (odd nonlinear characteristic), $Y(ju)$ is an odd function of u . In the following discussion only odd characteristics are considered. The nonlinearity input is taken in the form

$$x = A \sin \omega t + B \sin \gamma \omega t \quad (5.1-23)$$

where γ is *not* rational. From Eq. (5.1-21) the corresponding output is

$$\begin{aligned} y &= \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(ju) \exp(jAu \sin \omega t + jBu \sin \gamma \omega t) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(ju) \exp(jAu \sin \omega t) \exp(jBu \sin \gamma \omega t) du \end{aligned} \quad (5.1-24)$$

A key point in this development, due to Rice (Ref. 20), is the use of an expansion of the exponentials in series of the form (Ref. 5)

$$\begin{aligned} \exp(j\alpha \sin \beta) &= \cos(\alpha \sin \beta) + j \sin(\alpha \sin \beta) \\ &= \sum_{n=0}^{\infty} \epsilon_n J_{2n}(\alpha) \cos 2n\beta + 2j \sum_{n=0}^{\infty} J_{2n+1}(\alpha) \sin(2n+1)\beta \end{aligned} \quad (5.1-25)$$

where ϵ_n denotes the Neumann factor

$$\epsilon_n = \begin{cases} 1 & \text{for } n = 0 \\ 2 & \text{for } n = 1, 2, 3, \dots \end{cases}$$

and $J_\nu(\alpha)$ is the Bessel function of order ν and argument α . Substituting for the exponential terms in Eq. (5.1-24) gives

$$\begin{aligned} &\exp(jAu \sin \omega t) \exp(jBu \sin \gamma \omega t) \\ &= \left[\sum_{m=0}^{\infty} \epsilon_m J_{2m}(Au) \cos 2m\omega t + 2j \sum_{m=0}^{\infty} J_{2m+1}(Au) \sin(2m+1)\omega t \right] \\ &\quad \times \left[\sum_{n=0}^{\infty} \epsilon_n J_{2n}(Bu) \cos 2n\gamma \omega t + 2j \sum_{n=0}^{\infty} J_{2n+1}(Bu) \sin(2n+1)\gamma \omega t \right] \end{aligned} \quad (5.1-26)$$

Since $Y(ju)$ is an odd function of u , only the odd terms in Eq. (5.1-26) contribute to the value of the integral in Eq. (5.1-24). All even functions of u in Eq. (5.1-26), when multiplied by $Y(ju)$, result in integrands in Eq. (5.1-24) which are odd functions of u ; the value of the integral over the doubly infinite range is then zero. Using the fact that Bessel functions of odd

order are odd, and those of even order are even, yields that the portion of $\exp(jAu \sin \omega t) \exp(jBu \sin \gamma \omega t)$ displaying an odd dependence upon u is

$$2j \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \epsilon_m J_{2m}(Au) J_{2n+1}(Bu) \cos 2m\omega t \sin (2n+1)\gamma\omega t \\ + 2j \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \epsilon_n J_{2n}(Bu) J_{2m-1}(Au) \cos 2n\gamma\omega t \sin (2m+1)\omega t \quad (5.1-27)$$

Employing the trigonometric identity

$$\cos \alpha \sin \beta = \frac{1}{2} [\sin (\alpha + \beta) - \sin (\alpha - \beta)]$$

enables rewriting Eq. (5.1-27) in the form

$$j \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \epsilon_m J_{2m}(Au) J_{2n+1}(Bu) [\sin (2m + 2n\gamma + \gamma)\omega t - \sin (2m - 2n\gamma - \gamma)\omega t] \\ + j \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \epsilon_n J_{2n}(Bu) J_{2m+1}(Au) \\ \times [\sin (2m + 2n\gamma + 1)\omega t - \sin (2n\gamma - 2m - 1)\omega t] \quad (5.1-28)$$

Suppose that we seek the first-harmonic gain of the nonlinearity. Then the only terms in Eq. (5.1-28) of interest are those involving $\sin \omega t$ or $\sin (-\omega t)$. Considering γ irrational, such terms arise (over the allowable values of the indices m and n) only for the combination $m = 0, n = 0$. Had the nonlinearity gain to the input component of frequency $\gamma\omega$ been sought, again only $m = 0, n = 0$ would be pertinent. For these cases, therefore, the only terms of interest in Eq. (5.1-28) are

$$2jJ_0(Bu)J_1(Au) \sin \omega t + 2jJ_0(Au)J_1(Bu) \sin \gamma\omega t \quad (5.1-29)$$

Finally, from Eqs. (5.1-24) and (5.1-29), the *frequency-independent* TSIDF for the fundamental output term is given by

$$N_A(A, B) = \frac{j}{\pi A} \int_{-\infty}^{\infty} Y(ju) J_0(Bu) J_1(Au) du \quad (5.1-30)$$

Similarly,

$$N_B(A, B) = \frac{j}{\pi B} \int_{-\infty}^{\infty} Y(ju) J_0(Au) J_1(Bu) du \quad (5.1-31)$$

We may conclude that the frequency-independent TSIDF expressions for each input component are *functionally identical*. That is,

$$N_A(A, B) = N_B(B, A) \quad (5.1-32)$$

The only distinction between the two inputs in this case is their relative amplitudes! This conclusion is equally evident from Eqs. (5.1-14) and (5.1-15).

The foregoing procedure can be extended to the case of harmonically related input sinusoids simply by the inclusion of a phase-shift variable in one of the input components and the consideration of rational values of γ (Ref. 18).

Example 5.1-3 As an example of the application of this method, consider the *ideal-relay characteristic*, defined by

$$y(x) = \begin{cases} -De^{\sigma x} & x < 0 \\ 0 & x = 0 \\ De^{-\sigma x} & x > 0 \end{cases}$$

For $\sigma = 0$ this characterization reduces to that for the ideal relay. However, for σ positive, albeit infinitesimal, $y(x)$ as defined above becomes Fourier-transformable. Thus, from Eq. (5.1-22),

$$\begin{aligned} Y(ju) &= \lim_{\sigma \rightarrow 0} \left(\int_{-\infty}^{0-} -De^{\sigma x} e^{-jux} dx + \int_{0-}^{0+} 0 dx + \int_{0+}^{\infty} De^{-\sigma x} e^{-jux} dx \right) \\ &= \lim_{\sigma \rightarrow 0} \left(\frac{1}{j} \frac{2Du}{\sigma^2 + u^2} \right) \\ &= \frac{2D}{ju} \end{aligned} \quad (5.1-33)$$

Equations (5.1-31) and (5.1-33) are evidence that $N_B(A, B)$ is given by

$$N_B(A, B) = \frac{2D}{\pi B} \int_{-\infty}^{\infty} \frac{J_0(Au) J_1(Bu)}{u} du \quad (5.1-34)$$

Evaluation of this integral yields

$$N_B(A, B) = \begin{cases} \frac{8D}{\pi^2 B} E\left(\frac{A}{B}\right) & \text{for } \frac{B}{A} > 1 \\ \frac{8D}{\pi^2 B} \frac{A}{B} \left[E\left(\frac{B}{A}\right) - \left(1 - \frac{B^2}{A^2}\right) K\left(\frac{B}{A}\right) \right] & \text{for } \frac{B}{A} < 1 \end{cases} \quad (5.1-35)$$

where the complete elliptic integrals E and K are as defined previously.

The reader may be interested to compare this result with that obtained via double Fourier series expansion. Using the definition $k = B/A$, both results are seen to be identical.

Generally speaking, the transform approach to TSIDF calculation is quite useful. The transformed functions $Y(ju)$ are readily obtained. The major difficulty with this formulation is in evaluating the inversion integrals of Eqs. (5.1-30) and (5.1-31). Often, this can be accomplished in closed form in terms of known functions (gamma functions, hypergeometric functions, and elliptic integrals).

POWER-SERIES EXPANSION

A particularly simple TSIDF calculation mechanism for odd nonlinearities with non-harmonically-related input sinusoids results from the development of $N_A(A, B)$ and $N_B(A, B)$ [Eqs. (5.1-30) and (5.1-31)] in power series. The major steps in this development are briefly sketched in the ensuing presentation.

Let A denote the larger input sinusoid amplitude, and B denote the smaller. Then, provided that the necessary derivatives exist, $N_A(A, B)$ can be expanded in a Taylor series about $B = 0$.

$$N_A(A, B) = \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{\partial^n N_A(A, B)}{\partial B^n} \right|_{B=0} B^n \quad (5.1-37)$$

The derivatives are computed in the form

$$\frac{\partial^n N_A(A, B)}{\partial B^n} = \frac{j}{\pi A} \int_{-\infty}^{\infty} Y(ju) \frac{\partial^n J_0(Bu)}{\partial B^n} J_1(Au) du \quad (5.1-38)$$

Expressing $J_0(Bu)$ in its power series,

$$J_0(Bu) = \sum_{k=0}^{\infty} \frac{(-1)^k (Bu)^{2k}}{2^{2k} (k!)^2} \quad (5.1-39)$$

one can readily demonstrate that

$$\left. \frac{\partial^n J_0(Bu)}{\partial B^n} \right|_{B=0} = \begin{cases} 0 & \text{for } n \text{ odd} \\ \frac{(-1)^{n/2} n! u^n}{2^n [(n/2)!]^2} & \text{for } n \text{ even} \end{cases} \quad (5.1-40a)$$

$$(5.1-40b)$$

Thus Eq. (5.1-37) can be rewritten as

$$N_A(A, B) = \frac{j}{\pi A} \sum_{p=0}^{\infty} \frac{(-1)^p B^{2p}}{2^{2p} (p!)^2} \int_{-\infty}^{\infty} Y(ju) J_1(Au) u^{2p} du \quad (5.1-41)$$

The dummy variable n has been replaced by p , where $n = 2p$. The integral in this formulation has a surprisingly simple interpretation. This can be seen by first noting that $N_A(A, 0)$, the leading term in the series, is the single-input-sinusoid DF, $N(A)$.

$$\begin{aligned} N(A) &= \lim_{B \rightarrow 0} N_A(A, B) \\ &= \frac{j}{\pi A} \int_{-\infty}^{\infty} Y(ju) J_1(Au) du \end{aligned} \quad (5.1-42)$$

Use has been made of the fact that $J_0(0) = 1$. Repeated differentiation of this equation with respect to A yields relationships which can be so combined that the integral in Eq. (5.1-41) is expressed in terms of the DF and its derivatives! The result of this process is

$$N_A(A, B) = \sum_{p=0}^{\infty} \frac{B^{2p}}{2^{2p}(p!)^2} V_p(A) \quad (5.1-43a)$$

where $V_p(A)$ is computed *recursively* as

$$V_{p+1}(A) = \frac{d^2 V_p(A)}{dA^2} + \frac{3}{A} \frac{dV_p(A)}{dA} \quad (5.1-43b)$$

$$\text{with} \quad V_0(A) = N(A) \quad (5.1-43c)$$

In a completely similar way one can show that

$$N_B(A, B) = \sum_{p=0}^{\infty} \frac{B^{2p}}{2^{2p}p!(p+1)!} W_p(A) \quad (5.1-44a)$$

$$\text{where} \quad W_{p+1}(A) = \frac{d^2 W_p(A)}{dA^2} + \frac{1}{A} \frac{dW_p(A)}{dA} \quad (5.1-44b)$$

$$\text{and} \quad W_0(A) = N(A) + \frac{A}{2} \frac{dN(A)}{dA} \quad (5.1-44c)$$

Although these results may look cumbersome, they are actually quite simple to employ. Their utility is demonstrated by the following examples.

Example 5.1-4 Compute the TSIDFs for a cubic characteristic, $y = x^3$.

The corresponding DF is $N(A) = \frac{3}{4}A^2$. Thus, from Eqs. (5.1-43c) and (5.1-44c),

$$V_0 = \frac{3}{4}A^2 \quad \text{and} \quad W_0 = \frac{3}{2}A^2$$

Employing the recursive formulations for V_{p+1} and W_{p+1} [Eqs. (5.1-43b) and (5.1-44b)] gives

$$V_1 = W_1 = 6$$

$$V_p = W_p = 0 \quad \text{for all } p \geq 2$$

Substitution of these results into the power-series TSIDF expressions [Eqs. (5.1-43a) and (5.1-44a)] gives

$$N_A(A, B) = \frac{3}{4}A^2 + \frac{3}{2}B^2$$

$$N_B(A, B) = \frac{3}{2}A^2 + \frac{3}{4}B^2$$

These expressions are valid for all A, B and are identical with the cubic characteristic TSIDFs computed previously by direct expansion.

Example 5.1-5 Compute the TSIDFs for an ideal relay with drive levels $\pm D$.

Proceeding as above, but with $N(A) = 4D/\pi A$, yields

$$\begin{aligned} V_0 &= \frac{4D}{\pi A} & W_0 &= \frac{2D}{\pi A} \\ V_1 &= \frac{-4D}{\pi A^3} & W_1 &= \frac{2D}{\pi A^3} \\ V_2 &= \frac{-12D}{\pi A^5} & W_2 &= \frac{18D}{\pi A^5} \\ &\dots\dots\dots \end{aligned}$$

The TSIDF power-series expressions are therefore

$$\begin{aligned} N_A(A,B) &= \frac{4D}{\pi A} \left[1 - \frac{1}{4} \left(\frac{B}{A} \right)^2 - \frac{3}{64} \left(\frac{B}{A} \right)^4 - \frac{5}{256} \left(\frac{B}{A} \right)^6 - \dots \right] \\ N_B(A,B) &= \frac{2D}{\pi A} \left[1 + \frac{1}{8} \left(\frac{B}{A} \right)^2 + \frac{3}{64} \left(\frac{B}{A} \right)^4 + \frac{25}{1,024} \left(\frac{B}{A} \right)^6 + \dots \right] \end{aligned}$$

The ratio test for infinite series can be employed to determine that each of the above expansions is convergent for $(B/A)^2 < 1$. These expressions are identical with those determined by double Fourier series expansion [Eqs. (5.1-19a) and (5.1-19b)].

Example 5.1-6 Compute the TSIDFs for the harmonic nonlinearity, $y = \sin mx$.

The required DF is $N(A) = 2J_1(mA)/A$. Application of the differential relationships

$$\frac{dJ_1(mA)}{dA} = -\frac{J_1(mA)}{A} + mJ_0(mA) \qquad \frac{dJ_0(mA)}{dA} = -J_1(mA)$$

yields

$$\begin{aligned} V_0 &= \frac{2J_1(mA)}{A} & W_0 &= mJ_0(mA) \\ V_1 &= \frac{-2m^2J_1(mA)}{A} & W_1 &= -m^2J_0(mA) \\ V_2 &= \frac{2m^4J_1(mA)}{A} & W_2 &= m^4J_0(mA) \\ &\dots\dots\dots \\ V_p &= (-1)^p 2m^{2p} \frac{J_1(mA)}{A} & W_p &= (-1)^p m^{2p+1} J_0(mA) \end{aligned}$$

Substituting for V_p in Eq. (5.1-43a) gives

$$\begin{aligned} N_A(A,B) &= \frac{2}{A} J_1(mA) \sum_{p=0}^{\infty} \frac{(-1)^p (mB)^{2p}}{2^{2p} (p!)^2} \\ &= \frac{2}{A} J_1(mA) J_0(mB) \end{aligned}$$

upon identification of the infinite series for $J_0(mB)$. Similarly,

$$N_B(A,B) = \frac{2}{B} J_1(mB) J_0(mA)$$

These expressions exist for all A , B and are indeed the exact TSIDFs for the harmonic characteristic (cf. Prob. 5-3).

OTHER METHODS OF TSIDF CALCULATION

One obvious means of TSIDF calculation is by a digital-computer-implemented Fourier series analysis of the actual nonlinearity output. This approach has been used by Jaffe (Ref. 13), Elgerd (Ref. 8), and others for the study of various nonlinear elements. It has the disadvantage that closed-form solutions are never arrived at. On the other hand, it has the far-reaching advantage that it is applicable to *any* nonlinearity whose output is pointwise-determinable in terms of its input.

There remains one additional TSIDF calculation of interest. We defer this to the next chapter, where it is proved that the TSIDF can be computed as the DF of a DIDF, the last-mentioned describing function being the subject of that chapter.

Normalized TSIDF graphs for five common nonlinearities are presented in Appendix D.

5.2 SUBHARMONIC OSCILLATIONS

Systems with suitable nonlinear characteristics can respond to an input sinusoid by producing an output whose lowest-frequency component is a submultiple of the input frequency. The lowest-frequency component is usually at or near the system natural frequency of oscillation. This frequency response phenomenon is thus known as *subharmonic resonance*. Subharmonic frequencies as low as one-thirteenth of the input frequency have been observed in simple systems. Because of the resonant nature of the phenomenon, the subharmonic component is often of large amplitude, causing the system output to bear little resemblance to a sinusoid of the input frequency. Interestingly enough, just outside the region of subharmonic oscillation, the sinusoidally forced system output may be so small as to be practically zero. Complete treatments of subharmonic response phenomena in the cases of smooth and piecewise-linear nonlinearities can be found elsewhere (Refs. 11, 18). For ease in presentation, the TSIDF treatment here is based upon a cubic nonlinearity.

We have already computed the one-third-subharmonic TSIDF in systems with a cubic nonlinearity [Eq. (5.1-7b)]. It is rewritten here:

$$N_B(A, B, \frac{1}{3}, \theta) = \frac{3}{4}(B^2 + 2A^2 + AB \cos 3\theta - jAB \sin 3\theta) \quad (5.2-1)$$

Call the phase shift of this TSIDF β . It follows that ($k = B/A$)

$$\tan \beta = \frac{-k \sin 3\theta}{2 + k^2 + k \cos 3\theta} \quad (5.2-2)$$

This function, maximized over θ , is

$$\tan \beta_{\max \theta} = \frac{-k}{\sqrt{4 + 3k^2 + k^4}} \quad (5.2-3)$$

which can be further extremized with respect to k . Considering real k , Eq. (5.2-3) yields, for $k^2 = 2$,

$$\begin{aligned} \beta_{\max \theta, k} &= \pm \tan^{-1} \sqrt{\frac{1}{7}} \\ &\approx \pm 21^\circ \end{aligned} \quad (5.2-4)$$

Hence the *maximum possible* TSIDF phase shift is 21° . To avoid a one-third-subharmonic oscillation, the loop linear elements must have a phase lag *not exceeding* 159° . Under this condition the total phase shift is less than 180° , and oscillation cannot ensue (see Fig. 5.2-1).

Now consider the frequency locus $L(j\omega)$, for which one-third-subharmonic oscillations are possible at input frequencies between $3\omega_1$ and $3\omega_2$ radians/sec. Corresponding to some input frequency ω within this range, $L(j\omega/3)$ provides in excess of 159° phase lag. $N_B(A, B, \frac{1}{3}, \theta)$ need therefore only provide an additional phase lag of *less than* 21° for a one-third-subharmonic oscillation to take place. The governing equations are (for a unity feedback configuration)

$$A = \frac{M_r}{|1 + N_B(A, B, \frac{1}{3}, \theta)L(j\omega)|} \quad (5.2-5)$$

$$\text{and} \quad L(j\omega/3) = \frac{-1}{N_B(A, B, \frac{1}{3}, \theta)} \quad (5.2-6)$$

At any given ω , a range of subharmonic-oscillation amplitudes can exist as M_r is varied. Stability of a particular subharmonic oscillation is treated in terms of Eqs. (5.2-5) and (5.2-6) and the usual quasi-static describing function argument. It is to be noted that, as the phase of $N_B(A, B, \frac{1}{3}, \theta)$ is small for small values of B , the subharmonic oscillation is not self-starting, and must be initiated by some transient within the system.

Example 5.2-1 Determine all pairs of input amplitude and frequency such that the system of Fig. 5.2-2 can exhibit subharmonic oscillations.

We note that Eq. (5.2-6) implies the angle and magnitude conditions

$$\angle L(j\omega/3) + \angle N_B(A, B, \frac{1}{3}, \theta) = -\pi \quad (5.2-7a)$$

$$\text{and} \quad |L(j\omega/3)| |N_B(A, B, \frac{1}{3}, \theta)| = 1 \quad (5.2-7b)$$

Equation (5.2-7a) can be manipulated in the specific case at hand to yield

$$\frac{k \sin 3\theta}{2 + k^2 + k \cos 3\theta} = \tan \left(\frac{\pi}{2} - 2 \tan^{-1} \frac{(\omega\tau/3)(\alpha - 1)}{1 + \alpha(\omega\tau/3)^2} \right) \quad (5.2-8)$$

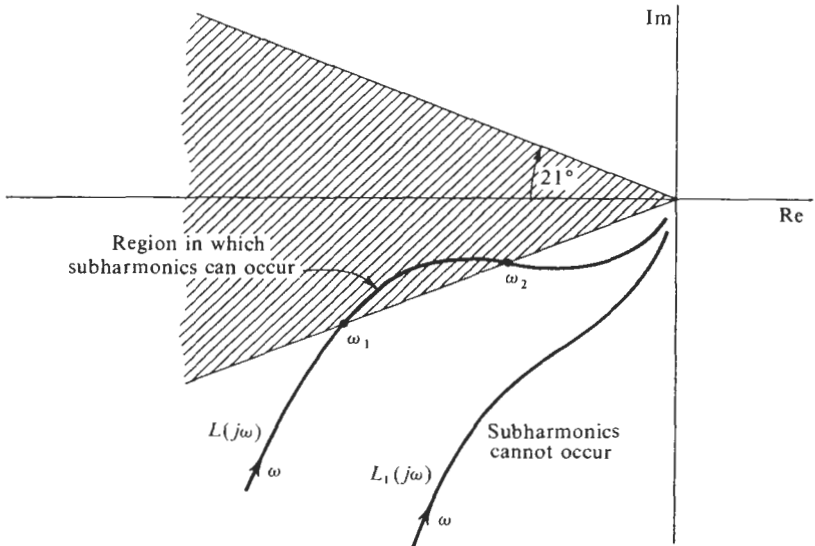


Figure 5.2-1 The 21° rule for elimination of one-third-subharmonic generation in systems with a cubic nonlinearity. (Adapted from West, Ref. 23.)

which is, in fact, a quadratic in k for fixed ω , θ . Similarly, Eq. (5.2-7b) can be manipulated to yield

$$A = \sqrt{\frac{4\omega}{9K} \frac{1 + (\alpha\omega\tau/3)^2}{[1 + (\omega\tau/3)^2]\sqrt{(2 + k^2 + k \cos 3\theta)^2 + (k \sin 3\theta)^2}}} \quad (5.2-9)$$

The last relationship required corresponds to Eq. (5.2-5). This gives three equations in the three unknowns, A , B , θ , provided that ω is held constant.

Insofar as the three equations arrived at are themselves particularly nonlinear, their closed-form solution is not sought. However, the desired solution can still be obtained. By picking values of θ in the range $0 \leq \theta \leq 120^\circ$, the corresponding values of k are specified in terms of Eq. (5.2-8). Only positive real values of k are of interest. Next, Eq. (5.2-9)

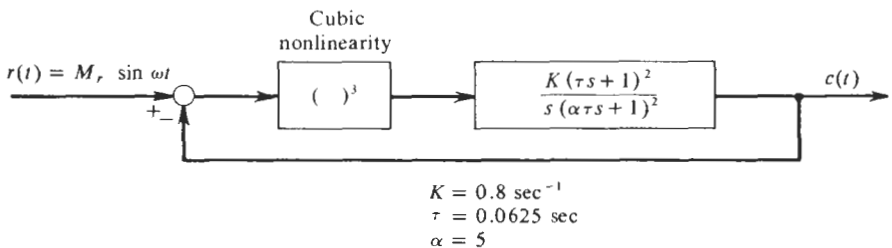


Figure 5.2-2 Third-order system with cubic nonlinearity.

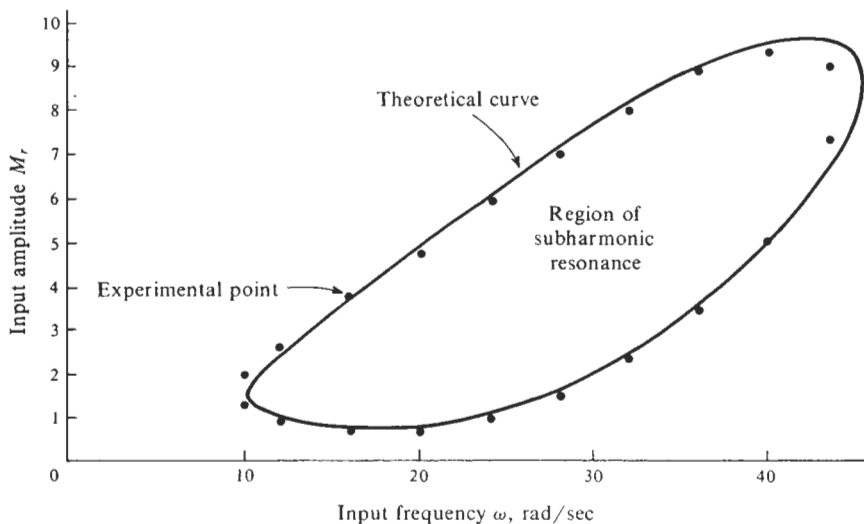


Figure 5.2-3 Region of subharmonic resonance for example system.

yields the values of A which correspond to the pairs of values (θ, k) . Finally, the last equation is used to determine the corresponding M_r .

As ω is varied, the complete range of solutions is traversed. Theoretical results and excellent analog-computer verification are illustrated in Fig. 5.2-3. A typical input-output waveform pair is illustrated in Fig. 5.2-4a. The "resonance" is apparent. As the input amplitude is either increased or decreased, a point is reached at which the subharmonic resonance disappears. Figure 5.2-4b depicts the decay of the subharmonic mode due to a slight decrease in the input amplitude. Observe that the steady-state response beyond the subharmonic region, in this case at least, is imperceptible.

It is often stated, erroneously, that only odd harmonics can occur in systems containing an odd nonlinearity. This interesting error appears to be the consequence of one major aspect of approximation techniques, namely, you get *only* what you ask for. Thus, in the cubic nonlinearity system of Example 5.2-1, even subharmonics (of order one-half) were indeed observed during simulation; however, a *bias* was associated with the corresponding nonlinearity input. Unless this bias is specifically provided for in the model nonlinearity input waveform, approximation methods such as TSIDF analysis fail to predict the even subharmonic mode.

Finally, it is to be noted that the frequency domain criterion for avoiding subharmonics (i.e., 21° rule for systems with a cubic nonlinearity) is far more readily applied than detailed determination of the region in M_r, ω coordinates throughout which subharmonics can occur. In fact, analytic descriptions of the subharmonic TSIDFs for common nonlinearities are generally not available, although some have been obtained experimentally. For the

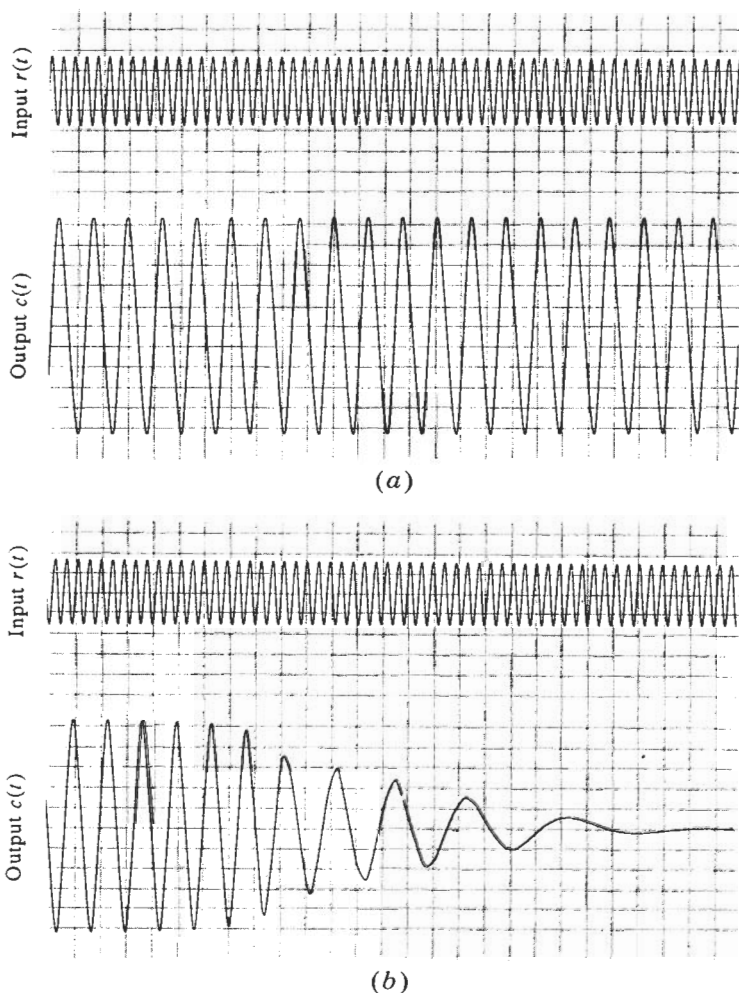


Figure 5.2-4 (a) Illustration of subharmonic resonance. (b) Decay of a subharmonic mode.

piecewise-linear limiter ($D = \delta = 1$), Douce and King (Ref. 6) have determined the bounds of the one-third-subharmonic TSIDF to be such that the condition for avoiding subharmonic resonance is as illustrated in Fig. 5.2-5. That for the ideal relay follows directly (cf. Prob. 5-8).

5.3 FREQUENCY RESPONSE COUNTEREXAMPLES

The DF calculation of frequency response in Chap. 3 is prefaced with a note of caution. It is stated there that certain non-limit-cycling systems break into a limit cycle oscillation when forced by a sinusoidal input (thus rendering invalid the DF calculation of frequency response), and also that certain

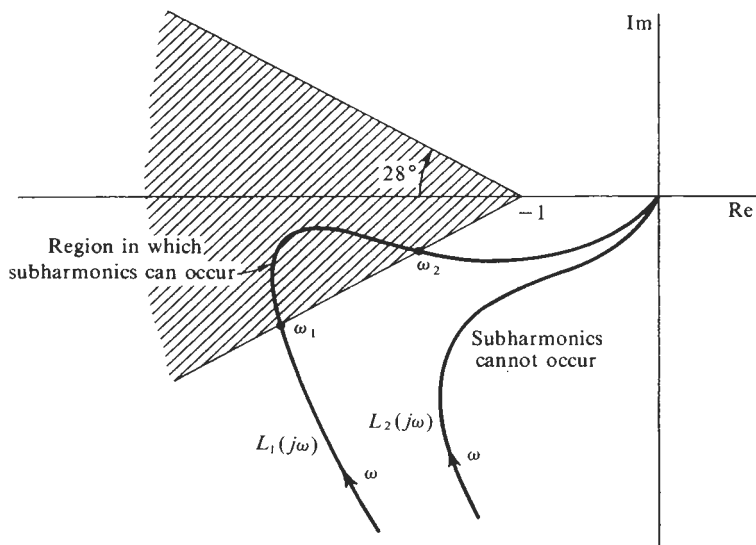


Figure 5.2-5 The 28° rule for elimination of one-third-subharmonic generation in systems containing a limiter. (Adapted from Douce, Ref. 7.)

limit cycling systems are quenched by the introduction of a sinusoidal input (thus permitting DF calculation of frequency response). In this section examples of each of the above phenomena are presented. The emphasis in presentation is on a physical interpretation of cause and effect. For further information the reader is referred to the interesting work of Gibson and Sridhar (Ref. 10), from which the following examples are drawn.

Example 5.3-1 *Limit cycle induction.* Consider the system of Fig. 5.3-1. This is a special case of Example 3.1-1 (that is, $\zeta = \omega_n = \delta = D = 1$), where it is shown by DF utilization that no limit cycle takes place when $K < 3.14$. For larger values of K a limit cycle of frequency $\omega_0 = 1$ exists.

Now set $K = 2$, so that $L(j1) = -1$, and apply an input $r = M_r \sin \omega t$. To test for

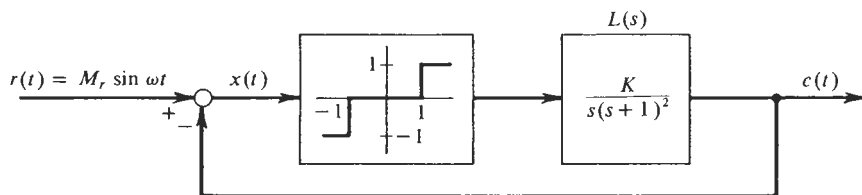


Figure 5.3-1 An example of limit cycle induction.

the possibility of a limit cycle, model the nonlinearity input according to

$$x = A \sin \omega t + B \sin t \quad (5.3-1)$$

where B is the limit cycle amplitude, should one exist. The requirement for a limit cycle in TSIDF terms is

$$N_B(A, B)L(j1) = -1$$

$$\text{or simply,} \quad N_B(A, B) = 1 \quad (5.3-2)$$

Whereas the DF for the relay with dead zone has a maximum value less than unity [such that no limit cycle can circulate according to the DF counterpart of Eq. (5.3-2)], the corresponding TSIDF can have values far in excess of unity. In particular, a range of A and B exist in which Eq. (5.3-2) is precisely satisfied; *a limit cycle can exist!*

The effect of an *externally* applied sinusoid has been to increase the (*internal*) loop gain to the point where an instability (i.e., limit cycle) can occur. To see the mechanism whereby the TSIDF gain can exceed unity, consider the value $A \approx 1.0$. The nonlinearity output never departs from zero, but input excursions carry right up to the discontinuities. Now a *very small* added signal (B) causes the output to jump to ± 1 ; hence the gain to the very small input under this circumstance is very large. This accounts for the behavior of the curve $\alpha = 1.0$ near the origin in Fig. D.3. Although not shown, there exists a continuous range of values $1.0 < \alpha < 1.3$ in which $N_B(A, B) \geq 1$. Therefore we conclude that all combinations of M_r and ω which result in $1.0 < A < 1.3$ are accompanied by an input-induced limit cycle.

Example 5.3-2 Limit cycle quenching. Consider the system of Fig. 5.3-2. This is a special case of Example 3.7-6 (that is, $\zeta = \omega_n = D = 1$), where it is noted that a limit cycle of frequency $\omega_0 = 1$ exists for all values of K .

Once again, set $K = 2$, apply an input $r = M_r \sin \omega t$, and model the system error signal according to Eq. (5.3-1). The condition for the limit cycle to continue is then satisfaction of Eq. (5.3-2). Figure D.1, a normalized plot of the ideal-relay TSIDF, illustrates that for *any* specified value of B , $N_B(A, B)$ decreases as A increases. It follows that there exists a value of A , which we denote $A_{\min}$, such that, for any value of A in the range $A_{\min} < A < \infty$, $N_B(A, B)$ is always less than unity, and no limit cycle can circulate.

$A_{\min}$ is readily calculated. In the region of its maximum value, $N_B(A, B)$ is given by Eq. (5.1-35), namely,

$$\begin{aligned} N_B(A, B) &= \frac{8}{\pi^2 B} E\left(\frac{A}{B}\right) \quad \frac{B}{A} > 1 \\ &= \frac{8}{\pi^2 A} x E(x) \end{aligned}$$

where $x = A/B$. With the aid of the differential relationship

$$\frac{dE(x)}{dx} = \frac{1}{x} [E(x) - K(x)]$$

the value of x which maximizes $N_B(A, B)$, denoted $x_{\max}$, is found to satisfy

$$2E(x_{\max}) = K(x_{\max})$$

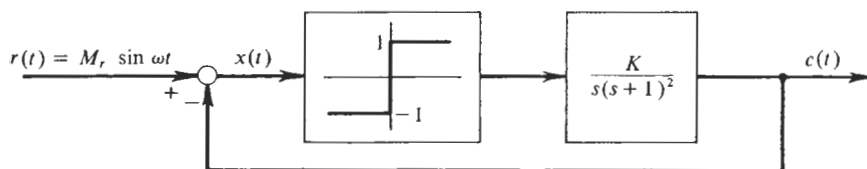


Figure 5.3-2 An example of limit cycle quenching.

Standard tables of the complete elliptic integrals can be consulted to yield that

$$x_{\max} \approx 0.909$$

at which point the maximum value of $N_B(A, B)$ is computed as

$$N_B(A, B)_{\max} = \frac{0.855}{A}$$

This results in the determination that $A_{\max} = 0.855$. Hence all combinations of M_r and ω for which $A > 0.855$ drop the loop gain at the limit cycle frequency to the point where the limit cycle is quenched (e.g., extinguished). These combinations are determined with DF theory, since there is now but one sinusoid with which to contend.

Circumstances under which a DF treatment of frequency response is insufficient can be deduced from the foregoing examples. Non-limit-cycling systems which can support a limit cycle at some values of loop gain (either higher or lower than the no-limit-cycle setting) must be checked by a TSIDF calculation to ensure the absence of a limit cycle. Limit cycling systems which can be quenched by either a raised or lowered loop gain should also be subject to TSIDF calculation to determine the range in which a frequency response exists. This phenomenon, called *signal stabilization*, is treated in more detail in Chap. 6.

5.4 MULTIPLE LIMIT CYCLES

Nonlinear systems often exhibit more than a single stable limit cycle. In such multiple limit cycle systems the possibilities exist that either one limit cycle predominates, several limit cycles occur one at a time, or several limit cycles occur simultaneously. A TSIDF argument is required for determination of the circumstances leading to the first of these alternatives.

Consider the relay system of Fig. 5.4-1a for which DF theory tells us that the limit cycles at frequencies ω_A (higher) and ω_B (lower) are stable. If the higher-frequency limit cycle is excited, the following relationship must hold:

$$N_A(A, B) |L(j\omega_A)| = 1 \quad (5.4-1)$$

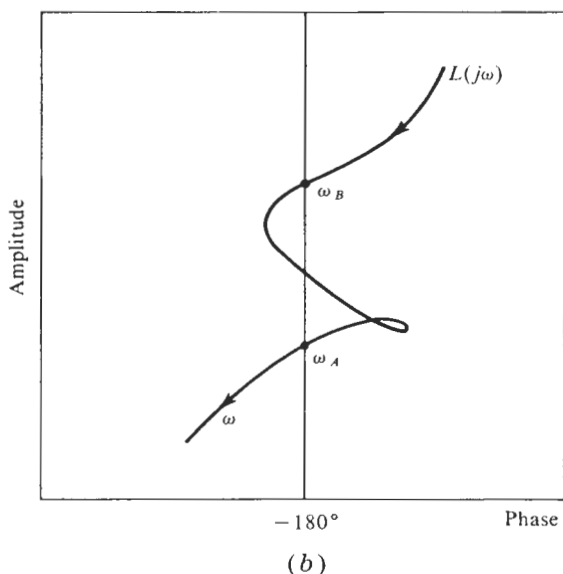
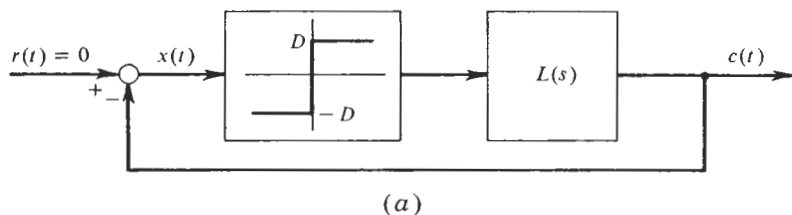


Figure 5.4-1 (a) Multiple limit cycle system. (b) Corresponding amplitude-phase plot.

where A is the limit cycle amplitude and $B = 0$. Now, using as a model for the nonlinearity input

$$x = A \sin \omega_A t + B \sin \omega_B t \quad (5.4-2)$$

where $B \ll A$, the TSIDF is employed to determine the nature of stability of the perturbed low-frequency limit cycle mode. The procedure is quite direct. Use of Eq. (5.1-20) enables rewriting Eq. (5.4-1) in the form

$$N_B(A, B) |L(j\omega_B)| = \frac{1}{2} \left| \frac{L(j\omega_B)}{L(j\omega_A)} \right| \quad (5.4-3)$$

from which it is deduced that the perturbed low-frequency mode is *unstable* if

$$\left| \frac{L(j\omega_B)}{L(j\omega_A)} \right| > 2 \quad (5.4-4)$$

that is, the oscillation grows. Hence, when Eq. (5.4-4) is satisfied, the high-frequency limit cycle alone is *not* a stable systematic mode. Otherwise it is.

Now apply the same argument to test the reverse situation. This almost always shows the perturbed high-frequency mode to be *stable* in the presence of a low-frequency limit cycle since the condition $|L(j\omega_A)/L(j\omega_B)| > 2$ is rare indeed. In those circumstances where this mode is *unstable*, the linear element's frequency locus is highly resonant in the vicinity of ω_A . Non-linearity output harmonic content near ω_A is consequently not negligible, and as a result, the TSIDF linearization is inappropriate. A somewhat different argument is required to ascertain conditions of stability. To facilitate analysis, the case where only the fundamental and the harmonic component nearest the higher-frequency oscillation mode is of significant amplitude can be treated. This is the approach taken in Ref. 1.

5.5 INCREMENTAL-INPUT DESCRIBING FUNCTION

A special case of the TSIDF occurs when the amplitude of one nonlinearity input sinusoid is much smaller than the amplitude of the other.¹ Under this circumstance, a simple closed-form solution exists for the nonlinearity gain to the small-amplitude input component; it is called the *incremental-input describing function*. This particular describing function is of considerable interest. It has application to the study of oscillation stability (including stability of forced sinusoidal response and limit cycle stability), and importantly, it is an excellent approximation to the TSIDF in the case of widely differing input amplitudes. Following the nomenclature used by Bonenn (Ref. 4), we consider separately the cases of synchronous (i.e., identical frequency) inputs and nonsynchronous (i.e., different frequency) inputs.

DERIVATION OF THE INCREMENTAL-INPUT DESCRIBING FUNCTION

Synchronous inputs Let the two sinusoidal inputs to a static nonlinearity be given by

$$x = A \sin \omega t + \epsilon \sin (\omega t + \theta) \quad (5.5-1)$$

where the second term represents the *incremental input*, that is,

$$\frac{\epsilon}{A} \ll 1 \quad (5.5-2)$$

¹ The content of this section can be viewed as a direct result of TSIDF expansion in a power series (Sec. 5.1). Rather than simply setting B to zero in Eqs. (5.1-43a) and (5.1-44a), a brief and more direct derivation of the describing functions of interest is presented here.

Expanding Eq. (5.5-1) and regrouping terms, x can be put in the form

$$\begin{aligned} x &= (A + \epsilon \cos \theta) \sin \omega t + (\epsilon \sin \theta) \cos \omega t \\ &= \sqrt{(A + \epsilon \cos \theta)^2 + (\epsilon \sin \theta)^2} \sin \left(\omega t + \tan^{-1} \frac{\epsilon \sin \theta}{A + \epsilon \cos \theta} \right) \end{aligned} \quad (5.5-3)$$

Using the relative amplitude constraint [Eq. (5.5-2)], the expressions for input magnitude and phase can be further simplified. Dropping second- and higher-order terms in ϵ/A , we have

$$\sqrt{(A + \epsilon \cos \theta)^2 + (\epsilon \sin \theta)^2} \approx A + \epsilon \cos \theta \quad (5.5-4)$$

$$\text{and} \quad \tan^{-1} \frac{\epsilon \sin \theta}{A + \epsilon \cos \theta} \approx \frac{\epsilon}{A} \sin \theta \quad (5.5-5)$$

Hence Eq. (5.5-1) can be written in the alternative form

$$x \approx (A + \epsilon \cos \theta) \sin \left(\omega t + \frac{\epsilon}{A} \sin \theta \right) \quad (5.5-6)$$

If $N(A) = n_p(A) + jn_q(A)$ is the DF for the nonlinearity, and has a continuous first derivative with respect to A , the output is (prime denoting differentiation with respect to A)

$$\begin{aligned} \text{Output} &\approx (A + \epsilon \cos \theta) [n_p(A + \epsilon \cos \theta)] \sin \left(\omega t + \frac{\epsilon}{A} \sin \theta \right) \\ &\quad + (A + \epsilon \cos \theta) [n_q(A + \epsilon \cos \theta)] \cos \left(\omega t + \frac{\epsilon}{A} \sin \theta \right) \\ &\approx (A + \epsilon \cos \theta) [n_p(A) + \epsilon \cos \theta n'_p(A)] \left(\sin \omega t + \frac{\epsilon}{A} \sin \theta \cos \omega t \right) \\ &\quad + (A + \epsilon \cos \theta) [n_q(A) + \epsilon \cos \theta n'_q(A)] \left(\cos \omega t - \frac{\epsilon}{A} \sin \theta \sin \omega t \right) \\ &\approx A [n_p(A) \sin \omega t + n_q(A) \cos \omega t] \\ &\quad + \epsilon [n_p(A) \sin (\omega t + \theta) + n_q(A) \cos (\omega t + \theta)] \\ &\quad + A \cos \theta (n'_p(A) \sin \omega t + n'_q(A) \cos \omega t) \end{aligned} \quad (5.5-7)$$

where again only first-order terms in ϵ/A have been retained. Defining the incremental-input describing function as the complex ratio of output terms due to ϵ divided by the corresponding input term, we finally arrive at the

desired result,

$N_i(A, \theta)$ = incremental-input describing function

$$\begin{aligned}
 &= \frac{\epsilon [n_p(A)e^{j(\omega t + \theta)} + n_q(A)e^{j(\omega t + \theta + \pi/2)} + A \cos \theta (n'_p(A)e^{j\omega t} + n'_q(A)e^{j(\omega t + \pi/2)})]}{\epsilon e^{j(\omega t + \theta)}} \\
 &= n_p(A) + jn_q(A) + A[n'_p(A) + jn'_q(A)] \cos \theta e^{-j\theta} \\
 &= N(A) + AN'(A) \cos \theta e^{-j\theta}
 \end{aligned} \tag{5.5-8}$$

Observe that the incremental-input describing function is independent of ϵ and is derivable directly from the ordinary DF for the nonlinearity. By rearranging the θ dependence of $N_i(A, \theta)$, one may show that

$$N_i(A, \theta) = N(A) + \frac{AN'(A)}{2} (1 + e^{-j2\theta}) \tag{5.5-9}$$

in which form it is evident that, for fixed A , the tip of the vector $N_i(A, \theta)$ traces out a circle in the polar plane (see Fig. 5.5-1).

Nonsynchronous inputs The nonlinearity input is now taken as

$$x = A \sin \omega t + \epsilon \sin \gamma \omega t \tag{5.5-10}$$

where the frequency ratio γ is *irrational*, and the condition expressed by Eq. (5.5-2) still holds. The phase angle θ has been discarded since x now possesses an aperiodic waveform. In fact, by considering θ to be a uniformly distributed random variable, one can derive the incremental-input describing function pertinent to the present case simply by averaging $N_i(A, \theta)$ of the synchronous case [Eq. (5.5-9)] with respect to θ . The result, independent of θ , is given by (Ref. 4)

$$N_i(A) = N(A) + \frac{AN'(A)}{2} \tag{5.5-11}$$

This is precisely the relationship arrived at by setting B to zero in Eq. (5.1-44a). That is,

$$N_i(A) = W_0(A) \tag{5.5-12}$$

where $W_0(A)$ is the first term in the expansion of $N_B(A, B)$ in powers of B .

APPLICATION TO JUMP RESONANCE PHENOMENA

A natural application for the synchronous incremental-input describing function lies in the study of jump resonance phenomena (Sec. 3.3). For the nonlinear system of Fig. 5.5-2a, the possibility of jump resonance can be studied in terms of the stability of a sinusoidal perturbation of frequency ω

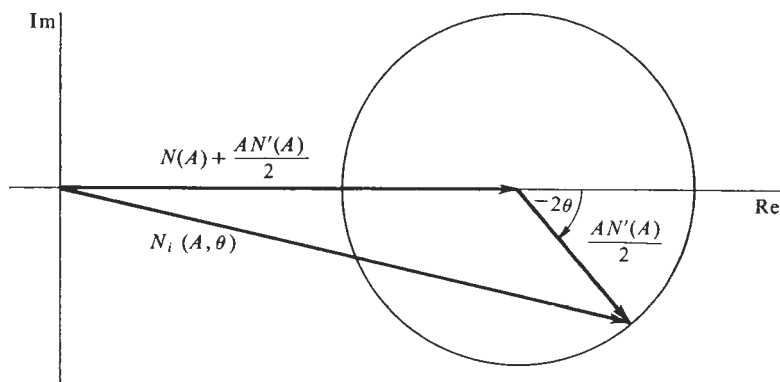


Figure 5.5-1 Complex representation of $N_i(A, \theta)$ for synchronous inputs. Illustrated for the case where $N(A)$ is real.

about a steady-state forced oscillation at the same frequency. The characteristic equation of the system seen by such a perturbation is (Fig. 5.5-2b)

$$1 + N_i(A, \theta)L(j\omega) = 0 \quad (5.5-13)$$

or

$$L(j\omega) = -\frac{1}{N_i(A, \theta)} \quad (5.5-14)$$

For fixed A , the quantity $-1/N_i(A, \theta)$ plays the role of a *stability point*, or more properly, a *stability curve*, since θ can take on any value from 0 to 2π radians. Whereas with the ordinary DF an instability (limit cycle) is indicated when the locus $L(j\omega)$ passes through the *point* $-1/N(A)$, now an instability (jump resonance) is indicated when the locus $L(j\omega)$ passes through any portion of the *curve* $-1/N_i(A, \theta)$. This new point of view regarding the

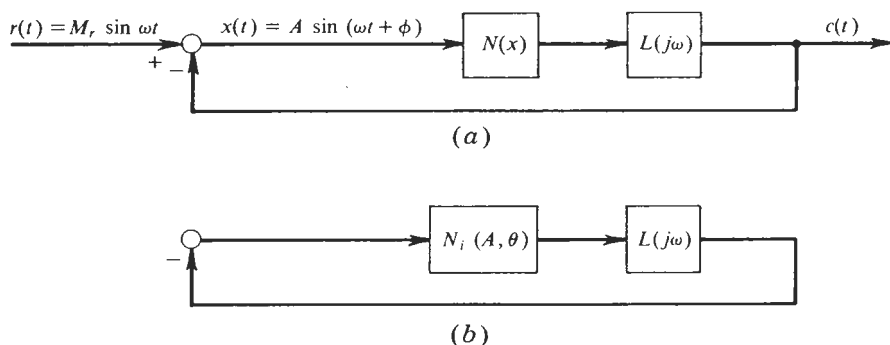


Figure 5.5-2 (a) Nonlinear system. (b) Corresponding incremental system.

jump resonance phenomenon is easily demonstrated as being equivalent to that presented earlier, in Sec. 3.3.

Example 5.5-1 Derive the equations for the contours in the polar plane along which jump resonance can occur.

Expressing $L(j\omega)$ in the form $U(\omega) + jV(\omega)$, Eq. (5.5-13) can be written as

$$1 + \left[N(A) + \frac{AN'(A)}{2} (1 + e^{-j2\theta}) \right] [U(\omega) + jV(\omega)] = 0 \quad (5.5-15)$$

Equating the real and imaginary parts on each side gives

$$\begin{aligned} 1 + \left[N(A) + \frac{AN'(A)}{2} (1 + \cos 2\theta) \right] U(\omega) + \left[\frac{AN'(A)}{2} \sin 2\theta \right] V(\omega) &= 0 \\ \left[N(A) + \frac{AN'(A)}{2} (1 + \cos 2\theta) \right] V(\omega) - \left[\frac{AN'(A)}{2} \sin 2\theta \right] U(\omega) &= 0 \end{aligned}$$

Eliminating θ between these equations and manipulating yields

$$\left[U(\omega) + \frac{1}{2} \left(\frac{1}{N(A)} + \frac{1}{N(A) + AN'(A)} \right) \right]^2 + V(\omega)^2 = \left[\frac{AN'(A)}{2N(A)[N(A) + AN'(A)]} \right]^2 \quad (5.5-16)$$

Geometrically interpreted, the contours of constant A are circles, with center coordinates $(-\frac{1}{2}\{1/N(A) + 1/[N(A) + AN'(A)]\}, 0)$ and radii $|AN'(A)/2N(A)[N(A) + AN'(A)]|$.

It is readily seen that Eqs. (5.5-16) and (3.3-12) are identical. This establishes the equivalence of both ways of viewing the condition for jump resonance. The reader is referred to Sec. 3.3 for further discussion of jump resonance phenomena.

APPLICATION TO TRANSIENT OSCILLATIONS

This section is concerned with the relative stability of small perturbations about steady-state oscillations in nonlinear systems. The systems are assumed describable by equations of the type

$$Q(s)x + P(s)y(x, \dot{x}) = 0 \quad (5.5-17)$$

where $s = d/dt$, and $P(s)$ and $Q(s)$ are polynomials in s . The differential operator corresponding to the system linear elements' transfer function is $L(s) = P(s)/Q(s)$, and the nonlinearity output is characterized by $y(x, \dot{x})$.

For the study of relative stability, the test signal applied to Eq. (5.5-17) is chosen in the form¹

$$x(t) = Ae^{j\omega_0 t} + \epsilon e^{at} e^{j(\omega_0 t + \theta)} \quad (5.5-18)$$

¹ By convention the *real* test signal is understood to be the imaginary part of this *complex* expression. The derivation closely follows Bonenn (Ref. 4).

where ϵ is an arbitrary but small parameter. Substituting for $x(t)$ in Eq. (5.5-17) yields, after making the describing function approximation,

$$Q(s)(Ae^{j\omega_0 t} + \epsilon e^{\sigma t} e^{j(\omega_0 t + \theta)}) + P(s)(AN(A)e^{j\omega_0 t} + \epsilon N_i(A, \theta)e^{\sigma t} e^{j(\omega_0 t + \theta)}) = 0 \quad (5.5-19)$$

The dominant terms in this equation, of order A , represent the steady-state oscillation and sum exactly to zero. For the remaining perturbation terms, application of the differential-operator rules (Ref. 19),

$$L(s)[e^{\sigma t} f(t)] = e^{\sigma t} L(s + \sigma)[f(t)] \quad (5.5-20)$$

$$L(s)[e^{j\omega t}] = e^{j\omega t} L(j\omega) \quad (5.5-21)$$

$$\text{yields} \quad \epsilon e^{\sigma t} e^{j(\omega_0 t + \theta)} [Q(\sigma + j\omega_0) + N_i(A, \theta)P(\sigma + j\omega_0)] = 0 \quad (5.5-22)$$

To satisfy equality with zero requires that the term in brackets vanishes; the perturbation signal itself is arbitrary. Dividing this term by $Q(\sigma + j\omega_0)$ gives, as the characteristic equation of the perturbed system,

$$1 + N_i(A, \theta)L(\sigma + j\omega_0) = 0 \quad (5.5-23)$$

where ω_0 is the steady-state oscillation frequency, and σ is the damping factor of interest. Equation (5.5-23) is a generalization of Eq. (5.5-13), the latter implicitly assuming $\sigma = 0$. The transition from Eq. (5.5-13) to (5.5-23) is analogous to the transition from stability determination by frequency methods to stability characterization in terms of poles and zeros. Of course, it is assumed in writing Eq. (5.5-23) that $N_i(A, \theta)$ is the appropriate gain to associate with an incremental input $\epsilon \exp(\sigma t) \sin \omega_0 t$. For σ sufficiently close to zero, this approximation is certainly valid. The nature of the approximation is identical with that made in Chap. 4, where $N(A)$ is used to characterize the nonlinearity in response to a sinusoid of slowly varying amplitude and frequency. The point of view adopted here, however, is quite different from that in Chap. 4.

A graphical procedure is employed to extract the desired solution of Eq. (5.5-23). It is best demonstrated by an example.

Example 5.5-2 Calculate the damping factor associated with limit cycle perturbations in the system of Fig. 5.5-3.

Recall of the DF for a preload nonlinearity enables writing the steady-state limit cycle equation as

$$\left(\frac{4}{\pi A} + 1 \right) \frac{1}{j\omega_0(j\omega_0 + 1)^2} = -1$$

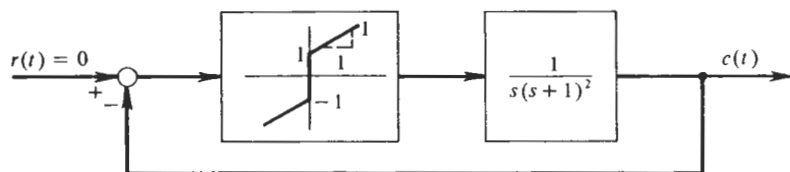


Figure 5.5-3 Limit cycling system.

This equation is satisfied by $\omega_0 = 1$ and $A = 4/\pi$. Now, from Eq. (5.5-9), the synchronous incremental-input describing function for the preload nonlinearity is

$$\begin{aligned} N_i(A, \theta) &= \frac{4}{\pi A} + 1 + \frac{A}{2} \left(-\frac{4}{\pi A^2} \right) (1 + e^{-j2\theta}) \\ &= \frac{3}{2} - \frac{1}{2} e^{-j2\theta} \end{aligned}$$

To facilitate solution, Eq. (5.5-23) is best put in the form

$$-N_i(A, \theta) = \frac{1}{L(\sigma + j\omega_0)}$$

Writing this equation in terms of the present problem gives the following complex relationship:

$$-\frac{3}{2} + \frac{1}{2} e^{-j2\theta} = (\sigma + j)(\sigma + 1 + j)^2$$

Since the left-hand side is a function of θ only, and the right-hand side is a function of σ only, this set of two equations in two unknowns can be readily solved by plotting each side as a function of the pertinent variable and determining intersections of the resultant loci. This is illustrated in Fig. 5.5-4, where we find the solutions $\sigma = 0$ and $\sigma \approx -0.1$. Notably, the quasi-static solution of Chap. 4 also leads to the value $\sigma = -0.1$.

It can be seen in general that at $\omega = \omega_0$, Eq. (5.5-23) has two solutions for σ . One of these always occurs at $\sigma = 0$, and corresponds to the steady-state change in the limit cycle. The other, at negative σ for stable limit cycles, corresponds to the time constant of the limit cycle transient. It is the measure of relative stability which we seek. The analog-computer result shown in Fig. 5.5-4 attests to the accuracy of this calculation, providing an experimental time constant of about 10 sec (that is, $\sigma_{\text{exper}} \approx -0.1 \text{ sec}^{-1}$).

APPLICATION TO TSIDF APPROXIMATION

An important use for the nonsynchronous incremental-input describing function occurs in TSIDF approximation for the case of widely differing input amplitudes. The approximating relationships are, from Eq. (5.5-7),

$$\tilde{N}_A(A, B) = N(A) \quad (5.5-24)$$

and from Eq. (5.5-8),

$$\tilde{N}_B(A, B) = N_i(A) \quad (5.5-25)$$

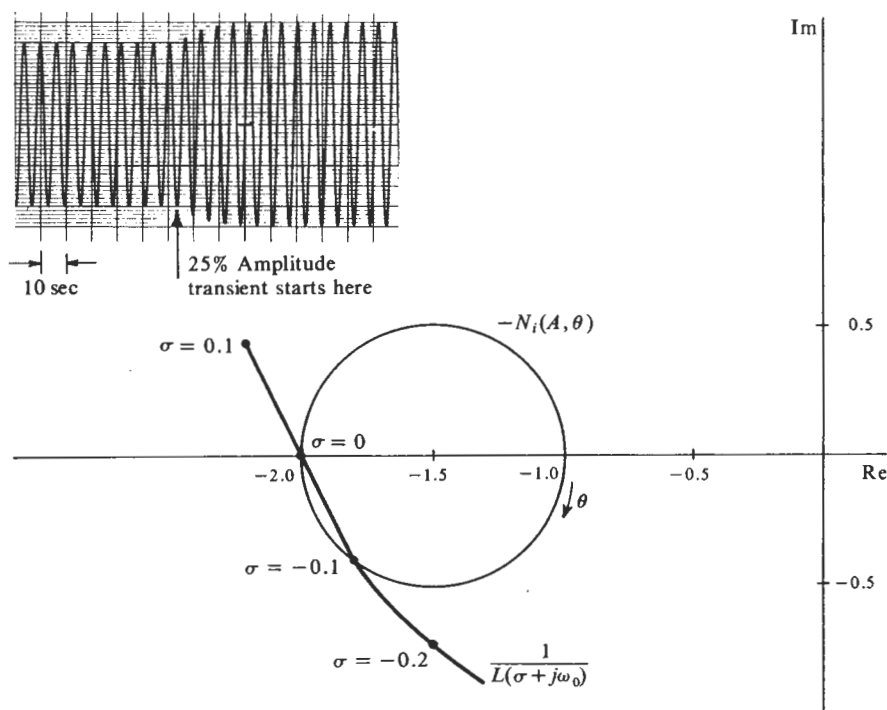


Figure 5.5-4 Graphical solution for the limit cycle transient damping factor. Insert shows analog-computer results

where a tilde denotes the approximation. The requirement is that B be sufficiently small for these approximations to hold. For the majority of nonlinear characteristics, this implies an upper bound on the ratio B/A . For nonlinear characteristics periodic in x (such as the harmonic nonlinearity $y = \sin mx$), this implies a limit on the absolute value of B .

Example 5.5-3 Determine the requirement on B under which the harmonic nonlinearity TSIDFs can be approximated to within 10 percent accuracy by the DF and incremental-input describing function [Eqs. (5.5-24) and (5.5-25)].

For the harmonic nonlinearity $y = \sin mx$, it is readily demonstrated (see Prob. 5-3) that the exact TSIDFs are

$$N_A(A, B) = \frac{2}{A} J_0(mB) J_1(mA) \quad (5.5-26)$$

$$N_B(A, B) = \frac{2}{B} J_0(mA) J_1(mB) \quad (5.5-27)$$

According to the proposed approximations, it follows from Eq. (2.3-23) that

$$\tilde{N}_A(A, B) = \frac{2}{A} J_1(mA) \quad (5.5-28)$$

and from Eq. (5.5-11),

$$\begin{aligned} \tilde{N}_B(A, B) &= \frac{2}{A} J_1(mA) + \frac{A}{2} \frac{d}{dA} \left[\frac{2}{A} J_1(mA) \right] \\ &= mJ_0(mA) \end{aligned} \quad (5.5-29)$$

Comparing Eqs. (5.5-26) and (5.5-28) yields that the exact and approximating TSIDFs differ by the factor $J_0(mB)$. This factor is bounded by 1.0 and 0.9 for mB in the range $|mB| < 0.65$. Now comparing Eqs. (5.5-27) and (5.5-29), we see that the factor of difference is $2J_1(mB)/mB$. This factor is bounded by 1.0 and 0.9 for mB in the range $|mB| < 0.89$. Hence the requirement on B is

$$|mB| < 0.65$$

which is, indeed, independent of A , as noted previously.

The relative ease with which the TSIDF approximation can be calculated certainly promotes its use as an analytical tool. Applications for this particular form of linearization are both diverse and numerous. A typical application is illustrated below.

Example 5.5-4 An optical-beam-riding antitank missile has its position control loop closed by a human operator. The on-off rate control loop employs a 400-cps carrier rate gyro feedback signal and associated amplification, demodulation, and compensation, as shown in Fig. 5.5-5. Assume that the airframe dynamics from δ to ψ are characterized by the transfer function

$$\frac{\psi}{\delta}(s) = \frac{50(s+1)}{s(s^2+2s+25)}$$

It is found that the in-flight missile trajectory oscillates about the optical tracker-target line of sight, producing a signal $e = 25 \sin 6t$ volts. Find the amplitude of missile control-surface deflection due to a 1-volt demodulator output noise at the carrier frequency.

The nonlinearity input is modeled according to

$$x \approx A \sin 6t + B \sin 2,512t$$

We first have to calculate the amplitude A of the low-frequency nonlinearity input component. This is given by the solution to the relationship

$$A = \left| \frac{\frac{0.1s+1}{s+1} 25}{1 + \frac{0.1s+1}{s+1} N(A) \frac{50(s+1)}{s^2+2s+25}} \right|_{s=j6}$$

where the approximation to $N_A(A, B)$ is the DF for the on-off element with dead zone

$$\tilde{N}_A(A, B) = \frac{4}{\pi A} \sqrt{1 - \left(\frac{0.5}{A} \right)^2}$$

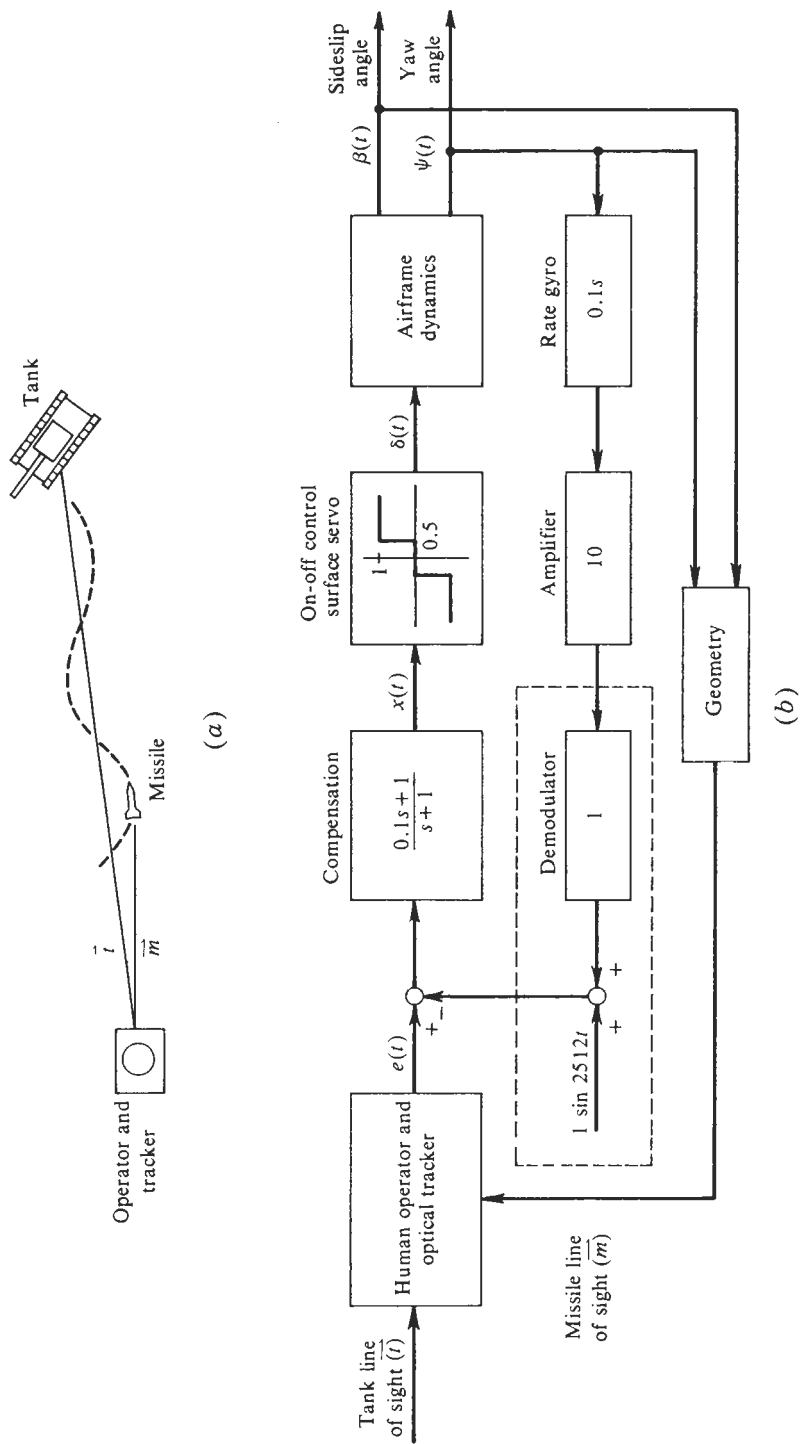


Figure 5.5-5 (a) Antitank missile problem geometry. (b) Control-system configuration.

The solution to these equations is readily found to be $A = 2.5$. Next the incremental-input describing function gain must be calculated in approximation to $N_B(A, B)$.

$$\begin{aligned}\tilde{N}_B(A, B) &= \frac{4}{\pi A} \sqrt{1 - \left(\frac{0.5}{A}\right)^2} + \frac{A}{2} \frac{d}{dA} \left[\frac{4}{\pi A} \sqrt{1 - \left(\frac{0.5}{A}\right)^2} \right] \\ &= \frac{2}{\pi A} \sqrt{1 - \left(\frac{0.5}{A}\right)^2} + \frac{0.5}{\pi A^3 \sqrt{1 - (0.5/A)^2}}\end{aligned}$$

This expression is evaluated at the appropriate value for A .

$$\tilde{N}_B(A, B) \Big|_{A=2.5} = 0.26$$

The system seen by the sinusoidal noise voltage has thus been linearized. Neglecting the small fed-back high-frequency signal, the required control-surface noise-deflection amplitude δ_n is calculated as

$$\delta_n \approx 1 \times \left| \frac{j(0.1)(2,512) + 1}{j2,512 + 1} \right| \times 0.26 = 0.026 \text{ radian}$$

This result is particularly interesting in view of the fact that δ_n would be zero if e were zero. Nonlinearity modification by one input as seen by another input is a phenomenon which can be put to practical use. We shall meet it repeatedly in the remainder of the text.

It is of some interest to check the accuracy of the TSIDF approximation employed. Using earlier notation, we have $A \approx 2.5$, $B \approx 0.1$. Entering Fig. D.1 at the curve labeled $A/\delta = 5$ (which is not shown, but interpolation is sufficient for our purposes here), we find at $B/\delta = 0.2$ the value $N_B(A, B)\delta/D \approx 0.13$, or $N_B(A, B) \approx 0.26$. This is certainly an excellent check on our approximation. It is also possible to check the value of $N_A(A, B)$, although to use the above-mentioned figure, we must employ Eq. (5.1-32). Hence, in place of $N_A(A, B)$, we choose to evaluate $N_B(B, A)$. This places us just below the curve labeled $A/\delta = 0$ at the abscissa value $B/\delta = 5$. The result is the determination that $N_A(A, B) \approx 0.50$. This justifies the use of $\tilde{N}_A(A, B)$, which also has the value 0.50 at $A = 2.5$.

5.6 ADDITIONAL TSIDF APPLICATIONS

By straightforward describing function technique one may use the TSIDF to study the complete forced harmonic response of a limit cycling system and, similarly, the response of a non-limit-cycling system to two simultaneously applied sinusoids. Interpretation of the nonlinearity input signal component parts is evident in each of these cases.

A conditionally stable system is one in which the linear elements' frequency locus has some frequency region displaying greater than 180° phase lag and, simultaneously, a gain in excess of unity. If, for example, such a system contains a limiter, a limit cycle can exist, following a suitable input transient. DF application can determine the limit cycle amplitude and frequency, but not the amplitudes and frequencies of input sinusoids which excite the limit cycle. TSIDF theory is readily applied here (Ref. 22).

Other TSIDF applications include the consideration of harmonic content in loop oscillations (i.e., DF correction term) and the analysis of systems with two nonlinearities separated by filters which insufficiently attenuate harmonics so that ordinary DF application is inaccurate (Ref. 23).

It must be remarked that, in general, TSIDF calculation for harmonically related input sinusoids is laborious. Where alternative methods of analysis obviate the need for this calculation, they may well be preferred. On the other hand, TSIDF calculation in the case of non-harmonically-related input sinusoids is quite easily accomplished. Calculation of the incremental-input describing function proceeds with no difficulty whatever. Once calculated, the employment of various TSIDFs for nonlinear-system study proceeds with the simplicity characteristic of describing function techniques.

REFERENCES

1. Amsler, B. E., and R. E. Gorozdos: On the Analysis of Bi-stable Control Systems, *IRE Trans. Autom. Control*, vol. AC-4 (December, 1959).
2. Bennett, W. R.: New Results in the Calculation of Modulation Products, *Bell System Tech. J.*, vol. 12 (1933), pp. 228-243.
3. Bonenn, Z.: Stability of Forced Oscillations in Nonlinear Feedback Systems, *IRE Trans. Autom. Control*, vol. AC-3 (December, 1958), pp. 109-111.
4. Bonenn, Z.: Relative Stability of Oscillations in Non-linear Control Systems, *Proc. IFAC*, Basel, Switzerland (August, 1963), p. 214/1-4.
5. Courant, R., and D. Hilbert: "Methods of Mathematical Physics," Interscience Publishers, Inc., New York, 1953, vol. 1, p. 475.
6. Douce, J. L., and R. E. King: Instability of a Nonlinear Conditionally Stable System Subjected to a Sinusoidal Input, *Trans. AIEE*, pt. II, *Appl. Ind.* (January, 1959), pp. 665-670.
7. Douce, J. L.: Discussion of a paper by Ogata, *Trans. ASME*, vol. 80, no. 8 (November, 1958), p. 1808.
8. Elgerd, O. I.: High-frequency Signal Injection: A Means of Changing the Transfer Characteristics of Nonlinear Elements, WESCON, 1962.
9. Gibson, J. E.: "Nonlinear Automatic Control," McGraw-Hill Book Company, New York, 1963.
10. Gibson, J. E., and R. Sridhar: A New Dual-input Describing Function and an Application to the Stability of Forced Oscillations, *Trans. AIEE*, pt. II, *Appl. Ind.* (May, 1963), pp. 65-70.
11. Hayashi, C.: "Nonlinear Oscillations in Physical Systems," McGraw-Hill Book Company, New York, 1964.
12. Huey, R. M., O. Pawloff, and T. Glucharoff: Extension of the Dual-input Describing-function Technique to Systems Containing Reactive Nonlinearity, *J. IEE*, London, vol. C-107 (June, 1960), pp. 334-341.
13. Jaffe, R. C.: Causal and Statistical Analyses of Dithered Systems Containing Three-level Quantizers, M.S. thesis, Massachusetts Institute of Technology, Cambridge, Mass., August, 1959.
14. Kalb, R. M., and W. R. Bennett: Ferromagnetic Distortion of a Two-frequency Wave, *Bell System Tech. J.*, vol. 14 (1935), pp. 322-359.

15. Ludeke, C. A.: The Generation and Extinction of Subharmonics, *Proc. Symp. Non-linear Circuit Analysis*, vol. 2 (April, 1953), Polytechnic Institute of Brooklyn, New York.
16. MacColl, L. A.: "Fundamental Theory of Servomechanisms," D. Van Nostrand Company, Inc., Princeton, N.J., 1945.
17. Oldenburger, R., and R. C. Boyer: Effects of Extra Sinusoidal Inputs to Nonlinear Systems, *Trans. ASME, J. Basic Eng.* (December, 1962), pp. 559-570.
18. Oldenburger, R., and R. Nicholls: Stability of Subharmonic Oscillations in Non-linear Systems, *Proc. JACC*, Minneapolis, Minn. (June, 1963), pp. 675-680.
19. Piaggio, H. T. H.: "Differential Equations," G. Bell & Sons, Ltd., London, 1946, p. 32.
20. Rice, S. O.: Mathematical Analysis of Random Noise, *Bell System Tech. J.*, vol. 24, pt. IV (1945), pp. 109-156.
21. West, J. C., and J. L. Douce: The Mechanism of Sub-harmonic Generation in a Feedback System, *J. IEE*, London, paper 1693 M, vol. B-102 (July, 1955), pp. 569-574.
22. West, J. C., J. L. Douce, and R. K. Livesley: The Dual Input Describing Function and Its Use in the Analysis of Non-linear Feedback Systems, *J. IEE*, London, vol. B-103 (July, 1956), pp. 463-474.
23. West, J. C.: "Analytical Techniques for Non-linear Control Systems," D. Van Nostrand Company, Inc., Princeton, N.J., 1960.

PROBLEMS

- 5-1. Calculate the TSIDFs in the cases of harmonically- and nonharmonically-related input sinusoids for the polynomial hard-spring characteristic given by

$$y = x + 0.3x^3$$

- 5-2. By employing the method of double Fourier series expansion, show that, for the half-wave linear rectifier,

$$y = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

the first three output Fourier coefficients are (Ref. 2)

$$P_{00} = \frac{4A}{\pi^2} [2E(k) - (1 - k^2)K(k)]$$

$$P_{01} = \frac{kA}{2}$$

$$P_{10} = \frac{A}{2}$$

- 5-3. Demonstrate that for the nonlinearity

$$y = \sin mx$$

the function $Y(ju)$ can be obtained in the form

$$Y(ju) = \frac{\pi}{j} [\delta(u - m) - \delta(u + m)]$$

and hence that the TSIDFs are given by

$$N_B(A, B) = \frac{2}{B} J_0(mA) J_1(mB)$$

$$N_A(A, B) = \frac{2}{A} J_0(mB) J_1(mA)$$

[Hint: Use the relationship

$$\frac{1}{\pi} \lim_{\sigma \rightarrow 0} \frac{\sigma}{\sigma^2 + \alpha^2} = \delta(\alpha)$$

where $\delta(\alpha)$ denotes the unit-impulse function of argument α .]

- 5-4. Derive Eqs. (5.1-44a) to (5.1-44c). Use the result to evaluate the TSIDFs for an odd square-law nonlinearity, $y = x|x|$.
- 5-5. Compute the approximate TSIDFs for an ideal relay. According to the exact TSIDF expressions, what is the range of validity of the approximation?
- 5-6. What is the upper limit on input frequency which you would recommend for the system of Fig. 5-1 such that one-third subharmonics are not to occur? Design a linear compensator to be located at station y such that *no* one-third subharmonics can occur.

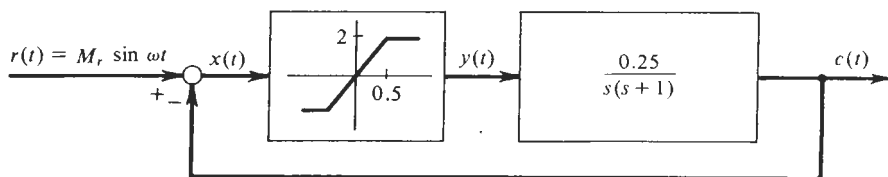


Figure 5-1 Nonlinear system with saturation.

- 5-7. Demonstrate that the TSIDF loci $-1/N_B(A, B, \frac{1}{3}, \theta)$ for a cubic nonlinearity are circles described by ($k = B/A$)

$$\text{Center} = \left(-\frac{4}{3A^2} \frac{2+k^2}{k^4+3k^2+4}, 0 \right)$$

$$\text{Radius} = \frac{4}{3A^2} \frac{k}{k^4+3k^2+4}$$

and plot these loci for fixed A at several values of k in the ranges $0 < k < 1$ and $1 < k < 8$.

- 5-8. By noting that any piecewise-linear limiter with drive levels $\pm D$ and input breakpoints $\pm \delta$ can be represented in terms of a unity drive level, unity breakpoint limiter, preceded by a linear gain $1/\delta$ and followed by a gain D , show that, as a consequence of Fig. 5.2-5, the rule for avoidance of subharmonic oscillations in systems containing a forward-path ideal relay is as illustrated in Fig. 5-2.

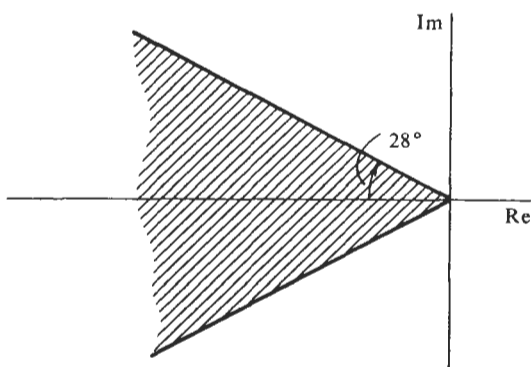


Figure 5-2 The 28° rule for elimination of subharmonics in systems containing an ideal relay.

5-9. The linear elements in an ideal-relay system are characterized by

$$L(s) = \frac{60(0.25s + 1)^2}{s(s + 1)^2(0.05s + 1)^2}$$

Is the high-frequency limit cycle indicated on an amplitude-phase plot of $L(j\omega)$ and $-1/N(A)$ a mode which can exist by itself?

- 5-10. A certain surface ship has a base-rate-motion-isolated rocket-launching platform. The elements of the platform roll-rate control loop are as illustrated in Fig. 5-3. Under an extremely rough sea condition the roll-rate control-loop disturbing torque signal T_d is given by $T_d = 1.5 \sin t$. Calculate the amplitude of the noise torquing signal at station T due to a 1-volt 400-cps stray pickup signal n .

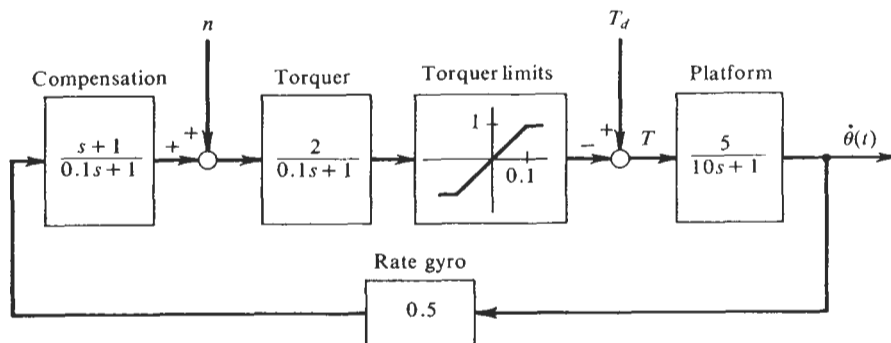


Figure 5-3 Launching platform roll-rate control loop.

- 5-11. A ground-based missile contains an inertial navigation system with a coarse-leveling loop as illustrated in Fig. 5-4. Instrument errors and gimbal servo dynamics have been neglected; the inertial platform is treated as a pure integrator. The quantity θ is the tilt angle (radians), measured with respect to the local vertical. Because of the effect of wind gusting on the missile structure, the accelerometer senses a disturbance input $d = 0.2g \sin 2\pi t + 0.02g \sin 8\pi t$. The accelerometer has a "notch" of $0.1g$ width and an otherwise linear characteristic of slope 1 volt/g. Calculate θ in terms of two sinusoidal components, and thereby determine the accuracy of this coarse-leveling mode. How can the accuracy be improved?

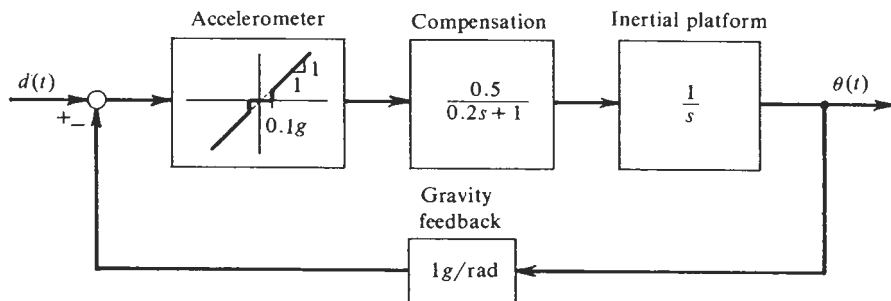


Figure 5-4 Missile inertial-navigation-system erection loop.

- 5-12. For the system of Example 4.4-1, compute the damping factor associated with a limit cycle perturbation by means of the incremental-input describing function. To facilitate an analytic solution to this problem, note that the graphical construction employed can be interpreted as the solution for the intersections between a line and a circle. Compute the slope of the "line" [that is, $L(\sigma + j\omega_0)^{-1}$] at the point where $\sigma = 0$, and thus deduce the approximate result, $\sigma \approx -\zeta\omega_n/(1 + 4\zeta^2)$ for large ζ .
- 5-13. Derive the incremental-input describing function associated with the k th incremental component of the nonlinearity input

$$x = A \sin \omega t + \sum_{k=1}^n \epsilon_k \sin(\omega t + \theta_k)$$

and put it in the standard form

$$N_{ik}(A, \theta_k) = N(A) + \frac{AN'(A)}{2} (1 + e^{-j2\theta_k})$$

- 5-14. A linear time-varying differential equation of considerable physical importance is the Mathieu equation,

$$\ddot{x} + [\omega_n^2 - M \sin(t + \theta)]x = 0$$

A block diagram appropriate for this system appears in Fig. 5-5. We wish to study the stability of the parametrically excited oscillations of this system by describing function methods.¹

¹ A. Leonhard, Extension of the Describing Function Method to the Investigation of Parametric Oscillations, *Proc. IFAC*, Basel, Switzerland (August, 1963), pp. 187/1-7.

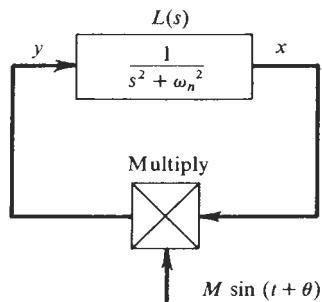


Figure 5-5 Block diagram of the Mathieu equation.

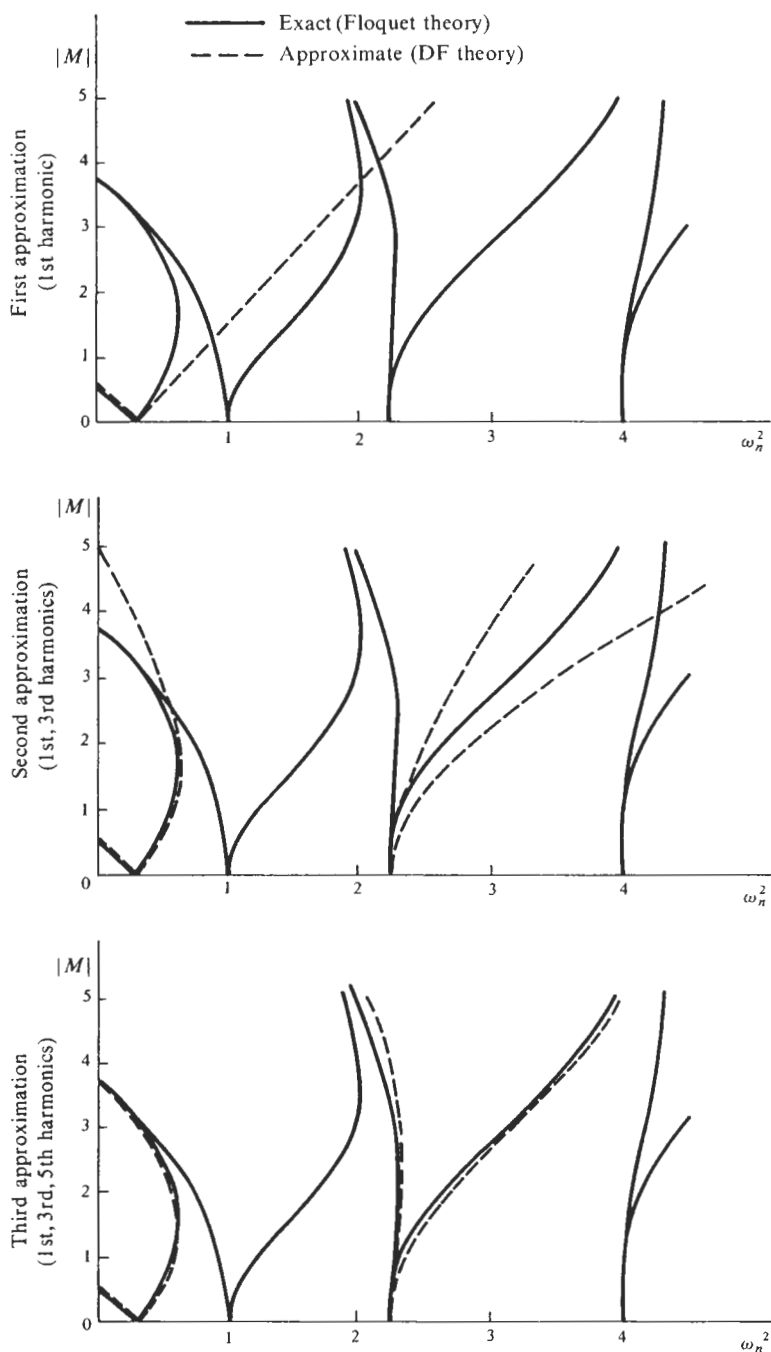


Figure 5-6 Stability boundaries for the Mathieu equation ($\omega = \frac{1}{2}$).

- (a) Assuming the system to be in a state of neutral stability given by

$$x = A \sin \omega t$$

show that an appropriate DF model for the multiplicative gain is

$$N(A, M, \theta) = \frac{M}{2} e^{j(\theta + \pi/2)}$$

(This requires the argument that only $\omega = \frac{1}{2}$ yields solutions of consequence.) Next, invoke the DF requirement for self-sustained system oscillation to find

$$M = 2[\omega_n^2 - (\frac{1}{2})^2]$$

This equation for the stability boundary is conventionally plotted in the form $|M|$ versus ω_n^2 . Such has been illustrated in Fig. 5-6 against a background of the exact stability boundaries for the Mathieu equation as determined by Floquet theory.

- (b) Now consider the more representative steady-state self-oscillation waveform

$$x = A_1 \sin \omega t + A_3 \sin (3\omega t + \psi)$$

Show that the appropriate TSIDF model for the multiplicative gain is

$$N_1(A_1, A_3, M, \theta) = \frac{M}{2} e^{j(\theta + \pi/2)} + \frac{MA_3}{2A_1} e^{j(\theta - \psi + \pi/2)}$$

$$N_3(A_1, A_3, M, \theta) = -\frac{MA_1}{2A_3} e^{j(\theta - \psi + \pi/2)}$$

and derive, as the equation for the stability boundary,

$$\left[\frac{M}{2} e^{j(\theta + \pi/2)} + \left(\frac{M}{2} \right)^2 L(j\frac{3}{2}) \right] L(j\frac{1}{2}) = 1$$

with the corresponding solution

$$M = \pm [\omega_n^2 - (\frac{3}{2})^2] \pm \sqrt{[\omega_n^2 - (\frac{3}{2})^2]^2 + 4[\omega_n^2 - (\frac{3}{2})^2][\omega_n^2 - (\frac{1}{2})^2]}$$

This result is plotted in Fig. 5-6, where, in addition to an improvement in the shape of the approximately derived solution, *two* distinct V-shaped boundaries appear.

- (c) Finally, consider the case in which the first, third, and fifth harmonics are included in the self-oscillation waveform model, viz.,

$$x = A_1 \sin \omega t + A_3 \sin (3\omega t + \psi_3) + A_5 \sin (5\omega t + \psi_5)$$

Determine the *three* appropriate describing function gains of the multiplicative term, and demonstrate that the stability-boundary equation takes the form

$$\left[\frac{M}{2} e^{j(\theta + \pi/2)} + \frac{(M/2)^2}{1/[L(j\frac{3}{2})] - (M/2)^2 L(j\frac{5}{2})} \right] L(j\frac{1}{2}) = 1$$

The solution of this equation has been plotted in Fig. 5-6, where it is clearly seen that the describing function analysis yields an excellent approximation to the exact solution. A third V-shaped boundary arises from the point $\omega_n^2 = 6.25$, although this is not shown.

- 5-15. The incorporation of a dc term in the self-oscillation waveform of the parametric oscillator of Prob. 5-14 yields additional stability-boundary solutions of interest.

(a) Choose

$$x = A_0 + A_1 \sin \omega t$$

and by following the lead of Prob. 5-14, develop the dc and first-harmonic gains of the multiplicative term (now using the argument that only $\omega = 1$ yields solutions of interest), and by manipulating, obtain the results

$$M \text{ is arbitrary} \quad \text{for } \omega_n^2 = 1$$

$$M = \sqrt{2\omega_n^2(\omega_n^2 - 1)} \quad \text{for } \omega_n^2 > 1$$

(b) Show that for the three-term waveform model

$$x = A_0 + A_1 \sin \omega t + A_2 \sin (2\omega t + \psi_2)$$

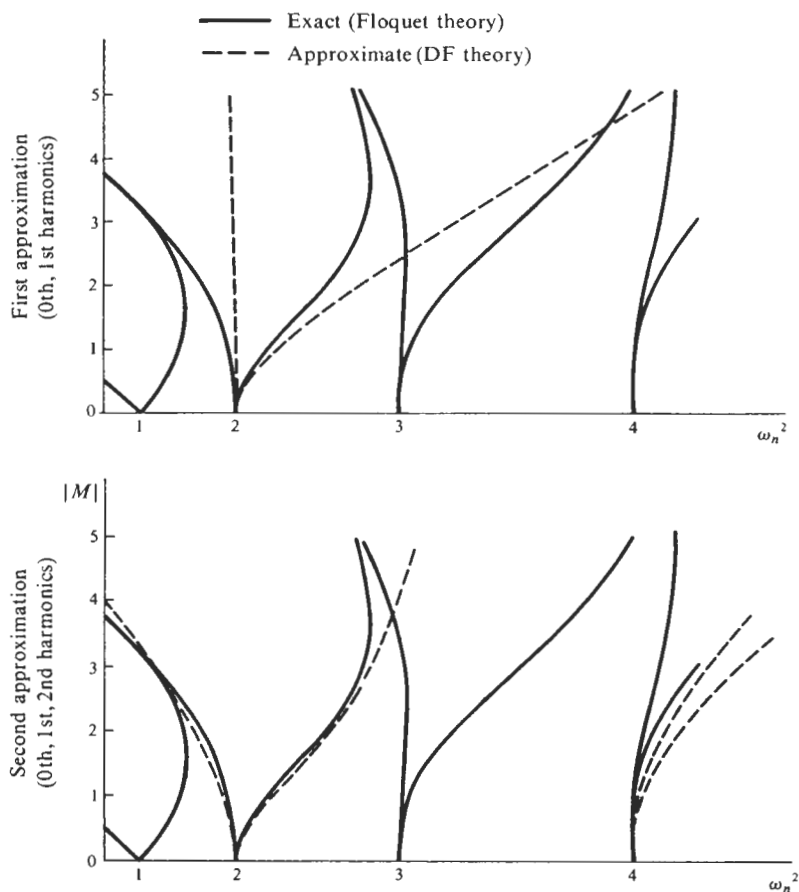


Figure 5-7 Stability boundaries for the Mathieu equation ($\omega = 1$).

the stability-boundary solutions are

$$M = \begin{cases} 2\sqrt{\omega_n^2 - 1}\sqrt{\omega_n^2 - 4} & \text{for } 1 > \omega_n^2 > 4 \\ 2\sqrt{\frac{\omega_n^2(\omega_n^2 - 1)(\omega_n^2 - 4)}{3\omega_n^2 - 8}} & \text{for } 1 < \omega_n^2 < \frac{8}{3} \quad \text{and} \quad \omega_n^2 > 4 \end{cases}$$

These approximate solutions are plotted in Fig. 5-7. Note that the boundaries arising under the assumption $\omega = 1$ with a self-oscillation waveform dc term are located in between the boundaries determined under the assumption $\omega = \frac{1}{2}$ without a dc term.

- 5-16. Make an argument which demonstrates that the area between the stability boundaries and the abscissa corresponds to the region of *stable* operation, whereas that above the boundaries denotes *unstable* operation, for the Mathieu equation.
- 5-17. For the system of Fig. 5-8, find the region in the space of M_r and ω in which the system would display a stable response to the sinusoidal input, and would not limit-cycle.

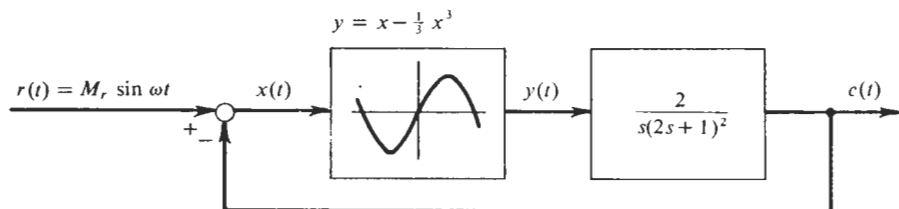


Figure 5-8

- 5-18. Write a general expression for the steady-state response $c(t)$ to the sinusoidal input for the system of Fig. 5-9—an expression valid for all combinations of input amplitude and frequency resulting in a response at x which is of small amplitude relative to the amplitude of the limit cycle at x .

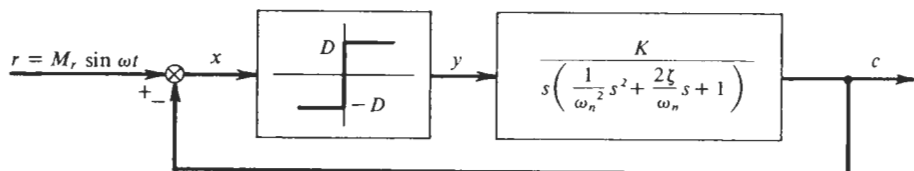


Figure 5-9

- 5-19. For a cubic nonlinearity, $y = x^3$, show that the *three*-sinusoid-input describing function gains corresponding to the input

$$x(t) = A_1 \sin \omega t + A_3 \sin (3\omega t + \theta_3) + A_5 \sin (5\omega t + \theta_5)$$

are:¹

$$N_{A_1} = \frac{3}{4} \left\{ \left[A_1^2 + 2A_3^2 + 2A_5^2 - A_1A_3 \cos \theta_3 \right. \right. \\ \left. \left. - \frac{A_3^2A_5}{A_1} \cos (2\theta_3 - \theta_5) - 2A_3A_5 \cos (\theta_5 - \theta_3) \right] \right. \\ \left. + j \left[-A_1A_3 \sin \theta_3 + \frac{A_3^2A_5}{A_1} \sin (2\theta_3 - \theta_5) - 2A_3A_5 \sin (\theta_5 - \theta_3) \right] \right\}$$

$$N_{A_3} = \frac{3}{4} \left\{ \left[2A_1^2 + A_3^2 + 2A_5^2 - \frac{A_1^3}{3A_3} \cos \theta_3 \right. \right. \\ \left. \left. - \frac{A_1^2A_5}{A_3} \cos (\theta_5 - \theta_3) + 2A_1A_5 \cos (\theta_5 - 2\theta_3) \right] \right. \\ \left. + j \left[\frac{A_1^3}{3A_3} \sin \theta_3 - \frac{A_1^2A_5}{A_3} \sin (\theta_5 - \theta_3) + 2A_1A_5 \sin (\theta_5 - 2\theta_3) \right] \right\}$$

$$N_{A_5} = \frac{3}{4} \left\{ \left[2A_1^2 + 2A_3^2 + A_5^2 - \frac{A_1^2A_3}{A_5} \cos (\theta_5 - \theta_3) + \frac{A_1A_3^2}{A_5} \cos (2\theta_5 - \theta_3) \right] \right. \\ \left. + j \left[-\frac{A_1^2A_3}{A_5} \sin (\theta_5 - \theta_3) + \frac{A_1A_3^2}{A_5} \sin (2\theta_5 - \theta_3) \right] \right\}$$

5-20. A nonlinear feedback system is comprised of a cubic nonlinearity, $y = x^3$, and fourth-order linear elements

$$L(s) = \frac{10s^2}{[(s + 0.01)^2 + 0.9^2][(s + 0.01)^2 + 1.1^2]}$$

- (a) Show that DF arguments lead to the conclusion that the system does not limit-cycle. Based upon consideration of the shape of the $L(j\omega)$ locus, would you suspect the DF conclusion to be in error?
- (b) Now assume a nonlinearity input of the form

$$x(t) = A_1 \sin \omega t + A_3 \sin (3\omega t + \theta)$$

and by simultaneously satisfying the four conditions

$$\begin{aligned} |N_{A_1}| |L(j\omega)| &= 1 & \angle N_{A_1} + \angle L(j\omega) &= \pm \pi \\ |N_{A_3}| |L(j3\omega)| &= 1 & \angle N_{A_3} + \angle L(j3\omega) &= \pm \pi \end{aligned}$$

show that this TSIDF argument *does* predict a limit cycle, given by

$$x(t) = 1.13 \sin 0.608t + 0.605 \sin (1.824t + 0.083)$$

This oscillator problem is one of a class of problems studied by Fitts, who showed that *several* modes of oscillation can indeed occur. The *first mode*, in which the predominant harmonic is the third, was measured as

$$x(t) = 1.13 \sin 0.604t + 0.548 \sin (1.812t + 0.093)$$

- (c) Outline how you would solve for the so-called *second mode* of oscillation, in which the first, third, and fifth harmonics appear?

¹ Fitts, R. E.: Linearization of Nonlinear Feedback Systems, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Mass., June, 1966.

6 DUAL-INPUT DESCRIBING FUNCTION (DIDF)

6.0 INTRODUCTION

The two-sinusoid-input describing function (TSIDF) of Chap. 5 is certainly again brought to mind by the title of this chapter, as are all other describing functions which simultaneously accommodate two nonlinearity input waveforms. Because of the specific utility of the particular describing function of this class to be discussed presently, however, we reserve for it the otherwise general appellation *dual-input describing function*, and related abbreviation DIDF.

TSIDF application to nonlinear control systems is conceptually limited by requiring that only sinusoids be described in the nonlinearity test input. For this reason an alternative linearization accommodating two inputs was sought by researchers interested, among other things, in the approximate forced response behavior of certain nonlinear systems. The DIDF, as we shall now see, is a physically motivated linearization of a nonlinearity which readily permits study, among other things, of the forced responses of *limit cycling* nonlinear systems. In what follows it is required that command

inputs do not cause the limit cycle to terminate, an assumption that can be verified when under suspicion. The nonlinearity input is then comprised of the limit cycle and a component due to the command input. By assuming that the component due to the command input varies little during a limit cycle period, one can formulate a nonlinearity linearization similar in concept to the TSIDF, *but far simpler to calculate*. Hence the DIDF model input waveform is a bias plus a sinusoid, the latter component effectively serving to linearize the nonlinearity gain to the former.

The idea of linearizing nonlinear characteristics by means of an additive sinusoid is not a new one. MacColl (Ref. 12) described the use of such a signal in a motor-drive system incorporating a relay. Loeb (Ref. 10) has suggested that *any* nonlinear system can be treated in this manner. Lozier (Ref. 11) is credited with a method of treating oscillating control systems using this linearization, an interpretation later independently arrived at by Li and Vander Velde (Ref. 9) in connection with limit cycling adaptive feedback control system applications. Other papers by Gelb (Refs. 3-5) followed this interpretation and further developed the dynamic characterization of limit cycling systems. Oldenburger experimentally discovered the effects of an *additive* high-frequency low-amplitude input to a control system, and subsequently provided analytical justification via DIDF considerations for these effects in several papers on "signal stabilization" (Refs. 16-18). Popov (Ref. 20) and later Popov and Pal'tov (Ref. 21) have published books in which DIDF harmonic linearization is treated.

In this chapter, following general formulation and calculation of several specific DIDFs, the forced response of a class of limit cycling systems is considered. The role of poles and zeros in the description of limit cycling systems is established in this context, and application of the theory is made to a class of adaptive control systems. Adaptive roll control for a missile is considered as an example in this class. It is then shown that limit cycles in systems with an *asymmetric* nonlinearity are easily determined. The uses of artificial dither for the compensatory purposes of linearization and signal stabilization are treated. Finally, a brief TSIDF derivation is developed in DIDF terms.

Before embarking upon a discussion of DIDF formulation, we digress to examine a system quite different from those discussed in the remainder of this chapter.

Example 6.0-1 Apart from limit cycling systems, the DIDF model input waveform can arise in quite a number of different ways. For illustration, consider the problem of determining the near-circular-orbit period of an earth satellite. The coupled radial and tangential force equations are

$$\ddot{r} - r\dot{\psi}^2 = -\frac{\mu}{r^2} \quad (6.0-1)$$

and

$$r\ddot{\psi} + 2\dot{r}\dot{\psi} = 0 \quad (6.0-2)$$

or since Eq. (6.0-2) can be written as $(1/r) d/dt(r^2\dot{\psi}) = 0$, we can write

$$r^2\dot{\psi} = h \quad (6.0-3)$$

where r and ψ are as defined in Fig. 6.0-1, μ is a constant defining the specific gravitational force, and h is the constant specific angular momentum of the orbit. Eliminating $\dot{\psi}$ from Eq. (6.0-1) by means of Eq. (6.0-3) gives

$$\ddot{r} - \frac{h^2}{r^3} = -\frac{\mu}{r^2} \quad (6.0-4)$$

A natural approximate solution to this conservative equation proceeds as follows: First, the geometry of the problem suggests a solution of the form ($\psi = \omega t$)

$$r = R + \delta_R \cos \psi \quad (6.0-5)$$

where R and δ_R are constant. Substituting this expression in Eq. (6.0-4) gives

$$-\omega^2 \delta_R \cos \psi - \frac{h^2}{R^3} \left(1 + \frac{\delta_R}{R} \cos \psi\right)^{-3} = -\frac{\mu}{R^2} \left(1 + \frac{\delta_R}{R} \cos \psi\right)^{-2} \quad (6.0-6)$$

The assumption of a near-circular orbit implies that $\delta_R/R \ll 1$. Correspondingly, the nonlinear terms in Eq. (6.0-6) can be approximated by appropriate first-order expansions, viz.,

$$-\omega^2 \delta_R \cos \psi - \frac{h^2}{R^3} \left(1 - \frac{3\delta_R}{R} \cos \psi\right) \approx -\frac{\mu}{R^2} \left(1 - \frac{2\delta_R}{R} \cos \psi\right) \quad (6.0-7)$$

Balancing harmonics on each side of this equation (a concept which is integral to DF analysis of previous chapters) gives the relationships

$$R = \frac{h^2}{\mu} \quad \text{and} \quad \omega^2 = \frac{3h^2}{R^4} - \frac{2\mu}{R^3} \quad (6.0-8)$$

Solution of these two equations for the orbit period T finally yields

$$T = \frac{2\pi}{\omega} = 2\pi \frac{h^3}{\mu^2} \quad (6.0-9)$$

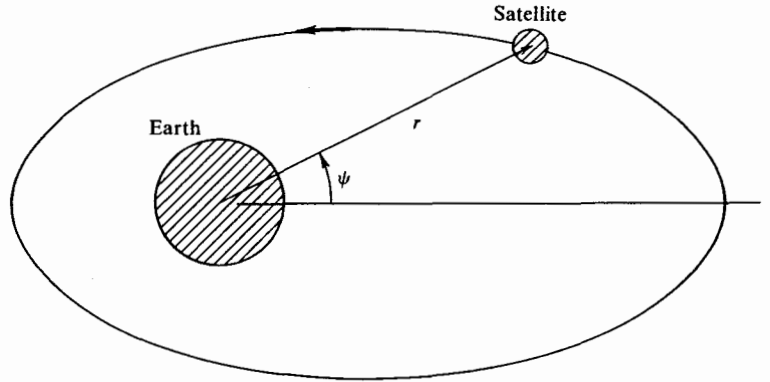


Figure 6.0-1 Earth-satellite geometry.

This compares favorably with the exact result (where orbit eccentricity = δ_R/R)

$$T_{\text{exact}} = 2\pi \frac{h^3}{\mu^2} \left[1 - \left(\frac{\delta_R}{R} \right)^2 \right]^{-\frac{3}{2}} \quad (6.0-10)$$

The ease with which this approximate solution is obtained is characteristic of the harmonic linearization approach. This accounts, in large measure, for the popularity and widespread use of describing function techniques.

6.1 MATHEMATICAL FORMULATION OF THE DIDF

MOTIVATION

Consider the system of Fig. 6.1-1 to be in a limit cycle state of period T . Now let a "slowly varying," but otherwise arbitrary, input which satisfies the inequality

$$T \left| \frac{dr(t)}{dt} \right| \ll A \quad (6.1-1)$$

be applied, where A is the amplitude of the limit cycle oscillation at the input to N . From this inequality it can be seen that a slowly varying function is taken as one which changes relatively little with respect to A over the period T . This coarse, somewhat restrictive definition is sufficient for our present purposes.

In this case a typical set of input and output waveforms are illustrated in Fig. 6.1-2a, where we see that the output follows the input, on the average, to within some dynamic following error. With the exception of an additive limit cycle component this is the input-output relationship ordinarily encountered in control systems of all types, where the function of the system is simply to reproduce the input waveform at a higher power level at the output. One cycle of the associated nonlinearity input waveform is shown in Fig. 6.1-2b. This model is the starting point for DIDF derivation. It plays a role completely analogous to the single-sinusoid model of DF usage, or the two-sinusoid model of TSIDF usage. Let us also note that precisely the same model is arrived at in general consideration of the limit cycle behavior of a system with an asymmetric nonlinearity, for in this case the

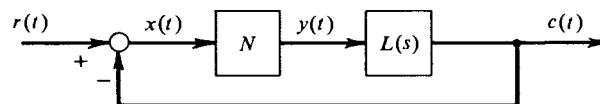


Figure 6.1-1 Limit-cycling system.

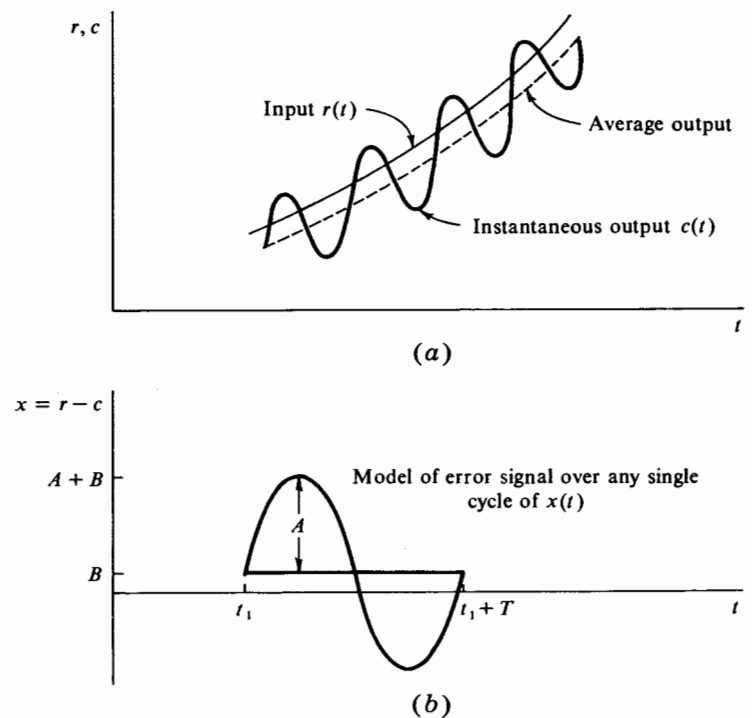


Figure 6.1-2 (a) Typical input and output waveforms with (b) associated error-signal model.

closed loop can develop a bias input to N in addition to the limit cycle. Other physically significant situations in which a true bias appears at the nonlinearity input, in addition to a sinusoid, include a system with an integration in the linear part responding to a ramp input while limit cycling, a limit cycling system responding to a constant disturbance input, and a non-limit-cycling system responding to a sinusoidal command and a constant disturbance input.

FORMULATION

The input to the nonlinearity is taken to be

$$x(t) = B + A \sin(\omega t + \theta) \quad (6.1-2)$$

From the statistical point of view taken in Chap. 1, B , A , and ω are considered determined by the nature of the system and its inputs. Thus the only random variable in the characterization of $x(t)$ is the phase angle θ ,

which has the uniform distribution over one cycle. The expectations which appear in the expressions for the describing functions are in this case just single integrations over a 2π interval in θ .

The expressions for the describing functions resulting from the uncoupled relations of Eq. (1.5-21) are clearly applicable in this case. The nonlinearity input is the sum of two components: $x_1(t)$ may be taken to be the bias, and $x_2(t)$ the sinusoid. With B a deterministic quantity, and the sinusoid $x_2(t)$ having zero mean, these two input components are obviously uncorrelated. With the describing function for the sinusoidal input component interpreted as the sum of an in-phase plus quadrature gain, as was discussed in Chap. 1 and again in Chap. 2, the output of the optimum quasi-linear approximator to a nonlinearity with this input is

$$y_a(t) = N_B B + n_p A \sin(\omega t + \theta) + n_q A \cos(\omega t + \theta) \quad (6.1-3)$$

The required gains, or describing functions, can readily be calculated for quite general nonlinearities. In recognition of this we indicate in the notation a dependence of nonlinearity output on the nonlinearity input $x(t)$ and its derivative.

$$y(t) = y[x(t), \dot{x}(t)] \quad (6.1-4)$$

This, however, is not to be taken as a restriction on the form of nonlinearities which can be treated. The only requirement is that one be able to define the nonlinearity output as the phase of the input sinusoid traverses a full cycle. The describing functions will in general depend on the bias level, the amplitude of the sinusoid, and its frequency.

The approximating gain to the bias input component is given by Eq. (1.5-27).

$$\begin{aligned} N_B(A, B, \omega) &= \frac{1}{B} \overline{y(0)} \\ &= \frac{1}{2\pi B} \int_0^{2\pi} y(B + A \sin \theta, A\omega \cos \theta) d\theta \end{aligned} \quad (6.1-5)$$

The real and imaginary parts of the complex approximating gain to the sinusoidal input component are given by Eqs. (1.5-36).

$$\begin{aligned} n_p(A, B, \omega) &= \frac{2}{A} \overline{y(0) \sin \theta} \\ &= \frac{1}{\pi A} \int_0^{2\pi} y(B + A \sin \theta, A\omega \cos \theta) \sin \theta d\theta \end{aligned} \quad (6.1-6)$$

$$\begin{aligned} n_q(A, B, \omega) &= \frac{2}{A} \overline{y(0) \cos \theta} \\ &= \frac{1}{\pi A} \int_0^{2\pi} y(B + A \sin \theta, A\omega \cos \theta) \cos \theta d\theta \end{aligned} \quad (6.1-7)$$

These describing functions, defined by the property of minimum mean-squared approximation error, are seen to be identical with the result of expanding the nonlinearity output, for the assumed input, into its Fourier series, and relating the corresponding input and output terms. Thus the mean output is related to the mean input by the gain N_B , and the fundamental output component is related to the sinusoidal input component by the complex gain N_A , which has real and imaginary parts n_p and n_q , respectively. This is a specialization of more general properties derived in Chap. 1. For any nonlinearity input, N_B is always the gain which equates the mean output of the quasi-linear approximator to the mean output of the nonlinearity. And for any nonlinearity input which does not include harmonically related sinusoids, N_A always represents the amplitude and phase relation between a sinusoidal input component and the harmonic component of the same frequency in the expectation of the nonlinearity output, computed by averaging over all random parameters except the phase angle of that sinusoid. In this case there are no random parameters other than the phase angle of the input sinusoid; so the expectation of the nonlinearity output referred to above is just the output itself.

It is evident from this formulation that in the limit, as $B \rightarrow 0$, $n_p(A, B, \omega)$ and $n_q(A, B, \omega)$ approach $n_p(A, \omega)$ and $n_q(A, \omega)$, the DF characterization of N . Also in the limit, as $B \rightarrow 0$, the gain to the bias approaches a value independent of B , which is termed the *incremental-input describing function*. It is defined by

$$\begin{aligned} N_B(A, \omega) &= \lim_{B \rightarrow 0} [N_B(A, B, \omega)] \\ &= \lim_{B \rightarrow 0} \left[\frac{1}{2\pi B} \int_0^{2\pi} y(B + A \sin \theta, A \omega \cos \theta) d\theta \right] \end{aligned} \quad (6.1-8)$$

Intuition leads one to suspect that this form of the incremental-input describing function is equivalent to that presented in Sec. 5.5 for non-harmonically related input sinusoids (i.e., asynchronous case). This equivalence is easily demonstrated.

Example 6.1-1 *Proof of the equality $N_i(A) = N_B(A, 0)$.* Application of L'Hospital's rule to the indeterminate form of the incremental-input describing function [Eq. (6.1-8)], specialized to static single-valued nonlinearities, yields

$$\begin{aligned} N_B(A, 0) &= \frac{1}{2\pi} \lim_{B \rightarrow 0} \left[\frac{1}{B} \int_0^{2\pi} y(B + A \sin \theta) d\theta \right] \\ &= \frac{1}{2\pi} \int_0^{2\pi} y'(A \sin \theta) d\theta \end{aligned} \quad (6.1-9)$$

where $y'(A \sin \theta)$ stands for the derivative of y with respect to its argument, computed at the point where the argument takes the value $A \sin \theta$. It is to be noted that Eq. (6.1-9)

provides a useful means for computing $N_B(A,0)$ directly, without the intermediate step of an explicit limiting process (that is, $B \rightarrow 0$). Only in the case of discontinuous nonlinear characteristics, where y' contains impulse functions, is it again necessary to consider a limiting process (cf. Prob. 6-2).

Next, $N_i(A)$ is determined in the following way: By definition

$$\begin{aligned} N_i(A) &= \frac{1}{\pi A} \int_0^{2\pi} y(A \sin \theta) \sin \theta \, d\theta + \frac{A}{2} \frac{d}{dA} \left[\frac{1}{\pi A} \int_0^{2\pi} y(A \sin \theta) \sin \theta \, d\theta \right] \\ &= \frac{1}{2\pi A} \int_0^{2\pi} y(A \sin \theta) \sin \theta \, d\theta + \frac{1}{2\pi} \int_0^{2\pi} y'(A \sin \theta) \sin^2 \theta \, d\theta \end{aligned} \quad (6.1-10)$$

The first integral can be integrated by parts to yield

$$\begin{aligned} \int_0^{2\pi} y(A \sin \theta) \sin \theta \, d\theta &= -y(A \sin \theta) \cos \theta \Big|_0^{2\pi} + \int_0^{2\pi} A y'(A \sin \theta) \cos^2 \theta \, d\theta \\ &= A \int_0^{2\pi} y'(A \sin \theta) \cos^2 \theta \, d\theta \end{aligned} \quad (6.1-11)$$

Inserting Eq. (6.1-11) into Eq. (6.1-10) gives

$$\begin{aligned} N_i(A) &= \frac{1}{2\pi} \int_0^{2\pi} y'(A \sin \theta) \cos^2 \theta \, d\theta + \frac{1}{2\pi} \int_0^{2\pi} y'(A \sin \theta) \sin^2 \theta \, d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} y'(A \sin \theta) \, d\theta \end{aligned} \quad (6.1-12)$$

whence it is concluded that

$$N_i(A) = N_B(A,0) \quad (6.1-13)$$

DISCUSSION

The DIDF incremental-input describing function representation of N leads to an extremely simple and very useful description of the input-output dynamics of limit cycling systems. We shall thoroughly explore this description, following several example DIDF calculations.

The approximation used in that analysis, in addition to the normal describing function approximation, is that the bias function employed in DIDF calculation is generalized to represent arbitrary functions of time within a restricted class. This class evidently can include functions which are *slowly varying*; so over any one period of the sinusoid the function appears essentially constant. But the relation just derived, Eq. (6.1-13), states that the effective gain of the nonlinearity to a small sinusoid of any frequency in the presence of another sinusoid is equal to the effective gain of the nonlinearity to a small bias in the presence of a sinusoid. This suggests the possibility that the effective gain of the nonlinearity to a small signal of arbitrary form in the presence of a larger sinusoid may be the same, equal to

the incremental-input describing function. This suggestion is given further strength by the results of Sec. 1.5, which showed that the effective gain of a static, single-valued nonlinearity to a small signal in the presence of *any* other independent input signals is the same in the limit as the small signal approaches zero, whether it be a bias, a sinusoid, or a random process. Thus we also include in the class of arbitrary functions for which the incremental-input describing function can be employed, functions which are *small*. Usually this "smallness" condition requires that the arbitrary signal be small compared with another signal at the nonlinearity input, such as the sinusoid in the DIDF formulation. In some instances, however, the smallness must be with respect to the characteristics of the nonlinearity (cf. Example 5.5-3).

6.2 DIDF CALCULATION

Because of the great similarity between DF and DIDF calculations, only a few selected DIDF calculations are presented. These are chosen from the ranks of frequency-independent nonlinearities which are piecewise-linear symmetric, piecewise-linear asymmetric, and polynomial, as well as frequency-dependent nonlinearities. First, let us examine a general calculation of some importance.

Consider a static, but otherwise arbitrary, nonlinearity $y = y(x)$. Using the obvious relationship $d(B + A \sin \psi) = A \cos \psi d\psi$, it follows that $n_q(A, B)$ can be written as

$$\begin{aligned}
 n_q(A, B) &= \frac{1}{\pi A} \int_0^{2\pi} y(B + A \sin \psi) \cos \psi d\psi \\
 &= \frac{1}{\pi A^2} \oint y(B + A \sin \psi) d(B + A \sin \psi) \\
 &= \frac{1}{\pi A^2} \oint y(x) dx \\
 &= \frac{-S}{\pi A^2}
 \end{aligned} \tag{6.2-1}$$

(Note: We have returned to the dummy variable ψ to be consistent with previous chapters)

where S is the area enclosed by $y(x)$ as x varies through a complete cycle between the limits $B + A$ and $B - A$. The reader can convince himself with the aid of a simple sketch that the minus sign in Eq. (6.2-1) is indeed appropriate.

Several conclusions of importance can be drawn from this result. The most obvious is that $n_q(A, B) = 0$ for *any* memoryless nonlinearity, since here $S = 0$. Also, since $n_q(A, B)$ is equal to $n_q(A)$ when B is set equal to zero, it follows that Eq. (6.2-1) is suitable for calculating the phase-shifting component of the DF, $n_q(A)$.

Now we proceed to specific nonlinearities.

IDEAL RELAY

For this nonlinearity the input and output waveforms are illustrated in Fig. 6.2-1. One can see clearly that, because of the presence of a positive B , the points at which the nonlinearity output switches between D and $-D$ are altered in such a way that a net positive bias component results in y . From Eqs. (6.1-6) and (6.1-7), in which frequency dependence is dropped, we get for this *static* nonlinearity

$$\begin{aligned}
 n_p(A, B) &= \frac{1}{\pi A} \int_0^{2\pi} y(B + A \sin \psi) \sin \psi \, d\psi \\
 &= \frac{1}{\pi A} \left[\int_0^{\pi+\psi_1} D \sin \psi \, d\psi + \int_{\pi+\psi_1}^{2\pi-\psi_1} (-D) \sin \psi \, d\psi \right. \\
 &\quad \left. + \int_{2\pi-\psi_1}^{2\pi} D \sin \psi \, d\psi \right] \\
 &= \frac{4D}{\pi A} \cos \psi_1 \\
 &= \frac{4D}{\pi A} \sqrt{1 - \left(\frac{B}{A}\right)^2} \tag{6.2-2}
 \end{aligned}$$

$$\begin{aligned}
 \text{and} \quad n_q(A, B) &= \frac{1}{\pi A} \int_0^{2\pi} y(B + A \sin \psi) \cos \psi \, d\psi \\
 &= \frac{1}{\pi A} \left[\int_0^{\pi+\psi_1} D \cos \psi \, d\psi + \int_{\pi+\psi_1}^{2\pi-\psi_1} (-D) \cos \psi \, d\psi \right. \\
 &\quad \left. + \int_{2\pi-\psi_1}^{2\pi} D \cos \psi \, d\psi \right] \\
 &= 0
 \end{aligned}$$

This last result is obtained with even greater ease by inspection of Eq. (6.2-1). The limit cycle DIDF is indeed non-phase-shifting, as one expects in the

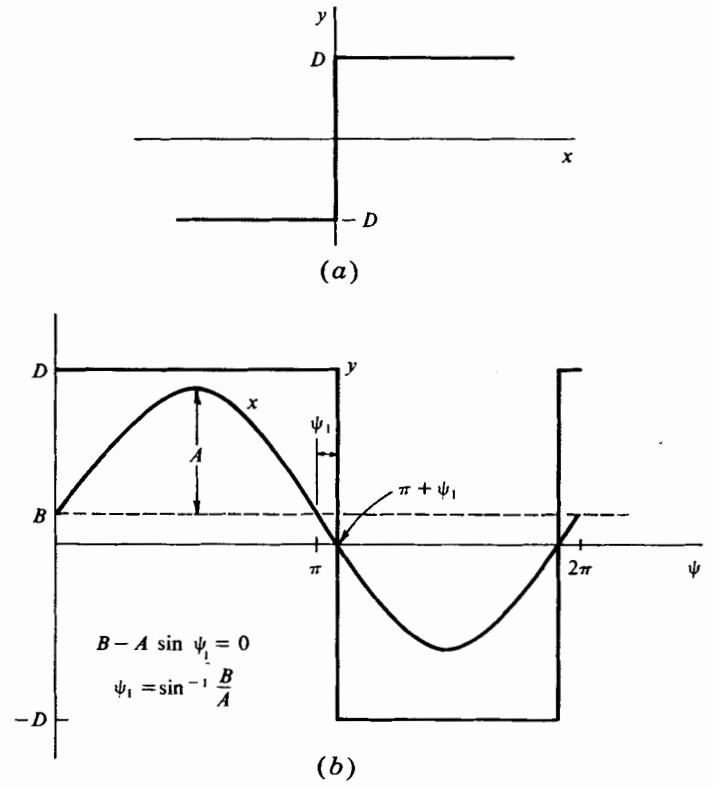


Figure 6.2-1 (b) Input and output waveforms for the ideal relay (a).

case of this memoryless nonlinearity. From Eq. (6.1-5) the signal DIDF is easily determined as

$$\begin{aligned}
 N_B(A, B) &= \frac{1}{2\pi B} \int_0^{2\pi} y(B + A \sin \psi) d\psi \\
 &= \frac{1}{2\pi B} \left[\int_0^{\pi + \psi_1} D d\psi + \int_{\pi + \psi_1}^{2\pi - \psi_1} (-D) d\psi + \int_{2\pi - \psi_1}^{2\pi} D d\psi \right] \\
 &= \frac{2D}{\pi B} \psi_1 \\
 &= \frac{2D}{\pi B} \sin^{-1} \frac{B}{A}
 \end{aligned} \tag{6.2-3}$$

The incremental-input describing function follows directly.

$$\begin{aligned} N_B(A) &= \lim_{B \rightarrow 0} N_B(A, B) \\ &= \frac{2D}{\pi A} \end{aligned} \quad (6.2-4)$$

The ease with which these calculations have been effected is striking by comparison with the corresponding TSIDF calculations of the previous chapter. As stated earlier, it is desired to let B represent a multitude of signals in addition to a simple bias. As an illustration of the case where B represents a sinusoid, let us examine the above ideal-relay results in order to get a feel for the conditions under which the DIDF is an acceptable substitute for the TSIDF. From the TSIDF ideal-relay calculation in the general case where the signal frequency is not a rational fraction of the limit cycle frequency, it is shown in Eqs. (5.1-19) that ($A > |B|$)

$$\begin{aligned} N_A(A, B) &= \frac{4D}{\pi A} \left[1 - \frac{1}{4} \left(\frac{B}{A} \right)^2 - \frac{3}{64} \left(\frac{B}{A} \right)^4 - \cdots \right] \\ N_B(A, B) &= \frac{2D}{\pi A} \left[1 + \frac{1}{8} \left(\frac{B}{A} \right)^2 + \frac{3}{64} \left(\frac{B}{A} \right)^4 + \cdots \right] \end{aligned} \quad (6.2-5)$$

Corresponding DIDF expressions [Eqs. (6.2-2) and (6.2-3)] can be expanded to yield

$$\begin{aligned} N_A(A, B) &= \frac{4D}{\pi A} \left[1 - \frac{1}{2} \left(\frac{B}{A} \right)^2 - \frac{1}{8} \left(\frac{B}{A} \right)^4 - \cdots \right] \\ N_B(A, B) &= \frac{2D}{\pi A} \left[1 + \frac{1}{6} \left(\frac{B}{A} \right)^2 + \frac{3}{40} \left(\frac{B}{A} \right)^4 + \cdots \right] \end{aligned} \quad (6.2-6)$$

Examining TSIDF calculations for the cases of rational, as well as irrational, frequency ratios and including *all* relative phase shifts leads to the conclusion that the ideal-relay DIDF result is within 5 percent of the TSIDF result under the conditions

$$\begin{aligned} \text{Amplitude-ratio condition:} \quad & \frac{B}{A} < \frac{1}{3} \\ \text{Frequency-ratio condition:} \quad & \gamma < \frac{1}{3} \end{aligned} \quad (6.2-7)$$

These rough quantitative statements apply as well to a wide range of common nonlinearities. They are therefore adopted as guideposts in a limit cycling system input-output characterization.

At this point another interpretation of the significance of the describing function $N_B(A, B)$ can be stated. Given that there is a limit cycle of amplitude A , this quantity explicitly accounts for the transmission of slowly varying signals through the nonlinearity in the presence of the limit cycle.

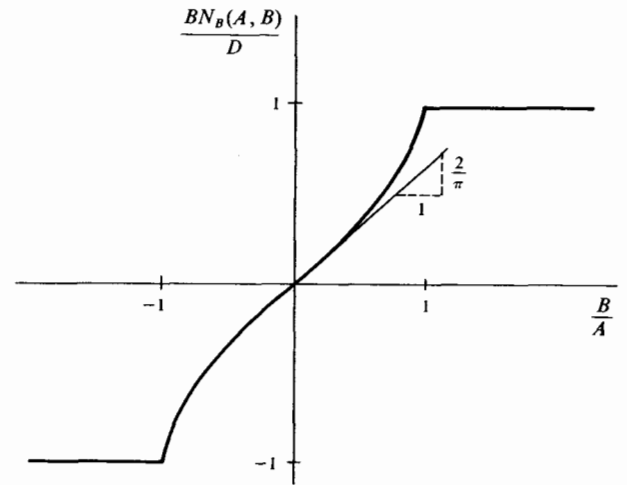


Figure 6.2-2 Equivalent nonlinear element for signal transmission through an ideal relay in the presence of a limit cycle.

It defines an *equivalent nonlinear element* (sometimes called a “modified nonlinearity”) in the sense that one can divorce limit cycle considerations from consideration of nonlinearity signal transmission, provided that A is constant. A normalized plot of $N_B(A, B)$ appears in Fig. 6.2-2. Insofar as the values of B/A encountered may be small, this *equivalent nonlinear element* can be replaced by an *equivalent linear element*, given by its slope at the origin, $2D/\pi A$. This is the physical interpretation of the incremental-input describing function computed earlier.

RECTANGULAR HYSTERESIS

This piecewise-linear characteristic, possessing memory, leads to a square-wave output, as shown in Fig. 6.2-3. It is clear that the fundamental component of the square wave is *not* in phase with the sinusoidal part of x . Thus the limit cycle DIDF is sought in the compact form

$$\begin{aligned}
 N_A(A, B) &= \frac{j}{\pi A} \int_0^{2\pi} y(B + A \sin \psi) e^{-j\psi} d\psi \\
 &= \frac{j}{\pi A} \left[\int_0^{\psi_1} (-D) e^{-j\psi} d\psi + \int_{\psi_1}^{\pi+\psi_2} D e^{-j\psi} d\psi + \int_{\pi+\psi_2}^{2\pi} (-D) e^{-j\psi} d\psi \right] \\
 &= \frac{2D}{\pi A} (e^{-j\psi_1} + e^{-j\psi_2}) \\
 &= \frac{2D}{\pi A} (e^{-j \sin^{-1}[(\delta-B)/A]} + e^{-j \sin^{-1}[(\delta+B)/A]}) \quad (6.2-8)
 \end{aligned}$$

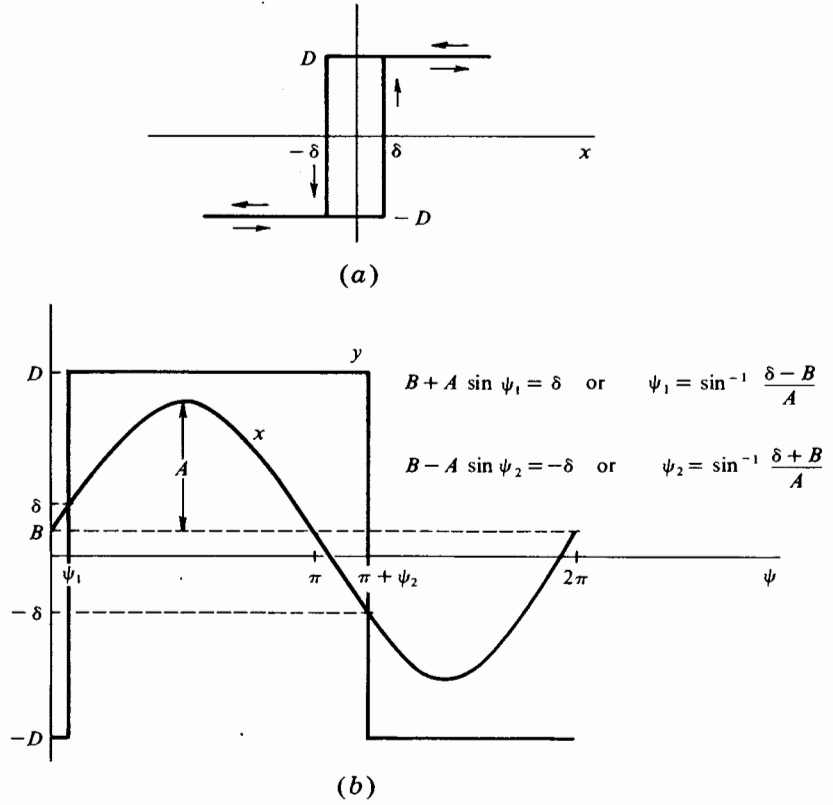


Figure 6.2-3 (b) Input and output waveforms for the rectangular hysteresis nonlinearity (a).

This result can, of course, be expanded to place in evidence either the real and imaginary or magnitude and phase-shift limit cycle DIDF terms. It is to be noted that, as $B \rightarrow 0$, Eq. (6.2-8) reduces to the DF derived for the rectangular hysteresis nonlinearity, Eq. (2.3-26).

The signal DIDF and incremental-input describing function follow directly.

$$\begin{aligned}
 N_B(A, B) &= \frac{1}{2\pi B} \int_0^{2\pi} y(B + A \sin \psi) d\psi \\
 &= \frac{1}{2\pi B} \left[\int_0^{\psi_1} (-D) d\psi + \int_{\psi_1}^{\pi + \psi_2} D d\psi + \int_{\pi + \psi_2}^{2\pi} (-D) d\psi \right] \\
 &= \frac{D}{\pi B} (\psi_2 - \psi_1) \\
 &= \frac{D}{\pi B} \left[\sin^{-1} \left(\frac{\delta + B}{A} \right) - \sin^{-1} \left(\frac{\delta - B}{A} \right) \right] \quad (6.2-9)
 \end{aligned}$$

Expressing the incremental-input describing function calculation as a process of differentiation yields

$$\begin{aligned}
 N_B(A) &= \lim_{B \rightarrow 0} N_B(A, B) \\
 &= \frac{2D}{\pi A} \lim_{B/A \rightarrow 0} \frac{\sin^{-1}(\delta/A + B/A) - \sin^{-1}(\delta/A - B/A)}{2B/A} \\
 &= \frac{2D}{\pi A} \frac{d}{d(\delta/A)} \sin^{-1} \frac{\delta}{A} \\
 &= \frac{2D}{\pi A} \frac{1}{\sqrt{1 - (\delta/A)^2}}
 \end{aligned} \tag{6.2-10}$$

TWO-SEGMENT PIECEWISE-LINEAR ASYMMETRIC NONLINEARITY

This memoryless characteristic is shown in Fig. 6.2-4. By proper choice of m_1 and m_2 , it can be made to represent an absolute-value device ($m_2 = -m_1$), a rectifier voltage-current characteristic ($m_1 \gg m_2 > 0$), and so forth. DIDF calculation proceeds easily. Since the nonlinearity is memoryless, it results that ($|B| \leq A$)

$$\begin{aligned}
 N_A(A, B) &= \frac{1}{\pi A} \int_0^{2\pi} y(B + A \sin \psi) \sin \psi \, d\psi \\
 &= \frac{1}{\pi A} \left[\int_0^{\pi+\psi_1} m_1(B + A \sin \psi) \sin \psi \, d\psi \right. \\
 &\quad \left. + \int_{\pi+\psi_1}^{2\pi-\psi_1} m_2(B + A \sin \psi) \sin \psi \, d\psi \right. \\
 &\quad \left. + \int_{2\pi-\psi_1}^{2\pi} m_1(B + A \sin \psi) \sin \psi \, d\psi \right] \\
 &= \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{\pi} \left(2 \frac{B}{A} \cos \psi_1 + \psi_1 - \frac{\sin 2\psi_1}{2} \right) \\
 &= \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{\pi} \left[\sin^{-1} \left(\frac{B}{A} \right) + \frac{B}{A} \sqrt{1 - \left(\frac{B}{A} \right)^2} \right]
 \end{aligned} \tag{6.2-11}$$

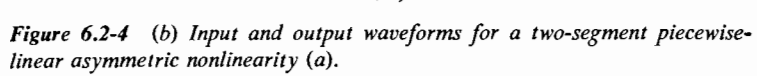
The term in brackets has already been found to occur frequently in DF calculations [denoted $f(B/A)$ in Sec. 2.3]. In terms of the previously introduced notation, Eq. (6.2-11) can be written as

$$N_A(A, B) = \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{2} f\left(\frac{B}{A}\right)$$

It is to be observed that the above results are valid only for a restricted range of B . Outside of this range, inspection yields

$$N_A = \begin{cases} m_1 & \text{for } B > A \\ m_2 & \text{for } B < -A \end{cases} \tag{6.2-12}$$

$$\tag{6.2-13}$$


$$N_P(A.B) =$$

$$\begin{aligned}
 N_B(A, B) &= \frac{1}{2\pi B} \int_0^{2\pi} y(B + A \sin \psi) d\psi \\
 &= \frac{1}{2\pi B} \left[\int_0^{\pi+\psi_1} m_1(B + A \sin \psi) d\psi + \int_{\pi+\psi_1}^{2\pi-\psi_1} m_2(B + A \sin \psi) d\psi \right. \\
 &\quad \left. + \int_{2\pi-\psi_1}^{2\pi} m_1(B + A \sin \psi) d\psi \right] \\
 &= \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{\pi} \left(\psi_1 + \frac{A}{B} \cos \psi_1 \right) \\
 &= \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{\pi} \frac{A}{B} \left[\frac{B}{A} \sin^{-1} \left(\frac{B}{A} \right) + \sqrt{1 - \left(\frac{B}{A} \right)^2} \right] \quad (6.2-14)
 \end{aligned}$$

The term in brackets here differs from that in Eq. (6.2-11), but as in the case of $f(B/A)$, this new term occurs repeatedly in DIDF calculation. It is thus designated $(\pi/2)g(B/A)$, in which case Eq. (6.2-14) can be written as

$$N_B(A, B) = \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{2} \frac{A}{B} g\left(\frac{B}{A}\right)$$

Clearly, in the case of this nonlinearity, the incremental-input describing function defined by Eq. (6.1-8) is meaningless, for an output bias appears even in the absence of an input bias. A more meaningful quantity is the gain to vanishingly small perturbations about that particular input bias B^0 which results in zero output bias. B^0 satisfies the relationship

$$N_B(A, B^0) = 0 \quad \text{for } B^0 \neq 0 \quad (6.2-15)$$

Another meaningful quantity in this instance is the perturbation in output bias caused by a perturbation about zero of the input bias.

POLYNOMIAL-TYPE NONLINEARITY

The class of nonlinearities under consideration is comprised of the odd functions

$$y(x) = c_n x^n \quad (6.2-16)$$

where n is an odd integer. The general formula for the limit cycle DIDF for a memoryless nonlinearity yields

$$\begin{aligned} N_A(A, B) &= \frac{1}{\pi A} \int_0^{2\pi} y(B + A \sin \psi) \sin \psi \, d\psi \\ &= \frac{c_n}{\pi A} \int_0^{2\pi} (B + A \sin \psi)^n \sin \psi \, d\psi \end{aligned} \quad (6.2-17)$$

Applying the binomial theorem in expansion of the integrand and integrating gives

$$\begin{aligned} N_A(A, B) &= \frac{c_n}{\pi A} \int_0^{2\pi} \left[\sum_{k=0}^n \frac{n!}{(n-k)! k!} (A \sin \psi)^{n-k} B^k \right] \sin \psi \, d\psi \\ &= \frac{c_n}{\pi A} \sum_{k=0}^n \frac{n!}{(n-k)! k!} A^{n-k} B^k \int_0^{2\pi} (\sin \psi)^{n-k} \sin \psi \, d\psi \\ &= \frac{c_n}{\pi} \sum_{k=0}^n \frac{n!}{(n-k)! k!} A^{n-k-1} B^k \int_0^{2\pi} (\sin \psi)^{n-k+1} \, d\psi \end{aligned} \quad (6.2-18)$$

Two necessary intermediate results are

$$\text{For } k \text{ odd:} \quad \int_0^{2\pi} (\sin \psi)^{n-k+1} \, d\psi = 0 \quad (6.2-19)$$

For k even:
$$\int_0^{2\pi} (\sin \psi)^{n-k+1} d\psi = 4 \int_0^{\pi/2} (\sin \psi)^{n-k+1} d\psi$$

$$= 2\sqrt{\pi} \frac{\Gamma\left(\frac{n-k+2}{2}\right)}{\Gamma\left(\frac{n-k+3}{2}\right)} \quad (6.2-20)$$

where $\Gamma(\lambda)$ is the gamma function of argument λ . Thus

$$N_A(A, B) = \frac{c_n}{\pi} 2\sqrt{\pi} \sum_{k=0,2,4,\dots}^n \frac{n!}{(n-k)! k!} A^{n-k-1} B^k \frac{\Gamma\left(\frac{n-k+2}{2}\right)}{\Gamma\left(\frac{n-k+3}{2}\right)}$$

$$= \frac{2c_n}{\sqrt{\pi}} \sum_{k(\text{even})=0}^{n-1} \frac{n!}{(n-k)! k!} A^{n-k-1} B^k \frac{\Gamma\left(\frac{n-k+2}{2}\right)}{\Gamma\left(\frac{n-k+3}{2}\right)} \quad (6.2-21)$$

The signal DIDF is computed as follows:

$$N_B(A, B) = \frac{1}{2\pi B} \int_0^{2\pi} y(B + A \sin \psi) d\psi$$

$$= \frac{c_n}{2\pi B} \int_0^{2\pi} (B + A \sin \psi)^n d\psi$$

$$= \frac{c_n}{2\pi B} \int_0^{2\pi} \left[\sum_{k=0}^n \frac{n!}{(n-k)! k!} (A \sin \psi)^{n-k} B^k \right] d\psi$$

$$= \frac{c_n}{2\pi B} \sum_{k=0}^n \frac{n!}{(n-k)! k!} A^{n-k} B^k \int_0^{2\pi} (\sin \psi)^{n-k} d\psi \quad (6.2-22)$$

The integral in the summation contributes only for k odd; hence

$$N_B(A, B) = \frac{c_n}{2\pi} 2\sqrt{\pi} \sum_{k=1,3,5,\dots}^n \frac{n!}{(n-k)! k!} A^{n-k} B^{k-1} \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{\Gamma\left(\frac{n-k+2}{2}\right)}$$

$$= \frac{c_n}{\sqrt{\pi}} \sum_{k(\text{odd})=1}^n \frac{n!}{(n-k)! k!} A^{n-k} B^{k-1} \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{\Gamma\left(\frac{n-k+2}{2}\right)} \quad (6.2-23)$$

Examining the limit of this expression as $B \rightarrow 0$ enables identification of the incremental-input describing function. The summation is first expanded, yielding one term in $B^0 = 1$ and $(n-1)/2$ other terms which disappear in the limit as $B \rightarrow 0$. Thus

$$\begin{aligned} N_B(A) &= \lim_{B \rightarrow 0} [N_B(A, B)] \\ &= \frac{nc_n}{\sqrt{\pi}} A^{n-1} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left[\frac{(n+1)}{2}\right]} \end{aligned} \quad (6.2-24)$$

A very common odd nonlinearity is of the form

$$y = x^3 \quad (6.2-25)$$

Using the above results enables finding the DIDFs for the limit cycle and signal as

$$N_A(A, B) = \frac{3}{4}A^2 + 3B^2 \quad (6.2-26)$$

$$N_B(A, B) = \frac{3}{2}A^2 + B^2 \quad (6.2-27)$$

NONLINEAR CLEGG INTEGRATOR

Discussion of this dual-mode nonlinear integrator can be found in Chap. 2 (see Fig. 2.4-3). Figure 6.2-5 shows the input and discontinuous output over one complete cycle, $-\psi_1 \leq \psi \leq 2\pi - \psi_1$. The output of the Clegg integrator is determined in two pieces: First, by integrating the input waveform from $-\psi_1$, the point at which the input turns positive, to the literal variable ψ (for $-\psi_1 \leq \psi < \pi + \psi_1$):

$$\begin{aligned} y(B + A \sin \psi, A\omega \cos \psi) &= \int_{-\psi_1/\omega}^{\psi/\omega} (B + A \sin \psi) d\left(\frac{\psi}{\omega}\right) \\ &= \frac{1}{\omega} \int_{-\psi_1}^{\psi} (B + A \sin \psi) d\psi \\ &= \frac{1}{\omega} [B(\psi + \psi_1) + A(\cos \psi_1 - \cos \psi)] \end{aligned} \quad (6.2-28)$$

and second, by integrating *with zero initial conditions* from $\pi + \psi_1$, the point at which the input turns negative, to ψ (for $\pi + \psi_1 \leq \psi < 2\pi - \psi_1$):

$$\begin{aligned} y(B + A \sin \psi, A\omega \cos \psi) &= \frac{1}{\omega} \int_{\pi+\psi_1}^{\psi} (B + A \sin \psi) d\psi \\ &= \frac{1}{\omega} [B(\psi - \psi_1 - \pi) - A(\cos \psi_1 + \cos \psi)] \end{aligned} \quad (6.2-29)$$

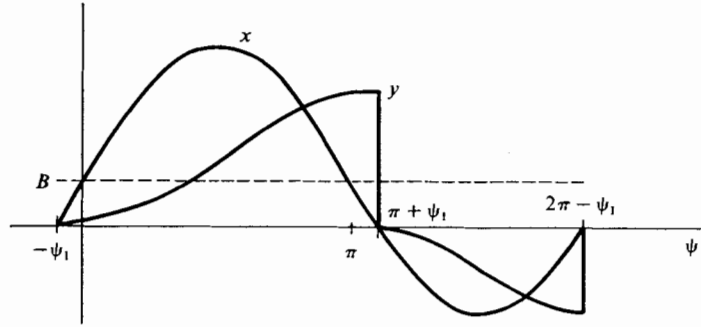


Figure 6.2-5 Input and output waveforms for the nonlinear Clegg integrator.

The frequency-dependent limit cycle DIDF is computed as follows:

$$\begin{aligned}
 N_A(A, B, \omega) &= \frac{j}{\pi A} \int_{-\psi_1}^{2\pi - \psi_1} y(B + A \sin \psi, A\omega \cos \psi) e^{-j\psi} d\psi \\
 &= \frac{j}{\pi A \omega} \int_{-\psi_1}^{\pi + \psi_1} [B(\psi + \psi_1) + A(\cos \psi_1 - \cos \psi)] e^{-j\psi} d\psi \\
 &\quad + \frac{j}{\pi A \omega} \int_{\pi + \psi_1}^{2\pi - \psi_1} [B(\psi - \psi_1 - \pi) - A(\cos \psi_1 + \cos \psi)] e^{-j\psi} d\psi \\
 &= \frac{4}{\pi \omega} \left[1 + \left(\frac{B}{A} \right) \sin^{-1} \left(\frac{B}{A} \right) \sqrt{1 - \left(\frac{B}{A} \right)^2} - \left(\frac{B}{A} \right)^2 \right. \\
 &\quad \left. - j \frac{\pi}{4} \left(1 + 2 \left(\frac{B}{A} \right)^2 \right) \right] \quad (6.2-30)
 \end{aligned}$$

where the interval over which the DIDF is evaluated is chosen, for convenience, as $-\psi_1 \leq \psi < 2\pi - \psi_1$, instead of $0 \leq \psi < 2\pi$, and the relationship $\psi_1 = \sin^{-1}(B/A)$ is employed. Observe that, as it should in the limit as $B \rightarrow 0$, the limit cycle DIDF reduces to the DF computed in Chap. 2:

$$N_A(A, B = 0, \omega) = \frac{4}{\pi \omega} \left(1 - j \frac{\pi}{4} \right)$$

Following the same procedure gives the signal DIDF and incremental-input describing function as

$$\begin{aligned}
 N_B(A, B, \omega) &= \frac{1}{2\pi B} \int_{-\psi_1}^{2\pi - \psi_1} y(B + A \sin \psi, A\omega \cos \psi) d\psi \\
 &= \frac{2}{\pi \omega} \left[\left(\frac{\pi}{2} \right)^2 + \left[\sin^{-1} \left(\frac{B}{A} \right) \right]^2 + \left(\frac{A}{B} \right) \sin^{-1} \left(\frac{B}{A} \right) \sqrt{1 - \left(\frac{B}{A} \right)^2} \right] \\
 &\quad (6.2-31)
 \end{aligned}$$

and

$$\begin{aligned}
 N_B(A, \omega) &= \lim_{B \rightarrow 0} N_B(A, B, \omega) \\
 &= \frac{1}{\omega} \left(\frac{4 + \pi^2}{2\pi} \right) \\
 &= \frac{2.21}{\omega}
 \end{aligned} \tag{6.2-32}$$

Thus, in the region of small B/A , both the limit cycle and signal DIDFs are not only independent of B but of A as well. Only frequency exists as a DIDF parameter. This is certainly reminiscent of the behavior of a linear integrator. The remaining differences between these *linearized* nonlinear integrator transfer functions and the single transfer function of a linear integrator are what makes the Clegg integrator a particularly useful compensatory device.

Additional DIDF calculations are tabulated in Appendix C. The frequent appearance there of the functions $f(B/A)$ and $g(B/A)$ attest to their value as a shorthand notation.

6.3 FORCED RESPONSE OF LIMIT CYCLING NONLINEAR SYSTEMS

In this section an input-output model which accounts for transient as well as frequency response behavior is developed for limit cycling systems. Needless to say, the approximate DIDF analysis employed results in certain restrictions on the use of this model. Special attention is therefore devoted to the question of its range of validity. The results sought are approximations, such as may be of convenient use in analysis and design work. The arguments presented are both heuristic and abbreviated.

Basically, what we should like to argue is the equivalence of systems *a* and *b* of Fig. 6.3-1. That is, the original nonlinearity is to be modeled by its incremental-input describing function, and the remaining effect of the limit cycle (in addition to its modulating effect on the original nonlinear element) is to appear in the linearized system as an additive output term. The reason for employing the incremental-input describing function rather than the signal DIDF is, of course, that the equivalent system thereby becomes totally linear. A prefilter has been associated with the system to allow for a reshaping of the input amplitude spectrum such that the DIDF nonlinearity characterization is valid over the range of inputs anticipated.

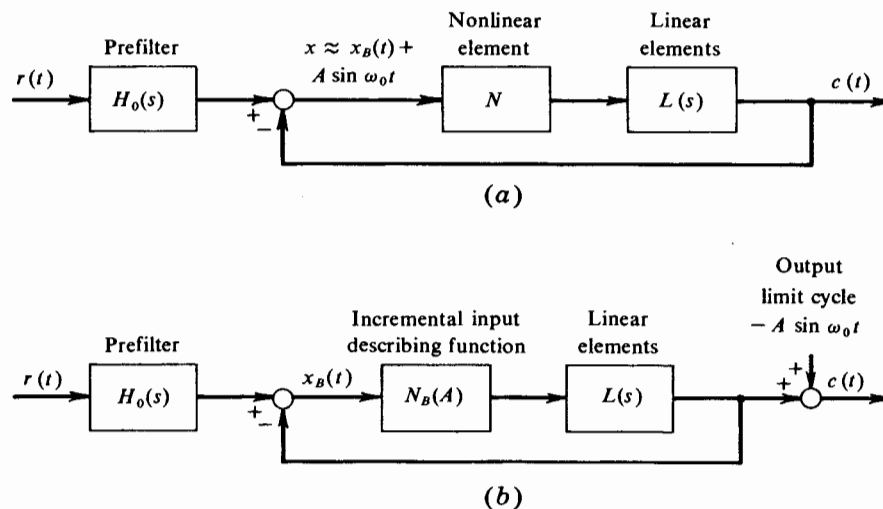


Figure 6.3-1 (a) A limit cycling nonlinear system. (b) Its equivalent-linear-system model.

POLE-ZERO SYSTEM CHARACTERIZATION

In the following discussion it is assumed that the nonlinearity input signal consists of a sinusoidal component due to the limit cycle plus another component due to the input signal (Fig. 6.3-1a). That is,

$$\begin{aligned} x(t) &\approx x_B(t) + x_A(t) \\ &\approx x_B(t) + A \sin \omega_0 t \end{aligned}$$

It is our intent here to discuss the meaning of poles and zeros as applied to the linear system model in Fig. 6.3-1b. A static nonlinearity is assumed. Consider the imaginary and real axes separately.

Along the imaginary axis we are concerned with sinusoidal response characteristics. Hence we consider the nonlinearity input signal to consist of two sinusoids, one due to the limit cycle and the other to system response to the input signal. This results in precisely the TSIDF situation studied earlier. According to the TSIDF analysis of Sec. 5.1, the gain to each sinusoid is frequency-independent provided that the sinusoidal frequencies are nonharmonically related. Since for present purposes any deterministic linking of these frequencies is not envisioned, the assumption of an irrational frequency ratio is not at all restrictive. If, further, the amplitude-ratio condition of Eq. (6.2-7) is imposed, the result for many nonlinearities is that the limit cycle amplitude is independent of the forcing signal $[N_A(A, B) \rightarrow N_A(A)]$, and the gain to the smaller sinusoid is independent of its own

amplitude $[N_B(A, B) \rightarrow N_B(A)]$. Under these circumstances the system input-output description is indeed linear.

Let us now turn our attention to the axis of real exponentials. Again we assume satisfaction of the amplitude-ratio condition just cited, where the ratio now refers to peak exponential amplitude divided by peak limit cycle amplitude. In considering the nonlinearity gain to an exponential in the presence of a sinusoid, it is immediately apparent that the time duration of the exponential relative to a limit cycle period is a significant factor. Exponentials of long duration such as 10 or more limit cycle periods are certainly well represented in the DIDF input signal model consisting of sinusoid plus bias. Equation (6.1-1) tends to be satisfied in this instance. On the other hand, exponentials of sufficiently short duration can take place during various phases of the sinusoid, and the responses would be quite different. For example, in the case of an ideal relay, an exponential occurring during the limit cycle amplitude peaking would evoke essentially no additional nonlinearity output. The "gain" to such a transient signal is near zero. If, on the other hand, the same short-duration exponential occurs near a limit cycle zero crossing, the nonlinearity output indeed reflects its presence, and thus leads to a substantially larger "gain." The important fact, however, is the possible *time dependence* of nonlinearity gain to the transient signal. *As the exponential duration increases, the time dependence of this gain decreases.* Figure 6.3-2 illustrates one particular situation. Calling $\tau_{\min}$ the minimum acceptable value of exponential time constant (i.e., the time constant corresponding to maximum allowable dependence of the signal gain upon time), one could argue that the exponential should continue for a minimum of two limit cycle periods, viz.,

$$4\tau_{\min} \approx 2 \frac{2\pi}{\omega_0}$$

or

$$\frac{1}{\tau_{\min}} \approx \frac{\omega_0}{3} \quad (6.3-1)$$

In this event the maximum delay (T_d in Fig. 6.3-2) is approximately 25 percent of the total exponential time duration. This value is somewhat arbitrary, but its implications will be fully apparent in the development to follow. With this order-of-magnitude calculation we proceed directly to an interpretation of the significance of poles and zeros in the complex s plane input-output system description.

Consider the s plane as divided into the three regions shown in Fig. 6.3-3. From the previous heuristic development we argue by extension that region I, to the left of the line defined by $\sigma = -\omega_0/3$, is the space in which closed-loop system poles display residues which depend upon the time they are excited. Region III, the right half-plane, cannot contain any closed-loop poles

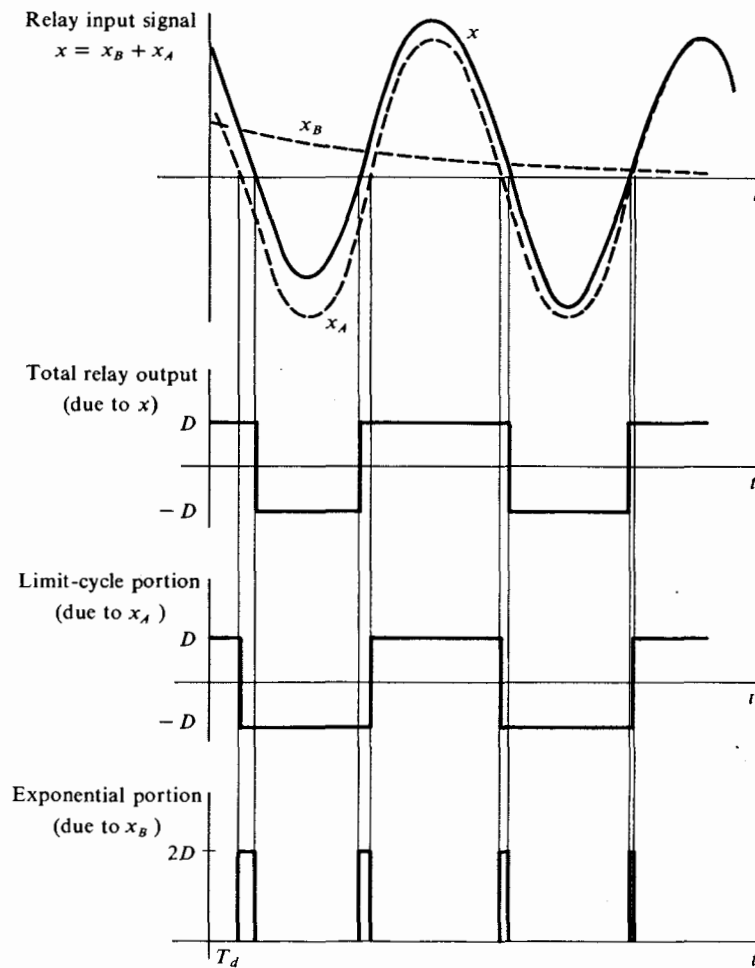


Figure 6.3-2 Response of an ideal relay to an exponential signal in the presence of a limit cycle.

since they would eventually violate the amplitude-ratio condition. The remaining region, designated region II, is the space wherein closed-loop poles are taken to correspond to approximately *linear time-invariant* response modes.

The addition of a zero $(1 + \tau s)$ to a function $F(s)$ yields, on a linear basis, a time function given by $f(t)$ plus $\tau df(t)/dt$, where $f(t) = \mathcal{L}^{-1}[F(s)]$. Since the time derivative of a sinusoid is another sinusoid of the same frequency, and since the time derivative of an exponential is another exponential with the same time constant, it is clear that the presence of zeros in no

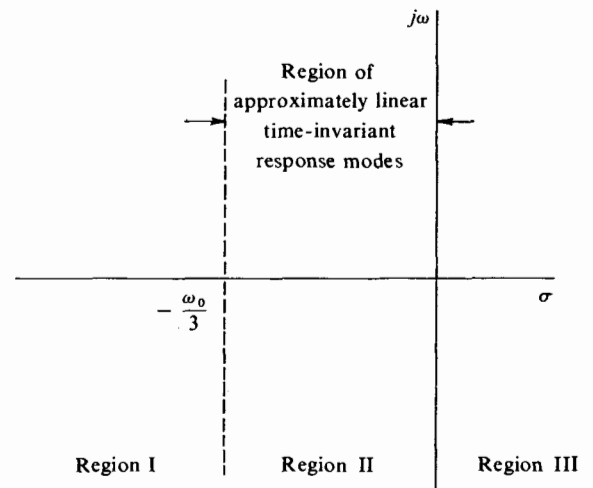
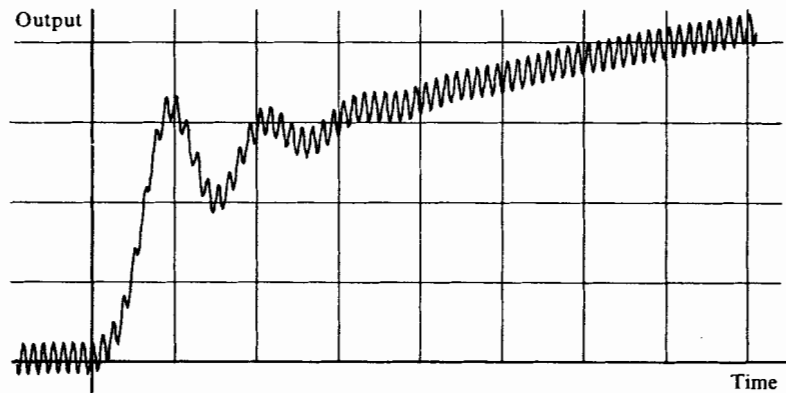


Figure 6.3-3 Regional division of the s plane.

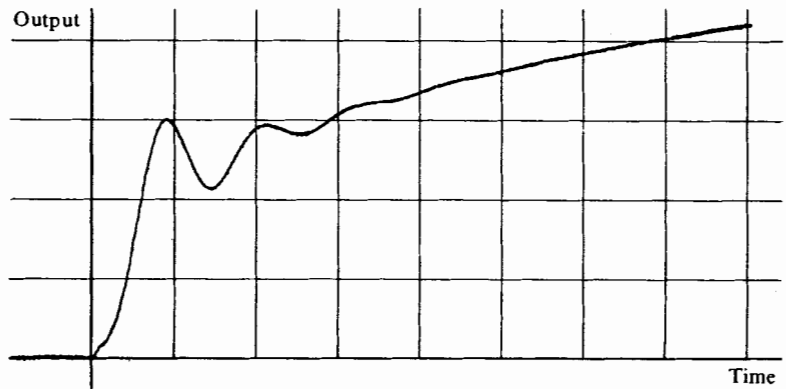
way alters the regional division of the s plane as given by Fig. 6.3-3. In fact, zeros may merely be thought of as altering the residues which accrue to the poles within the s plane. Note that all zeros must be used, regardless of the regional division in which they lie. These must be dealt with, therefore, both to secure a desired system time response and to ensure continuous satisfaction of the amplitude-ratio condition.

A limit cycling control system for a high-performance aircraft was simulated on an analog computer. The nonlinearity in this control system was an ideal relay. In order to test the equivalent gain concept for the relay in the presence of a limit cycle, transient responses for the limit cycling system and for the equivalent linear system in which the relay was replaced with a linear gain of magnitude $k = 2D/\pi A$ were studied. All predominant response modes were determined analytically to be within region II in the s plane, which thus requires an identical limit cycling system and equivalent-linear-system performance. Experimental results are shown in Fig. 6.3-4, in which the ordinate and abscissa scales for both responses are identical. Such results demonstrate the validity of the extension of pole-zero concepts to limit cycling systems, and tend to substantiate the DIDF model.

With a knowledge of the physical significance of poles and zeros for limit cycling systems, we are in a position to exploit the root-locus method, which has proved so useful in conventional linear theory. Recalling the closed-loop portion of the system of Fig. 6.3-1b and considering N to be a variable gain element enables construction of the locus of roots as in linear servo theory. The locus shape so derived is valid only insofar as N_B is non-phase-shifting. For $\sigma < -\omega_0/3 \text{ sec}^{-1}$ there is some phase shift added to signals



(a) Limit cycling control system



(b) Equivalent linear control system

Figure 6.3-4 Transient responses of a particular limit cycling control system and its analytically computed linear equivalent.

passing through the nonlinearity, and the actual root-locus position is uncertain to this extent. Locus gain calibration, in the usual sense, is also no longer meaningful; that is, by varying some process or compensation gain factor, the system pole locations cannot be changed. Instead, the open-loop signal gain automatically adjusts to a constant value, related to that value which establishes the limit cycle. This point will be further explored later on.

Example 6.3-1 Find the DIDF linearized equivalent system corresponding to the relay control system of Fig. 6.3-5. Assume that the prefilter has been chosen such that for all expected $r(t)$ the amplitude-ratio condition at $x(t)$ is satisfied.

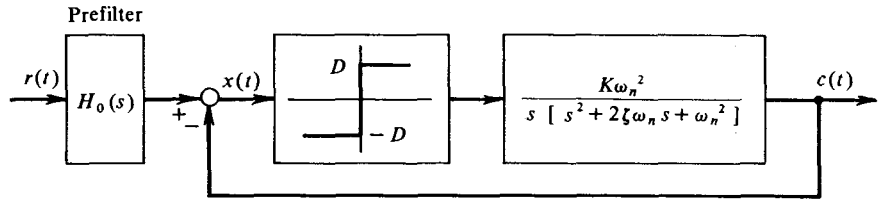


Figure 6.3-5 Third-order relay control system.

The uncalibrated root locus of the limit cycling loop is derived by assuming the relay to act as a non-phase-shifting gain, with the result shown in Fig. 6.3-6. The limit cycle frequency is determined as the point at which the locus crosses the $j\omega$ axis. Alternatively, by the DF methods of Chap. 3, the limit cycle amplitude and frequency are determined from the characteristic equation

$$s^3 + 2\zeta\omega_n s^2 + \omega_n^2 s + N_A K \omega_n^2 = 0$$

to be (Table 3.1-1)

$$\omega_n = \omega_0$$

and

$$A = \frac{2DK}{\pi\zeta\omega_n}$$

The limit cycle DIDF has been taken as

$$N_A \approx \frac{4D}{\pi A}$$

which is the value of $N_A(A, B, \omega)$ determined by Eq. (6.2-2), valid to 5 percent for all $B/A < \frac{1}{2}$. Using the above-determined value of A and the signal DIDF (incremental-input describing function),

$$N_B \approx \frac{2D}{\pi A}$$

computed from Eq. (6.2-3) by again using the fact that $B/A < \frac{1}{2}$, yields the characteristic equation of the linearized limit cycling loop as

$$s^3 + 2\zeta\omega_n s^2 + \omega_n^2 s + \zeta\omega_n^3 = 0$$

This equation has one real and two complex-conjugate roots, denoted by the squares on the three root-locus branches in Fig. 6.3-6. For small ζ , the three roots are approximately located within region II at the positions

$$s_{1,2,3} \approx -\zeta\omega_n, -\frac{\zeta\omega_n}{2} \pm j\omega_n \sqrt{1 - \frac{\zeta^2}{4}}$$

with the result that

$$\frac{C}{R}(s) \approx H_0(s) \frac{\zeta\omega_n^3}{(s + \zeta\omega_n)(s^2 + \zeta\omega_n s + \omega_n^2)}$$

where the prefilter transfer function has been included. Note that the gain constants K and D do not appear in the input-output transfer function. For this reason the input-output dynamics do not change with changing K and D ; the system is *adaptive* with respect to these parameters. This point is more fully explored in the following sections.

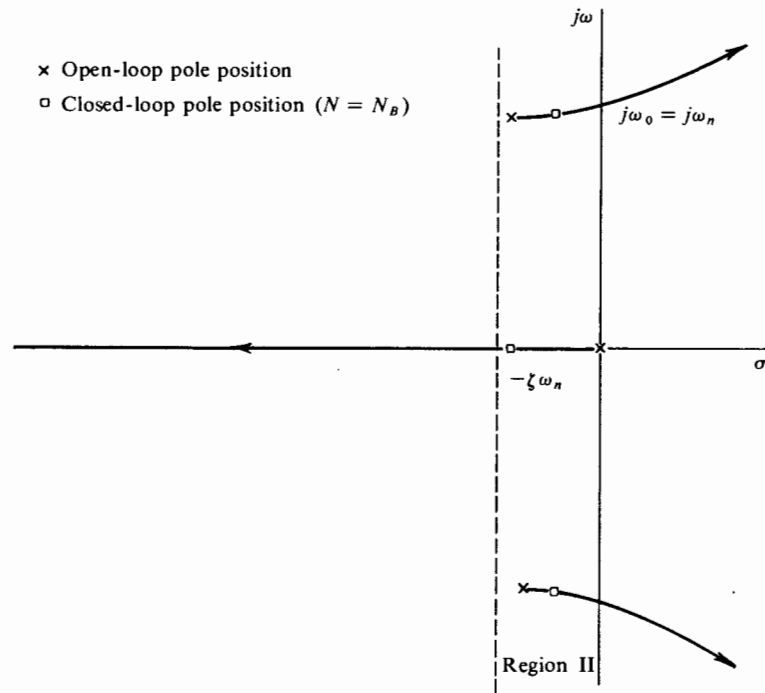


Figure 6.3-6 Root-locus diagram for the closed-loop part of the system in Fig. 6.3-5.

STABILITY

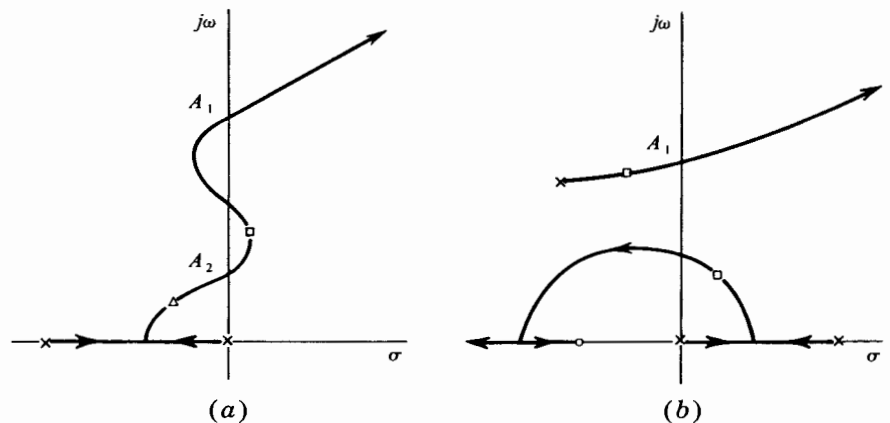
In the design of a closed-loop system for any long-term regulatory action, the first and most important specification is that the system possess no unstable modes. Linear systems with greater than 180° of open-loop phase lag, or stated differently, with root-locus branches in the right half-plane, are unstable for certain choices of open-loop gain.

On the other hand, limit cycling systems may not display similar linear unstable modes, whatever the open-loop gain setting. Consider, for example, a closed-loop control system containing an ideal relay (cf. Example 6.3-1). As the process static sensitivity K increases, the limit cycle amplitude at the nonlinearity input increases proportionally. Thus the gain N_B , inversely proportional to A , decreases, and the product KN_B is automatically held constant. This example evidences that, whereas in a linear system eventually an instability would occur, in the corresponding limit cycling system, the closed-loop dynamics do not even vary. The only price paid is the increase in limit cycle amplitude; and for cases in which this may be troublesome, Sec. 4.4 outlines the basis for its automatic regulation.

For the majority of common nonlinearities (i.e., those displaying saturation), N_B is less than N_A for $B < A$. In such cases unstable closed-loop

modes can occur only if the system is of a form which can become unstable for a decrease of open-loop gain from that value which sustains the limit cycle. The root loci for two systems of this class are illustrated in Fig. 6.3-7. In the first case, Fig. 6.3-7a, a conditionally stable system is illustrated. The quasi-static stability theory of Chap. 3 determines that of the three root-locus $j\omega$ -axis crossings, two correspond to stable limit cycles. These are labeled A_1 and A_2 , where $A_2 > A_1$. Let us assume the system to be in limit cycle state A_1 . Correspondingly, the dynamics of small-signal propagation through the loop are determined by the position of all closed-loop roots associated with the nonlinearity gain $N_B(A_1)$. Since the arrows along the locus correspond to the direction of increasing open-loop gain, it follows from the relationship $N_B(A_1) < N_A(A_1)$ that the root of interest can lie on either side of the $j\omega$ axis, depending upon the gain calibration of the locus. For the situation depicted in Fig. 6.3-7a, this root falls in the right half-plane, indicating unstable closed-loop dynamics. This leads immediately to the conclusion that limit cycle state A_1 is actually *unstable*; only the larger-amplitude limit cycle can occur! Observe that a similar argument proves the system stable to small signals in the presence of the larger-amplitude limit cycle.

Arguing along similar lines for the unstable open-loop system of Fig. 6.3-7b yields that there may be no stable state whatever, depending upon the position of the lower branch root at the nonlinearity gain $N_B(A_1)$. In the



- × Open-loop pole
- Open-loop zero
- Closed-loop root corresponding to $N_B(A_1)$
- △ Closed-loop root corresponding to $N_B(A_2)$

Figure 6.3-7 Root-locus plots for a conditionally stable system (a) and an unstable open-loop system (b), both containing a static memoryless nonlinearity.

illustration it is depicted as an unstable mode. However, if a stable situation does arise, it *must* do so at the smaller limit cycle amplitude A_1 .

Let us note in passing that limit cycling control *can* be applied successfully to open-loop unstable processes. It is essential in this case to design the system not only for a stable limit cycle but also for stable closed-loop modes at the gain prescribed by N_B . In a laboratory at the Massachusetts Institute of Technology, a limit cycling system was constructed to control an inverted pendulum, with quite successful operation.

STEADY-STATE FORCED ERRORS

In response to *harmonic* forcing the steady-state forced errors are determined directly from the DIDF linearized equivalent system. To the extent that the complete representation of $N_B(A, B)$ is used in such analyses, rather than just its slope at the origin, all results obtained (in the case of unrelated frequencies) will be *identical* with TSIDF results. Section 6.8 provides the justification for this statement. It is of interest, in addition, to note the ease with which steady-state errors resulting from *aperiodic* inputs may be determined. We demonstrate by example.

Example 6.3-2 Find the steady-state following error produced by the relay system of Fig. 6.3-5 when the input is (a) a step of magnitude R , and (b) a ramp of magnitude Rt . Assume a first-order prefilter with time constant τ .

(a) The steady-state prefilter output is a constant, R ; hence the following error of the overall system is equal to the constant-input following error of the limit cycling loop. For the limit cycling loop to be in steady state, a zero-average-value relay output is required as a result of the open-loop integration in $L(s)$. This condition can be satisfied only if $B = 0$, in which case *there is no steady-state following error*. This, of course, should come as no surprise, again because of the open-loop integration in $L(s)$.

(b) In the case of a ramp input, the steady-state prefilter output is another ramp, $R(t - \tau)$, displaying a following error $R\tau$. In order for the limit cycling loop to be in steady state, an average relay output equal to R/K is required, since this produces an output ramp which tracks the input. Employing the exact expressions for N_B and N_A , we get

$$\frac{R}{K} = BN_B(A, B) = \frac{2D}{\pi} \sin^{-1} \frac{B}{A}$$

or

$$\frac{B}{A} = \sin \frac{\pi R}{2DK}$$

and since $\omega_0 = \omega_n$,

$$\frac{4D}{\pi A} \sqrt{1 - \left(\frac{B}{A}\right)^2} \frac{K}{2\zeta\omega_n} = 1$$

This last condition is simply the limit cycle magnitude condition. Solving the above two equations, we get for the limit cycle amplitude

$$A = \frac{2DK}{\pi\zeta\omega_n} \cos \frac{\pi R}{2DK}$$

and for the input following error

$$B = \frac{DK}{\pi\zeta\omega_n} \sin \frac{\pi R}{DK}$$

Hence the ramp following error of the overall system is

$$\text{Ramp following error} = R\tau + \frac{DK}{\pi\zeta\omega_n} \sin \frac{\pi R}{DK}$$

Clearly, the above is only valid for $DK \geq R$. It is impossible to command any faster response from the given system. If, in addition, the limit cycle amplitude is to be relatively independent of the ramp input (say, a maximum deviation of 5 percent from the unforced condition), a drive capability $(DK)_{\min}$ is required for which

$$\cos \frac{\pi R}{2(DK)_{\min}} = 0.95$$

or

$$(DK)_{\min} \approx 5R \quad (6.3-2)$$

Notice that in this example the DDF nonlinearity input model of a bias plus a sinusoid is approximate only to the extent that higher harmonics of the limit cycle sinusoid are omitted. The bias term is a true representation of the form of signal appearing at the nonlinearity input.

A CLOSED-LOOP PROPERTY: INHERENT ADAPTIVITY

The open-loop signal transfer function OLTF($j\omega$), for the closed-loop portion of the basic limit cycling control-system configuration of Fig. 6.3-8, is

$$\begin{aligned} \text{OLTF}(j\omega) &= N_B H_1(j\omega) H_2(j\omega) H_3(j\omega) L(j\omega) \\ &= N_B H(j\omega) L(j\omega) \end{aligned} \quad (6.3-3)$$

where N_B is the signal gain of the nonlinearity. Presuming satisfaction of the amplitude-ratio condition enables drawing the locus for this OLTF($j\omega$) on the amplitude-phase plane (gain-phase plane). Here N_B can be considered as a variable gain, shifting the system OLTF($j\omega$) vertically.

A first look at this method of representation for systems containing an ideal relay is facilitated by recalling the results

$$N_A \approx \frac{4D}{\pi A} \quad \text{and} \quad N_B \approx \frac{2D}{\pi A} \quad (6.3-4)$$

Correspondingly,

$$\frac{N_B}{N_A} \approx \frac{1}{2} \quad (6.3-5)$$

The interesting fact¹ that the frequency locus OLTF($j\omega$) always passes

¹ First discovered by Lozier (Ref. 11).

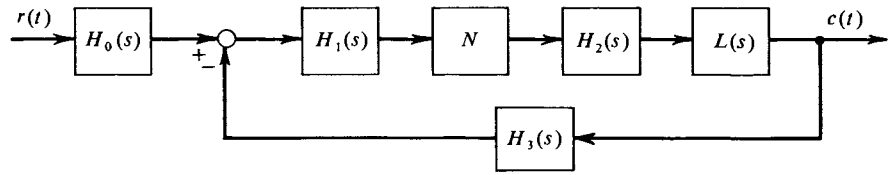


Figure 6.3-8 Basic limit cycling control-system structure.

through the point $\frac{1}{2} \angle -180^\circ$ [or $(-6 \text{ db}, -180^\circ)$] on the amplitude-phase plane, to within the accuracy of Eq. (6.3-5), is easily proved as follows:

$$\begin{aligned} \text{OLTF}(j\omega) &= N_B H(j\omega) L(j\omega) \\ &= \frac{1}{2} N_A H(j\omega) L(j\omega) \end{aligned} \quad (6.3-6)$$

But in order to sustain a limit cycle,

$$N_A H(j\omega_0) L(j\omega_0) = -1 \quad (6.3-7)$$

therefore

$$\text{OLTF}(j\omega_0) = -\frac{1}{2} \quad (6.3-8)$$

The frequency locus in Fig. 6.3-9 has been drawn for a typical linear process in a limit cycling loop.

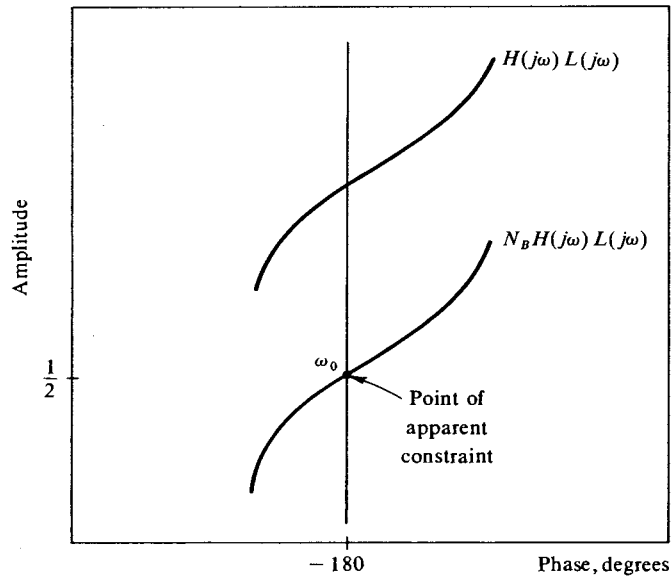


Figure 6.3-9 Typical linear elements, and associated OLTF ($j\omega$) of a relay limit cycling system containing these linear elements.

Seeing how one point of the $OLTF(j\omega)$ locus is constrained on the amplitude-phase plane, even in the face of changes in process characteristics, affords a first feeling for *inherent adaptivity* in a limit cycling control system. It is to be noted that Eq. (6.3-7) yields the magnitude and frequency of the limit cycle, and that for static, single-valued nonlinearities this frequency is also the frequency at which the phase of $H(j\omega)L(j\omega)$ is -180° . In the more general case for which the nonlinearity limit cycle DIDF is phase-shifting, the point of apparent constraint becomes

$$OLTF(j\omega_0) = -\frac{N_B}{|N_A|} e^{-j\angle N_A} \quad (6.3-9)$$

where the representation

$$N_A = |N_A| e^{j\angle N_A} \quad (6.3-10)$$

is employed.

6.4 A SCHEME FOR PARAMETER-ADAPTIVE CONTROL

Throughout this section, ξ is used to denote an "environmental parameter" such as that which causes a changing description for the controlled process $L(s, \xi)$. Thus ξ may be thought of as relating to such varied phenomena as temperature, pressure, altitude, aging, load fluctuations, and so forth. Because all physical controlled processes do in some way respond to a fluctuating environment, the need for some form of closed-loop environment insensitivity may exist. The conceptual approach to obtaining environment insensitivity beyond that provided by linear feedback configurations is termed *parameter-adaptive control*. Its objective may be stated, analytically, as the intent to continuously satisfy Eq. (6.4-1):

$$\frac{C}{R}(s, \xi) \approx \frac{C}{R}(s) \quad (6.4-1)$$

where C/R denotes the overall transfer function of the system under consideration. An approximate equality is indicated since attempts to achieve exact equality require perfect and complete *identification information*, that is, identification of the precise changes in $L(s, \xi)$. In the general case this requirement is either impractical or impossible.

In a limit cycling system the limit cycle itself is the carrier of identification information. This can be seen from Eq. (6.3-7), rewritten for convenience:

$$N_A(A)H(j\omega_0)L(j\omega_0, \xi) = -1 \quad (6.4-2)$$

The solution of this equation is

$$\text{Limit cycle at input to } N = A(\xi) \sin [\omega_0(\xi)t + \varphi] \quad (6.4-3)$$

whence we see that the limit cycle amplitude and frequency relate implicitly to the environmental parameter ξ . Although it requires a closed pool in which to circulate, the limit cycle is related *directly* to open-loop behavior. In contrast to other self-contained approaches to identification, the limit cycle adaptive control information is not attenuated by the negative feedback characteristic of the control loop. On the other hand, one limit cycle carries but two pieces of information.

Following identification is *adaptive control action*, during which the loop compensation is adjusted to maintain the limit cycle amplitude and frequency at reference values. This has the effect of constraining the amplitude and phase of the open-loop transfer function at the reference frequency, which in turn provides a good measure of environment insensitivity for the closed-loop system in many cases. Figure 6.4-1 illustrates the general limit cycling adaptive control system. The identification-data takeoff point need not be the output, and the adaptive control action feedback need not be arranged as shown. This figure is meant to be schematic only.

In this form of adaptive control there is no need to externally generate an identification test signal; the system itself provides the limit cycle. Further, there is no need for mechanizing some nonphysical performance index; the limit cycle amplitude and frequency themselves become the performance indices, and are readily available. Moreover, all instrumentation is particularly simple.

It has already been mentioned that the (ideal-relay) limit cycling control system input-output transfer is totally insensitive to controlled-process static sensitivity (zero frequency gain) perturbations. This turns out to be the case, to a first order, for many system nonlinearities besides the ideal relay.

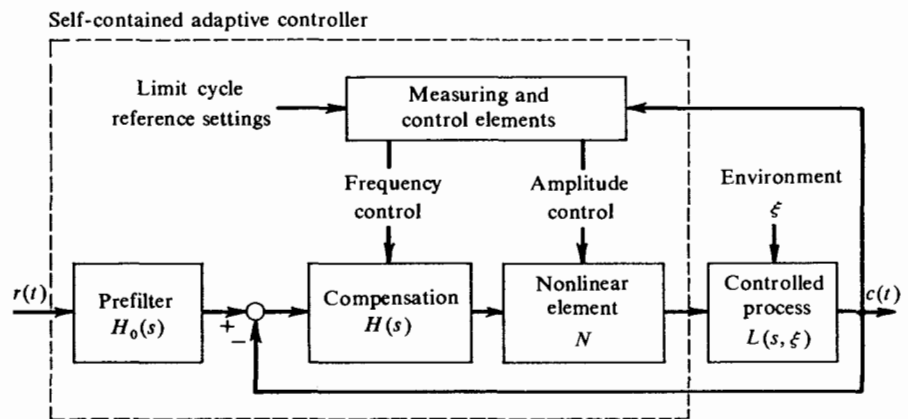


Figure 6.4-1 Limit cycling parameter-adaptive control system.

Note that this display of nearly perfect parameter adaptivity in response to controlled-process static sensitivity perturbations is a property of a limit cycling loop, the identification and adaptive control action operations occurring simultaneously and inextricably, within the nonlinearity. Controlled-process time-constant perturbations generally appear as both limit cycle amplitude and frequency perturbations, according to Eq. (6.4-2). It is the function of the measuring and control elements of Fig. 6.4-1 to sense and resolve these perturbations before directing any adaptive control action. A *sensitivity function* argument which quantitatively describes the amount of adaptivity associated with various limit cycling adaptive control schemes appears in the literature (Ref. 4).

The first commercial exploitation of a *limit cycling adaptive control system* (as such) was advanced by the Minneapolis-Honeywell Regulator Co. It is discussed in an interesting paper by Shuck (Ref. 22). The final system, an autopilot for a high-performance aircraft, was obtained by way of a long simulation program. Theoretical calculations were subsequently published (Ref. 3, pp. 294-300), which substantiated the adaptive capability of this experimental autopilot. Other applications of this technique have been determined, for instance, in connection with adaptive roll control of an air-ground missile. This application is discussed in Sec. 6.5. We digress here to consider briefly the matter of limit cycling loop compensation, both in general and with the *adaptive* control aspect in mind.

COMPENSATION

It is presumed in this chapter that the system of concern has its limit cycle mode excited at all times. This allows a simple analytic representation for the nonlinearity in the closed loop, and leads to a tractable mathematical system description. Thus a *series prefilter* is included in the limit cycling control-system structure. Its fundamental purpose is to reshape command input amplitude spectra at the input to the nonlinearity such that the amplitude-ratio condition as given by Eq. (6.2-7) is satisfied. It is clear that the prefilter design is a function of the properties of the expected class of input signals, the limit cycle, and the overall system specifications. The basic system structure is therefore as shown in Fig. 6.3-7. $H_0(s)$, $H_1(s)$, $H_2(s)$, and $H_3(s)$ denote linear filters, as is the controlled process (plant) $L(s)$. The nonlinearity N may be considered as a nonlinear filter which changes its transmittance to low-frequency input terms as the limit cycle changes. Note that it need *not* be a power element.

Compensation of a limit cycling loop requires that the nonlinearity drive level, prefilter, and loop linear filters be so chosen that system specifications are satisfied over the range of expected inputs. Thus, in the case of an ideal-relay servo, D must be chosen large enough so that the maximum values

of output derivatives which can be commanded are sufficient for following the input. In addition, if the limit cycle is to continue essentially unchanged in response to the input (a requirement for loop parameter adaptivity), the ratio of command-input-forced following error to limit cycle amplitude (B/A) must remain small. This requires a value of D which is roughly five times the input n th derivative, for a type- n plant of unity static sensitivity [Eq. (6.3-2)]. On the other hand, increasing D results in a proportional increase in A , for which a maximum specification is likely to exist at some station around the control loop.

Linear compensation is then selected to obtain the desired limit cycle frequency and amplitude. In practice, a design specification which requires a limit cycle frequency of about ten times (or greater) the desired overall system bandwidth is readily fulfilled. The position of the linear compensation is of the utmost importance. For example, if to increase the limit cycle frequency requires a *lead* network, it is most advantageously placed at H_1 , because placement at H_2 or H_3 would either maximize the limit cycle amplitude at $c(t)$ or unnecessarily "slow down" the system response. Of course, there are systems in which H_1 is fixed or otherwise inaccessible. In such cases an alternative must be sought.

SUMMARY

The limit cycling adaptive control technique can be applied to all linear processes for which the existence of a continuous limit cycle is acceptable. Depending upon the nature of the variant portions of the controlled-process transfer function, limit cycling adaptive control may yield parameter-adaptive systems for which the required instrumentation is simple relative to all other current schemes. The accompanying small expense of instrumentation, coupled with comparable (and often superior) system adaptability in a number of practical situations, clearly indicates the limit cycling adaptive technique as a highly competitive approach to parameter-adaptive control. The major limitation is the compromise which usually must be struck between system drive capability (ability to follow inputs) and limit cycle amplitude. Compensation of appropriate design can often improve this compromise.

6.5 APPLICATION TO AN ADAPTIVE MISSILE ROLL CONTROL SYSTEM

One of the difficulties in designing a control system for a missile is the fact that missile aerodynamic characteristics vary greatly over the flight profile. Thus a control system which is satisfactory at one flight condition may be

completely unsatisfactory at another. One obvious solution to this problem is programmed control, in which system compensation is varied as a function of some externally measured parameters of the flight condition (such as Mach number, dynamic pressure, altitude, etc.). Generally, however, it is more desirable to have the control system "adapt" to the changing missile characteristics without the necessity of external measurements. This approach overcomes the problems of operation under uncertain environmental conditions and required a priori correlation between external measurements and actual missile characteristics, both of which can invalidate programmed control. The self-inspecting approach to design of control systems with variable controlled elements (plants) is termed *parameter-adaptive design*. In this section the limit cycling system approach to parameter-adaptive control is applied to the missile roll control problem.

In general, using this design technique, the control loop containing the variable plant dynamics is caused to limit cycle by the inclusion of a nonlinear element. In so doing, the loop thereafter automatically displays a certain insensitivity with respect to plant parameter perturbations. In the specific case at hand the nonlinearity which causes the system limit cycle is taken as part of the system fixed elements, the resulting insensitivity mentioned above providing *perfect* parameter adaptation.

HYPOTHETICAL MISSILE AND SYSTEM SPECIFICATIONS

The hypothetical vehicle to be considered is a wingless air-to-ground missile with a cruciform tail. Control moments are obtained by deflecting control surfaces on the trailing edges of the tail fins. To avert cross coupling between the vertical control surfaces and pitching motion, and between the horizontal surfaces and yawing motion, it is necessary to *roll-stabilize* the missile. Roll control moments are obtained by deflecting the control surfaces differentially.

The general equation of motion about the roll axis is

$$I_x \ddot{\varphi} + L_{\dot{\varphi}} \dot{\varphi} + L_{\varphi} \varphi = L_{\delta} \delta + L_D \quad (6.5-1)$$

where I_x = missile moment of inertia

$L_{\dot{\varphi}}$ = damping coefficient

L_{φ} = elastic restraint coefficient

L_{δ} = control torque coefficient

L_D = disturbance torques

The elastic restraint torque $L_{\varphi} \varphi$ may occur when the missile has an angle of attack and an angle of sideslip. However, it is usually so small that it may be neglected, resulting in the roll equation given in most standard texts:

$$\ddot{\varphi} + \frac{L_{\dot{\varphi}}}{I_x} \dot{\varphi} = \frac{L_{\delta}}{I_x} \delta + \frac{L_D}{I_x} \quad (6.5-2)$$

For the missile under consideration, the damping in roll is assumed to be quite small, such that the transform of Eq. (6.5-2) may be approximated by

$$\begin{aligned}\varphi(s) &\approx \frac{1}{s^2} \frac{L_\delta}{I_x} \delta(s) + \frac{1}{s^2} \frac{1}{I_x} L_D(s) \\ &= \frac{k_1(\xi)}{s^2} \delta(s) + \frac{k_2(\xi)}{s^2} L_D(s)\end{aligned}\quad (6.5-3)$$

where s is the Laplace transform variable, and ξ is an environmental parameter. We are primarily concerned with the transfer relationship between roll angle and control-surface deflection. Since the control gain L_δ is roughly proportional to dynamic pressure, which may vary greatly over the flight profile, and in addition the moment of inertia decreases as fuel is burned, it is assumed that these variations cause $k_1(\xi)$ to vary a hundredfold. Thus

$$0.5 < k_1(\xi) < 50 \quad (6.5-4)$$

It is further assumed that although ξ can vary, the dynamics of variations encountered are slow relative to the limit cycle amplitude transient time.

It must be mentioned at this point that the simplified model for roll dynamics is chosen merely to expedite a presentation of the limit cycling adaptive-control philosophy. The general method to be discussed can be extended easily to the more complex situation of Eq. (6.5-1). This is primarily due to the fact that limit cycling control systems behave like high-gain linear systems, without the conventional problem of linear instability.

The control surfaces are driven by a nonlinear actuator, consisting of a solenoid valve with external orifice control and a hydraulic ram. The corresponding equation relating control-surface position δ to the valve input signal x can be written as

$$\dot{\delta}(\pm x) = \pm D \quad (6.5-5)$$

or in "transfer function" form,

$$\delta(s) = \frac{D}{s} \operatorname{sgn} x \quad (6.5-6)$$

The magnitude D is controlled by varying the size of the effective solenoid valve orifice. The nonlinearity considered is, in essence, an ideal relay with drive levels $\pm D$ followed by an integrator.

For the sake of subsequent reference to a specific system, the following arbitrarily chosen control-system specifications are selected:

- Roll step response percent overshoot < 15 percent
- Roll response time (to within ± 5 percent of final value) < 5 sec
- Limit cycle amplitude at output must be constant $\leq \alpha$ deg

where α , a number related to mission requirements, need not be specified further for this example.

LIMIT CYCLING MINOR LOOP

Consider the limit cycling loop of Fig. 6.5-1, where for the moment disturbance torques are ignored. The forward path contains the hydraulic actuator and the missile airframe dynamics, and also some unspecified linear compensatory filter $H_1(s)$. The feedback element is a rate gyro with gain factor $1/k$.

With this choice of loop elements, the limit cycle frequency can be selected as desired. The equation governing the limit cycle for any loop with nonlinearity N and linear elements $HL(s)$ is

$$N_A(A)HL(j\omega_0) = -1$$

In the present case, therefore,

$$\frac{4D}{\pi A} H_1(j\omega_0) \frac{k_1(\xi)}{k} \frac{1}{(j\omega_0)^2} = -1 \quad (6.5.7)$$

which can be resolved into the real equations

$$\frac{4D}{\pi A} |H_1(j\omega_0)| \frac{k_1(\xi)}{k} \frac{1}{\omega_0^2} = 1 \quad (6.5-8)$$

and

$$\angle H_1(j\omega_0) = 0 \quad (6.5-9)$$

In anticipation of a system response time of about 4 sec and in conjunction with hypothetical missile structural capabilities, $\omega_0 = 20$ radians/sec is

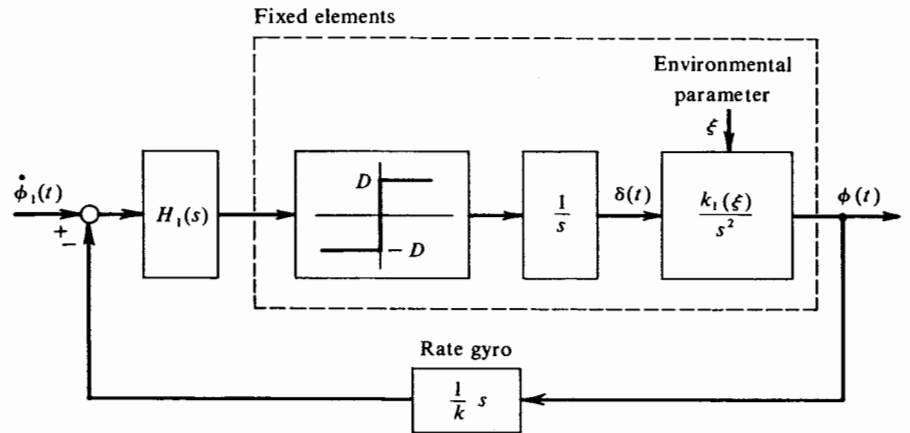


Figure 6.5-1 Limit cycling inner loop.

chosen. Clearly, to achieve a stable limit cycle, $H_1(s)$ must be a lead-lag network. It may be chosen as

$$H_1(s) = \frac{s+1}{0.1s+1} \frac{0.0025s+1}{0.025s+1} \quad (6.5-10)$$

which satisfies Eq. (6.5-9) at $\omega_0 = 20$ radians/sec. A root-locus plot for the limit cycling loop is shown in Fig. 6.5-2. The closed-loop poles marked off correspond to a value of equivalent open-loop gain equal to half that which sustains the limit cycle, in accordance with the previously established relationship for an ideal relay [Eq. (6.3-5)]. From the root-locus plot the minor-loop transfer function is determined as

$$\begin{aligned} \frac{\varphi}{\varphi_1}(s) &\approx \frac{k}{s} \frac{(1.042)(214)(s+1)}{(s+1.042)(s^2+3.5s+214)} \\ &\approx \frac{214k}{s(s^2+3.5s+214)} \end{aligned} \quad (6.5-11)$$

The approximate cancellation of the open-loop zero by the closed-loop pole is a manifestation of the high-gain character of limit cycling control loops mentioned earlier. The pole at the origin in this transfer function is due to the feedback element. Thus the application of the theory of limit cycling systems to this limit cycling loop leads to a simple conclusion regarding its input-output dynamical properties. *Note that these properties are independent of $k_1(\xi)$; the basis for a parameter-adaptive control system has been achieved.*

ROLL POSITION OUTER LOOP

Design of the linear outer loop is a matter of selecting a series prefilter which causes satisfaction of the amplitude-ratio condition at the input to the non-linearity over the expected class of inputs and simultaneously provides a suitable input-output transfer for the entire system. Thus $H_0(s)$ in Fig. 6.5-3 must be determined. It is a prefilter to the limit cycling loop, but internal to the roll position loop. A second-order prefilter suffices for the problem at hand. Thus

$$H_0(s) = \frac{\omega_f^2}{s^2 + 2\zeta_f\omega_f s + \omega_f^2} \quad (6.5-12)$$

To satisfy the frequency-ratio requirement we may conservatively choose $\omega_f = 0.1\omega_0 = 2$ radians/sec. That this choice is conservative follows from the fact that the prefilter output consists of forced responses and normal modes, wherein the normal modes (for this hypothetical missile problem) are assumed to have significant energy in a higher frequency band than the forced responses. Both ζ_f and k remain to be chosen. These may be

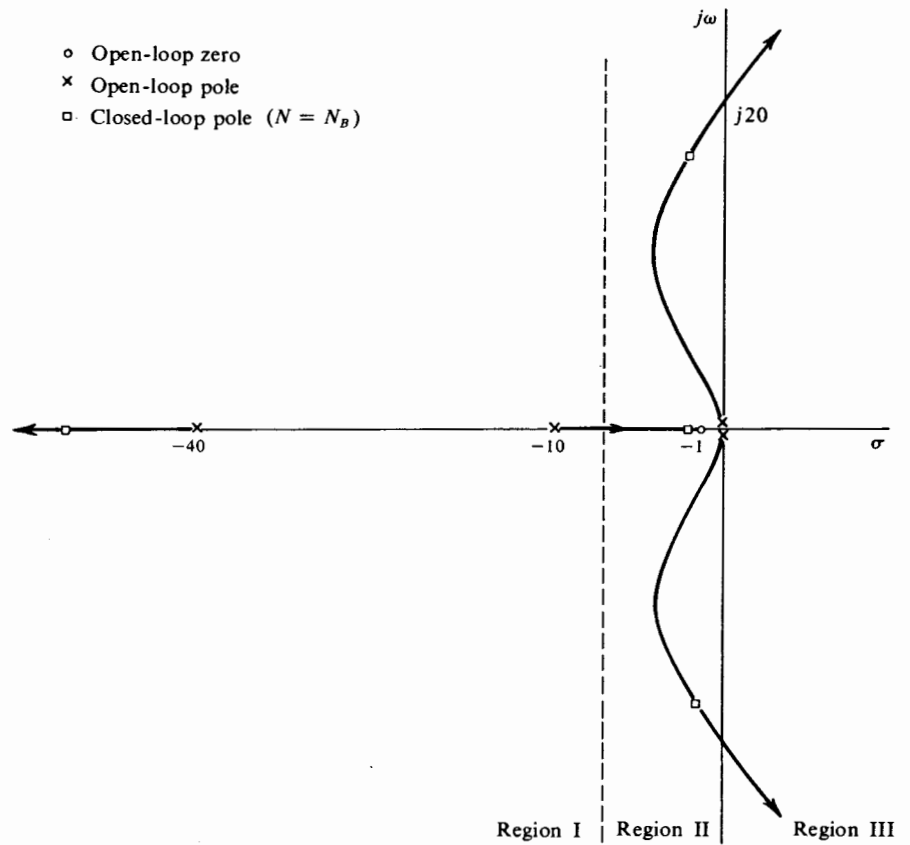


Figure 6.5-2 Minor-loop root-locus plot.

obtained by consideration of the amplitude-ratio condition. In particular, we can choose φ_{cmd} to be a step function (a reasonable choice in this roll problem), replace N by its linear equivalent, and solve for $x(t)$ in terms of the amplitude of φ_{cmd} , ζ_f , and k . It has been determined that the range of choices for ζ_f and k given by

$$\begin{aligned} \zeta_f &> 0.6 \\ 4\zeta_f &> k > 0.7 \end{aligned} \quad (6.5-13)$$

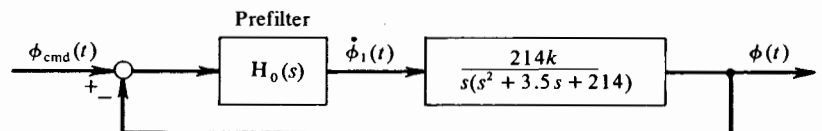


Figure 6.5-3 Roll position outer loop.

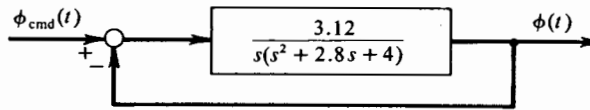


Figure 6.5-4 Equivalent overall system.

results simultaneously in system stability and satisfaction of the amplitude-ratio condition. One set of values within this range which also provides the desired system transient response is

$$\zeta_f = 0.7 \quad k = 0.78 \quad (6.5-14)$$

The resulting equivalent overall system is shown in Fig. 6.5-4, where the complex pole pair of the inner loop has been dropped from further consideration, being well beyond the bandwidth of the system.

AMPLITUDE-REGULATING LOOP

If design is terminated at this point we shall have a system which satisfies all specifications except that of output limit cycle amplitude. In fact, as the missile sensitivity $k_1(\xi)$ varies over a 100:1 range, the limit cycle amplitude at station φ follows proportionally, whereas that at station δ (Fig. 6.5-1) remains constant. Since we have made available a variable nonlinearity (variable orifice), we are at liberty to reverse the above situation (see Fig. 6.5-5).

The active device $M[A]$ measures the output limit cycle amplitude and adjusts the valve orifice so that this limit cycle amplitude remains constant at α deg (see specifications). Details of the general requirements for construction of $M[A]$ and the resultant speed of adaptation of the overall system can be deduced from the content of Sec. 4.4. Note that $M[A]$ serves to establish limit cycle amplitude regulation at the system output. This, of course, implies that the limit cycle amplitude at the control surface varies over a 100:1 range. Therefore D is chosen so that, at the smallest value for $k_1(\xi)$, the limit cycle amplitude at δ is acceptable. As $k_1(\xi)$ increases, this limit cycle amplitude decreases, thus always remaining acceptable. Setting D also relates to the amplitude-ratio condition, with a slight circular nature to this problem arising. A solution is reached when the amplitude-ratio condition at x is not violated for the largest expected system transient input, when the limit cycle at station δ is acceptable, and when the overall system response fits specifications. Thereafter the system response is invariant.

Note that the regulation of the limit cycle amplitude in no way alters the

system input-output behavior as long as the amplitude-ratio condition is always satisfied.

We may digress a moment to discuss the disturbance response properties of the control system at this point. Equation (6.5-3) indicates the manner in which disturbance torques enter the system. Figure 6.5-5 details this situation. To determine the effect of disturbance torques on roll angle it is necessary to specify, first, the energy distribution in $L_D(s)$. This is due to the nonlinear filtering qualities of an apparently "dithered" relay seen by any signal. For low-frequency disturbance inputs which also do not violate the amplitude-ratio condition at x , the system behaves as a linear feedback system, and its transfer characteristic may be so determined. The steady-state error, for example, in response to any constant disturbance is zero. For high-frequency disturbance inputs the limit cycling system appears open-loop, so that

$$\frac{\varphi}{L_D}(s) \approx \frac{k_2(\xi)}{s^2} \quad (6.5-15)$$

For exceedingly large transient disturbances the limit cycle may temporarily terminate with the "relay" hard over (i.e., control surface slewing at maximum rate in one direction to overcome the disturbance).

EXPERIMENTAL RESULTS

An analog computer was used in a study of the system of Fig. 6.5-5. The function of $M[A]$ was performed manually, according to the same rule which would govern the actual circuit. Figure 6.5-6 shows the system roll transient response for a number of conditions over the 100:1 range in $k_1(\xi)$. Note that in all cases the time to peak is approximately 2.8 sec and the overshoot is 12 percent. For the system of Fig. 6.5-4 a purely linear calculation predicts corresponding values of 2.9 sec and 12.5 percent. Thus the argument that the overall system behaves linearly is well supported. Further, observe that the limit cycle amplitude at the output is so small as to be barely discernible on the analog-computer recordings. Figure 6.5-7 shows δ at intervals over the same range. Note the ordinate scales. It is seen that the correlation between theory and experiment is excellent.

Finally, recall that the time-domain specifications were chosen arbitrarily, merely to allow numerical reference throughout the control-system design. The ultimate capability of the hypothetical airframe is by no means exhausted. In the final analysis this depends upon structural considerations and control-surface slewing capability. In fact, recognizing that the control-surface

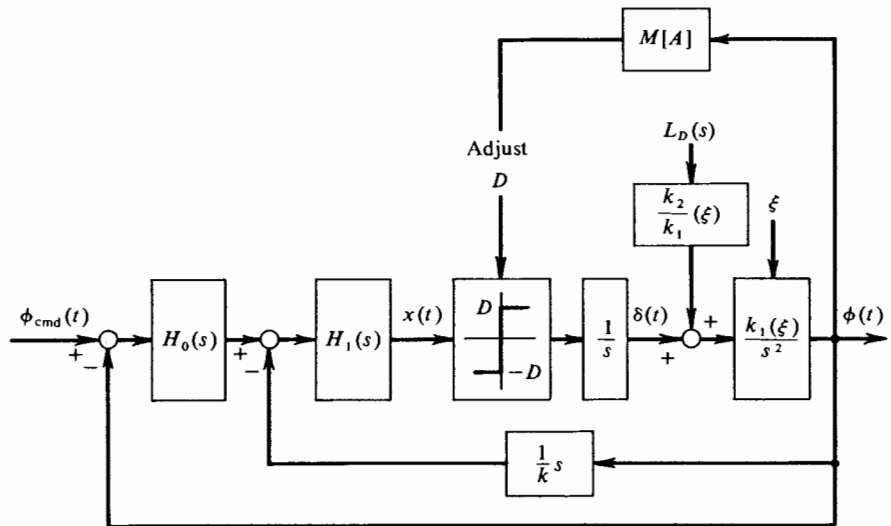


Figure 6.5-5 Proposed control configuration.

limit cycle amplitude is largest when $k_1(\xi)$ is smallest, and that this condition may exist for only a small part of the flight, quite significant decreases in system response time may be obtained. In any case, the simplicity in design of this system is clear, indicating the potential practicality of its application to the missile roll-stabilization problem.

6.6 LIMIT CYCLES IN SYSTEMS WITH AN ASYMMETRIC NONLINEARITY

A very useful property of the DIDF formulation is the means by which systems with an asymmetric¹ nonlinearity can be studied. As we shall presently see, this application is quite natural, and conceivably could have accounted for the DIDF development in the first place.

Consider the feedback system of Fig. 6.6-1 in which N is an asymmetric nonlinear element and $L(s)$ is the loop linear part. If this loop is in a limit cycle state, it is clear that $x(t)$ will contain a bias term [for our present purposes we choose $r(t) = 0$]. This follows since, in the assumed absence of any bias in $x(t)$, $y(t)$ will indeed develop a bias and contradict this assumption. Such behavior, of course, results solely from our assertion that the nonlinearity is asymmetric. In a symmetric nonlinearity system no biases appear, a fact which is implicit in the formulation of the limit cycle theories of previous chapters.

¹ By asymmetric is meant one which is not odd; that is, $y(x) \neq -y(-x)$.

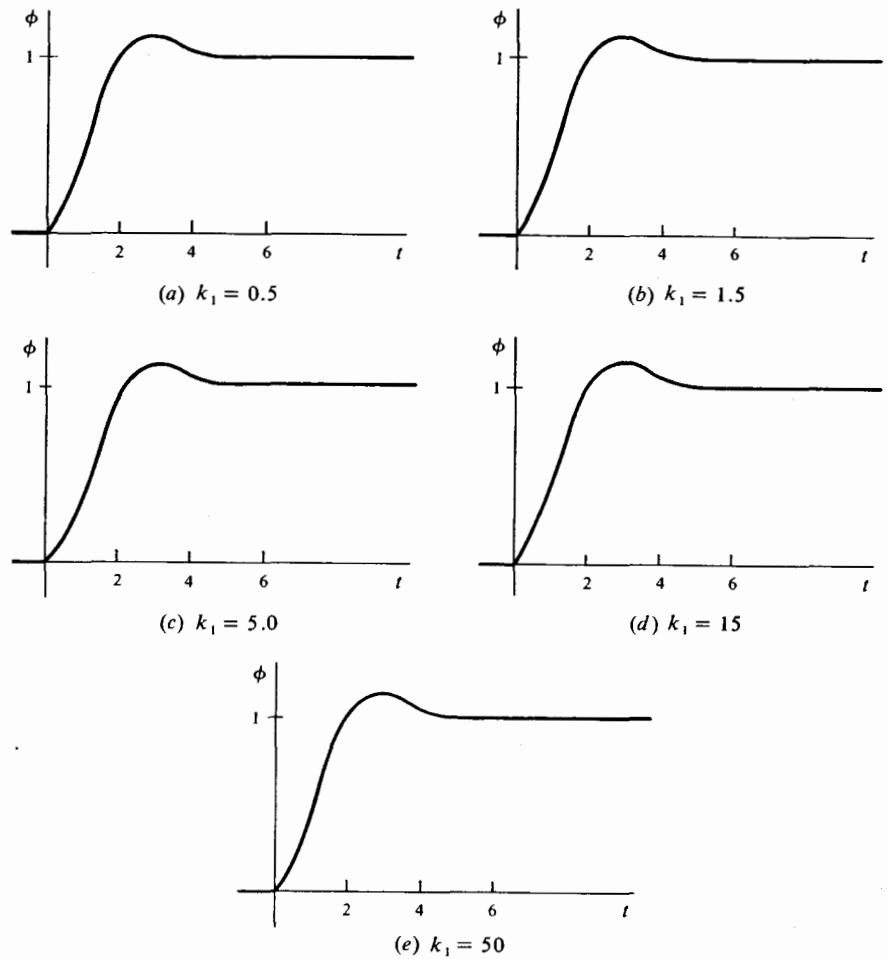


Figure 6.5-6 Missile roll responses taken over the entire flight envelope.

For a sinusoid at $x(t)$ to propagate unattenuated around the loop requires

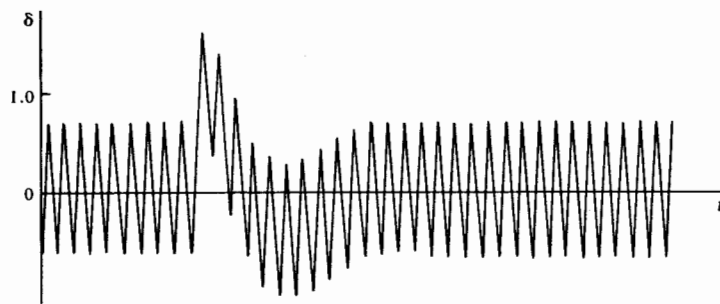
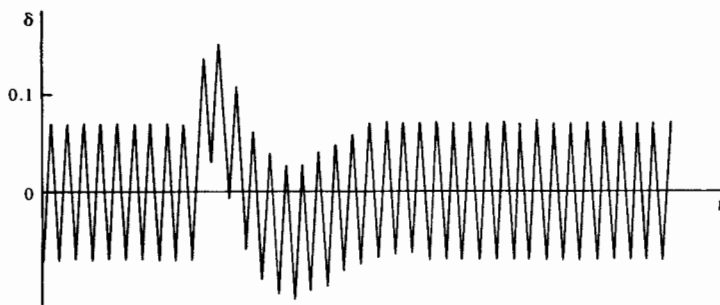
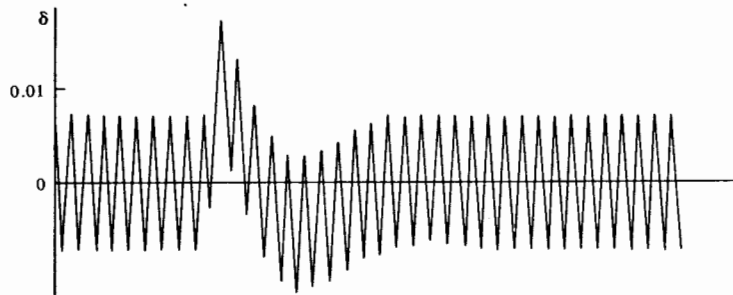
$$N_A(A, B, \omega_0)L(j\omega_0) = -1 \quad (6.6-1)$$

as in past limit cycle formulations. In addition, however, it is required that a bias (B) at $x(t)$ also propagate identically around the loop. Thus

$$BN_B(A, B, \omega_0)L(j0) = -B \quad (6.6-2)$$

This identity is automatically satisfied if $B = 0$, an uncommon situation, and hence is of little interest in that case. For $B \neq 0$, however, we must have

$$N_B(A, B, \omega_0)L(j0) = -1 \quad (6.6-3)$$

(a) $k_1 = 0.5$ (b) $k_1 = 5.0$ (c) $k_1 = 50$ **Figure 6.5-7** Control-surface responses taken over the entire flight envelope.

the second of *two* conditions which must be fulfilled in order that the system of Fig. 6.6-1 sustain a limit cycle. $L(j0)$ denotes, of course, the static sensitivity (dc gain) of the linear elements.

It is true, in general, that the periodic output of an asymmetric nonlinearity in a limit cycling system contains *even* harmonics. In particular, some *second* harmonic may be present. To whatever extent this is true, the filter

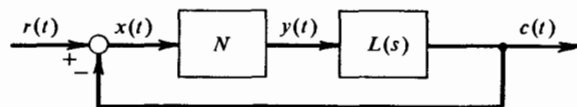


Figure 6.6-1 System with an asymmetric nonlinearity.

hypothesis need obviously be altered in the direction of requiring *more* filtering of the nonlinearity output for continued validity of the DIDF linearization. Otherwise, all remains as before.

When $L(s)$ has no open-loop integrations (poles at the origin), $L(j0)$ is finite and the application of Eq. (6.6-3) is straightforward. If $L(s)$ has one or more open-loop integrations, $|L(j0)| \rightarrow \infty$, and we require $N_B(A, B, \omega_0) = 0$ [alternatively, $BN_B(A, B, \omega_0) = 0$ since $B \neq 0$] as the only possible solution of Eq. (6.6-3). If, on the other hand, $L(s)$ has one or more open-loop differentiations (zeros at the origin), the only possible means for satisfying the limit cycle conditions are indicated by Eq. (6.6-2), which calls for $B = 0$. In each of these cases, stability of the indicated bias value must be determined. Table 6.6-1 summarizes the possible circumstances.

TABLE 6.6-1 REQUIREMENTS ON B FOR A LIMIT CYCLE TO EXIST†

$L(s)$	Requirement
$s^n L_1(s) \quad n \geq 1$	$B = 0$
$L_1(s)$	$N_B(A, B, \omega_0) L_1(j0) = -1$
$\frac{L_1(s)}{s^m} \quad m \geq 1$	$BN_B(A, B, \omega_0) = 0$

† $L_1(s)$ has neither poles nor zeros at the origin.

A simple example serves to clarify the use of this table.

Example 6.6-1 Determine the limit cycle state of the system in Fig. 6.6-2.

The asymmetric nonlinearity is effectively a biased-output ideal relay, for which $N_A(A, B)$ and $N_B(A, B)$ appear in Appendix C. For $|B/A| \leq 1$

$$N_A(A, B) = \frac{2D}{\pi A} \sqrt{1 - \left(\frac{B}{A}\right)^2}$$

$$N_B(A, B) = \frac{D}{2\pi B} \left(\pi + 2 \sin^{-1} \frac{B}{A} \right)$$

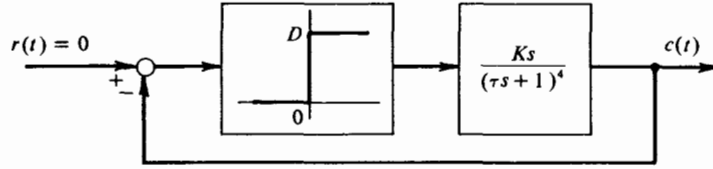


Figure 6.6-2 Example system with a biased-output ideal-relay nonlinearity.

Table 6.6-1 evidences that $B = 0$ is a limit cycle requirement. Hence, from Eq. (6.6-1) it follows that

$$\frac{2D}{\pi A} \frac{Kj\omega_0}{(j\tau\omega_0 + 1)^4} = -1$$

Solution yields

$$\omega_0 = \frac{2.41}{\tau}$$

and

$$A = 0.033 \frac{DK}{\tau}$$

It remains to be shown that the condition of zero nonlinearity input bias is a stable one. To do this, assume a perturbation about zero of the input bias, and the associated perturbation about $D/2$ of the output bias. This gives the incremental-input describing function applicable to the present example.

$$\begin{aligned} N_B(A, B = 0) &\triangleq \lim_{\Delta B \rightarrow 0} \frac{\Delta B N_B(A, \Delta B) - D/2}{\Delta B} \\ &= \frac{D}{\pi A} \lim_{\Delta B \rightarrow 0} \frac{\sin^{-1}(\Delta B/A)}{\Delta B/A} \\ &= \frac{D}{\pi A} \\ &= \frac{9.65\tau}{K} \end{aligned}$$

We are thus interested in stability of the system illustrated in Fig. 6.6-3. The open-loop elements of this linear feedback system are

$$\text{OLTF}(s) = 9.65 \frac{\tau s}{(\tau s + 1)^4}$$

By use of linear analysis it is readily verified that this system is indeed stable. Hence, and in summary, the limit cycle state determined above is stable. An analog simulation of this system bears out these findings, and indicates solution accuracies of about 5 percent.

If the linear elements of this example are replaced by either

$$\frac{K}{(\tau s + 1)^3} \quad \text{or} \quad \frac{K}{s(\tau s + 1)^2}$$

the DIDF analysis leads to incompatible requirements, from which it is to be concluded that a limit cycle is not possible. Again, analog simulation substantiates these conclusions.

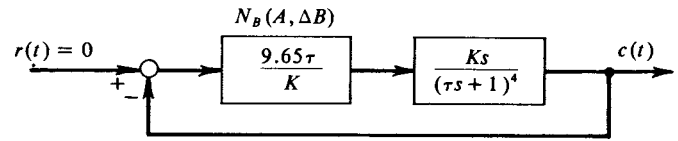


Figure 6.6-3 Bias stability study of example system.

6.7 ARTIFICIAL DITHER AND SIGNAL STABILIZATION

The vehicle by which DIDF formulation, calculation, and usage have been conveyed thus far in this chapter has been the limit cycling system. Other interpretations of the bias and sinusoidal part of the nonlinearity input signal exist, and they account for quite different behavioral aspects of nonlinear systems. Consider, for example, a non-limit-cycling system with an input

$$r(t) = r_B(t) + R \sin \omega_r t$$

where $|\dot{r}_B(t)| \ll R\omega_r/2\pi$. Provided that the application of this signal does not stimulate a systematic oscillation and that the amplitude of the low-frequency part of the nonlinearity input is less than that of the sinusoidal part, DIDF linearization of the system nonlinearity provides a means for determining the forced response behavior of the system. Although it is an easy matter to check the amplitude-ratio requirement, it is quite difficult to determine whether the input *causes* the system to oscillate at frequencies other than ω_r . Describing function verification would call for a nonlinearity linearization for inputs comprised of a bias and *two* sinusoidal inputs. In theory, this can certainly be achieved. Practically, this may go beyond the point of diminishing returns. For this reason attention is confined to situations where simple DIDF linearization is useful by itself.

ARTIFICIAL DITHER

The employment of high-frequency signal injection for the purpose of altering the *apparent* characteristics of a nonlinearity in a closed-loop system is an exceedingly useful compensatory device. The high-frequency signal is referred to as *artificial dither*, or just *dither*.¹ The technique of dithering for linearizing purposes has been known for some time.

Artificial dither implies intentionally applied dither. Unintentional introduction of dither into control systems, however, is not at all uncommon. Examples are 60- or 400-cps electrical pickup in an instrument servo, and

¹ In this chapter all dither waveforms considered are periodic. The use of random dither is discussed in Chap. 7.

mechanical vibration in a missile control-surface servo. Under the circumstances to be delineated presently, such unintentional dither can be readily accounted for as well.

In what follows it is assumed that the loop linear elements attenuate the high-frequency dither to the point where only an insignificant dither frequency signal makes the return trip to its originating place. Three types of dither are considered separately.

Triangular-wave dither Consider the application of symmetric triangular-wave dither to an *arbitrary* memoryless odd saturating nonlinearity (Fig. 6.7-1a). Let $e(t)$ represent the total nonlinearity input, comprised of signal $x(t)$ and dither $d(t)$. This symmetric nonlinearity is defined by

$$\begin{aligned} e &\leq -\delta & y &= -D \\ -\delta < e < \delta & y(e) &= -y(-e), \text{ otherwise arbitrary} \\ e &\geq \delta & y &= D \end{aligned} \quad (6.7-1)$$

Input and output waveforms are shown in Fig. 6.7-2, where for input values in the range $-\delta < e < \delta$ the (arbitrary) output is shown dotted. Since the

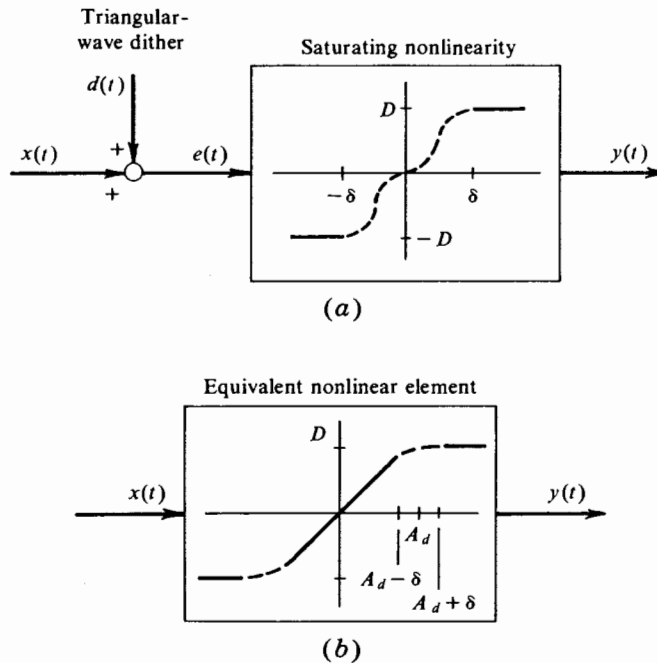


Figure 6.7-1 (a) Application of triangular-wave dither to an arbitrary odd saturating nonlinearity and (b) the equivalent nonlinear element.

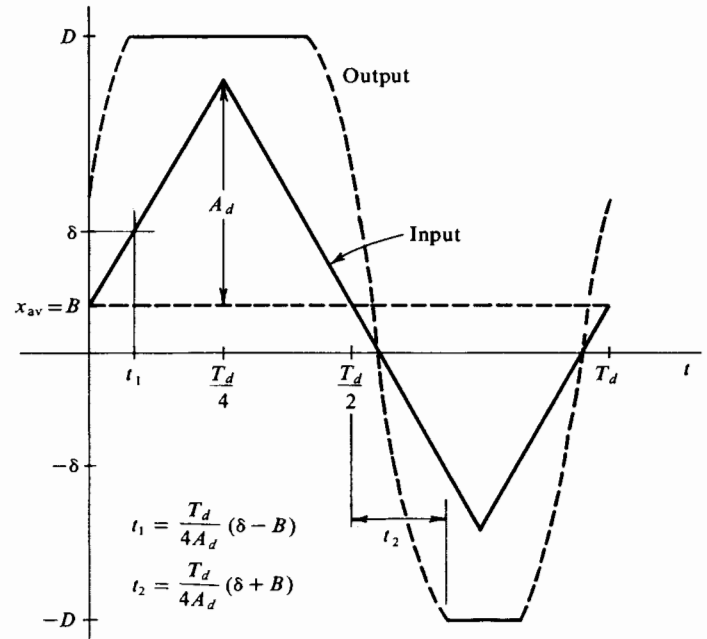


Figure 6.7-2 Input and output waveforms for an odd saturating nonlinearity with triangular-wave dither.

dither frequency is by definition significantly higher than the frequencies of interest in x , the input signal is modeled as a bias over one dither cycle.

The signal gain of this device is, by analogy with previous DIDF formulations,

$$N_B(A_d, B) = \frac{y_{av}}{B} \quad (6.7-2)$$

In terms of the parameters of the illustration, the average output is given by (T_d = dither period)

$$\begin{aligned} y_{av} &= \frac{1}{T_d} \left[D \left(\frac{T_d}{2} - 2t_1 \right) - D \left(\frac{T_d}{2} - 2t_2 \right) \right] \\ &= \frac{BD}{A_d} \end{aligned} \quad (6.7-3)$$

Notice that no account need be taken of the dotted portion of the nonlinearity output, since over a complete cycle it has an average value of zero. It is for this reason that arbitrary nonlinearity characteristics can be admitted,

provided they are odd. Hence, for $|B| < A_d - \delta$, the signal gain of this device is

$$N_B(A_d, B) = \frac{D}{A_d} \quad (6.7-4)$$

which, remarkably, is independent of δ . The inequality requirement on B points out that we have assumed the nonlinearity output to saturate positively and negatively during every dither cycle. For values of B in the range

$$A_d - \delta < |B| < A_d + \delta$$

the signal gain depends upon the particular nonlinearity under consideration. Values of B satisfying

$$|B| > A_d + \delta$$

lead to a constantly saturated nonlinearity output. Figure 6.7-1b summarizes these results by illustrating the resulting equivalent nonlinear element. Of great consequence is the fact that *the equivalent nonlinear element appears absolutely linear about the origin, irrespective of the shape of the original nonlinearity about the origin*. The range of linearity can be increased simply by increasing A_d , accompanied, of course, by a decrease in linear gain. For sufficiently large input signals the equivalent nonlinear element saturates at the output saturation value of the original nonlinearity.

By very similar reasoning it can be demonstrated that these results carry over to the entire class of odd saturating nonlinearities with memory.

Square-wave dither Provided that the dither plus bias drive the output to saturation on the positive *and* negative parts of every dither cycle, the average output of the nonlinear element is identically zero for all nonlinearities given by Eqs. (6.7-1). Continuing as before yields, for the arbitrary symmetric nonlinearity with square-wave dither shown in Fig. 6.7-3a, the equivalent nonlinearity of Fig. 6.7-3b. A_d is the square-wave amplitude. As before, dotted line segments refer to portions of the characteristics which vary from one nonlinearity to another.

In this case, the noteworthy fact is that *the equivalent nonlinear element possesses a dead zone about the origin, regardless of the shape of the original nonlinearity about the origin*. Again, the results apply to nonlinearities with memory, as well as to those without memory.

Sinusoidal dither The case of sinusoidal dither is identical with that for which the DIDF calculations of previous sections were performed. No generalization can be made here, except that, for sufficiently large A_d , the effect of sinusoidal dither is similar to that of triangular-wave dither in that *all* nonlinearities are linearized about the origin. In the case of sinusoidal dither, the assumption of negligible fed-back dither signal need not be held.

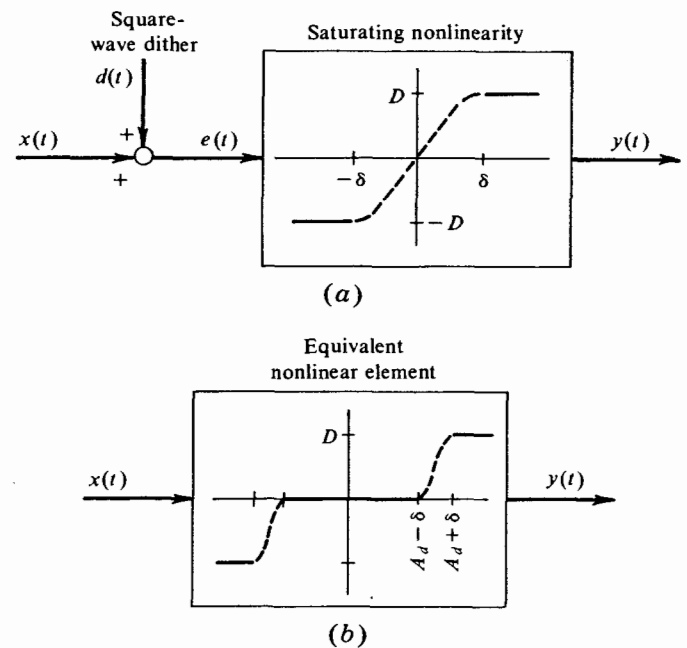


Figure 6.7-3 (a) Application of square-wave dither to an arbitrary odd saturating nonlinearity and (b) the equivalent nonlinear element.

With additional labor, but with no new analytic means required, the feedback sinusoid can be included in determining the total sinusoidal input to the nonlinearity.

Additional remarks It follows from the discussion of the effect of artificial dither that the ideal-relay characteristic, for example, can be manipulated as shown in Fig. 6.7-4. Similar results hold for many other nonlinearities. Thus a practical means for altering the properties of a control loop is that of simply changing the *waveform* of an additive dither signal. In two of the three cases shown, the effect of this dither is to *linearize* the nonlinearity for small input signals, the triangular dither providing a larger linear region than the sinusoidal dither. This technique is profitably employed in the compensation of nonlinear control systems, often for the purpose of linearizing the highly nonlinear force-velocity effects due to friction. Since the apparent nonlinearity characteristics are dependent upon A_d , the use of dither of fixed waveshape but variable amplitude allows for continuous *abscissa scaling* of each of the apparent nonlinearity characteristics illustrated. In a following servo, for example, if A_d were instrumented to be a function nearly inversely proportional to the servo error signal, the result would be a high loop gain

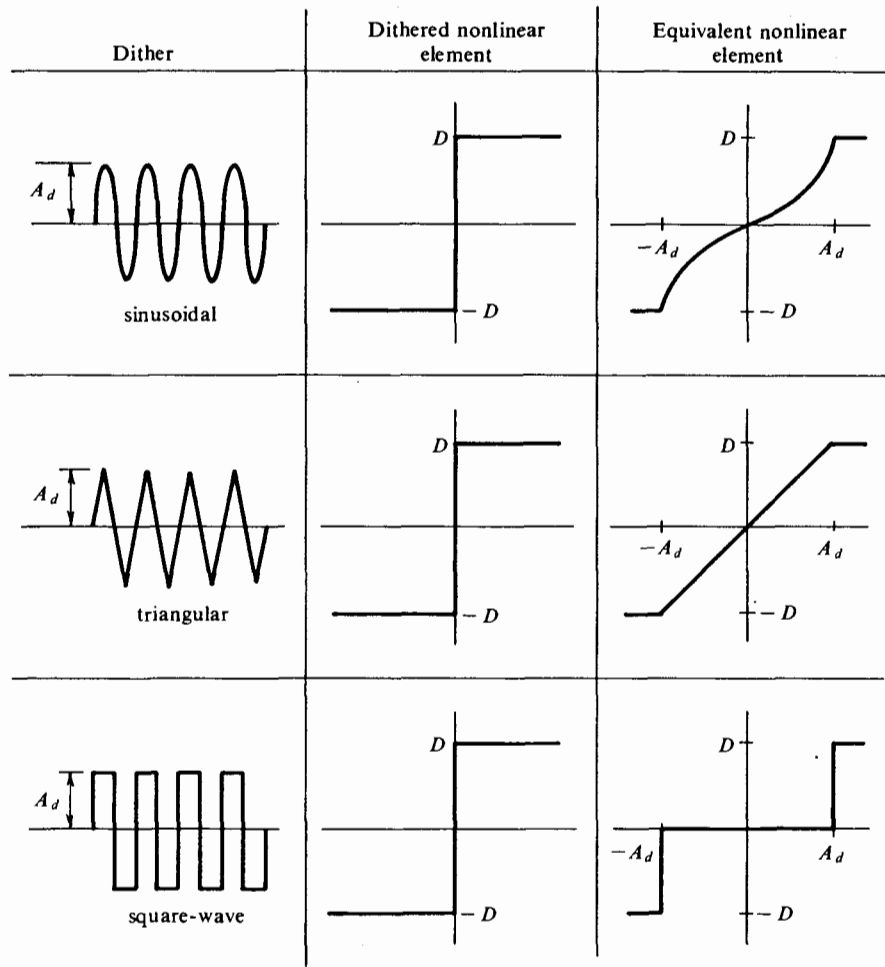


Figure 6.7-4 Ideal relay with three types of dither and the associated equivalent nonlinear characteristics.

at large error signals and a lower loop gain at small error signals. In such a way one can achieve the continuous analog of certain dual-mode servos which attempt to give near-maximum-effort performance without displaying a limit cycle.

SOME APPLICATIONS FOR ARTIFICIAL DITHER

In addition to the more common uses of artificial dither such as linearization of coulomb friction, threshold, and hysteresis effects, Naslin (Ref. 15) describes several other interesting applications.

Consider the use of dither for the purpose of providing a proportional torque zone in a relay-controlled split-field series motor (see Fig. 6.7-5). Assuming the motor to be fed by a constant current source (thus minimizing back emf effects), its output torque can take on only two values, $\pm T$, depending upon the state of the switch. The switch has two additive inputs, a signal $x_B(t)$ and dither $d(t)$. By previous DIDF arguments the average output torque is seen to be a linear function of the input signal $x_B(t)$ for small values of $|x_B/A_d|$. This concept of torque control can also be applied to other forms of energy. Basically, it amounts to a form of pulse-width modulation. Observe that grid-controlled rectifiers and some magnetic amplifiers operate in the very same manner.

Next, consider the use of dither for controlling small guided missiles by means of spoilers. Spoilers are vanes that disturb normal airfoil airflow, thus producing aerodynamic forces. The average aerodynamic force becomes a linear function of the input signal as the result of relay-controlling spoiler position such that it is either fully retracted or fully protruding, and supplying a linearizing dither signal to the relay, in addition to a signal input. Hence a linear region of control is effected.

Another application involves a position servo incorporating a photocell error detector with an *absolute-value* characteristic. This device was built for the purpose of providing an instrumented signal representation of the height of a mercury column, in a form suitable for telemetering. The servomanometer functional diagram is shown in Fig. 6.7-6. A lead-screw positioned carriage houses a photocell which detects whatever light flux (shown dotted) passes through the mercury-column-supported slit and over or under the 50-cps sinusoidally oscillating carriage-mounted diaphragm. The photocell output is filtered for recovery of whatever 50-cps content it contains, then amplified and fed to the control winding of a two-phase motor

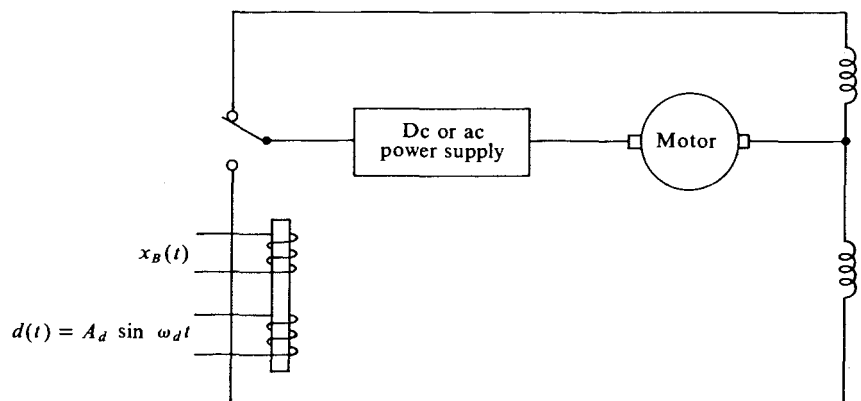


Figure 6.7-5 Relay-controlled split-field series motor. (Naslin, Ref. 15.)

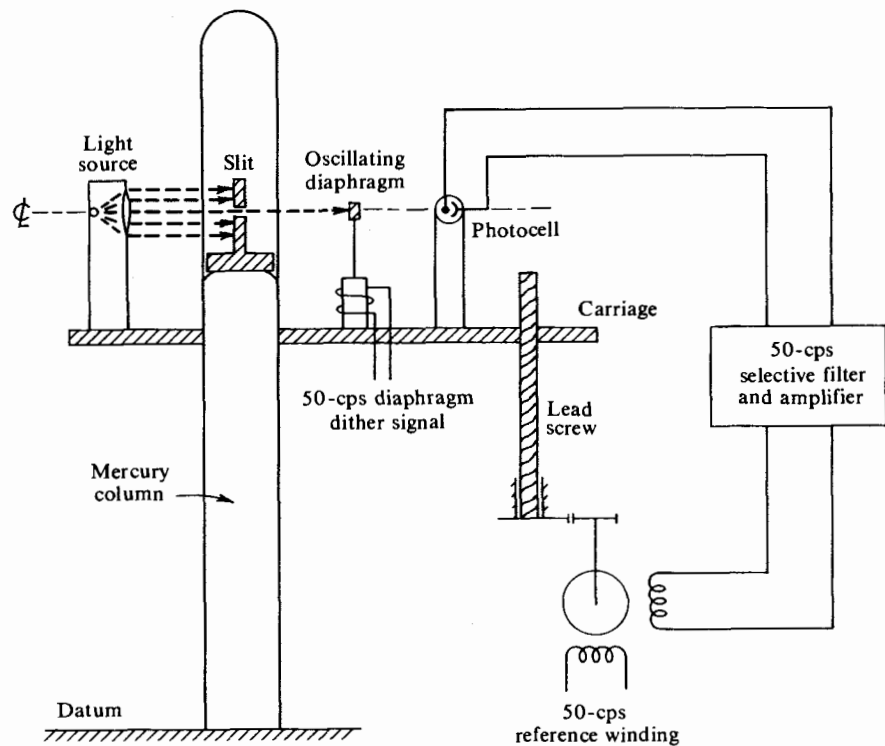


Figure 6.7-6 Servomanometer. (Naslin, Ref. 15.)

which drives the lead screw. Lead-screw angular position is then the instrumented quantity which is sensed by a shaft-angle encoder and telemetered.

To understand the behavior of this system it is necessary to determine the relationship between photocell output and slit-position input (see Fig. 6.7-7). Specializing DIDF calculations already presented for the two-segment piecewise-linear asymmetric nonlinearity to the case of an absolute-value characteristic ($m_1 = -m_2 = m$), and assuming $|e/A_d| < 1$, where $e = r - c$, yields [Eq. (6.2-11)]

$$\begin{aligned}
 N_A(A_d, e) &= \frac{2m}{\pi} \left[\sin^{-1} \left(\frac{e}{A_d} \right) + \frac{e}{A_d} \sqrt{1 - \left(\frac{e}{A_d} \right)^2} \right] \\
 &\approx \frac{4me}{\pi A_d}
 \end{aligned} \tag{6.7-5}$$

and [Eq. (6.2-14)]

$$N_B(A_d, e) = \frac{2m}{\pi} \left[\sin^{-1} \left(\frac{e}{A_d} \right) + \frac{A_d}{e} \sqrt{1 - \left(\frac{e}{A_d} \right)^2} \right] \approx \frac{2mA_d}{\pi e} \quad (6.7-6)$$

Hence the photocell output is given by

$$y(t) \approx \frac{2mA_d}{\pi} + \frac{4me}{\pi} \sin 100\pi t \quad (6.7-7)$$

where second-order and double- and higher-frequency terms have been dropped. Equation (6.7-7) exhibits a most interesting effect, namely, that *the amplitude of the first-harmonic term in $y(t)$ is directly proportional to the positioning error e* . The photocell plus dither act, in effect, as a modulator. Operation of the remainder of the system is clear, and need not be discussed further. The overall system accuracy is reported to be 0.1 mm of vertical mercury motion, or 0.002 psi.

Other diverse applications for artificial dither include extremal control, adaptive control, and signal stabilization. The last-mentioned application is briefly discussed next.

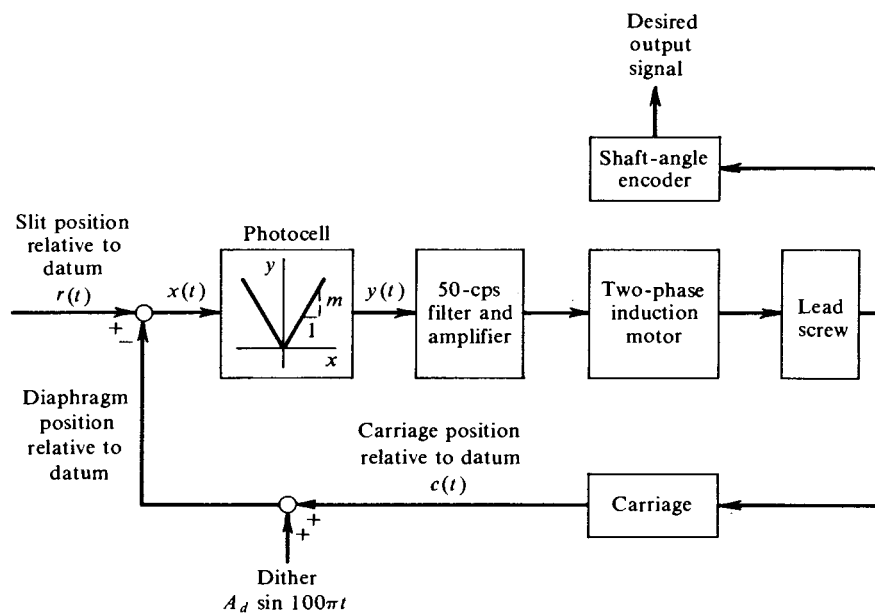


Figure 6.7-7 Block diagram of the servomanometer.

SIGNAL STABILIZATION

Since the injection of dither has the *effect* of altering nonlinear characteristics in closed-loop systems, it is reasonable to expect that limit cycles in nonlinear systems can be turned on, altered, turned off, and, in general, controlled by proper choice of dither waveshape. The use of dither to turn limit cycles off is referred to as *signal stabilization*. It has been extensively investigated by Oldenburger and his students (Refs. 16–19), among the first to discover this phenomenon experimentally and, subsequently, to provide analytical justification. They point out that the analytical justification for signal stabilization differs from that given by Minorsky (Ref. 13) relating to the asynchronous quenching of systems characterized by nonlinear differential equations.

Investigation of signal stabilization via describing function theory can be executed as a two-stage process. First, by use of DIDF theory, the dither and original nonlinear element are replaced with an equivalent nonlinear element, whose form implicitly accounts for the presence of dither, but which no longer explicitly displays the dither signal. Second, the resulting system is made the object of a DF analysis to reveal the presence or absence of a limit cycle. This rather simple investigation scheme is demonstrated by the following example.

Example 6.7-1 Find the minimum sinusoidal dither amplitude $A_{d,min}$ for which the system of Fig. 6.7-8a just fails to develop a limit cycle, given that $\omega_d \gg 1$.¹

First, the relay with dither is replaced by its equivalent nonlinear element (signal DIDF) (Fig. 6.7-8b). From this point analysis proceeds by determination of the DF for this equivalent nonlinear element and subsequent performance of a DF limit cycle analysis of the resulting system. Note that the limit cycle, if any, plays the role previously played by system error *signals*, and the dither plays the role otherwise belonging to a limit cycle.

The DF of the signal DIDF is determined as follows ($k = A/A_d < 1$):

$$N(A) = \frac{4}{\pi A} \int_0^{\pi/2} \frac{2D}{\pi} \sin^{-1}(k \sin \psi) \sin \psi \, d\psi$$

Integrating by parts and grouping terms yields

$$\begin{aligned} N(A) &= \frac{8D}{\pi^2 A} \left(\frac{1}{k} \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 \psi} \, d\psi + \frac{k^2 - 1}{k} \int_0^{\pi/2} \frac{d\psi}{\sqrt{1 - k^2 \sin^2 \psi}} \right) \\ &= \frac{8D}{\pi^2 A_d k^2} [E(k) + (k^2 - 1)K(k)] \end{aligned} \quad (6.7-8)$$

where $E(k)$ and $K(k)$ denote the complete elliptic integrals of the first and second kind, respectively [Eq. (5.1-18)].

¹ This system was studied by Oldenburger and Boyer (Ref. 19), although they did not determine the required DF analytically. Their results are, of course, identical with those given here.

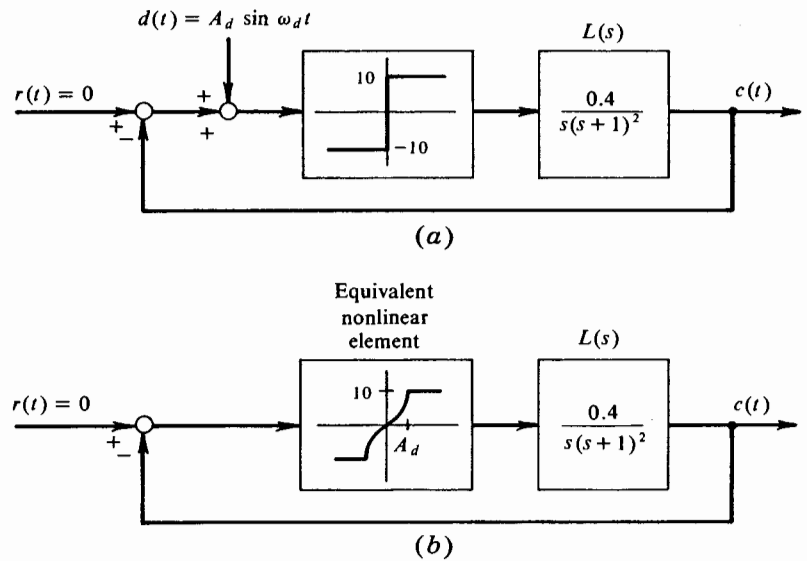


Figure 6.7-8 (a) Relay control system with stabilizing signal added. (b) Equivalent nonlinear system.

Similarly, it is readily shown that for $k > 1$ the DF is given by

$$N(A) = \frac{8D}{\pi^2 A_d k} E\left(\frac{1}{k}\right) \quad (6.7-9)$$

A normalized plot of $N(A)$ over a range of k is shown in Fig. 6.7-9. Since all devices which ultimately saturate at fixed levels possess DFs which are asymptotic to the ideal-relay DF for large inputs, this curve is included for reference.

Returning to the problem, it is to be observed that any limit cycle which takes place must do so at a frequency $\omega_0 = 1$,¹ where the plant phase shift is -180° . At this frequency the remaining condition for a limit cycle is

$$N(A) |L(j1)| = 1$$

Since $|L(j1)| = 0.2$, it follows that

$$N(A) = 5$$

Considering $-1/N(A)$ as plotted in the amplitude-phase plane, it is clear that the lowest point of this locus corresponds to the peak in the curve of Fig. 6.7-9. So adjusting this peak such that the $-1/N(A)$ locus never cuts the $L(j\omega)$ locus guarantees the absence of a limit cycle (to whatever extent one is willing to make quantitative guarantees based on approximate analysis). The required condition is

$$N_{\max} < 5 \quad \text{or} \quad \left(\frac{A_d N}{D}\right)_{\max} < \frac{5A_d}{D}$$

¹ For which reason the dither frequency was specified as greater than 1; that is, $\omega_d \gg \omega_0$. So one can neglect the fed-back dither when computing the equivalent nonlinearity.

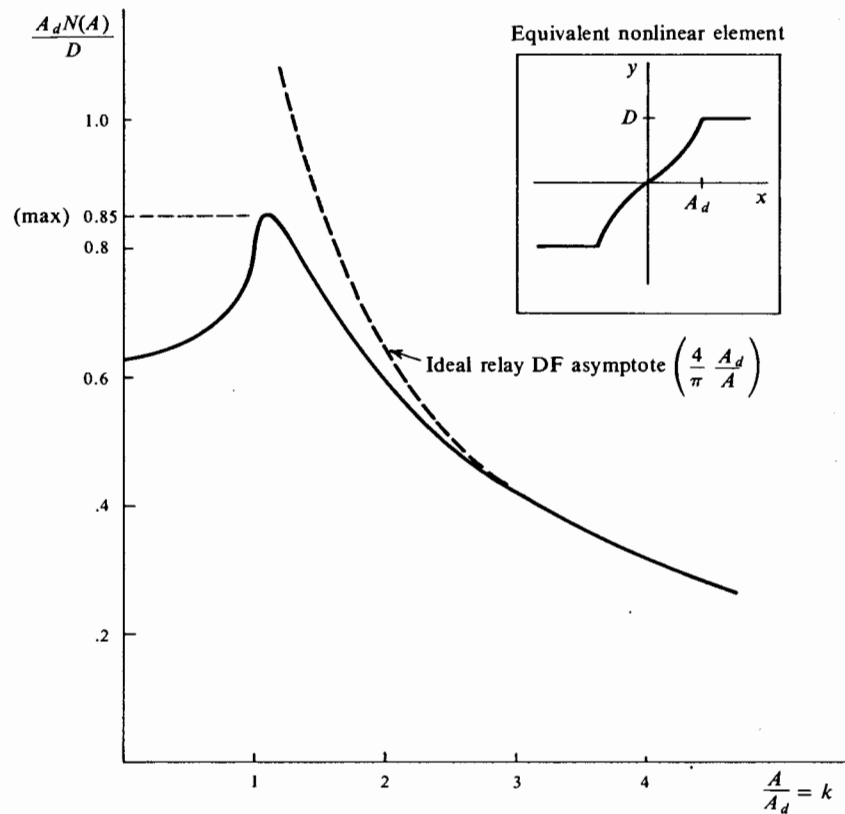


Figure 6.7-9 Equivalent nonlinear element DF for a sinusoidally dithered ideal-relay characteristic.

From the DF curve, we find $(A_d N/D)_{\max} = 0.85$; and from the problem statement $D = 10$; whence the condition on A_d for signal stabilization is

$$A_d > \frac{10(0.85)}{5} = 1.7 \text{ units}$$

or

$$A_{d,\min} = 1.7 \text{ units} \quad (6.7-10)$$

For values of A_d between 1.26 and 1.7 units, two limit cycles are indicated, of which only that of larger amplitude is stable. For values of A_d less than 1.26 units, a single stable limit cycle occurs.

Analysis as presented in the foregoing example is subject to the usual DF and DIDF limitations, such as the assumption of a single time-invariant loop nonlinearity, the absence of nonlinearity subharmonic generation, and satisfaction of the filter hypothesis. In addition, it is a convenience if only a negligible amount of dither returns via the feedback loop to the nonlinearity

input. As a rule of thumb, Oldenburger et al. suggest that the dither frequency be at least 10 times the highest possible limit cycle frequency, an assumption that is readily verified during analysis. This is roughly the same rule of thumb which ought to be used in design of a limit cycling control system, where the frequency ratio of 10 there refers to limit cycle frequency over highest significant input frequency.

The analytic study of signal stabilization as described above is contingent upon our ability to determine DFs for the equivalent nonlinear element under consideration. No difficulty is likely to arise in obtaining the equivalent nonlinear element itself. The DF can then be calculated by using the approximation techniques of Sec. 2.6. Occasionally, the DF can be readily determined analytically. This is shown for the ideal relay in Example 6.7-1, and can also be done for odd polynomial nonlinearities. In the case of a sinusoidally dithered cubic characteristic, for instance, the equivalent nonlinear element is given by [Eq. (6.2-27)]

$$y = \frac{3}{2}A_d^2x + x^3 \quad (6.7-11)$$

Employing Eq. (2.3-21) yields immediately the DF as

$$N(A) = \frac{3}{2}A_d^2 + \frac{3}{4}A^2 \quad (6.7-12)$$

6.8 TSIDF CALCULATION VIA THE DF OF A DIDF

The reader may have already noticed that the DF calculations made in the previous section yielded the exact TSIDF linearization for the ideal-relay and cubic characteristics. This is by no means just coincidence; one *can* compute the TSIDF for a nonlinearity by obtaining the DF of the equivalent nonlinear element. This is readily proved.

TSIDF CALCULATION

We have seen the output of a nonlinearity expanded in a double Fourier series (Sec. 5.1). In the case of single-valued frequency-independent nonlinearities with non-harmonically-related input sinusoids, the TSIDF for one of the two input sinusoids is given by [Eq. (5.1-15)]

$$\begin{aligned} N_B(A, B) &= \frac{1}{2\pi^2 B} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} y(A \sin \psi_1 + B \sin \psi_2) \sin \psi_2 d\psi_1 d\psi_2 \\ &= \frac{1}{\pi B} \int_{-\pi}^{\pi} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} y(A \sin \psi_1 + B \sin \psi_2) d\psi_1 \right] \sin \psi_2 d\psi_2 \quad (6.8-1) \end{aligned}$$

Notice that ψ_2 is held *constant* in the integral within brackets; the argument of y is equivalently a sinusoid plus a bias. In DIDF notation, this bracketed term is represented as $BN_B(A,B)$, the bias portion of the nonlinearity output. Alternatively, it is the output of the equivalent nonlinear element. The remainder of Eq. (6.8-1) is, of course, the DF formulation. Hence the TSIDF is indeed the DF of the equivalent nonlinear element! This proof must be restricted to the case of non-harmonically-related input sinusoids, although it can be generalized to include frequency-dependent nonlinearities. This is, in fact, a special case of the general property of *any other* independent inputs expressed in Eq. (1.5-41).

TSIDF APPROXIMATION

It is of some interest to determine the circumstances under which the DIDF itself is a suitable approximation to the TSIDF. The conditions under which this use can be made of the DIDF surely cannot be stated very generally. Fortunately, it is possible to determine qualitatively the range of usefulness of this approximation without doing the more tedious calculation of the TSIDF.

Two approaches to this problem appear in Ref. 1. We note the result here, without laboring through the related mathematics. It is, simply, that if the DIDFs $N_A(A,B)$ and $N_B(A,B)$ are nearly independent of B for some range of $|B| \leq A$, then the approximate TSIDFs given by $N_A(A,0)$ and $N_B(A,0)$ are valid in this range. Since these are seen to be the leading terms in the TSIDF power-series expansions of Sec. 5.1, this approximation warrants no further discussion.

6.9 BASIS FOR HIGHER-ORDER APPROXIMATIONS

One can obtain higher-order DIDF approximations in several ways. For nonlinearities which are nonanalytic, a procedure entirely equivalent to the refined DF approximation of Sec. 3.7 can be followed. In this instance the fed-back residual would be treated as altering the limit cycle DIDF only, leading to a second-approximation limit cycle DIDF. In the case of analytic nonlinearities a second-order approximation can be obtained by perturbing the original DF solution and satisfying perturbation-equation first-order terms derived from the system differential equation. Finally, for limit cycling relay control systems with either command inputs which result in error-signal bias terms or asymmetric relays or both, Tsytkin's method, presented in Sec. 3.8, can be modified and otherwise directly extended to yield exact solutions.

REFERENCES

1. Atherton, D. P., G. F. Turnbull, A. Gelb, and W. E. Vander Velde: Discussion of the Double Input Describing Function (DIDF) for Unrelated Sinusoidal Signals, *IEEE Trans. Autom. Control*, vol. AC-9, no. 2 (April, 1964), pp. 197-198.
2. Furman, G. G.: Removing the Noise from the Quantization Process by Dithering: Linearization, *RAND Mem.* RM-3271-PR (February, 1963), pp. 1-40.
3. Gelb, A.: "The Analysis and Design of Limit Cycling Adaptive Automatic Control Systems," Sc.D. thesis, Massachusetts Institute of Technology, Cambridge, Mass., August, 1961.
4. Gelb, A.: The Foundation for Limit Cycling Adaptive Design, *Proc. Northeast Electronics Res. Eng. Meeting* (November, 1961), pp. 76-77.
5. Gelb, A.: The Dynamic Input-Output Analysis of Limit Cycling Control Systems, *Proc. JACC, New York University*, New York (June, 1962), pp. 9.3-1-9.3-11.
6. Gelb, A., and T. C. Blaschke: Design for Adaptive Roll Control of a Missile, *J. Aeron. Astronautics*, vol. 9, no. 4 (Winter Issue, 1962), pp. 99-105.
7. Gelb, A., and W. E. Vander Velde: On Limit Cycling Control Systems, *IEEE Trans. Autom. Control*, vol. AC-8, no. 2 (April, 1963), pp. 142-157.
8. Korolev, N. A.: Pulse Stabilization of Automatic Control Relay Systems, *Automation and Remote Control*, vol. 18, no. 5 (April, 1958), pp. 435-446.
9. Li, Y. T., and W. E. Vander Velde: Philosophy of Nonlinear Adaptive Systems, *Proc. First IFAC Congr.*, Moscow, U.S.S.R., 1960, Butterworth Scientific Publications, London.
10. Loeb, J. M.: A General Linearizing Process for Nonlinear Control Systems, *Manual and Autom. Control*, 1952, pp. 274-284, Butterworth Scientific Publications, London.
11. Lozier, J. C.: Carrier-controlled Relay Servos, *Elec. Eng.*, vol. 69 (December, 1950), pp. 1052-1056.
12. MacColl, L. A.: "Fundamental Theory of Servomechanisms," D. Van Nostrand Company, Inc., Princeton, N.J., 1945.
13. Minorsky, N.: On Asynchronous Action, *J. Franklin Inst.*, vol. 59 (March, 1955), pp. 209-219.
14. Mishkin, E., and L. Braun, Jr.: "Adaptive Control Systems," McGraw-Hill Book Company, New York, 1961.
15. Naslin, P.: A Simplified Theory of Feedback Control Systems, Part 12, *Process Control and Automation*, vol. 6, no. 6 (June, 1959), pp. 273-277.
16. Oldenburger, R.: Signal Stabilization of a Control System, *Trans. ASME*, vol. 79 (August, 1957), pp. 1869-1872.
17. Oldenburger, R., and C. C. Liu: Signal Stabilization of a Control System, *Trans. AIEE*, vol. 78 (May, 1959), pp. 96-100.
18. Oldenburger, R., and T. Nakada: Signal Stabilization of Self-oscillating Systems, *IRE Trans. Autom. Control*, vol. AC-6, no. 3 (September, 1961), pp. 319-325.
19. Oldenburger, R., and R. C. Boyer: Effects of Extra Sinusoidal Inputs to Nonlinear Systems, *Trans. ASME*, ser. D, *J. Basic Eng.*, vol. 84, no. 4 (December, 1962), pp. 559-570.
20. Popov, E. P.: "The Dynamics of Automatic Control Systems," Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962.
21. Popov, E. P., and N. P. Pal'tov: "Priblizhennyye metody issledovaniya nelineinykh avtomaticheskikh sistem," Gosudarstvennoe Izdatel'stvo Fiziki i Matematicheskoi Literatury, Moscow, 1960. English translation: "Approximate Methods for Analyzing Non-linear Automatic Systems," Wright Patterson Air Force Base, Ohio, Translation Services Branch, Foreign Technology Division, January, 1963.

22. Shuck, O. H.: Honeywell's History and Philosophy in the Adaptive Control Field, *Wright Air Develop. Center, Tech. Rept.*, January, 1959, Wright Patterson Air Force Base, Ohio.

PROBLEMS

- 6-1. Show that any single-valued *asymmetric* nonlinearity can be represented as the parallel combination of one odd [$y(x) = -y(-x)$] and one even [$y(x) = y(-x)$] *symmetric* nonlinearity. What are the corresponding odd and even elements which comprise the two-segment piecewise-linear asymmetric nonlinearity

$$y(x) = \begin{cases} m_1 x & x \geq 0 \\ m_2 x & x < 0 \end{cases}$$

Calculate the DIDFs for these odd and even elements and sum the results, thus arriving at Eqs. (6.2-11) and (6.2-14).

- 6-2. (a) Show that the incremental-input describing function for a limiter of input breakpoints $\pm\delta$ and output saturation levels $\pm D$ is given by

$$N_B(A) = \frac{2D}{\pi\delta} \sin^{-1} \frac{\delta}{A}$$

Perform this calculation twice, once by taking the limit of $N_B(A, B)$ as $B \rightarrow 0$, and once by applying Eq. (6.1-9) directly.

(b) Compute the incremental-input describing function for an ideal relay directly in terms of Eq. (6.1-9). Note that the weights (i.e., strengths) of the impulse functions in y' are not unity.

- 6-3. Compute the limit cycle and signal DIDFs for an asymmetric polynomial nonlinearity described by

$$y(x) = \begin{cases} cx^2 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- 6-4. Compute signal and limit cycle DIDF expressions for the energy-storing nonlinear element described by

$$y(x, \dot{x}) = x(\dot{x})^2$$

- 6-5. Show that N_B for an odd square-law characteristic is given by

$$N_B = \frac{A^2}{\pi B} \left\{ \left[1 + 2 \left(\frac{B}{A} \right)^2 \right] \sin^{-1} \left(\frac{B}{A} \right) + 3 \frac{B}{A} \sqrt{1 - \left(\frac{B}{A} \right)^2} \right\}$$

and hence derive the result that for small B/A the DF of the equivalent nonlinear element is $1.27A$.

- 6-6. Show that when n is an even integer, the DIDFs of Eqs. (6.2-21) and (6.2-23) are in error by a factor of the order of $(B/A)^{2n}$. In particular, for $n = 2$, show that $N_A(A, B)$ is in error by the amount $(A/3\pi)(B/A)^4$.

- 6-7. A frequently used measure of the dependence of closed-loop transmission F_{cl} upon an environmental parameter ξ is the sensitivity function $S_\xi^{F_{cl}}$. It is defined as the ratio of per-unit change in F_{cl} to per-unit change in ξ , viz.,

$$S_\xi^{F_{cl}} = \frac{dF_{cl}/F_{cl}}{d\xi/\xi}$$

Now consider two unity feedback systems with third-order linear elements $L(s)$,

one with an additional linear gain K , and one with a nonlinearity N . Assume $L(s)$ to depend in some way upon ξ . Let F_{ol} denote the open-loop transfer function.

(a) For the linear system show that

$$S_{\xi}^{F_{cl}} = \frac{1}{1 + F_{ol}} S_{\xi}^L$$

(b) For the nonlinear system show that

$$S_{\xi}^{F_{cl}} = \frac{1}{1 + F_{ol}} (S_{\xi}^L + S_{\xi}^{N_B})$$

(c) If ξ affects only the gain constant of $L(s)$, and the nonlinearity is an ideal relay ($N_B \approx 2D/\pi A$), show that $S_{\xi}^{F_{cl}} = 0$. (Hint: A useful relationship is $S_{\xi}^{F_{cl}} = S_{\xi}^{F_{ol}} S_{\xi}^{F_{cl}}$).

6-8. (a) Find the complete description of the equivalent nonlinear element describing a relay with dead zone subjected to triangular-wave dither.

(b) Repeat (a) for a limiter with square-wave dither.

6-9. Find the equivalent nonlinear characteristic corresponding to an ideal relay with a sawtooth dither input, shown in Fig. 6-1.

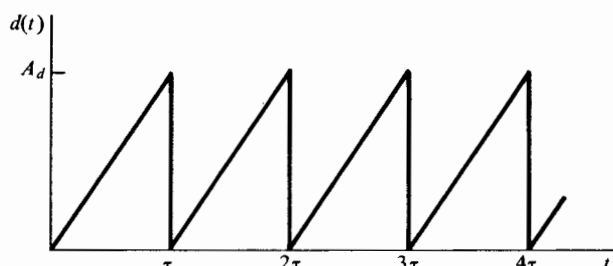


Figure 6-1 Sawtooth dither waveform.

6-10. The voltage induced in the antenna of a radio compass varies with the antenna orientation with respect to the transmitter, as shown in Fig. 6-2. Devise and analyze a system for making the radio compass direction sensitive (Naslin, *op. cit.*).

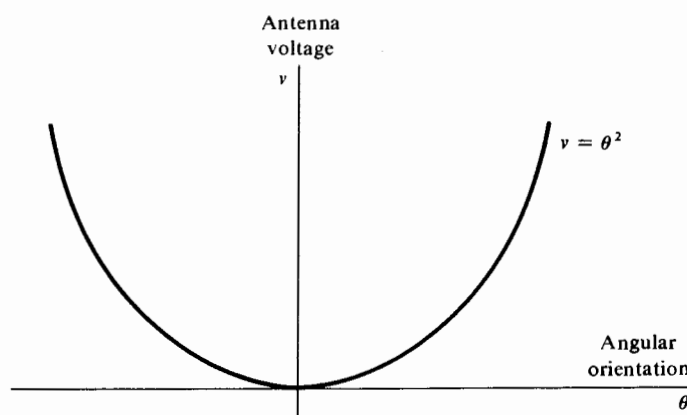


Figure 6-2 Even symmetric antenna voltage-orientation characteristic.

- 6-11. Determine the three nonlinear simultaneous equations which describe limit cycle behavior of the system shown in Fig. 6-3. Discuss an approach to the graphical solution of these equations. Plot the normalized quantities, A/δ , B/δ , $\omega_0\tau$ versus $DK\tau/\delta$.

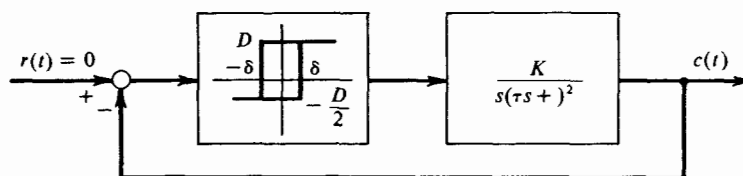


Figure 6-3 System with an output-biased rectangular hysteresis nonlinearity.

- 6-12. Find the limit cycle state of the system of Fig. 6-4 containing an asymmetric relay characteristic, if

(a)
$$L(s) = \frac{sK}{(\tau s + 1)^4}$$

(b)
$$L(s) = \frac{K}{(\tau s + 1)^3}$$

(c)
$$L(s) = \frac{K}{s(\tau s + 1)^2}$$

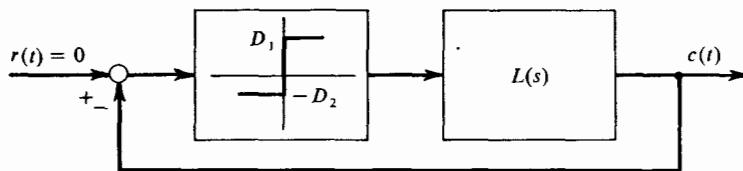


Figure 6-4 System with an asymmetric relay nonlinearity.

- 6-13. Describe the limit cycle state of a system containing an input-biased ideal relay with symmetric drive levels, described by

$$y(x) = \begin{cases} D & x > \delta \\ 0 & x = \delta \\ -D & x < \delta \end{cases}$$

and the linear elements

$$L(s) = \frac{K}{s(\tau s + 1)^2}$$

- 6-14. Design a simple adaptive controller for the variable plant $L(s) = 10\tau/(\tau s + 1)^3$, where $0.1 < \tau < 10$. Draw and interpret the closed-loop frequency response curves for the resultant system in the extremes $\tau = 0.1$ and $\tau = 10$. Compute the ramp response of the adaptive system. Design a simple gain-compensated linear closed-loop system for $L(s)$, and compare this system to the adaptive system.

- 6-15. The system of Fig. 6-5 has a square-law error detector. It can be made to respond to input signals by biasing the input signal off zero. Plot the bias at the output $c(t)$, and the input-output frequency response function for all significant frequencies with $B = 5$ units and $r(t) = 5 \sin \omega t$. What would be the effect of a negative bias?

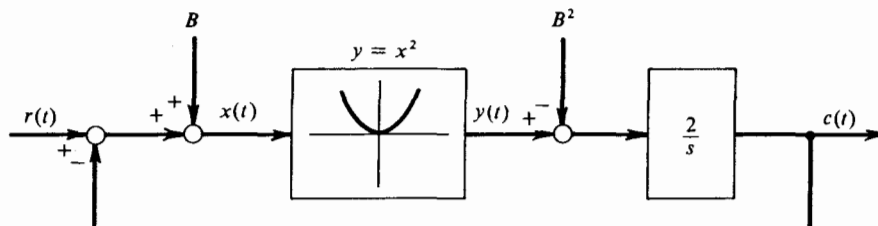


Figure 6-5

- 6-16. Consider the system of Fig. 6-6. Determine the limit cycle frequency and amplitude with the describing function method for $\tau_1 = 0$. Select a τ_1 to reduce the limit cycle amplitude to one-half of the uncompensated value, and determine the average steady-state error when the compensated system is driven by a ramp function $r(t) = 5t$.

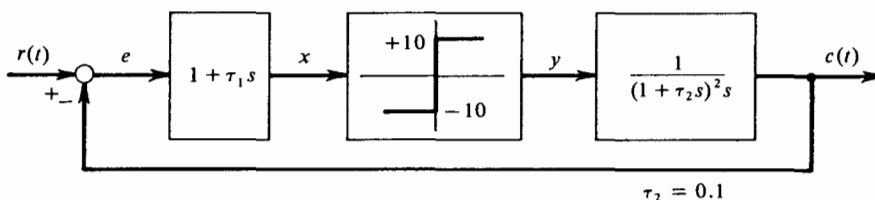


Figure 6-6

- 6-17. Design compensation for the inverted pendulum-balancing system shown in Fig. 6-7. The specifications are that the amplitude of the limit cycle at θ must be no greater than 5° and the error between θ and θ_r must be no greater than 1° for all constant $|\theta_r| \leq 30^\circ$. A current switch is controlled by the sign of $x(t)$. The steady-state current produces a torque $M = 10$ lb-ft, but there is a first-order lag ($\tau = 0.1$ sec) in the current, and hence torque, buildup. The pendulum is essentially ideal, with length 1 ft and weight 10 lb. It is to be controlled to an angle off vertical (inverted) given by θ_r .

For your design, plot the average error, the limit cycle amplitude and frequency, and some measure of the dominant system response characteristics as functions of θ_r for $0 \leq \theta_r \leq 30^\circ$.

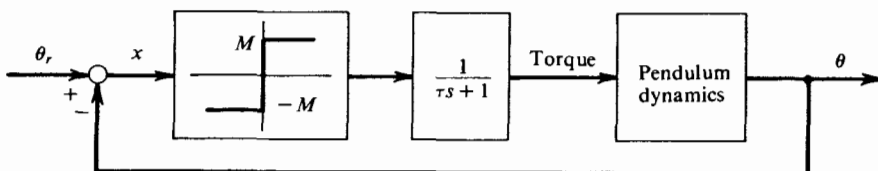


Figure 6-7

- 6-18. The laser gyroscope exhibits an input-rate (ω_i) output-frequency difference (Δf) characteristic which displays a dead zone due to "lock-in" (frequency entrainment), which arises from coupling between the two counterrotating laser beams. One method of enabling use of this gyro at low input rates is to apply a mechanical input-rate dither, to linearize the dead zone.

Assume that the nonlinear laser characteristic is described by

$$\begin{aligned}\Delta f &= K\omega_i & \omega_i < -\delta, \omega_i > \delta \\ &= 0 & -\delta \leq \omega_i \leq \delta\end{aligned}$$

Describe the equivalent gyro characteristic resulting from an applied dither signal of the form

$$d(t) = A_d \sin \omega_d t \quad A_d > \delta$$

7 DESCRIBING FUNCTIONS FOR NONLINEAR SYSTEMS WITH RANDOM INPUTS

7.0 INTRODUCTION

The preceding chapters have dealt with approximate descriptions of nonlinearities having inputs consisting of the sums of two commonly considered signal forms, sinusoids and constants. We wish now to add a third form to this repertory of input signals, a random process. The study of nonlinear systems with random inputs is of consequence both because the design of many high-performance systems is significantly influenced by the presence of undesired random noise and because the signal inputs expected in many operational circumstances do not permit description in deterministic terms. Realistic input signals often can be characterized only as members of an ensemble of possible input functions, the ensemble having certain statistical properties which are known or can be estimated.

Actually, the general theory of describing functions developed in Chap. 1 was couched in statistical terms at the outset. This was done so most signal forms of interest at the nonlinearity input could be included in a single format. To this end, sinusoidal and bias signals have been treated as special

cases of random processes. This led to describing functions for these signals which have the more familiar interpretation in terms of harmonic analysis.

In this chapter we consider, in addition to constants and sinusoids, random signals of the ordinary sort which do not have characteristic waveshapes. In this case a statistical approach is clearly essential, and no alternative deterministic interpretation seems possible. There is a considerable literature on the subject of quasi-linearization of nonlinear elements driven by random signals. Much of this is related to random signals having finite power density spectra, which rules out consideration of biases and sinusoids. This literature is briefly reviewed in the following section, after which the calculation and application of describing functions for systems with random inputs is discussed. A review of probability theory, random variables, and random processes appears in Appendix H for convenient reference.

7.1 STATISTICAL LINEARIZATION

A time-honored procedure for dealing in a practical way with nonlinear systems is to construct a linear model to approximate the nonlinearities. Thus, for example, a small-signal linearization consists of expanding the nonlinear function in a Taylor series about some operating point and retaining only the linear terms in the analysis. If the signals at the input to the nonlinearity are not small enough to permit this simple form of linearization, one can do better by allowing the linear approximator to depend on the input. A linear operator whose parameters depend on certain static characteristics of the input is referred to as *quasi-linear*. Quasi-linearization can be accomplished only if the form of the signal at the input to the nonlinearity is assumed in advance. The assumption of a *sinusoidal signal* leads to the ordinary describing function; that analysis is appropriately referred to as *harmonic linearization*.

Similarly, if we deal with nonlinear systems having *random inputs* by use of quasi-linear approximators to the nonlinearities, we refer to the procedure as *statistical linearization*. There is a variety of ways in which such an approximation may be formed. The quasi-linear approximator cannot model all the input-output transfer characteristics of the nonlinearity; so one must choose which of these characteristics to use as the basis for definition of the approximator. This choice does not seem as evident in the case of the random input as it did with a sinusoidal input.

If the nonlinearity is of dynamic character, a dynamic linear operator must somehow be defined as the approximator. If the nonlinearity is static, one is strongly tempted to approximate it with a static linear operator—a gain. As reasonable as this sounds, it still represents a somewhat arbitrary choice. One of the most important statistical characteristics of the output of the

nonlinearity with a random input, its power spectral density function (or equivalently, its autocorrelation function), cannot be duplicated at the output of the approximating operator if it is just a gain. If the assumption is made that the random process at the input to the nonlinearity is gaussian, the autocorrelation function of the output can readily be expressed as a power series in the input autocorrelation function as in Eq. (H-77). The Fourier transform of this autocorrelation function is the power density spectrum of the output of the nonlinearity, and the ratio of this function to the input power density spectrum defines the square of the magnitude of a transfer function which can be taken as a dynamic linear approximation to the static nonlinearity. The virtue of this approximation is that it produces a power density spectrum at its output equal to that of the nonlinearity it approximates. Several authors have employed this method of quasi-linearization. Pupkov (Ref. 36) gives in graphical form the first five coefficients, a_k , required for the series expansion of the output autocorrelation function for several common nonlinearities. Culver and Mesarovic (Ref. 10) suggest this procedure as a practical alternative to a still more general, but much more complicated, method of dynamic linearization. Kazakov (Ref. 18) summarizes the conditions which define two dynamic linear operators to approximate a static nonlinearity, one to operate on the mean and the other on the random component of the input, so as to match at their outputs the mean value and autocorrelation function of the output of the nonlinearity. He states, however, that investigations by a number of authors have not conclusively shown that the performance of closed-loop systems which include a static nonlinearity is better predicted by the use of a dynamic, as opposed to a static, linear approximation to the nonlinearity. Caron (Ref. 8) has taken a different approach to the end of matching the autocorrelation functions at the outputs of the nonlinearity and its approximator. He employs as his quasi-linear model of the nonlinearity a static gain chosen to equate the input-output cross-correlation functions for model and nonlinearity, to the output of which he adds a random signal calculated to equate the autocorrelation functions. In the application of this procedure, however, he predicts signals that have negative values for their power spectral density functions over portions of the frequency range. This physically meaningless result remains unexplained.

Actually, the distinction between a static and dynamic quasi-linear approximator for a static nonlinearity is probably of lesser consequence to the accuracy of the representation of a nonlinear feedback system than the assumption which is common to both—that the random process at the input to the nonlinearity is gaussian. This assumption seems essential to permit a practicable attack on all but the simplest of problems. The procedure required to calculate, rather than assume, the distribution of amplitudes at the input to the nonlinearity is described and illustrated by

Chuang and Kazda (Ref. 9). For systems of higher order than second, their approach can be pursued to its conclusion only in special cases, and in these cases the procedure is complicated enough to defy the interpretation and understanding of results which are essential to the system designer. We are left, then, with the necessity of assuming the form of the distribution function for the nonlinearity input; clearly, the only cases in which this can be done are those in which the input to the system is gaussian, and the nature of the feedback system is such that the input to the nonlinearity can also be assumed gaussian. The situations in which this assumption seems reasonable are discussed in Sec. 7.3; we simply note here that the same distortion products (nonlinear terms) in the expansion of the autocorrelation function for the output of a nonlinearity with a gaussian input which account for the difference in the forms of the power spectral density functions at input and output also change the form of the distribution function. If the linear part of the system, in returning this nonlinearity output back to its input, fails to filter these higher-frequency distortion products adequately, a dynamic representation of the static nonlinearity would be in order, but *not* one which depends on the assumption of a gaussian input to the nonlinearity.

If one agreed on the use of a static gain to approximate a static nonlinearity, there would still remain several reasonable choices for the criterion under which this gain is to be chosen. One line of approach was initiated by Booton (Refs. 5, 6), who approximated a nonlinearity with a random input by a gain chosen to minimize the mean-squared difference between the outputs of the nonlinearity and its approximator. This concept was extended by Somerville and Atherton (Ref. 42) to a multiple-gain representation of a nonlinearity, with an input having components of several characteristic forms, such as a random process and a sinusoid. Each component of the input is transmitted through a separate gain, and these gains are chosen to minimize the mean-squared error in the total approximation. It has been demonstrated in Sec. 1.5 that minimization of the mean-squared error in the approximation is achieved by the gain which makes the covariance between the component of input that the gain multiplies and the output of the approximating gain equal to the covariance between that component of input and the output of the nonlinearity. Merhav (Ref. 26) suggests the use of covariance equivalence as a more satisfying definition for equivalent linearization, but the results of this definition are identical with those based on minimization of mean-squared approximation error. Another approach is to equate the mean-squared values at the outputs of the nonlinearity and its approximator. Barrett and Coales (Ref. 4) described this procedure, which they attributed to Burt. Later Axelby (Ref. 3) represented a nonlinearity with an input bias signal and random process by two gains chosen to reproduce the mean and mean-squared values of the output of the nonlinearity. Kazakov (Ref. 17) discussed both this method and Booton's at

an early date. A third doctrine, espoused by Sawaragi and Takahashi (Ref. 39), and later by Sridhar and Oldenburger (Ref. 43), requires the calculation of the autocorrelation function of the output of the nonlinearity for the assumed input. This is done in series form in which there appear terms that are linearly related to the input autocorrelation function, higher-ordered distortion terms, and cross-modulation terms, if more than one input component is present. The approximating gains are selected to reproduce at their outputs the linear terms in the expansion of the nonlinearity output autocorrelation function.

Smith (Refs. 40, 41) attacked the problem of nonlinear feedback systems with random inputs, using the method of functional analysis. With this analytic tool and extensive experimentation, he compared a number of quasi-linear methods for the approximation of nonlinearities. He concluded that if the assumption of a gaussian input to the nonlinearity holds to a good approximation, all these methods give very good results. If one of the methods fails, they all do, since the most important reason for the failure is the failure of the gaussian assumption which is common to all. In these cases for the systems which Smith studied, the method of Booton or its extension by Somerville and Atherton was somewhat to be preferred. Since, in addition, it is preferable on the basis of the computation required, the method of quasi-linearization based on minimization of the mean-squared approximation error, or the method of covariance equivalence, has been pursued in this book.

Having decided upon this criterion, we take satisfaction in noting that the question of a static vs. dynamic quasi-linear approximator need not be considered a separate issue. The most general linear approximator was taken as the starting point in Chap. 1: a parallel set of dynamic linear operators, one to operate on each of the distinguishable components of input. The necessary and sufficient conditions which these operators must satisfy to minimize the mean-squared approximation error were then derived, and those cases in which the optimum transfer is just a static gain appeared as a consequence of the theory. Booton (Ref. 5) had noted the important fact that for an unbiased gaussian input to a static single-valued nonlinearity, the optimum (in the least-mean-squared-error sense) quasi-linear approximator is a static gain. Later Nuttall (Ref. 32) investigated this property more generally.

It should be observed that entirely different avenues of approach can be taken to the study of certain nonlinear systems with random inputs, approaches which do not employ a quasi-linear approximation for the nonlinearity at all. For relay control systems it is possible to write the switching conditions for the relay in terms of the time response of the linear part of the system, and the random input. From this the second-order statistics of the half-periods of the oscillation and of any signal in the system

can be calculated. This is a direct extension of the methods of Tsytkin and others, which were summarized in Sec. 3.8. This procedure, using different analytic formulations, has been discussed by Katkovnik and Pervozvanskii (Ref. 16), Morosanov (Ref. 28), and Tsytkin (Ref. 45). It does not fall within the theme of this book, however, and will not be considered further.

7.2 CALCULATION OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs)

We now address our attention to the calculation of random-input describing functions (RIDFs) for specific nonlinearities with rather general inputs. A nonlinear device with an input consisting of the sum of components of several characteristic forms presents a complicated situation. A quasi-linear approximation to the nonlinearity in this case results in a set of approximating gains, each of which is dependent upon all the input components. One might expect the calculation of these gains to be a rather formidable task; in this the reader will not be disappointed. However, the degree to which this theory proves useful to the practicing engineer depends directly upon his facility to calculate, and then to manipulate, the required RIDFs. This section is intended to be of assistance in the former problem.

It was determined in Chap. I that the most general nonlinearity input which need be considered is the sum of a single bias component, a single gaussian component, and an arbitrary number of sinusoids. For the purpose of this section and Appendix E, we shall take the input to be the sum of a bias, a gaussian signal, and a single sinusoid. For those cases in which it is necessary to include more than one sinusoid, the extension of these results is evident, but of course the complexity of the problem increases. In this case, then, the nonlinearity input is

$$x(t) = B + r(t) + A \sin(\omega t + \theta) \quad (7.2-1)$$

B , A , and ω are deterministic parameters; $r(0)$ is a random variable having a normal distribution; and θ is a random variable having a uniform distribution over an interval of 2π radians. The expectations in the definitions of the approximating gains are thus taken over the distributions of $r(0)$, which we shall just write as r , and θ . The joint probability density function for these random variables is

$$f_2(r, \theta) = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (7.2-2)$$

over the intervals $-\infty < r < \infty$ and $0 \leq \theta \leq 2\pi$. In this expression, σ is the standard deviation of the unbiased stationary gaussian random process $r(t)$.

In this section, attention is centered on static single-valued nonlinearities for which the output $y(t)$ can be written unambiguously as $y[x(t)]$. The gain to the bias component of the input is, from Eq. (1.5-27),

$$\begin{aligned} N_B &= \frac{1}{B} \overline{y[x(0)]} \\ &= \frac{1}{(2\pi)^{\frac{1}{2}} \sigma B} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr y(B + r + A \sin \theta) \exp\left(-\frac{r^2}{2\sigma^2}\right) \end{aligned} \quad (7.2-3)$$

The gain to the random component of the input, according to Eq. (1.5-51), is

$$\begin{aligned} N_R &= \frac{1}{\sigma^2} \overline{y[x(0)]r(0)} \\ &= \frac{1}{(2\pi)^{\frac{1}{2}} \sigma^3} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr y(B + r + A \sin \theta) r \exp\left(-\frac{r^2}{2\sigma^2}\right) \end{aligned} \quad (7.2-4)$$

The gain to the sinusoidal component of the input is given by Eq. (1.5-36) to be

$$\begin{aligned} N_A &= \frac{2}{A} \overline{y[x(0)] \sin \theta} \\ &= \frac{2}{(2\pi)^{\frac{1}{2}} \sigma A} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr y(B + r + A \sin \theta) \sin \theta \exp\left(-\frac{r^2}{2\sigma^2}\right) \end{aligned} \quad (7.2-5)$$

We shall first document several useful nonlinearities for which analytic RIDFs can be derived. This is followed by a numerical integration procedure simple enough for use in hand calculation. In addition to this, it is clear that a standard digital-computer program for RIDF calculation, together with certain function subroutines, will be a most valuable asset to one who is to make regular use of RIDF theory. This matter is also discussed.

POLYNOMIAL NONLINEARITIES

Nonlinear functions of the form

$$y(x) = \sum_n c_n x^n \quad (7.2-6)$$

can be integrated to analytic expressions for the RIDFs. It is possible to do the r and θ integrations in either order, but the algebra is a little simpler if the r integration is done first. This integration is expedited by writing the nonlinear function in terms of the Hermite polynomials, to take advantage of their orthogonality. As given in Appendix H, the first few of these

polynomials are

$$H_0(u) = 1$$

$$H_1(u) = u$$

$$H_2(u) = u^2 - 1$$

$$H_3(u) = u^3 - 3u$$

and in general,

$$H_{k+1}(u) = uH_k(u) - kH_{k-1}(u)$$

The orthogonality conditions are

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H_k(u) H_j(u) \exp\left(-\frac{u^2}{2}\right) du = \begin{cases} k! & j = k \\ 0 & j \neq k \end{cases} \quad (7.2-7)$$

If the integrals involved in the expressions for N_B , N_R , and N_A are to be put in the form of Eq. (7.2-7), the variable r must be normalized with respect to the standard deviation σ .

$$v = \frac{r}{\sigma}$$

Thus the first step is to rewrite the original nonlinear function in terms of Hermite polynomials, with r replaced by σv .

$$\begin{aligned} y(B + r + A \sin \theta) &= \sum_n c_n (r + B + A \sin \theta)^n \\ &= \sum_n d_n H_n(v) \end{aligned} \quad (7.2-8)$$

where the d_n will be functions of $B + A \sin \theta$ and σ . This step can always be taken by starting with the highest power of r (or v) and working toward the lowest.

Example 7.2-1 Express the cubic nonlinearity in the form of Eq. (7.2-8).

$$y(x) = x^3 \quad (7.2-9)$$

Thus

$$\begin{aligned} y(B + r + A \sin \theta) &= (\sigma v + B + A \sin \theta)^3 \\ &= \sigma^3 v^3 + 3\sigma^2 v^2 (B + A \sin \theta) \\ &\quad + 3\sigma v (B + A \sin \theta)^2 + (B + A \sin \theta)^3 \\ &= \sigma^3 H_3(v) + 3(B + A \sin \theta) \sigma^2 H_2(v) \\ &\quad + 3[(B + A \sin \theta)^2 \sigma + \sigma^3] H_1(v) \\ &\quad + [(B + A \sin \theta)^3 + 3(B + A \sin \theta) \sigma^2] H_0(v) \\ &= \sum_{n=0}^3 d_n H_n(v) \end{aligned} \quad (7.2-10)$$

With the nonlinearity expressed in terms of Hermite polynomials, the r integration—or with the change of variable, the v integration—is trivial.

$$\begin{aligned} N_B &= \frac{1}{(2\pi)^{\frac{3}{2}}B} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dv \left[\sum_n d_n H_n(v) \right] H_0(v) \exp\left(-\frac{v^2}{2}\right) \\ &= \frac{1}{2\pi B} \int_0^{2\pi} d_0(B + A \sin \theta, \sigma) d\theta \end{aligned} \quad (7.2-11)$$

$$\begin{aligned} N_R &= \frac{1}{(2\pi)^{\frac{3}{2}}\sigma} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dv \left[\sum_n d_n H_n(v) \right] H_1(v) \exp\left(-\frac{v^2}{2}\right) \\ &= \frac{1}{2\pi\sigma} \int_0^{2\pi} d_1(B + A \sin \theta, \sigma) d\theta \end{aligned} \quad (7.2-12)$$

$$\begin{aligned} N_A &= \frac{2}{(2\pi)^{\frac{3}{2}}A} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dv \left[\sum_n d_n H_n(v) \right] H_0(v) \sin \theta \exp\left(-\frac{v^2}{2}\right) \\ &= \frac{1}{\pi A} \int_0^{2\pi} d_0(B + A \sin \theta, \sigma) \sin \theta d\theta \end{aligned} \quad (7.2-13)$$

The remaining θ integration is an ordinary harmonic analysis of the functions d_0 and d_1 , which, when expanded, are simply polynomial expressions in $A \sin \theta$. The required integrals have already been considered in Chap. 6, and are rewritten here in the form

$$\begin{aligned} \int_0^{2\pi} (\sin \theta)^n d\theta &= 0 & n \text{ odd} \\ &= 2\sqrt{\pi} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)} & n \text{ even and } \geq 0 \quad (7.2-14) \\ &= 2\pi \frac{(1)(3)(5) \cdots (n-1)}{(2)(4)(6) \cdots (n)} & n \text{ even and } \geq 2 \end{aligned}$$

Use of this integral permits the RIDFs to be calculated in closed analytic form.

Example 7.2-1 (continued) Complete the calculation of the RIDFs for the cubic nonlinearity. We have

$$\begin{aligned} d_0 &= (B + A \sin \theta)^3 + 3(B + A \sin \theta)\sigma^2 \\ &= A^3 \sin^3 \theta + 3A^2B \sin^2 \theta + 3A(B^2 + \sigma^2) \sin \theta + B^3 + 3B\sigma^2 \end{aligned}$$

and
$$d_1 = 3[(B + A \sin \theta)^2 \sigma + \sigma^3]$$

$$= 3\sigma(A^2 \sin^2 \theta + 2AB \sin \theta + B^2 + \sigma^2)$$

Thus
$$N_B = \frac{1}{2\pi B} \int_0^{2\pi} d_0 d\theta$$

$$= \frac{1}{B} \left(\frac{3}{2} A^2 B + B^3 + 3B\sigma^2 \right)$$

$$= B^2 + 3\sigma^2 + \frac{3}{2} A^2 \quad (7.2-15)$$

$$N_R = \frac{1}{2\pi \sigma} \int_0^{2\pi} d_1 d\theta$$

$$= \frac{3\sigma}{\sigma} \left(\frac{1}{2} A^2 + B^2 + \sigma^2 \right)$$

$$= 3B^2 + 3\sigma^2 + \frac{3}{2} A^2 \quad (7.2-16)$$

and
$$N_A = \frac{1}{\pi A} \int_0^{2\pi} d_0 \sin \theta d\theta$$

$$= \frac{2}{A} \left[\frac{3}{8} A^3 + \frac{3}{2} A(B^2 + \sigma^2) \right]$$

$$= 3B^2 + 3\sigma^2 + \frac{3}{4} A^2 \quad (7.2-17)$$

It may be noted that the expressions for N_B and N_A reduce to the corresponding DIDFs of Eqs. (6.2-27) and (6.2-26) in the case of $\sigma = 0$.

No other type of nonlinearity yields RIDFs in so simple a form as this.

A modified form of polynomial nonlinearity is often of interest, a function which includes both odd and even powers of x but which is defined as an odd function.

$$y(x) = \sum_n c_n x^n \quad x \geq 0$$

$$y(-x) = -y(x) \quad (7.2-18)$$

The terms which are odd powers of x have the same analytic definition over the full range of x , and can be integrated by the procedure given above. However, the terms which are even powers of x have a different analytic definition for $x < 0$ than for $x \geq 0$ and require a different treatment. The general term of order n , with n even, would be written

$$y(x) = \begin{cases} x^n & x \geq 0 \\ -x^n & x < 0 \end{cases} \quad (7.2-19)$$

The r integrations in the expressions for N_B and N_A are then

$$\begin{aligned}
 & \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y(B + r + A \sin \theta) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y(\sigma v + B + A \sin \theta) \exp\left(-\frac{v^2}{2}\right) dv \\
 &= \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{B + A \sin \theta}{\sigma}\right) (\sigma v + B + A \sin \theta)^n \exp\left(-\frac{v^2}{2}\right) dv \\
 &\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{B + A \sin \theta}{\sigma}\right) (\sigma v + B + A \sin \theta)^n \exp\left(-\frac{v^2}{2}\right) dv \quad (7.2-20)
 \end{aligned}$$

Expansion of the integrand yields both odd and even powers of v . The required integrals are then of the form

$$\frac{1}{\sqrt{2\pi}} \int v^n \exp\left(-\frac{v^2}{2}\right) dv = -\frac{1}{\sqrt{2\pi}} \int v^{n-1} d\left[\exp\left(-\frac{v^2}{2}\right)\right] \quad (7.2-21)$$

Repeated integration by parts gives the result

$$\begin{aligned}
 & \frac{1}{\sqrt{2\pi}} \int v^n \exp\left(-\frac{v^2}{2}\right) dv \\
 &= -\frac{1}{\sqrt{2\pi}} \left[v^{n-1} \exp\left(-\frac{v^2}{2}\right) + (n-1)v^{n-3} \exp\left(-\frac{v^2}{2}\right) \right. \\
 &\quad \left. + \cdots + (n-1)(n-3) \cdots (2) \exp\left(-\frac{v^2}{2}\right) \right] \quad n \text{ odd and } \geq 3 \\
 &= -\frac{1}{\sqrt{2\pi}} \left[v^{n-1} \exp\left(-\frac{v^2}{2}\right) + (n-1)v^{n-3} \exp\left(-\frac{v^2}{2}\right) \right. \\
 &\quad \left. + \cdots + (n-1)(n-3) \cdots (1)v \exp\left(-\frac{v^2}{2}\right) \right] \\
 &\quad + (n-1)(n-3) \cdots (1) \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{v^2}{2}\right) dv \quad n \text{ even and } \geq 2
 \end{aligned} \quad (7.2-22)$$

This last integration in the expression for even n cannot be written in closed form. It is tabulated as a function of the upper limit of integration for a lower limit of $-\infty$, as in Burington (Ref. 7). We shall be dealing repeatedly with this function, and for convenience shall designate it the *probability integral*.

$$\text{PI}(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^v \exp\left(-\frac{u^2}{2}\right) du \quad (7.2-23)$$

The integrand of the probability integral is the normal probability function which also appears throughout this work. We shall refer to it simply as the *probability function*.

$$\text{PF}(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) \quad (7.2-24)$$

The integral above is conveniently written in terms of these functions.

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} \int v^n \exp\left(-\frac{v^2}{2}\right) dv \\ &= -[v^{n-1} + (n-1)v^{n-3} + \cdots + (n-1)(n-3)\cdots(2)]\text{PF}(v) && n \text{ odd and } \geq 3 \\ &= -\text{PF}(v) && n = 1 \\ &= -[v^{n-1} + (n-1)v^{n-3} + \cdots + (n-1)(n-3)\cdots(1)v]\text{PF}(v) \\ &\quad + (n-1)(n-3)\cdots(1)\text{PI}(v) && n \text{ even and } \geq 2 \\ &= \text{PI}(v) && n = 0 \end{aligned} \quad (7.2-25)$$

This integral form suffices for the r integration in the expressions for N_B and N_A , as given in Eq. (7.2-20), and also for the r integration in the expression for N_R , which is of similar form. However, the θ integration cannot now be done analytically.

This solution is adequate for cases in which there is no sinusoidal component in the nonlinearity input, and thus no θ integration in the expressions for N_B and N_R . In that case,

$$x(t) = B + r(t) \quad (7.2-26)$$

$$N_B = \frac{1}{\sqrt{2\pi\sigma}B} \int_{-\infty}^{\infty} y(B+r) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \quad (7.2-27)$$

$$N_R = \frac{1}{\sqrt{2\pi\sigma^3}} \int_{-\infty}^{\infty} y(B+r)r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \quad (7.2-28)$$

Example 7.2-2 The ideal-relay characteristic

$$y(x) = \begin{cases} D & x > 0 \\ 0 & x = 0 \\ -D & x < 0 \end{cases} \quad (7.2-29)$$

has in this case the RIDFs

$$\begin{aligned}
 N_B &= \frac{-D}{\sqrt{2\pi}B} \int_{-\infty}^{-B/\sigma} \exp\left(-\frac{v^2}{2}\right) dv + \frac{D}{\sqrt{2\pi}B} \int_{-B/\sigma}^{\infty} \exp\left(-\frac{v^2}{2}\right) dv \\
 &= \frac{D}{B} \left[-\text{PI}\left(-\frac{B}{\sigma}\right) + 1 - \text{PI}\left(-\frac{B}{\sigma}\right) \right] \\
 &= \frac{D}{B} \left[2\text{PI}\left(\frac{B}{\sigma}\right) - 1 \right]
 \end{aligned} \tag{7.2-30}$$

and

$$\begin{aligned}
 N_R &= \frac{-D}{\sqrt{2\pi}\sigma} \int_{-\infty}^{-B/\sigma} v \exp\left(-\frac{v^2}{2}\right) dv + \frac{D}{\sqrt{2\pi}\sigma} \int_{-B/\sigma}^{\infty} v \exp\left(-\frac{v^2}{2}\right) dv \\
 &= \frac{D}{\sigma} \left[\text{PF}\left(-\frac{B}{\sigma}\right) + \text{PF}\left(-\frac{B}{\sigma}\right) \right] \\
 &= 2 \frac{D}{\sigma} \text{PF}\left(\frac{B}{\sigma}\right)
 \end{aligned} \tag{7.2-31}$$

In the preceding example, certain evident properties of the probability function and probability integral were used:

$$\text{PF}(\infty) = \text{PF}(-\infty) = 0 \tag{7.2-32}$$

$$\text{PI}(\infty) = 1 \tag{7.2-33}$$

$$\text{PI}(-\infty) = 0 \tag{7.2-34}$$

$$\text{PF}(-v) = \text{PF}(v) \tag{7.2-35}$$

$$\text{PI}(-v) = 1 - \text{PI}(v) \tag{7.2-36}$$

If the RIDFs are required for nonlinearities of this modified polynomial form with a sinusoidal input component as well, the θ integration requires integrating terms of the form $\text{PF}[(B + A \sin \theta)/\sigma]$ and $\text{PI}[(B + A \sin \theta)/\sigma]$. This must be done numerically. We shall return to the question of numerical integration for RIDFs later in this section.

HARMONIC NONLINEARITIES

Nonlinear functions of the form

$$y(x) = \sum_n c_n \sin m_n x \tag{7.2-37}$$

permit analytic evaluation of the RIDFs. The representation of the nonlinear function in terms of an integral transform is especially convenient in this instance. This procedure was employed in Chap. 5, both for analytic

arguments and for TSIDF evaluation; we shall employ the same transform here.

$$y(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(ju) \exp(jux) du \quad (7.2-38)$$

where
$$Y(ju) = \int_{-\infty}^{\infty} y(x) \exp(-jux) dx \quad (7.2-39)$$

in those cases for which the transform exists. Writing $y(x)$ in terms of its transform in Eqs. (7.2-3) to (7.2-5) gives, for the case of a bias plus gaussian plus single-sinusoid input,

$$\begin{aligned} N_B &= \frac{1}{(2\pi)^{\frac{1}{2}} \sigma B} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} du Y(ju) \exp \left[-\frac{r^2}{2\sigma^2} + ju(B+r+A \sin \theta) \right] \\ &= \frac{1}{(2\pi)^{\frac{1}{2}} B} \int_{-\infty}^{\infty} du Y(ju) \exp(juB) \int_0^{2\pi} d\theta \exp(juA \sin \theta) \int_{-\infty}^{\infty} dv \\ &\quad \times \exp \left[-\frac{1}{2}(v^2 - 2ju\sigma v) \right] \end{aligned} \quad (7.2-40)$$

Completing the square in the exponent for the v integral gives

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dv \exp \left[-\frac{1}{2}(v^2 - 2ju\sigma v) \right] = \exp \left(-\frac{\sigma^2 u^2}{2} \right) \quad (7.2-41)$$

From the definition of the Bessel function,

$$J_n(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp[j(x \sin \theta - n\theta)] d\theta \quad (7.2-42)$$

the θ integral is recognized as $J_0(Au)$. Putting these results in Eq. (7.2-40), we have

$$N_B = \frac{1}{2\pi B} \int_{-\infty}^{\infty} Y(ju) J_0(Au) \exp \left(-\frac{\sigma^2 u^2}{2} + juB \right) du \quad (7.2-43)$$

This form of the result is quite convenient. Alternatively, one can complete the square in the exponent to get

$$N_B = \frac{1}{\sqrt{2\pi B}} \text{PF} \left(\frac{B}{\sigma} \right) \int_{-\infty}^{\infty} Y(ju) J_0(Au) \exp \left[-\frac{1}{2}(\sigma u - jB/\sigma)^2 \right] du \quad (7.2-44)$$

Using similar manipulations, the gain to the gaussian signal is written

$$\begin{aligned} N_R &= \frac{1}{(2\pi)^{\frac{1}{2}} \sigma^3} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} du Y(ju) r \exp \left[-\frac{r^2}{2\sigma^2} + ju(B+r+A \sin \theta) \right] \\ &= \frac{1}{(2\pi)^{\frac{1}{2}} \sigma} \int_{-\infty}^{\infty} du Y(ju) \exp(juB) \int_0^{2\pi} d\theta \exp(juA \sin \theta) \int_{-\infty}^{\infty} dv v \\ &\quad \times \exp \left[-\frac{1}{2}(v^2 - 2ju\sigma v) \right] \\ &= \frac{j}{2\pi} \int_{-\infty}^{\infty} Y(ju) J_0(Au) u \exp \left(-\frac{\sigma^2 u^2}{2} + juB \right) du \end{aligned} \quad (7.2-45)$$

Alternatively,

$$N_R = \frac{j}{\sqrt{2\pi}} \text{PF}\left(\frac{B}{\sigma}\right) \int_{-\infty}^{\infty} Y(ju) J_0(Au) u \exp \left[-\frac{1}{2} \left(\sigma u - j \frac{B}{\sigma} \right)^2 \right] du \quad (7.2-46)$$

Finally,

$$\begin{aligned} N_A &= \frac{2}{(2\pi)^{\frac{1}{2}} \sigma A} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} du Y(ju) \sin \theta \\ &\quad \times \exp \left[-\frac{r^2}{2\sigma^2} + ju(B + r + A \sin \theta) \right] \\ &= \frac{2}{(2\pi)^{\frac{1}{2}} A} \int_{-\infty}^{\infty} du Y(ju) \exp(juB) \int_0^{2\pi} d\theta \sin \theta \exp(juA \sin \theta) \int_{-\infty}^{\infty} dv \\ &\quad \times \exp \left[-\frac{1}{2}(v^2 - 2ju\sigma v) \right] \\ &= \frac{j}{\pi A} \int_{-\infty}^{\infty} Y(ju) J_1(Au) \exp \left(-\frac{\sigma^2 u^2}{2} + juB \right) du \end{aligned} \quad (7.2-47)$$

or

$$N_A = \sqrt{\frac{2}{\pi}} \frac{j}{A} \text{PF}\left(\frac{B}{\sigma}\right) \int_{-\infty}^{\infty} Y(ju) J_1(Au) \exp \left[-\frac{1}{2} \left(\sigma u - j \frac{B}{\sigma} \right)^2 \right] du \quad (7.2-48)$$

In the θ integration for N_A , $\sin \theta$ is written in terms of $\exp(j\theta)$ and $\exp(-j\theta)$. Integration then gives a term $J_{-1}(Au)$ and a term $J_1(Au)$, which can be collected, using the property $J_{-n}(x) = -J_n(x)$ for n odd.

In the case of each of these gains, the original double-integral expression involving the nonlinear function has been reduced to a single integral involving the transform of the nonlinear function. For some nonlinearities $Y(ju)$ is easier to integrate than $y(x)$.

Example 7.2-3 Calculate the RIDFs for the harmonic nonlinearity. The general term in the sum of Eq. (7.2-37) is

$$y(x) = \sin mx \quad (7.2-49)$$

The transform of this is

$$\begin{aligned} Y(ju) &= \int_{-\infty}^{\infty} \sin mx \exp(-jux) dx \\ &= \frac{1}{2j} \int_{-\infty}^{\infty} [\exp[j(m-u)x] - \exp[-j(m+u)x]] dx \\ &= \frac{\pi}{j} [\delta(u-m) - \delta(u+m)] \end{aligned} \quad (7.2-50)$$

This function can readily be integrated according to Eqs. (7.2-43), (7.2-45), and (7.2-47) to give analytic expressions for the RIDFs for the general harmonic-term nonlinear function.

$$\begin{aligned}
 N_B &= \frac{1}{2\pi B} \int_{-\infty}^{\infty} \frac{\pi}{j} [\delta(u-m) - \delta(u+m)] J_0(Au) \exp\left(-\frac{\sigma^2 u^2}{2} + jBu\right) du \\
 &= \frac{1}{2Bj} \left[J_0(Am) \exp\left(-\frac{\sigma^2 m^2}{2} + jBm\right) - J_0(-Am) \exp\left(-\frac{\sigma^2 m^2}{2} - jBm\right) \right] \\
 &= \frac{1}{B} J_0(Am) \exp\left(-\frac{m^2 \sigma^2}{2}\right) \frac{1}{2j} [\exp(jBm) - \exp(-jBm)] \\
 &= \frac{1}{B} \sin mB \exp\left(-\frac{m^2 \sigma^2}{2}\right) J_0(mA) \quad (7.2-51)
 \end{aligned}$$

$$\begin{aligned}
 N_R &= \frac{j}{2\pi} \int_{-\infty}^{\infty} \frac{\pi}{j} [\delta(u-m) - \delta(u+m)] J_0(Au) u \exp\left(-\frac{\sigma^2 u^2}{2} + jBu\right) du \\
 &= \frac{1}{2} \left[J_0(Am) m \exp\left(-\frac{\sigma^2 m^2}{2} + jBm\right) + J_0(-Am) m \exp\left(-\frac{\sigma^2 m^2}{2} - jBm\right) \right] \\
 &= J_0(Am) m \exp\left(-\frac{m^2 \sigma^2}{2}\right) \frac{1}{2} [\exp(jBm) + \exp(-jBm)] \\
 &= m \cos mB \exp\left(-\frac{m^2 \sigma^2}{2}\right) J_0(mA) \quad (7.2-52)
 \end{aligned}$$

$$\begin{aligned}
 N_A &= \frac{j}{\pi A} \int_{-\infty}^{\infty} \frac{\pi}{j} [\delta(u-m) - \delta(u+m)] J_1(Au) \exp\left(-\frac{\sigma^2 u^2}{2} + jBu\right) du \\
 &= \frac{1}{A} \left[J_1(Am) \exp\left(-\frac{\sigma^2 m^2}{2} + jBm\right) - J_1(-Am) \exp\left(-\frac{\sigma^2 m^2}{2} - jBm\right) \right] \\
 &= \frac{1}{A} J_1(Am) \exp\left(-\frac{m^2 \sigma^2}{2}\right) [\exp(jBm) + \exp(-jBm)] \\
 &= \frac{2}{A} \cos mB \exp\left(-\frac{m^2 \sigma^2}{2}\right) J_1(mA) \quad (7.2-53)
 \end{aligned}$$

In this development, use is made of the property of Bessel functions that $J_n(x)$ is an odd function of x if n is odd and is an even function of x if n is even.

These are the RIDFs for the general term of a harmonic nonlinearity. If a nonlinear function is a linear combination of terms of this type, as in Eq. (7.2-37), the RIDFs are the corresponding linear combinations of the terms given in Eqs. (7.2-51) to (7.2-53).

PIECEWISE-LINEAR NONLINEARITIES

A great many nonlinearities of common interest are composed of a sequence of straight-line segments. If we allow the possibility of discontinuous jumps between these segments, this family of nonlinearities assumes even greater proportions. It will be of value to treat this family in as general a way as possible.

To this end consider a general linear segment as pictured in Fig. 7.2-1. Any nonlinear function consisting of straight-line segments with or without jumps between them can be decomposed into segments of this form. If the contribution of this general segment to the integrals defining the RIDFs can be calculated, the RIDFs for any piecewise-linear nonlinearity could be synthesized as the sums of such contributions. Unfortunately, it will not be possible to carry out this program analytically in the case of the three-component input. In the double-integral expressions for the RIDFs [Eqs. (7.2-3) to (7.2-5)], either the r or the θ integration can be carried out for the general linear segment, but in each case this gives rise to functions which do not permit the second integration to be done analytically. This suffices for the case where either the sinusoid or the gaussian process is missing from the input, but if both are present, the RIDF calculation must be completed by numerical integration.

The limits of integration for a nonlinear function of finite extent are much simpler to handle if the r integration is done first. The general linear segment of Fig. 7.2-1 is nonzero over the interval $x_1 < x < x_2$. The r and θ integrations must then be carried out over the region in the r, θ plane, which satisfies simultaneously the inequalities

$$x_1 < B + r + A \sin \theta < x_2$$

and

$$0 < \theta < 2\pi \quad (7.2-54)$$

This region is pictured in Fig. 7.2-2 for the case $A < (x_2 - x_1)/2$. The θ integration, if done first, would be over different intervals of θ for r in different regions, and these intervals are different for larger values of A relative to $x_2 - x_1$. On the other hand, if the r integration is done first, it is always over the interval $x_1 - B - A \sin \theta < r < x_2 - B - A \sin \theta$. The subsequent θ integration is then simply over the interval $0 < \theta < 2\pi$.

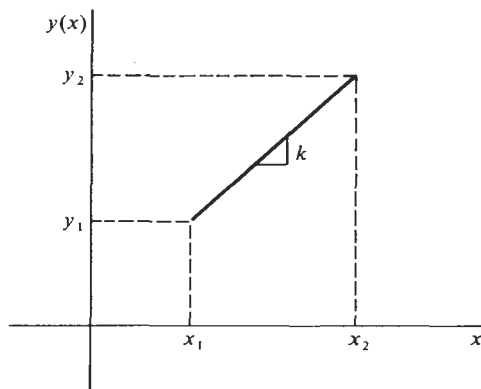


Figure 7.2-1 General segment of a piecewise-linear nonlinearity.

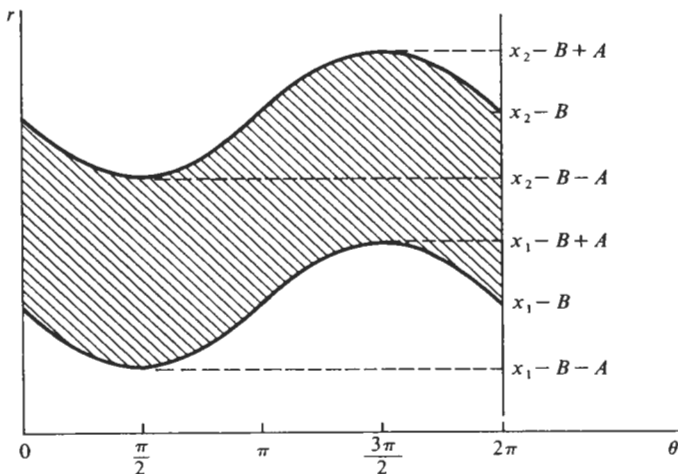


Figure 7.2-2 Region of integration for the general linear segment.

In the expressions for N_B and N_A [Eqs. (7.2-3) and (7.2-5)], the r integration is of the form

$$I_1 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y(B + r + A \sin \theta) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \quad (7.2-55)$$

For the general linear segment of Fig. 7.2-1, this becomes

$$\begin{aligned} I_1 &= \frac{1}{\sqrt{2\pi}} \int_{v_1}^{v_2} [y_1 + k(B + \sigma v + A \sin \theta - x_1)] \exp\left(-\frac{v^2}{2}\right) dv \\ &= [y_1 + k(B + A \sin \theta - x_1)][\text{PI}(v_2) - \text{PI}(v_1)] + k\sigma[\text{PF}(v_1) - \text{PF}(v_2)] \\ &= [y_1 - k\sigma v_1][\text{PI}(v_2) - \text{PI}(v_1)] + k\sigma[\text{PF}(v_1) - \text{PF}(v_2)] \end{aligned} \quad (7.2-56)$$

using the integral form of Eq. (7.2-25). v_1 and v_2 are the end points of the integration, given by

$$v_1 = \frac{1}{\sigma} (x_1 - B - A \sin \theta) \quad (7.2-57)$$

$$v_2 = \frac{1}{\sigma} (x_2 - B - A \sin \theta) \quad (7.2-58)$$

This form of I_1 holds for any linear segment of finite extent, or one which extends to $x \rightarrow +\infty$, in which case $v_2 \rightarrow +\infty$. To be applicable to a

segment which extends to $x \rightarrow -\infty$, Eq. (7.2-56) must be rewritten in terms of y_2 .

$$y_2 = y_1 + k(x_2 - x_1) \quad (7.2-59)$$

$$\begin{aligned} I_1 &= [y_2 - k(x_2 - x_1) + k(B + A \sin \theta - x_1)][PI(v_2) - PI(v_1)] \\ &\quad + k\sigma[PF(v_1) - PF(v_2)] \\ &= (y_2 - k\sigma v_2)[PI(v_2) - PI(v_1)] + k\sigma[PF(v_1) - PF(v_2)] \end{aligned} \quad (7.2-60)$$

This also holds for any segment of finite extent, or one which extends to $x \rightarrow -\infty$, in which case $v_1 \rightarrow -\infty$.

The r integration in the expression for N_R [Eq. (7.2-4)] is of the form

$$\begin{aligned} I_2 &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y(B + r + A \sin \theta)r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\ &= \frac{1}{\sqrt{2\pi}} \int_{v_1}^{v_2} [y_1 + k(B + \sigma v + A \sin \theta - x_1)]v \exp\left(-\frac{v^2}{2}\right) dv \\ &= [y_1 - k\sigma v_1][PF(v_1) - PF(v_2)] + k\sigma[v_1 PF(v_1) - v_2 PF(v_2)] \\ &\quad + k\sigma[PI(v_2) - PI(v_1)] \\ &= y_1[PF(v_1) - PF(v_2)] + k\sigma[PI(v_2) - PI(v_1) - (v_2 - v_1)PF(v_2)] \end{aligned} \quad (7.2-61)$$

In terms of y_2 this is rewritten

$$I_2 = y_2[PF(v_1) - PF(v_2)] + k\sigma[PI(v_2) - PI(v_1) - (v_2 - v_1)PF(v_1)] \quad (7.2-62)$$

When evaluating Eq. (7.2-61) for a segment which extends to plus infinity, it should be noted that

$$\lim_{v_2 \rightarrow \infty} (v_2 - v_1)PF(v_2) = 0 \quad (7.2-63)$$

$$\text{Similarly,} \quad \lim_{v_1 \rightarrow -\infty} (v_2 - v_1)PF(v_1) = 0 \quad (7.2-64)$$

These integrals, I_1 and I_2 , are functions of the parameters which characterize the linear segment, and of σ , B , and $A \sin \theta$. Indicating in the notation only the dependence on $A \sin \theta$, the RIDFs for the segment are now written as

$$N_B = \frac{1}{2\pi B} \int_0^{2\pi} I_1(A \sin \theta) d\theta \quad (7.2-65)$$

$$N_R = \frac{1}{2\pi\sigma} \int_0^{2\pi} I_2(A \sin \theta) d\theta \quad (7.2-66)$$

$$N_A = \frac{1}{\pi A} \int_0^{2\pi} I_1(A \sin \theta) \sin \theta d\theta \quad (7.2-67)$$

which are simply in the form of a harmonic analysis of the functions I_1 and I_2 . As in the case of modified polynomial nonlinearities, the integrals of $\text{PF}(x)$ and $\text{PI}(x)$ are required. For a piecewise-linear nonlinearity which consists of the sum of a number of these segments, the individual contributions to the RIDFs can be calculated according to Eqs. (7.2-65) to (7.2-67) and then added, or more efficiently, the individual I_1 and I_2 contributions can be added first and then integrated at once with respect to θ .

If the nonlinearity input is taken to include only a bias and a gaussian signal, the integrals I_1 and I_2 constitute a complete solution for the RIDFs. In that case

$$x(t) = B + r(t) \quad (7.2-68)$$

$$\begin{aligned} N_B &= \frac{1}{\sqrt{2\pi}\sigma B} \int_{-\infty}^{\infty} y(B+r) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\ &= \frac{1}{B} I_1(A=0) \end{aligned} \quad (7.2-69)$$

$$\begin{aligned} N_R &= \frac{1}{\sqrt{2\pi}\sigma^3} \int_{-\infty}^{\infty} y(B+r)r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\ &= \frac{1}{\sigma} I_2(A=0) \end{aligned} \quad (7.2-70)$$

These contributions to the RIDFs corresponding to each linear segment can very easily be added to give a closed-form expression for the approximating gains for any piecewise-linear nonlinearity.

Example 7.2-4 Calculate the RIDFs for the relay with dead-zone nonlinearity (see Fig. 7.2-3) for the case of a bias plus gaussian input ($A=0$).

$$\text{Segment 1:} \quad \begin{cases} x_1 = -\infty & v_1 = -\infty \\ x_2 = -\delta & v_2 = \frac{-\delta - B}{\sigma} \\ y_2 = -D & k = 0 \end{cases}$$

From Eqs. (7.2-60) and (7.2-62) it follows that

$$\begin{aligned} I_1(A=0) &= -D \left[\text{PI}\left(\frac{-\delta - B}{\sigma}\right) - \text{PI}(-\infty) \right] \\ &= -D \left[1 - \text{PI}\left(\frac{\delta + B}{\sigma}\right) \right] \end{aligned}$$

and

$$\begin{aligned} I_2(A=0) &= -D \left[\text{PF}(-\infty) - \text{PF}\left(\frac{-\delta - B}{\sigma}\right) \right] \\ &= D \text{PF}\left(\frac{\delta + B}{\sigma}\right) \end{aligned}$$

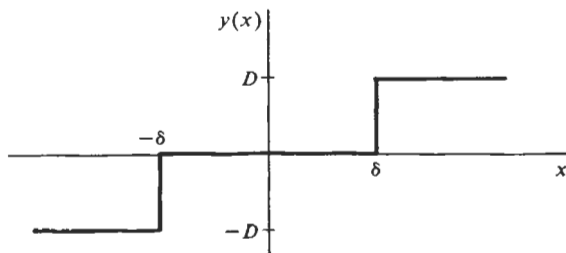


Figure 7.2-3 Nonlinear characteristic for Example 7.2-4.

Segment 2:

$$\begin{cases} x_1 = \delta & v_1 = \frac{\delta - B}{\sigma} \\ x_2 = \infty & v_2 = \infty \\ y_1 = D & k = 0 \end{cases}$$

Similarly, Eqs. (7.2-56) and (7.2-61) yield

$$\begin{aligned} I_1(A = 0) &= D \left[\text{PI}(\infty) - \text{PI}\left(\frac{\delta - B}{\sigma}\right) \right] \\ &= D \left[1 - \text{PI}\left(\frac{\delta - B}{\sigma}\right) \right] \end{aligned}$$

and

$$\begin{aligned} I_2(A = 0) &= D \left[\text{PF}\left(\frac{\delta - B}{\sigma}\right) - \text{PF}(\infty) \right] \\ &= D \text{PF}\left(\frac{\delta - B}{\sigma}\right) \end{aligned}$$

Finally, Eqs. (7.2-69) and (7.2-70) give

$$N_B = \frac{D}{B} \left[\text{PI}\left(\frac{\delta + B}{\sigma}\right) - \text{PI}\left(\frac{\delta - B}{\sigma}\right) \right] \quad (7.2-71)$$

$$N_R = \frac{D}{\sigma} \left[\text{PF}\left(\frac{\delta + B}{\sigma}\right) + \text{PF}\left(\frac{\delta - B}{\sigma}\right) \right] \quad (7.2-72)$$

These expressions reduce to the RIDFs for the ideal relay [Eqs. (7.2-30) and (7.2-31)] in the case $\delta = 0$.

NUMERICAL INTEGRATION FOR HAND CALCULATION

The majority of nonlinear functions do not permit an analytic integration of the expressions defining the RIDFs. The user must then be prepared to calculate RIDFs through numerical, graphical, or other means of integration. One evident procedure for approximate RIDF calculation is to approximate

the given nonlinear function by one for which the RIDFs can more easily be calculated. Either a polynomial or harmonic-function approximation to the nonlinearity will lead to analytic expressions for approximate RIDFs using the results derived above. A straight-line-segment approximation permits one integration to be done analytically.

However, the approximation of one function by others is in itself a substantial task; so we look to other means of RIDF calculation which use the given nonlinear function directly. The most convenient technique for general use is *numerical*. We present, first, an approximate procedure simple enough for hand calculation, and then consider the problem of more exact numerical integration by machine calculation.

The double-integral expressions for the RIDFs in the case of a three-component input [Eqs. (7.2-3) to (7.2-5)] can be integrated in either order. If the r integration is done first, these expressions might be written

$$N_B = \frac{1}{B} \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y(B + r + A \sin \theta) \exp \left(-\frac{r^2}{2\sigma^2} \right) dr \right] d\theta \quad (7.2-73)$$

$$N_R = \frac{1}{\sigma} \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\infty} y(B + r + A \sin \theta) r \exp \left(-\frac{r^2}{2\sigma^2} \right) dr \right] d\theta \quad (7.2-74)$$

$$N_A = \frac{1}{A} \frac{1}{\pi} \int_0^{2\pi} \left[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y(B + r + A \sin \theta) \exp \left(-\frac{r^2}{2\sigma^2} \right) dr \right] \sin \theta d\theta \quad (7.2-75)$$

If the θ integration were done first, these expressions would be written

$$N_B = \frac{1}{B} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_0^{2\pi} y(B + r + A \sin \theta) d\theta \right] \exp \left(-\frac{r^2}{2\sigma^2} \right) dr \quad (7.2-76)$$

$$N_R = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_0^{2\pi} y(B + r + A \sin \theta) d\theta \right] r \exp \left(-\frac{r^2}{2\sigma^2} \right) dr \quad (7.2-77)$$

$$N_A = \frac{1}{A} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \left[\frac{1}{\pi} \int_0^{2\pi} y(B + r + A \sin \theta) \sin \theta d\theta \right] \exp \left(-\frac{r^2}{2\sigma^2} \right) dr \quad (7.2-78)$$

In either case there appear just four types of integrals:

$$I_1(c) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y(c+r) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \quad (7.2-79)$$

$$I_2(c) = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\infty} y(c+r)r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \quad (7.2-80)$$

$$I_3(c) = \frac{1}{2\pi} \int_0^{2\pi} y(c+A \sin \theta) d\theta \quad (7.2-81)$$

$$I_4(c) = \frac{1}{\pi} \int_0^{2\pi} y(c+A \sin \theta) \sin \theta d\theta \quad (7.2-82)$$

In an r integration, $c = B + A \sin \theta$, and in a θ integration, $c = B + r$; but in either case, c is just a constant during that integration. The present functions I_1 and I_2 are consistent with those used before in the case of the general segment of a piecewise-linear nonlinearity.

Any one of these RIDFs is now calculated by first evaluating the inner integral for a number of values of c . The resulting function of c is plotted to permit interpolating for other values of c as required in the evaluation of the outer integral. In the calculation of the second integral, this function of c resulting from the first integration plays the role of a new nonlinear function. Each of these expressions can be interpreted in terms of a "modified nonlinearity" if one desires to do so. For example, in the calculation of N_B and N_R , using Eqs. (7.2-76) and (7.2-77), the inner integral is of the form $I_3(c)$. This function may be considered a modified nonlinearity in which the effect of the sinusoid on the gains of the original nonlinearity to the bias and the gaussian signal has been accounted for. The outer integrals which involve $I_3(B+r)$ in the integrand are exactly the expressions for N_B and N_R for the nonlinear function $I_3(x)$, having an input $x(t) = B + r(t)$. Similarly, in the calculation of N_B and N_A , using Eqs. (7.2-73) and (7.2-75), the outer integrals are exactly the expressions for N_B and N_A for the nonlinear function $I_1(x)$ with the input $x(t) = B + A \sin(\omega t + \theta)$, where $I_1(c)$ resulting from the inner integral may be interpreted as a modified nonlinearity, which accounts for the effect of the random process on the gains of the original nonlinearity to the bias and sinusoid. This modified-nonlinearity concept, which has been referred to by many writers [Somerville and Atherton (Ref. 42) were probably first among them] extends also to cases of additional sinusoids at the input.

We now develop an approximate integration procedure for each of the

required integral forms. For the first [Eq. (7.2-79)] we write

$$\begin{aligned}
 I_1(c) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y(c+r) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\
 &= \int_{-\infty}^{\infty} y(c+\sigma v) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv \\
 &= \int_0^1 y(c+\sigma v) d[\text{PI}(v)] \\
 &\approx \delta[\text{PI}(v)] \sum_i y(c+\sigma v_i)
 \end{aligned} \tag{7.2-83}$$

where the integral is approximated by a finite sum, and equal increments of $\text{PI}(v)$ are assumed. The v_i are the values of v at which the nonlinear function is evaluated. There are a number of reasonable ways to choose the values of v_i to use. One would be to take the mid-range value of v corresponding to each incremental range of $\text{PI}(v)$. Probably better than this, on the average, would be the values of v_i which give the correct answer for some average form of $y(x)$, say, a linear function in each interval. This is the same approach taken in Chap. 2 for the approximate numerical calculation of the DF. This choice of the v_i results in the trapezoidal integration rule, approximating the integral of the actual $y(x)$ by the integral of a function which is equal to $y(x)$ at each end of the integration step and linear between these points. The exact integral over a linear segment of $y(x)$ is given in Eq. (7.2-56). The corresponding contribution to the sum of Eq. (7.2-83) can be equated to this to determine the appropriate value of v_i . The result of this calculation is

$$v_i = \frac{\text{PF}(v_1) - \text{PF}(v_2)}{\text{PI}(v_2) - \text{PI}(v_1)} \tag{7.2-84}$$

The interpretation of this expression is as follows: A value for $\delta[\text{PI}(v)]$ is chosen to give the desired number of intervals. Ten intervals, for example, correspond to $\delta[\text{PI}(v)] = 0.1$. The first interval corresponds to $\text{PI}(v)$ ranging from 0 to 0.1; the second to $\text{PI}(v)$ ranging from 0.1 to 0.2; etc. For each of these intervals, v_1 is the value of v corresponding to $\text{PI}(v)$ at the lower end of the interval, and v_2 is the value of v corresponding to $\text{PI}(v)$ at the upper end of the interval. v_i is then the value of v , intermediate to v_1 and v_2 , at which $y(c+\sigma v_i)$ should be evaluated to yield the correct integral over this interval if $y(x)$ were linear over the interval. $\text{PI}(v)$ ranges monotonically from 0 to 1; so $\delta[\text{PI}(v)] = 1/n$, where n is the number of intervals to be used. Equation (7.2-83) is then written

$$I_1(c) \approx \frac{1}{n} \sum_{i=1}^n y(c+\sigma v_i) \tag{7.2-85}$$

The values of v_i for $n = 10$ and 20 are given in Table 7.2-1.

The second integral [Eq. (7.2-80)] is written

$$\begin{aligned}
 I_2(c) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y(c+r)r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\
 &= \int_{-\infty}^{\infty} y(c+\sigma v) \frac{1}{\sqrt{2\pi}} v \exp\left(-\frac{v^2}{2}\right) dv \\
 &= \int_{-\infty < v < \infty} y(c+\sigma v) d[-PF(v)] \\
 &\approx |\delta[PF(v)]| \sum_i S_i y(c+\sigma v_i) \\
 &\approx \frac{1}{n} \sqrt{\frac{2}{\pi}} \sum_{i=1}^n S_i y(c+\sigma v_i) \quad n \text{ even} \quad (7.2-86)
 \end{aligned}$$

In this case the additional factor S_i is required to indicate the sign of $\delta[-PF(v)]$ in each interval; thus $S_i = -1$ for $v_i < 0$ and $+1$ for $v_i > 0$. $PF(v)$ ranges from 0 to $1/\sqrt{2\pi}$ in the first $n/2$ steps and then back to 0 in the next $n/2$. The values of v_i which yield the correct integral over an interval for which $y(x)$ is linear are given by

$$v_i = \frac{v_2 PF(v_2) - v_1 PF(v_1) - PI(v_2) + PI(v_1)}{PF(v_2) - PF(v_1)} \quad (7.2-87)$$

v_1 and v_2 refer in this case to the left- and right-hand values of v at the ends of each interval of $PF(v)$. The v_i for $n = 10$ and 20 are tabulated in Table 7.2-1.

The third integral [Eq. (7.2-81)] is treated in a similar way. Note that

$$\begin{aligned}
 I_3(c) &= \frac{1}{2\pi} \int_0^{2\pi} y(c+A \sin \theta) d\theta \\
 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} y(c+A \sin \theta) d\theta \quad (7.2-88)
 \end{aligned}$$

for any static single-valued nonlinearity, which is a restriction that has been observed throughout this section. Thus

$$\begin{aligned}
 I_3(c) &\approx \frac{\delta\theta}{\pi} \sum_i y(c+A \sin \theta_i) \\
 &\approx \frac{1}{n} \sum_{i=1}^n y(c+A \sin \theta_i) \quad (7.2-89)
 \end{aligned}$$

The appropriate values of $\sin \theta_i$ for trapezoidal integration accuracy are

$$\sin \theta_i = \frac{\cos \theta_1 - \cos \theta_2}{\theta_2 - \theta_1} \quad (7.2-90)$$

TABLE 7.2-1 ARGUMENT VALUES FOR TRAPEZOIDAL INTEGRATION OF

I_1 integration		I_2 integration			
$n = 10$	$n = 20$	$n = 10$		$n = 20$	
v_i	v_i	S_i	v_i	S_i	v_i
	-2.0627			-1	-2.5455
	-1.4472			-1	-1.9551
	-1.1532			-1	-1.6677
	-0.9361			-1	-1.4502
	-0.7563			-1	-1.2643
-1.7550	-0.5983	-1	-2.2503	-1	-1.0937
-1.0446	-0.4541	-1	-1.5590	-1	-0.9280
-0.6773	-0.3189	-1	-1.1790	-1	-0.7578
-0.3865	-0.1892	-1	-0.8429	-1	-0.5680
-0.1260	-0.0627	-1	-0.4354	-1	-0.3028
0.1260	0.0627	+1	0.4354	+1	0.3028
0.3865	0.1892	+1	0.8429	+1	0.5680
0.6773	0.3189	+1	1.1790	+1	0.7578
1.0446	0.4541	+1	1.5590	+1	0.9280
1.7550	0.5983	+1	2.2503	+1	1.0937
	0.7563			+1	1.2643
	0.9361			+1	1.4502
	1.1532			+1	1.6677
	1.4472			+1	1.9551
	2.0627			+1	2.5455

where θ_1 and θ_2 are the left- and right-hand values of θ at the ends of each interval of θ . The values of $\sin \theta_i$ for $n = 10$ and 20 are tabulated in Table 7.2-1.

The fourth integral [Eq. (7.2-82)] is written

$$\begin{aligned}
 I_4(c) &= \frac{1}{\pi} \int_0^{2\pi} y(c + A \sin \theta) \sin \theta \, d\theta \\
 &= \frac{2}{\pi} \int_{-\pi/2}^{\pi/2} y(c + A \sin \theta) \sin \theta \, d\theta \\
 &= \frac{2}{\pi} \int_{-\pi/2 < \theta < \pi/2} y(c + A \sin \theta) d(-\cos \theta) \\
 &\approx \frac{2}{\pi} |\delta \cos \theta| \sum_i S_i y(c + A \sin \theta_i) \\
 &\approx \frac{1}{n} \frac{4}{\pi} \sum_{i=1}^n S_i y(c + A \sin \theta_i) \quad n \text{ even} \quad (7.2-91)
 \end{aligned}$$

S_i in this case carries the sign of $-\delta \cos \theta$, and thus is -1 for $\sin \theta_i < 0$ and $+1$ for $\sin \theta_i > 0$. $\cos \theta$ ranges from 0 to 1 in the first $n/2$ steps and back

CERTAIN RIDF EXPRESSIONS

I_3 integration		I_4 integration			
$n = 10$	$n = 20$	$n = 10$		$n = 20$	
$\sin \theta_i$	$\sin \theta_i$	S_i	$\sin \theta_i$	S_i	$\sin \theta_i$
	-0.9959			-1	-0.9983
	-0.9714			-1	-0.9882
	-0.9229			-1	-0.9678
	-0.8518			-1	-0.9362
	-0.7596			-1	-0.8924
-0.9836	-0.6488	-1	-0.9933	-1	-0.8344
-0.8874	-0.5220	-1	-0.9520	-1	-0.7590
-0.7042	-0.3823	-1	-0.8634	-1	-0.6600
-0.4521	-0.2332	-1	-0.7095	-1	-0.5239
-0.1558	-0.0784	-1	-0.4088	-1	-0.2936
0.1558	0.0784	+1	0.4088	+1	0.2936
0.4521	0.2332	+1	0.7095	+1	0.5239
0.7042	0.3823	+1	0.8634	+1	0.6600
0.8874	0.5220	+1	0.9520	+1	0.7590
0.9836	0.6488	+1	0.9933	+1	0.8344
	0.7596			+1	0.8924
	0.8518			+1	0.9362
	0.9229			+1	0.9678
	0.9714			+1	0.9882
	0.9959			+1	0.9983

to 0 in the next $n/2$. The values of $\sin \theta_i$ for trapezoidal integration accuracy are

$$\sin \theta_i = \frac{\frac{1}{2}(\sin 2\theta_2 - \sin 2\theta_1) - \theta_2 + \theta_1}{2(\cos \theta_2 - \cos \theta_1)} \quad (7.2-92)$$

θ_1 and θ_2 are the left- and right-hand values of θ at the ends of each interval of $\cos \theta$. The $\sin \theta_i$ are tabulated in Table 7.2-1 for $n = 10$ and 20.

It might be noted that if $y(x)$ were odd, and $c = 0$, which would be true in the case of a single input component—either a gaussian or a sinusoid— I_1 and I_3 would be either zero or irrelevant, and I_2 and I_4 could be calculated by integrating over half the range and multiplying the result by 2. In any case, these numerical integrations can be performed quite rapidly by hand. The form of these approximate integrals is especially convenient in that the values of $y(x)$ which are calculated or read from a graph need not be multiplied by any constants; they are simply added to or subtracted from the total. Having a graph of $y(x)$ and an adding machine, one can read the required values and add them very quickly.

The accuracy of these approximate integration rules has been checked in the case of the cubic nonlinearity, $y(x) = x^3$, which is somewhat typical of smooth nonlinearities, and in the case of the relay with dead zone shown in

TABLE 7.2-2 ACCURACY OF THE TRAPEZOIDAL INTEGRATION RULES

	Cubic nonlinearity		Relay nonlinearity	
	Max error	Av error	Max error	Av error
Integral 1:				
$n = 10$	4.0%	2.5%	7.0%	2.2%
$n = 20$	1.6	1.0	4.5	1.3
Integral 2:				
$n = 10$	6.8	3.4	6.8	1.5
$n = 20$	3.1	1.5	3.9	1.2
Integral 3:				
$n = 10$	0.8	0.4	3.5	0.9
$n = 20$	0.2	0.1	3.9	0.7
Integral 4:				
$n = 10$	1.1	0.4	4.3	0.9
$n = 20$	0.3	0.1	3.6	0.6

Note: For the cubic nonlinearity, the errors are given in percent of the true value. For the relay nonlinearity, some of the true values are zero. The errors are thus given in percent of the maximum value of the integral over the 25 values calculated. These maximum values are

$$I_{1,\max} = 1.000$$

$$I_{2,\max} = 0.781$$

$$I_{3,\max} = 1.000$$

$$I_{4,\max} = 1.246$$

Fig. 7.2-3, which is somewhat characteristic of nonlinearities with a discontinuous jump. Each of the four integrals was evaluated at 25 combinations of values for its two parameters. These parameters c , σ , and A (or c/δ , σ/δ , A/δ , in the case of the relay with dead zone) were taken in all combinations of the values 0.2, 0.5, 1, 2, and 5. The average error over the 25 points and the maximum error is given in Table 7.2-2 for each of these nonlinearities and for $n = 10$ and 20. For these cases, the 10-point integration rule yields better than 7 percent accuracy, and the 20-point rule, better than 5 percent.

RIDF CALCULATION BY DIGITAL COMPUTATION

The RIDFs may be calculated as accurately as desired by digital computation using standard numerical integration techniques. This subject needs no discussion here. We simply note that the double integrations required to evaluate the RIDFs for a three-component input can be done much faster if one of the integrations can be done analytically. This was found to be possible for all piecewise-linear nonlinearities and for the modified polynomial nonlinearities, which include even-power terms but are defined to be

odd functions. This represents a very significant family of nonlinear functions. In each case, if the r integration is done first, the integrand for the θ integration includes terms in $\text{PF}(c + A \sin \theta)$ and $\text{PI}(c + A \sin \theta)$. If one has subroutines available in his library to calculate these functions, the integration can be performed very simply. Since $\text{PF}(x)$ is just an exponential function, a standard exponential subroutine which is available in most computing centers is sufficient. For completeness, however, we suggest an expression which can be used for this purpose. $\text{PI}(x)$ is less generally available, and we develop an expression for this function which is convenient for machine computation. These functions are pictured in Fig. 7.2-4 for the positive range of x ; they are defined over the negative range by the fact that $\text{PF}(x)$ is even and the relation given in Eq. (7.2-36):

$$\text{PI}(-x) = 1 - \text{PI}(x)$$

An approximation for $\text{PF}(x)$ which is very easy to program and reasonably efficient for machine calculation is derived from the power-series expansion of the function about the origin. This expansion suggests the form

$$\begin{aligned} f_F(x) &= \frac{1/\sqrt{2\pi}}{1 + a_1 x^2 + a_2 x^4 + \cdots + a_n x^{2n}} \\ &= \frac{0.39894229}{1 + \{a_1 + [a_2 + \cdots + (a_{n-1} + a_n x^2)x^2 \cdots]x^2\}x^2} \quad (7.2-93) \end{aligned}$$

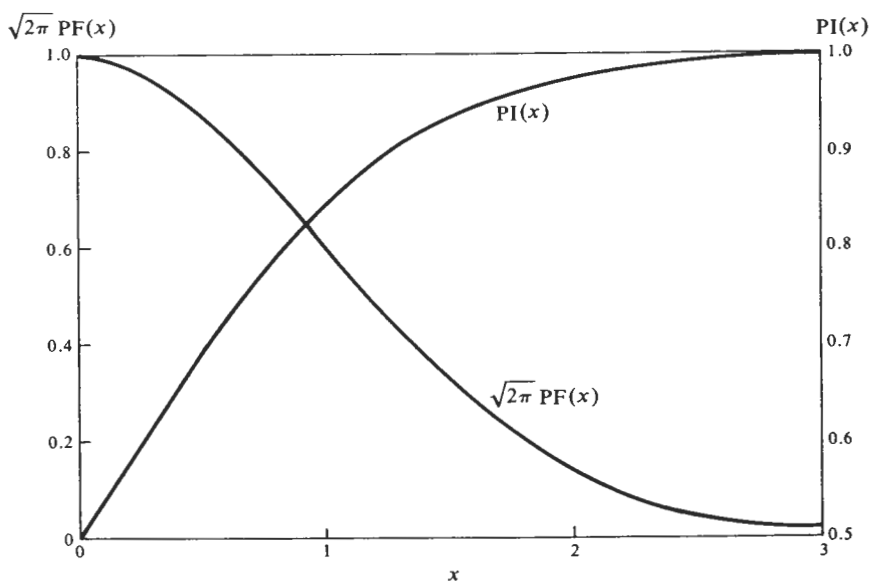


Figure 7.2-4 Normal probability functions.

Using the second, or nested, form, the denominator of this function can be programmed in Fortran by a single DO statement. If this function is to be used primarily for the computation of RIDFs, six-digit accuracy in the approximation is surely adequate, since no more than four-digit accuracy would be of any value in the results of the computation. For $f_F(x)$, as given in Eq. (7.2-93), to approximate PF(x) with a maximum error of $1 \cdot 10^{-6}$ over the full range of x , requires 14 terms in the denominator. This accuracy is not achieved using values of the a_i appropriate to the series expansion of PF(x) about the origin, but with a_i values chosen to minimize the sum of squared errors in the approximation at many points over the range of x . With the coefficients optimized in this way, the maximum approximation error decreases by a factor of about 2, with each additional term in the expansion. A set of 13 coefficients which yield $1 \cdot 10^{-6}$ accuracy is given in Table 7.2-3. Note that this approximating function is valid for positive and negative values of x .

The function PI(x) is surprisingly difficult to approximate well with simple functional forms. For computational efficiency it was found wise to use different approximating functions over two different intervals of x . The following functional forms are suggested:

$$\begin{aligned} f_I(x)_1 &= 0.5 + b_1x + b_2x^3 + b_3x^5 + \cdots + b_nx^{2n-1} \\ &= 0.5 + \{b_1 + [b_2 + \cdots + (b_{n-1} + b_nx^2)x^2 \cdots]x^2\}x \quad |x| \leq x_M \\ f_I(x)_2 &= 1 - \frac{1}{c_1x + c_2x^3 + c_3x^5 + \cdots + c_nx^{2n-1}} \\ &= 1 - \frac{1}{\{c_1 + [c_2 + \cdots + (c_{n-1} + c_nx^2)x^2 \cdots]x^2\}x} \quad x > x_M \end{aligned} \quad (7.2-94)$$

The first of these functions is suggested by the form of the power-series expansion of PI(x) about the origin. The second is motivated by the fact that, for large values of x ,

$$1 - \text{PI}(x) \sim \frac{1}{x} \text{PF}(x) \quad (7.2-95)$$

where this asymptotic equality implies that the ratio of the two members tends to unity as x tends toward infinity. Even with the range of the function broken into two parts, and with the coefficients b_i and c_i optimized on the basis of minimum sum of squared errors, 20 coefficients are required in each of these functions to achieve an error no greater than $1 \cdot 10^{-6}$ over the full range of x . The appropriate point at which to break the range of x is in this case $x_M = 3.3$. Sets of coefficients which yield $1 \cdot 10^{-6}$ accuracy are listed in Table 7.2-3. Note that the first of these functions, for x of smaller magnitude, is valid for both positive and negative values of x . The second

TABLE 7.2-3 COEFFICIENTS IN THE APPROXIMATIONS FOR PF(x) AND PI(x)

$PF(x) \approx f_F(x) = \frac{0.39894229}{1 + \{a_1 + [a_2 + \cdots + (a_{12} + a_{13}x^2)x^2 \cdots]x^2\}x^2}$	
$a_1 = 0.50000000$	$a_8 = 0.96881187 \cdot 10^{-7}$
$a_2 = 0.12500000$	$a_9 = 0.53822881 \cdot 10^{-8}$
$a_3 = 0.20833333 \cdot 10^{-1}$	$a_{10} = 0.26911439 \cdot 10^{-9}$
$a_4 = 0.26041665 \cdot 10^{-2}$	$a_{11} = 0.12232172 \cdot 10^{-10}$
$a_5 = 0.26041664 \cdot 10^{-3}$	$a_{12} = 0.50847426 \cdot 10^{-12}$
$a_6 = 0.21701386 \cdot 10^{-4}$	$a_{13} = 0.37188551 \cdot 10^{-13}$
$a_7 = 0.15500990 \cdot 10^{-5}$	
$PI(x) \approx f_I(x)_1 = 0.5 + \{b_1 + [b_2 + \cdots + (b_{19} + b_{20}x^2)x^2 \cdots]x^2\}x \quad x \leq 3.3$	
$b_1 = 0.39894229$	$b_{11} = 0.51124335 \cdot 10^{-11}$
$b_2 = -0.66490381 \cdot 10^{-1}$	$b_{12} = -0.21217608 \cdot 10^{-12}$
$b_3 = 0.99735570 \cdot 10^{-2}$	$b_{13} = 0.81334163 \cdot 10^{-14}$
$b_4 = -0.11873282 \cdot 10^{-2}$	$b_{14} = -0.28965156 \cdot 10^{-15}$
$b_5 = 0.11543468 \cdot 10^{-3}$	$b_{15} = 0.96312710 \cdot 10^{-17}$
$b_6 = -0.94446554 \cdot 10^{-5}$	$b_{16} = -0.30032944 \cdot 10^{-18}$
$b_7 = 0.66596923 \cdot 10^{-6}$	$b_{17} = 0.88214403 \cdot 10^{-20}$
$b_8 = -0.41226666 \cdot 10^{-7}$	$b_{18} = -0.19304226 \cdot 10^{-21}$
$b_9 = 0.22735293 \cdot 10^{-8}$	$b_{19} = 0.18588157 \cdot 10^{-26}$
$b_{10} = -0.11301169 \cdot 10^{-9}$	$b_{20} = 0.24283477 \cdot 10^{-25}$
$PI(x) \approx f_I(x)_2 = 1 - \frac{1}{\{c_1 + [c_2 + \cdots + (c_{19} + c_{20}x^2)x^2 \cdots]x^2\}x} \quad x > 3.3$	
$c_1 = 0.15145329 \cdot 10^2$	$c_{11} = 0.66769241 \cdot 10^{-9}$
$c_2 = 0.21862164 \cdot 10^1$	$c_{12} = 0.30408705 \cdot 10^{-10}$
$c_3 = 0.37773824$	$c_{13} = 0.13631897 \cdot 10^{-11}$
$c_4 = 0.56444044 \cdot 10^{-1}$	$c_{14} = 0.73926276 \cdot 10^{-13}$
$c_5 = 0.67908929 \cdot 10^{-2}$	$c_{15} = 0.59690240 \cdot 10^{-14}$
$c_6 = 0.66819702 \cdot 10^{-3}$	$c_{16} = 0.56745842 \cdot 10^{-15}$
$c_7 = 0.55222069 \cdot 10^{-4}$	$c_{17} = 0.33853563 \cdot 10^{-16}$
$c_8 = 0.39226156 \cdot 10^{-5}$	$c_{18} = -0.37560769 \cdot 10^{-17}$
$c_9 = 0.24387498 \cdot 10^{-6}$	$c_{19} = 0.32913919 \cdot 10^{-21}$
$c_{10} = 0.13458967 \cdot 10^{-7}$	$c_{20} = 0.18022685 \cdot 10^{-20}$

function, for x of larger magnitude, is valid only for positive x ; Eq. (7.2-36) is then used to define the function for negative values.

The integral expression for the describing function gain to each input component is singular for a zero magnitude of that component. Thus one may expect the expression to be poorly behaved numerically for small magnitudes of input. An accurate reference point for small signals is provided by the limiting forms of the describing functions derived in Sec. 1.5. For static single-valued nonlinearities, Eq. (1.5-63) states

$$\lim_{B \rightarrow 0} N_B = N_R(B = 0)$$

and Eq. (1.5-66) gives

$$\lim_{A \rightarrow 0} N_A = N_R(A = 0)$$

The expression for N_R remains well behaved for small values of A and B . No similar expression is known for N_R , but differentiation of Eq. (1.5-59) with respect to B reveals the fact that

$$N_R = \frac{d}{dB} (BN_B) \quad (7.2-96)$$

This relation is true in general; in particular, it remains well-behaved for small values of σ .

Another approach to approximate RIDF calculation has been suggested by Atherton (Ref. 2). Using the transform of the nonlinear function, he calculates the RIDFs for a general quantizer nonlinearity in which the steps in both input and output variables are arbitrary. This leads to expressions for the gains in the form of sums of terms which involve the confluent hypergeometric function and the gamma function, among others. These results define the RIDFs for a stepwise approximation to any nonlinearity, and if the required functions are prepared as subroutines for digital computation, the approximation can be made as good as desired.

Having developed adequate means of calculating RIDFs, we now apply them to study the behavior of nonlinear feedback systems with random and other inputs. For convenient reference the three-input RIDFs for several common nonlinearities are presented in Appendix E.

7.3 FEEDBACK SYSTEMS WITH RANDOM SIGNALS AND NOISE

function, for x of larger magnitude, is valid only for positive x ; Eq. (7.2-36) is then used to define the function for negative values.

The integral expression for the describing function gain to each input component is singular for a zero magnitude of that component. Thus one may expect the expression to be poorly behaved numerically for small magnitudes of input. An accurate reference point for small signals is provided by the limiting forms of the describing functions derived in Sec. 1.5. For static single-valued nonlinearities, Eq. (1.5-63) states

$$\lim_{B \rightarrow 0} N_B = N_R(B = 0)$$

and Eq. (1.5-66) gives

$$\lim_{A \rightarrow 0} N_A = N_R(A = 0)$$

The expression for N_R remains well behaved for small values of A and B . No similar expression is known for N_R , but differentiation of Eq. (1.5-59) with respect to B reveals the fact that

$$N_R = \frac{d}{dB} (BN_B) \quad (7.2-96)$$

This relation is true in general; in particular, it remains well-behaved for small values of σ .

Another approach to approximate RIDF calculation has been suggested by Atherton (Ref. 2). Using the transform of the nonlinear function, he calculates the RIDFs for a general quantizer nonlinearity in which the steps in both input and output variables are arbitrary. This leads to expressions for the gains in the form of sums of terms which involve the confluent hypergeometric function and the gamma function, among others. These results define the RIDFs for a stepwise approximation to any nonlinearity, and if the required functions are prepared as subroutines for digital computation, the approximation can be made as good as desired.

Having developed adequate means of calculating RIDFs, we now apply them to study the behavior of nonlinear feedback systems with random and other inputs. For convenient reference the three-input RIDFs for several common nonlinearities are presented in Appendix E.

7.3 FEEDBACK SYSTEMS WITH RANDOM SIGNALS AND NOISE

describing functions and by the required satisfaction of a number of simultaneous nonlinear algebraic relations which define the solution. With the thought that it may be prudent to approach these details gradually, this section is restricted to the class of problems in which there is only a single identifiable signal form at the input to the nonlinearity—a gaussian random process. This very important class of problems includes cases in which the system responds to input signals and noise if both are characterized by gaussian processes. We shall even be able to design the Wiener-optimum compensation for such systems, but let us begin with the basic problem of a nonlinear system responding to a single gaussian input.

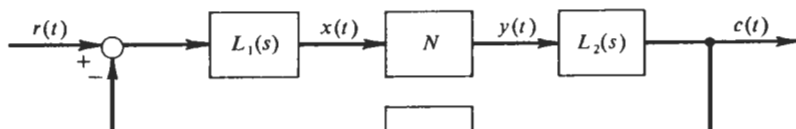
THE BASIC PROBLEM

The system configuration is shown in Fig. 7.3-1. The system input $r(t)$ is a stationary gaussian process whose autocorrelation function $\varphi_{rr}(\tau)$, or equivalently, whose power spectral density function $\Phi_{rr}(\omega)$, is given. If the nonlinearity input $x(t)$ is to be modeled adequately as a gaussian process, we must assume that the system is not limit cycling. This is a situation precisely paralleling the study of the frequency response of a nonlinear system using the single-sinusoid-input DF as in Chap. 3, and the same observations made earlier regarding the possibility of a system which does not limit-cycle in the absence of input exhibiting one in the presence of an input, or the no-input limit cycle of a system being quenched by an input, are applicable here. With the assumption of no limit cycle, $x(t)$ is just the random response due to the input, and this response is considered gaussian for the purpose of calculating the describing function for the nonlinearity.

The single-random-input describing function is given by Eq. (7.2-4) with A and B equal to zero.

$$\begin{aligned} N_R &= \frac{1}{\sqrt{2\pi}\sigma^3} \int_{-\infty}^{\infty} ry(r) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} vy(\sigma v) \exp\left(-\frac{v^2}{2}\right) dv \end{aligned} \quad (7.3-1)$$

This integral can be expressed analytically for a number of nonlinearities, in addition to those for which the double-integral expressions of the preceding



section can be evaluated. Some of these expressions are tabulated in Appendix E.

With the effect of the nonlinearity on the input random process expressed in terms of its RIDF, the system has a quasi-linear description. Standard linear theory can then be used to evaluate the mean-squared value of the signal that appears at any point in the loop. In particular, the standard deviation σ of $x(t)$ is defined by the square root of

$$\sigma^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(\omega) d\omega \quad (7.3-2)$$

$$\text{where} \quad \Phi_{xx}(\omega) = \Phi_{rr}(\omega)H(j\omega)H(-j\omega) \quad (7.3-3)$$

$$\text{and} \quad H(j\omega) = \frac{L_1(j\omega)}{1 + N_R L_1(j\omega)L_2(j\omega)L_3(j\omega)} \quad (7.3-4)$$

The power spectral density function $\Phi(\omega)$ was introduced as a function of the real frequency ω because only in this case does the function have a clear physical significance: it is the power per unit frequency interval of the harmonic components of the function in the vicinity of ω . However, control engineers are more accustomed to functions expressed in terms of the Laplace transform variable s . Equations (7.3-2) to (7.3-4) can be written in terms of this variable by simply introducing the change of variable $s = j\omega$.

$$\sigma^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \Phi_{xx}(s) ds \quad (7.3-5)$$

$$\Phi_{xx}(s) = \Phi_{rr}(s)H(s)H(-s) \quad (7.3-6)$$

$$H(s) = \frac{L_1(s)}{1 + N_R L_1(s)L_2(s)L_3(s)} \quad (7.3-7)$$

Strictly speaking, the symbol $\Phi(s)$ is improper since it does not refer to $\Phi(\omega)$ with ω replaced by s ; rather, ω is replaced by s/j . However, the symbol is convenient; it is in common use; and no confusion should arise. For rational spectra, ω appears only in even powers; so $\Phi(s)$ is derived from $\Phi(\omega)$ simply by replacing ω^2 by $-s^2$.

The integral expression for σ^2 [Eq. (7.3-5)] can be evaluated analytically if $\Phi_{xx}(s)$ is a rational function of s^2 . This will be true of any case in which the spectrum $\Phi_{rr}(s)$ is rational in s^2 and the transfer function $H(s)$ is rational in s . For this case, the integral has been evaluated by a number of writers and appears tabulated in many books and journal articles. Table 7.3-1 lists the results for low-ordered spectra; more comprehensive tables may be found in

TABLE 7.3-1 TABULATED INTEGRALS FOR RATIONAL SPECTRA

$$V_n = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{C(s)C(-s)}{D(s)D(-s)} ds$$

$$C(s) = c_0 + c_1s + c_2s^2 + \cdots + c_{n-1}s^{n-1}$$

$$D(s) = d_0 + d_1s + d_2s^2 + \cdots + d_ns^n$$

$$V_1 = \frac{c_0^2}{2d_0d_1}$$

$$V_2 = \frac{c_1^2d_0 + c_0^2d_2}{2d_0d_1d_2}$$

$$V_3 = \frac{c_2^2d_0d_1 + (c_1^2 - 2c_0c_2)d_0d_3 + c_0^2d_2d_3}{2d_0d_3(d_1d_2 - d_0d_3)}$$

$$V_4 = \frac{c_3^2(d_0d_1d_2 - d_0^2d_3) + (c_2^2 - 2c_1c_3)d_0d_1d_4 + (c_1^2 - 2c_0c_2)d_0d_3d_4 + c_0^2(d_2d_3d_4 - d_1d_4^2)}{2d_0d_4(d_1d_2d_3 - d_0d_3^2 - d_1^2d_4)}$$

$$V_5 = \frac{1}{2\Delta} [c_4^2m_0 + (c_3^2 - 2c_2c_4)m_1 + (c_2^2 - 2c_1c_3 + 2c_0c_4)m_2 + (c_1^2 - 2c_0c_2)m_3 + c_0^2m_4]$$

$$m_0 = \frac{1}{d_5} (d_3m_1 - d_1m_2) \quad m_3 = \frac{1}{d_0} (d_2m_2 - d_4m_1)$$

$$m_1 = d_1d_2 - d_0d_3 \quad m_4 = \frac{1}{d_0} (d_2m_3 - d_4m_2)$$

$$m_2 = d_1d_4 - d_0d_5 \quad \Delta = d_0(d_1m_4 - d_3m_3 + d_5m_2)$$

simultaneous solution defines the rms value of $x(t)$ for the given input. The most convenient way to effect the solution is to choose a value of σ , the rms value of $x(t)$, and determine the corresponding value of N_R from Eq. (7.3-1). Equation (7.3-5) is then integrated using this value of N_R . The mean-squared value of $r(t)$ can be kept in literal form since it just appears as a multiplying factor for the integral. It is then a simple matter to solve for the rms value of $r(t)$, which results in the value of σ originally chosen. This process is repeated with other values of σ until the relation between rms input and rms value of $x(t)$ is well enough defined to permit interpolation among the results to find the response to a given input.

The solution of σ_r for a given σ_x will always be single-valued. However, different values of σ_x may lead to the same value of σ_r ; that is, for a given input, a *multiple-valued response* may be indicated. Here again is a situation which parallels the sinusoidal response characteristics of similar systems. In that case, a multiple-valued response was termed jump resonance. In this case such a response characteristic is usually called *dual-mode operation*,

error mode tends to be maintained consistently, depending on initial conditions. In the random input case, as one would expect, the system changes apparently at random between the large and small error modes of operation.

Having the value of σ_x , and thus of N_R , determined for any input, the response at any other point in the loop, such as at the output $c(t)$, can be determined by linear methods. The limitation on this is that there must be adequate filtering of the higher-frequency nongaussian content in the output of the nonlinearity. The quasi-linear description of the nonlinearity used here does not model accurately the power spectrum, the distribution, or the mean-squared value of the process at the output of the nonlinearity. The more linear filtering this process experiences, however, the more nearly these characteristics will approach those predicted by this analysis. This too is entirely equivalent to the sinusoidal response situation.

Example 7.3-1 As an illustration of the procedure just described, consider the second-order system shown in Fig. 7.3-2. The input is a gaussian process, having the power density spectrum

$$\Phi_{rr}(s) = \frac{k}{a^2 - s^2}$$

Using V_1 of Table 7.3-1, the mean-squared value of this input is found to be

$$\sigma_r^2 = \frac{k}{2a}$$

Thus the spectrum can be rewritten to display the mean-squared input directly.

$$\Phi_{rr}(s) = \frac{2a\sigma_r^2}{a^2 - s^2} \quad (7.3-8)$$

The transfer from input to $x(t)$ is

$$H_x(s) = \frac{X(s)}{R(s)} = \frac{s^2}{s^2 + N_R s + N_R}$$

in which the RIDF is used to characterize the gain of the nonlinearity to its random input.

$$\Phi_{xx}(s) = \frac{2a\sigma_r^2}{a^2 - s^2} H_x(s)H_x(-s)$$

Taking out the factor $2a\sigma_r^2$, the function $C(s)/D(s)$ in the integral form of Table 7.3-1 is identified as

$$\frac{C(s)}{D(s)} = \frac{1}{a + s} \frac{s^2}{s^2 + N_R s + N_R} = \frac{s^2}{s^3 + (a + N_R)s^2 + (1 + a)N_R s + aN_R}$$

V_3 from this table then gives

$$\sigma_x^2 = \frac{2a\sigma_r^2 a(1 + a)N_R^2}{s^3 + (a + N_R)s^2 + (1 + a)N_R s + aN_R} = \frac{a\sigma_r^2}{s^3 + (a + N_R)s^2 + (1 + a)N_R s + aN_R} \quad (7.3-9)$$

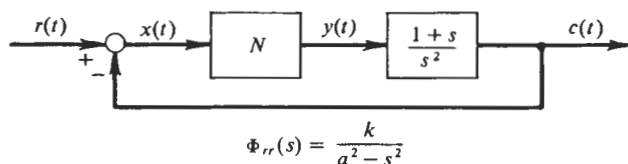


Figure 7.3-2 System of Example 7.3-1.

Equation (7.2-16) with $A = B = 0$ gives the RIDF as

$$N_R = 3\sigma_x^2 \quad (7.3-10)$$

Now for any given bandwidth of the input spectrum, say, $a = 1 \text{ sec}^{-1}$, Eqs. (7.3-9) and (7.3-10) can be solved. For example, take $\sigma_x = 1$ unit. Then

$$N_R = 3$$

$$\sigma_r = 1\sqrt{3 + \frac{1}{2}} = 1.87 \text{ units}$$

This calculation is repeated to define a curve of σ_x versus σ_r for a given a . Then the spectrum of the output is

$$\Phi_{cc}(s) = \frac{2a\sigma_r^2}{a^2 - s^2} H_c(s)H_c(-s)$$

$$H_c(s) = \frac{C(s)}{R(s)} = \frac{N_R(s+1)}{s^2 + N_R s + N_R}$$

and the mean-squared output is

$$\sigma_c^2 = \frac{\sigma_r^2}{N_R} \frac{N_R + a/(1+a)}{N_R + a^2/(1+a)}$$

An example of the response of this nonlinear system to a gaussian input is shown in Fig. 7.3-3, together with the response of the corresponding quasi-linearized system to the same input. The evident close correlation between the outputs, $c(t)$, of the two systems indicates the success of the quasi-linear description of the operation of the nonlinearity. Notice that the signal $y(t)$ is not well modeled. This effect is even more striking if the nonlinearity is an ideal relay, as shown in Fig. 7.3-4.

ACCURACY

As part of a more comprehensive study, Smith (Refs. 40, 41) checked the accuracy of this analysis, using the system of Fig. 7.3-2 and three different nonlinearities. For each nonlinearity he computed the theoretical relation between σ_r and σ_x for a range of input rms values and for three different bandwidths of the input spectrum. He then measured the actual response of the system as simulated with an analog computer. His results are shown in Figs. 7.3-5 to 7.3-7. The agreement between predicted and observed

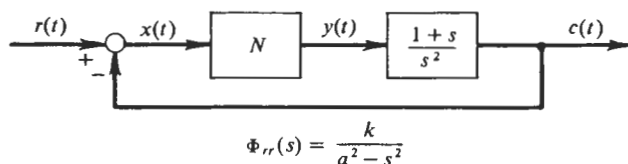


Figure 7.3-2 System of Example 7.3-1.

Equation (7.2-16) with $A = B = 0$ gives the RIDF as

$$N_R = 3\sigma_x^2 \quad (7.3-10)$$

Now for any given bandwidth of the input spectrum, say, $a = 1 \text{ sec}^{-1}$, Eqs. (7.3-9) and (7.3-10) can be solved. For example, take $\sigma_x = 1$ unit. Then

$$N_R = 3$$

$$\sigma_r = 1\sqrt{3 + \frac{1}{2}} = 1.87 \text{ units}$$

This calculation is repeated to define a curve of σ_x versus σ_r for a given a . Then the spectrum of the output is

$$\Phi_{cc}(s) = \frac{2a\sigma_r^2}{a^2 - s^2} H_c(s)H_c(-s)$$

$$H_c(s) = \frac{C(s)}{R(s)} = \frac{N_R(s+1)}{s^2 + N_R s + N_R}$$

and the mean-squared output is

$$\sigma_c^2 = \frac{\sigma_r^2}{N_R} \frac{N_R + a/(1+a)}{N_R + a^2/(1+a)}$$

An example of the response of this nonlinear system to a gaussian input is shown in Fig. 7.3-3, together with the response of the corresponding quasi-linearized system to the same input. The evident close correlation between the outputs, $c(t)$, of the two systems indicates the success of the quasi-linear description of the operation of the nonlinearity. Notice that the signal $y(t)$ is not well modeled. This effect is even more striking if the nonlinearity is an ideal relay, as shown in Fig. 7.3-4.

ACCURACY

As part of a more comprehensive study, Smith (Refs. 40, 41) checked the accuracy of this analysis, using the system of Fig. 7.3-2 and three different nonlinearities. For each nonlinearity he computed the theoretical relation between σ_r and σ_x for a range of input rms values and for three different bandwidths of the input spectrum. He then measured the actual response of the system as simulated with an analog computer. His results are shown in Figs. 7.3-5 to 7.3-7. The agreement between predicted and observed

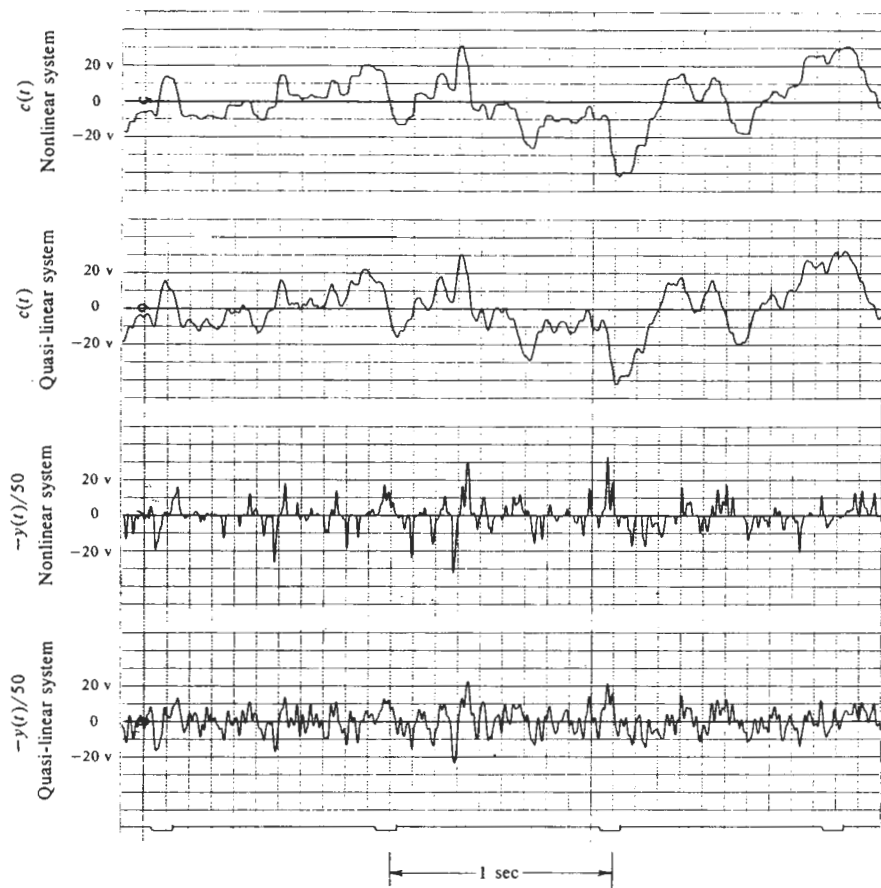


Figure 7.3-3 Sample of response to a random input. Fig. 7.3-2 system, cubic nonlinearity.

decrease can be observed from these results. Note, first, that in every instance in which the data lie close to the line $\sigma_x = \sigma_r$, the accuracy is excellent. This is the condition in which the rms error is nearly as large as the rms input, which implies that the feedback in the system is not effective. This occurs mostly for the wideband input. If the feedback signal is small compared with the input signal, $x(t)$ has nearly the properties of $r(t)$, including its normal distribution. This is the condition under which the analysis was developed. Only when σ_x is appreciably smaller than σ_r does the accuracy decrease, and then not in every case. Note that for the limiter system (Fig. 7.3-7), with $a = 1$ and small rms input, σ_x is much smaller than

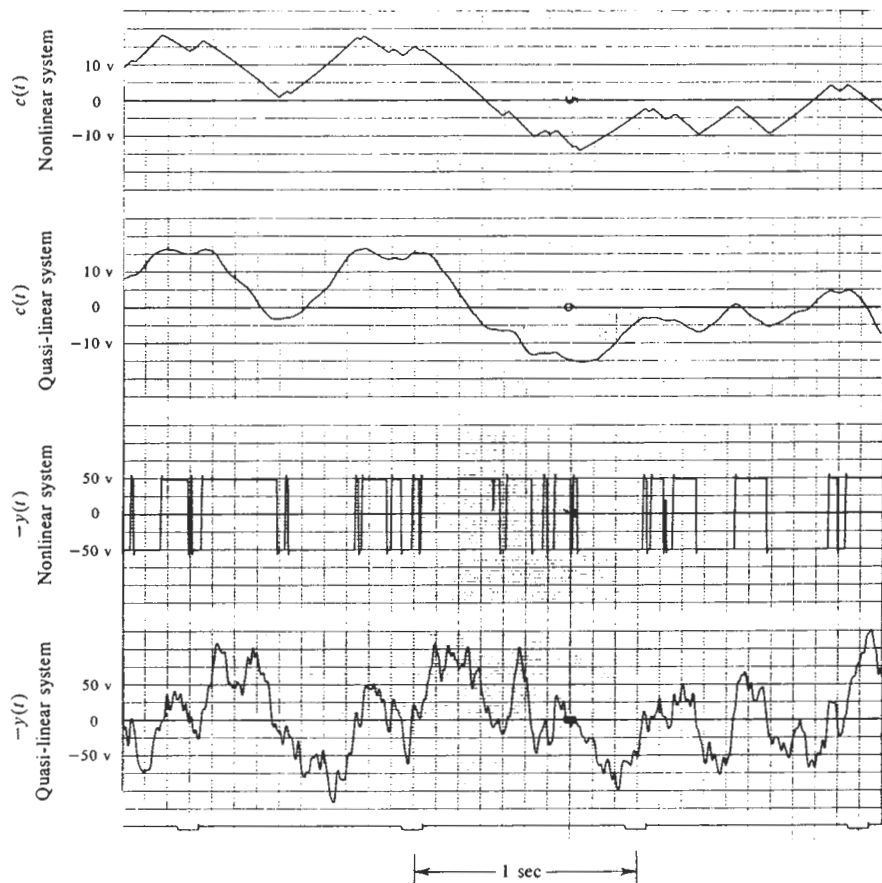


Figure 7.3-4 Sample of response to a random input. System of Fig. 7.3-2, ideal-relay nonlinearity.

An interpretation of these results can be based on the distribution of the signal at the input to the nonlinearity. With a gaussian input to the system, $x(t)$ can have a distribution which differs significantly from the normal only if an appreciable amount of the distortion generated by the nonlinearity is fed back around the loop to the input to the nonlinearity. An appreciable amount of fed-back signal is implied if the rms error is much smaller than the rms input; this is indeed the intent of the feedback configuration. But in the case of the limiter system with rms error small relative to the value of x at which limiting occurs, the limiter is only rarely driven into the non-

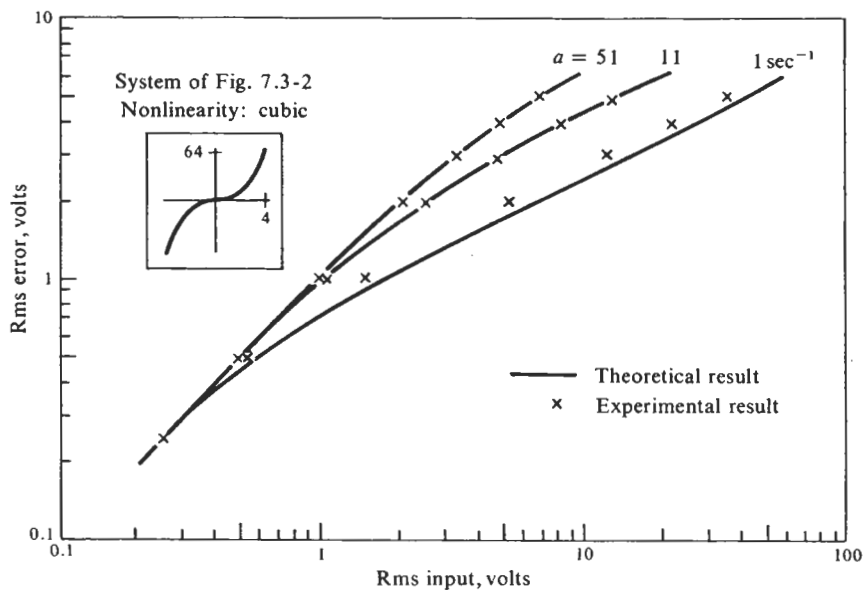
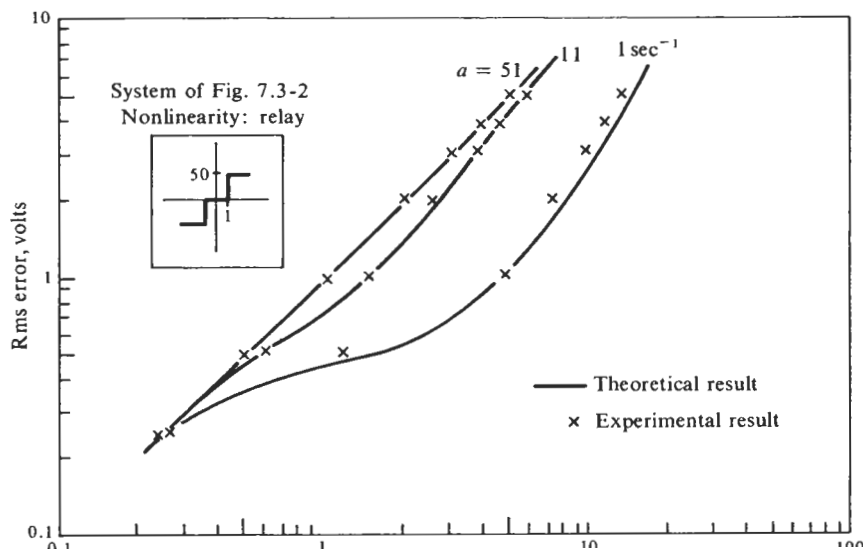


Figure 7.3-5 Response to gaussian input. (Adapted from Smith, Ref. 40.)



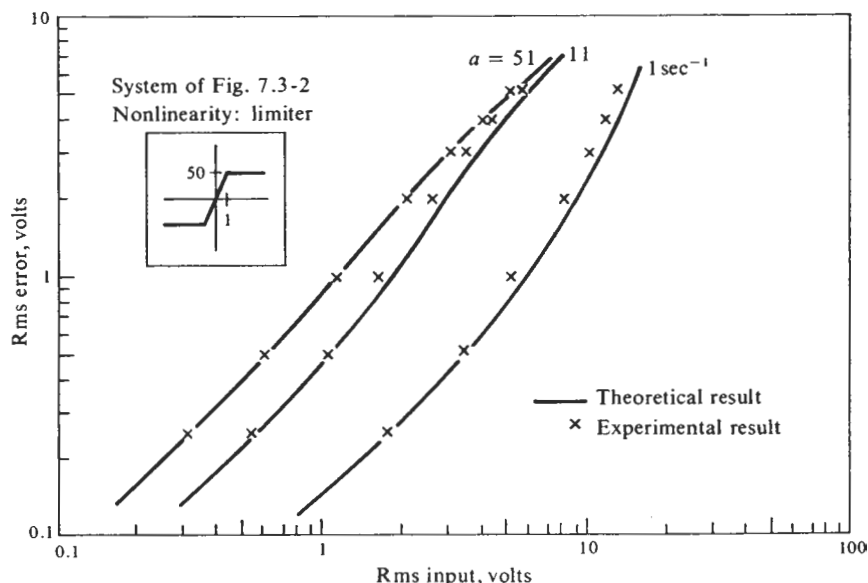


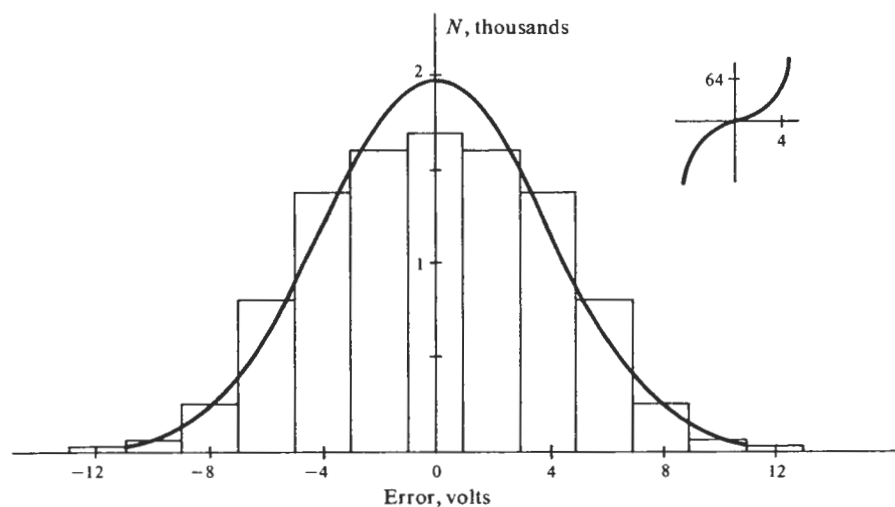
Figure 7.3-7 Response to gaussian input. (Adapted from Smith, Ref. 40.)

must be appreciable compared with the input signal, and it must contain significant nonlinear distortion.

As an indication of the sensitivity of the accuracy of this analysis to the distribution of $x(t)$, histograms showing the distribution of the nonlinearity input for three points in Figs. 7.3-5 to 7.3-7, where the theoretical values are most in error, are given in Figs. 7.3-8 to 7.3-10. The normal distribution having the same rms value is shown in each case for reference. These distributions differ quite significantly from the normal, yet the rms error was predicted with less than 10 percent error in each of these cases.

Nikiforuk (Ref. 30) and Nikiforuk and West (Ref. 31) used narrowband filtering to measure the power density spectrum of the output of second-order systems with hard-spring and limiter nonlinearities. Comparison of the output spectrum with the input permits evaluation of the magnitude of the system transfer function over the important frequency range. This experimentally determined frequency response function was compared with that predicted by this analysis and found to agree with better than 20 percent accuracy, except in a low-frequency region, where they questioned their experimental results.

It might be said in summary that the accuracy of the RIDF in dealing with a single random input is of the same order as the accuracy of the ordinary DF



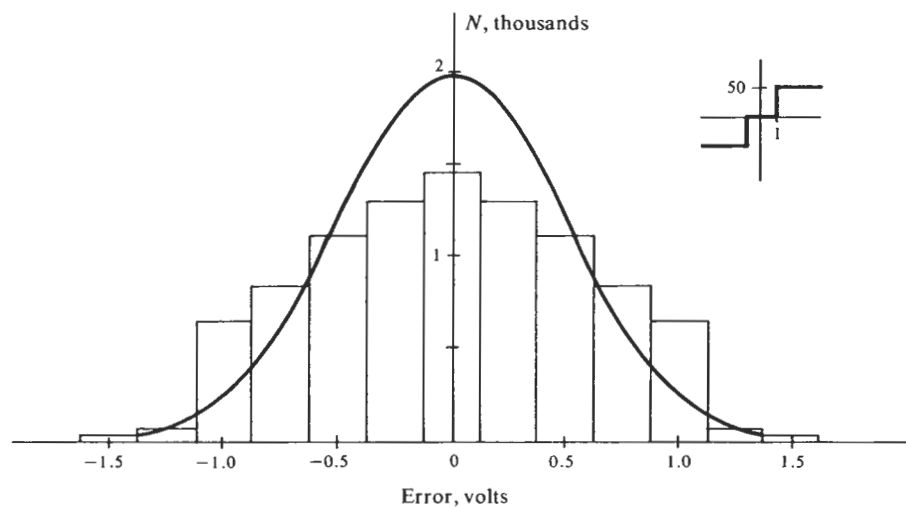
System of Fig. 7.3-2

Rms error = 4.0 volts

Nonlinearity: cubic

 $a = 1 \text{ sec}^{-1}$

Figure 7.3-8 Distribution of 10,000 error samples. (Adapted from Smith, Ref. 40.)



System of Fig. 7.3-2

Rms error = 0.5 volt

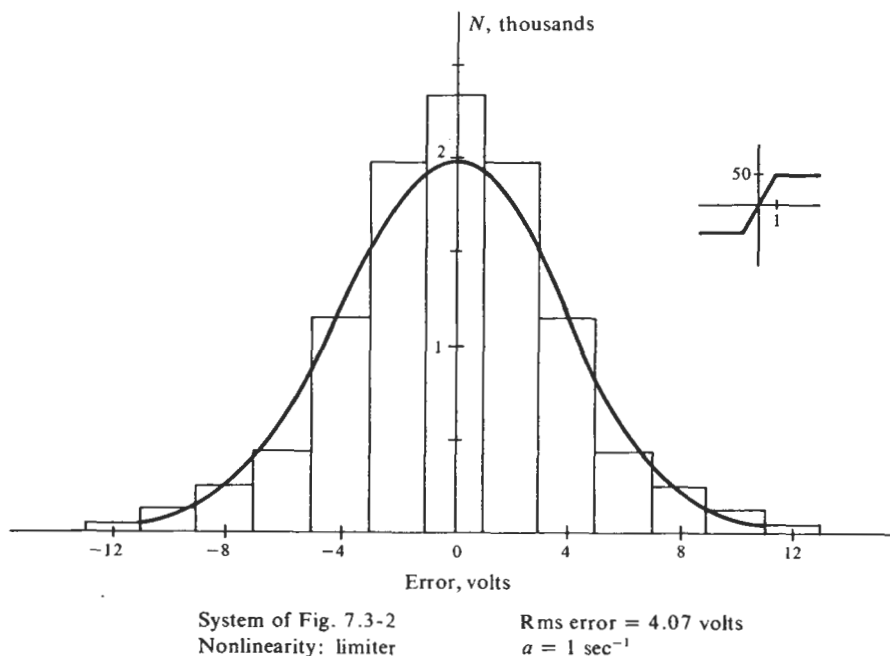


Figure 7.3-10 Distribution of 10,000 error samples. (Adapted from Smith, Ref. 40.)

plants would provide better filtering of nonlinear distortion and permit even better analytic description. When accuracy is degraded, it is due to inadequate filtering of the nongaussian nonlinearity output, which then violates the assumption of a normal distribution at the input to the nonlinearity. Any theory which is to be more accurate than this over a broad range of problems will have to deal with the possibility of a distorted distribution for the nonlinearity input. This, unfortunately, poses a problem which is likely to remain intractable for practical purposes.

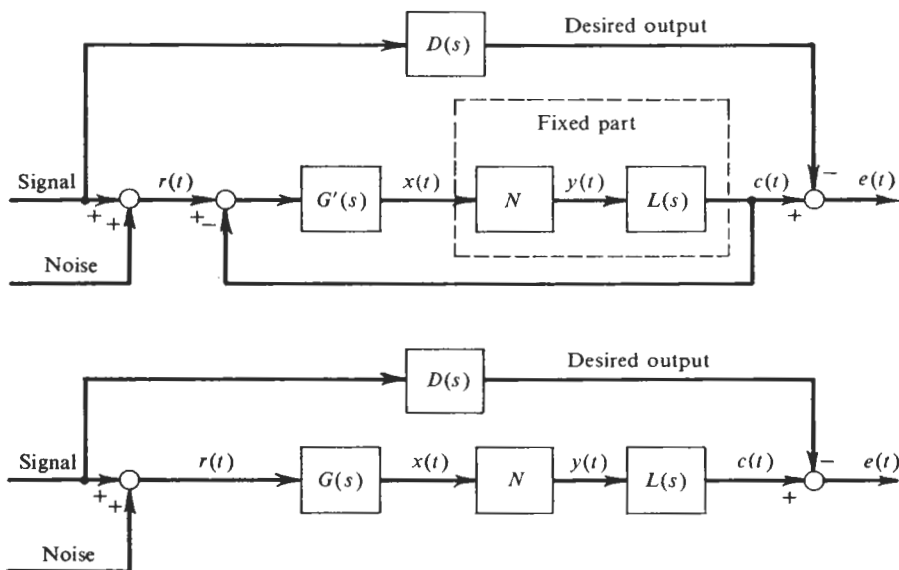
OPTIMUM LINEAR COMPENSATION

This section has thus far considered the *analysis* of a nonlinear system subjected to a random input. We turn now to the problem of the *synthesis*, or design, of such systems under stated sets of conditions. One of the classics among optimal design problems is the Wiener problem of the separation of a signal from noise with a linear filter. The extension of the solution to the Wiener problem to the design of optimum linear compensation for a system

The conditions of the problem are these:

1. The system responds to an input which is the sum of a signal and noise, both of which are considered stationary gaussian random processes with given power spectral density functions.
2. The desired output of the system is a function derived from the signal component of the input by some linear invariant operation, not necessarily physically realizable.
3. The criterion for optimality is minimum steady-state mean-squared error, which is the difference between the actual output of the system and the desired output.
4. The system contains an invariant stable linear part and a static single-valued nonlinear part which are considered fixed. The optimum linear, physically realizable compensation is to be designed for this system.

The system configuration is shown in Fig. 7.3-11. N and $L(s)$ characterize the fixed nonlinear and linear parts of the system; $G'(s)$ is the compensator to be designed. The original feedback configuration is recast in an open-chain configuration with a known relationship between the new cascade compensator $G(s)$ and the original system compensator $G'(s)$. This is the configuration to which familiar filter theory is applicable were it not for the



nonlinear element in the chain. Using the theory of this chapter, the nonlinearity is characterized by its RIDF, which depends on σ_x . But the mean-squared value of $x(t)$ is not known until $G(s)$ is determined, and that depends on the gain of the nonlinearity. A logical way out of this dilemma is provided by the constrained optimum-design technique given by Newton, Gould, and Kaiser (Ref. 29). If a value is chosen for σ_x , N_R is determined, and the constrained-optimum-design solution gives the function $G(s)$ which minimizes the mean-squared error subject to the constraint that the mean-squared value of $x(t)$ is as assumed. This solution is repeated for a number of values of the constrained variable σ_x^2 . The mean-squared error corresponding to each of these is noted, and the solution which yields the smallest mean-squared error is optimum.

The solution to the constrained-optimum-design problem specialized to the configuration of Fig. 7.3-11 is given here for convenient reference. For the development of this solution, the reader is referred to Newton, Gould, and Kaiser (*op. cit.*).

With reference to Fig. 7.3-11, the quantity to be minimized is $\overline{e(t)^2} + \lambda \overline{x(t)^2}$, where λ is a Lagrange multiplier. We are considering unbiased variables in this section; so $\overline{x(t)^2} = \sigma_x^2$. The solution proceeds in the following steps:

1. Choose a value of σ_x . Calculate the corresponding value of N_R from

$$N_R = \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} y(\sigma_x v) v \exp\left(-\frac{v^2}{2}\right) dv$$

2. Form the function

$$H(s) = N_R^2 L(s)L(-s) + \lambda$$

and factor it.

$$H(s) = H(s)_L H(s)_R$$

$H(s)_L$ contains all the factors of $H(s)$ which define poles or zeros in the left half-plane, and $H(s)_R$ contains all the factors of $H(s)$ which define poles or zeros in the right half-plane. Any poles or zeros on the $j\omega$ axis are moved just off the axis by the addition of a real part, ϵ . For example, if $H(s)$ has the factors $1/(s)(-s)$, they are replaced by $1/(s + \epsilon)(-s + \epsilon)$, the first factor becoming a factor of $H(s)_L$ and the second a factor of $H(s)_R$.

3. Form the power spectral density functions $[r(t) = s(t) + n(t)]$

$$\Phi_{rs}(s) = \Phi_{ss}(s) + \Phi_{ns}(s)$$

$$\Phi_{rr}(s) = \Phi_{ss}(s) + \Phi_{ns}(s) + \Phi_{sn}(s) + \Phi_{nn}(s)$$

and factor the input spectrum.

4. Form the function

$$M(s) = \frac{D(s)N_R L(-s)\Phi_{rs}(s)}{H(s)_R \Phi_{rr}(s)_R}$$

and separate it into

$$M(s) = M(s)_L + M(s)_R$$

where $M(s)_L$ has its poles only in the left half-plane, and $M(s)_R$ has its poles only in the right half-plane. The zeros of these functions may lie in either half-plane. This step is carried out by partial fraction expansion of $M(s)$ and adding together those fractions which define poles in the left half-plane to form $M(s)_L$. $M(s)_R$ need not be evaluated explicitly.

5. The optimum cascade compensation function is

$$G(s) = \frac{M(s)_L}{H(s)_L \Phi_{rr}(s)_L}$$

This is a function of the Lagrange multiplier λ .

6. Use the constraint condition

$$\sigma_x^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \Phi_{rr}(s) G(s) G(-s) ds$$

to evaluate λ , and hence $G(s)$, to yield the value of σ_x assumed in step 1.

7. Calculate the mean-squared error.

$$\bar{e}^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \Phi_{ee}(s) ds$$

$$\begin{aligned} \Phi_{ee}(s) &= E(s)E(-s)\Phi_{ss}(s) + N_R^2 G(s)L(s)G(-s)L(-s)\Phi_{nn}(s) \\ &\quad + N_R G(s)L(s)E(-s)\Phi_{sn}(s) + N_R G(-s)L(-s)E(s)\Phi_{ns}(s) \\ E(s) &= N_R G(s)L(s) - D(s) \end{aligned}$$

The first two terms in this expression for the error power spectral density function are of the form of the integral tabulated in Table 7.3-1. The sum of the last two terms is also of this form.

This solution is repeated for enough values of σ_x to define a curve of $\bar{e}^2$ versus σ_x . The value of σ_x which yields the minimum $\bar{e}^2$ is then used in the final solution for the compensator $G(s)$. The solution to the problem of the design of the optimum linear compensation function for a nonlinear system of the form considered here is tedious, to be sure, but requires no new concepts and does provide a very satisfying solution to a rather

We might just note that describing function theory is applicable to systems with nonlinear parts separated by sufficient linear filtering regardless of what analytic form one chooses to describe the system. Time-variable systems or nonstationary system statistics are more conveniently treated with a *state space* characterization of the system than with a time-variable extension of integral transform theory. Under the assumption of known distribution for the components of a system state vector, the instantaneous nonlinearities involved in the system differential equation of state and in any nonlinear measurements taken on the system can be approximated as quasi-linear gain matrices using a multidimensional extension of Chap. 1 theory. In the case of a *non-limit-cycling nonlinear system driven by gaussian noise*, these gain matrices depend on the mean value and covariance matrix of the system state vector, and of any input noises. Having a quasi-linear description of the nonlinear system, linear filter theory in the form of the *Kalman filter* is employed directly to obtain approximate indications of the time-varying first- and second-order statistics of the system.

OPTIMUM NONLINEAR COMPENSATION

A somewhat different problem is presented by a system having a given fixed part which is just asked to follow a random input as well as possible in the least-mean-squared-error sense. Consider any linear compensation to be fixed for simplicity in instrumentation. Then, for a given input spectrum, it is a simple matter to determine the optimum forward gain. Suppose now that we should like the system to respond optimally to inputs of various power levels; specifically, consider the shape and bandwidth of the input spectrum to remain fixed but scaled in amplitude to correspond to different rms input levels. For each input power level there will be a different optimum value of forward gain if the system fixed part contains a nonlinearity. What is needed, then, is a forward gain which depends on its input rms value, that is, a nonlinear element. If one permits some liberty in the application of RIDF theory, it is possible to design such a nonlinear compensator in an optimal way. The technique was first documented by Douce and King (Ref. 13).

The configuration of the system to be considered is shown in Fig. 7.3-12. The input signal is a stationary gaussian process with a power density spectrum which is of a prescribed shape but with a variable amplitude scaling. The fixed part includes the controlled member, its drive system, and any linear compensation which has been designed and is held fixed in this process. The nonlinearity N_1 is to be designed to yield a minimum mean-

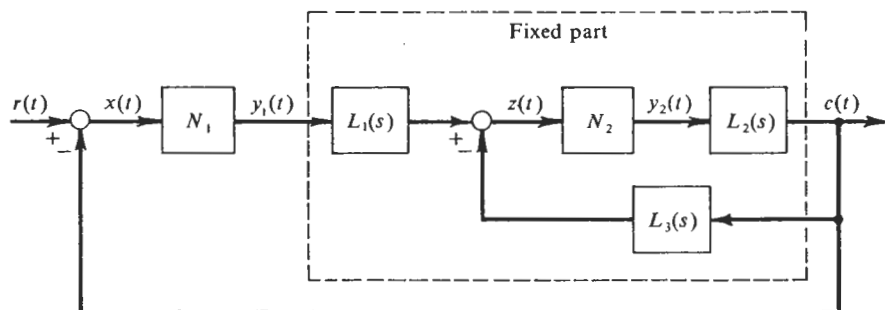


Figure 7.3-12 System configuration for design of optimum nonlinear compensation.

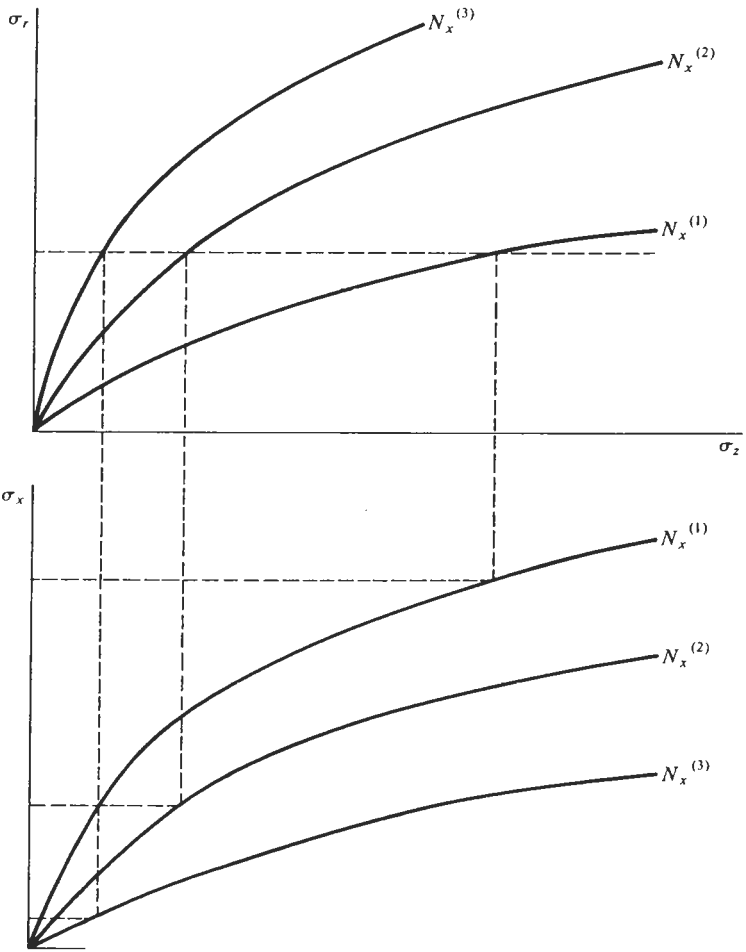
This may well imply a coarser degree of approximation than has been employed before in this section, since the output of the first nonlinearity is propagated to the input to the second without benefit of the heavy filtering of the controlled member. In some instances, $L_1(s)$ may be unity or even a lead function; so the distribution of $z(t)$ surely would not be normal, as assumed in the calculation of the RIDF for N_2 . The only justification for proceeding is that the RIDFs for many nonlinearities, and in particular for the limiter which is of common interest in this context, are not strongly dependent on the distribution of the input. With the understanding that the accuracy of the approximation for N_2 may not be up to the standard for RIDF usage, the solution proceeds in simple steps.

1. Choose a value for N_x , the RIDF for the unknown nonlinearity $y_1(x)$. Also choose a value for σ_z , and solve for the σ_r which results in this value of σ_z . This step is just like the "basic problem" discussed at the beginning of this section. Calculate the corresponding value of σ_x .
2. Repeat step 1 for enough values of N_x and σ_z to prepare plots of the form shown in Fig. 7.3-13.
3. At a given value of σ_r , use these plots to determine the value of σ_x corresponding to each choice of N_x . Plot these data for several values of σ_r as in Fig. 7.3-14.
4. For each input rms value there appears on these curves an optimum value of N_x (unless the solution is trivial). This is the value which minimizes the rms error. Cross-plot the optimum values of N_x against the corresponding values of σ_x , as shown in Fig. 7.3-14. This is the optimum RIDF for the nonlinearity $y_1(x)$.
5. Find a nonlinear function which has (nearly) the RIDF just determined.

In their illustrative example, Douce and King (Ref. 13) found that a nonlinearity of the form

$$y_1(x) = \begin{cases} ax^b & x \geq 0 \\ -y_1(-x) & x < 0 \end{cases}$$

with the proper choices of a and b , gave a suitable approximation to the required RIDF. The RIDF for this nonlinearity is given in Appendix E. Testing a simulation of the resulting system, Douce and King found the nonlinear compensator to reduce the system mean-squared error compared



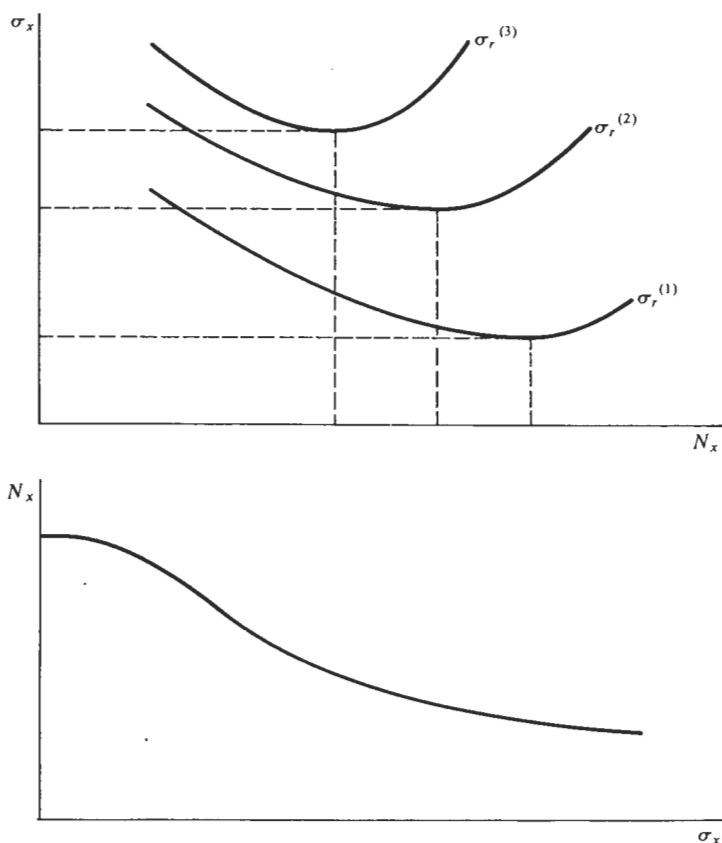


Figure 7.3-14 Results of steps 3 and 4 of solution procedure for optimum nonlinear compensation.

with the uncompensated nonlinear system by a factor which ranged up to 3:1 for some input conditions. Furthermore, the experimental results were in good agreement with the theoretical predictions, in spite of the fact that in their configuration there was no filter between N_1 and N_2 , as shown in Fig. 7.3-12.

7.4 FEEDBACK SYSTEMS WITH RANDOM AND OTHER INPUTS

We now extend our attention to nonlinear systems having sinusoids and biases propagating around the loop, in addition to random processes. The class of problems which now lies within the range of our consideration is

truly impressive, and although the details of problem solution become more laborious with each step of increasing generality, the theory of multiple-input describing functions still provides practical answers to important questions that pertain to the design of real systems. It is now essential that one have the gains of the nonlinearity to the various components of its input calculated in advance, and perhaps plotted for ease of use. The RIDFs for several important nonlinearities are given in Appendix E; the methods of Sec. 7.2 can be used to calculate such data for others.

With these data at hand, it is a relatively simple matter to determine the effect of a random dither on a system limit cycle, to design the dither to meet a specification on limit cycle amplitude, or to calculate the amount of dither required to quench the limit cycle altogether. One can determine the characteristics of a noise input which will excite a limit cycle in a system which does not display one in the absence of input. The bandwidth and peak resonance of the sinusoidal response characteristic of a nonlinear system are altered by the addition of noise. It is possible to eliminate a jump resonance characteristic with noise. These phenomena are predictable within the present context. Systems with asymmetric nonlinearities can be considered, as well as systems with biased responses to inputs. For example, the average and mean-squared steady-state following error can be calculated for a limit cycling system responding to a ramp command and a random disturbance. Many other combinations of circumstances will give rise to the basic situation being considered in this section: a signal at the input to the system nonlinearity which is well modeled by the sum of a bias, a sinusoid, and a gaussian process.

USE OF RANDOM DITHER

Consider, first, the effect of a random noise, either intentionally supplied as a dither or otherwise appearing in a system, on the limit cycle and sinusoidal response characteristics of a system of the form shown in Fig. 7.3-1. If this system is excited by a gaussian process only, any sinusoid appearing at $x(t)$ would have to be due to a limit cycle, and the performance of the system would be defined by the following relations:

$$N_A(A, \sigma_x) L_1 L_2 L_3(j\omega) = -1 \quad (7.4-1)$$

$$\sigma_x^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} H(s) H(-s) \Phi_{rr}(s) ds \quad (7.4-2)$$

$$H(s) = \frac{L_1(s)}{1 + N_R(A, \sigma_x) L_1 L_2 L_3(s)} \quad (7.4-3)$$

These describing functions, N_A and N_R , pertain to the given nonlinearity driven by an input $x(t)$, which is modeled as the sum of a sinusoid and a gaussian process. No bias is being considered at the moment, which

implies an unbiased input to the system and an odd nonlinearity. The limit cycle condition [Eq. (7.4-1)] can be simplified in view of the basic restriction which has been observed throughout this chapter; namely, only static single-valued nonlinearities have been under consideration. Under this restriction, all the describing functions are real-valued static gains; so it is known in advance that if a limit cycle is to exist, all the required phase shift must be supplied by the linear part of the system. The limit cycle frequency, if any, is then the frequency at which the sinusoidal response function for the open-loop linear part of the system has 180° of phase lag. Call this frequency ω_0 . The amplitude of the open-loop linear sinusoidal response function at this frequency is known, and the complex Eq. (7.4-1) is reduced to the real equation

$$N_A(A, \sigma_x) = \frac{1}{|L_1 L_2 L_3(j\omega_0)|} \quad (7.4-4)$$

Equations (7.4-2) and (7.4-4) are now to be solved simultaneously for A and σ_x if the input spectrum $\Phi_{rr}(s)$ is given. The solution defines the system limit cycle in the presence of the gaussian input and the input-output transfer characteristics of the limit cycling system to the gaussian input.

If the random input is applied as a dither for the purpose of controlling the amplitude of the limit cycle in the absence of other inputs, the dither may be applied either at the input or at some other station in the loop, such as directly at the input to the nonlinearity. Equations (7.4-4) and (7.4-2), with $\Phi_{rr}(s)$ interpreted as the power density spectrum for the dither input, remain the same regardless of the point at which the dither is injected. In every case $H(s)$ used in Eq. (7.4-2) is the closed-loop transfer function from the point where the dither is injected to the input to the nonlinearity, with the nonlinearity replaced by N_R . The *synthesis* problem of determining the amount of dither required to meet a specification on limit cycle amplitude is particularly simple to carry out. If A is specified and graphs of N_A and N_R as functions of A and σ are available, Eq. (7.4-4) can be solved immediately for σ_x . Then, with A and σ_x known, and if the shape of the power density spectrum of the random dither is chosen, Eq. (7.4-2) can be solved directly for the required rms value of the dither, no iterations or additional graphing being required. One caution must be observed, however. We shall see that nonlinear systems often display more than one mode of operation in their limit cycle behavior in the presence of random noise or dither. These modes correspond to multiple solutions to the equations defining the response. When a design is tentatively completed, the designer must check for the possibility of an additional solution; and if one exists, he must assure himself that the alternative limit cycle mode also meets the specification.

As an illustration of these procedures, Kwatny (Ref. 20) studied a system with a third-order linear part and an ideal-relay nonlinearity. The configuration is shown in Fig. 7.4-1. The input $r(t)$ was a gaussian random

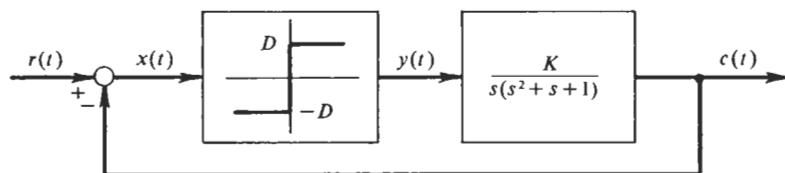


Figure 7.4-1 Kwatny's example system. (Ref. 20.)

process with power density spectrum

$$\Phi_{rr}(s) = \frac{2\tau\sigma_r^2}{1 - \tau^2 s^2} \quad (7.4-5)$$

The RIDFs for the ideal relay with an input consisting of the sum of a gaussian process and a sinusoid are shown abbreviated from Appendix E in Fig. 7.4-2. We shall present the results for the case $\sigma_r = 5$ units and $\tau = 1$ sec with the gain of the system, KD , a parameter. For small values of this parameter no solution of Eq. (7.4-4) is possible. That is, for values of σ_x consistent with the given input and the low values of KD , there is no value of A for which $N_A(A, \sigma_x)$ is large enough to satisfy Eq. (7.4-4). Over a range of gains, then, this system is predicted to display no limit cycle, although it is clear that it would limit-cycle for any gain if the noise were not present. The transfer of noise through the system in the absence of a limit cycle is described by Eq. (7.4-2), with N_R given by the $A/\sigma = 0$ point in Fig. 7.4-2. For values of KD large enough to yield a solution to the limit cycle condition [Eq. (7.4-4)], there are in fact two solutions. Thus one predicts two possible modes of response: one with a relatively small A and large σ_x , the other with a relatively large A and small σ_x .

The theoretically predicted performance of this system is summarized in Fig. 7.4-3, which is a plot of the rms output of the system against system gain, KD , each normalized by the rms value of the input random process, σ_r . The mean-squared output, which is the only quantity that can conveniently be checked experimentally, includes both the transfer of noise to the output and the limit cycle at that point.

$$\overline{c(t)^2} = \frac{1}{2}A^2 + \text{mean-squared noise at } c$$

In the lower range of KD , there is just one mode of response with no limit cycle. The rms output is due entirely to the transfer of the input noise to $c(t)$. In the higher range of KD , the two solutions are shown with the limit cycle amplitude indicated along the two branches of the curve.

The actual performance of this system as observed in an analog simulation is entirely consistent with these predictions. Experimentally observed values of rms output are also shown on Fig. 7.4-3. In the region where no limit

cycle is predicted, the experimental and theoretical values of rms output differ by less than 10 percent. In the region where a limit cycle and two-mode operation are predicted, the observed rms output lies between the two theoretical values, suggesting an averaging between the two modes of operation. In his study of the actual output time histories, Kwatny (Ref. 20) observed the tendency to shift in a random manner between modes which would approximate those predicted. At the high end of the range of system gain, the tendency toward the larger-amplitude limit cycle mode was more pronounced.

In the case of a non-limit-cycling system, the gaussian-plus-sinusoid-input describing functions are useful in the study of the sinusoidal response characteristics of a nonlinear system with random noise or dither, or the random signal response characteristics of a nonlinear system with sinusoidal noise or dither. The noise transfer expression [Eq. (7.4-2)] is still applicable in this case, but the limit cycle condition [Eq. (7.4-4)] is replaced by the expression for the magnitude of the sinusoid response function of the closed-loop quasi-linearized system from system input to nonlinearity input.

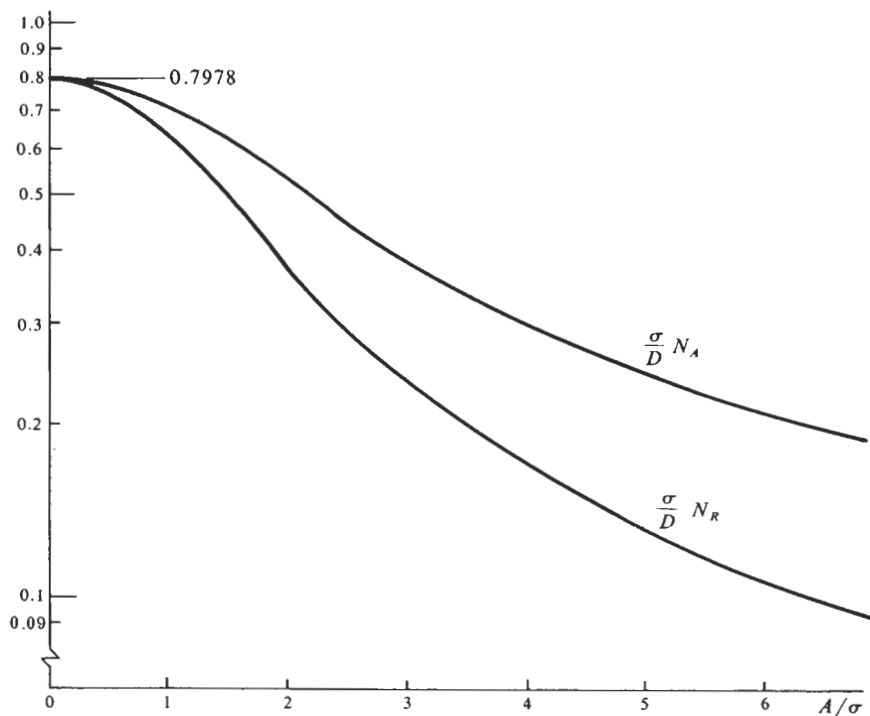


Figure 7.4-2 Gaussian-plus-sinusoid-input describing functions for the ideal-relay nonlinearity.

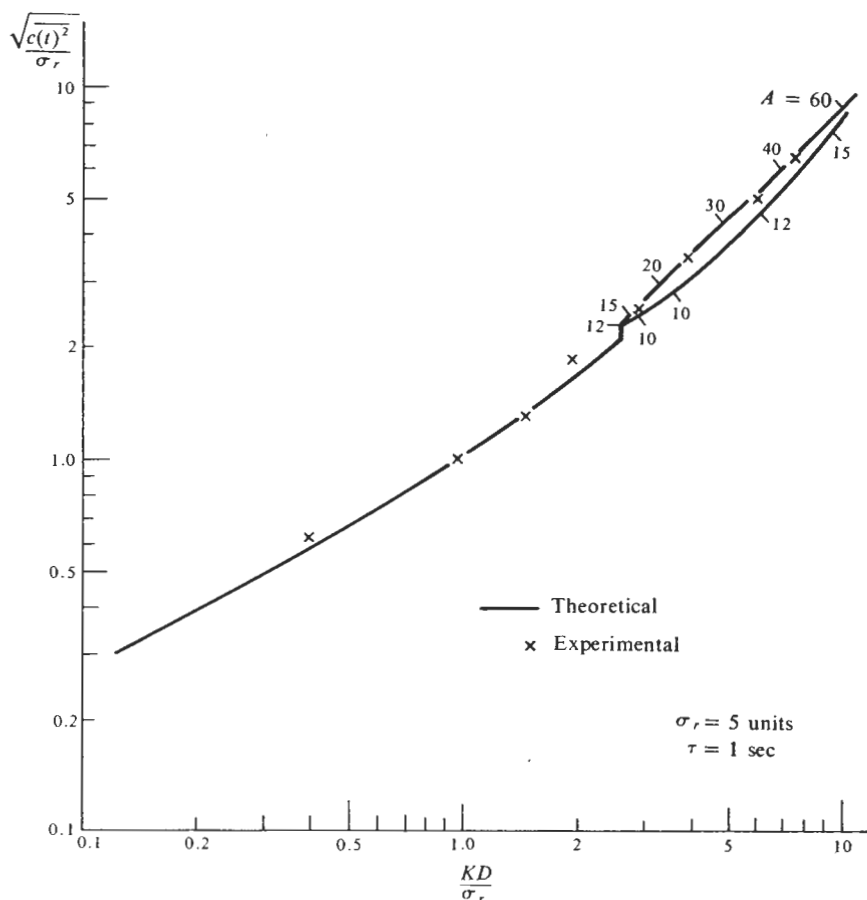


Figure 7.4-3 Performance of Kwatny's example system. (Adapted from Ref. 20.)

Nikiforuk and West (Ref. 31) show experimental results for the former in the case of a second-order system with a limiter nonlinearity. Their frequency response curves for different noise bandwidths or rms values show shifts in peak magnitude and in the frequency for peak magnitude which are completely understandable in terms of the decrease in the gain of the limiter to the sinusoid, with an increase in the rms noise at the input to the limiter. They also note that noise or random dither tended to quench the jump resonance in the frequency response characteristic. As the noise power was increased, the upper jump frequency tended to move down toward the lower until the jump phenomenon was eliminated.

These system characteristics, which are well explained by two-input RIDF theory, are difficult to think about qualitatively because of the complex

interrelation among N_A , N_R , A , and σ_x . The situation is considerably simplified if either A or σ_x can reasonably be treated as constant. This can be done in cases where the input sinusoid is very high frequency or the input noise is very wideband compared with the bandwidth of the quasi-linearized system. In such cases the system acts open-loop with respect to the high-frequency signal, since only a trivial amount is fed back. For example, if the input noise is wideband, σ_x can be determined once and for all, and the gain of the nonlinearity to the sinusoid is given by the curve of N_A versus A for the appropriate value of σ . From that point on, the problem is just like a single-sinusoid-input problem, but with a DF modified by the presence of the noise. The effect of the noise is invariably to "smooth out" the N_A curve, which is the qualitative explanation for the observed quenching of a jump resonance. The noise can either increase or decrease the gain of the nonlinearity to the sinusoid, depending on whether the slope of the nonlinearity generally increases or decreases with increasing input. For example, the presence of noise increases N_A in the case of a polynomial or hard-spring nonlinearity; it decreases N_A in the case of a soft-spring, limiter, or ideal-relay nonlinearity. Other nonlinearities have a combination of these effects. The relay with dead zone, for example, has a small slope (zero) for small signals, then a large slope (infinite), and then a small slope (zero) for large signals. Noise in this case increases N_A for A smaller than the dead zone, and decreases N_A for A much larger than the dead zone. Furthermore, the largest increase in N_A for small A is realized by an intermediate amount of noise such that σ_x is comparable with the dead zone.

Oldenburger and Sridhar (Refs. 33, 43) neglected the fed-back noise in their studies of random-signal stabilization. In the context of a more complete theory of quasi-linearization, this point of view is treated as an approximation which, when applicable, affords a significant simplification in problem solution. In view of the fact that the presence of noise can either increase or decrease the gain of the nonlinearity to a sinusoid, depending on the shape of the nonlinearity, it is easy to construct examples of systems in which noise will reduce or quench the limit cycle that would exist in the absence of input, and other examples in which noise will excite a limit cycle in a system which does not limit-cycle in the absence of input. The effect of the noise is very much like the effect of a second sinusoid in the system, and these examples are similar to the examples of sinusoidally quenched or excited limit cycles discussed in Sec. 5.4.

APPLICABILITY OF THE SINGLE-RANDOM-INPUT DESCRIBING FUNCTION

Use of the single-random-input describing function to study the response of a nonlinear system to a random input is always made under the same condition

as that required for use of the single-sinusoid-input describing function to study the response of a nonlinear system to a sinusoidal input, namely, that the system does not limit-cycle in the presence of the input. The two-sinusoid-input describing function is required to test for a limit cycle in the presence of an input sinusoid, and the gaussian-plus-sinusoid-input describing function is required to test for a limit cycle in the presence of a random input. The condition for the applicability of the single-random-input describing function was stated previously in the discussion of the use of dither: no limit cycle is indicated if for the σ_x corresponding to the given input and system configuration, there is no value of A for which $N_A(A, \sigma_x)$ is as large as the reciprocal of the magnitude of the transfer function for the open-loop linear part at the frequency for 180° phase lag. This simply says that there is no solution to Eq. (7.4-4). This condition is simply stated; the difficulty in applying it is the determination of σ_x . A number of nonlinearities have the property that the maximum value of $N_A(A, \sigma)$ as a function of A for a given σ occurs at $A = 0$. This is true, for example, of the ideal relay (see Fig. 7.4-2), the limiter, and soft-spring characteristics generally. In this case, the following point of view constitutes a most appealing pitfall: there can be no limit cycle if $N_A(A, \sigma_x)$ for $A = 0$ is less than $1/|L(j\omega_0)|$, where $L(s)$ is the transfer function for the linear part of the system. But for $A = 0$, σ_x can be determined using the simpler single-random-input describing function. Furthermore, it was shown previously [see Eq. (1.5-66)] that

$$\lim_{A \rightarrow 0} N_A(A, \sigma) = N_R(0, \sigma)$$

so the single-random-input describing function would seem to provide a check on its own applicability. The indicated procedure is to calculate σ_x using the single-random-input describing function; note that the corresponding $N_R(\sigma)$ is equal to $N_A(0, \sigma)$, and see if this value is less than $1/|L(j\omega_0)|$. The more complicated gaussian-plus-sinusoid-input describing function need not be calculated at all, provided one is sure in advance that for any given σ , $N_A(A, \sigma)$ is never greater than $N_A(0, \sigma)$.

This argument, specialized to particular cases, was given by Kwatny (Ref. 20) and Gibson and Sridhar (Ref. 14), and in each case led to obviously incorrect results. The basic difficulty is the presence of *multiple solutions*; the user of describing function theory must always be alert to this possibility. The nature of the problem can be illustrated by reference to the system of Fig. 7.4-1. Suppose, first, that the rms input σ_r is taken as the variable parameter, and we wish to find the range of σ_r for which no limit cycle exists. $1/|L(j\omega_0)|$ for this system is $1/K$ and, for a specified K , is a known constant. The single-random-input describing function for the ideal relay is given by Eq. (7.2-31) with $B = 0$.

$$N_R(\sigma_x) = \sqrt{\frac{2}{\pi}} \frac{D}{\sigma_x} \quad (7.4-6)$$

According to the argument given above, we are looking for the range of σ_r which yields values of σ_x such that $N_R(\sigma_x) < 1/K$. But this is exactly the condition for stability of the system, with the nonlinearity replaced by the gain N_R . This interpretation of the condition for no limit cycle in terms of the stability of the quasi-linearized random-input-only system is not unique to this example; it is true of any system which is stable for all gains less than a critical value. However, all values of σ_r from zero to infinity yield stable systems in this example if random-noise-only theory is applied. This can be seen by noting that for large values of σ_r , the system cannot follow the input and σ_x tends toward σ_r . As σ_r is decreased, σ_x decreases, and thus N_R increases. As N_R approaches the value which would render the system unstable, corresponding to a known finite value of σ_x , the transfer of noise from $r(t)$ to $x(t)$ approaches infinity. Thus N_R approaching the critical value for stability corresponds to σ_r approaching zero. All positive values of σ_r , then, even vanishingly small values, yield values of N_R in the stable range, and the simplified limit cycle condition would say that no limit cycle exists since $N_A(0, \sigma_x)$, which is equal to N_R , is smaller than $1/|L(j\omega_0)|$. This does indeed constitute a solution to the mathematical problem, but it does violence to the physical problem. We know the no-input system limit cycles, and the limit cycle must persist for some range of small values of σ_r . If the theory is to be right, there must be another solution (or solutions) to the mathematical problem for small values of σ_r . The nature of the alternative solution can be seen by referring to the gaussian-plus-sinusoid-input describing functions of Fig. 7.4-2. If one admits the possibility of a limit cycle with $A > 0$, the effect of the sinusoid on N_R is to decrease it. For values of N_R close to the critical value for stability, a decrease in N_R means a decrease in σ_x . But the decrease in σ_x increases N_A to a value large enough to support the limit cycle. This alternative solution with $A > 0$ defines the mode of operation we should expect to see in practice, and the solution can be found only by admitting the possibility of a limit cycle to the analysis.

A comparable situation exists if the input is fixed and the system gain KD is the variable parameter. If noise-only theory is used, all values of KD from zero to infinity are found to yield stable quasi-linearized systems, and in every case $N_A(0, \sigma_x)$ is less than $1/|L(j\omega_0)|$. The nature of the solution as $KD \rightarrow \infty$ is that σ_x goes to infinity (and thus N_R goes to zero) with KD in such a way that the quasi-linearized system just approaches the point of instability. However, for values of KD above one value, there are not only one, but two, additional solutions with nonzero values of A . As noted earlier, the actual system displays both of these modes.

In summary, there seems to be no reliable shortcut to the determination of the validity of random-input-only describing function theory. Except in the case of low-ordered systems which obviously cannot limit-cycle, the possibility of a limit cycle can be tested only by including a sinusoid in the analysis, and this requires full use of multiple-input describing function theory.

A THREE-INPUT ILLUSTRATION

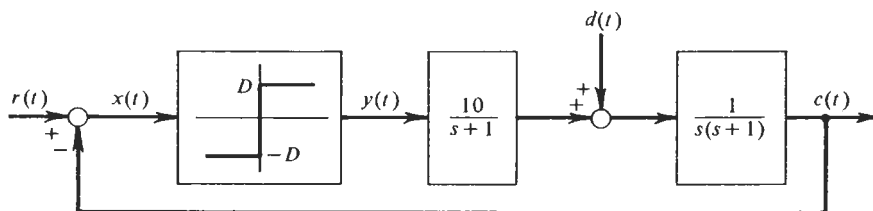
The presence of all three of the signal forms considered here at the input to the nonlinearity does not alter the nature of the analysis, only the labor required to apply it. The original nonlinear system is replaced by three quasi-linear systems, one to process each of the signal forms. In each of these quasi-linear systems, the nonlinearity is represented by the corresponding RIDF: $N_B(A, B, \sigma)$, $N_A(A, B, \sigma)$, and $N_R(A, B, \sigma)$. The nonlinear character of the system is represented in the interdependence of these effective gains. The three quasi-linear systems yield three relations, one each for A , B , and σ , in terms of the original system configuration and its inputs. These three expressions and the expressions for N_B , N_A , and N_R , which will usually be represented by graphed data, must be solved simultaneously.

As an illustration of the recommended procedure, consider the system of Fig. 7.4-4. This system is responding to a ramp input $r(t)$ and a gaussian disturbance $d(t)$. We wish to determine the mean value and standard deviation of the error in the steady state. The steady-state error response to the ramp is a constant due to the integration in the plant. In addition, it is clear that, for small enough disturbance inputs, the system will limit-cycle. Thus $x(t)$ must be modeled as the sum of a bias, a sinusoid, and a random signal, and the three-input RIDFs must be employed. A sample of this nonlinearity input, $x(t)$, for particular choices of system parameters and input characteristics, is shown in Fig. 7.4-5. The bias in this signal is clearly evident, as is the limit cycle, and the irregularity of the record indicates the presence of the random noise.

The quasi-linear system which processes the bias signal has the same configuration as the original system, but the nonlinearity is replaced by $N_B(A, B, \sigma)$, and the only input is the ramp function $r(t)$. Because of the integration in the plant, the steady-state response of this quasi-linear system is evidently

$$10BN_B = k \quad (7.4-7)$$

The quasi-linear system which processes the sinusoid has the same configuration as the original system, but the nonlinearity is replaced by $N_A(A, B, \sigma)$,



$$r(t) = kt$$

$$\Phi_{dd}(s) = \frac{2\tau\sigma_d^2}{1 - \tau^2 s^2}$$

Figure 7.4-4 Three-input example system.

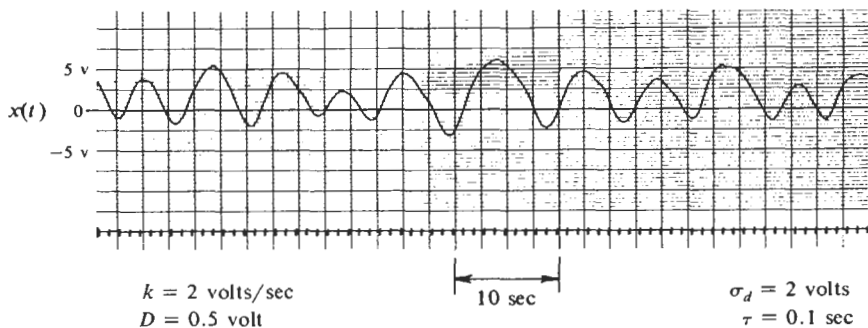


Figure 7.4-5 Sample of steady-state response. System of Fig. 7.4-4.

and there is no input. If this quasi-linear system is to display an undamped oscillatory mode corresponding to a limit cycle of the nonlinear system, the gain of the open-loop transfer function at the frequency for which it has 180° of phase lag must be unity. That frequency in this case is 1 radian/sec, and the limit cycle condition is

$$5N_A = 1 \quad (7.4-8)$$

The quasi-linear system which processes the random disturbance also has the same configuration as the original system, but the nonlinearity is replaced by $N_R(A, B, \sigma)$, and the only input is the gaussian disturbance $d(t)$. The σ which the three RIDFs depend on is the standard deviation of the random component of $x(t)$, the nonlinearity input. It is given by

$$\sigma^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{2\tau\sigma_d^2}{1 - \tau^2 s^2} H(s)H(-s) ds \quad (7.4-9)$$

$$H(s) = -\frac{s+1}{s^3 + 2s^2 + s + 10N_R} \quad (7.4-10)$$

This integral can be evaluated using Table 7.3-1 to give an explicit expression for σ^2 in terms of N_R and σ_d^2 if τ is given. This expression is

$$\sigma^2 = \sigma_d^2 \left[\frac{\frac{10}{\tau} \left(2 + \frac{1}{\tau}\right) N_R + \left(2 + \frac{1}{\tau}\right) \left(1 + \frac{2}{\tau}\right) - \left(10N_R + \frac{1}{\tau}\right)}{10N_R \left[\left(10N_R + \frac{1}{\tau}\right) \left(1 + \frac{2}{\tau}\right) \left(2 + \frac{1}{\tau}\right) - \frac{10}{\tau} N_R \left(2 + \frac{1}{\tau}\right)^2 - \left(10N_R + \frac{1}{\tau}\right)^2 \right]} \right] \quad (7.4-11)$$

The solution is most readily carried out by assuming a value for σ . Then, for each of the values of B/σ for which the describing functions are plotted in Appendix E, the value of A/σ which gives $N_A = 0.2$ can be read from the graph, as well as the value of A/σ which gives $BN_B = k/10$. These sets of values are plotted versus B/σ , and the intersection indicates the values of A and B which result in satisfaction of Eqs. (7.4-7) and (7.4-8) for the chosen value of σ . This defines N_R , and Eq. (7.4-11) is solved for the value of σ_d^2 which would result in the assumed value of σ . This procedure is repeated for several choices of σ to define the response of the system to a range of values of σ_d^2 . One can then interpolate among these data to define the response for a given value of σ_d^2 .

For some values of σ and B there will be no value of A for which N_A is as large as 0.2. This indicates no possibility of a limit cycle with those values of σ and B . For other choices of σ and B there may be more than one value of A for which the limit cycle condition can be satisfied. This indicates the possibility of multiple solutions, although one branch of such solutions may not be consistent with the requirement on gain to the bias. Another indication of multiple solutions can appear at the last step, where the relation between σ and σ_d may show more than one value of σ , and hence of A and B , corresponding to a single value of σ_d .

Having found A , B , and σ for given ramp and disturbance inputs, the mean error is

$$\overline{x(t)} = B \quad (7.4-12)$$

and the standard deviation of the error is

$$\sigma_x = \sqrt{\sigma^2 + \frac{1}{2}A^2} \quad (7.4-13)$$

If multiple solutions are indicated, experience has shown that the actual system operation usually shifts randomly among the possible modes, and measured statistics represent some kind of average among the values of these statistics in the different modes.

GENERALIZATION OF THE BIAS FUNCTION

The utility of multiple-input RIDFs is greatly broadened if one generalizes the bias which has been included as one of the characteristic forms of function at the input to the nonlinearity, allowing it to become an arbitrary function of time, within certain restrictions.

$$B \rightarrow B(t)$$

Within the necessary restrictions which must be placed on this extension of concept, one gains the tremendous advantage of being freed from the necessity of specifying in advance the form of each of the components of

signal at the nonlinearity input. Thus one can study, in a restricted sense to be sure, system stability, general transient responses, and other response characteristics that do not give rise to signal forms which can be known in advance.

The conditions under which the gain of a nonlinearity to a bias function, B , can reasonably be interpreted as the gain of that nonlinearity to an arbitrary function, $B(t)$, were discussed in Chap. 6 for the case in which B , or $B(t)$, appears at the nonlinearity input, together with a sinusoid. Our present consideration of this subject differs only in that random signals, as well, may be present. Although the presence of a random signal does not change the basic nature of this problem, it does complicate the writing of quantitative conditions under which the generalized interpretation of B is accurate. Qualitatively, one may say that if the nonlinearity input is observed over an interval long enough to include one period of the sinusoid and the interval of significant correlation of the random signal, the function $B(t)$ should appear essentially constant. But more than this is involved. The gains of the nonlinearity to the sinusoid and to the random signal depend on B ; so if B changes, the system sinusoidal and random-signal response characteristics change. But we propose still to employ steady-state sinusoidal response characteristics and stationary (that is, steady-state) random response characteristics for the quasi-linearized system. So $B(t)$ should also change slowly enough so that the system can be thought of as progressing through a continuous sequence of steady states as regards its sinusoidal and random response.

The generalization of B to $B(t)$ can be justified, not only in the case where $B(t)$ changes slowly enough, but also in the case where $B(t)$ is small enough. For small enough $B(t)$, N_A and N_R are little influenced by the changing B , and the system displays its steady-state sinusoidal and random responses. Moreover, we note from Eq. (1.5-63) that

$$\lim_{B \rightarrow 0} N_B = N_R(B = 0)$$

and from Eq. (1.5-66) that

$$\lim_{A \rightarrow 0} N_A = N_R(A = 0)$$

This says that the gain of a nonlinearity to a bias signal in the presence of any assortment of other signals is the same as the gain of that nonlinearity to a sinusoid in the presence of the other signals if both the bias and the sinusoid are small enough; and in fact this gain is the same as the gain of the nonlinearity to the random component of its input, which may be large or small. This suggests the thought that *the gain of the nonlinearity to a small signal in the presence of other signals is the same regardless of the shape of the small*

signal, and that this gain is the gain of the nonlinearity to the random component of input in the absence of the small signal. Just how small $B(t)$ must be to qualify on the grounds of "smallness" rather than "slowness" cannot be stated in general; it depends on the nonlinearity. For those nonlinearities whose RIDFs are documented in Appendix E, the reader can see the range of B for which N_A and N_R are essentially independent of B , and N_B is nearly equal to N_R , and the range of A for which N_A is essentially equal to N_R .

The gain of the nonlinearity to small signals $B(t)$ is just what is needed to study the stability of a system for small signals in the presence of others. The interesting kind of behavior which this allows one to test for is comparable with the case indicated in Chap. 6 of a system which has a stable limit cycle mode, but in the presence of that limit cycle the system is unstable for small signals. In this case we can consider a system which is responding to a random input and is also circulating one or more sinusoids which may result either from inputs or a limit cycle or both. In the presence of these signals, we wish to test the stability of the system to small signals. This test has a particularly simple interpretation when a random signal is present. For clarity, let us consider just one sinusoid. The stability of the system to small signals is given by the stability of the quasi-linearized system in which the nonlinearity is replaced by $N_B(A, B, \sigma)$ with $B = 0$. The appropriate values of A and σ must be determined using the quasi-linearized system which propagates the sinusoid, in which the nonlinearity is replaced by $N_A(A, B, \sigma)$ with $B = 0$, and the quasi-linearized system which propagates the random signal, in which the nonlinearity is replaced by $N_R(A, B, \sigma)$ with $B = 0$. But once the simultaneous relations which define A and σ are solved, the stability to small signals is evident since

$$\lim_{B \rightarrow 0} N_B(A, B, \sigma) = N_R(A, 0, \sigma)$$

So if the quasi-linearized system which propagates the random signal is stable, the system is stable for arbitrary small signals in the presence of the random signal and sinusoid.

Popov (Ref. 35) solved this problem, taking a somewhat more complicated point of view. He considered a system with a bias and a random process and used the same gains as the describing functions $N_B(B, \sigma)$ and $N_R(B, \sigma)$ defined here to model the effect of the nonlinearity in transferring the bias and the random process. For the given random input to the system and for every arbitrary choice of B , the rms value of the random process at the nonlinearity input can be determined, using standard quasi-linear techniques. This gives a value of σ corresponding to every choice of B . Using this $\sigma(B)$, the gain to the bias function, $N_B(B, \sigma)$, is determined simply as a function of B . Now one can study a new nonlinear system in which the input to the

nonlinearity is B generalized to $B(t)$ and the output is BN_B . This system has no random process appearing explicitly, but the effect of the random input to the original system is accounted for in the modification of the original nonlinearity to the new nonlinearity. For the study of the stability of the system in the presence of the random input, Popov uses small-signal linearization of the new nonlinearity about its origin. That is, he replaces the nonlinearity by a gain which is the slope of BN_B with respect to B at $B = 0$. But we now know, from Eq. (7.2-96), that

$$\frac{d}{dB}(BN_B) = N_R$$

so the gain derived in this way is exactly $N_R(\sigma)$ for the value of σ corresponding to $B = 0$! This procedure for the study of stability is then equivalent to the more direct approach described above.

A procedure similar to this can be employed to treat more complicated problems which depend on the assumption that the gain of the nonlinearity to a small or slowly varying signal $B(t)$ is equal to the gain of the nonlinearity to a bias signal B in the presence of the same additional signals. Smith (Ref. 40) gives an example of the calculation of the transient response of a nonlinear system to a ramp input which is corrupted by an additive noise. The first part of the procedure is the determination of the equivalent nonlinearity for a noise-free system in which the effect of the noise in the original system is accounted for in the transformation of the original nonlinearity to the new nonlinearity. This is done as described above by considering the output of the new nonlinearity to be BN_B , in which the dependence upon σ has been eliminated by use of $N_R(B, \sigma)$ to solve for the value of σ corresponding to each value of B . Having a new nonlinear system subjected to a noise-free ramp input, one can solve for the response in what seems to be the most convenient way. This is the most difficult step, since it may not admit of solution through quasi-linearization. We might note that if the problem were such that harmonic linearization of the new nonlinearity would lead to a solution, the result of the harmonic linearization of the new nonlinearity $BN_B(B)$ would be exactly $N_A(A, \sigma)$, which might better have been calculated in the first place. Smith completed his example by treating the new nonlinearity as piecewise-linear in two ranges. This gives the mean response of the system to the noisy ramp input. This mean response has the interpretation that the response at every time is the mean, or average, of all possible responses at that time. It is thus the time-varying mean over the ensemble of responses to all possible inputs, each input consisting of the sum of the deterministic ramp function and a member of the ensemble of all possible noise functions. An estimate of probable deviations from this mean response can now be calculated by determining the noise transferred to the point where the response is measured, using $N_R(B, \sigma)$ and the value of σ corresponding to each value of B in the mean response. If the response is being measured at

the input to the nonlinearity, this step is trivial, since the rms value of the noise at that point is already known as a function of B .

The system treated by Smith in his example is shown in Fig. 7.4-6. The mean response can be measured only by repeating the solution many times with random noise at the input and averaging over these responses at each point in time. Smith averaged over 25 recorded responses to determine the mean error response to the noisy ramp input. The results are shown in Fig. 7.4-7. Plotted in this figure are the theoretically computed mean error, the measured mean error at 0.1-sec intervals, and the 50 percent probable band for the distribution of the measured mean. Considering the fact that this calculation depends upon the approximations ordinarily involved in the application of describing function theory, and in addition uses stationary theory to process obviously nonstationary statistics, and uses the gain of the nonlinearity to the constant B as its gain to the slowly varying mean signal, the agreement seems quite good.

This use of RIDF analysis has little justification in theory. However, it does provide a means of attack on a very complicated class of problems—an attack which is both workable and somewhat satisfying physically.

7.5 ALTERNATE APPROACH FOR NONLINEARITIES WITH MEMORY

Throughout this chapter, attention has been restricted to static single-valued nonlinearities. This is not due to any limitation in the theory of multiple-input describing functions. The general describing function theory developed in Chap. 1 is valid for dynamic and multiple-valued nonlinearities as well, and in fact, application of the theory poses no great hurdle if nonlinearity inputs consisting of a bias and sinusoids only are considered. The difficulty with random inputs to a memory-type nonlinearity is purely a practical one: the optimum quasi-linear filter to pass the random input component is a dynamic filter in this case, and its determination requires solution of an

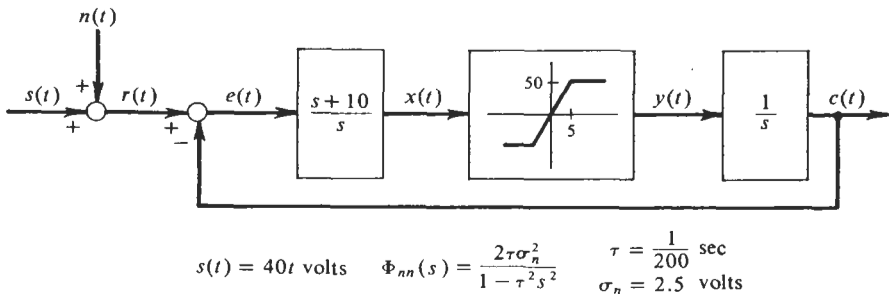


Figure 7.4-6 Smith's example system. (Ref. 40.)

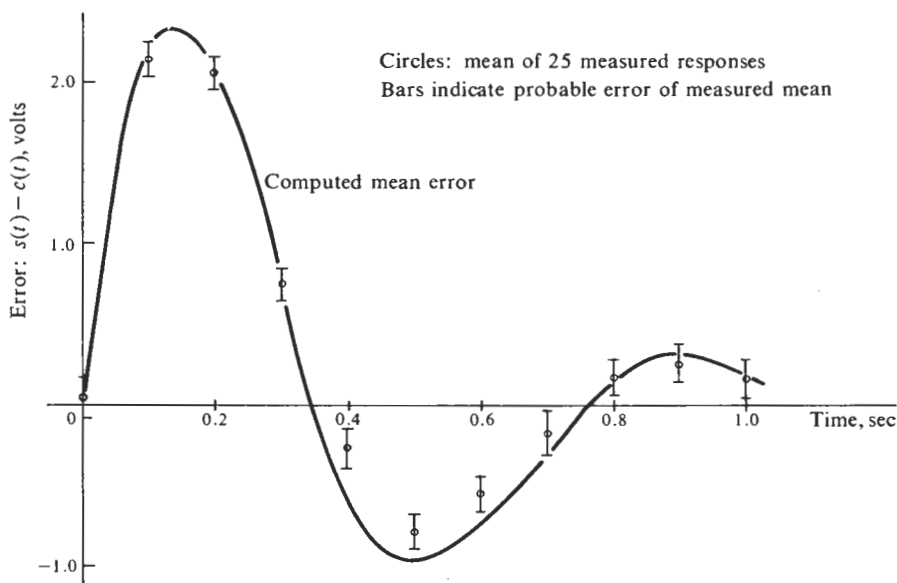


Figure 7.4-7 Results of Smith's example. (Ref. 40.)

integral equation of the Wiener-Hopf form [Eq. (1.5-44)]. The statistics which must be provided to effect the solution are the autocorrelation function for the random input and the cross correlation between the present output of the nonlinearity and all prior values of the random input. Just the evaluation of these statistics seems a forbidding chore, to say nothing of solving the integral equation. All this was found to simplify greatly in the case of static single-valued nonlinearities.

An entirely different approach to the determination of quasi-linear approximators to dynamic or multiple-valued nonlinearities with random inputs may be preferable. Because of the difficulty of calculating the distribution of random signals in dynamic nonlinear systems, we are constrained, as a practical matter, to assume the random signal appearing at the input to the nonlinearity to be gaussian. This assumption depends on the low-pass-filtering properties of the linear part of the system. But any gaussian random process can be expressed in the form of sinusoids: either as a sinusoidal function with random amplitude and phase or as the limit of a Fourier series with random coefficients (Ref. 38). With this description of the random input process, one can propagate the sinusoids representing the gaussian input, together with the actual sinusoidal input components and bias, through the nonlinearity, and identify the output components which are correlated with the input components. These relations, when averaged over the distributions of the random variables associated with the representation

of the input gaussian process, define the quasi-linear approximator based on the property of cross-correlation equivalence. These techniques are discussed by Levadi and Cosgriff (Ref. 24) and Mahalanabis and Nath (Ref. 25).

It may well be that the labor involved in quasi-linear approximation of dynamic and implicit nonlinearities enclosed in feedback systems renders the technique of little value in this case. The whole virtue of describing function analysis, as opposed, for example, to computer simulation of the system under study, is the understanding one gains of the system performance characteristics and the basis it provides for the rational design of systems. When the labor required to apply the technique exceeds some reasonable bound—and this bound depends strongly on each individual's facility in the technique—one's understanding of the system tends to be obscured, and the value of the analytic procedure in system design is decreased. This simply underscores the fact that we are dealing with very complicated problems. The impressive thing is the breadth of the range of complex and practically significant system situations in which quasi-linear techniques can be applied with reasonable facility.

REFERENCES

1. Atherton, D. P.: The Evaluation of the Response of Single-valued Non-linearities to Several Inputs, *Proc. IEE*, vol. C-109, no. 15 (March, 1962), pp. 146-157.
2. Atherton, D. P.: Rapid Evaluation of the Auto-correlation Function of the Output of Single-valued Non-linearities in Response to Sinusoidal and Gaussian Signals, *Proc. IEE*, vol. C-109, no. 16 (September, 1962), pp. 656-664.
3. Axelby, G. S.: Random Noise with Bias Signals in Nonlinear Devices, *IRE Trans. Autom. Control*, vol. AC-4, no. 2 (November, 1959), pp. 167-172.
4. Barrett, J. F., and J. F. Coales: An Introduction to the Analysis of Non-linear Control Systems with Random Inputs, *Proc. IEE*, vol. C-103, no. 3 (November, 1955), pp. 190-199.
5. Booton, R. C., Jr.: Nonlinear Control Systems with Random Inputs, *IRE Trans. Circuit Theory*, vol. CT-1, no. 1 (March, 1954), pp. 9-17.
6. Booton, R. C., Jr., M. V. Mathews, and W. W. Seifert: Nonlinear Servomechanisms with Random Inputs, *MIT DACL Rept.* 70, August, 1953.
7. Burington, R. S.: "Handbook of Mathematical Tables and Formulas," 4th ed., McGraw-Hill Book Company, New York, 1964.
8. Caron, J. Y.: Errors in the Quasi-linearization Technique for Nonlinear Systems, M.S. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, Cambridge, Mass., 1955.
9. Chuang, K., and L. F. Kazda: A Study of Nonlinear Systems with Random Inputs, *AIEE Trans.*, pt. II, vol. 78, no. 42 (May, 1959), pp. 100-105.
10. Culver, W. J., and M. D. Mesarovic: Dynamic Statistical Linearization, Joint Automatic Control Conference, paper 14-5, New York University, New York, 1962.
11. Davenport, W. B., Jr., and W. L. Root: "Introduction to Random Signals and Noise," McGraw-Hill Book Company, New York, 1958.
12. Douce, J. L.: A Note on the Evaluation of the Response of a Non-linear Element to

- Sinusoidal and Random Signals, *Proc. IEE*, vol. C-105, no. 7 (March, 1958), pp. 88-92.
13. Douce, J. L., and R. E. King: The Effect of an Additional Non-linearity on the Performance of Torque-limited Control Systems Subjected to Random Inputs, *Proc. IEE*, vol. C-107, no. 12 (September, 1960), pp. 190-197.
 14. Gibson, J. E., and R. Sridhar: The Response of Nonlinear Closed Loop Systems to Random Inputs, *Trans. ASME, J. Basic Eng.*, vol. D-86, no. 1 (March, 1964), pp. 132-138.
 15. Gusev, M. I.: Determining the Effect of Regular Signal Dynamics in the Statistical Linearization Method, *Automation and Remote Control*, vol. 21, no. 11 (May, 1961), pp. 1093-1100.
 16. Katkovnik, V. Y., and A. A. Pervozvanskii: Relay System with Self-oscillating Mode of Operation Disturbed by Random Signals, *Automation and Remote Control*, vol. 22, no. 5 (November, 1961), pp. 517-523.
 17. Kazakov, I. Y.: Approximate Probabilistic Analysis of the Operational Accuracy of Essentially Nonlinear Automatic Systems, *Automation and Remote Control*, vol. 17, no. 5 (1956), pp. 385-409.
 18. Kazakov, I. Y.: Problems of the Theory of Statistical Linearization and Its Applications, *Preprints of IFAC 1960 Moscow Confer.*, 1960, pp. 1968-1972, Butterworth Scientific Publications, London.
 19. Kislov, B. D.: Reduced Equivalent Amplification Factor of a Nonlinear Element in the Presence of Noise, *Automation and Remote Control*, vol. 21, no. 8 (March, 1961), pp. 804-809.
 20. Kwatny, H. G.: An Investigation of the Limit Cycle Characteristics of Nonlinear Systems in the Presence of a Random Noise, M.S. thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, May, 1962.
 21. Laning, J. H., Jr., and R. H. Battin: "Random Processes in Automatic Control," McGraw-Hill Book Company, New York, 1956.
 22. Lee, Y. W.: "Statistical Theory of Communication," John Wiley & Sons, Inc., New York, 1960.
 23. Leland, H. R.: Input-Output Cross-correlation Functions for Some Memory-type Nonlinear Systems with Gaussian Inputs, *AIEE Trans.*, pt. II, vol. 79, no. 49 (July, 1960), pp. 219-223.
 24. Levadi, V. S., and R. L. Cosgriff: A Describing Function for Nonlinear Systems with Memory Subject to Random Input, *IEEE Trans.*, pt. II, *Appl. Ind.*, March, 1964, no. 71, pp. 73-76.
 25. Mahalanabis, A. K., and A. K. Nath: On the DIDFs of a Nonlinear Element with Memory, *IEEE Trans. Autom. Control*, vol. AC-11, no. 1 (January, 1966), pp. 138-139.
 26. Merchav, S. J.: Equivalent Gain of Single-valued Nonlinearities with Multiple Inputs, *IRE Trans. Autom. Control*, vol. AC-7, no. 5 (October, 1962), pp. 122-124.
 27. Merkulova, E. P.: The Problem of Optimizing Systems Which Contain Essentially Nonlinear Elements, *Automation and Remote Control*, vol. 20, no. 10 (June, 1960), pp. 1303-1313.
 28. Morosanov, I. S.: Effect of Fluctuations on Extremal Relay Systems under Self-oscillation Conditions, *Automation and Remote Control*, vol. 21, no. 9 (April, 1961), pp. 884-890.
 29. Newton, G. C., L. A. Gould, and J. F. Kaiser: "Analytic Design of Linear Feedback Controls," John Wiley & Sons, Inc., New York, 1957.
 30. Nikiforuk, P. N.: Response of a Particular Nonlinear Control System to Random Signals, *AIEE Trans.*, pt. II, vol. 75 (January, 1957), pp. 419-422.
 31. Nikiforuk, P. N., and J. C. West: The Describing Function Analysis of a Nonlinear Servo Mechanism Subject to Stochastic Signals and Noise, *Proc. IEE*, vol. C-104, no. 5 (March, 1957), pp. 193-203.

32. Nuttall, A. H.: Theory and Application of the Separable Class of Random Processes, *M.I.T. Res. Lab. Electronics, Tech. Rept.* 343, May, 1958.
33. Oldenburger, R., and R. Sridhar: Signal Stabilization of a Control System with Random Inputs, *AIEE Trans.*, pt. II, vol. 80, no. 57 (November, 1961), pp. 260-267.
34. Pastel, M. P.: Analyzing Nonlinear Systems with Random Inputs, *Control Eng.*, February, 1962, pp. 113-117.
35. Popov, E. P.: On Evaluating the Quality of Nonlinear Automatic Systems with Random Interference, *Automation and Remote Control*, vol. 20, no. 10 (June, 1960), pp. 1281-1287.
36. Pupkov, K. A.: Method of Investigating the Accuracy of Essentially Nonlinear Automatic Control Systems by Means of Equivalent Transfer Functions, *Automation and Remote Control*, vol. 21, no. 2 (October, 1960), pp. 126-139.
37. Pyatnitskii, G. I.: The Effect of Stationary Random Processes on Automatic Control Systems Containing Essentially Nonlinear Elements, *Automation and Remote Control*, vol. 21, no. 4 (November, 1960), pp. 328-332.
38. Rice, S. O.: "Mathematical Analysis of Random Noise," *Bell System Tech. J.*, vols. 23 and 24, 1944 and 1945. Reprinted in N. Wax, "Selected Papers in Noise and Stochastic Processes," Dover Publications, Inc., New York, 1954.
39. Sawaragi, Y., and S. Takahashi: Response of Control Systems Containing Zero-memory Non-linearity to Sinusoidal and Gaussian Inputs, *Proc. Heidelberg Conf. Autom. Control*, 1956, pp. 271-274.
40. Smith, H. W.: Analysis of Nonlinear Feedback Systems with Random Inputs, Sc.D. thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Cambridge, Mass., June, 1961. Revised as "Approximate Analysis of Randomly Excited Nonlinear Controls," Research Monograph 34, The M.I.T. Press, Cambridge, Mass., 1966.
41. Smith, H. W.: The Applicability of Quasi-linear Methods to Nonlinear Feedback Systems with Random Inputs, *Proc. Second IFAC Congr.*, Basel, Switzerland, August, 1963.
42. Somerville, M. J., and D. P. Atherton: Multi-gain Representation for a Single-valued Nonlinearity with Several Inputs and the Evaluation of Their Equivalent Gains by a Cursor Method, *Proc. IEE*, vol. C-105, no. 8 (July, 1958), pp. 537-549.
43. Sridhar, R., and R. Oldenburger: Stability of a Nonlinear Feedback System in the Presence of Gaussian Noise, *Trans. ASME, J. Basic Eng.*, March, 1962, pp. 61-70.
44. Thompson, W. E.: The Response of a Non-linear System to Random Noise, *Proc. IEE*, vol. C-102, no. 1 (September, 1954), pp. 46-48.
45. Tsympkin, Y. Z.: The Effect of Random Interference on a Periodic State in Relay Automatic Systems, *Soviet Phys.*, vol. 6, no. 7 (January, 1962), pp. 562-564.
46. Vander Velde, W. E.: A Nonlinear Method of Automatic Control in the Presence of Random Noise, Sc.D. thesis, Massachusetts Institute of Technology, Department of Aeronautical Engineering, May, 1956.
47. West, J. C., and P. N. Nikiforuk: The Response of Remote-position Control Systems with Hard-spring Non-linear Characteristics to Step Function and Random Inputs, *Proc. IEE*, vol. B-102, no. 5 (October, 1954), pp. 575-593.

PROBLEMS

- 7-1. An alternative means of statistical linearization for static single-valued nonlinearities is to define an equivalent gain as the ratio of rms nonlinearity output to rms input. Consider the nonlinearity input to be an unbiased gaussian process, and compute this

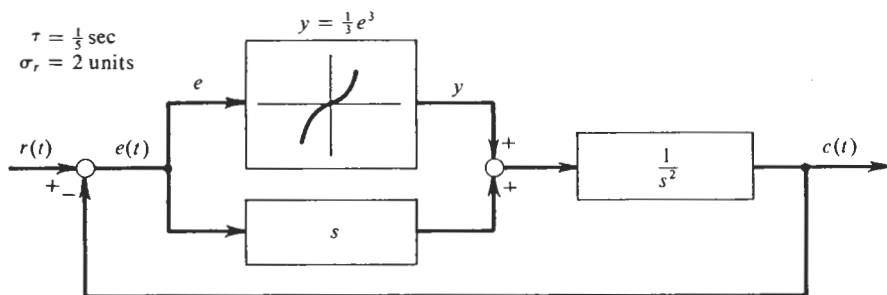
equivalent gain for the cubic nonlinearity, nonlinearity 18 of Appendix E. The integral form of Eq. (7.2-25) will be helpful in this calculation. Compare the result with N_R given in Appendix E.

- 7-2. Compute the rms equivalent gain for the relay with dead zone, nonlinearity 3 of Appendix E, and compare the result with N_R for a range of values of σ/δ .
- 7-3. Compute the rms equivalent gain for the limiter, nonlinearity 7 of Appendix E, and compare the result with N_R for a range of values of σ/δ .
- 7-4. Compute the gaussian-plus-bias-input RIDFs, $N_R(\sigma, B)$ and $N_B(\sigma, B)$, for the piecewise-linear nonlinearity, nonlinearity 13 of Appendix E, using Eqs. (7.2-69) and (7.2-70) and the previously given expressions for I_1 and I_2 .
- 7-5. Note how N_R and N_B can be obtained for nonlinearities 3 to 5, 7 to 9, and 12 of Appendix E by specializing the results of Prob. 7-4.
- 7-6. Calculate N_R and N_B for nonlinearity 7 of Appendix E by specializing the results for nonlinearity 6.
- 7-7. Calculate N_R and N_B for nonlinearity 8 of Appendix E by writing the nonlinear function as the sum of a linear function plus nonlinearity 7. Note that this could also be done in the case of the three-input RIDFs, using the graphed data for the limiter.
- 7-8. Calculate N_R and N_B for nonlinearity 11 of Appendix E by specializing the results for nonlinearity 6, and by expressing the nonlinear function as the sum of two nonlinearities of the form of nonlinearity 8.
- 7-9. Calculate N_R and N_B for nonlinearity 13 of Appendix E by expressing the nonlinear function as the sum of a linear function plus nonlinearity 12, and by expressing it as the sum of nonlinear functions of the forms 3 and 9.
- 7-10. Calculate approximate values for the gaussian-plus-bias-input RIDFs for the limiter, nonlinearity 7 of Appendix E, for several choices of σ/δ and B/δ , using the integration formulas (7.2-85) and (7.2-86). Try both 10- and 20-point integration formulas, and compare the results with the analytic answers given in Appendix E.
- 7-11. Verify the relation

$$\frac{d}{dB} (BN_B) = N_R$$

in the case of the two nonlinearities for which analytic forms of the three-input RIDFs are given in Sec. E-3.

- 7-12. What is the steady-state mean-squared error in the system of Fig. 7-1 when driven by a gaussian process having the given power spectral density function?



$$\Phi_{rr}(s) = \frac{2\tau\sigma_r^2}{1 - \tau^2 s^2}$$

Figure 7-1

- 7-13. Using the results of Prob. 7-1, compute the theoretical value of σ_r corresponding to several choices of σ_x and a for which experimental points are plotted in Fig. 7.3-5. Compare these theoretical results, using the rms equivalent gain for the nonlinearity, with the experimental values and the theoretical results, using the describing function of this book (Fig. 7.3-5), and note that in almost every instance the describing function defined here gives more accurate results.
- 7-14. Do the same comparison as in Prob. 7-13, using the results of Prob. 7-2 and the data of Fig. 7.3-6.
- 7-15. Do the same comparison as in Prob. 7-13, using the results of Prob. 7-3 and the data of Fig. 7.3-7.
- 7-16. Determine the optimum linear compensator $G'(s)$ and the resulting mean-squared error for the system of Fig. 7-2. The desired output is the signal component of the input.

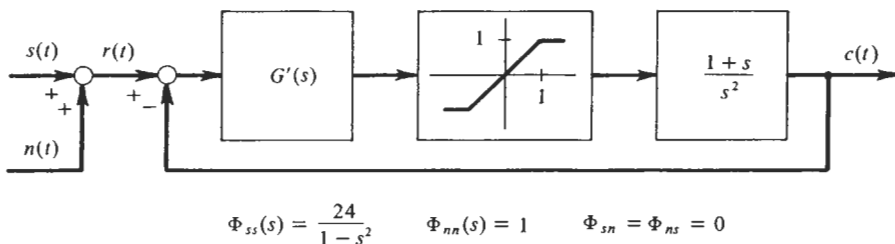


Figure 7-2

- 7-17. Determine the minimum rms value of gaussian dither, $d(t)$, which assures that the system of Fig. 7-3 will not limit-cycle in the absence of other inputs.

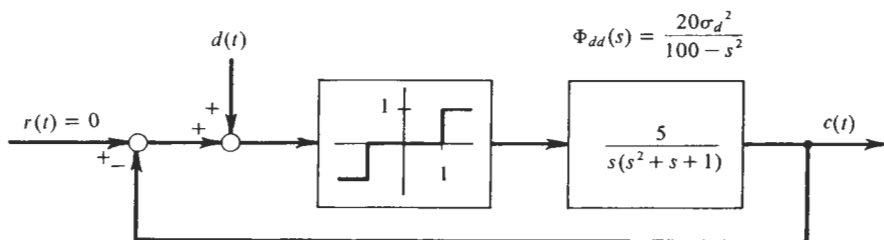


Figure 7-3

- 7-18. The system of Fig. 7-4 responds primarily to constant input commands and random disturbances. Plot the mean and standard deviation of the error, $e(t)$, for constant inputs ranging from 0 to 25 units.

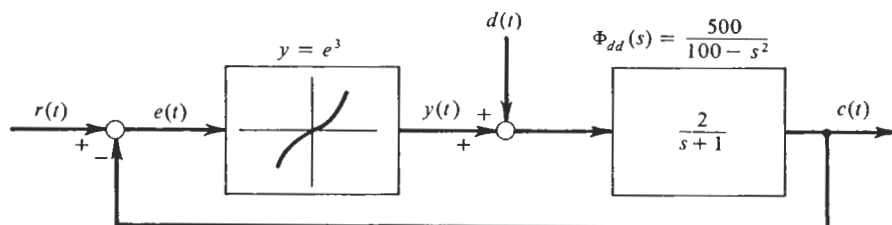


Figure 7-4

7-19. The system of Fig. 7-5 responds to a ramp command input and a gaussian disturbance input. What are the steady-state mean and variance of $x(t)$ under the following conditions:

$$\begin{aligned} m &= 2 \text{ units/sec} \\ D &= 2 \text{ units} \\ K &= 2 \text{ sec}^{-1} \\ \Phi_{dd}(s) &= 5 \text{ units}^2/\text{cps} \end{aligned}$$

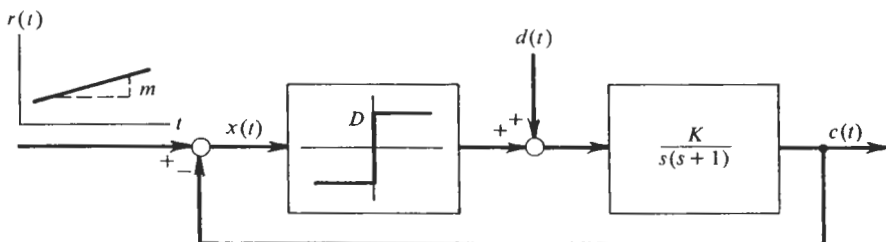


Figure 7-5

7-20. The system of Fig. 7-6 responds to a sinusoidal command input, a constant disturbance input, and a gaussian measurement noise in the feedback signal generator. What values of R , D , σ_n result in the error, $x(t)$, having a bias component of 1 unit, a sinusoidal component with amplitude 2 units, and a gaussian component with rms value of 1 unit?

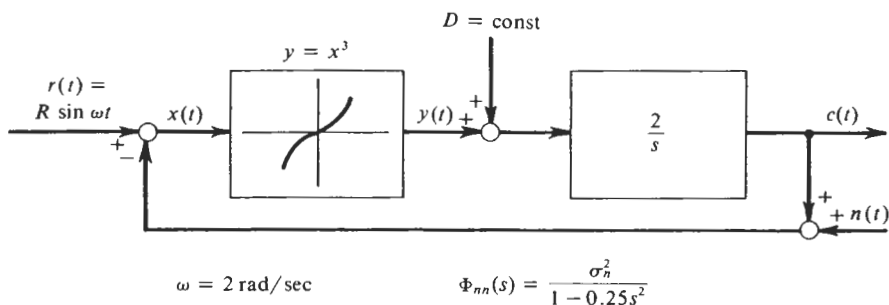


Figure 7-6

- 7-21. Solve the three-input illustration problem of Sec. 7.4 (system of Fig. 7.4-4) for the characteristics of the error, $x(t)$, in the case of the following parameter values:

$$k = 1 \text{ unit/sec}$$

$$D = 0.5 \text{ unit}$$

$$\tau = 0.1 \text{ sec}$$

$$\sigma_d = 1 \text{ unit}$$

- 7-22. Determine the modified nonlinearity for the system of Fig. 7.4-6 which accounts for the presence of noise in the system, and which can subsequently be used to evaluate the response to other inputs under the generalization of the bias signal to an arbitrary small or slow signal.

8 NONOSCILLATORY TRANSIENTS IN NONLINEAR SYSTEMS

8.0 INTRODUCTION

In general, describing function concepts are directed primarily at the *steady-state* responses of nonlinear systems. The test inputs used to develop most describing functions are therefore based upon signal forms which are expected in steady-state system operation. We have seen this to be the case in the formulation of the sinusoidal-input describing function (DF), two-sinusoid-input describing function (TSIDF), and dual-input describing function (DIDF) of previous chapters. In the latter instance the linearization employed also permitted determination of the *transient* response of a *limit cycling* nonlinear system under certain conditions. When a random process was considered, its statistical properties were determined on a steady-state basis. If, on the other hand, an approximate solution for the *transient* response of a *non-limit-cycling* nonlinear system is desired, we are led to consider *aperiodic* nonlinearity test inputs.

In this chapter we treat two describing functions based on aperiodic test inputs. One is the *transient-input describing function* due to Chen (Refs.

3, 4). It is formulated in a way considerably different from that presented in earlier sections of the text. Second is the *exponential-input describing function*. This new describing function is based on an exponential test input and follows very closely the describing function framework established earlier in the text. Both describing functions, emphasizing simplicity in use, are intended to facilitate the design of nonlinear systems.

8.1 TRANSIENT-INPUT DESCRIBING FUNCTION

The philosophy of quasi-linearization employed here is adopted specifically for the purpose of studying the transient response of nonlinear systems. For convenience, in transient investigation we take the input to be a step. This assumption, however, need not exclude the consideration of another form of input, which may be regarded as generated by the step response of a certain linear network. Nonlinearities under consideration are taken to be piecewise-linear as a result of the approximation procedure to be presented. These are assumed to be followed by linear low-pass elements, as in conventional describing function theory.

At the outset, a problem of circular nature arises. In order to perform an appropriate quasi-linearization of a nonlinear element it is necessary to test that element with a signal which simulates the actual input; however, determination of the actual input depends upon the nonlinearity quasi-linearization, which has allowed the system investigation in the first place. This apparent difficulty may be resolved by recalling that, since the nonlinearity under consideration is piecewise-linear, resulting in a piecewise-linear overall system, the shifting of the nonlinearity operating point during the transient can be readily determined on a "marching" basis. A rough determination of the actual transient response in the first linear range via ordinary linear techniques determines the test function for the nonlinearity in the first two linear ranges. The quasi-linearization so obtained permits a rough evaluation of the actual transient response in the second linear range. This, in turn, determines the test function for the nonlinearity in the first three linear ranges, etc. This procedure is shown to provide useful guidance in design work through yielding a qualitative indication of the dependence of system transient response on system parameters.

The above-outlined quasi-linearization procedure is well illuminated by the following example. Consider the simple system of Fig. 8.1-1, containing a limiter and a pure integrator. Note, first, that this system cannot limit-cycle. Immediately following application of a step input of magnitude R (for $R > \delta$), the limiter output assumes the value D . Accordingly, the actual nonlinearity input is

$$x(t) = R - DKt \quad \text{for } 0 \leq t < \frac{R - \delta}{DK} \quad (8.1-1)$$

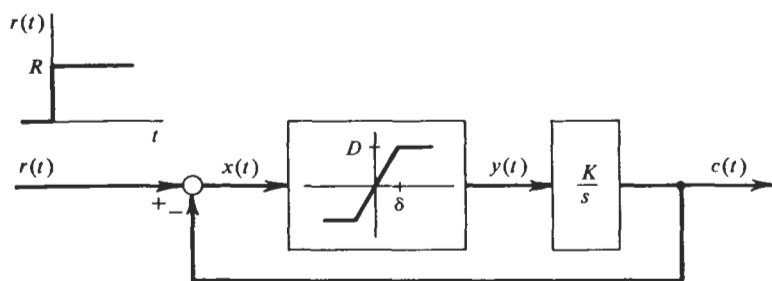


Figure 8.1-1 Simple nonlinear system.

which therefore is used as the test input. Figure 8.1-2 illustrates both nonlinearity input and output signals. At this point the appropriate transient-input describing function can be constructed. This is accomplished by obtaining the nonlinearity output as the summation of a static operation on its input less a *residual* $x_r(t)$, chosen so as to complete the required total output signal description. Figure 8.1-3a shows the resultant dynamic quasi-linear element, of which the static part is the transient-input describing function, $K_{eq} = D/\delta$. As is the case in this example, K_{eq} is generally chosen to be the slope of the nonlinearity at the origin. Figure 8.1-3b illustrates an equivalent block-diagram representation in which $x_r(t)$ is derived by a linear operation, $L_r(s)$, on the system input. $L_r(s)$ is derived as follows:

$$\begin{aligned}
 L_r(s) &= \frac{\mathcal{L}[x_r(t)]}{\mathcal{L}[r(t)]} \\
 &= \frac{\frac{(R - \delta)D}{\delta s} - \frac{D^2 K}{\delta s^2} \left[1 - \exp\left(-\frac{R - \delta}{DK} s\right) \right]}{\frac{R}{s}} \\
 &= \frac{(R - \delta)D}{R\delta} - \frac{D^2 K}{R\delta s} \left[1 - \exp\left(-\frac{R - \delta}{DK} s\right) \right] \quad (8.1-2)
 \end{aligned}$$

Having formulated the nonlinearity quasi-linearization in terms of the block diagram of Fig. 8.1-3b, the resultant overall linear system can be redrawn as in Fig. 8.1-4a. Consideration of operation reveals that the linear system of Fig. 8.1-4a is a *precise* duplicator of the input-output dynamics of its nonlinear predecessor over the *entire* class of input step functions, because the example nonlinear system response cannot overshoot. The nonlinearity mode of operation (for $R > \delta$) is therefore first in saturation, then always thereafter in the linear range. The transient-input describing function is simply the linear region gain of the nonlinearity, wherewith the residual has

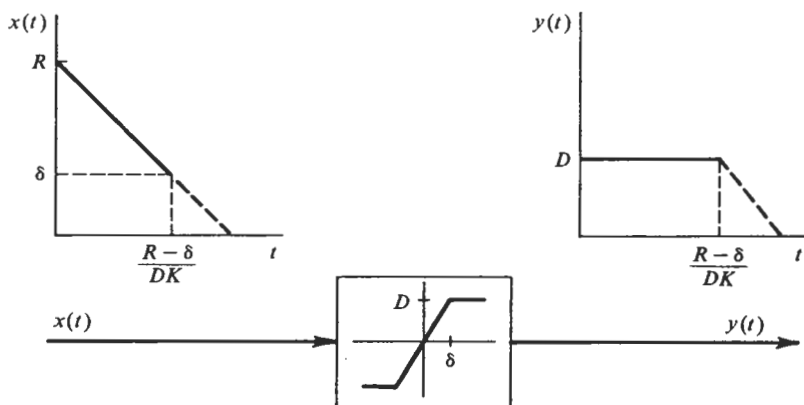
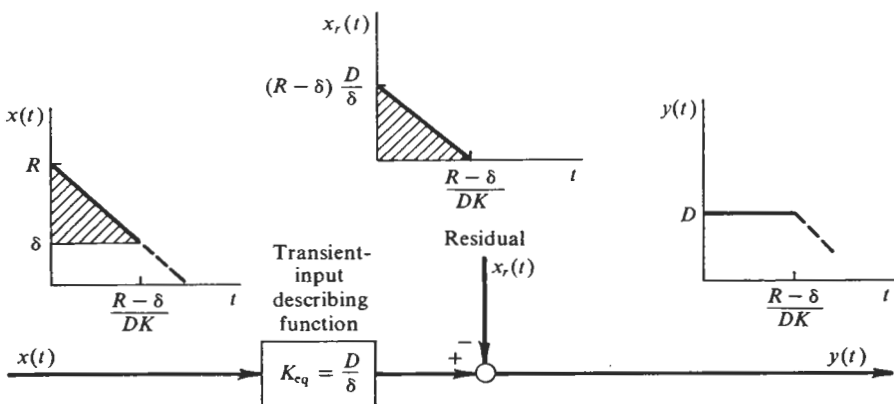


Figure 8.1-2 Testing nonlinearity by actual input. (Chen, Ref. 3.)

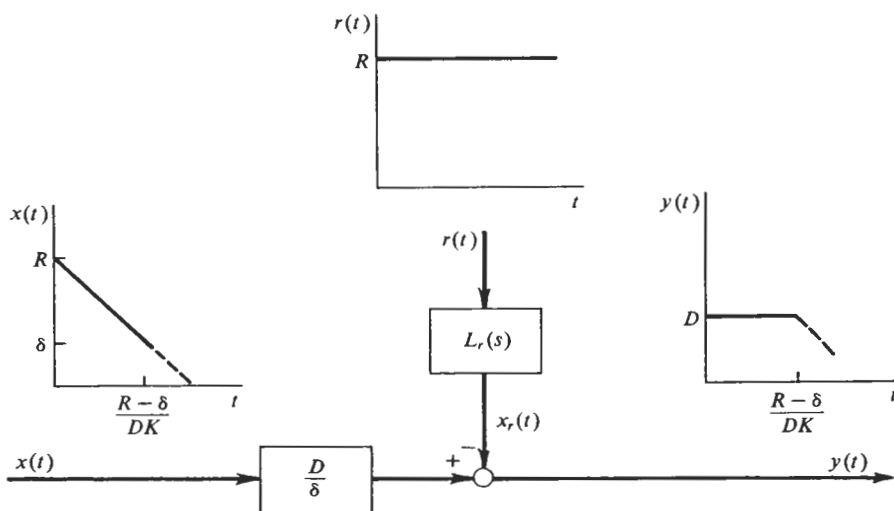
been chosen to correct the net linearized nonlinearity output while the nonlinearity is actually in the saturated operating region.

Quasi-linearization of the non-limit-cycling nonlinear system has been accomplished by the use of a transient-input describing function and an appropriate residual $x_r(t)$. In contrast to the describing functions of previous chapters, the transient-input describing function is itself not a function of the input signal; however, the residual in this case may *not* be ignored. *The residual alone accounts for nonlinear operation.* The effect of the nonlinearity is evident in both sections of the quasi-linearized system of Fig. 8.1-4b, which is a somewhat simplified arrangement of Fig. 8.1-4a. Further, since $L_r(s)$ is a function of R , the overall system clearly displays the input-signal dependence characteristic of all nonlinear systems.

Success of the quasi-linearization procedure described depends upon the extent to which the linearized nonlinearity output describes its actual counterpart. To be sure, it is always possible to generate the *exact* nonlinearity output by treating the transient problem in several stages, each one of which is exactly described by a linear constant-coefficient differential equation. This follows from the hypothesized piecewise-linear description of all nonlinearities under consideration. However, the intent here is to generate rapidly an *approximate* nonlinearity output based upon dominant modes and upon the empirical procedures regularly employed in the study of linear systems; and from there to extrapolate the nonlinear-system transient response. Since the value of the proposed quasi-linearization scheme lies primarily in its use in design for specified transient response of nonlinear systems, and since the transient response is commonly specified by approximation concepts derived from the theory of linear systems, it is desirable to examine these concepts.

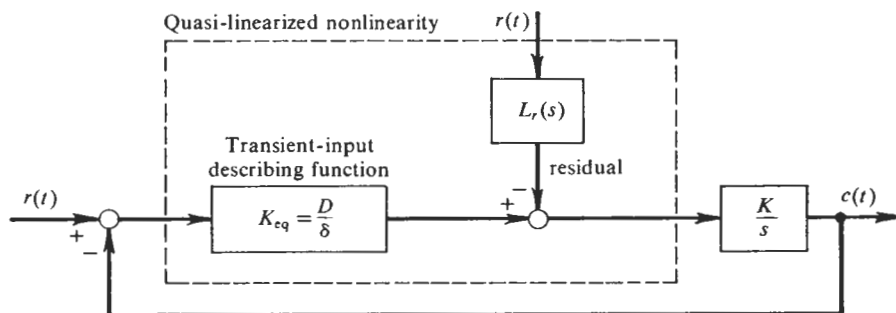


(a) Defining the transient-input describing function

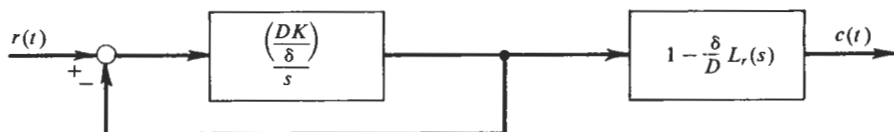


(b) Equivalent block diagram

Figure 8.1-3 Active quasi-linear element. (Adapted from Chen, Ref. 3.)



(a) Replacement of the nonlinearity



(b) Equivalent block-diagram representation

Figure 8.1-4 Quasi-linearized nonlinear system. (Adapted from Chen, Ref. 3.)

8.2 LINEAR-SYSTEM APPROXIMATIONS

Two distinct problem areas in which linear approximation techniques are employed are the derivation of a simplified pole-zero model for the quasi-linearized nonlinear system and determination of the transient response of this model. The former problem relates to the development of a simplified treatment for handling residuals, $x_r(t)$, as discussed below.

DEALING WITH THE RESIDUAL

It is possible to generate the approximate residual required in a given situation by graphical means. Consider, for example, a nonlinear system of the form shown in Fig. 8.1-1, but in which the transfer function of the linear element includes additional denominator dynamics. For transient responses which may overshoot, but in which the peak overshoot does not exceed δ , $x_r(t)$ may be derived from Fig. 8.2-1. At any time $t = t_i$, $x_r(t)$ is simply the difference between the output of a linear element of gain D/δ (the transient-input describing function for this example) and the actual limiter output. This gives the residual as an explicit function of the nonlinearity input x . Provided that x can be obtained as a function of time, $x_r(t)$ directly follows. At this point engineering approximation is essential for ultimate usability of the technique; $c(t)$ must be estimated. Of course, it is always possible to

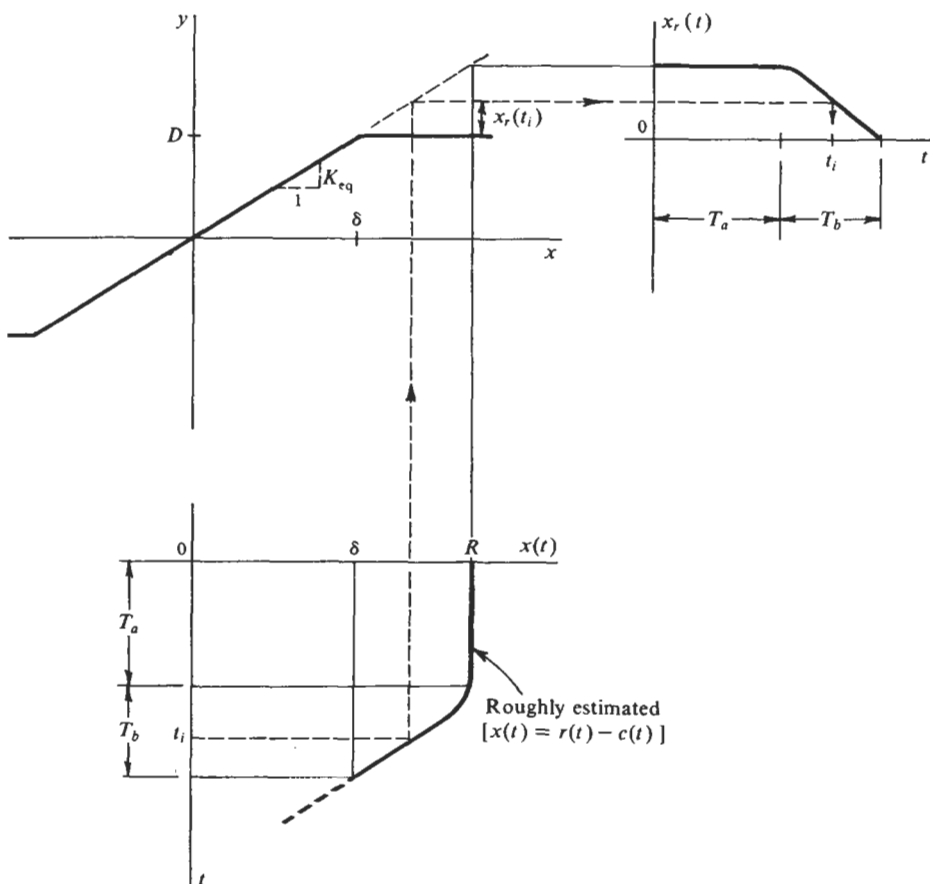


Figure 8.2-1 Example derivation of $x_r(t)$.

compute the required portion of $c(t)$ exactly. With the exception of the simplest cases, however, such a calculation is undesirable. In the hypothetical example under consideration, $c(t)$ often can be described by a dead time, followed by a ramplike function in the initial stages of the transient response. Such is the initial part of the step response of the linear elements. A crude estimate of the dead time serves to calibrate the time axis, which allows determination of $x_r(t)$, as indicated graphically in Fig. 8.2-1. T_a is the estimated output step-response dead time, and $T_a + T_b$ is the estimated time at which the output $c(t)$ reaches the value $R - \delta$. Thus we have derived a trapezoidal residual. It can be further approximated by straight-line segments, as in Fig. 8.2-2. Since this residual shape is of use in a variety of cases, we continue studying it.

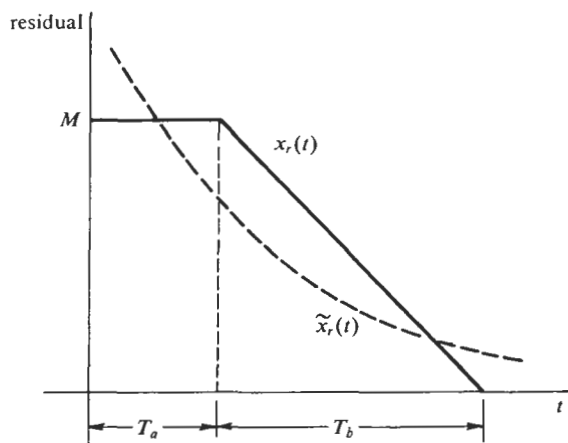


Figure 8.2-2 Approximation of a trapezoidal pulse by a single exponential function. (Chen, Ref. 3.)

The method of quasi-linearization employed defines the residual $x_r(t)$ as the step response of a linear network. Even in the simple example considered in Sec. 8.1, the transfer function of this linear network, $L_r(s)$, is found to be relatively complex. It is of advantage to find a finite nontranscendental linear-network transfer function with which to approximate $L_r(s)$. Such can be found by utilization of the Padé approximation (Ref. 6). The resulting network, denoted by $\tilde{L}_r(s)$, is required to have as its transfer function a ratio of polynomials in s . Its use is based on the assumption of low-pass loop linear elements, a requirement common to all describing function methods.

Returning to the trapezoidal residual of Fig. 8.2-2, the exact transfer function of the linear residual shaping network, $L_r(s)$, is observed to be

$$L_r(s) = \frac{M}{R} - \frac{M}{RT_b s} e^{-T_a s} (1 - e^{-T_b s}) \quad (8.2-1)$$

The approximating network, in general, is taken as

$$\tilde{L}_r(s) = \frac{a_0 + a_1 s + a_2 s^2 + \cdots + a_n s^n}{1 + b_1 s + b_2 s^2 + \cdots + b_n s^n} \quad (8.2-2)$$

$\tilde{L}_r(s)$ can be chosen equivalent to $L_r(s)$ in the Padé sense by expanding each in ascending powers of s and choosing the coefficients a_i and b_i to force equality up to terms of order $2n$ in s . This procedure assures identical output moments, up to order $2n$, in the time domain. In other words,

$$\int_0^\infty x_r(t) t^k dt = \int_0^\infty \tilde{x}_r(t) t^k dt \quad \text{for } 0 \leq k \leq 2n \quad (8.2-3)$$

The expansion of $L_r(s)$ is

$$\begin{aligned} L_r(s) &= \frac{M}{R} - \frac{M}{RT_b s} \sum_{k=0}^{\infty} (-1)^k \frac{T_a^k s^k}{k!} \left[1 - \sum_{l=0}^{\infty} (-1)^l \frac{T_b^l s^l}{l!} \right] \\ &= \frac{M}{R} \left(T_a + \frac{T_b}{2} \right) s - \frac{M}{R} \left(\frac{T_a^2}{2} + \frac{T_a T_b}{2} + \frac{T_b^2}{6} \right) s^2 + \dots \quad (8.2-4) \end{aligned}$$

For simplicity in the final analysis $\tilde{L}_r(s)$ is taken to be of the first order. The required expansion is

$$\begin{aligned} \tilde{L}_r(s) &= \frac{a_0 + a_1 s}{1 + b_1 s} \\ &= a_0 + (a_1 - a_0 b_1) s - b_1 (a_1 - a_0 b_1) s^2 - \dots \quad (8.2-5) \end{aligned}$$

Equations (8.2-4) and (8.2-5), compared, indicate the result that

$$\tilde{L}_r(s) = \frac{(M/R)(T_a + T_b/2)s}{1 + [(3T_a^2 + 3T_a T_b + T_b^2)/(6T_a + 3T_b)]s} \quad (8.2-6)$$

The step response of this network, $\tilde{x}_r(t)$, is sketched in Fig. 8.2-2. To obtain the proper initial value of $\tilde{x}_r(t)$ it is desirable to alter $\tilde{L}_r(s)$ slightly, by so choosing b_1 as to yield

$$\tilde{L}_r(s) = \frac{(M/R)(T_a + T_b/2)s}{1 + (T_a + T_b/2)s} \quad (8.2-7)$$

This function both supplies the correct initial value of $\tilde{x}_r(t)$ and possesses the same output response area (0th moment) as the ideal residual shaping network response of Eq. (8.2-1). The simplicity of this approximating network is evidenced upon application of it in the example system in the following section. Where the initial value of $\tilde{x}_r(t)$ is not of concern, a denominator time constant given by $(3T_a + 2T_b)/6$ yields a useful second approximation which is exact with respect to Eq. (8.2-6) at the extremes $T_a = 0$ or $T_b = 0$.

It is true that higher-order functions $\tilde{L}_r(s)$ can lead to far better approximations to $x_r(t)$. For example, it is possible to find a second-order filter with step response.

$$\tilde{x}_r(t) = A e^{-\alpha t} \cos(\beta t + \gamma) \quad (8.2-8)$$

which can be fairly closely fitted to $x_r(t)$ of Fig. 8.2-2. Certainly, the approximation is better than that obtained with the residual forming filter of Eq. (8.2-6). The price paid for this refinement is increased computational difficulty.

APPROXIMATIONS FOR LINEAR-SYSTEM TRANSIENT RESPONSE

The final form of the quasi-linearized nonlinear system is a rational linear transfer function, the parameters of which are determined by the original

nonlinear system and the input. All the well-known empirical rules linking transient response with either frequency response or pole-zero data can therefore be brought to bear on the design problem. Conventional measures of transient response, such as *rise time*, *time to peak*, *peak overshoot*, and *settling time*, are extremely useful in this endeavor. Their application is addressed to a system description in terms of *dominant* behavioral modes (e.g., a dominant complex pole pair). Since these are treated at length in standard servo texts, they are not repeated here (see, for example, Refs. 5 and 7).

An important concept in relating the time and frequency or complex domains relates to the regions in each which are dominantly responsible for response characteristics in the other. In particular, it is well known that the low-frequency region (corresponding to poles and zeros near the origin) bears heavily on the final stages of transient response; the high-frequency region (corresponding to poles and zeros far from the origin) bears heavily on the initial stages of transient response; and so on for the intermediate ranges of each. This concept underlies the usefulness of so-called *error coefficients*, which directly relate the response of a linear network to the input and its derivatives or integrals (Ref. 2).

In brief demonstration of the type of approximation to be employed, consider the following quick method of determining *delay time*, suggested by Chen (Ref. 3). A linear transfer function $L(s)$ is assumed describable in the form

$$L(s) = L_1(s) \frac{T_1 s + 1}{T_2 s + 1} \quad (8.2-9)$$

where the pole and zero at $-1/T_2$ and $-1/T_1$, respectively, are much farther from the origin than any poles and zeros of $L_1(s)$. $L(s)$ may thus be approximated, relative to its contribution to the final portion of the transient (that is, $t \gg 0$ or s small), by

$$L(s) \approx L_1(s) e^{T_1 s} e^{-T_2 s} \quad (8.2-10)$$

In this form it is clear that the transient response delay time increases with far removed left-half-plane poles and decreases with far removed left-half-plane zeros. Since poles and zeros are not commonly found under the limiting circumstances described, the rule of thumb applicable is to weight the net delay time by a factor of $\frac{1}{2}$. Thus, for the example in question,

$$T_d \approx \frac{1}{2}(T_2 - T_1) \quad (8.2-11)$$

Poles or zeros in the right half-plane may be readily accounted for by utilizing the proper sign in Eq. (8.2-11).

The delay-time concept is by itself most valuable in systems whose initial transient slope is zero or near zero. This slope is quite easily related to the

system poles and zeros. For the general linear network given by

$$L(s) = \frac{\prod_{i=1}^m (T_i s + 1)}{\prod_{j=1}^n (T_j s + 1)} \quad (8.2-12)$$

the slope of the transient response is

$$\frac{dc}{dt}(t) = \mathcal{L}^{-1} \left[R \frac{\prod_{i=1}^m (T_i s + 1)}{\prod_{j=1}^n (T_j s + 1)} \right] \quad (8.2-13)$$

which, at time zero, by the initial-value theorem of Laplace transform analysis, is simply

$$\frac{dc}{dt}(t=0) = \begin{cases} 0 & \text{for } n > m + 1 \\ R \frac{\prod_{i=1}^m (T_i)}{\prod_{j=1}^n (T_j)} & \text{for } n = m + 1 \\ \infty & \text{for } n < m + 1 \end{cases} \quad (8.2-14)$$

The suggested procedure for approximating the residual by a trapezoidal signal followed by the determination of an approximation to the transfer function which yields such a signal from a step input is certainly not the only possible line of approach. Any reasoning which leads to a suitable residual shaping network is admissible. The technique is suggested simply as a standardized procedure useful in a variety of instances.

Armed with a quasi-linear system and the rules of thumb so valuable in linear-system work, the designer may proceed to the main question of nonlinear-system transient response.

8.3 TRANSIENT RESPONSE OF NON-LIMIT-CYCLING NONLINEAR SYSTEMS

Utilizing the transient-input describing function and including the residual in the overall quasi-linear system formulation, we can develop a pole-zero description of the system dynamics. Some of the poles and zeros result as functions of the transient-input amplitude R , correctly indicating that the nonlinear-system response is generally input-amplitude-dependent.

The nonlinear system of Fig. 8.1-1, for example, leads to the quasi-linear

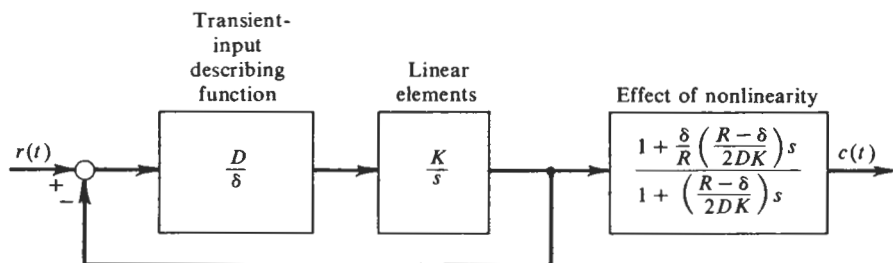


Figure 8.3-1 Quasi-linearized example system. Note that the effect of the nonlinearity appears as an external lag network ($R \geq \delta$).

system of Fig. 8.3-1. The derivation follows from Fig. 8.1-4b, in which $1 - (\delta/D)L_r(s)$ is replaced according to

$$\begin{aligned}
 1 - \frac{\delta}{D} L_r(s) &\approx 1 - \frac{\delta}{D} \tilde{L}_r(s) \\
 &= \frac{1 + \frac{\delta}{R} \left(\frac{R - \delta}{2DK} \right) s}{1 + \left(\frac{R - \delta}{2DK} \right) s} \quad (8.3-1)
 \end{aligned}$$

where $\tilde{L}_r(s)$ is obtained from Eq. (8.2-7) with $T_a = 0$, $T_b = (R - \delta)/DK$, and $M = (R - \delta)(D/\delta)$.

A sequence of pole-zero plots for the example system at the values $R = \delta$, 2δ , 5δ , 9δ is shown in Fig. 8.3-2. The progression of the pole-zero pattern with increasing R (denoted by arrows) clearly indicates the changing transient response character. For $R \leq \delta$ the system dynamics appear independent of R , and are so indicated. As R begins to exceed δ , a pole-zero pair (representing nonlinear operation) moves in toward the origin along the negative real axis. In the limit of increasing R , the system develops a pure integration, and the variable zero asymptotically approaches the value $-2DK/\delta$ or $-2/\tau$.

The initial slope of the approximated transient response for all $R \geq \delta$, as given by Eq. (8.2-14), is

$$\begin{aligned}
 \frac{dc}{dt}(t=0) &= R \frac{\frac{\delta}{R} \left(\frac{R - \delta}{2DK} \right)}{\left(\frac{\delta}{DK} \right) \left(\frac{R - \delta}{2DK} \right)} \\
 &= DK \quad (8.3-2)
 \end{aligned}$$

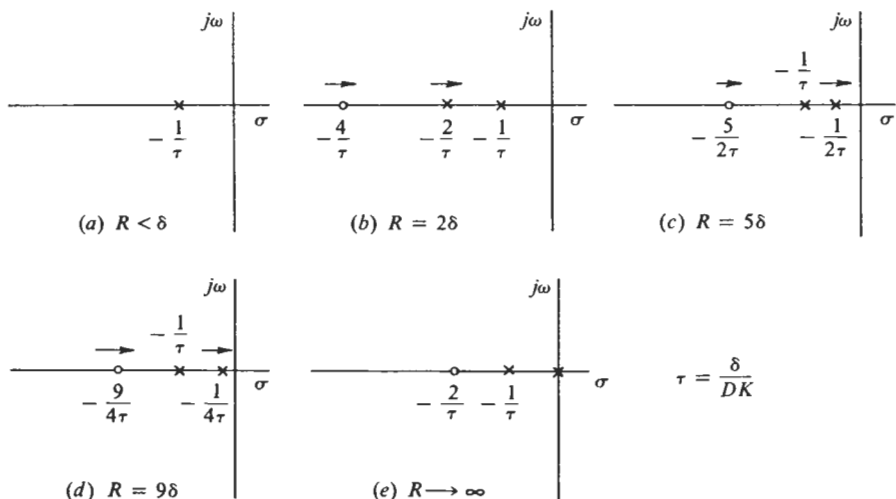


Figure 8.3-2 Pole-zero plots for various input amplitudes, R . (Arrows indicate variable pole and zero migration directions with increasing R .)

This value is exact. The transient response settling time (time to within 95 to 105 percent of final value) is approximately given by three dominant system time constants. For $R \geq 5\delta$ it takes the value

$$T_s \approx \frac{3(R - \delta)}{2DK} = \frac{3}{2} \left(\frac{R}{\delta} - 1 \right) \tau \quad (8.3-3)$$

Use of this type of approximation enables sketching the transient response of the quasi-linear system.

For the simple system of Fig. 8.3-1, the transient response is readily determined by linear theory as

$$\frac{1}{\delta} c(t) = \frac{R}{\delta} - \frac{(R - \delta)^2}{\delta(R - 3\delta)} \exp \left(-\frac{2DK}{R - \delta} t \right) + \frac{R + \delta}{R - 3\delta} \exp \left(-\frac{DK}{\delta} t \right) \quad (8.3-4)$$

This function is plotted in Fig. 8.3-3 for $R = \delta, 2\delta, 5\delta, 9\delta$, together with the exact transient responses of the original nonlinear system, given by

$$\frac{1}{\delta} c(t) = \begin{cases} \frac{DK}{\delta} t & 0 \leq t < \frac{R - \delta}{DK} \\ \frac{R}{\delta} \left[1 - \frac{\delta}{R} \exp \left(\frac{R - \delta}{\delta} \right) \exp \left(-\frac{DK}{\delta} t \right) \right] & t > \frac{R - \delta}{DK} \end{cases} \quad (8.3-5)$$

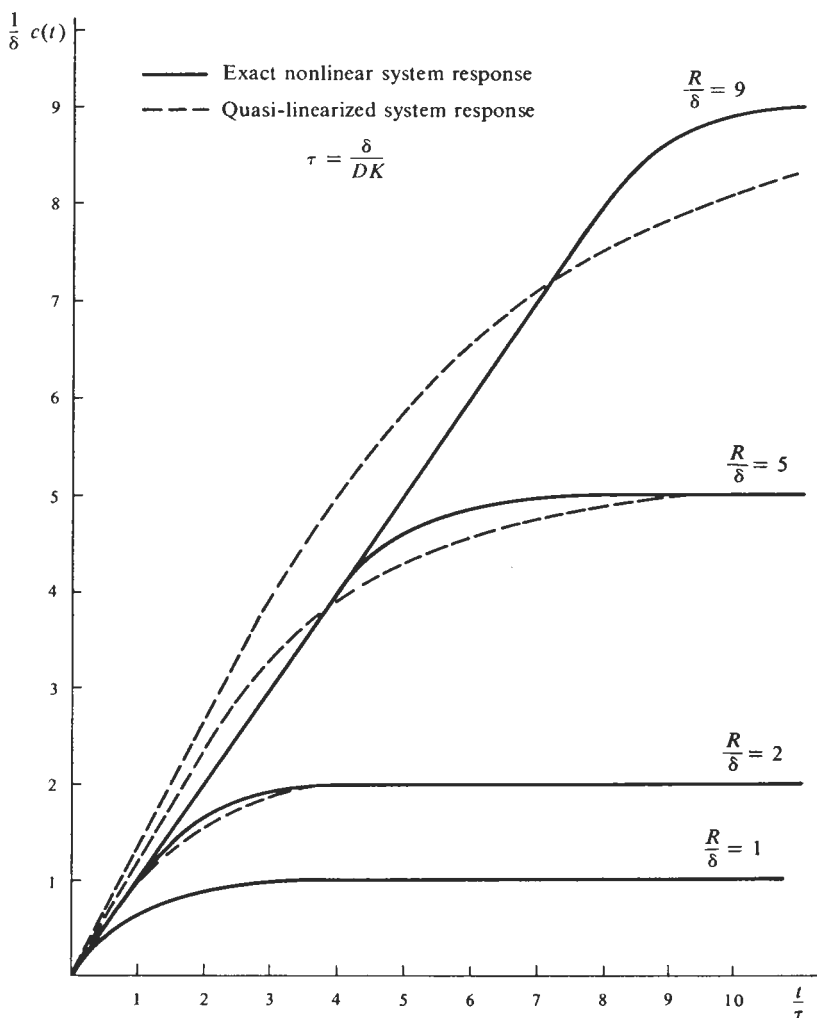


Figure 8.3-3 Exact and quasi-linear example-system responses.

At $R/\delta = 1$, of course, the exact and quasi-linearized responses coincide. With increasing R/δ , the difference between exact and approximated responses increases because of the limitations existing in any two-exponential fit to a straight line, the now dominant portion of the actual nonlinear-system response. In fact, in view of this difficulty, one must conclude that the simple linear-system approximation to the actual nonlinear-system response is rather good. All major aspects of the transient response are indeed accounted for in quasi-linearization. More important, the system poles

and zeros, and hence the transient response descriptors, are given in terms of system parameters.

The simple system used as the vehicle with which to convey the concept of quasi-linearization for transient response could well have been studied more exhaustively, and in fact with greater ease, by the phase-plane approach mentioned in Chap. 1. A third-order linear part to the example system would, however, require recourse to phase-space techniques; a fourth-order linear part could not be studied at all without substantial approximations reducing the order of the system. The transient-input describing function method, on the other hand, is not limited by the order of the system.

8.4 DESIGN PROCEDURE FOR A SPECIFIED TRANSIENT RESPONSE

The limiting factor in general utilization of the method described above relates to the difficulty in determining the residual, $x_r(t)$, and hence in obtaining a satisfactory residual shaping network $\tilde{L}_r(s)$ with which to account adequately for nonlinear behavior. Again it is pointed out that the nonlinear-system response solution can always be obtained by the piecewise-linear point of view, giving $x(t)$ analytically and exactly. From this comes the required $x_r(t)$; whence it follows that $L_r(s)$ can be exactly determined. Unfortunately, $L_r(s)$ generally results as a very complicated transcendental function. The use of a simplified nontranscendental residual shaping network $\tilde{L}_r(s)$, of no higher than second order, is mandatory for the eventual pencil-and-paper application of this method. The approximations which, in practice, must be made in the determination of $\tilde{L}_r(s)$ represent the most critical calculations incurred. Given that these can be executed according to reasonable engineering judgment, the resulting quasi-linear systems provide a useful analytical means for describing nonlinear-system transient behavior.

Chen (Ref. 4) suggests a very interesting technique which mitigates the above problem in design. Figure 8.4-1a illustrates the class of nonlinear systems to which this design procedure is applicable. Based on the given system specifications, one first seeks for a suitable hypothetical *totally linear system* and interprets this linear system in terms of a unity feedback configuration. Now, if the actual nonlinear system is to behave like the hypothetical linear system, the error signals in each instance must be quite similar. Hence the error signal derived from the linear system is used to test the nonlinearity, and thereby determine a suitable $x_r(t)$. This procedure is diagrammed in Fig. 8.4-2.

Once $x_r(t)$ is obtained, the procedure outlined in Sec. 8.2 is followed to give a suitable residual-forming network $\tilde{L}_r(s)$. Manipulating the now linear system block diagram as in Fig. 8.1-4 enables the residual-forming filter to be

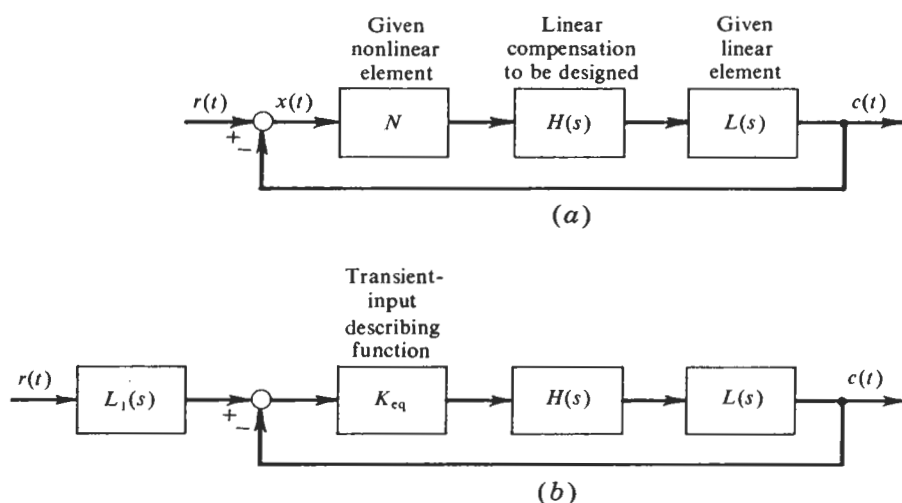


Figure 8.4-1 (a) Allowable control-system configuration. (b) Configuration of the quasi-linearized system. (Adapted from Chen, Ref. 4.)

transferred to a series connection with the closed loop. This defines $L_1(s)$ in Fig. 8.4-1b. The compensation network $H(s)$ is now chosen by purely linear design techniques to enable the quasi-linearized system to meet overall specifications.

At this point the compensated nonlinear system is simulated on an analog computer, and the response obtained. If the response does not meet specifications, either the transient-input describing function representation

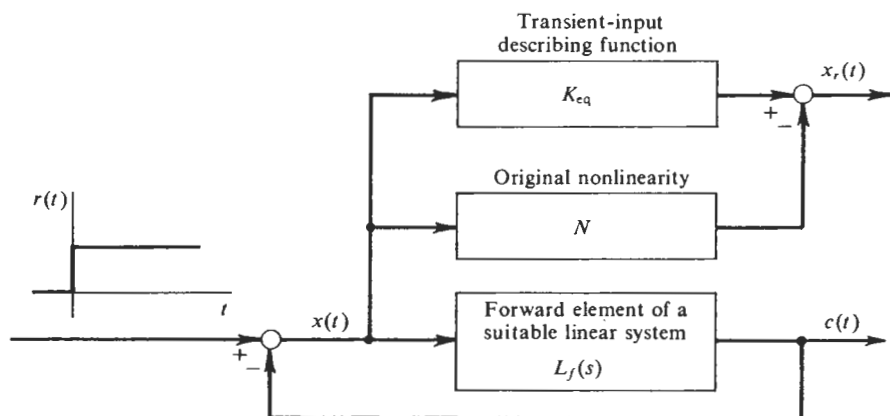


Figure 8.4-2 Computer setup for approximate determination of the residual. (Adapted from Chen, Ref. 4.)

can be changed slightly and the process repeated, or $H(s)$ can be refined. Regarding this procedure the reader should note that the analog computer has been applied in a systematic way. Intuition has not been heavily relied on, and a gross parameter-search procedure has been avoided.

8.5 EXPONENTIAL-INPUT DESCRIBING FUNCTION (EIDF)

In this section we present a new and quite useful technique of approximate solution for the transient response of nonlinear systems. The basis for this new approach, the *exponential-input describing function*, closely follows the main theme of this book. Thus, the ensuing presentation is brief and to the point. A related viewpoint can be found elsewhere (Ref. 1).

Consider the nonlinear system of Fig. 8.5-1a. If the output increases monotonically to its steady-state value when excited by a step input, the

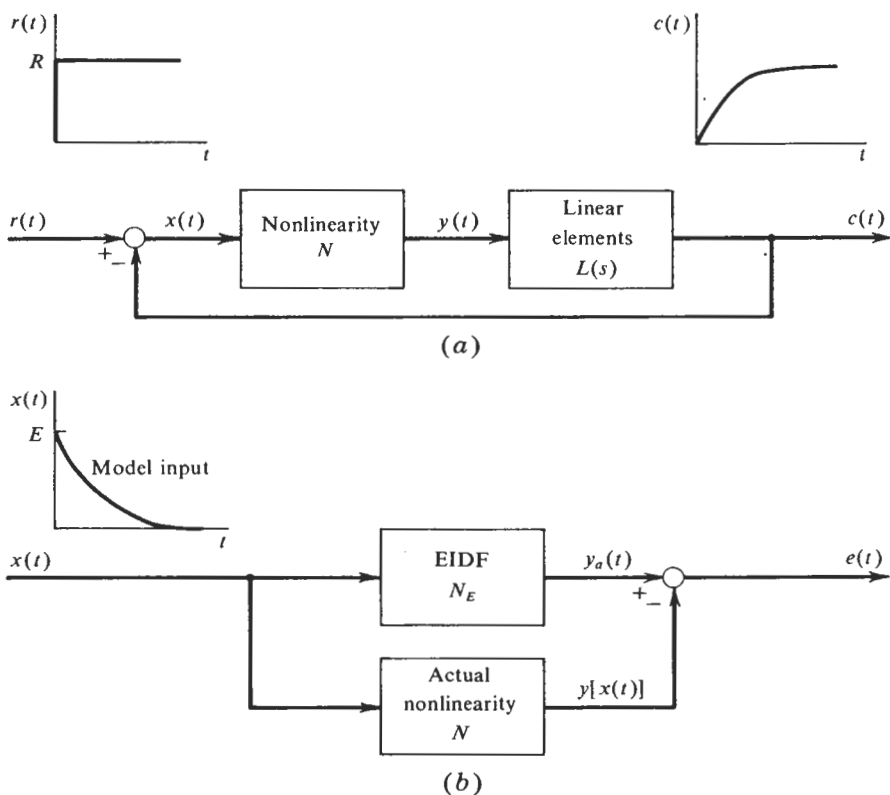


Figure 8.5-1 (a) Nonlinear system with nonoscillatory transient response. (b) Corresponding EIDF formulation.

signal $x(t)$ will correspondingly decrease in a monotonic way. Thus, we are led to consider a *model* input which is an *exponential*. The EIDF representation of the nonlinearity is arrived at by minimizing the integral-squared error in a linear approximation to the actual nonlinearity output; see Fig. 8.5-1b. The representation error, e , is

$$e(t) = N_E x(t) - y[x(t)] \quad (8.5-1)$$

where N_E is a fixed linear gain. The corresponding squared error, integrated over all time, is

$$\int_0^\infty e^2(t) dt = N_E^2 \int_0^\infty x^2(t) dt - 2N_E \int_0^\infty x(t)y[x(t)] dt + \int_0^\infty y^2[x(t)] dt \quad (8.5-2)$$

Minimizing this expression by differentiating with respect to N_E and setting the result to zero yields the EIDF,

$$N_E = \frac{\int_0^\infty x(t)y[x(t)] dt}{\int_0^\infty x^2(t) dt} \quad \text{for } x(t) = Ee^{-t/\tau} \quad (8.5-3)$$

Calculation of the EIDF proceeds easily. In the case of the sharp saturation (limiter) nonlinearity, for example, we get ($E > \delta$)

$$\begin{aligned} N_E(E) &= \frac{\int_0^{t_1} Ee^{-t/\tau} D dt + \int_{t_1}^\infty Ee^{-t/\tau} \frac{D}{\delta} Ee^{-t/\tau} dt}{\int_0^\infty E^2 e^{-2t/\tau} dt} \\ &= \frac{-ED\tau(e^{-t_1/\tau} - 1) + \frac{E^2 D}{\delta} \left(\frac{\tau}{2}\right) e^{-2t_1/\tau}}{\frac{\tau E^2}{2}} \end{aligned}$$

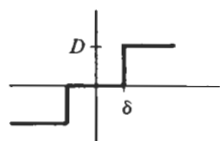
Noting that

$$e^{-t_1/\tau} = \frac{\delta}{E}$$

we obtain

$$N_E(E) = \frac{2D}{E} \left(1 - \frac{\delta}{2E}\right) \quad E > \delta \quad (8.5-4)$$

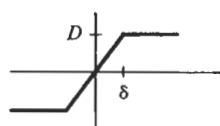
EIDFs for other common nonlinearities are presented in Fig. 8.5-2. Note that in the case of *all* static nonlinearities, the EIDF is *independent* of τ . This fact greatly facilitates its use, as we see in the following examples.



(a) Relay with dead zone

$$N_E = 0 \quad |E| \leq \delta$$

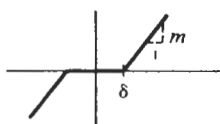
$$= \frac{2D}{|E|} \left(1 - \frac{\delta}{|E|}\right) \quad |E| > \delta$$



(b) Sharp saturation

$$N_E = \frac{D}{\delta} \quad |E| \leq \delta$$

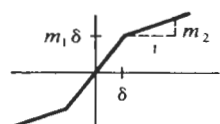
$$= \frac{2D}{|E|} \left(1 - \frac{\delta}{2|E|}\right)^2 \quad |E| > \delta$$



(c) Deadband gain

$$N_E = 0 \quad |E| \leq \delta$$

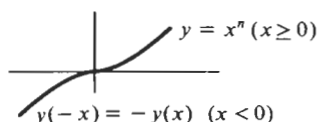
$$= m \left(1 - \frac{\delta}{|E|}\right)^2 \quad |E| > \delta$$



(d) Gain-changing element

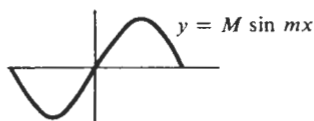
$$N_E = m_1 \quad |E| \leq \delta$$

$$= \frac{2\delta}{|E|} (m_1 - m_2) \left(1 - \frac{\delta}{2|E|}\right) + m_2 \quad |E| > \delta$$



(e) Polynomial nonlinearity

$$N_E = \frac{2}{n+1} |E|^{n-1}$$



(f) Harmonic nonlinearity

$$N_E = \frac{2M}{mE^2} (1 - \cos mE)$$

Figure 8.5-2 Exponential-input describing functions for some common nonlinearities.

Example 8.5-1 Solve for the transient response of the system illustrated in Fig. 8.1-1 by the exponential-input describing function method.

The first step is to replace the limiter by its exponential-input describing function,

$$N_E = \frac{2D}{|E|} \left(1 - \frac{\delta}{2|E|} \right)$$

Next, for the resulting linearized system, we write

$$C(s) = \frac{1}{\frac{s}{\frac{2DK}{R} \left(1 - \frac{\delta}{2R} \right)} + 1} \frac{R}{s}$$

where E , the peak exponential-input model amplitude, has been set equal to R . Inverse transforming to obtain $c(t)$ yields

$$c(t) = R \left\{ 1 - \exp \left[- \frac{2DK}{R} \left(1 - \frac{\delta}{2R} \right) t \right] \right\}$$

This solution, simpler than that of Eq. (8.3-4), is almost as good an approximation to the exact system response.

Example 8.5-2 Use the exponential-input describing function to compute the approximate transient response of the system illustrated in Fig. 4.1-2 when subject to the initial conditions $c(t=0) = c(0)$ and $\dot{c}(t=0) = 0$.

The exponential-input describing function for the ideal relay can be obtained by setting δ to zero in the corresponding expression for the relay with dead zone, Fig. 8.5-2. The result is

$$N_E = \frac{2D}{|E|} = \frac{2D}{c(0)}$$

where we have identified E , the peak value of the exponential-input model, as being equal to $-c(0)$. Replacing the ideal relay nonlinearity with the exponential-input describing function and treating the resulting linearized system by familiar transform techniques yields

$$C(s) = \frac{s + b}{s^2 + bs + KN_E} c(0)$$

Dividing the numerator and denominator of the right-hand side of this equation by b^2 and employing the normalized variables

$$p = \frac{s}{b} \quad \text{and} \quad V = \frac{DK}{b^2}$$

results in

$$C(p) = \frac{p + 1}{p^2 + p + 2V/c(0)} \frac{c(0)}{b}$$

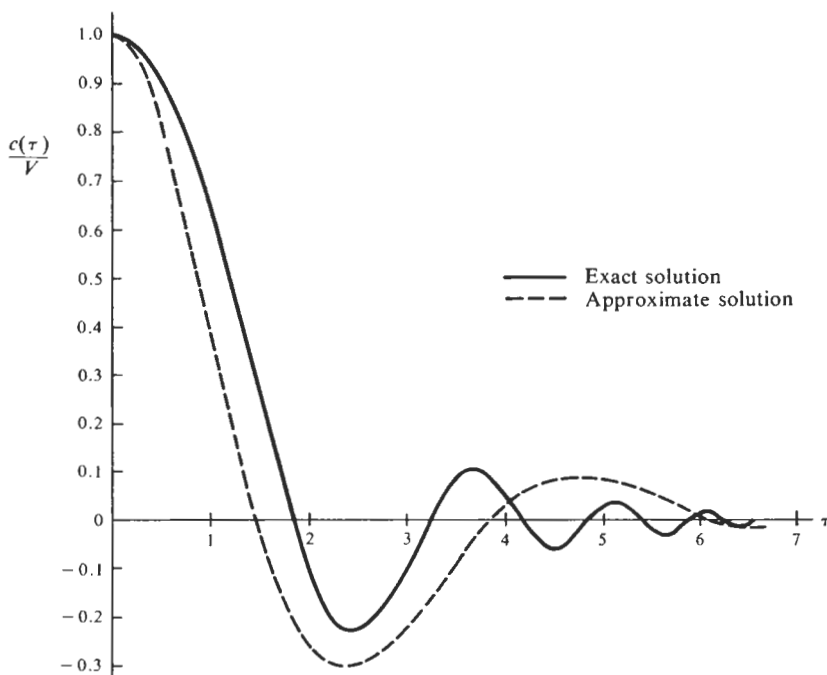


Figure 8.5-3 Solution of Example 8.5-2.

For the specific case in which $c(0) = V$ this becomes

$$C(p) = \frac{p+1}{p^2+p+2} \frac{V}{b}$$

which, upon inversion [keeping in mind the fact that $\mathcal{L}^{-1}\left[C\left(\frac{s}{b}\right)\right] = bc(bt)$], yields

$$c(\tau) = Ve^{-\tau/2} \left(\cos \frac{\sqrt{7}}{2} \tau + \frac{1}{\sqrt{7}} \sin \frac{\sqrt{7}}{2} \tau \right)$$

where $\tau = bt$. This result is plotted in Fig. 8.5-3, along with the exact system response.

The approximate transient response curve is seen to be quite good, considering the degree of ease with which it was obtained. The nature of the response as well as its approximate settling time are reasonably well determined. It must be kept in mind that the oscillatory $c(\tau)$ is the actual nonlinearity input, whereas an exponential was originally assumed for the purpose of quasi-linearization!

Certainly, the EIDF is an extremely simple device to use. For this reason it has significant use as a nonlinear-system design tool in the area of transient response, where only little of design value can be said by other means. On the other hand, it can be expected only to provide *gross* transient response characteristics.

From the way in which it has been formulated, it can be seen that the EIDF quasi-linearized system will always have the dynamics associated with a closed-loop system composed of the original linear elements and an input-sensitive gain factor. Thus, certain transient effects can *not* be observed. This was seen in Example 8.5-2, where the exact solution had a time-variable oscillation frequency. The techniques of Chap. 4 are required to ascertain that level of detail in the transient response. But, if a system designer wanted quick insight into the behavior of this system, the EIDF transient solution coupled with the DF steady-state limit cycle solution [zero (small) amplitude at infinite (high) frequency] surely do provide the required information. System compensation could then be designed on an analytic basis, to be checked by subsequent computer simulation.

REFERENCES

1. Bickart, T. A.: The Exponential Describing Function in the Analysis of Nonlinear Systems, *IEEE Trans. Autom. Control*, vol. AC-11, no. 3 (July, 1966), pp. 491-497.
2. Biernson, G.: A Simple Method for Calculating the Time Response of a System to an Arbitrary Input, *M.I.T. Servomech. Lab. Rept. 7138-R-3*, January, 1954.
3. Chen, K.: Quasi-linearization Techniques for Transient Study of Nonlinear Feedback Control Systems, *AIEE Trans.*, pt. II, *Appl. Ind.*, January, 1956, pp. 354-365.
4. Chen, K.: Quasi-linearization Design of Nonlinear Control Systems, *Proc. JACC*, Minneapolis, Minn., June, 1963, pp. 340-346.
5. Grabbe, E. M., S. Ramo and D. E. Wooldridge: "Handbook of Automation, Computation and Control," John Wiley & Sons, Inc., New York, 1958, chap. 22.
6. Stewart, J. L.: Generalized Padé Approximation, *Proc. IRE*, vol. 48, no. 12 (December, 1960), pp. 2003-2008.
7. Truxal, J. G.: "Automatic Feedback Control System Synthesis," McGraw-Hill Book Company, New York, 1955, chap. 5.

PROBLEMS

- 8-1. Show that for the nonlinear system of Fig. 8-1 the quasi-linearized transient response is

$$c(t) = (R - \delta) \left[1 - \exp \left(-\frac{DK}{R - \delta} t \right) \right]$$

and compare this with the exact transient response.

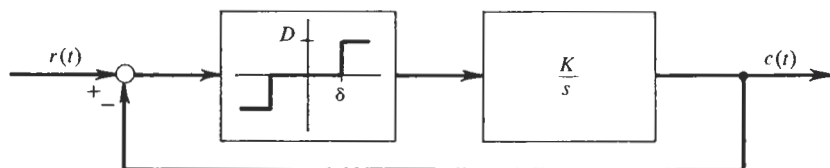


Figure 8-1 Nonlinear relay system.

8-2. Repeat Prob. 8-1 but with the transfer function $K/s(Ts + 1)$ rather than K/s . *Hint:*

$$\text{Determine by expanding an exponential, that } T_A^2 \approx \frac{[2T(R - \delta)]}{DK}.$$

8-3. The input-output specification for the nonlinear system of Fig. 8-2 is given in the form of a desired second-order response, with $\zeta = 0.7$ and $\omega_n = 1$. Design a compensation network $H(s)$ to achieve this response. [*Hint:* Approximate $x(t)$ by straight-line segments.]

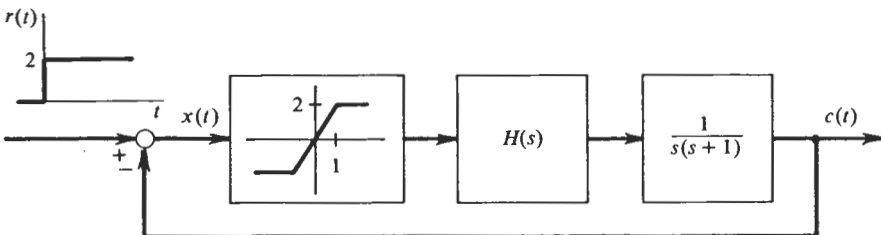


Figure 8-2 Example system for transient response compensation.

8-4. (a) Show that the EIDF technique is not restricted to systems with unity gain feedback links. What configuration restrictions do apply?

(b) Discuss the "filter hypothesis" as it relates to accuracy of the EIDF method.

(c) Consider the use of a more accurate EIDF model input given by $Ee^{-(t-T_d)/\tau}$ where T_d is the delay time of the loop linear elements. In what way does T_d affect calculation of the EIDF? How would you employ this formulation?

(d) Investigate the utility of a more complex model input than that presented in the text, namely, $Ee^{-t/\tau} \cos(\omega t + \alpha)$. Does this lead to a tractable EIDF calculation?

(e) What avenues of approach can you suggest for EIDF method accuracy enhancement?

8-5. Solve Probs. 8-1 to 8-3 using the EIDF method. What can be said about system output stand-off errors in the case of nonlinearities with dead zone?

8-6. Compute the EIDF for the dynamic nonlinearity given by

$$y(\dot{x}, x) = -x\dot{x}$$

How would you use this result in solving for the transient response of a system with linear elements $L(s) = K/[s(s + 1)]$?

9 OSCILLATIONS IN NONLINEAR SAMPLED-DATA SYSTEMS

9.0 INTRODUCTION

All the material of the preceding chapters has been concerned with systems which process signals *continuously* around the loop. This chapter takes under consideration the describing function analysis of nonlinear systems which at some point process *discrete samples* of signals. It should not come as a surprise that the analysis of nonlinear sampled-data systems is more complicated, or at least more laborious, than the corresponding analysis of nonlinear continuous-data systems. The treatment in this chapter considers just bias and single-sinusoid signals present at the input to the nonlinear part of the system. Even in the simplest case of a single sinusoid, the presence of another periodic process in the system—the sampling operation—gives rise to complications of the same kind as are encountered in the study of continuous systems with two sinusoidal components at the nonlinearity input. These complications are significant when the frequencies of the two periodic processes are rationally related, and this is the case of first importance in the study of nonlinear sampled-data systems.

Sampled-data systems have come into practical importance for a variety of reasons. The earliest of these had primarily to do with economy in the design and use of equipment. Many problems in impedance matching or power-level matching can be avoided if critical components are isolated—disconnected—most of the time, and the connection made only briefly at periodic intervals to read out a sample of the signal. The possibility of time-sharing one component among several systems also gives rise to a sampled form of signal processing. A major increase in interest in sampled-data systems was caused by the development of radar systems during the 1940s. Most radars provide information only in the form of periodic samples either because of a periodic scanning process or because of pulsed transmission of the microwave energy. A more recent surge of interest has been due to the increasing utilization of digital computers as controllers in feedback systems. In some areas of application, especially aerospace guidance and control, the use of discrete-data processors is often a practical necessity. Thus many system engineers find themselves concerned almost exclusively with the design of sampled-data systems. And as with continuous systems, these systems may be designed with, or otherwise may suffer from, a number of important nonlinear effects.

THE EFFECTS OF SAMPLING

In this chapter, as in most of the preceding material, attention is centered on systems which can be reduced to single-loop configurations having a single nonlinear part separated from the linear part. The linear part in this case may include any number of continuous linear elements and discrete, or pulsed, linear elements. The ordering of these elements around the loop is of some consequence to the application of describing function theory, because in this case higher-frequency, and possibly lower-frequency, components are generated, not only by the nonlinear part, but by the sampling operations as well. Consider the system configuration of Fig. 9.0-1a. In the study of steady-state oscillations in this system, the nonlinearity input, being the output of the continuous linear filter, may reasonably be taken as a sinusoid for the purpose of quasi-linearization. The output of the nonlinearity, $y(t)$, then contains harmonic components at the fundamental and higher harmonic frequencies. On a two-sided frequency scale, the harmonic components of $y(t)$ would in general have the frequencies $\pm k\omega$, $k = 0, 1, 2, \dots$, where ω is the frequency of the input sinusoid. The sampling operation modulates $y(t)$ with the frequencies $\pm l\omega_s$, $l = 0, 1, 2, \dots$, where ω_s is the frequency of closure of the sampling switch.¹ Thus $y^*(t)$ contains

¹ The reader who needs a basic treatment of the description of the sampling operation, the transfer of sampled signals through linear systems, and z-transform theory is directed to any one of a number of texts on the subject. Among them are Jury (Ref. 7), Kuo (Ref. 15), Raggazzini and Franklin (Ref. 22), and Tou (Ref. 27).

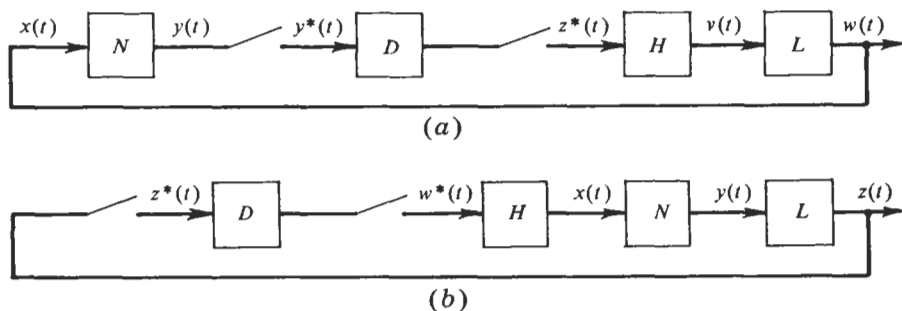


Figure 9.0-1 Nonlinear sampled-data system configurations. N = static nonlinear element; H = data hold; D = discrete linear element; L = continuous linear element. Asterisks denote sampled signals.

harmonic components with frequencies $\pm k\omega \pm l\omega_s$. These frequencies, which appear in the loop, determine to a large extent the possibility of successful application of describing function theory.

1. If the frequency ratio ω/ω_s is irrational, $y^*(t)$ is aperiodic. It contains a harmonic component with frequency ω ; in fact, that component is just $1/T_s$ times the fundamental component of the nonlinearity output. This may be seen from the familiar expression for the transform of $y^*(t)$:

$$Y^*(j\omega) = \frac{1}{T_s} \sum_{l=-\infty}^{\infty} Y[j(\omega + l\omega_s)] \quad (9.0-1)$$

If ω_s is not rationally related to ω , the only term in this sum with frequency ω is the primary term for $l = 0$. Thus, if describing function theory can be applied at all in the case of irrational frequencies, the describing function relating $x(t)$ to $y^*(t)$ is just $1/T_s$ times the ordinary single-sinusoid-input describing function for the nonlinearity which relates $x(t)$ to $y(t)$. The question of applicability is raised because $y^*(t)$ in this case may very well contain harmonic components with frequencies *lower* than ω . These components cannot be discarded on the basis of the filter hypothesis. The low-frequency components in $y^*(t)$ are due to higher harmonics of $y(t)$ which lie close to $l\omega_s$, and thus are modulated to frequencies near zero. Describing function theory would then seem to be applicable only if ω is so small that the effect of the sampling on the operation of the system is trivial.

2. If the frequency ratio ω/ω_s is rational, $\omega/\omega_s = m/n$, $y^*(t)$ is periodic with a frequency which is an integral multiple of ω/m . Again $y^*(t)$ may

contain harmonic components with lower frequencies than ω , and essentially the same comments about applicability of describing function theory made in regard to irrational frequencies are pertinent in this case.

3. If the frequency ratio ω/ω_s is a whole fraction, $\omega/\omega_s = 1/n$, $y^*(t)$ is periodic with frequency ω . It contains no harmonic component with frequency lower than ω , except possibly for a dc component, if $\omega < \frac{1}{2}\omega_s$. With the possible effects of a dc component taken into consideration, these are the right conditions for applicability of describing function theory, and the remainder of the chapter is restricted to this case. Fortunately, this includes a most important class of problems since the limit cycles which are most commonly observed in sampled nonlinear systems have periods which are whole multiples of the sampling period. It is important to observe that the component of frequency ω in $y^*(t)$ is not just $1/T_s$ times the corresponding component of $y(t)$. Thus it would be quite incorrect to employ a describing function which relates $x(t)$ to $y(t)$ and ignores the higher harmonics of a signal which is being sampled. *It is essential that the describing function characterize directly the relation between $x(t)$ and $y^*(t)$.*

As another illustration, consider the system configuration of Fig. 9.0-1b. In this case the higher harmonics in $y(t)$ are attenuated by the continuous linear filter before being sampled. Thus the modulating effect of the sampler on these harmonics may be of little importance. The greater question in this case is whether the input to the nonlinearity can be assumed a sinusoid. The hold does not provide very complete filtering of the high-frequency content of the sampled signal. Thus, unless there were additional filtering in the position of the hold, it might be necessary to characterize the transfer from $z(t)$ to $y(t)$ by a describing function—a task which promises to be laborious.

The following sections deal with the determination of and stability of limit cycle modes in sampled nonlinear systems, where the limit cycles tested have periods which are whole multiples of the sampling period. These are not the only limit cycles which may be possible in such systems but experience with both real and simulated systems has shown these to be by far the most commonly occurring modes. This does not exhaust the usefulness of describing function theory in its application to sampled nonlinear systems. But other applications, such as the study of forced sinusoidal response, must be considered carefully in each individual case because of the possibility of lower-frequency components, as discussed above under irrationally related frequencies, and the possible existence of limit cycle modes in addition to the forced response. A very important special-case system which can be dealt with by a simple extension of previous techniques is treated in the following section. Then, in the next, we turn to the study of limit cycles in more general systems.

9.1 LIMIT CYCLES IN SAMPLED TWO-LEVEL RELAY SYSTEMS

The material presented in Sec. 9.2 is readily applicable to the study of limit cycles in two-level relay systems, but these systems are of such importance that it seems worth exploiting the simpler approach which is possible in this case.

The configuration of the system is shown in Fig. 9.1-1. The two-level relay is shown as having possible hysteresis. A zero-order hold is considered to follow the sampling switch. A great many systems do employ a zero-order hold or simple clamp which clamps a sampled signal to a constant over the following sampling period. This analysis is not, however, limited to such systems. If the actual system does not include a hold, or uses a higher-order hold, the transfer function of that hold is included in the linear part, as shown in Fig. 9.1-1, along with the reciprocal of the zero-order-hold transfer function. The linear part may include any number of continuous and discrete linear elements.

THE DESCRIBING FUNCTION

One has free choice in deciding how much of the system to characterize with a describing function, so long as the nonlinear part is included. The analysis of this system is most like the analysis of continuous systems considered heretofore, if one chooses to represent the effect of the nonlinearity, the sampling switch, and the zero-order hold by a describing function. To this end, $x(t)$ is taken to be a sinusoid, unbiased to begin with, and the fundamental harmonic component of $z(t)$ is calculated. The frequencies we shall consider, according to the discussion of the preceding section, are whole fractions of the sampling frequency ω_s . Moreover, we shall center attention on the even fractions, $\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \dots$, since these are the limit cycle modes one might expect to see in the very common case in which the linear part of the system includes a pole at the origin, an integration. In that case $z(t)$ must be an unbiased function in any steady-state limit cycle with no input to the

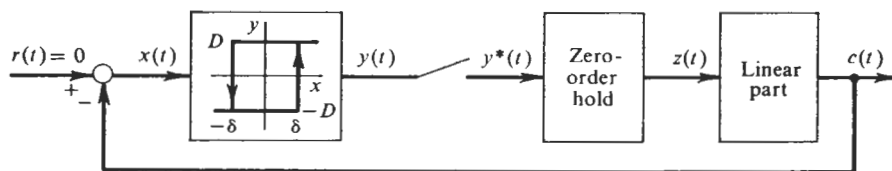


Figure 9.1-1 Two-level relay system configuration.

system. The drive signal into the linear part will then consist of a periodic cycle which includes an equal number of sampling periods of plus and minus drive. The only arrangement of these periods of plus and minus drive which is consistent with the sinusoid assumed as the input to the nonlinearity is n positive drive periods followed by n negative drive periods in the case of a cycle with period $2nT_s$, T_s being the sampling period. Such a cycle will be termed an n, n mode.

The input and output waveforms for the 2, 2 mode are shown in Fig. 9.1-2. $x(t)$ is a sinusoid with period $4T_s$, and $z(t)$ is a square wave with that period. The output of the hold, $y(t)$, is shown lagging the output of the nonlinearity, $z(t)$, because $y(t)$ is not in phase with the sampling points. The lag between the zero crossing of $y(t)$ and the next sampling point is not known a priori; evidently it can take any value between 0 and T_s in time or 0 and π/n in phase angle. The amplitude of the fundamental harmonic of $z(t)$ is $(4/\pi)D$, and the phase lag of that component relative to $x(t)$ is $\sin^{-1}(\delta/A) + \varphi$, where φ is the sampling lag. The describing function for the chain of elements—nonlinearity, sampling switch, and hold—is then

$$N(A, \varphi) = \frac{4D}{\pi A} \angle -\sin^{-1}\left(\frac{\delta}{A}\right) - \varphi \quad 0 < \varphi < \frac{\pi}{n}, A > \delta \quad (9.1-1)$$

This expression holds for an n, n mode of any order.

THE LINEAR PART

The remainder of the system, the linear part as shown in Fig. 9.1-1, is now characterized by its steady-state sinusoidal response at the frequency $(1/2n)\omega_s$.

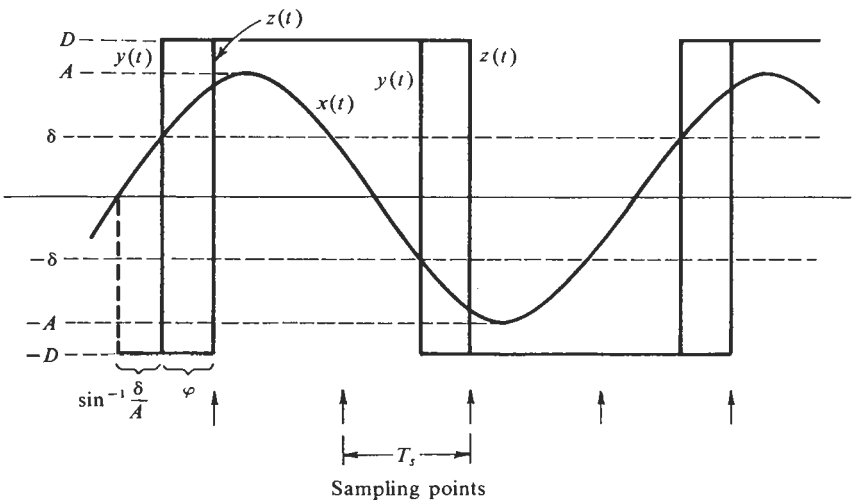


Figure 9.1-2 Signal waveforms for the 2, 2 mode.

If this is a continuous linear operator, the only requirement for applicability of describing function theory is that it attenuate the higher harmonics of $z(t)$ sufficiently to return essentially the fundamental sinusoid to $x(t)$. If this includes samplers and discrete linear elements, a more restrictive condition is placed on them. Consider the linear elements in Fig. 9.1-3. In this example, L_1 and L_2 represent continuous linear filters, whereas D represents a discrete linear filter. It may be mechanized as a pulsed analog filter, or perhaps as a digital computer solving a linear difference equation. L_1 and/or L_2 may include data holds. We wish to find the sinusoidal component of frequency ω in the steady-state output of this chain when the input is periodic with that frequency.

$$\begin{aligned} C(j\omega) &= L_2(j\omega)V^*(j\omega) \\ &= L_2(j\omega)D(j\omega)W^*(j\omega) \end{aligned} \quad (9.1-2)$$

But $W(j\omega) = L_1(j\omega)Z(j\omega)$

and
$$W^*(j\omega) = \frac{1}{T_s} \sum_{l=-\infty}^{\infty} L_1[j(\omega + l\omega_s)]Z[j(\omega + l\omega_s)]$$

so
$$C(j\omega) = \frac{1}{T_s} L_2(j\omega)D(j\omega) \sum_{l=-\infty}^{\infty} L_1[j(\omega + l\omega_s)]Z[j(\omega + l\omega_s)] \quad (9.1-3)$$

From this expression one can see that if $z(t)$ were just a sinusoid of frequency $\omega < \frac{1}{2}\omega_s$, none of the complementary components of $w^*(t)$ would have the frequency ω . The only component of $c(t)$ which would have the frequency ω is that due to the $l = 0$ term in the sum of Eq. (9.1-3). Thus the sinusoidal response function representing the fundamental transfer through the chain of elements in Fig. 9.1-3 would be just $(1/T_s)L_1(j\omega)D(j\omega)L_2(j\omega)$.

This result is complicated, however, by the fact that, in the system under study, $z(t)$ is a periodic function which includes harmonics in addition to the fundamental component. With ω an even fraction of ω_s , some of the odd harmonics of $z(t)$ are modulated to additional components of $w^*(t)$ at the frequency ω . The harmonics which contribute to the fundamental component after sampling are those with frequencies equal to $l\omega_s \pm \omega$, for all integers l . The effect of any significant contributors could be included by calculating the harmonics of $z(t)$, passing them through $L_1(jk\omega)$ (k is the order of the harmonic), and adding the term in Eq. (9.1-3). However, for

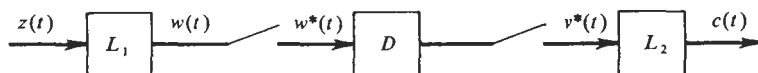


Figure 9.1-3 An example linear part.

simple application of the theory of this section, we must insist that the harmonic content of $z(t)$ be sufficiently attenuated before it reaches the first sampler in the linear part. An alternative procedure to treat this problem is discussed in the following section.

DETERMINATION OF MODES

Having a describing function $N(A, \varphi)$ to characterize the nonlinear part, sampler, and hold, and a steady-state sinusoidal response function $L(j\omega)$ to characterize the remaining linear part, the condition for the possible existence of a limit cycle mode is, as always,

$$1 + N(A, \varphi)L(j\omega) = 0 \quad (9.1-4)$$

The gain-phase plot is a convenient means of displaying the solutions to this equation. Although only the frequencies $(1/2n)\omega_s$ are of interest, it is often useful—especially for the later design of compensation—to plot the complete $L(j\omega)$ function. A typical curve is shown in Fig. 9.1-4. Solutions of Eq. (9.1-4) are represented by intersections of this curve of $L(j\omega)$ with $-1/N(A, \varphi)$. $N(A, \varphi)$, in this case, is defined only for the discrete set of frequencies $(1/2n)\omega_s$, and it depends both on A and n , the order of the mode. It is convenient to separate that part of N which depends only on A from that which depends on n for simplicity in plotting the function. Thus define

$$N'(A) = \frac{4D}{\pi A} \begin{cases} \text{---} \\ \text{---} \end{cases} -\sin^{-1} \frac{\delta}{A} \quad (9.1-5)$$

which is the describing function given in Eq. (9.1-1), except for the sampling lag φ . This φ may take any value in the range $(0, \pi/n)$, and since this range depends on n —and thus ω —the bands of possible sampling lags can conveniently be shown as lines originating at the point of $L(j\omega)$ for each frequency and extending a distance corresponding to π/n . With the sampling lag accounted for separately in this manner, the describing function which is plotted, $-1/N'(A)$, is nothing more than the describing function for the two-level relay with hysteresis as it appears in continuous systems.

The completed plot (Fig. 9.1-4) indicates all possible limit cycle modes. For the typical case shown, the only intersections of $L(j\omega)$ plus the sampling lag with $-1/N'(A)$ occur at the frequencies $\frac{1}{8}\omega_s$ and $\frac{1}{4}\omega_s$. Thus only the 3, 3 and 4, 4 modes are possible in this case. The higher-frequency modes are not possible because the nonlinearity and linear part have too much phase lag even if the sampling action contributes none, and the lower-frequency modes are not possible because even with the maximum possible sampling delay, the sinusoidal signal does not accumulate 360° of phase lag around the loop. The intersections indicating the possible modes are encircled in the figure. The frequency of each mode is indicated on the scale of $L(j\omega)$ at that

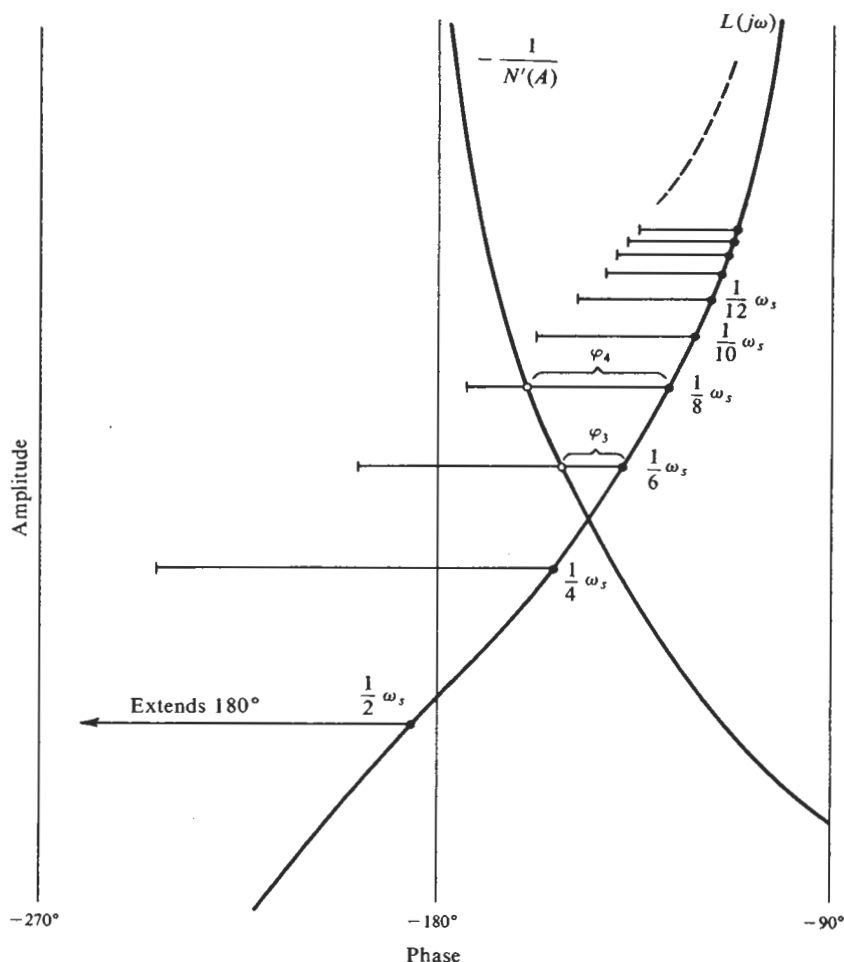


Figure 9.1-4 Gain-phase plot for sampled two-level relay system.

point; the amplitude of each mode at the input to the nonlinearity is indicated on the scale of $-1/N'(A)$ at that point; and the phase lag due to the sampling delay in each mode is indicated by the phase difference between $L(j\omega)$ and $-1/N'(A)$ at those points.

Notice that for frequencies much smaller than the sampling frequency, the possible limit cycle frequencies become closely spaced and the maximum sampling lag is small. In this low-frequency region the sampling has little effect on the behavior of the system.

When more than one limit cycle mode is possible, the limit cycle which will be observed depends on the prior history of the system variables. If the modes are stable (this matter is discussed in a later section), there is some

region of initial conditions from which the system will settle into each mode. These regions are often much smaller for some modes than for others, and so some modes are more likely to occur than others. In any case, a system must be designed so that all possible modes are acceptable. If one or more indicated modes are not acceptable, usually because the amplitude at some point in the system is too large, these modes must be eliminated by compensation.

DESIGN OF COMPENSATION

It is at this point in the design of systems that one of the great advantages of the use of describing functions becomes apparent. In most instances the compensation required to improve those performance characteristics which can be evaluated by the use of describing functions is quite evident. So it is in this instance. If the 4, 4 mode as indicated on Fig. 9.1-4 has an unacceptably large amplitude, it can be eliminated by providing at least $(45 - \varphi_4)$ deg of phase lead at the frequency $\frac{1}{8}\omega_s$ somewhere around the loop. After such compensation, the 3, 3 mode will still be possible, and in all likelihood one or both of the higher-frequency modes as well. If the 3, 3 mode is to be eliminated also, the compensation is designed to provide at least $(45 - \varphi_4)$ deg of lead at $\frac{1}{8}\omega_s$ and at least $(60 - \varphi_3)$ deg of lead at $\frac{1}{6}\omega_s$. The amplitudes of the remaining possible modes at various stations around the loop will differ with the location of the compensation, and this can be evaluated using just steady-state frequency response characteristics.

This linear compensation can be implemented with either a continuous or discrete compensator. Alternatively, the design problem might be to find that value of hysteresis, δ , in the switching characteristic which will allow only modes up to order n , for some specified value of n . In this case, too, the answer to the problem is fairly evident, using describing function theory. For the case pictured in Fig. 9.1-4, the 4, 4 mode can be eliminated by reducing δ , but other modes would remain possible since, in the limit as $\delta \rightarrow 0$, the curve of $-1/N'(A)$ extends along the entire phase = -180° line.

BIAS OFFSET

In the foregoing discussion, the limit cycle modes have been considered unbiased sinusoids. The fact that some systems may sustain a dc offset around part of the loop is another important difference between sampled and continuous systems. If the linear part of the system does not include an integrator, a steady-state bias signal could exist only if such a signal could regenerate itself when propagated around the loop. But this will not be possible with this nonlinearity and ordinary linear parts because a positive bias in $x(t)$ (refer to Fig. 9.1-1) will cause a positive bias, if any, in $z(t)$, and

this will be transferred through the linear part to a positive bias in $c(t)$. But this is inconsistent with the assumed positive bias in $x(t)$. Thus, if the system is stable to the low-frequency signal in the presence of the limit cycle, the bias will decay to zero at all stations around the loop.

If the linear part includes an integrator, the input to it, $z(t)$ in Fig. 9.1-1, must be unbiased in any steady-state mode with no input to the system. This does not, however, preclude a bias in $c(t)$ and $x(t)$. Any bias in these signals is possible so long as the output of the nonlinearity remains unbiased. The range of possible bias offsets can be seen in Fig. 9.1-5, where the signal waveforms through the nonlinearity and hold have been drawn for the 3, 3 mode of the system of Fig. 9.1-4, and the proper sampling lag, φ_3 , taken from that figure, is shown. The points in time at which the samples are taken are shown as dots on the $x(t)$ waveform. It is clear that the $x(t)$ curve could be shifted up or down by a small bias without changing the output $z(t)$ at all. The range of this possible offset, to which the system is insensitive, is determined by noting how far $x(t)$ can be shifted without changing $y(t)$ (see Fig. 9.1-1) at the sampling points. For the case shown in Fig. 9.1-5, it is the sampling points at which $z(t)$ switches which are critical, and the range of possible bias offset is indicated. In other cases, it may be one of the other sampled points which first causes a change in $z(t)$. The bias range shown in the figure is that of $x(t)$, the input to the nonlinearity. This range can be reflected to other signals in the linear part of the system, using the dc gain of the system between the two points. The signals between $z(t)$ and the integrator in the linear part remain unbiased.

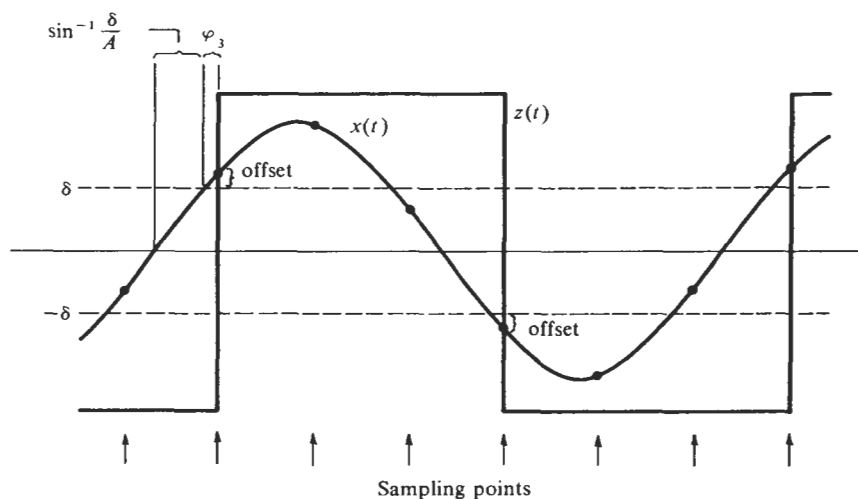


Figure 9.1-5 Nonlinearity input and output showing range of possible bias offset.

Compensation designed to improve limit cycle moding will affect the range of bias offset as well, sometimes adversely. In some systems, the possibility of a dc error which the system does not respond to is of greater consequence than the possible limit cycle modes. If the bias range is unacceptably large, compensation must be designed to reduce it. This is done by introducing (if possible) a larger dc gain relative to the limit cycle gain between the error point and the input to the nonlinearity. This can be achieved by a conventional lag-lead compensator which has high gain at low frequencies and unity gain with negligible phase lag at the limit cycle frequencies. An integrator with a bypass can eliminate the dc offset at the error point altogether.

9.2 LIMIT CYCLES IN OTHER SAMPLED NONLINEAR SYSTEMS

From the experience of the preceding section we observe some of the important characteristics of a describing function analysis of sampled nonlinear systems. The describing function was found to depend not only on the amplitude of the sinusoid assumed at the input to the nonlinearity, but also on the ratio of the sinusoidal frequency to the sampling frequency and the phase of the sampling points relative to the input sinusoid. The frequency-ratio and phase-angle dependencies were also found in the two-sinusoid-input describing function for continuous systems in the case of rationally related frequencies. The present situation is somewhat analogous to that: again, there are two periodic processes operative in the system, and their periods are rationally related. Instead of two periodic signals, however, the periodic processes in this case are one sinusoidal signal and the periodic sampling process.

Analysis of limit cycles in sampled two-level relay systems was found to be not much more difficult than the corresponding analysis of continuous two-level relay systems. This is due to the simple way in which the frequency-ratio and phase-angle dependencies enter that problem. The nonlinearity output in that case is known to be a square wave if it is nontrivial. Thus the fundamental amplitude of the output is independent of the frequency ratio or phase angle. The sampling phase simply adds directly to the phase angle of the describing function, and the frequency ratio plays no role other than to prescribe the bound on possible sampling lag.

For other nonlinearities, the basic concepts remain unchanged, but the details are more complicated because the whole waveform of the nonlinearity output changes with sampling phase angle. This complicates considerably the calculation of the describing function for a sampled nonlinearity for all input amplitudes, frequency ratios, and phase angles relative to the sampling points. However, once the computation is done and the

results graphed, the use of the describing function to analyze system limit cycles and to design compensation to meet specifications on possible limit cycle modes is just as easy and meaningful as it is for continuous systems.

TWO POINTS OF VIEW

The general system configuration under consideration is shown in Fig. 9.2-1. The configuration is characterized by a single loop with a single separable nonlinear part. The linear parts L_1 , L_2 , and L_3 provide continuous linear filtering, and may include samplers and discrete linear elements as well. The major requirement for applicability of describing function theory is that the signal which is assumed sinusoidal, $x(t)$ in this case, must indeed approximate that waveform. This means, in the case of sampled systems, that not only the harmonic content of the nonlinearity output, but also the harmonic content due to all samplers in the system, must be adequately attenuated in the linear filter which returns the signal to the nonlinearity input.

In addition to any samplers which may be included in the linear parts, one is shown specifically following the nonlinearity. It is a sampler such as this, operating directly at the input or output of a nonlinearity, which gives rise to the greatest difference between continuous- and sampled-nonlinear-system operation. For the system as shown, it would be a very poor approximation to use an ordinary describing function to characterize the transfer from $x(t)$ to $y(t)$, and characterize the rest of the system by its steady-state sinusoidal response at the fundamental frequency. The major error in that approach would be the neglect of the higher harmonics of $y(t)$ in determining the fundamental component of $y^*(t)$. The sampling operation modulates some of the higher harmonics of $y(t)$ down to fundamental-frequency components of $y^*(t)$. Thus any reasonable describing function approach must model directly the transfer from $x(t)$ to $y^*(t)$. A similar situation would hold if the sampler preceded the nonlinearity, as shown in Fig. 9.2-2a. In that case, the hold does not filter the harmonics because of the sampling operation sufficiently to justify the assumption of a sinusoid at $y(t)$, the input to the nonlinearity. Thus the describing function must be defined to characterize

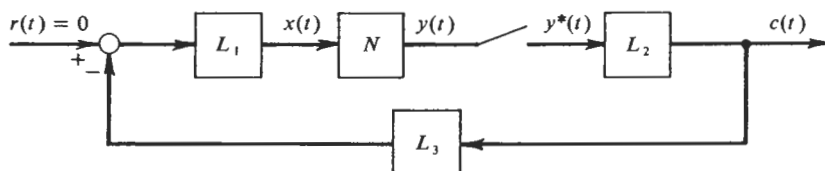


Figure 9.2-1 The general sampled-nonlinear-system configuration considered in this section. L_1 , L_2 , L_3 may include additional samplers and both continuous and discrete linear elements.

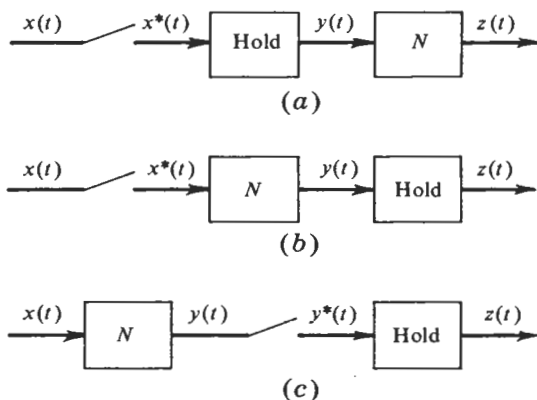


Figure 9.2-2 Equivalent arrangements for static, single-valued nonlinearities.

directly the transfer from $x(t)$ to $z(t)$. If the nonlinearity is static and single-valued so that the output depends only on the current value of the input, the three orderings of sampler, hold, and nonlinearity shown in Fig. 9.2-2 are equivalent. This implies that in arrangement *b* the nonlinear operation is carried out on the impulses at the input to yield impulses at the output whose areas or strengths are related to those at the input by the given nonlinear function. If the nonlinearity were dynamic or multiple-valued, the order of the arrangement would matter, and the describing function would have to be calculated for the given arrangement. For the present we shall refer to arrangement *c*, and include the hold in the linear part of the system. This is the arrangement shown in Fig. 9.2-1.

Two different approaches to the describing function analysis of systems such as that of Fig. 9.2-1 have been set forth. One is the same approach taken in the preceding section as applied to two-level relay systems. Chow (Ref. 2) documented this point of view, which was being developed by Russell (Ref. 23) at the same time. They defined a describing function for a nonlinearity, sampler, and zero-order hold in the classical way—the amplitude and phase relation between an assumed sinusoid at the input and the fundamental component of the periodic output. The use of the hold in the chain of elements for which the describing function was defined is not essential to this approach. Referring again to Fig. 9.2-1, one can better calculate a describing function to characterize just the transfer from $x(t)$ to $y^*(t)$ using the same point of view—representation of the fundamental harmonic response of the nonlinearity and sampler. This has the advantage of greater generality since it is applicable to the sampled nonlinearity regardless of what kind of hold, if any, is used in the system. If a hold does indeed follow the sampler, it is then included in the linear part of the system. As a lesser advantage, it may be noted that the calculation of the fundamental

component of $y^*(t)$ is particularly simple since that function is just a sequence of impulses. The describing function which relates an assumed sinusoid at the input to a nonlinearity to the fundamental component of the sampled output of the nonlinearity will here be termed a *sampled describing function*.

$$N(A, \varphi) = \text{sampled describing function} \\ = \frac{\text{phasor representation of fundamental component of } y^*(t)}{\text{phasor representation of } x(t)}$$

With the transfer from $x(t)$ to the fundamental component of $y^*(t)$ in Fig. 9.2-1 given by the sampled describing function, the linear part of the system is characterized by its steady-state sinusoidal response at the fundamental frequency.

Another approach to this problem was documented by Kuo (Refs. 14, 15). With the input to the nonlinearity, $x(t)$, assumed to be a sinusoid of a frequency which is a whole fraction of the sampling frequency, the sampled output of the nonlinearity, $y^*(t)$, is a periodic sequence of impulses with the same period as the input. The z transform of the output sequence can be written explicitly, and the ratio of that z transform to the z transform of the input sinusoid is defined by Kuo to be the *z -transform describing function* for the nonlinearity and sampler.

$$N^*(A, \varphi) = z\text{-transform describing function} \\ = \left[\frac{z \text{ transform of } y^*(t)}{z \text{ transform of } x(t)} \right]_{\text{Fundamental frequency}}$$

With the relation between samples of $x(t)$ to samples of $y(t)$ defined by the z -transform describing function, the linear part of the system is characterized by its sampled transfer function, and the equation of loop closure which defines a possible limit cycle mode is evaluated at the fundamental frequency.

These two points of view are basically different. The difference in results using the two approaches is trivial in some instances and quite significant in others. If one approach were superior on all counts, the other could be discarded at once. But this is not the case; so the control engineer should keep both techniques in his bag of analytic tools and understand when to use each one. The major advantage of the first approach, the sampled describing function, is its simplicity. The analytic manipulations involved, for example, in determining the range of system parameters for which a particular limit cycle mode is possible, are considerably simpler using the sampled describing function approach. This advantage in simplicity is due to the use of the ordinary sinusoidal response function for the linear part, rather than the sampled transfer function. For this same reason, the design of compensation to meet specifications on limit cycle modes is obvious when the sinusoidal

response function is used, but obscure when a new sampled transfer function must be calculated for the cascaded compensation and original linear part. The major advantage of the second approach, the z -transform describing function, is its better accuracy in some situations. Using this technique, the exact pulse sequence at the nonlinearity output is processed exactly through the linear part using z transforms. This can be done without approximation for any linear configuration of continuous elements, samplers, and discrete elements. The result is the z transform of the signal fed back to the nonlinearity input. This transform contains terms which define the fundamental frequency component of the fed-back samples, ripple terms, and the normal modes of the linear part of the system. This function is clearly not compatible with the simple sinusoid originally assumed at the nonlinearity input. But the process of equating the transform of the fed-back function to the transform of the originally assumed sinusoid *when both are evaluated at the fundamental frequency* serves to select just the fundamental component of the fed-back samples and equate it to the assumed sinusoid. Thus the sequence of samples of the fed-back signal is being accurately determined in this case, and it is the fundamental harmonic component of this exact pulse train which is being equated to the sinusoid originally assumed at the nonlinearity input.

By contrast, using the sampled describing function approach, the harmonics of $y^*(t)$ are dropped immediately, and only the fundamental component is passed through the linear part. If the linear part contains no additional samplers, the fundamental component of the signal fed back to the nonlinearity input is correctly determined by this procedure, and the only difference between the two techniques is the difference between the fundamental component of the continuous signal fed back to the nonlinearity and the fundamental component of the samples of that signal. Since the fundamental component of the samples of a sinusoid, for $\omega_s > 2\omega$, is just $1/T_s$ times the sinusoid itself, it can be seen that the two techniques must give very nearly the same results if the assumption which is common to both is well satisfied, namely, that the signal fed back to the nonlinearity closely approximates a sinusoid. In that case, the sampled describing function is to be preferred because of its simplicity. If the linear part fails to filter the fed-back signal well enough, the fundamental component of the samples of the fed-back signal may differ from the fundamental component of the continuous fed-back signal, and in that case the z -transform describing function is likely to give the more nearly correct answer because it recognizes the fed-back signal only at the sampling instants, as does the nonlinearity. The fundamental component of the continuous fed-back signal, on the other hand, is influenced by the shape of the signal between the sampling points; and this is of no consequence to the nonlinearity. In this case of inadequate filtering, however, the use of describing function theory by either method is risky.

When the linear part contains additional samplers and discrete linear elements, the case for the z -transform describing function is stronger. As noted above, the samples of the fed-back signal can be determined exactly, no matter what the configuration of the linear part. Dropping the higher harmonics of $y^*(t)$, as is done when using the sampled describing function, can be serious in this case because a sampler in the linear part will modulate some of these harmonics down to additional contributions to the fundamental frequency component. Thus, with additional samplers in the linear part, propagation of only the fundamental component of $y^*(t)$ through the sinusoidal response function for the linear part does not correctly determine the fundamental component of the fed-back signal at the nonlinearity input. The approximation will be good if the harmonics of $y^*(t)$ are well filtered before the signal encounters a sampler, but this is a requirement which need not be met for successful use of the z -transform describing function.

A basic limitation of the z -transform describing function should also be noted. Using that technique, one only processes information about the samples of the signals circulating in the system. Sinusoidal signals are defined by their samples in the sense that the fundamental component of the sampled sinusoid is a constant times the sinusoid itself only if the sampling frequency is greater than twice the frequency of the sinusoid. This means that the very important case of limit cycle modes which have a frequency just $\frac{1}{2}$ the sampling frequency cannot be analyzed by the z -transform describing function method.

Implementation of these two points of view will be clarified by an example analyzed by both methods.

Example 9.2-1 The system of Fig. 9.2-3 uses an ideal two-level relay as a controller. It drives a digital integrator which implements the rectangular-rule integration formula

$$v(kT_s) = v[(k-1)T_s] + y[(k-1)T_s]T_s \quad (9.2-1)$$

which has the z transform

$$V^*(z) = z^{-1}V^*(z) + z^{-1}Y^*(z)T_s$$

or

$$D^*(z) = \frac{V^*(z)}{Y^*(z)} = \frac{T_s}{z-1} \quad (9.2-2)$$

The output of the integrator is held constant between sampling instants, and the held signal feeds back through a continuous linear filter, which we take to be

$$L(s) = \frac{K}{\tau s + 1} \quad (9.2-3)$$

This example is deliberately chosen to have poor filtering of the fed-back signal so that the difference between the two describing function methods will be dramatized.

Determine the conditions under which the 2, 2 limit cycle mode is possible in this system.

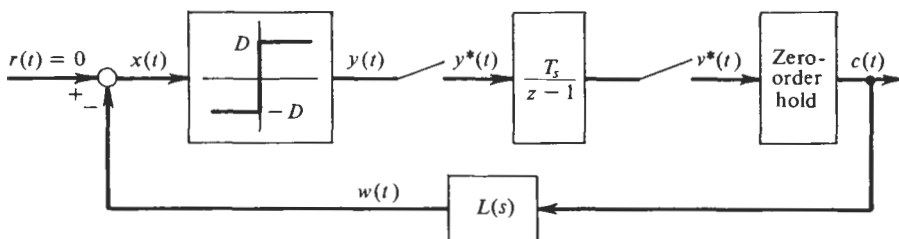


Fig. 9.2-3 System configuration for Example 9.2-1.

First solution The phase relation between the sinusoid assumed at $x(t)$ and the sampling points may vary between 0 and T_s in time or 0 and $\pi/2$ radians in angle. Take the time origin to coincide with one of the sampling points. Then, with respect to this origin, $x(t)$ is advanced in phase by an angle which may range between 0 and 90° .

$$x(t) = A \sin \left(\frac{\omega_s}{4} t + \varphi \right) \quad 0 < \varphi < 90^\circ \quad (9.2-4)$$

$y(t)$ is then a square wave with amplitude $\pm D$, and $y^*(t)$ a periodic sequence of impulses with strengths $\pm D$, as shown in Fig. 9.2-4. The fundamental component of $y^*(t)$, shown dashed in that figure, is obviously advanced in phase by 45° with respect to the time origin, and its amplitude can be computed directly as the amplitude of the fundamental frequency sine component having that phase angle.

$$\begin{aligned} |y_r^*(t)| &= \frac{2}{2T_s} \int_0^{2T_s} y^*(t) \sin \left(\frac{\omega_s}{4} t + 45^\circ \right) dt \\ &= \frac{1}{T_s} (D \sin 45^\circ + D \sin 135^\circ) \\ &= \sqrt{2} \frac{D}{T_s} \end{aligned} \quad (9.2-5)$$

In the calculation of the fundamental component of a signal which consists of or includes impulse functions, impulses should not be included at both end points of the interval of integration. The integration can include the impulse at either end point, but not both. From Eqs. (9.2-4) and (9.2-5), the sampled describing function for the ideal two-level relay in the case of the 2, 2 mode is seen to be

$$N(A, \varphi) = \frac{\sqrt{2}}{T_s} \frac{D}{A} \angle 45^\circ - \varphi \quad 0 < \varphi < 90^\circ \quad (9.2-6)$$

The sinusoidal response function for the linear part of this system is determined by the following calculation:

$$\begin{aligned} W(j\omega) &= L(j\omega)H(j\omega)V^*(j\omega) \\ &= L(j\omega)H(j\omega)D^*(j\omega)Y^*(j\omega) \end{aligned} \quad (9.2-7)$$

where $H(j\omega)$ is the sinusoidal response function for the zero-order hold. For this linear part, the fundamental component of $w(t)$ is correctly given by the fundamental component of $y^*(t)$, modified by the sinusoidal response function $D^*(j\omega)H(j\omega)L(j\omega)$, evaluated at the

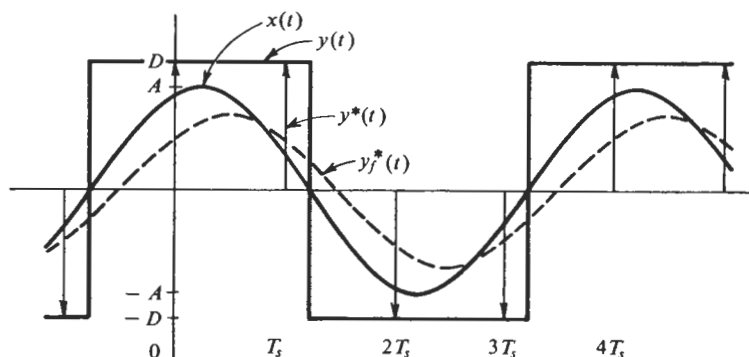


Figure 9.2-4 Nonlinearity input and output for Example 9.2-1.

frequency $\omega = \omega_s/4$. If the nonlinearity and sampler were followed by a continuous linear element, and then a sampler and discrete linear element, this would not be true. In evaluating $D^*(z)$ at $\omega = \omega_s/4$, we note $z = e^{sT_s} \rightarrow e^{j(1/4)\omega_s T_s} = j$. Also, the zero-order hold has the sinusoidal response function

$$H(j\omega) = \frac{1 - e^{-j\omega T_s}}{j\omega} \quad (9.2-8)$$

$$\begin{aligned} \text{so} \quad D^*HL\left(j\frac{\omega_s}{4}\right) &= \frac{T_s}{j-1} \frac{1+j}{j\frac{\pi}{2}\frac{1}{T_s}} \cdot \frac{K \angle -\tan^{-1} \frac{\pi}{2} \frac{\tau}{T_s}}{\sqrt{1 + \left(\frac{\pi}{2} \frac{\tau}{T_s}\right)^2}} \\ &= \frac{2}{\pi} T_s^2 \frac{K}{\sqrt{1 + \left(\frac{\pi}{2} \frac{\tau}{T_s}\right)^2}} \angle -180^\circ - \tan^{-1} \frac{\pi}{2} \frac{\tau}{T_s} \end{aligned} \quad (9.2-9)$$

Equating the fundamental component of $w(t)$ to $-x(t)$ requires

$$ND^*HL = -1$$

$$\text{or} \quad \frac{2\sqrt{2}}{\pi} \frac{KDT_s}{A} \frac{1}{\sqrt{1 + \left(\frac{\pi}{2} \frac{\tau}{T_s}\right)^2}} \angle 45^\circ - 180^\circ - \varphi - \tan^{-1} \frac{\pi}{2} \frac{\tau}{T_s} = 1 \angle -180^\circ$$

Solving for A and φ , we find

$$A = \frac{2\sqrt{2}}{\pi} \frac{KDT_s}{\sqrt{1 + \left(\frac{\pi}{2} \frac{\tau}{T_s}\right)^2}} \quad (9.2-10)$$

$$\varphi = 45^\circ - \tan^{-1} \frac{\pi}{2} \frac{\tau}{T_s} \quad (9.2-11)$$

Since φ must lie in the range $(0, 90^\circ)$, Eq. (9.2-11) can be satisfied for all values of τ/T_s in the range

$$-\frac{2}{\pi} < \frac{\tau}{T_s} < \frac{2}{\pi}$$

It will later be determined that the indicated limit cycle mode is stable only if the open-loop system is stable. Thus one would conclude from application of sampled describing function theory that the 2, 2 limit cycle mode may exist if $(0 < \tau/T_s < 2/\pi)$.

Second solution For application of z -transform describing function theory to this problem we first write the z transform of $x(t)$.

$$x(t) = A \cos \varphi \sin \frac{\omega_s}{4} t + A \sin \varphi \cos \frac{\omega_s}{4} t \quad (9.2-12)$$

Use of a standard table of z transforms gives

$$X^*(z) = \frac{Az}{z^2 + 1} (\cos \varphi + z \sin \varphi) \quad (9.2-13)$$

The sampled output of the nonlinearity is seen in Fig. 9.2-4 to have the transform

$$\begin{aligned} Y^*(z) &= D + Dz^{-1} - Dz^{-2} - Dz^{-3} + Dz^{-4} + \cdots \\ &= \frac{1 + z^{-1}}{1 + z^{-2}} D \\ &= \frac{z(z + 1)}{z^2 + 1} D \end{aligned} \quad (9.2-14)$$

The z -transform describing function is then

$$\begin{aligned} N^*(A, \varphi) &= \frac{Y^*(z)}{X^*(z)} \\ &= \frac{D}{A} \frac{z + 1}{\cos \varphi + z \sin \varphi} \end{aligned} \quad (9.2-15)$$

The sampled transfer function for the linear part of the system is calculated by standard procedures, again referring to a table of z transforms.

$$\begin{aligned} (D^*HL)^* &= \left(\frac{T_s}{z - 1} \frac{z - 1}{zs} \frac{K}{\tau s + 1} \right)^* \\ &= \frac{T_s}{z} \left[\frac{K}{s(\tau s + 1)} \right]^* \\ &= \frac{KT_s(1 - a)}{(z - 1)(z - a)} \end{aligned} \quad (9.2-16)$$

where

$$a = e^{-T_s/\tau} \quad (9.2-17)$$

Equating the fundamental component of the fed-back samples to $-x(t)$ requires

$$N^*(D^*HL)^*|_{w=\omega_s/4} = -1 \quad (9.2-18)$$

$$\frac{D}{A} \frac{j+1}{\cos \varphi + j \sin \varphi} \frac{KT_s(1-a)}{(j-1)(j-a)} = 1 \angle -180^\circ$$

Solving for A and φ yields, after some manipulation,

$$A = KDT_s \frac{1-a}{\sqrt{1+a^2}} \quad (9.2-19)$$

$$\varphi = -\tan^{-1} a \quad (9.2-20)$$

with a given by Eq. (9.2-17). Again, φ must lie in the range $(0, 90^\circ)$, but according to Eq. (9.2-20), it takes only negative values. Thus application of z -transform describing function theory leads to the conclusion that the 2, 2 limit cycle mode cannot be sustained by this system—a conclusion quite different from that reached using the sampled describing function previously.

Discussion Consideration of the shapes of the signals that would result in the linear part of this system for the given nonlinearity output readily confirms the conclusion reached by z -transform describing function analysis as correct. If the periodic pulse sequence $y^*(t)$ as shown in Fig. 9.2-4 were impressed on the linear part of this system, the steady-state response at $c(t)$ and $w(t)$ is shown in Fig. 9.2-5. It is clear that at $t = 0$, $w(t)$ must be greater than zero, and thus $x(t)$ must be less than zero. But this is inconsistent with a positive impulse in $y^*(t)$ at $t = 0$; so the mode is impossible. The failure of the sampled describing function method in this case is due to a rather slight difference in phase between the fundamental component of $w(t)$ and the fundamental component of $w^*(t)$. The fundamental component of $w(t)$, for small enough values of τ/T_s , crosses zero to the left of $t = 0$ and indicates that the mode is possible. But the fundamental component of $w^*(t)$ crosses zero to the right of $t = 0$ for all values of τ/T_s and correctly indicates that the mode is impossible. The fed-back signal in this case, especially for small values of τ/T_s , does not approximate a sinusoid at all well, and one would expect possible difficulty in the use of describing function theory.

As noted before, this example, using only first-order continuous linear filtering, was deliberately chosen to emphasize possible differences in results using the two describing function points of view. In cases where the signal returned to the nonlinearity is better filtered, and where there is no problem with sampling of harmonically distorted signals in the system linear part, the two procedures yield very similar results, both of which are in close agreement with exact results. As an indication of this, consider another example in which second-order continuous linear filtering exists.

Example 9.2-2 The system of Fig. 9.2-6 uses an ideal two-level relay controller and first-order digital lead compensation. The remainder of the linear part consists of a zero-order hold and a continuous linear part

$$L(s) = \frac{K}{s(\tau s + 1)} \quad \tau = T_s \quad (9.2-21)$$

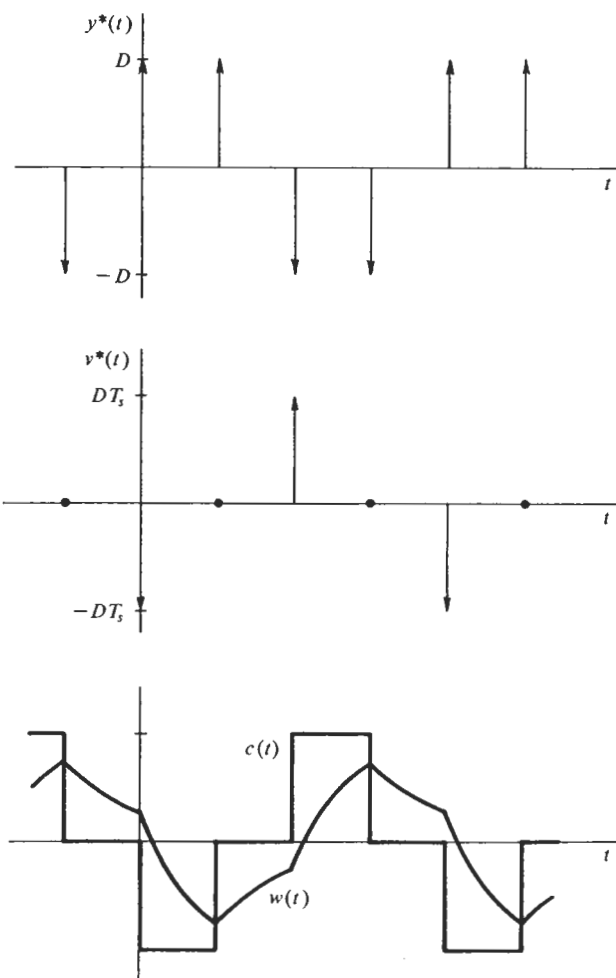


Figure 9.2-5 Signal waveforms in Example 9.2-1.

Determine the amount of lead (the value of a) necessary to render the 3, 3 limit cycle mode impossible.

Using procedures identical with those of the preceding example, we find

$$x(t) = A \sin \left(\frac{\omega_s}{6} t + \varphi \right) \quad 0 < \varphi < 60^\circ \quad (9.2-22)$$

Using the sampled describing function approach,

$$N(A, \varphi) = \frac{4D}{3AT_s} \underline{\angle 30^\circ - \varphi} \quad 0 < \varphi < 60^\circ \quad (9.2-23)$$

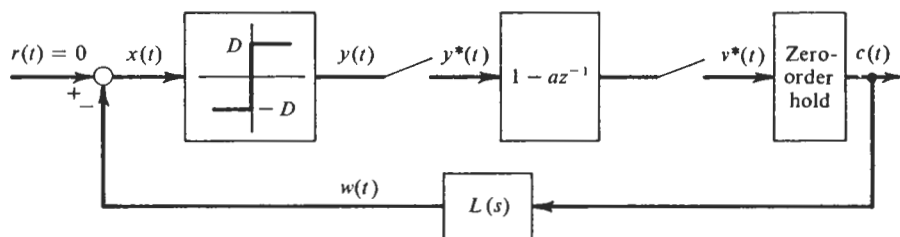


Figure 9.2-6 System configuration for Example 9.2-2.

and

$$D^*HL\left(j\frac{\omega_s}{6}\right) = \frac{K(1 - ae^{-j\pi/3})(1 - e^{-j\pi/3})}{-\left(\frac{\pi}{3}\frac{1}{T_s}\right)^2\left(1 + j\frac{\pi}{3}\frac{\tau}{T_s}\right)} \quad (9.2-24)$$

The solution of

$$ND^*HL = -1$$

gives

$$A = 0.839KDT_s \sqrt{(1 - 0.5a)^2 + (0.866a)^2} \quad (9.2-25)$$

$$\varphi = 43.7^\circ + \tan^{-1} \frac{0.866a}{1 - 0.5a} \quad (9.2-26)$$

The requirement that φ lie in the interval $(0, 60^\circ)$ dictates that the range of a for which the 3, 3 limit cycle mode is possible is

$$-2.46 < a < 0.289$$

Values of a outside this range render the mode impossible.

With the z-transform describing function approach we find

$$N^*(A, \varphi) = \frac{D}{A} \frac{z^2 + z + 1}{(z + 1)[0.866 \cos \varphi + (z - 0.5) \sin \varphi]} \quad 0 < \varphi < 60^\circ \quad (9.2-27)$$

and

$$(D^*HL)^* = KT_s \frac{(0.368z + 0.264)(z - a)}{z(z - 1)(z - 0.368)} \quad (9.2-28)$$

The solution of

$$N^*(D^*HL)^*|_{\omega=\omega_s/6} = -1$$

gives

$$A = 0.836KDT_s \sqrt{(0.5 - a)^2 + (0.866)^2} \quad (9.2-29)$$

$$\varphi = -15.9^\circ + \tan^{-1} \frac{0.866}{0.5 - a} \quad (9.2-30)$$

The requirement that φ lie in the interval $(0, 60^\circ)$ indicates in this instance that the range of a for which the 3, 3 limit cycle mode is possible is

$$-2.54 < a < 0.283$$

The calculated end points for this interval of a differ no more than 3 percent from those calculated by the sampled describing function technique. Also, in the two expressions for

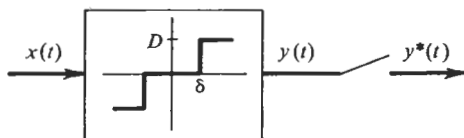
the amplitude of the limit cycle, Eqs. (9.2-25) and (9.2-29), the two radicals involving a are identical; so the expressions are seen to differ by just 0.4 percent.

Having discussed and illustrated these two points of view regarding describing function analysis of limit cycles in sampled nonlinear systems, we proceed to consider the calculation of these describing functions for nonlinearities other than the two-level relay, which has already been treated. An important observation should be made at the outset. In the preceding discussion, the point of view taken was that the samples of the nonlinearity output, or their fundamental component, are processed around the loop, and the fundamental component of the signal returned to the nonlinearity input is equated to the sinusoid originally assumed. When the z -transform describing function is calculated in advance and graphed for convenient use, the ratio of Y^* to X^* is immediately evaluated at the fundamental frequency. But the evaluation of Y^* at the fundamental frequency serves to select just the fundamental component from the $y^*(t)$ waveform. Also, evaluation of X^* at the fundamental frequency selects the fundamental component of $x^*(t)$. Since $x(t)$ is taken to be a sinusoid, the fundamental component of $x^*(t)$ is just $1/T_s$ times $x(t)$ for $T > 2T_s$. Thus, for $T > 2T_s$, the z -transform describing function is exactly T_s times the sampled describing function which relates the amplitude and phase of the fundamental component of $y^*(t)$ to $x(t)$. As noted before, for $T = 2T_s$ the z -transform describing function is not applicable; the sampled describing function is. Since in other cases the two describing functions are related by a known constant, only one calculation need be made. The easier calculation by far is that of the sampled describing function. The difference between the two points of view is then finally evidenced in the difference between the continuous and sampled transfer functions for the linear part of the system.

THE THREE-LEVEL RELAY

After the two-level relay, the nonlinearity of greatest importance in limit cycle analysis of sampled nonlinear systems is almost surely the three-level relay. In many instances in which a fixed drive level is desired, a zero level is included for the specific purpose of avoiding limit cycles in the absence of input. The nonlinearity under consideration is shown in Fig. 9.2-7. Calculation of the sampled describing function for this nonlinearity is in principle no different from the corresponding calculation for the two-level relay, but in practice is much more tedious because of the change in form of the output at the input level δ . This gives rise to a variety of different output mode shapes at every frequency, these modes depending on the amplitude and phase of the input. For most nonlinearities other than the two-level relay, the phase of the input sinusoid relative to the sampling points enters

Figure 9.2-7 The sampled three-level relay.



the problem in a more complicated way in that it affects the shape of the output waveform. The labor involved in describing function calculation is due to the necessity of enumerating all possible mode shapes and dealing with each one.

As an illustration of the calculation in the present case of the three-level relay, consider the possible modes of period $6T_s$. With the input assumed to be an unbiased sinusoid,

$$x(t) = A \sin(\omega t + \varphi) \quad (9.2-31)$$

With $\omega = \frac{1}{6}\omega_s = \pi/3T_s$,

$$x(t + 3T_s) = -x(t) \quad (9.2-32)$$

so the values of $y(t)$ at the first three sampling points determine $y(t)$ at all other sampling points. The sampling lag can range between zero and one sampling period, or

$$0 < \varphi < 60^\circ$$

in this case. The nonlinearity input at the times of the first three samples is then

$$x(0) = A \sin \varphi \quad (9.2-33)$$

$$x(T_s) = A \sin(60^\circ + \varphi) \quad (9.2-34)$$

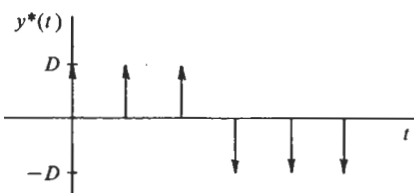
$$x(2T_s) = A \sin(120^\circ + \varphi) \quad (9.2-35)$$

In view of the limited range of φ , it is clear that $y(t)$ at these three sampling points can only be 0 or $+D$. This gives eight possible combinations of values, including the trivial zero output. But not all eight combinations are consistent with the sinusoidal form of the input. Notice that over the full range of φ , $x(T_s)$ is greater than $x(0)$ or $x(2T_s)$. Thus, if $y(T_s)$ is zero, $y(0)$ and $y(2T_s)$ cannot be D . This leaves four possible modes other than the trivial one. The four $y^*(t)$ waveforms are pictured in Fig. 9.2-8. Each of these modes is possible for a restricted range of φ which depends on the amplitude of the input. Take as an example the waveform *b*. The condition for $y(0) = D$ is

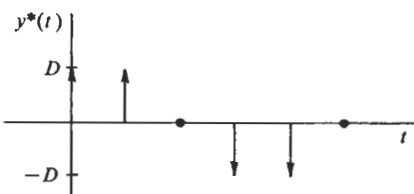
$$\begin{aligned} A \sin \varphi &> \delta \\ \text{or} \quad \varphi &> \Delta \end{aligned} \quad (9.2-36)$$

where we have defined

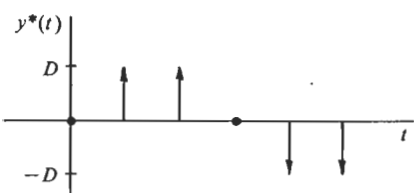
$$\Delta = \sin^{-1} \frac{\delta}{A} \quad (9.2-37)$$



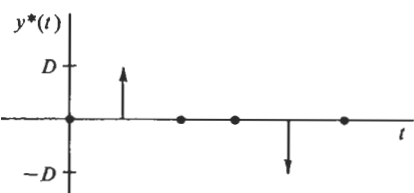
(a)



(b)



(c)



(d)

Figure 9.2-8 Possible modes for $T = 6T_s$.

The condition for $y(T_s) = D$ is

$$\Delta < \varphi + 60^\circ < 180^\circ - \Delta \quad (9.2-38)$$

and the condition for $y(2T_s) = 0$ is

$$\varphi + 120^\circ > 180^\circ - \Delta \quad (9.2-39)$$

The boundaries of these ranges of φ can be plotted against Δ , and a region in the φ, Δ space identified in which the conditions of Eqs. (9.2-36), (9.2-38), and (9.2-39) are simultaneously satisfied. This region is shown in Fig. 9.2-9 labeled (b). Also shown are the corresponding regions for the other waveforms of Fig. 9.2-8.

The describing function is now determined by calculating the fundamental component of each of the waveforms of Fig. 9.2-8. This is particularly easy to do since the harmonic analysis involves integrating functions which are sums of delta functions. The amplitude and phase relation between $x(t)$ and the fundamental component of $y^*(t)$ is then the sampled describing function, which is also $1/T_s$ times the z -transform describing function. For each of the modes of Fig. 9.2-8, this gives

$$(a) \quad N = \frac{4}{3} \frac{D}{AT_s} \angle 30^\circ - \varphi \quad (9.2-40a)$$

$$(b) \quad N = \frac{2}{\sqrt{3}} \frac{D}{AT_s} \angle 60^\circ - \varphi \quad (9.2-40b)$$

$$(c) \quad N = \frac{2}{\sqrt{3}} \frac{D}{AT_s} \angle -\varphi \quad (9.2-40c)$$

$$(d) \quad N = \frac{2}{3} \frac{D}{AT_s} \angle 30^\circ - \varphi \quad (9.2-40d)$$

The range of φ for which each of these modes can exist is given in terms of $\Delta = \sin^{-1}(\delta/A)$. Thus it is convenient to normalize N with δ and keep δ/A as the amplitude parameter. Limit cycle modes will be determined finally by plotting L versus $-1/N$; so the function $-1/N$ can best be plotted immediately.

$$(a) \quad \frac{D}{\delta T_s} \left(-\frac{1}{N} \right) = \frac{3}{4} \frac{A}{\delta} \angle \varphi - 210^\circ \quad (9.2-41a)$$

$$(b) \quad \frac{D}{\delta T_s} \left(-\frac{1}{N} \right) = \frac{\sqrt{3}}{2} \frac{A}{\delta} \angle \varphi - 240^\circ \quad (9.2-41b)$$

$$(c) \quad \frac{D}{\delta T_s} \left(-\frac{1}{N} \right) = \frac{\sqrt{3}}{2} \frac{A}{\delta} \angle \varphi - 180^\circ \quad (9.2-41c)$$

$$(d) \quad \frac{D}{\delta T_s} \left(-\frac{1}{N} \right) = \frac{3}{2} \frac{A}{\delta} \angle \varphi - 210^\circ \quad (9.2-41d)$$

Now, for every choice of A/δ in the range $(1, \infty)$, the magnitude of $-1/N$ in each mode is determined. Also, a range of phase angles for $-1/N$ is given by these expressions and the indicated range of φ in Fig. 9.2-9 at the value of $\Delta = \sin^{-1}(\delta/A)$. These describing function regions can conveniently be graphed on a gain-phase plot. Such a graph is shown for $T = 6T_s$ in Fig. 9.2-10. The regions within which the different modes exist are indicated. The possibility of a limit cycle of period $6T_s$ can be determined by evaluating the transfer function for the linear part of the system at the frequency $\frac{1}{6}\omega_s$.

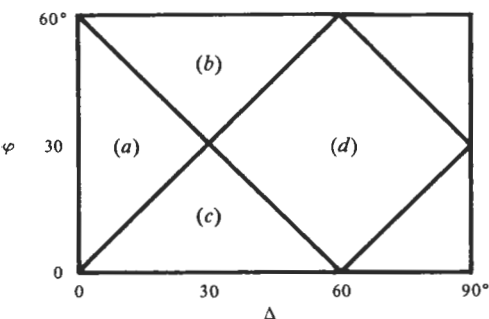


Figure 9.2-9 Regions of existence for the different modes.

multiplying it by $D/\delta T_s$, and noting the resulting complex number on this graph. If the point falls within one of the regions corresponding to a particular mode, the system can sustain that limit cycle mode. The indicated magnitude and phase of the describing function, together with the relations of Eqs. (9.2-41), then yield the amplitude and phase of the mode. If one desires to compensate the system to eliminate any or all limit cycle modes of this frequency, the necessary magnitude and phase of linear compensation at this frequency are obvious from the graph. Notice that in several regions more than one limit cycle mode can exist. Which mode a system will display, if any, depends on its initial conditions.

Calculation of the describing function for periods which are odd multiples of the sampling period is even more laborious than for even-multiple periods because no condition of symmetry such as Eq. (9.2-32) applies. This requires the value of $y(t)$ at each sampling point in the cycle to be considered independently, and gives rise to more possible modes. Some of these modes, which exist in different regions of the φ, Δ plane (corresponding to Fig. 9.2-9) map into the same describing function region (corresponding to Fig. 9.2-10). This occurs in the case of modes which have the same waveshape but are shifted in phase by one or more sampling periods or are traversed in the opposite direction.

In the case of even-multiple periods, all the possible $y^*(t)$ waveforms corresponding to unbiased sinusoidal inputs to the nonlinearity are unbiased functions. Both biased and unbiased waveforms are possible in the case of odd-multiple periods. However, only the unbiased modes are of consequence if $x(t)$ is assumed unbiased. If the linear part of the system includes an integrator (has a pole at the origin in the s plane), it is clear that $y^*(t)$ must be unbiased in any steady-state oscillation of the unexcited system. If the linear part has no pole at the origin, and also no zero, any bias in $y^*(t)$ would be propagated around the loop and appear at the nonlinearity input. Analysis of this situation would require two-input describing function theory. Only in the rare case of a linear part having a zero at the origin in the s

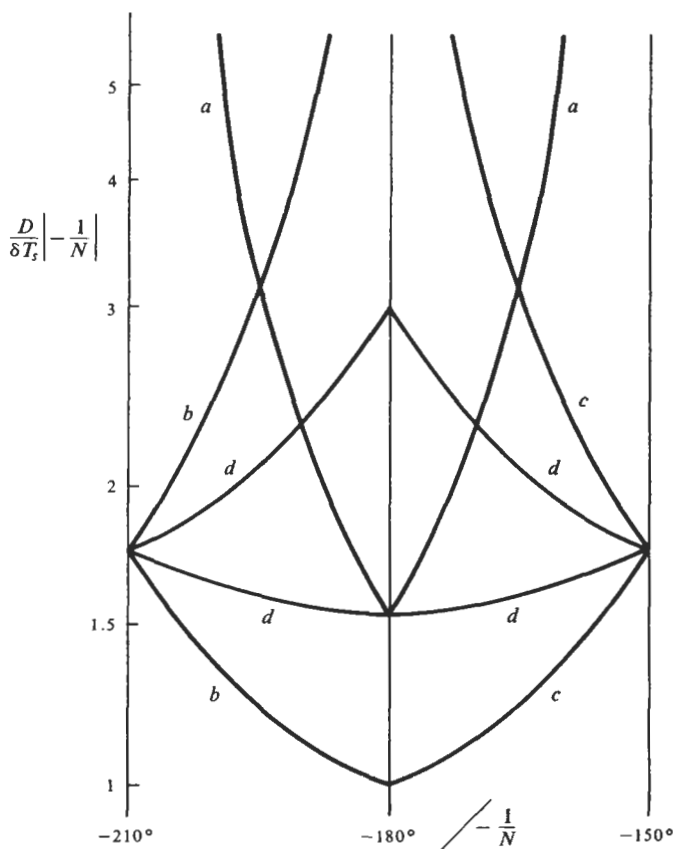


Figure 9.2-10 Sampled describing function for the three-level relay, $T = 6T_s$.

plane would a biased $y^*(t)$ waveform be consistent with an unbiased $x(t)$ for an unexcited system. Only the unbiased modes are included in the describing functions plotted in Appendix F.

Systems with a pole at the origin can sustain a biased output from the linear part even if the input to that part is unbiased. This leads to a range of possible null offsets, or bias levels in the nonlinearity input, which may remain uncorrected in the presence of any limit cycle mode. The situation in the case of the three-level relay is no different from that discussed earlier in the case of the two-level relay. Having determined the limit cycle modes which may exist under the assumption of an unbiased $x(t)$, the amplitude and phase angle of the sinusoid at $x(t)$ are known. Direct observation of

this sinusoid at the sampling points indicates how far the cycle may be biased up and down without changing the nonlinearity output. This defines the range of possible null offsets in the presence of a given limit cycle mode. It appears rather obvious, and is proved by Chow (Ref. 2), that no new unbiased $y^*(t)$ modes are introduced by the presence of a bias in $x(t)$.

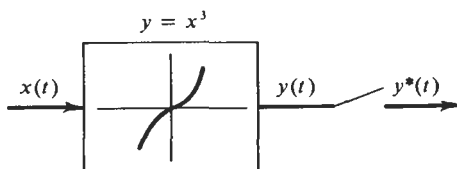
The accuracy of describing function analysis applied to sampled nonlinear systems is of the same order as in the case of continuous systems. Many writers have done experimental studies to evaluate this accuracy. Among them are Chow (Ref. 2), Dixon (Ref. 5), and Kuo (Refs. 14, 15), all of whom report that for sampled three-level relay systems with second-order linear parts, the error in describing function prediction of the amplitude of limit cycle modes ranges generally from 1 to 10 percent. The critical value of system gain, or nonlinearity dead zone, δ , for the existence of a particular mode is even better predicted: the error is usually less than 5 percent. There is, in the case of sampled systems, no error in the determination of limit cycle frequency.

In calculating describing functions for the sampled three-level relay, the labor involved increases rapidly with the period of the limit cycle modes considered. At the same time, the importance of the sampling operation decreases as the period of the cycle increases. Thus, for long-period modes, an approximation to the describing function is both feasible and highly desirable. From the form of the describing functions plotted in Appendix F, one can anticipate the bounds on $-1/N$, which were derived by Chow (Ref. 2) and Russell (Ref. 23). If the period of the oscillation is $T = nT_s$, the negative reciprocal describing function exhibits phase angles within a band $2\pi/n$ wide for n even, or π/n wide for n odd, these bands being centered around the angle $-\pi$. The minimum magnitude of the function $-(D/\delta T_s)(1/N)$ is 0.5 for $n = 2$, is 1.0 for some small values of n , and approaches $\pi/2 = 1.57$ for large n . The maximum magnitude is always infinite. Thus, except for $n = 2$, for which the exact describing function is given in Appendix F, one can lay out a rectangle on the gain-phase plot ranging from 1.0 to infinity in amplitude and from $-\pi - \pi/n$ to $-\pi + \pi/n$ in phase, and be assured that $-(D/\delta T_s)(1/N)$ corresponding to all modes of period nT_s will lie within that rectangle. If the design requirement is to avoid any limit cycle modes, this bound on the describing function can be used to account for all modes of longer period than those for which the describing function has been calculated.

OTHER NONLINEARITIES

Describing function calculation for nonlinearities which have the same analytic description for all inputs is quite simple. Take as an illustration the cubic nonlinearity and sampler pictured in Fig. 9.2-11. Again, $x(t)$ is

Figure 9.2-11 The sampled cubic nonlinearity.



taken to be a sinusoid as in Eq. (9.2-31). Thus

$$y(mT_s) = A^3 \sin^3(\omega mT_s + \varphi) \quad (9.2-42)$$

The fundamental component of $y^*(t)$ for $\omega = (1/n)\omega_s$ is

$$\begin{aligned} y^*(t)_f &= a \sin \omega t + b \cos \omega t \\ &= \sqrt{a^2 + b^2} \sin \left(\omega t + \tan^{-1} \frac{b}{a} \right) \end{aligned} \quad (9.2-43)$$

where

$$\begin{aligned} a &= \frac{2}{nT_s} \int_0^{nT_s} y^*(t) \sin \frac{1}{n} \omega_s t \, dt \\ &= \frac{2A^3}{nT_s} \sum_{m=0}^{n-1} \sin^3 \left(\frac{m}{n} 2\pi + \varphi \right) \sin \frac{m}{n} 2\pi \end{aligned} \quad (9.2-44)$$

and

$$b = \frac{2A^3}{nT_s} \sum_{m=0}^{n-1} \sin^3 \left(\frac{m}{n} 2\pi + \varphi \right) \cos \frac{m}{n} 2\pi \quad (9.2-45)$$

The sampled describing function is then

$$N = \frac{1}{A} \sqrt{a^2 + b^2} / \tan^{-1} (b/a) - \varphi \quad (9.2-46)$$

The expressions for a and b can be reduced by trigonometric manipulations, but this will not be pursued here. The amplitude of the input in this case affects only the magnitude of N , which is proportional to A^2 . For any order n of limit cycle mode, $(1/A^2)N$ describes a closed contour in the complex plane or on a gain-phase plot as φ varies from 0 to $(1/n)2\pi$. This contour is then just scaled in magnitude with A^2 .

Lepschy and Ruberti (Ref. 16) have calculated the describing function for the dead-zone-gain nonlinearity with a sampler and zero-order hold. This nonlinearity is interesting not only in its own right, but also because parallel combinations of dead-zone gains can be used to synthesize any piecewise-linear continuous nonlinearity. As always, the sum of describing functions for the parallel elements is the describing function for the combination. To facilitate addition of the describing functions for dead-zone-gain elements, Lepschy and Ruberti have plotted their results, not only in terms of magnitude and phase, but also in terms of real and imaginary parts.

Some nonlinearities, such as the limiter, are of less interest in the study of limit cycles in sampled nonlinear systems than in continuous nonlinear systems because of the very rare occurrence of *stable* limit cycle modes of interesting form in systems containing these nonlinearities. The stability of indicated limit cycle modes is the subject of the next section.

9.3 STABILITY OF LIMIT CYCLE MODES

The analysis of the preceding sections has determined essentially the possible states of dynamic equilibrium of unexcited sampled nonlinear systems of a certain class. These equilibrium states, or limit cycles, can be either stable or unstable in that a small perturbation from the self-sustaining periodic mode may tend either to decay or grow. The determination of limit cycle modes simply identifies those signal histories which will reproduce themselves when propagated around a closed-loop system. This says nothing about the stability of the modes, or if stable, about the range of initial conditions which will result in eventual capture to a particular mode.

The standard technique of perturbation analysis to study the stability of limit cycles has even wider applicability to sampled than to continuous systems. In either case one supposes a small perturbation $p(t)$ to exist at the input to the nonlinearity, in addition to the limit cycle $x(t)$. The output of the nonlinearity is then expanded approximately into the limit cycle output plus a perturbation which is linearly related to $p(t)$. In the case of continuous systems, this expansion can be done only for nonlinearities which are differentiable over the range of inputs experienced in the limit cycle. This rules out such commonplace nonlinearities as two- and three-level relays. In the case of sampled systems, the nonlinearity need not be differentiable over any continuous range. The only points on the nonlinear characteristic which are of consequence are those at which the samples are taken in the limit cycle mode under consideration. Even in the case of discontinuous nonlinearities such as relays, the slope of the nonlinearity is well defined at those discrete points where the samples are taken in the limit cycle.

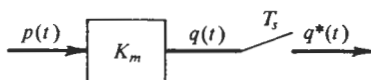
If the output of the nonlinearity can be expanded into the sum of the limit cycle output plus a perturbation which is linearly related to the input perturbation $p(t)$, the effect of the nonlinearity on the perturbation is characterized as a linear gain which depends on the limit cycle input $x(t)$, and hence is periodically time-varying. In the case of continuous systems, determination of the stability of linear systems containing a periodically time-varying gain is still a substantial chore. The same would be true of sampled systems, except that it is possible to transform the system from one containing a sampler with periodically varying gain into one containing multiple samplers with fixed gains in parallel. This complicates the configuration somewhat, but

permits the use of standard linear invariant theory to determine the stability of limit cycle modes.

Nease (Ref. 18) proves that the stability of a periodic solution of a nonlinear set of difference equations is given by the stability of the linear set which approximates the action of the system on a small perturbation about the periodic solution. Nease also proves a theorem which establishes the stability of a linear set of difference equations with periodically varying coefficients. Because of its more likely appeal to control-system designers, we prefer to present here a technique for the study of stability suggested by Pyshkin (Ref. 21). Suppose that a sampled nonlinear system has been analyzed by the method of the preceding section, and the existence of a limit cycle of period $T = nT_s$ has been established. It remains to determine whether this indicated limit cycle is stable or unstable. If the periodic signal $x(t)$ which appears at the nonlinearity input in the steady-state limit cycle is perturbed by a small additive signal $p(t)$, the perturbation in the nonlinearity output at the sampling times, $q(mT_s)$, is given to first order by $p(mT_s)$ times the slope of the nonlinear characteristic dy/dx , evaluated at $x(mT_s)$. Considering the operation of the system on the perturbations only, the system configuration is the original configuration, with the nonlinearity replaced by the periodically time-varying gain $(dy/dx)[x(mT_s)]$. This gain is defined at the sampling instants only, but it is of consequence only at those instants.

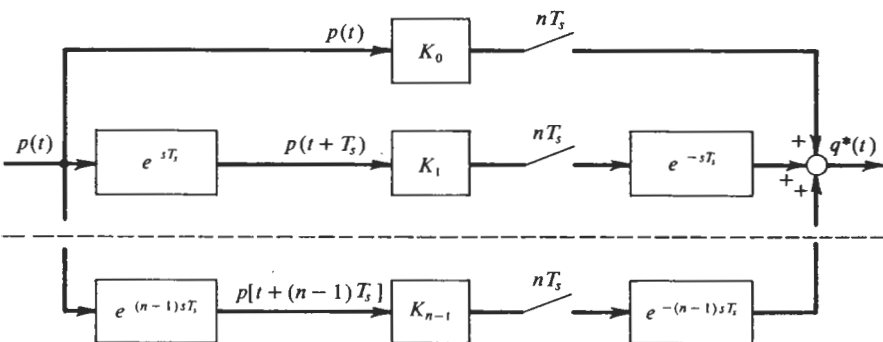
The system whose stability determines the stability of the limit cycle contains one sampler (in addition to any samplers in the linear part of the original system), with sampling period T_s , preceded by a gain which takes a set of n discrete values at the sampling times and repeats them sequentially. This variable-gain sampler is indicated in Fig. 9.3-1a. Rather than deal with the stability of a variable-gain system, we prefer to replace the variable-gain sampler with n samplers in parallel, each having a fixed gain and the sampling period nT_s . The equivalent set of samplers is shown in Fig. 9.3-1b. Because ordinary sampled-data system analysis depends on the assumption that all samplers in the system operate synchronously, the various samplers in this equivalent configuration are preceded by ideal predictors, so that $p(t)$ will be sampled at the appropriate times, and followed by ideal delays, so that the samples will appear at the output at the correct times. With this transformation, the system whose stability determines the stability of the limit cycle under consideration becomes a fixed-parameter linear sampled-data system, and ordinary linear sampled-data system theory is applicable.

The characteristic equation of a system containing a parallel set of predictors, samplers, and delays, as in Fig. 9.3-1b, involves z transforms of functions with predictors or delays of a fraction of the sampling period. Such transforms can be found in tables of "modified z transforms" (Ref. 7) or "advanced z transforms" (Ref. 22). It should also be noted that if an



$$K_m = K_0, K_1, \dots, K_{n-1}, K_0, K_1, \dots$$

(a) Single sampler with pK_0 , dically varying gain



(b) n samplers with fixed gains

Figure 9.3-1 Equivalent sampling systems.

unbiased even-period limit cycle mode (n even) is being tested for stability, and if the nonlinear characteristic is odd, then

$$\frac{dy}{dx} \left[x \left(mT_s + \frac{n}{2} T_s \right) \right] = \frac{dy}{dx} [x(mT_s)]$$

so the sequence of gains K_i , appearing in Fig. 9.3-1b, repeats after the first $n/2$ values. Thus only $n/2$ parallel paths are required, and each sampler operates with the period $(n/2)T_s$. These points can be illustrated by an example.

Example 9.3-1 Consider the sampled system containing a limiter shown in Fig. 9.3-2. The linear part includes a zero-order hold and an integrator. Test the stability of a limit cycle mode of period $T = 4T_s$, having the form $y(mT_s) = 1, a, -1, -a$ for $m = 0, 1, 2, 3$. The constant a can be any value between zero and 1.

First solution Just for illustration, work first with the full form of the equivalent sampling system for perturbation analysis as shown in Fig. 9.3-1b. The slope of the nonlinearity evaluated at the sampling points is

$$\frac{dy}{dx} [x(mT_s)] = 0, k, 0, k$$

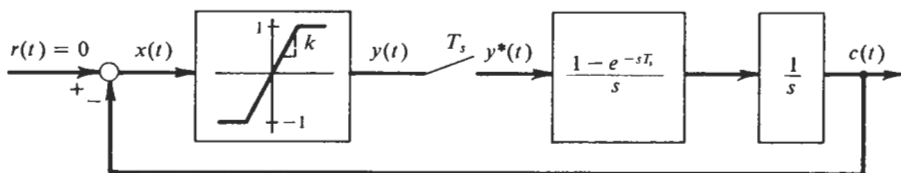


Figure 9.3-2 System for Example 9.3-1.

The full form of the linearly perturbed system is shown in Fig. 9.3-3. The characteristic equation for this system can be found in the following steps [$Q_i(s)$ is the Laplace transform of $q_i(t)$]:

$$Q_2 = -Q_2^* \frac{1 - e^{-sT_s}}{s^2} k - Q_4^* \frac{1 - e^{-sT_s}}{s^2} k e^{-2sT_s} \quad (9.3-1)$$

$$Q_4 = -Q_2^* \frac{1 - e^{-sT_s}}{s^2} k e^{2sT_s} - Q_4^* \frac{1 - e^{-sT_s}}{s^2} k$$

Equivalently,

$$(1 + F_1^*)Q_2^* + F_2^*Q_4^* = 0 \quad (9.3-2)$$

$$F_3^*Q_2^* + (1 + F_1^*)Q_4^* = 0$$

where

$$F_1^* = k \left(\frac{1 - e^{-sT/4}}{s^2} \right)^* \quad (9.3-3)$$

$$F_2^* = k \left(\frac{1 - e^{-sT/4}}{s^2} e^{-sT/2} \right)^*$$

$$F_3^* = k \left(\frac{1 - e^{-sT/4}}{s^2} e^{sT/2} \right)^*$$

The star indicates the z transform with respect to the sampling period $T = 4T_s$. As an illustration of this calculation, find the transform F_2^* .

$$\begin{aligned} F_2^* &= k \left(\frac{e^{-sT/2}}{s^2} - \frac{e^{-s\frac{3}{2}T}}{s^2} \right)^* \\ &= k \left(e^{-sT} \frac{e^{sT/2}}{s^2} - e^{-sT} \frac{e^{s\frac{3}{2}T}}{s^2} \right)^* \end{aligned}$$

This transform is found directly in a table of modified or advanced z transforms.

$$\begin{aligned} F_2^* &= k \left[z^{-1} \frac{\frac{1}{2}T + \frac{1}{2}Tz^{-1}}{(1 - z^{-1})^2} - z^{-1} \frac{\frac{3}{2}T + \frac{3}{2}Tz^{-1}}{(1 - z^{-1})^2} \right] \\ &= \frac{kT_s}{z - 1} \end{aligned} \quad (9.3-4)$$

Similarly, one finds

$$F_1^* = \frac{kT_s}{z - 1} \quad (9.3-5)$$

$$F_3^* = \frac{kT_s z}{z - 1} \quad (9.3-6)$$

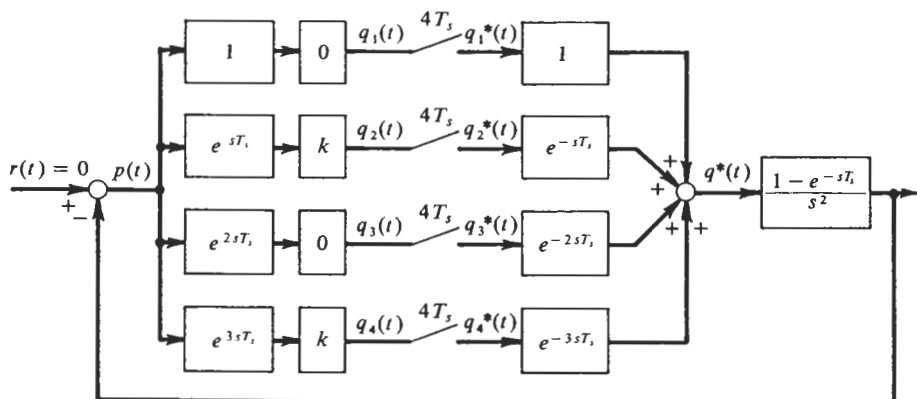


Figure 9.3-3 Full form of perturbed system of Example 9.3-1.

The characteristic equation of this system results from equating to zero the determinant of the coefficient matrix in Eqs. (9.3-2).

$$(1 + F_1^*)^2 - F_2^* F_3^* = 0$$

With Eqs. (9.3-4) to (9.3-6) this becomes

$$z = (1 - kT_s)^2 \quad (9.3-7)$$

The closed-loop pole takes values between 0 and 1 for values of kT_s in the range

$$0 < kT_s < 2$$

This is the range of stability for the linearly perturbed system. Thus, if the original system had an indicated limit cycle mode of the stated form for a value of kT_s in this range, the mode would be stable.

Second solution The analysis of this problem can be simplified because the limit cycle under consideration is an unbiased even-period mode and the nonlinearity is odd. Thus, only the first half of the parallel paths shown in Fig. 9.3-3 are needed if the period of the samplers is changed to $2T_s$. But this leaves only one path with a nonzero gain. This is the second path, which includes a one-unit predictor and a one-unit delay. With just a single path, a predictor and corresponding delay have no effect on system stability, and for the present purpose can be ignored. If there were multiple paths of this form, all predictors and delays could be shifted by the same time interval with no effect on the stability of the system.

The simplified form of the perturbed system is shown in Fig. 9.3-4. This is just an ordinary single-path linear sampled-data system formed from the original nonlinear system by replacing the limiter by its gain in the linear region and changing the sampling period from T_s to $2T_s$. The closed-loop root of this system is found to be located at

$$z = 1 - kT_s \quad (9.3-8)$$

Again the stable range is found to be

$$0 < kT_s < 2$$

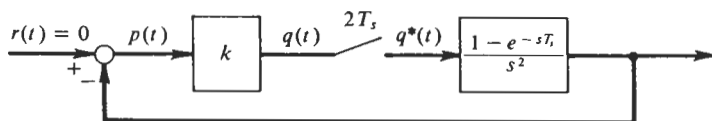


Figure 9.3-4 Simplified form of perturbed system of Example 9.3-1.

At the unstable boundary, $kT_s = 2$, the closed-loop pole in this case is located at $z = -1$, which implies a characteristic mode whose samples oscillate between constant plus and minus values. In the previous case, Eq. (9.3-7) shows the closed-loop pole to be located at $z = +1$ for $kT_s = 2$. This pole location implies a characteristic mode whose samples are all the same constant value. This is not an inconsistency since z in Eq. (9.3-8) is e^{s2T_s} , whereas z in Eq. (9.3-7) is e^{s4T_s} . The mode which oscillates between equal plus and minus values every $2T_s$ displays the same constant value when sampled every $4T_s$.

This example shows why limiters are somewhat uninteresting for limit cycle analysis of sampled nonlinear systems. If the system is designed to be stable within the linear range of the limiter, it is unusual for limit cycle modes of the partially saturated type considered in the example to exist. But if the gain or sampling period is increased to the point where such a mode does exist, it is quite likely that the system is unstable in the linear range with the sampling period T_s ; it is very rare that the system would be found stable with the sampling period $2T_s$ as required for stability of the limit cycle mode. Thus, for most limiter systems, the only *stable* limit cycle modes found are the fully saturated modes, in which the limiter acts just like a two-level relay.

One may also note the simple interpretation of this stability analysis in the case of *two- or three-level relay systems*. Since the slope of these nonlinear characteristics at any sampling point is zero, the linearized system which processes perturbations on any limit cycle mode has a gain preceding the sampler which is zero at every sampling instant. Thus, as far as perturbations are concerned, there is no loop closed around the linear part of the system. These modes are then stable if the open-loop linear part is stable, and unstable if the open-loop linear part is unstable.

9.4 EXACT VERIFICATION OF LIMIT CYCLE MODES

In Sec. 9.2 we discussed two points of view regarding the determination of possible limit cycle modes in sampled nonlinear systems, each depending on the describing function approximation that the input to the nonlinearity is a sinusoid. This approximation is a good one if the system has sufficient linear filtering of the harmonics generated by the nonlinearity and by the samplers. For the great majority of practical purposes, "sufficient" linear filtering is provided by second- or higher-order filters if the fundamental frequency is near or beyond the cutoff frequency of the filter.

In those cases in which one questions the adequacy of the linear filtering properties of a system for describing function analysis, he can verify exactly whether or not any suggested limit cycle can exist. Indeed, if one wishes, he can use an exact technique as his basic tool for limit cycle determination in the first place. The only objection to this is the greater labor involved and the uncertainty regarding how many possible modes to test. Using an exact technique, the test to determine whether or not any suggested limit cycle mode can exist in a system is an individual problem; it must be repeated for every suggested mode and for every system to be considered. The virtue of the describing function approach is that it allows separate characterization of the linear and nonlinear parts of the system. A particular nonlinearity can be considered quite independent of any system, and its describing function calculated for any number of modes. With these functions in hand, very likely graphed, one then need consider only the transfer characteristics of the linear parts of any systems which contain this nonlinearity to determine the existence of all the modes for which the describing function was originally calculated. In addition, the form of the describing function for the different modes considered, and the shape of the frequency response function for the linear part, often make it clear whether or not other modes are likely to exist.

The procedure for testing the existence of a suggested mode is simple and obvious in principle. The mode to be tested is characterized by a particular periodic sequence of samples at the nonlinearity output. This sequence is the input to the linear part of the system; together with the unknown initial conditions in the linear part, it determines the output of the linear part. But this output is the input to the nonlinearity. If a set of initial conditions can be found which in the steady state will produce a nonlinearity input consistent with the nonlinearity output originally assumed, this will demonstrate the possible existence of the mode. Simple input forms, such as steps and fundamental frequency sinusoids, can also be included in this analysis. Many writers have employed this procedure, the differences in their approaches being in the technique used to propagate the assumed nonlinearity output through the linear part and the means of handling the initial and steady-state conditions. Bergen (Ref. 1) suggested use of the Laplace transform of $y^*(t)$ and the transfer function for the linear part to get the Laplace transform of $x(t)$, including initial-condition terms. The z transform of this expression was taken, and the initial conditions chosen to eliminate the transient term, thus revealing the desired steady-state solution. Jury and Nishimura (Ref. 8) use z -transform theory to accumulate the incremental responses of the linear part to each impulse of $y^*(t)$, employ skip sampling to select the corresponding sampled value of $x(t)$ in each period of the cycle, and apply the final-value theorem to determine the steady-state values of these samples. Special consideration must be given to the possible bias level which can exist in the case of systems with an

integrator. Pai (Ref. 19) described the use of the z transform of the specified $y^*(t)$ and of the sampled transfer function for the linear part to calculate the z transform of $x(t)$. He also prescribed the form of the z transform of the steady-state oscillation at $x(t)$, and equated coefficients of like powers of z to derive the algebraic equations to be solved simultaneously for the samples of $x(t)$. Torng and Meserve (Refs. 17, 26) first expand the periodic sequence of samples of $y(t)$ into a series of orthogonal functions; they used sines and cosines with integral arguments. Using the difference equation for the linear part, they then solve for the coefficients in the expansion of the periodic sequence of samples of $x(t)$. Having these coefficients, the samples themselves are determined. No account need be taken of initial conditions since the samples of $y(t)$ and $x(t)$ are taken in the form of steady-state periodic sequences at the outset. Torng (Ref. 25) later suggested use of complex exponentials as the basis for the series expansion. Such an expansion is, in effect, a Fourier transformation, and the coefficients in the expansion enjoy a multiplicative input-output relation for linear systems.

Of all known techniques, two are described here as being perhaps the most direct. The most straightforward approach, conceptually, is the direct use of the difference equation for the linear part to write down a set of algebraic equations which determine the samples of $x(t)$. Interestingly enough, none of the writers referred to above used this procedure. The resulting equations must be solved simultaneously, but most of the procedures mentioned above lead to the same equations to be solved. The second technique described uses the "transform" of the $y^*(t)$ and $x^*(t)$ sequences, or their expansion into series of complex exponentials. This procedure does not require solution of equations; it is a direct, step-by-step calculation. However, the calculation requires a great deal of complex algebra, and it is not clear that the labor involved is less than in the case of the first procedure. It should be emphasized that any of these procedures give only necessary conditions for the existence of a limit cycle mode. Whether or not such a mode will actually be observed depends first of all upon whether the mode is stable or unstable; and if stable, it depends on the system initial conditions.

THE DIFFERENCE EQUATION METHOD

The object of our present endeavor is to see whether a postulated periodic sequence of samples of $y(t)$, the nonlinearity output, will result in a sequence of samples of $x(t)$, the nonlinearity input, which in the steady state will reproduce the postulated $y^*(t)$. Having the samples of $y(t)$ given, the most obvious way to calculate the samples of $x(t)$ is to make direct use of the difference equation for the linear part, which is evident from the z transform of its transfer function. This, in one sentence, is a complete description of the procedure. We proceed to an illustration.

Example 9.4-1 An example used to illustrate the sampled describing function and z -transform describing function techniques will be used here as well. This will afford an exact check on the results derived by the describing functions. The system under consideration is that of Example 9.2-2, for which the configuration is shown in Fig. 9.2-6. The question posed is the determination of the range of values of a , the parameter of the digital lead compensator, for which the 3, 3 limit cycle mode is impossible.

The postulated sequence of nonlinearity output samples is $y(mT_s) = D, D, D, -D, -D, -D$ for $m = 0, 1, 2, 3, 4, 5$. For simplicity, $y(mT_s)$ will hereafter be denoted y_m . This sequence is periodically repeating:

$$y_{m+6} = y_m \quad (9.4-1)$$

The sampled transfer function for the system linear part is

$$\begin{aligned} \frac{X(z)}{Y(z)} &= -K \left[\frac{(1 - az^{-1})(1 - z^{-1})}{s^2(\tau s + 1)} \right]^* \\ &= -KT_s \left[\frac{z - b - (\tau/T_s)(1 - b)(z - 1)}{(z - 1)(z - b)} \right] \frac{z - a}{z} \end{aligned} \quad (9.4-2)$$

using the notation

$$b = e^{-T_s/\tau} \quad (9.4-3)$$

As in Example 9.2-2, take $T_s/\tau = 1$. Then Eq. (9.4-2) implies the difference equation

$$x_m - 1.368x_{m-1} + 0.368x_{m-2} = -KT_s[0.368y_{m-1} + (0.264 - 0.368a)y_{m-2} - 0.264ay_{m-3}] \quad (9.4-4)$$

which holds for all m . The y_m are given for all m , and using the periodicity relation, [Eq. (9.4-1)], which holds for the x_m as well as for the y_m , one can let m take the values 0, 1, 2, ..., 5 in Eq. (9.4-4) and write down six relations among the six unknowns x_0 to x_5 .

$$\begin{aligned} x_0 - 1.368x_5 + 0.368x_4 &= -KDT_s(-0.632 + 0.632a) \\ x_1 - 1.368x_0 + 0.368x_5 &= -KDT_s(0.104 + 0.632a) \\ x_2 - 1.368x_1 + 0.368x_0 &= -KDT_s(0.632 - 0.104a) \\ x_3 - 1.368x_2 + 0.368x_1 &= -KDT_s(0.632 - 0.632a) \\ x_4 - 1.368x_3 + 0.368x_2 &= -KDT_s(-0.104 - 0.632a) \\ x_5 - 1.368x_4 + 0.368x_3 &= -KDT_s(-0.632 + 0.104a) \end{aligned} \quad (9.4-5)$$

These equations do not yield a unique solution for the x_m since only five are linearly independent. Thus five of the x_m can be expressed in terms of the sixth. If one solves for x_0 to x_4 , the result is

$$\begin{aligned} x_0 &= x_5 - KDT_s(-0.837 + 0.557a) > 0 \\ x_1 &= x_5 - KDT_s(-1.041 + 1.393a) > 0 \\ x_2 &= x_5 - KDT_s(-0.484 + 1.599a) > 0 \\ x_3 &= x_5 - KDT_s(0.354 + 1.043a) < 0 \\ x_4 &= x_5 - KDT_s(0.558 + 0.204a) < 0 \\ x_5 & < 0 \end{aligned} \quad (9.4-6)$$

The inequalities on the right have been added as the conditions required to sustain the postulated y_m sequence. The easiest way to interpret these six conditions is to plot the boundaries between admissible and inadmissible values of x_s as functions of a . The conditions on x_2 and x_3 are found to conflict if $a > 0.303$. The conditions on x_0 and x_3 conflict if $a < -2.45$. Thus

$$-2.45 < a < 0.303$$

is the range of values of a for which the 3, 3 limit cycle mode is possible in this system. Values of a outside this range render the mode impossible, either because of too much lead in the linear part or too much lag. These boundaries on the range of a may be compared with -2.46 , 0.289 , derived using the sampled describing function, and -2.54 , 0.283 , obtained with the z -transform describing function. The errors in these describing function results range from 0.4 to 6.6 percent.

For values of a within the range from -2.45 to 0.303 , there is a range of values of x_s for which all the inequalities of Eqs. (9.4-6) are satisfied. This range of freedom for x_s , and thus for all the x_m , represents the range of possible bias levels which can exist in $x(t)$ in the presence of the 3, 3 limit cycle mode. For example, if there were no digital lead, $a = 0$, the 3, 3 mode would be possible and x_s could take values ranging from 0 to $-0.484KDT_s$. The symmetrical limit cycle would have $x_s = -0.242KDT_s$, and the null offset or bias could range between $\pm 0.242KDT_s$.

In the case of an even-period mode such as this, a simpler procedure can be employed. The critical condition for the existence of a limit cycle mode always occurs for the unbiased, or symmetrical, mode. If the symmetry condition is postulated,

$$x_{m+3} = -x_m \quad (9.4-7)$$

just three of the x_m define the mode, and there is no remaining uncertainty due to bias level. Using Eq. (9.4-7) in the first three of Eqs. (9.4-5),

$$\begin{aligned} x_0 + 1.368x_2 - 0.368x_1 &= -KDT_s(-0.632 + 0.632a) \\ x_1 - 1.368x_0 - 0.368x_2 &= -KDT_s(0.104 + 0.632a) \\ x_2 - 1.368x_1 + 0.368x_0 &= -KDT_s(0.632 - 0.104a) \end{aligned} \quad (9.4-8)$$

These equations have a unique solution. Each of these x_m must be greater than zero to reproduce the postulated y_m sequence. The critical conditions are

$$\begin{aligned} x_0 &= KDT_s(0.595 + 0.243a) > 0 \\ x_2 &= KDT_s(0.242 - 0.799a) > 0 \end{aligned} \quad (9.4-9)$$

which again gives $-2.45 < a < 0.303$ as the condition for the existence of the 3, 3 mode. For values of a in this range, the mode is possible, and the possible range of bias levels is evident.

THE TRANSFORM METHOD

This is the technique described by Torng (Ref. 25). It centers attention not on the sequences of samples of $y(t)$ and $x(t)$ directly, but rather on the coefficients of the expansions of these sequences into series of complex exponentials with integral arguments. The virtue of this is the fact that a periodic sequence can be processed through a linear discrete filter more readily in terms of these coefficients than in terms of the sequences themselves,

since a multiplicative input-output relation exists for the coefficients. In this and all other respects, a direct analogy exists between this theory and the familiar Fourier transform theory for continuous functions and systems.

A discrete sequence of n values,

$$y_m = y_0, y_1, \dots, y_{n-1} \quad (9.4-10)$$

which repeats periodically with the period n , can be expressed as a series of n complex exponential terms.

$$y_m = \sum_{l=-k}^{k+1} Y_l \exp \left(j l \frac{2\pi}{n} m \right) \quad (9.4-11)$$

where k is defined by

$$n = \begin{cases} 2k + 1 & n \text{ odd} \\ 2k + 2 & n \text{ even} \end{cases} \quad (9.4-12)$$

and the convention $Y_{k+1} = 0$ if n is odd will be observed. The coefficients in this expansion, or transformation, are given by

$$Y_l = \frac{1}{n} \sum_{m=0}^{n-1} y_m \exp \left(-j l \frac{2\pi}{n} m \right) \quad (9.4-13)$$

From this expression we note the property

$$Y_{-l} = Y_l^* \quad (9.4-14)$$

so the coefficients with negative indices need not be calculated separately.

If this periodic sequence y_m is the input to a discrete linear filter, and x_m is the output, and if this filter has the sampled transfer function $D(z)$,

$$\frac{X(z)}{Y(z)} = D(z) \quad (9.4-15)$$

then the coefficients of the expansion of the x_m sequence are given by

$$X_l = D \left[\exp \left(j l \frac{2\pi}{n} \right) \right] Y_l \quad (9.4-16)$$

If the modified or advanced z transform for the linear filter is used in this relation for X_l , then the entire $x(t)$ history is defined by the inverse transform [Eq. (9.4-11)].

The procedure for testing the existence of a postulated limit cycle mode using the transform method is then to transform the postulated y_m sequence, using Eq. (9.4-13), and to calculate the X_l coefficients from Eq. (9.4-16) and the x_m sequence from Eq. (9.4-11). The conditions on the x_m which will sustain the postulated y_m sequence can then be applied. Steady-state conditions are assumed throughout, since the form of the transformation is applicable only to periodic sequences, and the input-output relation for linear systems gives the forced response only.

Example 9.4-2 Repeat Example 9.4-1, using the transform method. The y_m sequence to be tested is

$$y_m = D, D, D, -D, -D, -D \quad (9.4-17)$$

which repeats with period $n = 6$. According to Eq. (9.4-12), $k = 2$ in this case. Y_l must be calculated for $l = 0, 1, 2, 3$, according to Eq. (9.4-13). As an illustration of this calculation, consider Y_1 .

$$\begin{aligned} Y_1 &= \frac{D}{6} \left[1 + \exp\left(-j\frac{\pi}{3}\right) + \exp\left(-j\frac{\pi}{3}2\right) \right. \\ &\quad \left. - \exp\left(-j\frac{\pi}{3}3\right) - \exp\left(-j\frac{\pi}{3}4\right) - \exp\left(-j\frac{\pi}{3}5\right) \right] \\ &= \frac{D}{3} \left[1 + \exp\left(-j\frac{\pi}{3}\right) + \exp\left(-j\frac{2\pi}{3}\right) \right] \\ &= \frac{D}{3} (1 - j\sqrt{3}) \end{aligned} \quad (9.4-18)$$

The other Y_l are computed in the same way.

$$\begin{aligned} Y_0 &= 0 \\ Y_2 &= 0 \\ Y_3 &= \frac{D}{3} \end{aligned} \quad (9.4-19)$$

The sampled transfer function for the linear part of this system is found from Eq. (9.4-2), using Eq. (9.4-3) and $T_s/\tau = 1$, to be

$$\frac{X}{Y}(z) = \frac{-KT_s[0.368z^2 + (0.264 - 0.368a)z - 0.264a]}{z^3 - 1.368z^2 + 0.368z} \quad (9.4-20)$$

The X_l are now calculated from Eq. (9.4-16). Take X_1 as an illustration.

$$\begin{aligned} X_1 &= \frac{X}{Y} \left[\exp\left(j\frac{\pi}{3}\right) \right] Y_1 \\ &= \frac{-KT_s[0.368 \exp(j2\pi/3) + (0.264 - 0.368a) \exp(j\pi/3) - 0.264a][(D/3)(1 - j\sqrt{3})]}{\exp(j\pi) - 1.368 \exp(j2\pi/3) + 0.368 \exp(j\pi/3)} \\ &= 0.435KDT_s[0.670 + 0.264a + j(-0.691 + 0.926a)] \end{aligned} \quad (9.4-21)$$

Needless to say, a fair amount of algebra has been omitted between these last steps. X_2 and X_3 are calculated in the same way.

$$\begin{aligned} X_2 &= 0 \\ X_3 &= 0.0127KDT_s(1 + a) \end{aligned} \quad (9.4-22)$$

The expression for X_0 , however, is indeterminate.

$$\begin{aligned} X_0 &= \frac{X}{Y}(1)Y_0 \\ &= \frac{0}{0} \end{aligned} \quad (9.4-23)$$

X_0 will be carried as an unknown coefficient. This is the manifestation of the undetermined bias level which can exist in the presence of the limit cycle.

The x_m sequence can now be calculated from Eq. (9.4-11), using the property $X_{-l} = X_l^*$.

$$\begin{aligned}
 x_m &= 0.435KDT_s[0.670 + 0.264a - j(-0.691 + 0.926a)] \exp\left(-j\frac{\pi}{3}m\right) \\
 &\quad + X_0 + 0.435KDT_s[0.670 + 0.264a + j(-0.691 + 0.926a)] \exp\left(j\frac{\pi}{3}m\right) \\
 &\quad + 0.0127KDT_s(1+a) \exp(j\pi m) \\
 &= 0.435KDT_s \left[\frac{2.30}{KDT_s} X_0 + 0.0291(-1)^m(1+a) \right. \\
 &\quad \left. + 2 \operatorname{Re} \left\{ [0.670 + 0.264a + j(-0.691 + 0.926a)] \exp\left(j\frac{\pi}{3}m\right) \right\} \right] \quad (9.4-24)
 \end{aligned}$$

Substitution of the various values of m in this expression yields the required results.

$$\begin{aligned}
 x_0 &= KDT_s \left(\frac{X_0}{KDT_s} + 0.595 + 0.243a \right) > 0 \\
 x_1 &= KDT_s \left(\frac{X_0}{KDT_s} + 0.797 - 0.596a \right) > 0 \\
 x_2 &= KDT_s \left(\frac{X_0}{KDT_s} + 0.242 - 0.799a \right) > 0 \\
 x_3 &= KDT_s \left(\frac{X_0}{KDT_s} - 0.595 - 0.243a \right) < 0 \\
 x_4 &= KDT_s \left(\frac{X_0}{KDT_s} - 0.797 + 0.596a \right) < 0 \\
 x_5 &= KDT_s \left(\frac{X_0}{KDT_s} - 0.242 + 0.799a \right) < 0
 \end{aligned} \quad (9.4-25)$$

Except for the X_0 terms, these relations are identical with those derived by the difference equation method under the assumption of a symmetrical limit cycle, and they define the same range of values of a for which the postulated mode is possible.

Application of either of these exact methods of determining or verifying the existence of limit cycle modes is considerably more tedious than use of the describing function techniques discussed in Sec. 9.2, especially the sampled describing function. An exact technique would ordinarily be required only when one has reason to question describing function results. Lack of sufficient continuous linear filtering is the primary reason for such a question. Another reason is a marginal situation in which a system design is near the boundary between possibility and impossibility for a particular

mode. Regardless of how the necessary conditions for existence of a limit cycle are established, the theory of Sec. 9.3 can be applied to test the stability of the mode.

9.5 LIMIT CYCLES IN PULSE-WIDTH-MODULATED SYSTEMS

A control system employing a pulse-width modulator, hereafter abbreviated PWM, is a special case of a sampled nonlinear system, and the techniques of the preceding sections can readily be applied to such a system. The primary reason for the utilization of a PWM in a control system is to achieve an approximation to proportional control in spite of a drive system which is only capable of—or for some reason is employed in—an on-off mode of operation. Somewhat varying forms of pulse-width modulators have been used, but the most common form is the linear lead modulator, which will be considered for the purpose of illustration here. The operation of this modulator is pictured in Fig. 9.5-1, and is defined by the following relations, which hold for all $m = 0, 1, 2, \dots$:

$$y(t) = \begin{cases} D \operatorname{sgn} [x(mT_s)] & mT_s < t < mT_s + k |x(mT_s)| \\ 0 & mT_s + k |x(mT_s)| < t < (m+1)T_s \end{cases} \quad (9.5-1)$$

if $k |x(mT_s)| < T_s$. Otherwise,

$$y(t) = D \operatorname{sgn} [x(mT_s)] \quad mT_s < t < (m+1)T_s \quad (9.5-2)$$

if $k |x(mT_s)| > T_s$.

Possible variations on this form of PWM are of two types. First, the relation between the sampled value of $x(t)$ and the pulse width may be any well-defined relation whatever, rather than the proportional relation considered here. Second, the zero interval in each sampling period of $y(t)$ may be placed in the first part of each period, with the nonzero interval following in the second part. This is known as a lag PWM. Neither of these variations complicates the application of the analytic techniques discussed here. In any case, even if the relation between $x(mT_s)$ and pulse width is linear, and if the possibility of saturation is ignored, the PWM operation is a nonlinear operation, and a nonlinear theory must be employed to study the details of the performance of a system using a PWM.

A number of writers have detailed various approaches to the exact determination of limit cycles in PWM systems. The same point of view is used as in the case of other sampled nonlinear systems: the form of limit cycle mode is postulated, which defines the form of the nonlinearity output; this

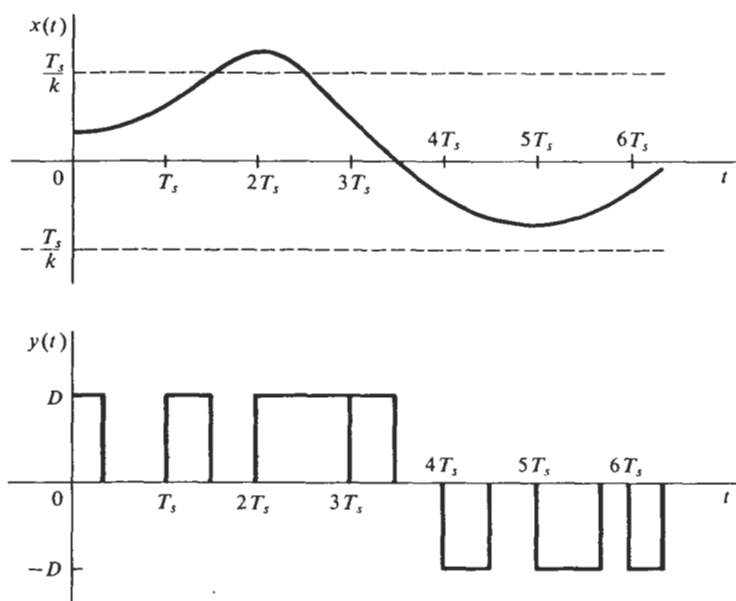


Figure 9.5-1 Operation of a linear lead PWM.

signal form is processed through the linear part of the system to the nonlinearity input; and conditions are applied which will produce the postulated output of the nonlinearity. Da-Chuan (Ref. 3) implemented this procedure for limit cycles of period $T = 2T_s$, using time response expressions directly. The procedure of Jury and Nishimura (Ref. 8), which was described in the preceding section in connection with other sampled nonlinear systems, can be applied equally well to PWM systems. Nease (Ref. 18) included PWM systems in his development of exact methods and general principles applicable to sampled nonlinear systems. The difference equation and transform methods described in Sec. 9.4 are not directly applicable to PWM systems because the input to the linear part of the system in this case is not a periodic train of impulses. However, direct time response, z transform, and Laplace transform techniques can all be used to calculate the response of the linear part to an input train of variable-width pulses.

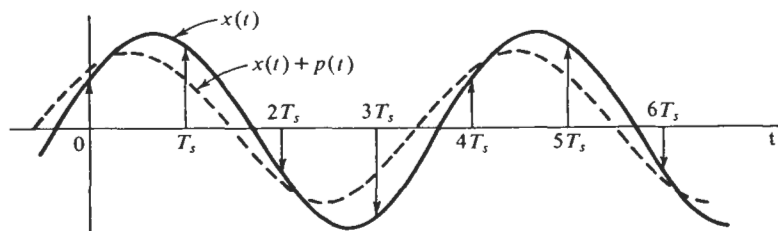
The describing function point of view can be applied to PWM systems in essentially the same way as it is applied to other sampled nonlinear systems. Delfeld and Murphy (Ref. 4) did this in an unnecessarily complicated way. Pyshkin (Ref. 21) took the straightforward approach, and treated limit cycles with periods $2T_s$ and $4T_s$. Calculation of the describing function for any particular form of PWM, such as the linear lead PWM considered for an illustration here, is quite simple in principle, the detail becoming tedious for

large-period modes because of the different modal forms of the same period which occur for different values of amplitude and sampling phase. The input to the modulator is taken in the standard form

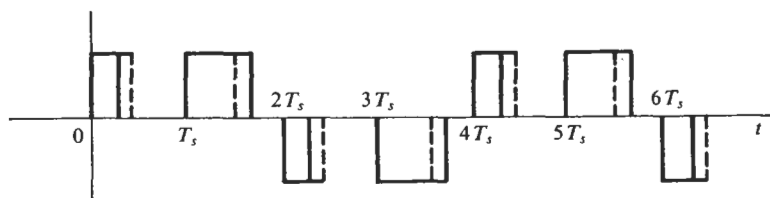
$$x(t) = A \sin(\omega t + \varphi) \quad (9.5-3)$$

with the time origin chosen at one of the sampling points. The frequencies of interest in limit cycle analysis are whole fractions of the sampling frequency—and in fact, just the even fractions, since one would not anticipate odd-period modes in a PWM system, especially if the linear part includes an integration. For any chosen frequency, the output of the PWM corresponding to any pair of values for A , φ is defined, and the fundamental harmonic component of this waveform is readily computed. Thus the describing function of the PWM is defined as a function of the ratio of the sinusoidal frequency to the sampling frequency, the amplitude of the input sinusoid, and the phase of that sinusoid relative to the sampling points. The negative reciprocal describing function for the linear lead PWM defined by Eqs. (9.5-1) and (9.5-2) is plotted in Appendix F for sinusoidal periods of 2, 4, 6, and 8 sampling periods. The possible existence of a limit cycle mode in a PWM system may be determined by placing the point corresponding to the transfer function for the linear part of the system, evaluated at fundamental frequency, on one of these plots. If the point lies within the region spanned by the describing function curves, one or more modes of that period are possible, and the corresponding amplitude and phase may be found by interpolation among the curves.

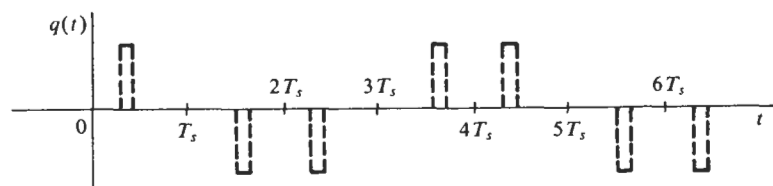
Any limit cycle found possible should be tested for stability. The technique described in Sec. 9.3 in relation to other sampled nonlinear systems can be applied, with a slight change of concept, to the present case as well. Consider a limit cycle $x(t)$ to experience a small additive perturbation $p(t)$. The output of the PWM is then a series of pulses with widths slightly perturbed from the pulse widths in the limit cycle. If the limit cycle signal at $y(t)$ is subtracted from this perturbed signal, the remaining perturbation $q(t)$ is a series of narrow pulses occurring near the ends of the pulses in the limit cycle. These signals are pictured in Fig. 9.5-2. In view of the fact that PWM systems are usually designed with a sampling period short compared with the response time of the linear part of the system, so the ripple at sampling frequency is well attenuated at the output of the system, the pulses in the output perturbation shown in Fig. 9.5-2c are very narrow compared with the linear-part response time, and can properly be approximated as impulses of the same area. The relation between input perturbation and output perturbation approximated as a sequence of impulses occurring at the ends of the pulses in the unperturbed limit cycle can then be linearized. In general, this requires the slope of the function which relates sampled input magnitude to pulse width. For the linear function of Eq.



(a) PWM input: limit cycle plus perturbation



(b) PWM output: limit cycle plus perturbation



(c) Perturbation at PWM output

Figure 9.5-2 Operation of a PWM with a perturbed limit cycle input.

(9.5-1), this slope is just k if the pulse is unsaturated, or zero if the pulse is saturated. If this slope, evaluated at the unperturbed limit cycle point $x(mT_s)$, is called K_m , then the perturbation $p(mT_s)$ gives rise to an approximate impulse of area $p(mT_s)K_mD$, occurring at the end of the pulse corresponding to $x(mT_s)$.

This linearized relation between input and output perturbations can now be represented by a parallel set of phased samplers with fixed gains, as was done in Sec. 9.3. For a limit cycle of period $T = nT_s$, where n is an even integer, $n/2$ such samplers are required, each with the sampling period $\frac{1}{2}nT_s$. The configuration is shown in Fig. 9.5-3. This linearized representation of the operation of the PWM on perturbations added to a limit cycle can be used with the transfer function for the linear part of the system to define a linear fixed-parameter sampled-data system whose stability determines the stability of the limit cycle under consideration. It may be noted that the gain K_m

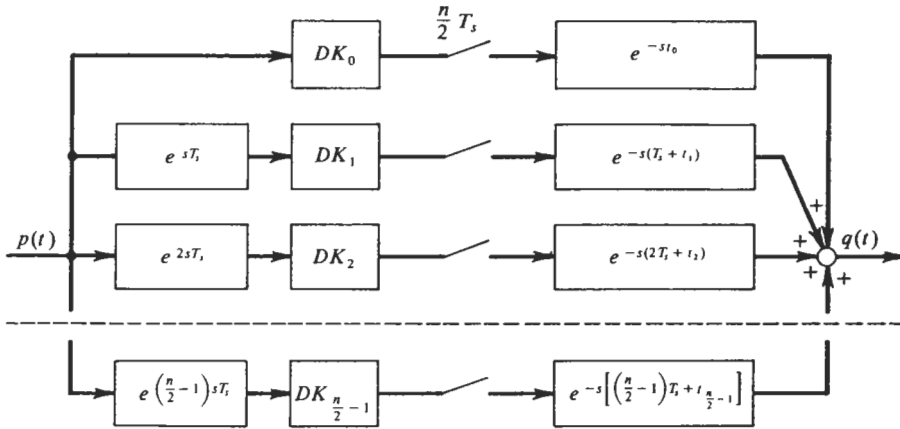


Figure 9.5-3 Linearized equivalent of PWM for perturbation analysis.

K_m = slope of function relating input magnitude to pulse width, evaluated at the limit cycle value $x(mT_s)$.

t_m = pulse width corresponding to the limit cycle value $x(mT_s)$.

associated with any saturated pulse is zero, and in effect the corresponding path is omitted from the configuration of Fig. 9.5-3. If a fully saturated limit cycle mode is considered, one in which every pulse is saturated, the PWM does not pass perturbations at all. Such a mode is thus stable if the open-loop linear part is stable, and unstable if the open-loop linear part is unstable.

REFERENCES

1. Bergen, A. R.: Discussion of Torng and Meserve (Ref. 26), *IRE Trans. Autom. Control*, vol. AC-5, no. 4 (September, 1960), pp. 304-305.
2. Chow, C. K.: Contactor Servomechanisms Employing Sampled Data, *Trans. AIEE*, pt. II, vol. 73 (March, 1954), pp. 51-61.
3. Da-Chuan, S.: On the Possibility of Certain Types of Oscillations in Sampled-data Control Systems, *Automation and Remote Control*, vol. 20, no. 1 (July, 1959), pp. 77-82.
4. Delfeld, F. R., and G. J. Murphy: Analysis of Pulse-width-modulated Control Systems, *IRE Trans. Autom. Control*, vol. AC-6, no. 3 (September, 1961), pp. 283-292.
5. Dixon, M. V.: Nonlinear Sampled-data System Analysis Using Describing Functions, M.S. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, Cambridge, Mass., January, 1965.
6. Gocłowski, J. C.: Analysis of a Sampled-data Relay Servo with Hysteresis, *NEREM Record*, 1963.
7. Jury, E. I.: "Sampled-data Control Systems," John Wiley & Sons, Inc., New York, 1958.

8. Jury, E. I., and T. Nishimura: On the Periodic Modes of Oscillation in Pulse-width-modulated Feedback Systems, *Trans. ASME, J. Basic Eng.*, March, 1962, pp. 71-84.
9. Kazakov, V. P.: Study of Pulsed Contactor Automatic Control System Dynamics, *Automation and Remote Control*, vol. 18, no. 1 (April, 1958), pp. 37-49.
10. Kazakov, V. P.: Influence of Hysteresis on the Mode of Periodic Processes in Pulsed Relay Systems, *Automation and Remote Control*, vol. 22, no. 5 (November, 1961), pp. 530-533.
11. Kinnen, I. E., and J. Tou: Analysis of Nonlinear Sampled-data Control Systems, *Trans. AIEE*, pt. II, vol. 78 (January, 1960), pp. 386-394.
12. Korshunov, Y. M.: The Analysis of Periodic States Due to Level Quantization of Signals in Automatic Digital Systems, *Automation and Remote Control*, vol. 22, no. 7 (December, 1961), pp. 778-787.
13. Korshunov, Y. M.: Plotting the Equivalent Complex Amplification Factor of a Nonlinear Pulse Element, *Automation and Remote Control*, vol. 23, no. 5 (November, 1962), pp. 537-547.
14. Kuo, B. C.: The z-transform Describing Function for Nonlinear Sampled-data Control Systems, *Proc. IRE*, vol. 48, no. 5 (May, 1960), pp. 941-942.
15. Kuo, B. C.: "Analysis and Synthesis of Sampled-data Control Systems," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.
16. Lepschy, A., and A. Ruberti: The Describing Function for the Study of Sampled-data Control Systems with a Piece-wise Nonlinearity, *Alta Frequenza*, vol. 32 (May, 1963), pp. 357-365.
17. Meserve, W. E., and H. C. Torng: Investigation of Periodic Modes of Sampled-data Control Systems Containing a Saturating Element, *Trans. ASME, J. Basic Eng.*, vol. D-83, no. 1 (March, 1961), pp. 77-81.
18. Nease, R. F.: Analysis and Design of Nonlinear Sampled-data Control Systems, *WADC Tech. Note 57-162, M.I.T. Servomech. Lab.*, June, 1957.
19. Pai, M. A.: Oscillations in Nonlinear Sampled-data Systems, *AIEE Paper CP 62-91*, December, 1961.
20. Pechorina, N. N.: The Stability of Pulse-width Modulated Automatic Control Systems, *Bull. Acad. Sci. U.S.S.R., Power Engineering and Automation*, no. 2, 1960.
21. Pyshkin, I. V.: Limit cycles in Pulse-width Modulated Systems, reprinted in *Theory Appl. Discrete Autom. Systems*, Academy of Sciences of the U.S.S.R., Moscow, 1960, pp. 132-150.
22. Ragazzini, J. R., and G. F. Franklin: "Sampled-data Control Systems," McGraw-Hill Book Company, New York, 1958.
23. Russell, F. A.: Discussion of Chow (Ref. 2), *Trans. AIEE*, pt. II, vol. 73 (March, 1954), pp. 62-64.
24. Simkin, M. M.: The Use of the Describing Function in Nonlinear Pulse Systems, *Automation and Remote Control*, vol. 22, no. 11 (April, 1962), pp. 1345-1353.
25. Torng, H. C.: Complete and Exact Identification of Self-sustained Oscillations in Relay Sampled-data Control Systems, *AIEE Paper CP 62-1059*, May, 1962.
26. Torng, H. C., and W. E. Meserve: Determination of Periodic Modes in Relay Servomechanisms Employing Sampled Data, *IRE Trans. Autom. Control*, vol. AC-5, no. 4 (September, 1960), pp. 298-303.
27. Tou, J. T.: "Digital and Sampled-data Control Systems," McGraw-Hill Book Company, New York, 1959.
28. Tsytkin, Y. Z.: Investigation of Stability of Periodic States in Nonlinear Pulse Automatic Systems, *Automation and Remote Control*, vol. 22, no. 6 (December, 1961), pp. 614-623.

PROBLEMS

- 9-1. Consider the sampled ideal relay shown in Fig. 9-1, with $x(t) = A \sin(\omega t + 45^\circ)$. The t scale has one of the sampling points at its origin. Sketch the waveforms of $x(t)$, $y(t)$, and $y^*(t)$ for several cycles of $x(t)$, in the case $\omega/\omega_s = 1/\pi$. Also indicate on a frequency scale the locations of the harmonic components of $y^*(t)$. What do you conclude about the applicability of describing function theory?

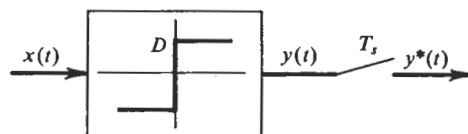


Figure 9-1

- 9-2. Repeat Prob. 9-1 with $\omega/\omega_s = \frac{2}{7}$.
 9-3. Repeat Prob. 9-1 with $\omega/\omega_s = \frac{1}{4}$.
 9-4. (a) For $T_s = 1$ sec, $KD = 10$, $\tau = 3$ sec, find the possible limit cycle modes of the system of Fig. 9-2 and the amplitude of each at $c(t)$.
 (b) What is the general relation between T_s and τ which will ensure that there can be no very large amplitude limit cycle?

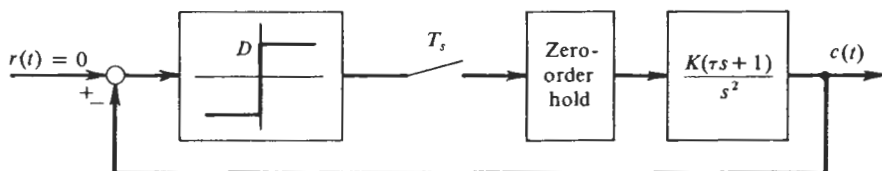


Figure 9-2

- 9-5. What limit cycle modes might the system of Fig. 9-3 display with no compensation, $G(s) = 1$? What is the maximum possible average offset at the output?

Design $G(s)$ so that this system has only the 1, 1 limit cycle mode. What now is the maximum possible average offset at the output? Suggest an additional compensation which will reduce this offset by a factor of 10.

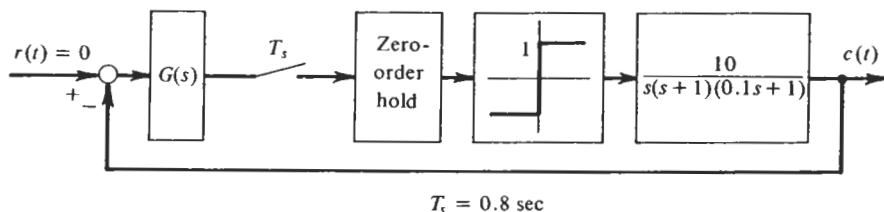


Figure 9-3

- 9-6. Find the maximum permissible relay hysteresis δ such that the system of Fig. 9-4 will display no limit cycle mode of lower frequency than the 3, 3 mode.

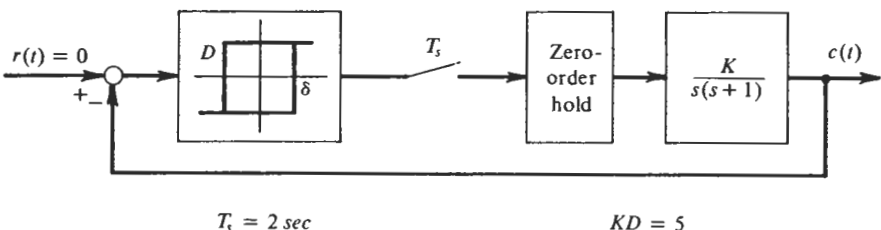


Figure 9-4

- 9-7. Calculate the z -transform describing function for the sampled relay with dead zone for each of the modes shown in Fig. 9.2-8. By comparison with Eqs. (9.2-40), verify in these cases the general relation

$$N^*(A, \varphi) = T_s N(A, \varphi) \quad \omega < \frac{1}{2}\omega_s$$

- 9-8. Use the sampled describing function method as presented in Sec. 9.2 to find the range of τ/T_s for which the 2, 2 limit cycle mode is possible in the system of Fig. 9-5. Interpret your result in terms of the graphical construction suggested for two-level relay systems in Sec. 9.1.

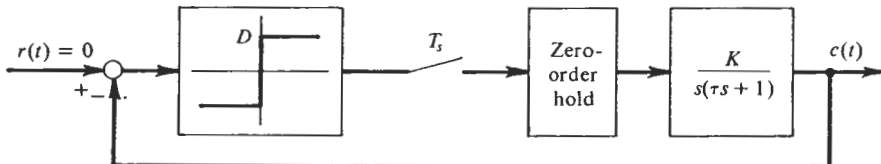


Figure 9-5

- 9-9. Solve Prob. 9-8 using the z -transform describing function method.
- 9-10. For the system of Fig. 9-5, derive the conditions which define the 1, 1 limit cycle mode, using both the sampled and z -transform describing function methods. Note that in this case, where $\omega = \frac{1}{2}\omega_s$, the z -transform describing function method gives a necessary condition, but not sufficient conditions to define the limit cycle.
- 9-11. The system of Fig. 9-6 uses a unit-sensitivity digital lead compensator to stabilize an inertia plant. Use the sampled describing function method to determine the range of τ/T_s which makes the 4, 4 limit cycle mode impossible.

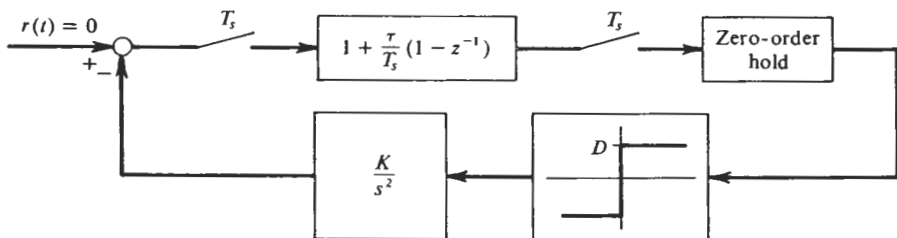


Figure 9-6

9-12. Solve Prob. 9-11, using the z -transform describing function method.

9-13. What is the minimum permissible value of dead zone, δ , which guarantees that the system of Fig. 9-7 will not limit-cycle in the absence of input?

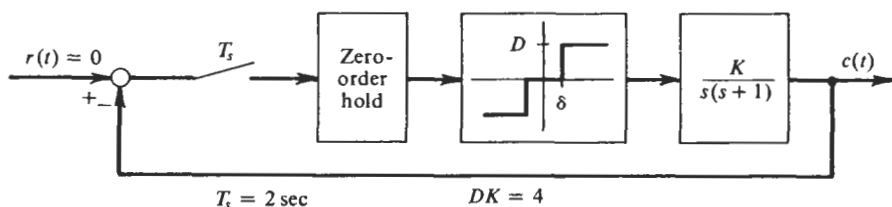


Figure 9-7

9-14. If $\delta = 1$, in the system of Fig. 9-7, what limit cycle modes are possible? Suggest a compensation which will eliminate all but the 1, 1 mode.

9-15. Determine the range of k for which a limit cycle of period $T = 4T_s$ and form $a, 0, -a, 0$ would be stable in the system of Fig. 9-8.

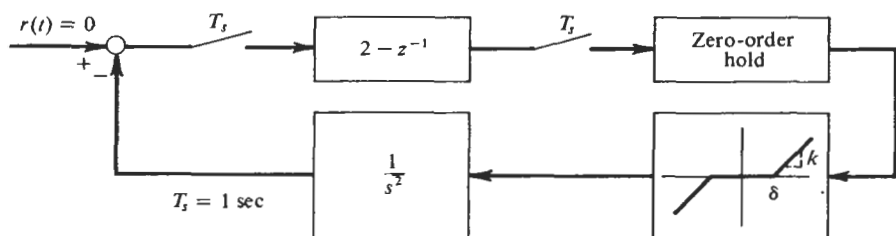


Figure 9-8

9-16. Use the difference-equation method to find the exact range of τ/T_s which makes the 4, 4 limit cycle mode impossible in the system of Fig. 9-6. Compare this result with those given by the describing functions in Probs. 9-11 and 9-12.

9-17. Repeat Prob. 9-16, using the transform method.

9-18. What is the maximum gain K that the pulse-width-modulated system of Fig. 9-9 can tolerate without exhibiting any of the limit cycle modes for which the describing function is given in Appendix F? Suggest a compensation which will permit this gain to be doubled without a limit cycle.

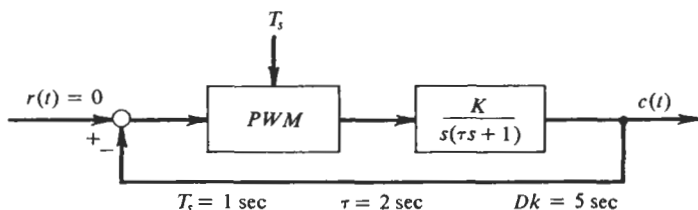


Figure 9-9

APPENDIX A

AMPLITUDE-RATIO-DECIBEL CONVERSION TABLE

This table is a tabulation of the relationship

$$\text{db} = 20 \log_{10} N$$

where

db = number of decibels

N = amplitude ratio

AMPLITUDE-RATIO-DECIBEL CONVERSION TABLE

N	0	1	2	3	4	5	6	7	8	9
0.01	-40.00	-39.17	-38.42	-37.72	-37.08	-36.48	-35.92	-35.39	-34.89	-34.42
0.02	-33.98	-33.56	-33.15	-32.77	-32.40	-32.04	-31.70	-31.37	-31.06	-30.75
0.03	-30.46	-30.17	-29.90	-29.63	-29.37	-29.12	-28.87	-28.64	-28.40	-28.18
0.04	-27.96	-27.74	-27.53	-27.33	-27.13	-26.94	-26.74	-26.56	-26.38	-26.20
0.05	-26.02	-25.85	-25.68	-25.51	-25.35	-25.19	-25.04	-24.88	-24.73	-24.58
0.06	-24.44	-24.29	-24.15	-24.01	-23.88	-23.74	-23.61	-23.48	-23.35	-23.22
0.07	-23.10	-22.97	-22.85	-22.73	-22.62	-22.50	-22.38	-22.27	-22.16	-22.05
0.08	-21.94	-21.83	-21.72	-21.62	-21.51	-21.41	-21.31	-21.21	-21.11	-21.01
0.09	-20.92	-20.82	-20.72	-20.63	-20.54	-20.45	-20.35	-20.26	-20.18	-20.09
0.1	-20.00	-19.17	-18.42	-17.72	-17.08	-16.48	-15.92	-15.39	-14.89	-14.42
0.2	-13.98	-13.56	-13.15	-12.77	-12.40	-12.04	-11.70	-11.37	-11.06	-10.75
0.3	-10.46	-10.17	-9.90	-9.63	-9.37	-9.12	-8.87	-8.64	-8.40	-8.18
0.4	-7.96	-7.74	-7.53	-7.33	-7.13	-6.94	-6.74	-6.56	-6.38	-6.20
0.5	-6.02	-5.85	-5.68	-5.51	-5.35	-5.19	-5.04	-4.88	-4.73	-4.58
0.6	-4.44	-4.29	-4.15	-4.01	-3.88	-3.74	-3.61	-3.48	-3.35	-3.22
0.7	-3.10	-2.97	-2.85	-2.73	-2.62	-2.50	-2.38	-2.27	-2.16	-2.05
0.8	-1.94	-1.83	-1.72	-1.62	-1.51	-1.41	-1.31	-1.21	-1.11	-1.01
0.9	-0.92	-0.82	-0.72	-0.63	-0.54	-0.45	-0.35	-0.26	-0.18	-0.09
1.0	0.00	0.09	0.17	0.26	0.34	0.42	0.51	0.59	0.67	0.75
1.1	0.83	0.91	0.98	1.06	1.14	1.21	1.29	1.36	1.44	1.51
1.2	1.58	1.66	1.73	1.80	1.87	1.94	2.01	2.08	2.14	2.21
1.3	2.28	2.35	2.41	2.48	2.54	2.61	2.67	2.73	2.80	2.86
1.4	2.92	2.98	3.05	3.11	3.17	3.23	3.29	3.35	3.41	3.46

1.5	3.52	3.58	3.64	3.69	3.75	3.81	3.86	3.92	3.97	4.03
1.6	4.08	4.14	4.19	4.24	4.30	4.35	4.40	4.45	4.51	4.56
1.7	4.61	4.66	4.71	4.76	4.81	4.86	4.91	4.96	5.01	5.06
1.8	5.11	5.15	5.20	5.25	5.30	5.34	5.39	5.44	5.48	5.53
1.9	5.58	5.62	5.67	5.71	5.76	5.80	5.85	5.89	5.93	5.98
2	6.02	6.44	6.85	7.23	7.60	7.96	8.30	8.63	8.94	9.25
3	9.54	9.83	10.10	10.37	10.63	10.88	11.13	11.36	11.60	11.82
4	12.04	12.26	12.47	12.67	12.87	13.06	13.26	13.44	13.62	13.80
5	13.98	14.15	14.32	14.49	14.65	14.81	14.96	15.12	15.27	15.42
6	15.56	15.71	15.85	15.99	16.12	16.26	16.39	16.52	16.65	16.78
7	16.90	17.03	17.15	17.27	17.38	17.50	17.62	17.73	17.84	17.95
8	18.06	18.17	18.28	18.38	18.49	18.59	18.69	18.79	18.89	18.99
9	19.08	19.18	19.28	19.37	19.46	19.55	19.65	19.74	19.82	19.91
10	20.00	20.83	21.58	22.28	22.92	23.52	24.08	24.61	25.11	25.58
20	26.02	26.44	26.85	27.23	27.60	27.96	28.30	28.63	28.94	29.25
30	29.54	29.83	30.10	30.37	30.63	30.88	31.13	31.36	31.60	31.82
40	32.04	32.26	32.47	32.67	32.87	33.06	33.26	33.44	33.62	33.80
50	33.98	34.15	34.32	34.49	34.65	34.81	34.96	35.12	35.27	35.42
60	35.56	35.71	35.85	35.99	36.12	36.26	36.39	36.52	36.65	36.78
70	36.90	37.03	37.15	37.27	37.39	37.50	37.62	37.73	37.84	37.95
80	38.06	38.17	38.28	38.38	38.49	38.59	38.69	38.79	38.89	38.99
90	39.08	39.18	39.28	39.37	39.46	39.55	39.65	39.74	39.82	39.91

APPENDIX B

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs)

The DF is given by (cf. Sec. 2.2)

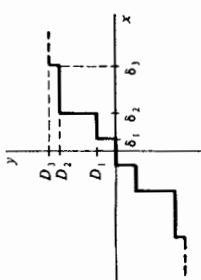
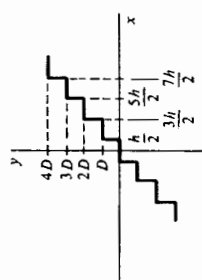
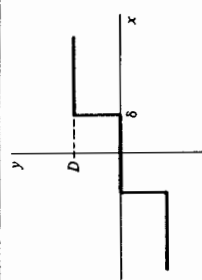
$$N(A, \omega) = n_p(A, \omega) + j n_q(A, \omega) = \frac{j}{\pi A} \int_0^{2\pi} y(A \sin \psi, A \omega \cos \psi) e^{-j\psi} d\psi$$

In this table we employ the "saturation function" (cf. Sec. 2.3) denoted by

$$\begin{aligned} f(\gamma) &= -1 & \gamma < -1 \\ &= \frac{2}{\pi} (\sin^{-1} \gamma + \gamma \sqrt{1 - \gamma^2}) & |\gamma| \leq 1 \\ &= 1 & \gamma > 1 \end{aligned}$$

This function is plotted in Fig. C.1.

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs) (Continued)

Nonlinearity	Comments	$n_p(A, \omega)$ and $n_q(A, \omega)$
 1. General odd quantizer	$A < \delta_1$ $\delta_{n+1} > A > \delta_n$	$n_p = 0$ $n_q = 0$ $n_p = \frac{4}{\pi A} \sum_{i=1}^n (D_i - D_{i-1}) \sqrt{1 - \left(\frac{\delta_i}{A}\right)^2}$ $n_q = 0$
 2. Uniform quantizer or granularity	$A < \frac{h}{2}$ $\frac{2n+1}{2} h > A > \frac{2n-1}{2} h$ See Fig. B.1 and Sec. 2.3	$n_p = 0$ $n_q = 0$ $n_p = \frac{4D}{\pi A} \sum_{i=1}^n \sqrt{1 - \left(\frac{2i-1}{2} \frac{h}{A}\right)^2}$ $n_q = 0$
 3. Relay with dead zone	$A < \delta$ $A > \delta$ See Fig. B.1	$n_p = 0$ $n_q = 0$ $n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2}$ $n_q = 0$

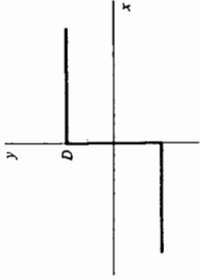
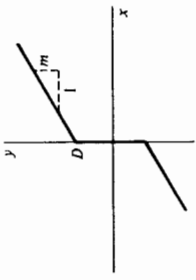
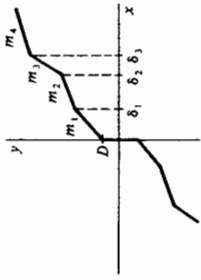
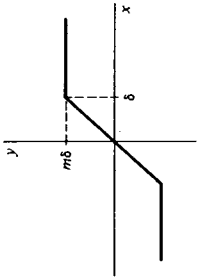
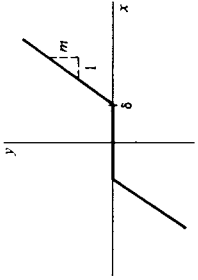
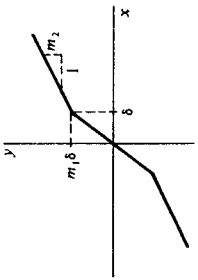
 4. Ideal relay	See Sec. 2.3	$n_p = \frac{4D}{\pi A}$ $n_q = 0$
 5. Preload	See Sec. 2.3	$n_p = \frac{4D}{\pi A} + m$ $n_q = 0$
 6. General piecewise-linear odd memoryless nonlinearity	$\delta_1 > A > 0$ $\delta_2 > A > \delta_1$ or, in a form valid for all A, $(\delta_{n+1} > A > \delta_n)$ See Sec. 2.3	$n_p = \frac{4D}{\pi A} + m_1$ $n_q = 0$ $n_p = \frac{4D}{\pi A} + (m_1 - m_2)f\left(\frac{\delta_1}{A}\right) + m_2$ $n_q = 0$ $n_p = \frac{4D}{\pi A} + \sum_{i=1}^n (m_i - m_{i+1})f\left(\frac{\delta_i}{A}\right) + m_{n+1}$ $n_q = 0$

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs) (Continued)

Nonlinearity	Comments	$n_p(A, \omega)$ and $n_q(A, \omega)$
 7. Saturation or limiter	See Fig. B.2 and Sec. 2.3	$n_p = mf\left(\frac{\delta}{A}\right)$ $n_q = 0$
 8. Dead zone or threshold	See Fig. B.2 and Sec. 2.3	$n_p = m\left[1 - f\left(\frac{\delta}{A}\right)\right]$ $n_q = 0$
 9. Gain-changing nonlinearity	See Sec. 2.3	$n_p = (m_1 - m_2)f\left(\frac{\delta}{A}\right) + m_2$ $n_q = 0$

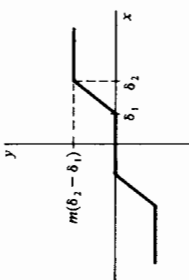
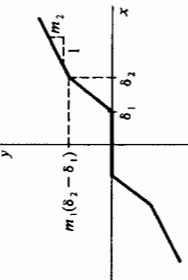
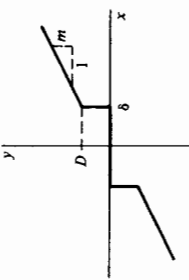
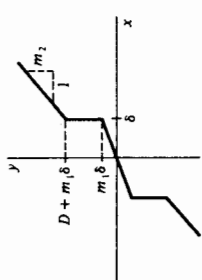
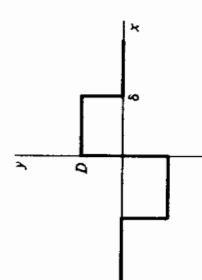
 10. Limiter with dead zone	See Sec. 2.3	$n_p = m \left[f \left(\frac{\delta_2}{A} \right) - f \left(\frac{\delta_1}{A} \right) \right]$ $n_q = 0$
 11. Gain-changing nonlinearity with dead zone		$n_p = -m_1 f \left(\frac{\delta_1}{A} \right) + (m_1 - m_2) f \left(\frac{\delta_2}{A} \right) + m_2$ $n_q = 0$
 12.	$A < \delta$ $A > \delta$	$n_p = 0$ $n_q = 0$ $n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A} \right)^2} + m \left[1 - f \left(\frac{\delta}{A} \right) \right]$ $n_q = 0$

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs) (Continued)

Nonlinearity	Comments	$n_p(A, \omega)$ and $n_q(A, \omega)$
 13.	$A < \delta$ $A > \delta$	$n_p = m_1$ $n_q = 0$ $n_p = (m_1 - m_3)f\left(\frac{\delta}{A}\right) + m_2 + \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2}$ $n_q = 0$
 14.	$A < \delta$ $A > \delta$	$n_p = \frac{4D}{\pi A}$ $n_q = 0$ $n_p = \frac{4D}{\pi A} \left[1 - \sqrt{1 - \left(\frac{\delta}{A}\right)^2} \right]$ $n_q = 0$
$y = c$ 15.		$n_p = 0$ $n_q = 0$
$y = x$ 16. Linear gain		$n_p = 1$ $n_q = 0$

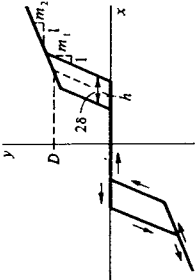
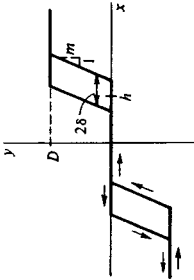
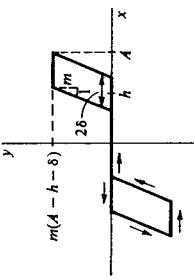
$y = x x $		$n_p = \frac{8}{3\pi} A$ $n_q = 0$
17. Odd square law	See Fig. B.3	
$y = x^3$		$n_p = \frac{3}{4} A^3$ $n_q = 0$
18. Cubic characteristic	See Fig. B.3	
$y = x^3 x $		$n_p = \frac{32}{15\pi} A^3$ $n_q = 0$
19. Odd quartic characteristic		
$y = x^5$		$n_p = \frac{5}{8} A^4$ $n_q = 0$
20. Quintic characteristic		
$y = x^5 x $		$n_p = \frac{64}{35\pi} A^5$ $n_q = 0$
21.		
$y = x^7$		$n_p = \frac{35}{64} A^6$ $n_q = 0$
22.		

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs) (Continued)

Nonlinearity	Comments	$n_p(A, \omega)$ and $n_q(A, \omega)$
23. $y = x^7 x $		$n_p = \frac{512}{315\pi} A^7$ $n_q = 0$
24. $y = x^n$	$n = 3, 5, 7, \dots$ See Fig. B.3 and Sec. 2.3	$n_p = \frac{n(n-2)(n-4) \cdots (3)}{(n+1)(n-1)(n-3) \cdots (4)} A^{n-1}$ $n_q = 0$
25. $y = x^{n-1} x $	$n = 2, 4, 6, \dots$ See Fig. B.3 and Sec. 2.3	$n_p = \frac{4}{\pi} \frac{n(n-2)(n-4) \cdots (2)}{(n+1)(n-1)(n-3) \cdots (3)} A^{n-1}$ $n_q = 0$
26. Odd square root $y = \sqrt{x} \quad (x \geq 0)$ $= -\sqrt{-x} \quad (x < 0)$	See Fig. B.3	$n_p = 1.11 A^{-1/2}$ $n_q = 0$
27. Cube root characteristic $y = x^{1/3}$		$n_p = 1.16 A^{-2/3}$ $n_q = 0$
28. $y = x^b \quad (x \geq 0)$ $= -(-x)^b \quad (x < 0)$	$b > -2$ $\Gamma(\text{arg.})$ is the gamma function. See Sec. 2.3	$n_p = \frac{2}{\sqrt{\pi}} \frac{\Gamma\left(\frac{b+2}{2}\right)}{\Gamma\left(\frac{b+3}{2}\right)} A^{b-1}$ $n_q = 0$

$y = M \sin mx$	$J_1(mA)$ is the Bessel function of order 1 for real arguments. See Fig. B.4 and Sec. 2.3	$n_p = 2M \frac{J_1(mA)}{A}$ $n_q = 0$
29. Harmonic nonlinearity $y = M \sinh mx$	$I_1(mA)$ is the modified Bessel function of order 1.	$n_p = 2M \frac{I_1(mA)}{A}$ $n_q = 0$
30. $y = 1 - e^{-cx} \quad (x \geq 0)$ $= -(1 - e^{cx}) \quad (x < 0)$	$I_1(cA)$ is the modified Bessel function of order 1 and $S_1(cA)$ is the modified Struve function of order 1. See Ref. 25 of Chap. 2	$n_p = \frac{2}{A} [I_1(cA) - S_1(cA)]$ $n_q = 0$
31. Exponential saturation $y = \frac{cx}{\sqrt{1 + (cx)^2}}$	$K(k)$ and $E(k)$ are the elliptic integrals of first and second kind, respectively.	$n_p = \frac{4}{\pi c A^2} \left[-\frac{1}{\sqrt{1 + (cA)^2}} K\left(\frac{cA}{\sqrt{1 + (cA)^2}}\right) + \sqrt{1 + (cA)^2} E\left(\frac{cA}{\sqrt{1 + (cA)^2}}\right) \right]$ $n_q = 0$
32. Algebraic saturation $y(t) = x(t - T_d)$	This phenomenon is linear, hence the DF is exactly the transfer function $\exp(-j\omega T_d)$.	$n_p = \cos \omega T_d$ $n_q = -\sin \omega T_d$
33. Time delay		

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs) (Continued)

Nonlinearity	Comments	$n_s(A, \omega)$ and $n_q(A, \omega)$
 34.	$m_1 > m_2$ $A \geq h + \frac{D}{m_1} + \frac{\delta m_1}{m_1 - m_2}$	$n_p = m_2 - \frac{m_1}{2} \left[f\left(\frac{h + \delta}{A}\right) + f\left(\frac{h - \delta}{A}\right) \right]$ $+ \left(\frac{m_1 - m_2}{2} \right) \left[f\left(\frac{h + D/m_1 + \delta m_1/(m_1 - m_2)}{A}\right) \right. \\ \left. + f\left(\frac{h + D/m_1 - \delta m_1/(m_1 - m_2)}{A}\right) \right]$ $n_q = -\frac{4D\delta}{\pi A^2}$
 35.	$A > h + \frac{D}{m} + \delta$	$n_p = \frac{m}{2} \left[-f\left(\frac{h + \delta}{A}\right) + f\left(\frac{h + \delta + D/m}{A}\right) + f\left(\frac{h - \delta + D/m}{A}\right) - f\left(\frac{h - \delta}{A}\right) \right]$ $n_q = -\frac{4D\delta}{\pi A^2}$
 36.	$A > h + \delta$	$n_p = \frac{m}{2} \left[1 + f\left(1 - \frac{2\delta}{A}\right) - f\left(\frac{h + \delta}{A}\right) - f\left(\frac{h - \delta}{A}\right) \right]$ $n_q = -\frac{4\delta m}{\pi A^2} (A - h - \delta)$

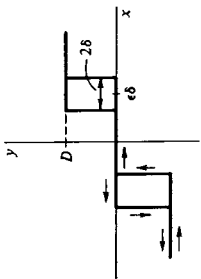
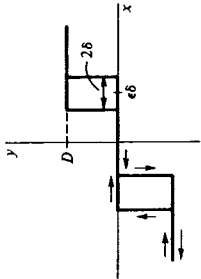
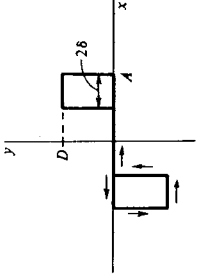
 37. (positive) Hysteresis	$A < \delta(1 + \epsilon)$ $A > \delta(1 + \epsilon)$ See Fig. B.5 and Sec. 2.3	$n_p = 0$ $n_q = 0$ $n_p = \frac{2D}{\pi A} \left[\sqrt{1 - \left(\frac{\delta}{A}\right)^2} (1 - \epsilon)^2 + \sqrt{1 - \left(\frac{\delta}{A}\right)^2} (1 + \epsilon)^2 \right]$ $n_q = -\frac{4D\delta}{\pi A^2}$
 38. (negative) Hysteresis	$A < \delta(\epsilon - 1)$ $\delta(\epsilon - 1) < A < \delta(\epsilon + 1)$ $A > \delta(\epsilon + 1)$	$n_p = 0$ $n_q = 0$ $n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} (1 - \epsilon)^2$ $n_q = 0$ $n_p = \frac{2D}{\pi A} \left[\sqrt{1 - \left(\frac{\delta}{A}\right)^2} (1 - \epsilon)^2 + \sqrt{1 - \left(\frac{\delta}{A}\right)^2} (1 + \epsilon)^2 \right]$ $n_q = \frac{4D\delta}{\pi A^2}$
 39.	$A > \delta$	$n_p = \frac{2D}{\pi A} \sqrt{1 - \left(1 - \frac{2\delta}{A}\right)^2}$ $n_q = -\frac{4D\delta}{\pi A^2}$

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs) (Continued)

Nonlinearity	Comments	$n_p(A,\omega)$ and $n_q(A,\omega)$
40.	$A > \delta + \frac{D}{m}$	$n_p = \frac{m}{2} \left[2 - f \left(\frac{\frac{D}{m} + \delta}{A} \right) + f \left(\frac{\frac{D}{m} - \delta}{A} \right) \right]$ $n_q = - \frac{4D\delta}{\pi A^2}$
41.	$m_1 > m_2$ $A > \frac{D}{m_1} + \frac{\delta m_1}{m_1 - m_2}$	$n_p = \frac{m_1 - m_2}{2} \left[f \left(\frac{\frac{D}{m_1} + \frac{m_1 \delta}{m_1 - m_2}}{A} \right) + f \left(\frac{\frac{D}{m_1} - \frac{m_1 \delta}{m_1 - m_2}}{A} \right) \right] + m_2$ $n_q = - \frac{4D\delta}{\pi A^2}$
42.	$A > \frac{D}{m} + \delta$	$n_p = \frac{m}{2} \left[f \left(\frac{\frac{D}{m} + \delta}{A} \right) + f \left(\frac{\frac{D}{m} - \delta}{A} \right) \right]$ $n_q = - \frac{4D\delta}{\pi A^2}$

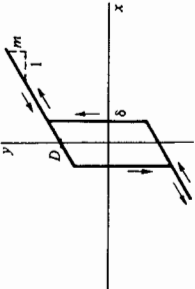
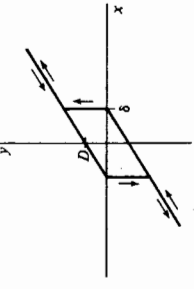
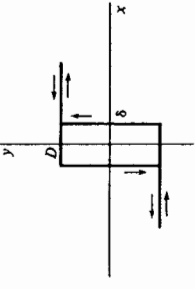
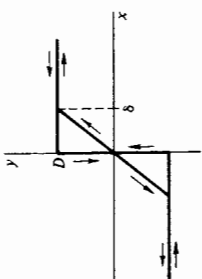
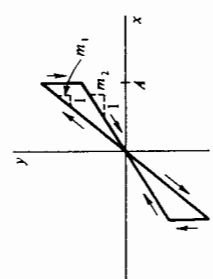
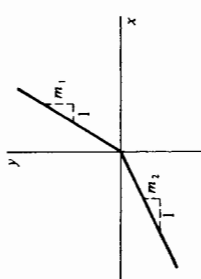
	$A > \delta$	$n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} + m$ $n_q = -\frac{4D\delta}{\pi A^2}$
43.	See Sec. 2.5	
	$A > \delta$	$n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2} + \frac{D}{\delta}$ $n_q = -\frac{4D\delta}{\pi A^2}$
44. Negative deficiency	See Fig. B.6	
	$A > \delta$	$n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta}{A}\right)^2}$ $n_q = -\frac{4D\delta}{\pi A^2}$
45. Rectangular hysteresis or toggle	See Fig. B.6 and Sec. 2.3	

TABLE OF SINUSOIDAL-INPUT DESCRIBING FUNCTIONS (DFs) (Continued)

Nonlinearity	Comments	$n_p(A, \omega)$ and $n_q(A, \omega)$
46.		$n_p = m$ $n_q = -\frac{4D}{\pi A}$
47.		$n_p = 0$ $n_q = -\frac{4D}{\pi A}$
48. Friction-controlled backlash	$A > \frac{b}{2}$ See Fig.B.7 and Sec. 2.3	$n_p = \frac{1}{2} \left[1 + f \left(1 - \frac{b}{A} \right) \right]$ $n_q = -\frac{1}{2} \left[\frac{b}{A} - \left(\frac{b}{A} \right)^2 \right]$

	$A > \delta$ Multivalued nonlinearity for which $n_q(A, \omega) = 0$.	$n_p = \frac{D}{2\delta} f\left(\frac{\delta}{A}\right) + \frac{2D}{\pi A}$ $n_q = 0$
	Multivalued nonlinearity for which the DF is independent of A.	$n_p = \frac{m_1 + m_2}{2}$ $n_q = \frac{m_1 - m_2}{\pi}$
	Asymmetric characteristic equivalent to the parallel combination of a linear gain $(m_1 + m_2)/2$ and an absolute value characteristic $(m_1 - m_2)/2$. The even part does not contribute to the DF.	$n_p = \frac{m_1 + m_2}{2}$ $n_q = 0$

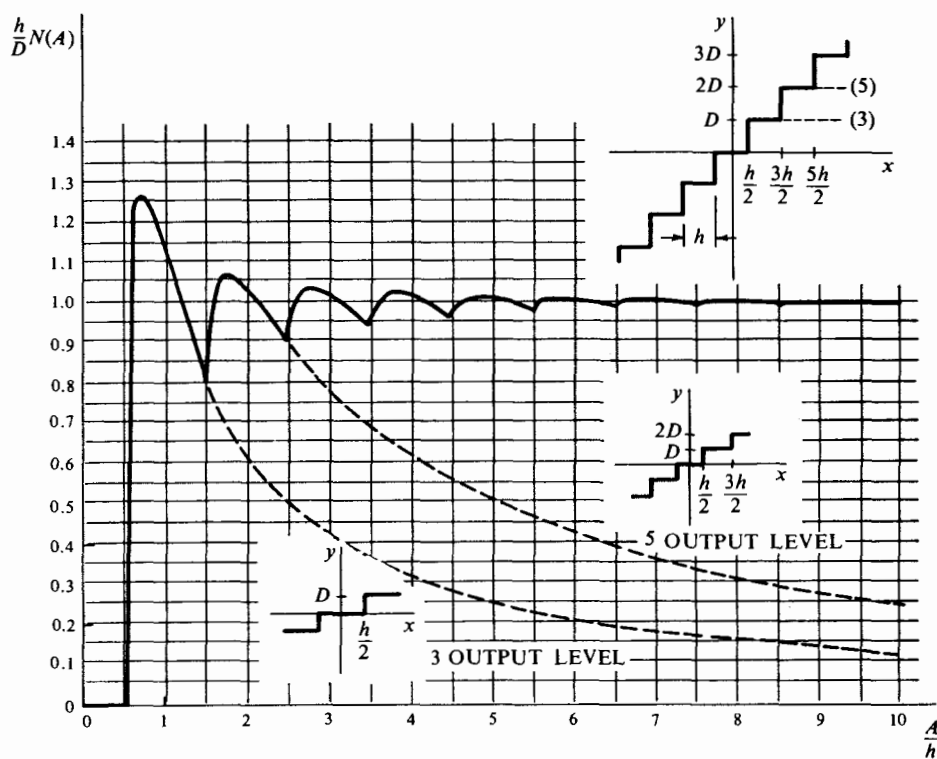


Figure B.1 Quantizer DF.

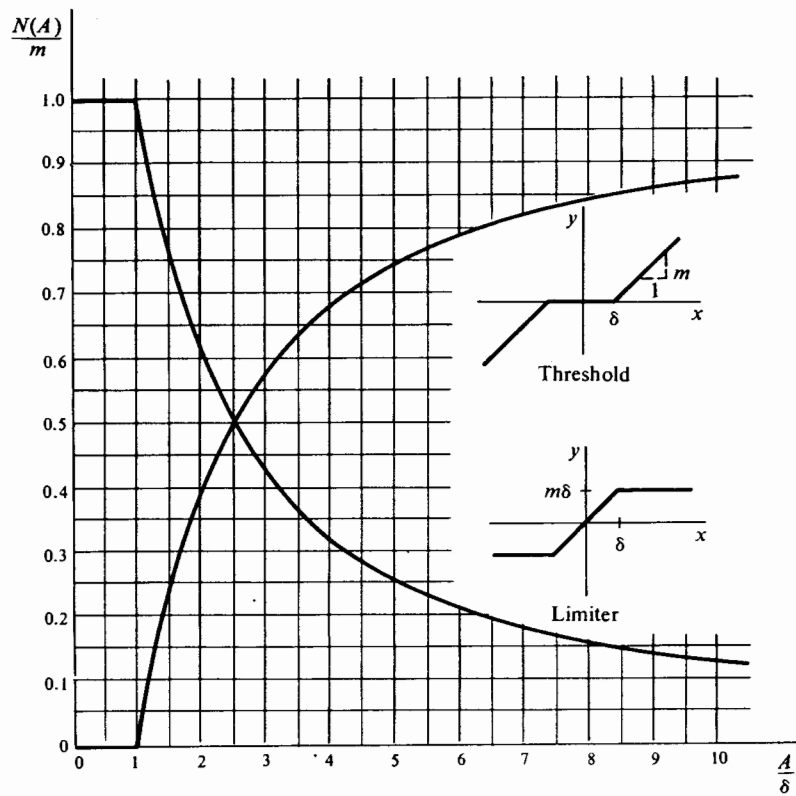


Figure B.2 DFs for limiter and threshold characteristics.

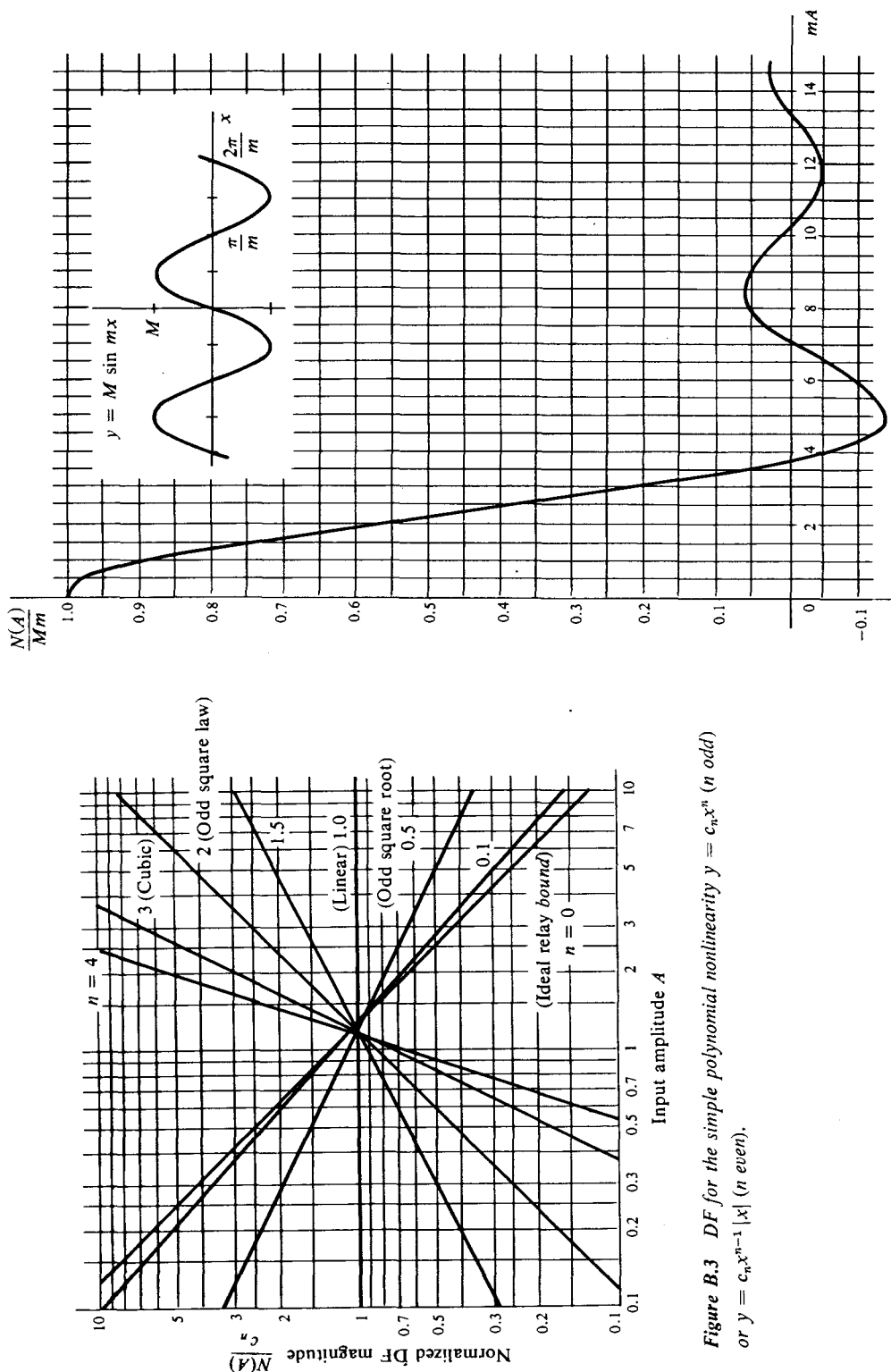


Figure B.3 DF for the simple polynomial nonlinearity $y = c_n x^n$ (n odd) or $y = c_n x^{n-1} |x|$ (n even).

Figure B.4 Normalized harmonic nonlinearity DF.

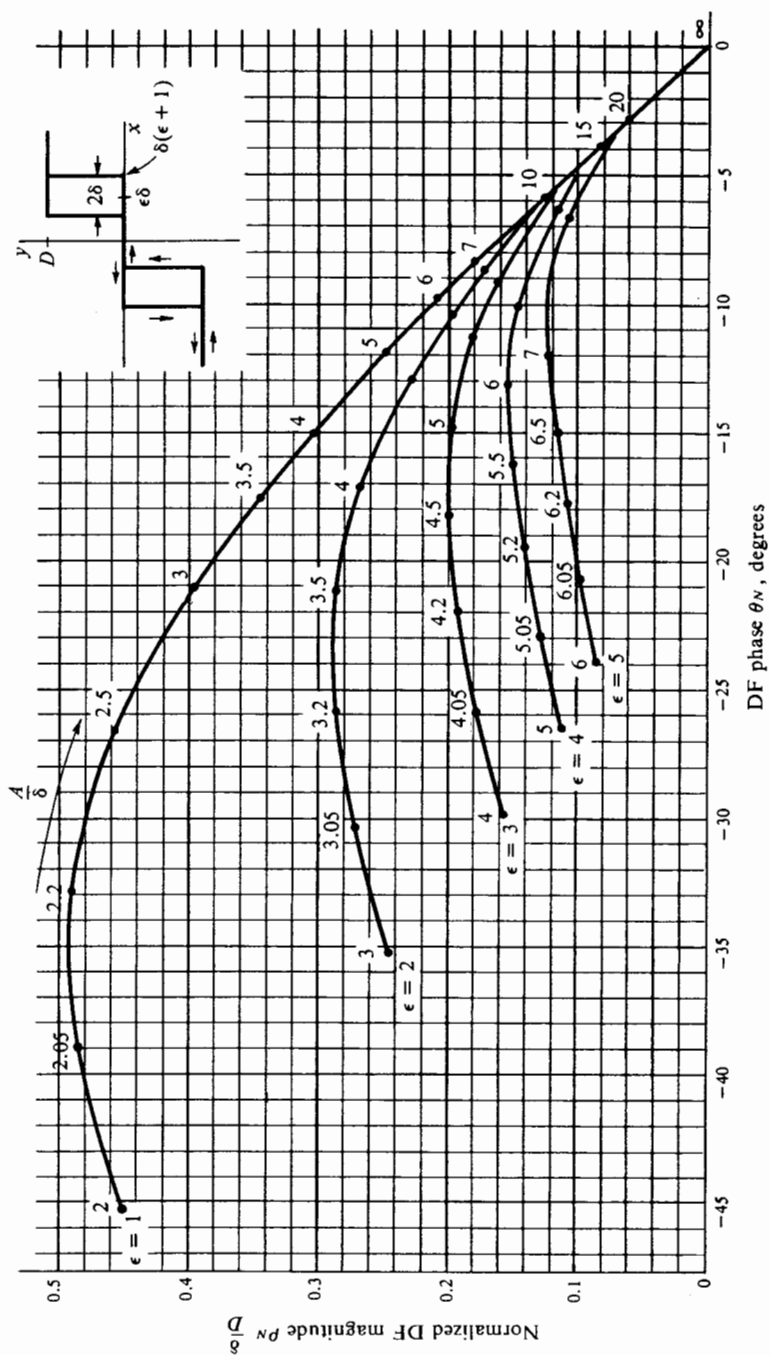


Figure B.5 DF magnitude vs. phase for hysteresis characteristics.

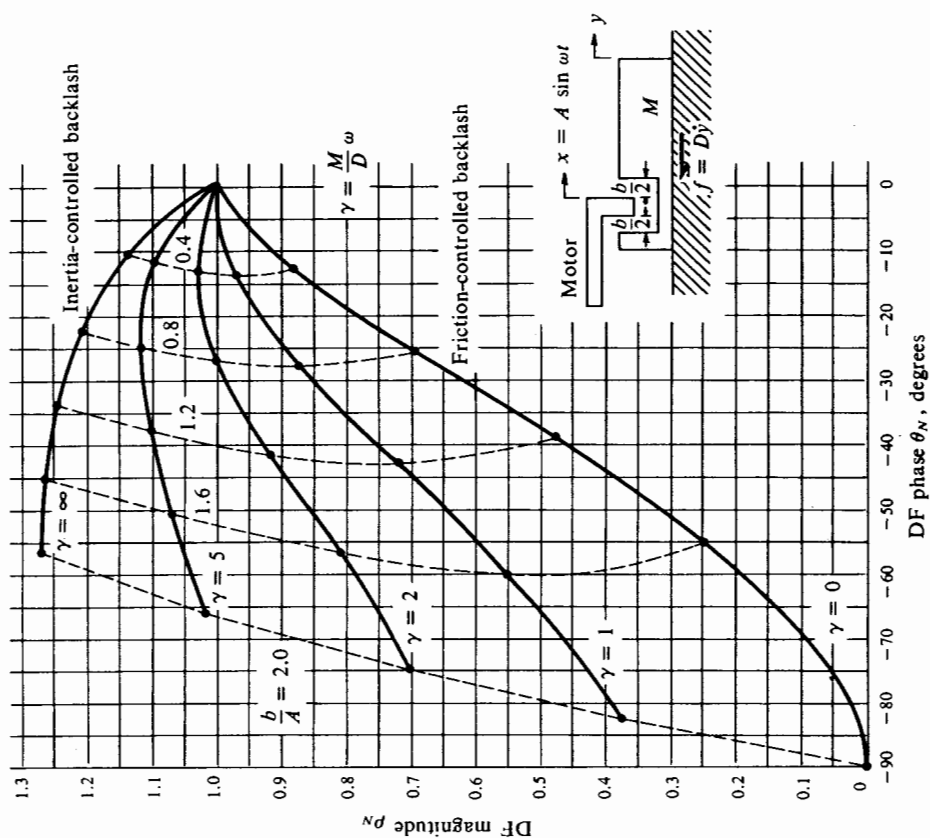


Figure B.7 DF for backlash with inertia and viscous-friction loading.

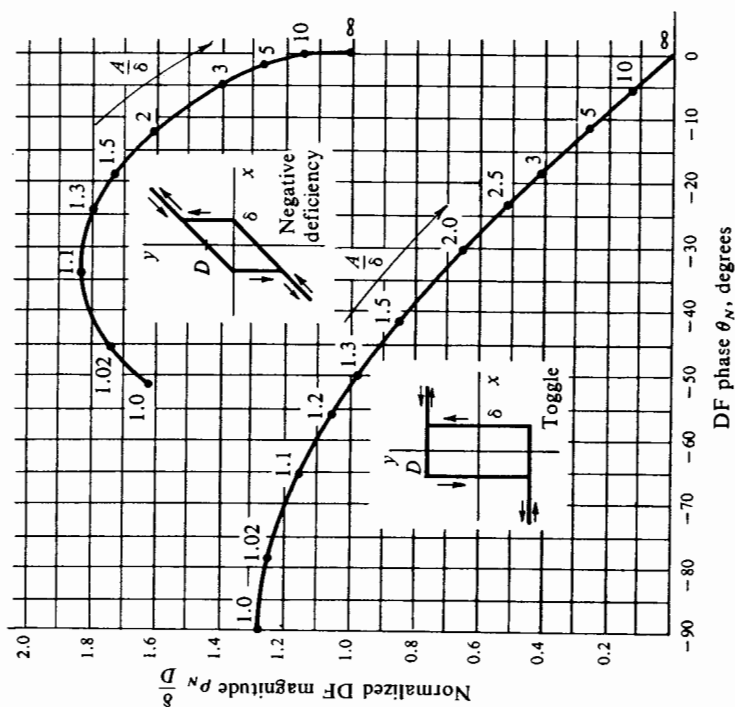


Figure B.6 DF magnitude vs. phase for toggle and negative deficiency characteristics.

APPENDIX C

TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIDFs)

The DIDF sinusoidal gain is given by (cf. Sec. 6.1)

$$N_A(A, B, \omega) = n_p(A, B, \omega) + j n_q(A, B, \omega) = \frac{j}{\pi A} \int_0^{2\pi} y(B + A \sin \psi, A \omega \cos \psi) e^{-j\psi} d\psi$$

and the corresponding dc gain is given by (cf. Sec. 6.1)

$$N_B(A, B, \omega) = \frac{1}{2\pi B} \int_0^{2\pi} y(B + A \sin \psi, A \omega \cos \psi) d\psi$$

In this table we employ the "saturation function" (cf. Sec. 2.3) denoted by

$$\begin{aligned} f(\gamma) &= -1 & \gamma < -1 \\ &= \frac{2}{\pi} (\sin^{-1} \gamma + \gamma \sqrt{1 - \gamma^2}) & |\gamma| \leq 1 \\ &= 1 & \gamma > 1 \end{aligned}$$

and the associated function (cf. Sec. 6.2)

$$\begin{aligned} g(\gamma) &= \frac{2}{\pi} (\gamma \sin^{-1} \gamma + \sqrt{1 - \gamma^2}) & |\gamma| \leq 1 \\ &= |\gamma| & |\gamma| > 1 \end{aligned}$$

These functions are plotted in Fig. C.1. Two additional functions of considerable use are

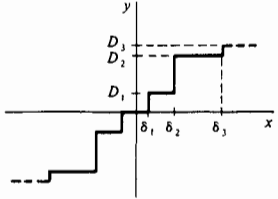
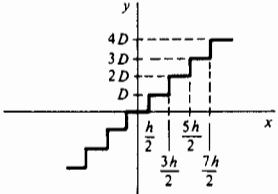
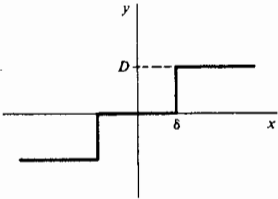
$$\begin{aligned} p(\gamma) &= -\frac{1}{2} & \gamma < -1 \\ &= \frac{1}{\pi} \sin^{-1} \gamma & |\gamma| \leq 1 \\ &= \frac{1}{2} & \gamma > 1 \end{aligned}$$

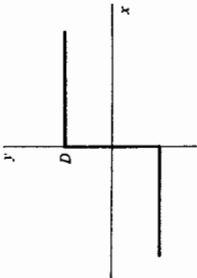
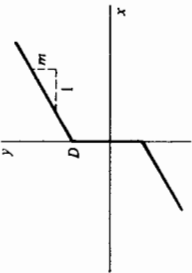
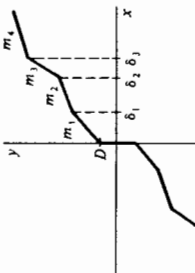
and

$$\begin{aligned} q(\gamma) &= \frac{2}{\pi} \sqrt{1 - \gamma^2} & |\gamma| \leq 1 \\ &= 0 & |\gamma| > 1 \end{aligned}$$

These functions are plotted in Fig. C.2.

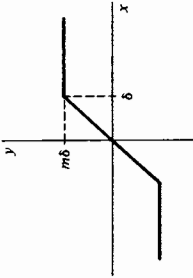
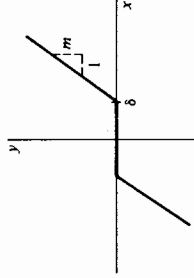
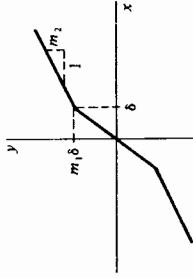
TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIDFs) (Continued)

Nonlinearity	Comments	$N_B(A, B, \omega)$, $n_p(A, B, \omega)$, and $n_q(A, B, \omega)$
 1. General odd quantizer	$\delta_{n+1} > A + B > \delta_n$ $D_0 = 0$	$N_B = \frac{1}{B} \sum_{i=1}^n (D_i - D_{i-1}) \left[p\left(\frac{\delta_i + B}{A}\right) - p\left(\frac{\delta_i - B}{A}\right) \right]$ $n_p = \frac{1}{A} \sum_{i=1}^n (D_i - D_{i-1}) \left[q\left(\frac{\delta_i + B}{A}\right) + q\left(\frac{\delta_i - B}{A}\right) \right]$ $n_q = 0$
 2. Uniform quantizer or granularity	$\frac{2n+1}{2} h > A + B $ $> \frac{2n-1}{2} h$	$N_B = \frac{D}{B} \sum_{i=1}^n \left[p\left(\frac{(2i-1)h + B}{A}\right) - p\left(\frac{(2i-1)h - B}{A}\right) \right]$ $n_p = \frac{D}{A} \sum_{i=1}^n \left[q\left(\frac{(2i-1)h + B}{A}\right) + q\left(\frac{(2i-1)h - B}{A}\right) \right]$ $n_q = 0$
 3. Relay with dead zone		$N_B = \frac{D}{B} \left[p\left(\frac{\delta + B}{A}\right) - p\left(\frac{\delta - B}{A}\right) \right]$ $n_p = \frac{D}{A} \left[q\left(\frac{\delta + B}{A}\right) + q\left(\frac{\delta - B}{A}\right) \right]$ $n_q = 0$

	$A > B $	See Sec. 6.2	$N_B = \frac{2D}{\pi B} \sin^{-1} \frac{B}{A}$ $n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{B}{A}\right)^2}$ $n_q = 0$
	$A > B $		$N_B = \frac{2D}{\pi B} \sin^{-1} \frac{B}{A} + m$ $n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{B}{A}\right)^2} + m$ $n_q = 0$
	$\delta_{n+1} > A + B > \delta_n$ and $A > B $		$N_B = \frac{2D}{\pi B} \sin^{-1} \frac{B}{A} + \frac{A}{2B} \sum_{i=1}^n (m_i - m_{i+1}) \left[g\left(\frac{\delta_i + B}{A}\right) - g\left(\frac{\delta_i - B}{A}\right) \right] + m_{n+1}$ $n_p = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{B}{A}\right)^2} + \frac{1}{2} \sum_{i=1}^n (m_i - m_{i+1}) \left[f\left(\frac{\delta_i + B}{A}\right) + f\left(\frac{\delta_i - B}{A}\right) \right] + m_{n+1}$ $n_q = 0$

6. General piecewise-linear odd memoryless nonlinearity

TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIDFs) (Continued)

Nonlinearity	Comments	$N_B(A, B, \omega)$, $n_s(A, B, \omega)$, and $n_q(A, B, \omega)$
	7. Saturation or limiter	$N_B = \frac{mA}{2B} \left[g\left(\frac{\delta+B}{A}\right) - g\left(\frac{\delta-B}{A}\right) \right]$ $n_p = \frac{m}{2} \left[f\left(\frac{\delta+B}{A}\right) + f\left(\frac{\delta-B}{A}\right) \right]$ $n_q = 0$
	8. Dead zone or threshold	$N_B = m \left\{ 1 - \frac{A}{2B} \left[g\left(\frac{\delta+B}{A}\right) - g\left(\frac{\delta-B}{A}\right) \right] \right\}$ $n_p = m \left\{ 1 - \frac{1}{2} \left[f\left(\frac{\delta+B}{A}\right) + f\left(\frac{\delta-B}{A}\right) \right] \right\}$ $n_q = 0$
	9. Gain-changing nonlinearity	$N_B = \frac{A}{2B} (m_1 - m_2) \left[g\left(\frac{\delta+B}{A}\right) - g\left(\frac{\delta-B}{A}\right) \right] + m_2$ $n_p = \frac{1}{2} (m_1 - m_2) \left[f\left(\frac{\delta+B}{A}\right) + f\left(\frac{\delta-B}{A}\right) \right] + m_2$ $n_q = 0$

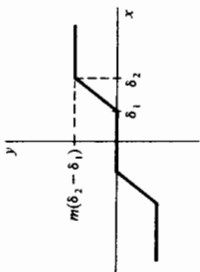
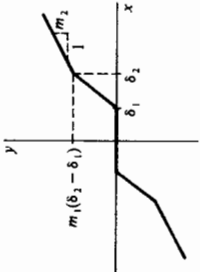
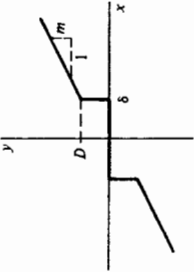
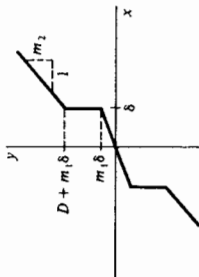
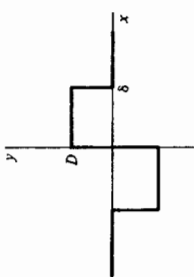
 10. Limiter with dead zone		$N_B = \frac{mA}{2B} \left[g\left(\frac{\delta_2 + B}{A}\right) - g\left(\frac{\delta_2 - B}{A}\right) - g\left(\frac{\delta_1 + B}{A}\right) + g\left(\frac{\delta_1 - B}{A}\right) \right]$ $n_p = \frac{m}{2} \left[f\left(\frac{\delta_2 + B}{A}\right) + f\left(\frac{\delta_2 - B}{A}\right) - f\left(\frac{\delta_1 + B}{A}\right) - f\left(\frac{\delta_1 - B}{A}\right) \right]$ $n_q = 0$
 11. Gain-changing nonlinearity with dead zone		$N_B = \frac{A}{2B} \left\{ (m_1 - m_2) \left[g\left(\frac{\delta_2 + B}{A}\right) - g\left(\frac{\delta_2 - B}{A}\right) \right] - m_1 \left[g\left(\frac{\delta_1 + B}{A}\right) - g\left(\frac{\delta_1 - B}{A}\right) \right] \right\} + m_2$ $n_p = \frac{m_1 - m_2}{2} \left[f\left(\frac{\delta_2 + B}{A}\right) + f\left(\frac{\delta_2 - B}{A}\right) \right] - \frac{m_1}{2} \left[f\left(\frac{\delta_1 + B}{A}\right) + f\left(\frac{\delta_1 - B}{A}\right) \right] + m_2$ $n_q = 0$
 12.		$N_B = m \left\{ 1 - \frac{A}{2B} \left[g\left(\frac{\delta + B}{A}\right) - g\left(\frac{\delta - B}{A}\right) \right] \right\} + \frac{D}{B} \left[p\left(\frac{\delta + B}{A}\right) - p\left(\frac{\delta - B}{A}\right) \right]$ $n_p = m \left\{ 1 - \frac{1}{2} \left[f\left(\frac{\delta + B}{A}\right) + f\left(\frac{\delta - B}{A}\right) \right] \right\} + \frac{D}{A} \left[q\left(\frac{\delta + B}{A}\right) + q\left(\frac{\delta - B}{A}\right) \right]$ $n_q = 0$

TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIFs) (Continued)

Nonlinearity	Comments	$N_B(A, B, \omega)$, $n_p(A, B, \omega)$, and $n_q(A, B, \omega)$
		$N_B = \frac{A}{2B} (m_1 - m_2) \left[g\left(\frac{\delta + B}{A}\right) - g\left(\frac{\delta - B}{A}\right) \right] + m_2 + \frac{D}{B} \left[p\left(\frac{\delta + B}{A}\right) - p\left(\frac{\delta - B}{A}\right) \right]$ $n_p = \frac{1}{2} (m_1 - m_2) \left[f\left(\frac{\delta + B}{A}\right) + f\left(\frac{\delta - B}{A}\right) \right] + m_2 + \frac{D}{A} \left[q\left(\frac{\delta + B}{A}\right) + q\left(\frac{\delta - B}{A}\right) \right]$ $n_q = 0$
	$A > B $	$N_B = \frac{D}{B} \left[\frac{2}{\pi} \sin^{-1} \frac{B}{A} - p\left(\frac{\delta + B}{A}\right) + p\left(\frac{\delta - B}{A}\right) \right]$ $n_p = \frac{D}{A} \left[\frac{4}{\pi} \sqrt{1 - \left(\frac{B}{A}\right)^2} - q\left(\frac{\delta + B}{A}\right) - q\left(\frac{\delta - B}{A}\right) \right]$ $n_q = 0$

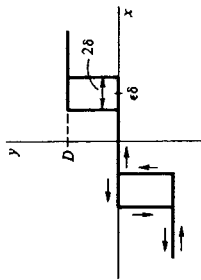
13.

14.

$y = c$		$N_B = \frac{c}{B}$ $n_p = 0$ $n_q = 0$
15.		
$y = x$		$N_B = 1$ $n_p = 1$ $n_q = 0$
16. Linear gain		
$y = x x $		$N_B = \frac{A^2}{\pi B} \left\{ \left[1 + 2 \left(\frac{B}{A} \right)^2 \right] \sin^{-1} \frac{B}{A} + 3 \frac{B}{A} \sqrt{1 - \left(\frac{B}{A} \right)^2} \right\}$ $n_p = \frac{8}{3\pi} A \left\{ \left[1 + \frac{1}{2} \left(\frac{B}{A} \right)^2 \right] \sqrt{1 - \left(\frac{B}{A} \right)^2} + \frac{3}{2} \frac{B}{A} \sin^{-1} \frac{B}{A} \right\}$ $n_q = 0$
17. Odd square law		
$y = x^3$		$N_B = \frac{3}{2} A^3 + B^3$ $n_p = \frac{3}{2} A^3 + 3B^2$ $n_q = 0$
18. Cubic characteristic	See Sec. 6.2	
$y = x^5$		$N_B = \frac{15}{8} A^4 + 5A^2 B^2 + B^4$ $n_p = \frac{5}{8} A^4 + \frac{15}{8} A^2 B^2 + 5B^4$ $n_q = 0$
20. Quintic characteristic		
$y = x^7$		$N_B = \frac{35}{8} A^6 + \frac{15}{8} A^4 B^2 + \frac{35}{8} A^2 B^4 + B^6$ $n_p = \frac{35}{8} A^6 + \frac{15}{8} A^4 B^2 + \frac{15}{8} A^2 B^4 + 7B^6$ $n_q = 0$
22.		

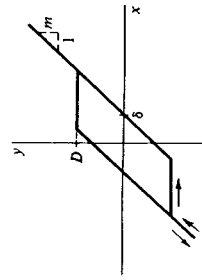
TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIDFs) (Continued)

Nonlinearity	Comments	$N_B(A, B, \omega)$, $n_p(A, B, \omega)$, and $n_q(A, B, \omega)$
$y = x^n$	$n = 3, 5, 7, \dots$	$N_B = \frac{1}{\sqrt{\pi}} \sum_{k(\text{odd})=1}^n \frac{n!}{(n-k)! k!} A^{n-k} B^{k-1} \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{\Gamma\left(\frac{n-k+2}{2}\right)}$ $n_p = \frac{2}{\sqrt{\pi}} \sum_{k(\text{even})=0}^{n-1} \frac{n!}{(n-k)! k!} A^{n-k-1} B^k \frac{\Gamma\left(\frac{n-k+2}{2}\right)}{\Gamma\left(\frac{n-k+3}{2}\right)}$ $n_q = 0$
24.	See Sec. 6.2	
$y = M \sin mx$		$N_B = \frac{M}{B} J_0(mA) \sin mB$ $n_p = \frac{2M}{A} J_1(mA) \cos mB$ $n_q = 0$
29. Harmonic nonlinearity		
$y = M \sinh mx$	I_0 and I_1 are modified Bessel functions of orders 0 and 1, respectively.	$N_B = \frac{M}{B} I_0(mA) \sinh mB$ $n_p = \frac{2M}{A} I_1(mA) \cosh mB$ $n_q = 0$
30.		



37. (positive) Hysteresis

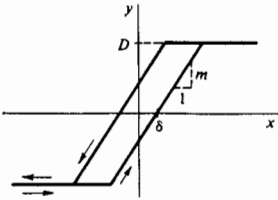
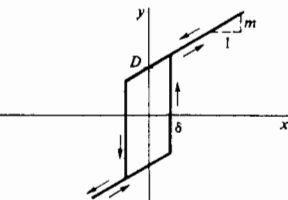
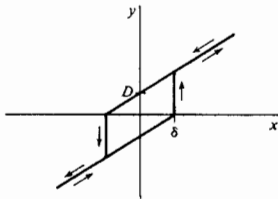
$$\begin{aligned}
 N_B &= 0 \\
 n_a &= 0 \\
 n_c &= 0 \\
 N_B &= \frac{D}{2\pi B} \left[\sin^{-1} \frac{(\epsilon + 1)\delta + B}{A} - \sin^{-1} \frac{(\epsilon + 1)\delta - B}{A} + \sin^{-1} \frac{(\epsilon - 1)\delta + B}{A} \right. \\
 &\quad \left. - \sin^{-1} \frac{(\epsilon - 1)\delta - B}{A} \right] \\
 n_p &= \frac{D}{\pi A} \left\{ \sqrt{1 - \left[\frac{(\epsilon + 1)\delta + B}{A} \right]^2} + \sqrt{1 - \left[\frac{(\epsilon + 1)\delta - B}{A} \right]^2} \right. \\
 &\quad \left. + \sqrt{1 - \left[\frac{(\epsilon - 1)\delta + B}{A} \right]^2} + \sqrt{1 - \left[\frac{(\epsilon - 1)\delta - B}{A} \right]^2} \right\} \\
 n_c &= -\frac{4D\delta}{\pi A^3}
 \end{aligned}$$



40.

$$\begin{aligned}
 N_B &= m \left\{ 1 + \frac{A}{4B} \left[-g \left(\frac{\frac{D}{m} + \delta + B}{A} \right) + g \left(\frac{\frac{D}{m} + \delta - B}{A} \right) + g \left(\frac{\frac{D}{m} - \delta + B}{A} \right) \right. \right. \\
 &\quad \left. \left. - g \left(\frac{\frac{D}{m} - \delta - B}{A} \right) \right] \right\} \\
 n_p &= m \left\{ 1 - \frac{1}{4} \left[f \left(\frac{\frac{D}{m} + \delta + B}{A} \right) + f \left(\frac{\frac{D}{m} + \delta - B}{A} \right) - f \left(\frac{\frac{D}{m} - \delta + B}{A} \right) \right. \right. \\
 &\quad \left. \left. - f \left(\frac{\frac{D}{m} - \delta - B}{A} \right) \right] \right\} \\
 n_c &= -\frac{4D\delta}{\pi A^3}
 \end{aligned}$$

TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIDFs) (Continued)

Nonlinearity	Comments	$N_B(A, B, \omega)$, $n_p(A, B, \omega)$, and $n_q(A, B, \omega)$
42. 	$A - B > \frac{D}{m} + \delta$	$N_B = \frac{m}{4} \left[g\left(\frac{\frac{D}{m} + \delta + B}{A}\right) - g\left(\frac{\frac{D}{m} + \delta - B}{A}\right) + g\left(\frac{\frac{D}{m} - \delta + B}{A}\right) - g\left(\frac{\frac{D}{m} - \delta - B}{A}\right) \right]$ $n_p = \frac{m}{4} \left[f\left(\frac{\frac{D}{m} + \delta + B}{A}\right) + f\left(\frac{\frac{D}{m} + \delta - B}{A}\right) + f\left(\frac{\frac{D}{m} - \delta + B}{A}\right) + f\left(\frac{\frac{D}{m} - \delta - B}{A}\right) \right]$ $n_q = -\frac{4D\delta}{\pi A^2}$
43. 	$A - B > \delta$	$N_B = \frac{D}{\pi B} \left(\sin^{-1} \frac{\delta + B}{A} - \sin^{-1} \frac{\delta - B}{A} \right) + m$ $n_p = \frac{2D}{\pi A} \left[\sqrt{1 - \left(\frac{\delta + B}{A}\right)^2} + \sqrt{1 - \left(\frac{\delta - B}{A}\right)^2} \right] + m$ $n_q = -\frac{4D\delta}{\pi A^2}$
44. Negative deficiency 	$A - B > \delta$	$N_B = \frac{D}{\pi B} \left(\sin^{-1} \frac{\delta + B}{A} - \sin^{-1} \frac{\delta - B}{A} \right) + \frac{D}{\delta}$ $n_p = \frac{2D}{\pi A} \left[\sqrt{1 - \left(\frac{\delta + B}{A}\right)^2} + \sqrt{1 - \left(\frac{\delta - B}{A}\right)^2} \right] + \frac{D}{\delta}$ $n_q = -\frac{4D\delta}{\pi A^2}$

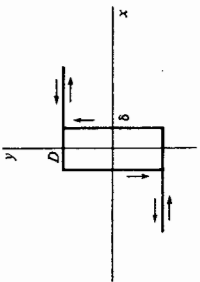
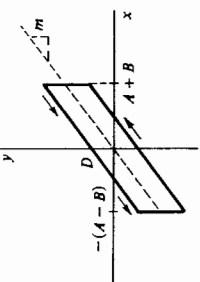
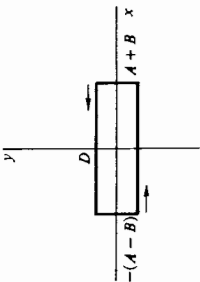
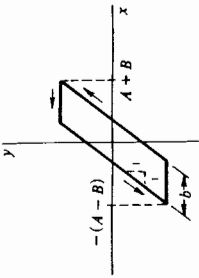
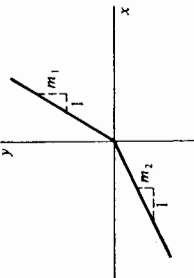
 45. Rectangular hysteresis or toggle	$A - B > \delta$ See Sec. 6.2	$N_B = \frac{D}{\pi B} \left(\sin^{-1} \frac{\delta + B}{A} - \sin^{-1} \frac{\delta - B}{A} \right)$ $n_p = \frac{2D}{\pi A} \left[\sqrt{1 - \left(\frac{\delta + B}{A} \right)^2} + \sqrt{1 - \left(\frac{\delta - B}{A} \right)^2} \right]$ $n_q = -\frac{4D\delta}{\pi A^2}$
 46.		$N_B = m$ $n_p = m$ $n_q = -\frac{4D}{\pi A}$
 47.		$N_B = 0$ $n_p = 0$ $n_q = -\frac{4D}{\pi A}$

TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIDFs) (Continued)

Nonlinearity	Comments	$N_B(A,B,\omega)$, $n_p(A,B,\omega)$, and $n_q(A,B,\omega)$
 48. Friction-controlled backlash		$N_B = 1$ $n_p = \frac{1}{2} \left[1 + f \left(1 - \frac{b}{A} \right) \right]$ $n_q = -\frac{1}{\pi} \left[2 \frac{b}{A} - \left(\frac{b}{A} \right)^2 \right]$
 51.	$B > A$ $-A < B < A$ $B < -A$ See Sec. 6.2	$N_B = m_1$ $n_p = m_1$ $n_q = 0$ $N_B = \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{2} \frac{A}{B} g \left(\frac{B}{A} \right)$ $n_p = \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{2} f \left(\frac{B}{A} \right)$ $n_q = 0$ $N_B = m_2$ $n_p = m_2$ $n_q = 0$

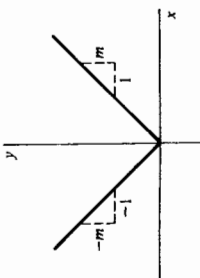
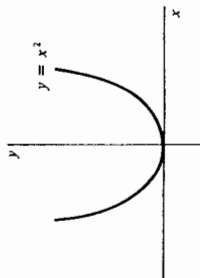
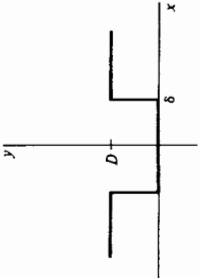
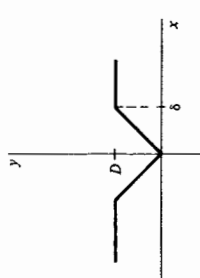
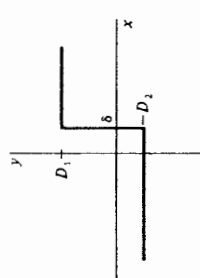
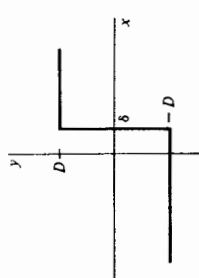
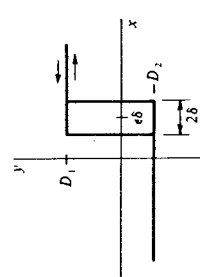
 52. Absolute value	$B > A$ $-A < B < A$ $B < -A$	$N_B = m$ $n_p = m$ $n_q = 0$ $N_B = m \frac{A}{B} g \left(\frac{B}{A} \right)$ $n_p = m f \left(\frac{B}{A} \right)$ $n_q = 0$ $N_B = -m$ $n_p = -m$ $n_q = 0$
 53. Square law		$N_B = \frac{1}{B} [B^2 + \frac{1}{2} A^2]$ $n_p = 2B$ $n_q = 0$
 54.	$A + B < \delta$ $A - B > \delta$	$N_B = 0$ $n_p = 0$ $n_q = 0$ $N_B = \frac{D}{\pi B} \left(\pi - \sin^{-1} \frac{\delta - B}{A} - \sin^{-1} \frac{\delta + B}{A} \right)$ $n_p = \frac{2D}{\pi A} \left[\sqrt{1 - \left(\frac{\delta - B}{A} \right)^2} - \sqrt{1 - \left(\frac{\delta + B}{A} \right)^2} \right]$ $n_q = 0$

TABLE OF DUAL-INPUT DESCRIBING FUNCTIONS (DIDFs) (Continued)

Nonlinearity	Comments	$N_B(A, B, \omega)$, $n_B(A, B, \omega)$, and $n_Q(A, B, \omega)$
 55.	$A + B < \delta$ $A - B > \delta$	N_B (see absolute-value case) n_B (see absolute-value case) $n_Q = 0$ $N_B = \frac{D}{2\delta} \left[-g\left(\frac{\delta - B}{A}\right) + 2g\left(\frac{B}{A}\right) - g\left(\frac{\delta + B}{A}\right) \right] + \frac{D}{B}$ $n_B = \frac{D}{\pi\delta} \left[f\left(\frac{\delta - B}{A}\right) + 2f\left(\frac{B}{A}\right) - f\left(\frac{\delta + B}{A}\right) \right]$ $n_Q = 0$
 56. Input- and output-biased ideal relay	$A - B > \delta$	$N_B = \frac{D_1 - D_2}{2B} - \frac{D_1 + D_2}{\pi B} \sin^{-1} \frac{\delta - B}{A}$ $n_B = \frac{2(D_1 + D_2)}{\pi A} \sqrt{1 - \left(\frac{\delta - B}{A}\right)^2}$ $n_Q = 0$
 57. Input-biased ideal relay	$A - B > \delta$	$N_B = \frac{2D}{\pi B} \sin^{-1} \frac{B - \delta}{A}$ $n_B = \frac{4D}{\pi A} \sqrt{1 - \left(\frac{\delta - B}{A}\right)^2}$ $n_Q = 0$

	$A - B > \delta(\epsilon + 1)$	$N_B = \frac{1}{2\pi B} \left\{ (D_1 + D_2) \left[\pi - \sin^{-1} \frac{\delta(\epsilon - 1) - B}{A} - \sin^{-1} \frac{\delta(\epsilon + 1) - B}{A} \right] - 2\pi D_2 \right\}$ $n_p = \frac{D_1 + D_2}{\pi A} \left\{ \sqrt{1 - \left[\frac{\delta(\epsilon + 1) - B}{A} \right]^2} + \sqrt{1 - \left[\frac{\delta(\epsilon - 1) - B}{A} \right]^2} \right\}$ $n_q = - \frac{2(D_1 + D_2)\delta}{\pi A^2}$

58. Input- and output-biased rectangular hysteresis

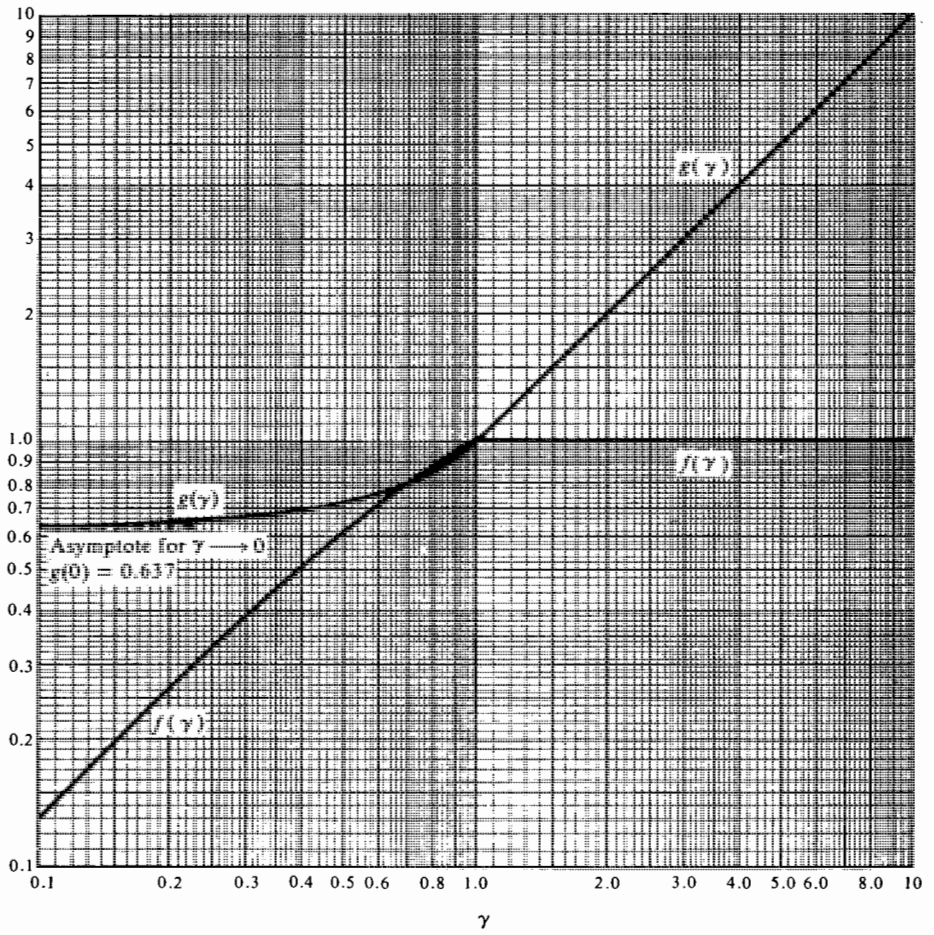


Figure C.1 Graphs of $f(\gamma)$ and $g(\gamma)$.

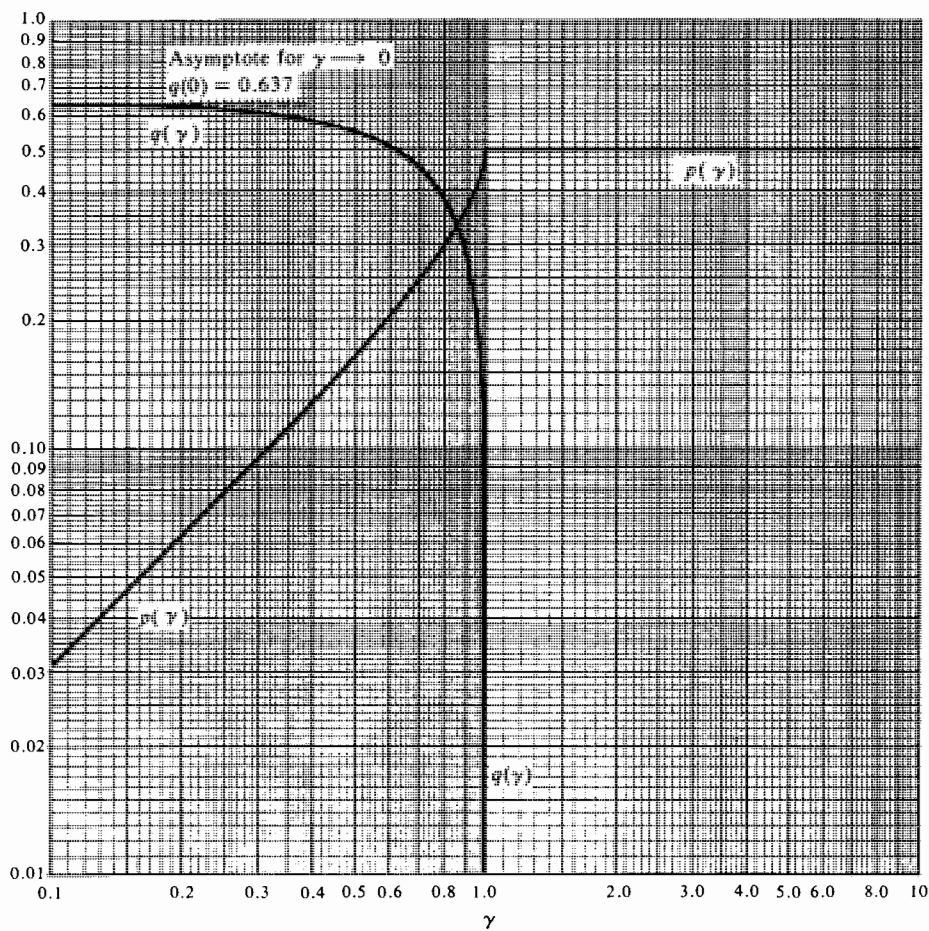


Figure C.2 Graphs of $p(\gamma)$ and $q(\gamma)$.

APPENDIX D

TABLE OF TWO-SINUSOID-INPUT DESCRIBING FUNCTIONS (TSIDFs)

The TSIDF can be represented by either of the following integral expressions (cf. Sec. 5.1)

$$N_B(A, B) = \frac{1}{2\pi^2 B} \int_{-\pi}^{\pi} d\psi_2 \sin \psi_2 \int_{-\pi}^{\pi} d\psi_1 y(A \sin \psi_1 + B \sin \psi_2)$$

or

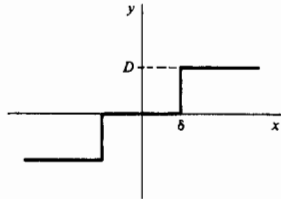
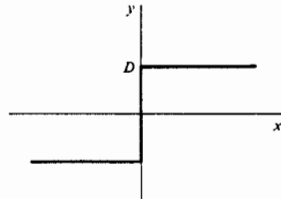
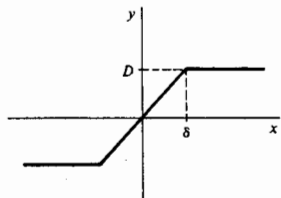
$$N_B(A, B) = \frac{j}{\pi B} \int_{-\infty}^{\infty} du J_0(Au) J_1(Bu) \int_{-\infty}^{\infty} dx y(x) e^{-jux}$$

where J_0 and J_1 are the Bessel functions of orders 0 and 1, respectively. Use of this table is facilitated by application of the relationship

$$N_A(A, B) = N_B(B, A)$$

All entries are for the case of nonharmonically related sinusoids, corresponding to the above integral formulations.

TABLE OF TWO-SINUSOID-INPUT DESCRIBING FUNCTIONS (TSIDFs) (Continued)

Nonlinearity	Comments	$N_B(A, B)$
 3. Relay with dead zone	${}_2F_1$ is the gaussian hypergeometric function. See Gibson and Sridhar, Ref. 10 of Chap. 5. See Fig. D.1	$\frac{4D}{\pi} \sum_{m=0}^{\infty} \frac{m \cos\left(2m \sin^{-1} \frac{\delta}{B}\right)}{\sqrt{1 - \left(\frac{\delta}{B}\right)^2}} \sum_{n=0}^{\infty} \frac{(-1)^n \Gamma(m+n)}{n! (m-n)! \Gamma(n+1)} \left(\frac{A}{B}\right)^{2n} {}_2F_1\left[-n, -n; 2; \left(\frac{B}{A}\right)^2\right]$ or, in integral form, $\frac{2D}{\pi B} \int_{-\infty}^{\infty} \frac{\cos \delta u}{u} J_0(Au) J_1(Bu) du$
 4. Ideal relay	$K(k)$ and $E(k)$ are the elliptic integrals of first and second kind, respectively. See Fig. D.2 and Sec. 5.1	$\frac{8D}{\pi^2 B} \frac{A}{B} \left\{ E\left(\frac{B}{A}\right) - \left[1 - \left(\frac{B}{A}\right)^2 \right] K\left(\frac{B}{A}\right) \right\} \quad \text{for } \frac{B}{A} < 1$ $\frac{8D}{\pi^2 B} E\left(\frac{A}{B}\right) \quad \text{for } \frac{B}{A} > 1$
 7. Limiter	See Gibson and Sridhar (op. cit.) See Fig. D.3	$\frac{2D}{\pi \delta} \sum_{m=0}^{\infty} \sin\left(\frac{2D}{\delta} \sin^{-1} \frac{\delta}{B}\right) \sum_{n=0}^{\infty} \frac{(-1)^n \Gamma(m+n)}{n! (m-n)! \Gamma(n+1)} \left(\frac{A}{B}\right)^{2n} {}_2F_1\left[-n, -n; 2; \left(\frac{B}{A}\right)^2\right]$ or, in integral form, $\frac{2D}{\pi \delta B} \int_{-\infty}^{\infty} \frac{\sin \delta u}{u^2} J_0(Au) J_1(Bu) du$

$y = x^3$		$\frac{3}{4}B^2 + \frac{3}{2}A^2$
18. Cubic characteristic	See Fig. D.4 and Sec. 5.1	
$y = M \sin mx$	J_0 and J_1 are the Bessel functions of orders 0 and 1, respectively.	$\frac{2MJ_0(mA)J_1(mB)}{B}$
29. Harmonic nonlinearity	See Fig. D.5	

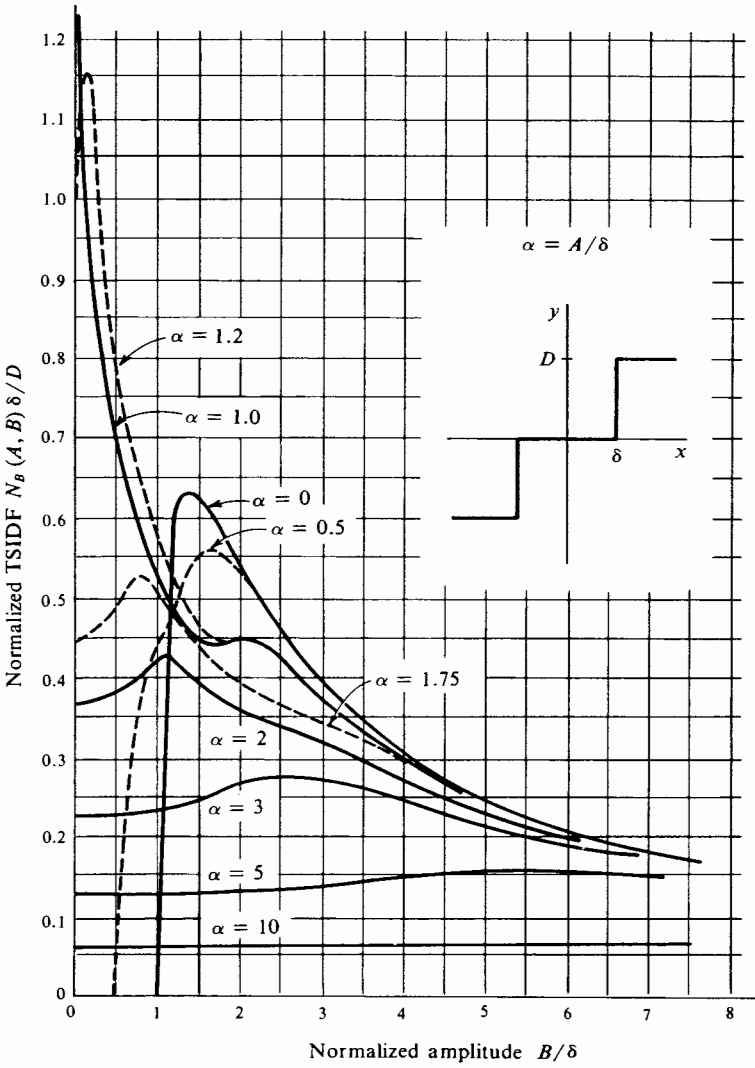


Figure D.1 Normalized relay with dead zone TSIDF. (Gibson and Sridhar, Ref. 10 of Chap. 5.)

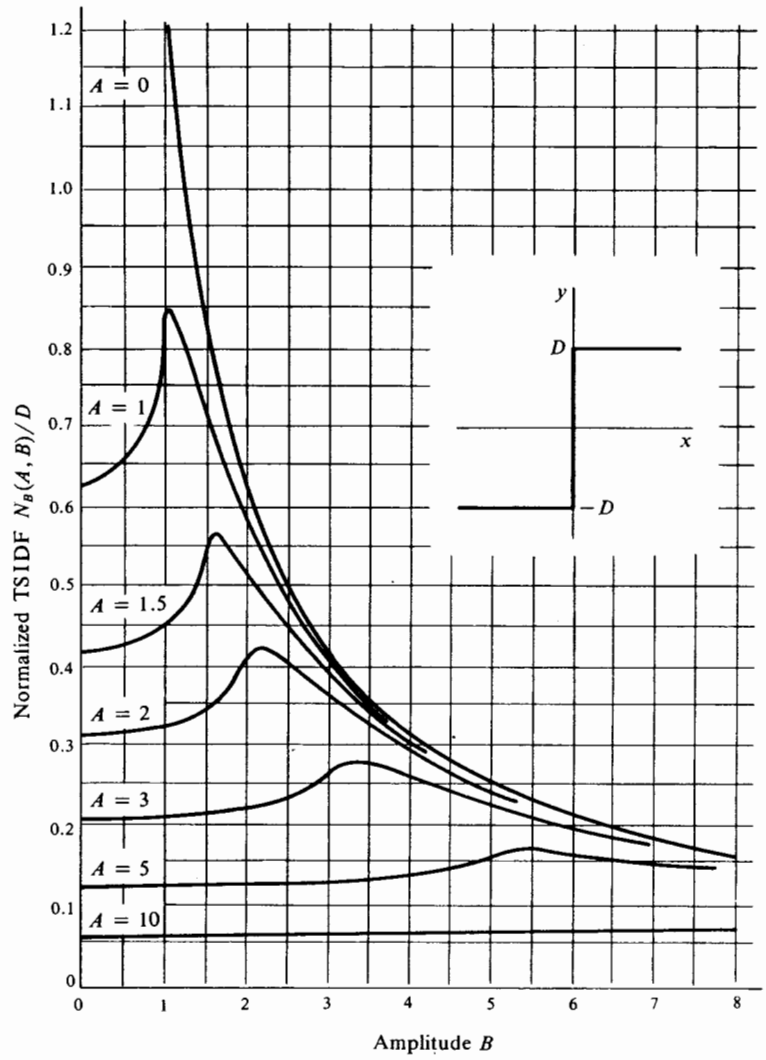


Figure D.2 Normalized ideal-relay TSIDF.

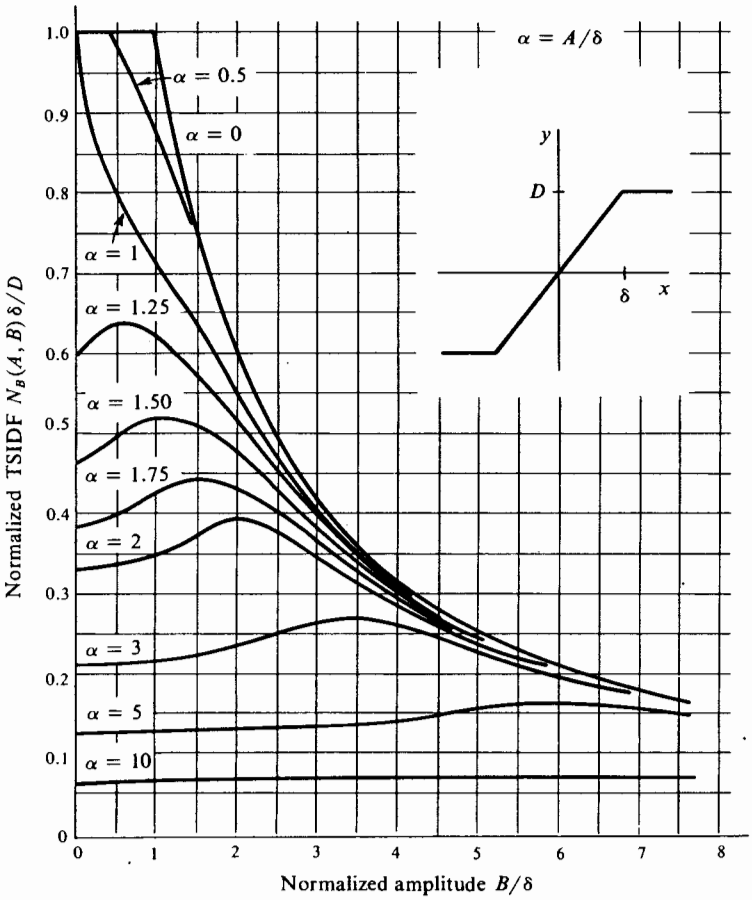


Figure D.3 Normalized limiter TSIDF. (Gibson and Sridhar, Ref. 10 of Chap. 5.)

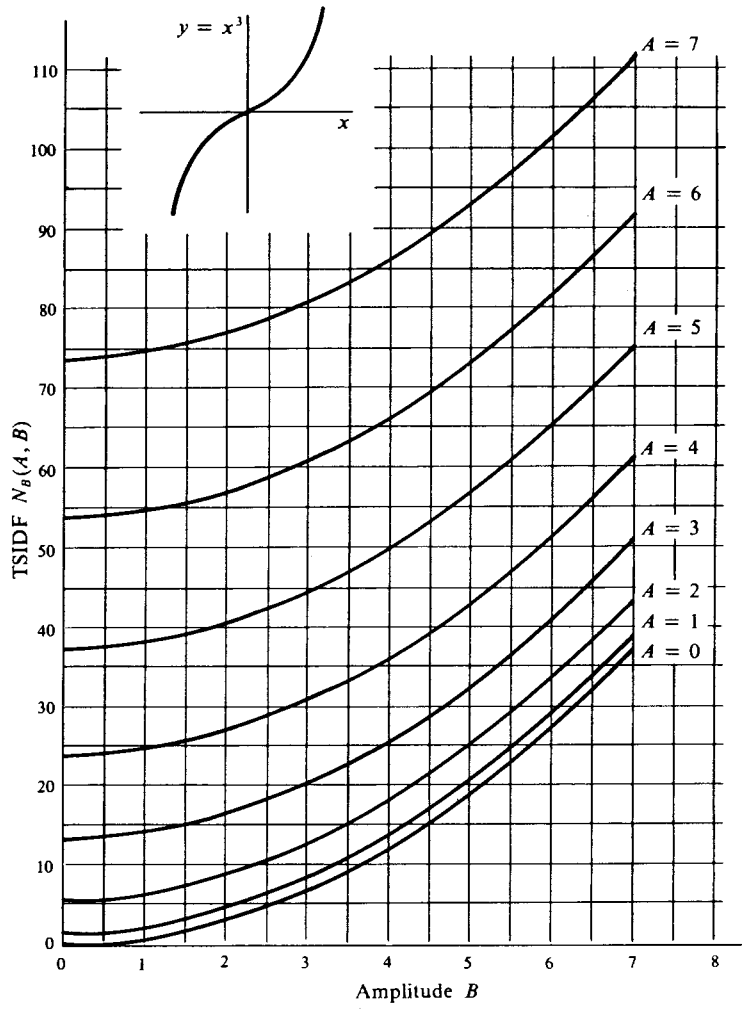


Figure D.4 Cubic characteristic TSIDF.

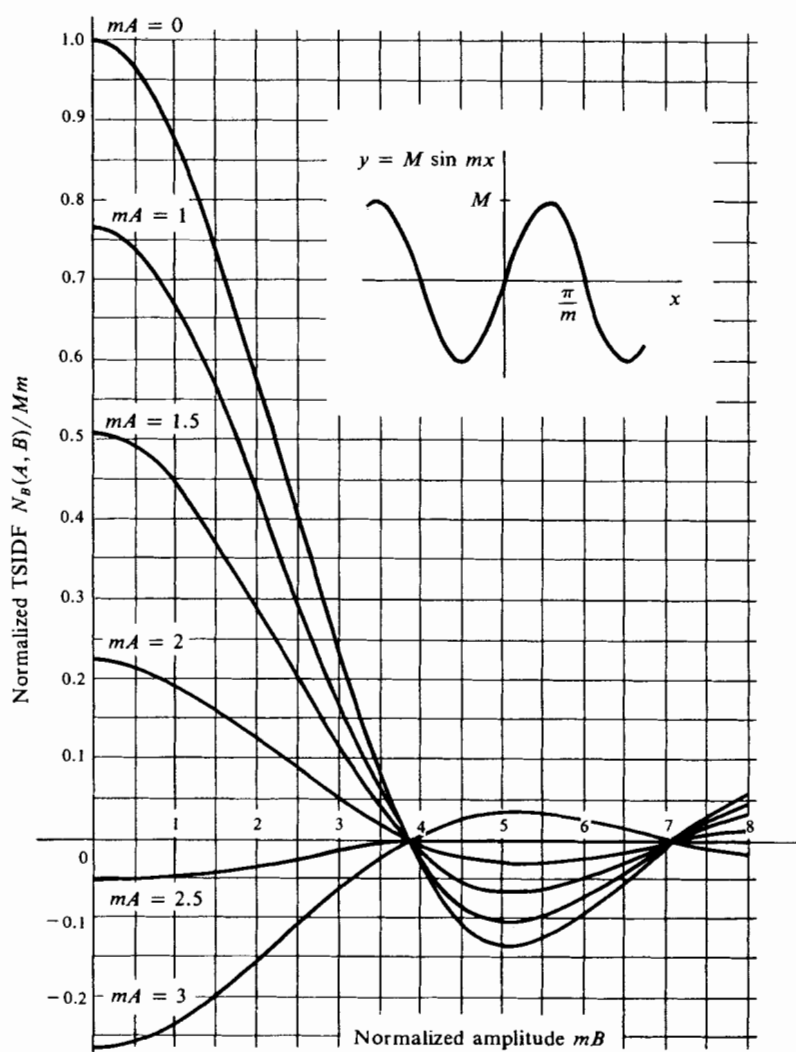


Figure D.5 Normalized harmonic nonlinearity TSIDF.

APPENDIX E

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs)

In this table we employ the probability function (cf. Sec. 7.2) denoted by

$$PF(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

and its integral, the probability integral, denoted by

$$PI(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{v^2}{2}\right) dv$$

These functions are plotted in Fig. E.2-1.

This table is given in three sections:

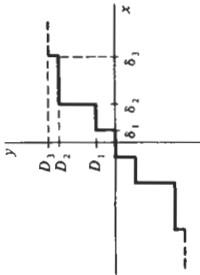
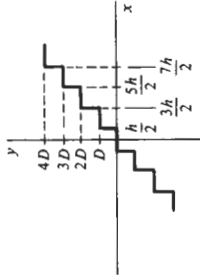
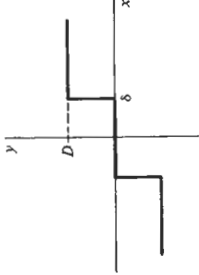
- E.1 Gaussian-input RIDFs
- E.2 Gaussian-plus-bias-input RIDFs
- E.3 Gaussian-plus-bias-plus-sinusoid-input RIDFs

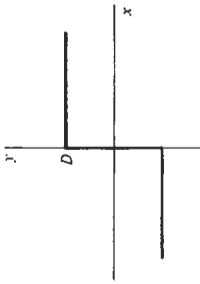
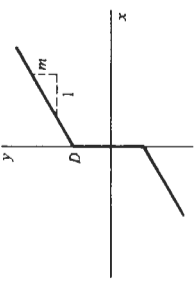
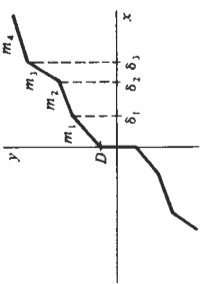
E.1 GAUSSIAN-INPUT RIDFs

$x(t) = r(t)$ an unbiased Gaussian process

$$N_x(\sigma) = \frac{1}{\sqrt{2\pi\sigma^3}} \int_{-\infty}^{\infty} y(r)r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma)$
 1. General odd quantizer	$D_0 = 0$ N is index of last quantizer level	$\frac{2}{\sigma} \sum_{i=1}^N (D_i - D_{i-1}) PF\left(\frac{\delta_i}{\sigma}\right)$
 2. Uniform quantizer	N is index of last quantizer level See Fig. E.1-1	$\frac{2D}{\sigma} \sum_{i=1}^N PF\left(\frac{2i-1}{2} \frac{h}{\sigma}\right)$
 3. Relay with dead zone	See Fig. E.1-1	$2 \frac{D}{\sigma} PF\left(\frac{\delta}{\sigma}\right)$

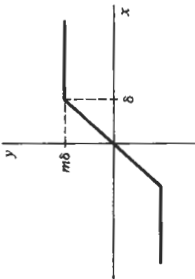
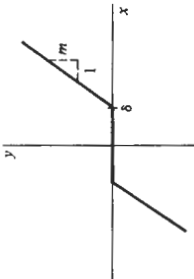
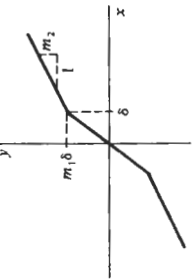
		$\sqrt{\frac{2}{\pi}} \frac{D}{\sigma}$	
		$\sqrt{\frac{2}{\pi}} \frac{D}{\sigma} + m$	
	N is index of last linear segment	$\sqrt{\frac{2}{\pi}} \frac{D}{\sigma} + 2 \sum_{i=1}^{N-1} (m_i - m_{i+1}) PI\left(\frac{\delta_i}{\sigma}\right) + 2m_N - m_1$	

4. Ideal relay

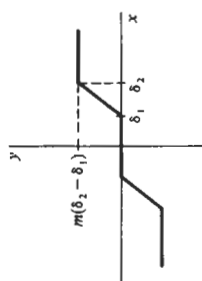
5. Preload

6. General piecewise-linear odd memoryless nonlinearity

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

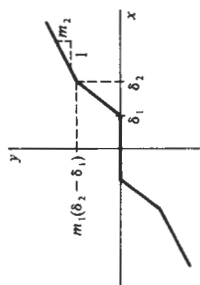
Nonlinearity	Comments	$N_R(\sigma)$
	See Fig. E.1-2	$m \left[2PI\left(\frac{\delta}{\sigma}\right) - 1 \right]$
	See Fig. E.1-2	$2m \left[1 - PI\left(\frac{\delta}{\sigma}\right) \right]$
		$m_1 + 2(m_2 - m_1) \left[1 - PI\left(\frac{\delta}{\sigma}\right) \right]$

9. Gain-changing nonlinearity



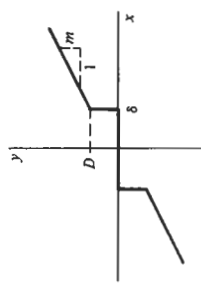
10. Limiter with dead zone

$$2m \left[PI \left(\frac{\delta_2}{\sigma} \right) - PI \left(\frac{\delta_1}{\sigma} \right) \right]$$



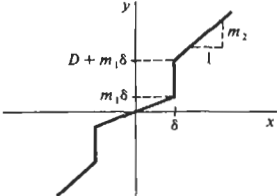
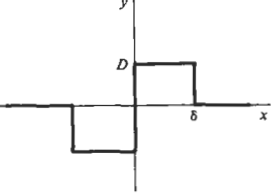
11. Gain-changing nonlinearity with dead zone

$$2(m_1 - m_2) PI \left(\frac{\delta_2}{\sigma} \right) - 2m_1 PI \left(\frac{\delta_1}{\sigma} \right) + 2m_2$$



$$2m \left[1 - PI \left(\frac{\delta}{\sigma} \right) \right] + 2 \frac{D}{\sigma} PF \left(\frac{\delta}{\sigma} \right)$$

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma)$
13. 		$m_1 + 2(m_2 - m_1) \left[1 - PF\left(\frac{\delta}{\sigma}\right) \right] + 2 \frac{D}{\sigma} PF\left(\frac{\delta}{\sigma}\right)$
14. 		$\sqrt{\frac{2}{\pi}} \frac{D}{\sigma} \left[1 - \sqrt{2\pi} PF\left(\frac{\delta}{\sigma}\right) \right]$
15. $y = c$		0
16. Linear gain $y = x$		1

$y = x x $		$\sqrt{\frac{2}{\pi}} 2\sigma$
17. Odd square law	See Fig. E.1-3	
$y = x^2$		$3\sigma^2$
18. Cubic characteristic	See Fig. E.1-3	
$y = x^3 x $		$\sqrt{\frac{2}{\pi}} 8\sigma^3$
19. Odd quartic characteristic	See Fig. E.1-3	
$y = x^4$		$15\sigma^4$
20. Quintic characteristic		
$y = x^5 x $		$\sqrt{\frac{2}{\pi}} 48\sigma^5$
21.		
$y = x^7$		$105\sigma^5$
22.		

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma)$
23. $y = x^7 x $		$\sqrt{\frac{2}{\pi}} 384\sigma^7$
$y = x^n$	$n = 3, 5, 7, \dots$	$n(n-2)(n-4) \cdots (1)\sigma^{n-1}$
24.	See Sec. 7.2	
$y = x^{n-1} x $	$n = 2, 4, 6, \dots$	$\sqrt{\frac{2}{\pi}} n(n-2)(n-4) \cdots (2)\sigma^{n-1}$
25.	See Sec. 7.2	
$y = \sqrt{x} \quad (x \geq 0)$ $= -\sqrt{-x} \quad (x < 0)$		$0.860\sigma^{-1/2}$
26. Odd square root	See Fig. E.1-3	
$y = x^{1/3}$		$0.830\sigma^{-2/3}$
27. Cube root characteristic		
$y = x^b \quad (x \geq 0)$ $= -(-x)^b \quad (x < 0)$	$\Gamma(x)$ is gamma function	$\sqrt{\frac{2}{\pi}} 2^{b/2} \Gamma\left(1 + \frac{b}{2}\right) \sigma^{b-1}$
28.		

$y = M \sin mx$		$\sqrt{2\pi} M m P F(m\sigma) = M m e^{-m^2 \sigma^2/2}$
29. Harmonic Nonlinearity	See Fig. E.1-4	
$y = M \sinh mx$		$M m e^{m^2 \sigma^2/2}$
30.		
$y = 1 - e^{-cx} \quad (x \geq 0)$ $= -(1 - e^{cx}) \quad (x < 0)$		$2c e^{c^2 \sigma^2/2} [1 - P I(c\sigma)]$
31. Exponential saturation		

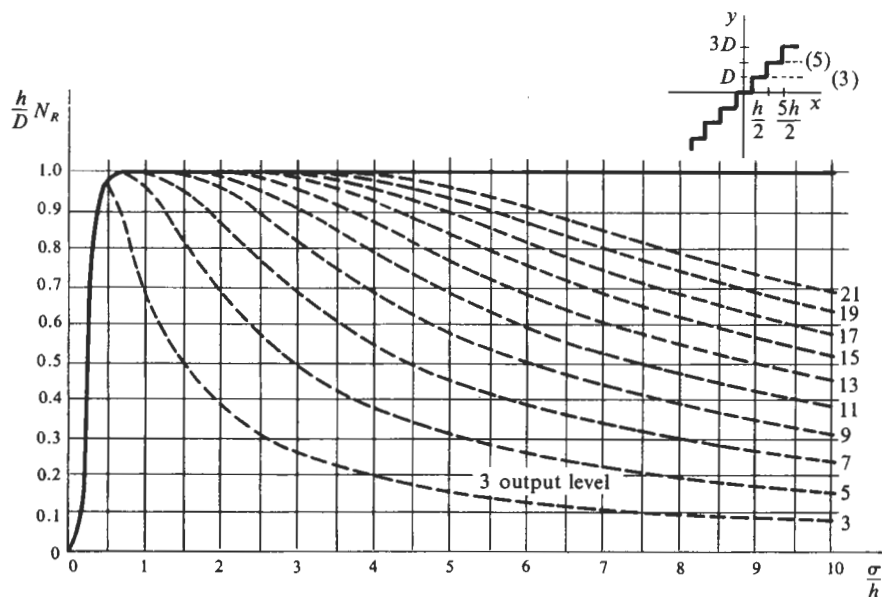


Figure E.1-1 Quantizer RIDF.

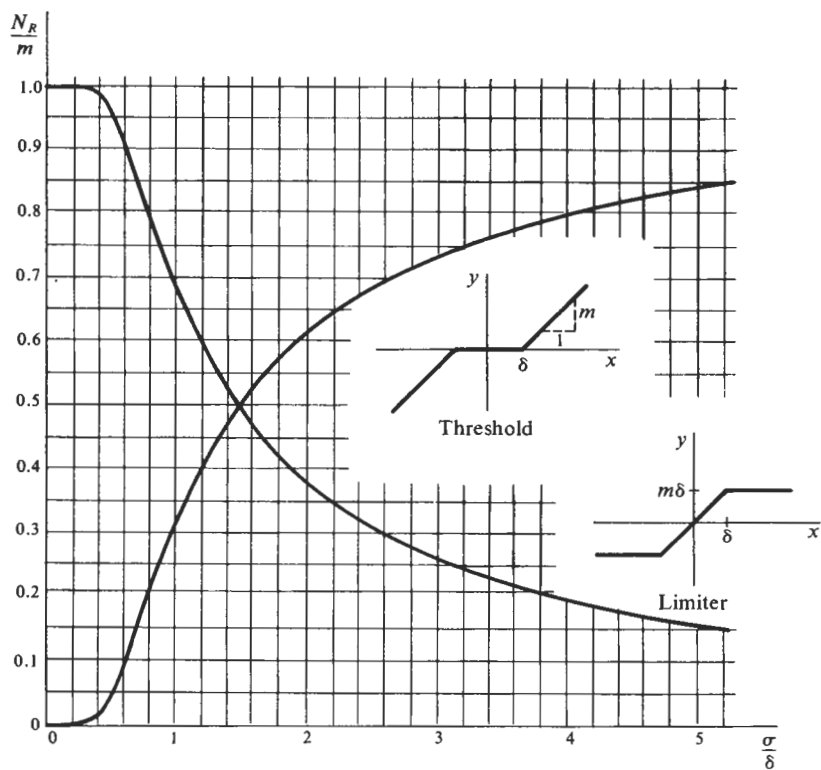


Figure E.1-2 RIDFs for limiter and threshold characteristics.

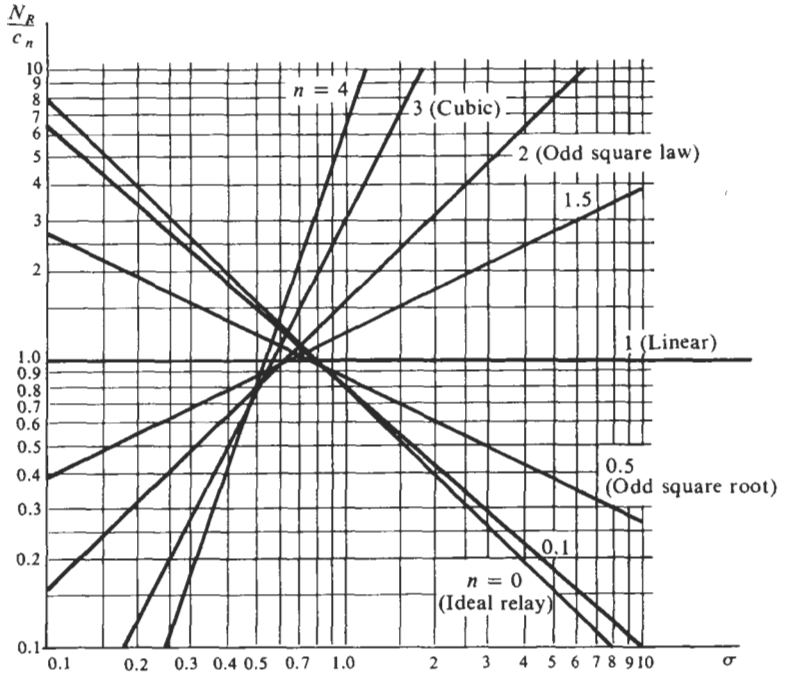


Figure E.1-3 RIDF for the simple polynomial nonlinearity $y = c_n x^n$ (n odd) or $y = c_n x^{n-1} |x|$ (n even).

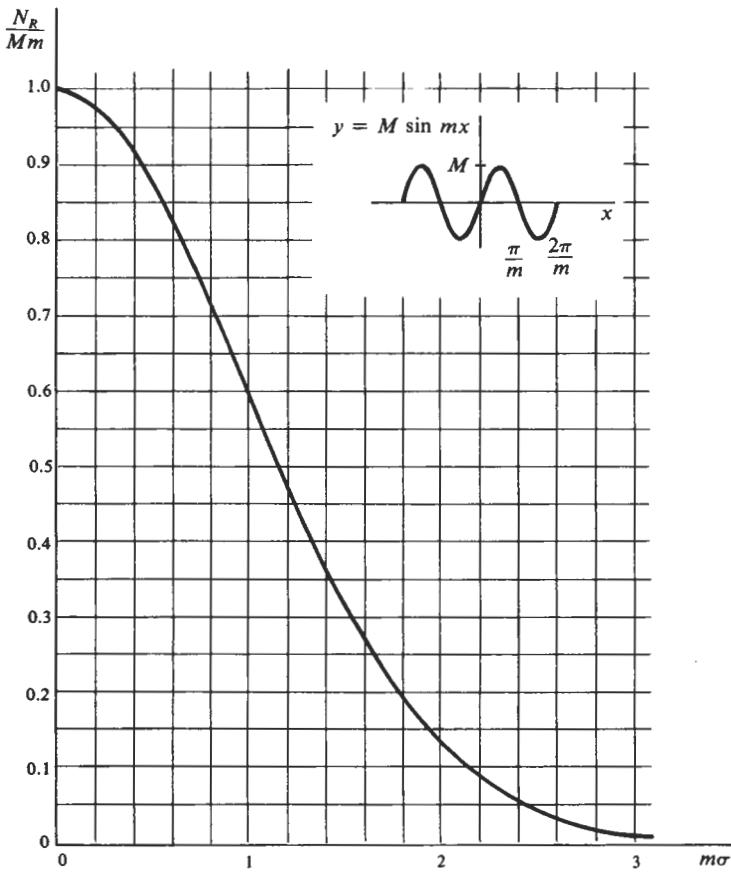


Figure E.1-4 Harmonic nonlinearity RIDF.

E.2 GAUSSIAN-PLUS-BIAS-INPUT RIDFs

$$x(t) = r(t) + B$$

The gain to the gaussian input component is given by:

$$N_B(\sigma, B) = \frac{1}{\sqrt{2\pi}\sigma^3} \int_{-\infty}^{\infty} y(r + B)r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

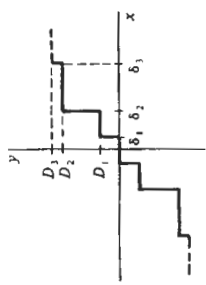
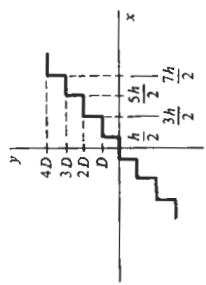
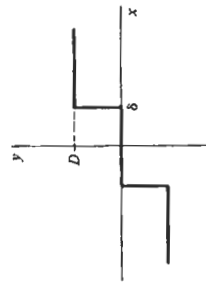
and the corresponding gain to the bias input component is:

$$N_B(\sigma, B) = \frac{1}{\sqrt{2\pi}\sigma B} \int_{-\infty}^{\infty} y(r + B) \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

This section uses the additional function $G(x) = xPI(x) + PF(x)$.

The functions $PF(x)$, $PI(x)$, and $G(x)$ are plotted in Fig. E.2-1.

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma, B)$ and $N_B(\sigma, B)$
 1. General odd quantizer	$D_0 = 0$ N is index of last quantizer level.	$N_R = \frac{1}{\sigma} \sum_{i=1}^N (D_i - D_{i-1}) \left[PF\left(\frac{\delta_i + B}{\sigma}\right) + PF\left(\frac{\delta_i - B}{\sigma}\right) \right]$ $N_B = \frac{1}{B} \sum_{i=1}^N (D_i - D_{i-1}) \left[PI\left(\frac{\delta_i + B}{\sigma}\right) - PI\left(\frac{\delta_i - B}{\sigma}\right) \right]$
 2. Uniform quantizer	N is index of last quantizer level.	$N_R = \frac{D}{\sigma} \sum_{i=1}^N \left[PF\left(\frac{2i-1}{2} \frac{h}{\sigma} + \frac{B}{\sigma}\right) + PF\left(\frac{2i-1}{2} \frac{h}{\sigma} - \frac{B}{\sigma}\right) \right]$ $N_B = \frac{D}{B} \sum_{i=1}^N \left[PI\left(\frac{2i-1}{2} \frac{h}{\sigma} + \frac{B}{\sigma}\right) - PI\left(\frac{2i-1}{2} \frac{h}{\sigma} - \frac{B}{\sigma}\right) \right]$
 3. Relay with dead zone		$N_R = \frac{D}{\sigma} \left[PF\left(\frac{\delta + B}{\sigma}\right) + PF\left(\frac{\delta - B}{\sigma}\right) \right]$ $N_B = \frac{D}{B} \left[PI\left(\frac{\delta + B}{\sigma}\right) - PI\left(\frac{\delta - B}{\sigma}\right) \right]$

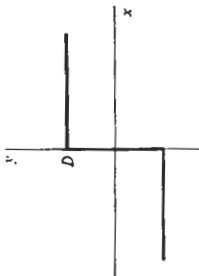
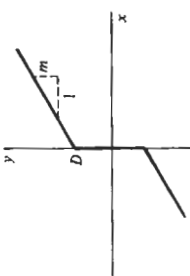
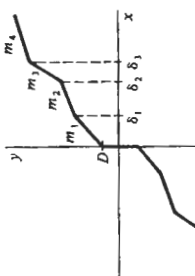
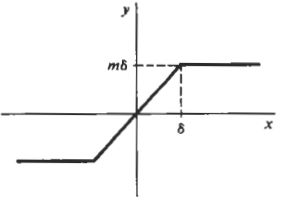
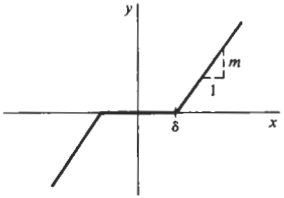
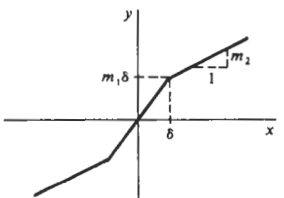
		$N_R = 2 \frac{D}{\sigma} PF \left(\frac{B}{\sigma} \right)$ $N_B = \frac{D}{B} \left[2PI \left(\frac{B}{\sigma} \right) - 1 \right]$
4. Ideal relay 		$N_R = 2 \frac{D}{\sigma} PF \left(\frac{B}{\sigma} \right) + m$ $N_B = \frac{D}{B} \left[2PI \left(\frac{B}{\sigma} \right) - 1 \right] + m$
5. Preload 	N is index of last linear segment.	$N_R = 2 \frac{D}{\sigma} PF \left(\frac{B}{\sigma} \right) + 2m_N - m_1 + \sum_{i=1}^{N-1} (m_i - m_{i+1}) \left[PI \left(\frac{\delta_i + B}{\sigma} \right) + PI \left(\frac{\delta_i - B}{\sigma} \right) \right]$ $N_B = \frac{D}{B} \left[2PI \left(\frac{B}{\sigma} \right) - 1 \right] + 2m_N - m_1 + \frac{\sigma}{B} \sum_{i=1}^{N-1} (m_i - m_{i+1}) \left[G \left(\frac{\delta_i + B}{\sigma} \right) - G \left(\frac{\delta_i - B}{\sigma} \right) \right]$
6. General piecewise-linear odd memoryless nonlinearity		

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma, B)$ and $N_B(\sigma, B)$
 7. Sharp saturation or limiter		$N_R = m \left[PI\left(\frac{\delta + B}{\sigma}\right) + PI\left(\frac{\delta - B}{\sigma}\right) - 1 \right]$ $N_B = m \left\{ \frac{\sigma}{B} \left[G\left(\frac{\delta + B}{\sigma}\right) - G\left(\frac{\delta - B}{\sigma}\right) \right] - 1 \right\}$
 8. Dead zone or threshold		$N_R = m \left[2 - PI\left(\frac{\delta + B}{\sigma}\right) - PI\left(\frac{\delta - B}{\sigma}\right) \right]$ $N_B = m \left\{ 2 - \frac{\sigma}{B} \left[G\left(\frac{\delta + B}{\sigma}\right) - G\left(\frac{\delta - B}{\sigma}\right) \right] \right\}$
 9. Gain-changing nonlinearity		$N_R = m_1 + (m_2 - m_1) \left[2 - PI\left(\frac{\delta + B}{\sigma}\right) - PI\left(\frac{\delta - B}{\sigma}\right) \right]$ $N_B = m_1 + (m_2 - m_1) \left\{ 2 - \frac{\sigma}{B} \left[G\left(\frac{\delta + B}{\sigma}\right) - G\left(\frac{\delta - B}{\sigma}\right) \right] \right\}$

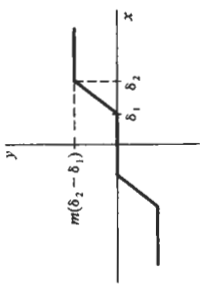
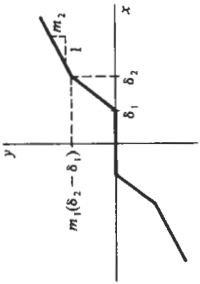
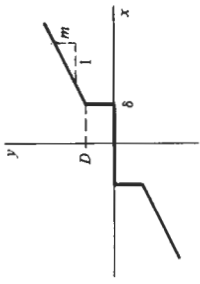
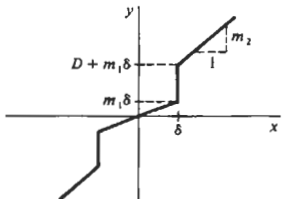
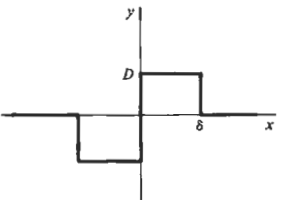
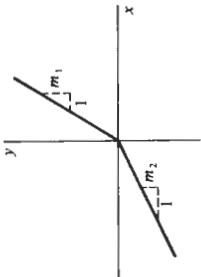
 10. Limiter with dead zone		$N_R = m \left[PI \left(\frac{\delta_2 + B}{\sigma} \right) + PI \left(\frac{\delta_2 - B}{\sigma} \right) - PI \left(\frac{\delta_1 + B}{\sigma} \right) - PI \left(\frac{\delta_1 - B}{\sigma} \right) \right]$ $N_B = m \frac{\sigma}{B} \left[G \left(\frac{\delta_2 + B}{\sigma} \right) - G \left(\frac{\delta_2 - B}{\sigma} \right) - G \left(\frac{\delta_1 + B}{\sigma} \right) + G \left(\frac{\delta_1 - B}{\sigma} \right) \right]$
 11. Gain-changing nonlinearity with dead zone		$N_R = 2m_2 - m_1 \left[PI \left(\frac{\delta_1 + B}{\sigma} \right) + PI \left(\frac{\delta_1 - B}{\sigma} \right) \right] +$ $(m_1 - m_2) \left[PI \left(\frac{\delta_2 + B}{\sigma} \right) + PI \left(\frac{\delta_2 - B}{\sigma} \right) \right]$ $N_B = 2m_2 + \frac{\sigma}{B} \left\{ -m_1 \left[G \left(\frac{\delta_1 + B}{\sigma} \right) - G \left(\frac{\delta_1 - B}{\sigma} \right) \right] + \right.$ $\left. (m_1 - m_2) \left[G \left(\frac{\delta_2 + B}{\sigma} \right) - G \left(\frac{\delta_2 - B}{\sigma} \right) \right] \right\}$
 12.		$N_R = m \left[2 - PI \left(\frac{\delta + B}{\sigma} \right) - PI \left(\frac{\delta - B}{\sigma} \right) \right] + \frac{D}{\sigma} \left[PF \left(\frac{\delta + B}{\sigma} \right) + PF \left(\frac{\delta - B}{\sigma} \right) \right]$ $N_B = m \left\{ 2 - \frac{\sigma}{B} \left[G \left(\frac{\delta + B}{\sigma} \right) - G \left(\frac{\delta - B}{\sigma} \right) \right] \right\} + \frac{D}{B} \left[PI \left(\frac{\delta + B}{\sigma} \right) - PI \left(\frac{\delta - B}{\sigma} \right) \right]$

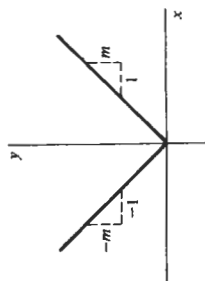
TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma, B)$ and $N_B(\sigma, B)$
13. 		$N_R = m_1 + \frac{D}{\sigma} \left[PF\left(\frac{\delta + B}{\sigma}\right) + PF\left(\frac{\delta - B}{\sigma}\right) \right] +$ $(m_2 - m_1) \left[2 - PI\left(\frac{\delta + B}{\sigma}\right) - PI\left(\frac{\delta - B}{\sigma}\right) \right]$ $N_B = m_1 + \frac{D}{B} \left[PI\left(\frac{\delta + B}{\sigma}\right) - PI\left(\frac{\delta - B}{\sigma}\right) \right] +$ $(m_2 - m_1) \left\{ 2 - \frac{\sigma}{B} \left[G\left(\frac{\delta + B}{\sigma}\right) - G\left(\frac{\delta - B}{\sigma}\right) \right] \right\}$
14. 		$N_R = \frac{D}{\sigma} \left[2PF\left(\frac{B}{\sigma}\right) - PF\left(\frac{\delta + B}{\sigma}\right) - PF\left(\frac{\delta - B}{\sigma}\right) \right]$ $N_B = \frac{D}{B} \left[2PI\left(\frac{B}{\sigma}\right) - PI\left(\frac{\delta + B}{\sigma}\right) + PI\left(\frac{\delta - B}{\sigma}\right) - 1 \right]$
15. $y = c$		$N_R = 0$ $N_B = \frac{c}{B}$
16. Linear gain $y = x$		$N_R = 1$ $N_B = 1$

$y = x \mid x \mid$		$N_R = 4\sigma PF\left(\frac{B}{\sigma}\right) + 2B\left[2PI\left(\frac{B}{\sigma}\right) - 1\right]$ $N_B = 2\sigma PF\left(\frac{B}{\sigma}\right) + B\left[1 + \left(\frac{\sigma}{B}\right)^2\right]\left[2PI\left(\frac{B}{\sigma}\right) - 1\right]$
17. Odd-square law		$N_R = 3\sigma^2 + 3B^2$ $N_B = 3\sigma^2 + B^2$
18. Cubic characteristic		$N_R = 15\sigma^4 + 30\sigma^2B^2 + 5B^4$ $N_B = 15\sigma^4 + 10\sigma^2B^2 + B^4$
20. Quintic characteristic		$N_R = 105\sigma^6 + 315\sigma^4B^2 + 105\sigma^2B^4 + 7B^6$ $N_B = 105\sigma^6 + 105\sigma^4B^2 + 21\sigma^2B^4 + B^6$
22.		
$y = x^n$	$n = 3, 5, 7, \dots$	$N_R = \sum_{k(\text{even})=0}^{n-1} \frac{n!}{k!(n-k)!} \sigma^{n-k-1} B^k (1)(3) \cdots (n-k)$ $N_B = B^{n-1} + \sum_{k(\text{odd})=1}^{n-2} \frac{n!}{k!(n-k)!} \sigma^{n-k} B^{k-1} (1)(3) \cdots (n-k-1)$
24.	See Sec. 7.2	

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

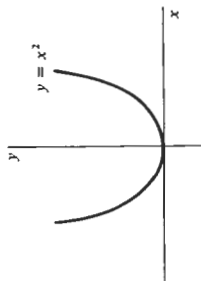
Nonlinearity	Comments	$N_R(\sigma, B)$ and $N_B(\sigma, B)$
$y = M \sin mx$		$N_R = Mm \cos mB e^{-m^2 \sigma^2 / 2}$ $N_B = \frac{M}{B} \sin mB e^{-m^2 \sigma^2 / 2}$
29. Harmonic nonlinearity		
$y = M \sinh mx$		$N_R = Mm \cosh mB e^{m^2 \sigma^2 / 2}$ $N_B = \frac{M}{B} \sinh mB e^{m^2 \sigma^2 / 2}$
30.		
$y = 1 - e^{-cx} \quad (x \geq 0)$ $= -(1 - e^{cx}) \quad (x < 0)$		$N_R = \frac{2}{\sigma} PF\left(\frac{B}{\sigma}\right) + \frac{1}{\sigma} e^{c^2 \sigma^2 / 2} \left\{ e^{cB} \left[c\sigma - c\sigma PI\left(c\sigma + \frac{B}{\sigma}\right) - PF\left(c\sigma + \frac{B}{\sigma}\right) \right] \right.$ $\left. + e^{-cB} \left[c\sigma - c\sigma PI\left(c\sigma - \frac{B}{\sigma}\right) - PF\left(c\sigma - \frac{B}{\sigma}\right) \right] \right\}$ $N_B = \frac{1}{B} \left[2PI\left(\frac{B}{\sigma}\right) - 1 \right] + \frac{1}{B} e^{c^2 \sigma^2 / 2} \left\{ e^{cB} \left[1 - PI\left(c\sigma + \frac{B}{\sigma}\right) \right] - e^{-cB} \left[1 - PI\left(c\sigma - \frac{B}{\sigma}\right) \right] \right\}$
31. Exponential saturation		
		$N_R = m_2 + (m_1 - m_2) PI\left(\frac{B}{\sigma}\right)$ $N_B = m_2 + \frac{\sigma}{B} (m_1 - m_2) G\left(\frac{B}{\sigma}\right)$



52. Absolute value

$$N_R = m \left[2PI \left(\frac{B}{\sigma} \right) - 1 \right]$$

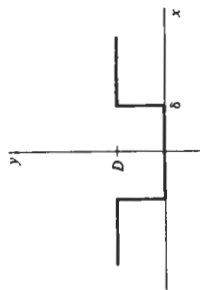
$$N_B = m \left[2 \frac{\sigma}{B} G \left(\frac{B}{\sigma} \right) - 1 \right]$$



53. Square-law

$$N_R = 2B$$

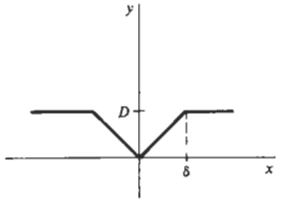
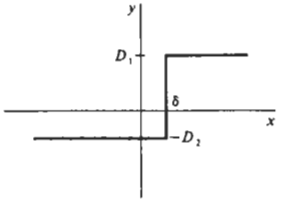
$$N_B = \frac{1}{B} [\sigma^2 + B^2]$$



$$N_R = \frac{D}{\sigma} \left[-PF \left(\frac{\delta + B}{\sigma} \right) + PF \left(\frac{\delta - B}{\sigma} \right) \right]$$

$$N_B = \frac{D}{B} \left[2 - PI \left(\frac{\delta + B}{\sigma} \right) - PI \left(\frac{\delta - B}{\sigma} \right) \right]$$

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma, B)$ and $N_B(\sigma, B)$
55. 		$N_R = \frac{D}{\delta} \left[2PI\left(\frac{B}{\sigma}\right) - 1 - PI\left(\frac{\delta+B}{\sigma}\right) + PI\left(\frac{\delta-B}{\sigma}\right) \right]$ $N_B = \frac{D}{B} \left\{ 2 - \frac{B}{\delta} + \frac{\sigma}{\delta} \left[2G\left(\frac{B}{\sigma}\right) - G\left(\frac{\delta+B}{\sigma}\right) - G\left(\frac{\delta-B}{\sigma}\right) \right] \right\}$
56. Biased ideal relay 		$N_R = \frac{D_1 + D_2}{\sigma} PF\left(\frac{\delta - B}{\sigma}\right)$ $N_B = \frac{D_1}{B} - \frac{D_1 + D_2}{B} PI\left(\frac{\delta - B}{\sigma}\right)$

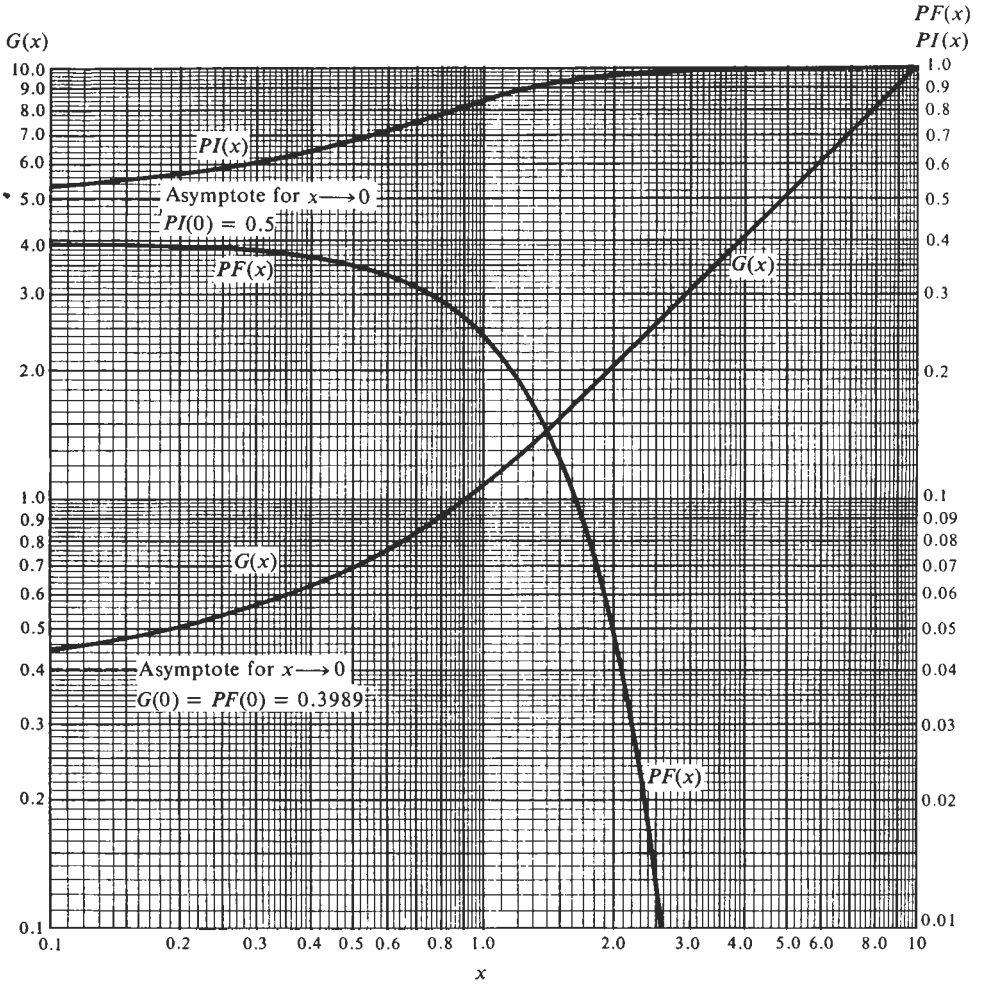


Figure E.2-1 Graphs of $PF(x)$, $PI(x)$, and $G(x)$.

E.3 GAUSSIAN-PLUS-BIAS-PLUS-SINUSOID-INPUT RIDFs

$$x(t) = r(t) + B + A \sin (\omega t + \theta)$$

The gain to the gaussian input component is given by

$$N_R(\sigma, B, A) = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma^3} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr y(r + B + A \sin \theta) r \exp \left(-\frac{r^2}{2\sigma^2} \right)$$

the gain to the bias input component is

$$N_B(\sigma, B, A) = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma B} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr y(r + B + A \sin \theta) \exp \left(-\frac{r^2}{2\sigma^2} \right)$$

and the corresponding gain to the sinusoid input component is

$$N_A(\sigma, B, A) = \frac{2}{(2\pi)^{\frac{3}{2}} \sigma A} \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dr y(r + B + A \sin \theta) \sin \theta \exp \left(-\frac{r^2}{2\sigma^2} \right)$$

TABLE OF RANDOM-INPUT DESCRIBING FUNCTIONS (RIDFs) (Continued)

Nonlinearity	Comments	$N_R(\sigma, B, A)$, $N_B(\sigma, B, A)$, and $N_A(\sigma, B, A)$
$y = x^3$ 18. Cubic characteristic		$N_R = 3\sigma^2 + 3B^2 + \frac{3}{8}A^2$ $N_B = 3\sigma^2 + B^2 + \frac{3}{8}A^2$ $N_A = 3\sigma^2 + 3B^2 + \frac{3}{4}A^2$
 29. Harmonic nonlinearity	J_0 and J_1 are the Bessel functions of orders 0 and 1, respectively.	$N_R = Mm \cos mB \exp\left(-\frac{m^2\sigma^2}{2}\right) J_0(mA)$ $N_B = \frac{M}{B} \sin mB \exp\left(-\frac{m^2\sigma^2}{2}\right) J_0(mA)$ $N_A = \frac{2M}{A} \cos mB \exp\left(-\frac{m^2\sigma^2}{2}\right) J_1(mA)$

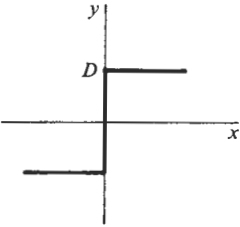


Figure E.3-1 Three-input RIDFs for the ideal-relay nonlinearity.

In 3 parts

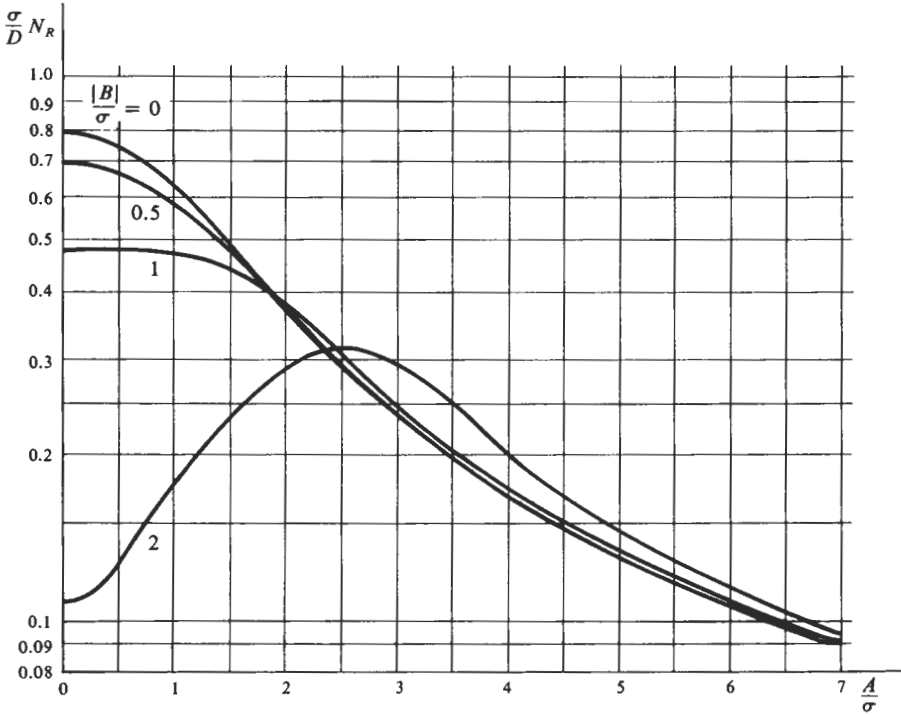


Figure E.3-1a Gain to the gaussian input component. (ideal relay)

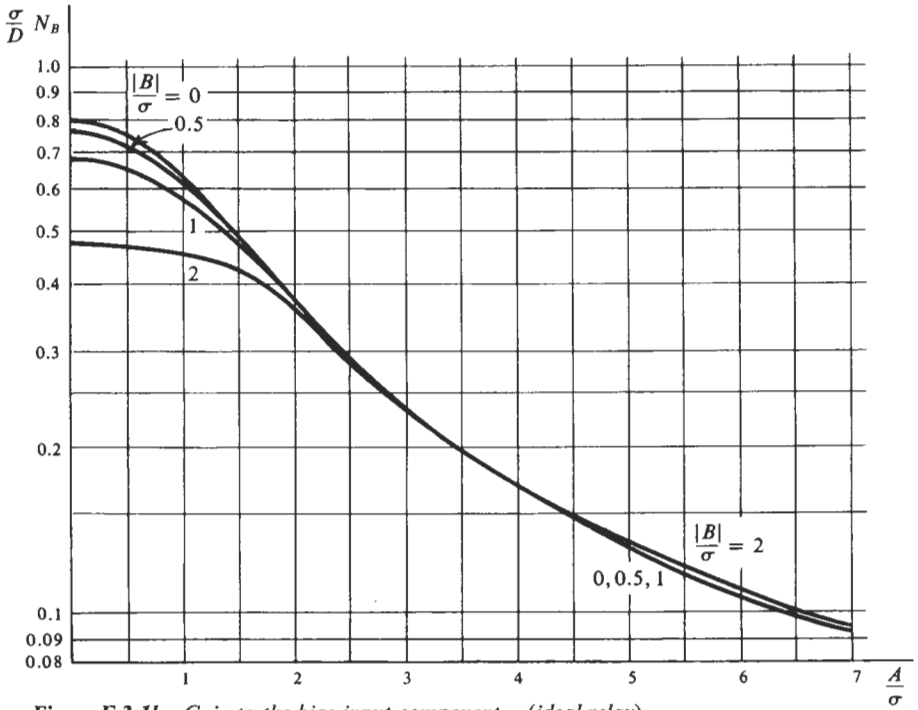


Figure E.3-1b Gain to the bias input component. (ideal relay)

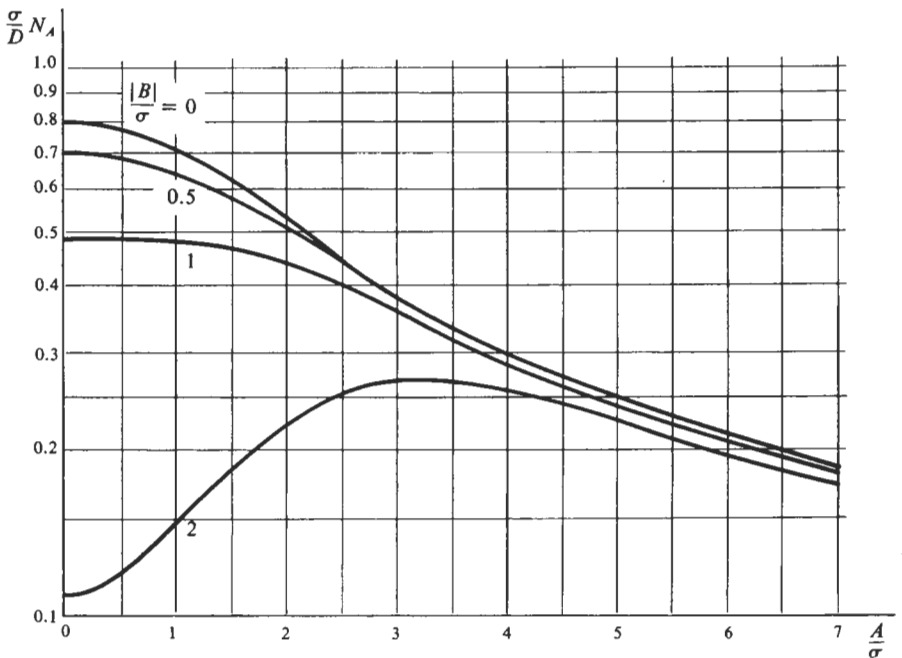


Figure E.3-1c Gain to the sinusoid input component. (ideal relay)

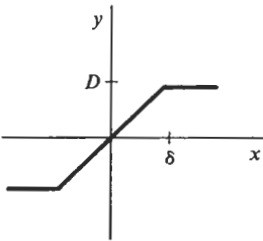


Figure E.3-2 Three-input RIDFs for the limiter nonlinearity.

In 11 parts

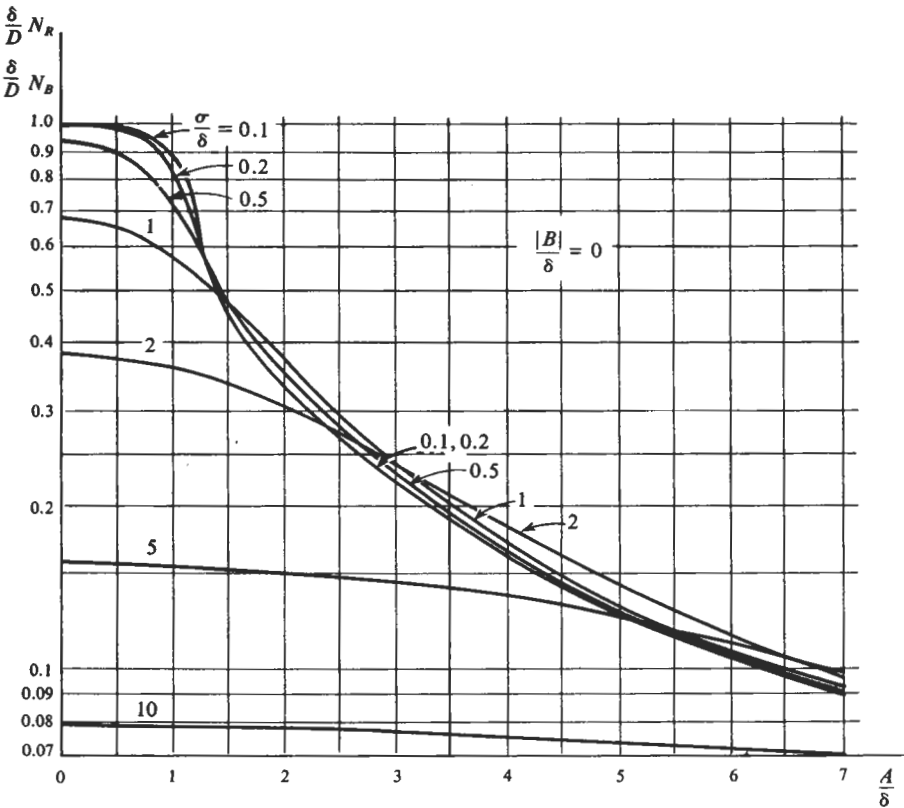


Figure E.3-2a Gain to the Gaussian and bias input components. (limiter, $|B|/\delta = 0$)

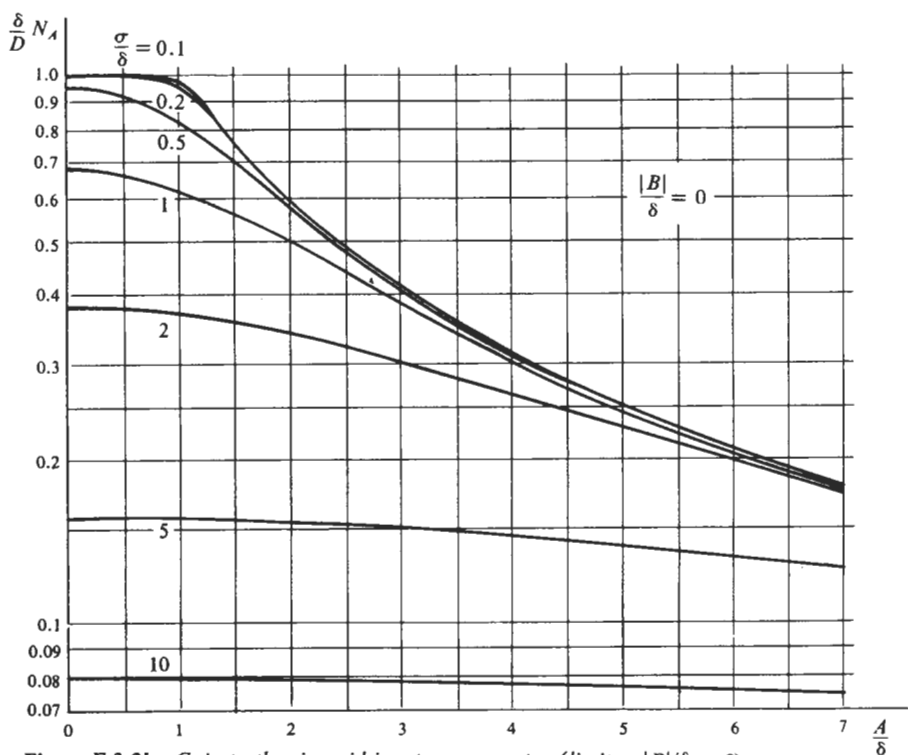


Figure E.3-2b Gain to the sinusoid input component. (limiter, $|B|/\delta = 0$)

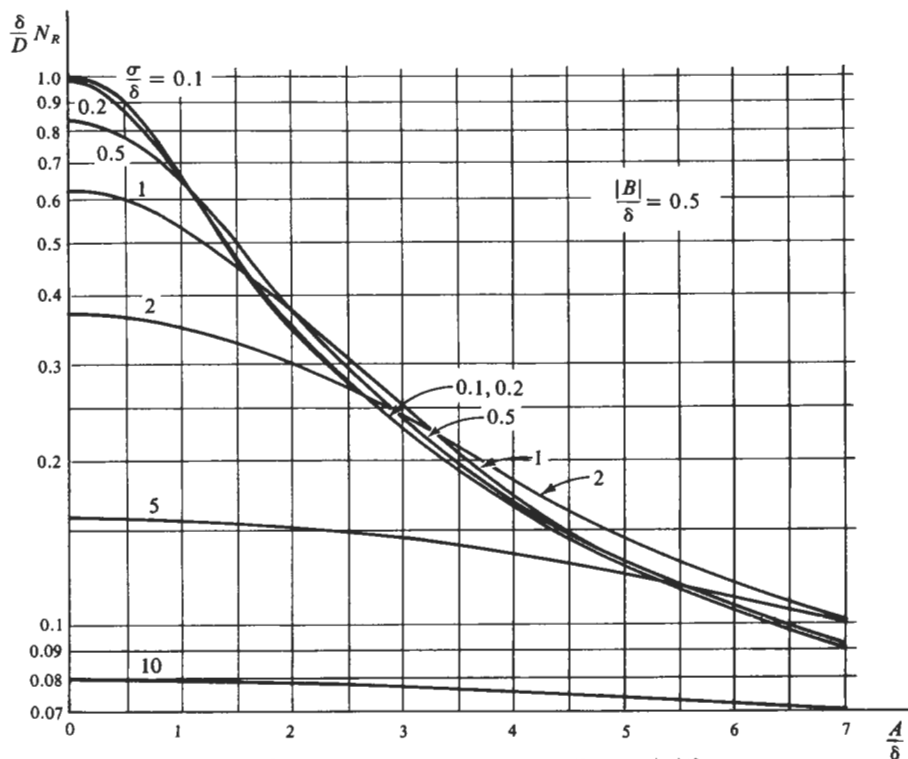


Figure E.3-2c Gain to the gaussian input component. (limiter, $|B|/\delta = 0.5$)

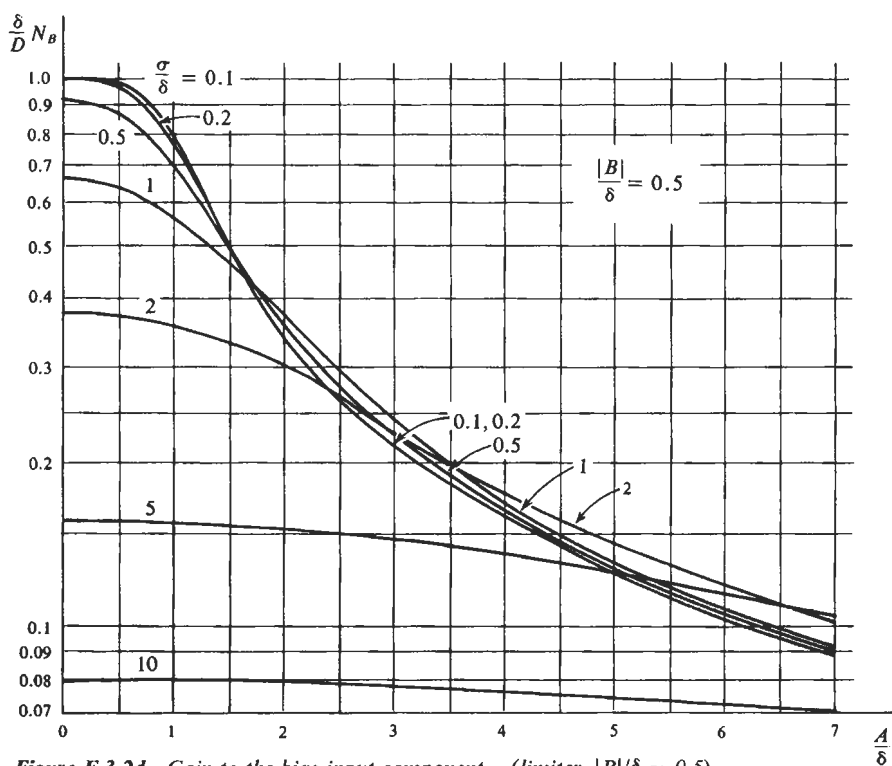


Figure E.3-2d Gain to the bias input component. (limiter, $|B|/\delta = 0.5$)

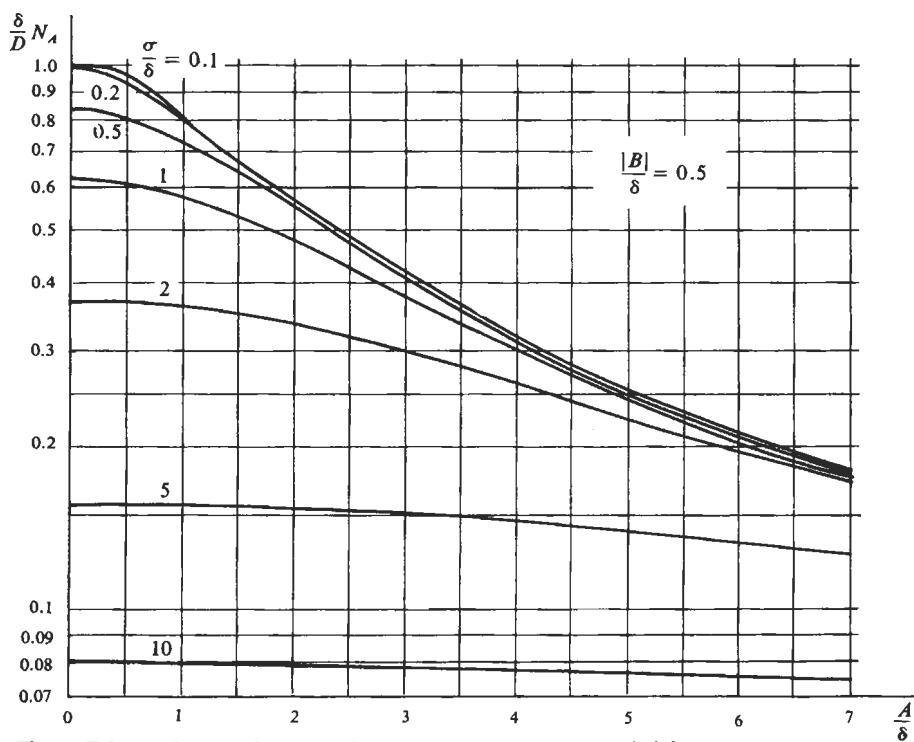


Figure E.3-2e Gain to the sinusoid input component. (limiter, $|B|/\delta = 0.5$)

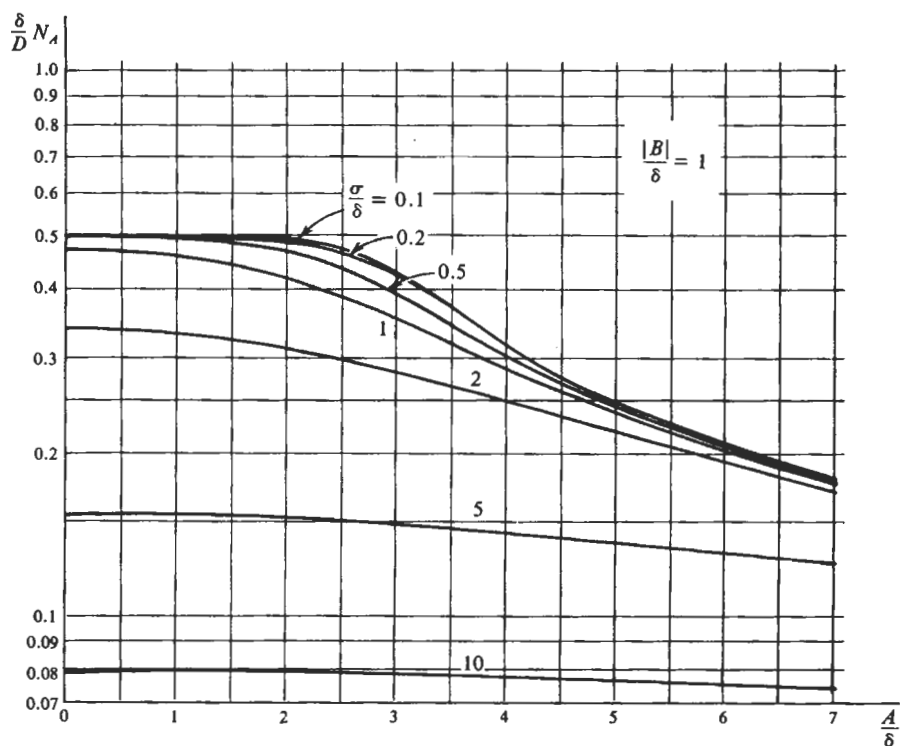


Figure E.3-2h Gain to the sinusoid input component. (limiter, $|B|/\delta = 1$)

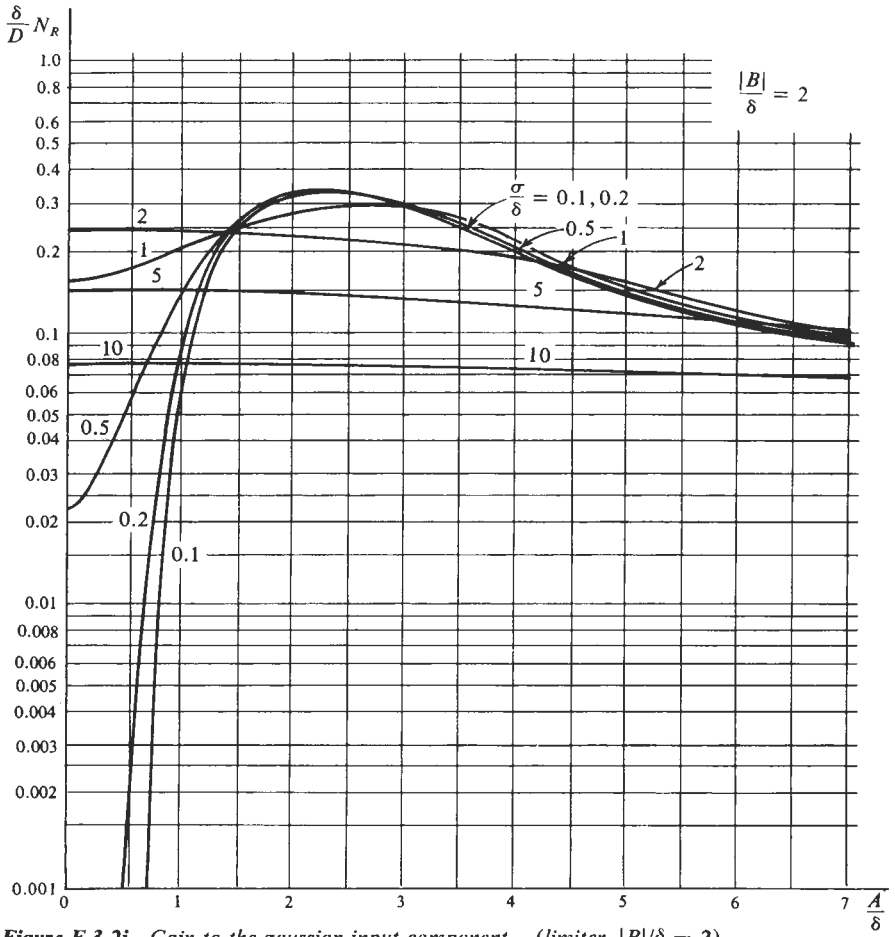


Figure E.3-2i Gain to the Gaussian input component. (limiter, $|B|/\delta = 2$)

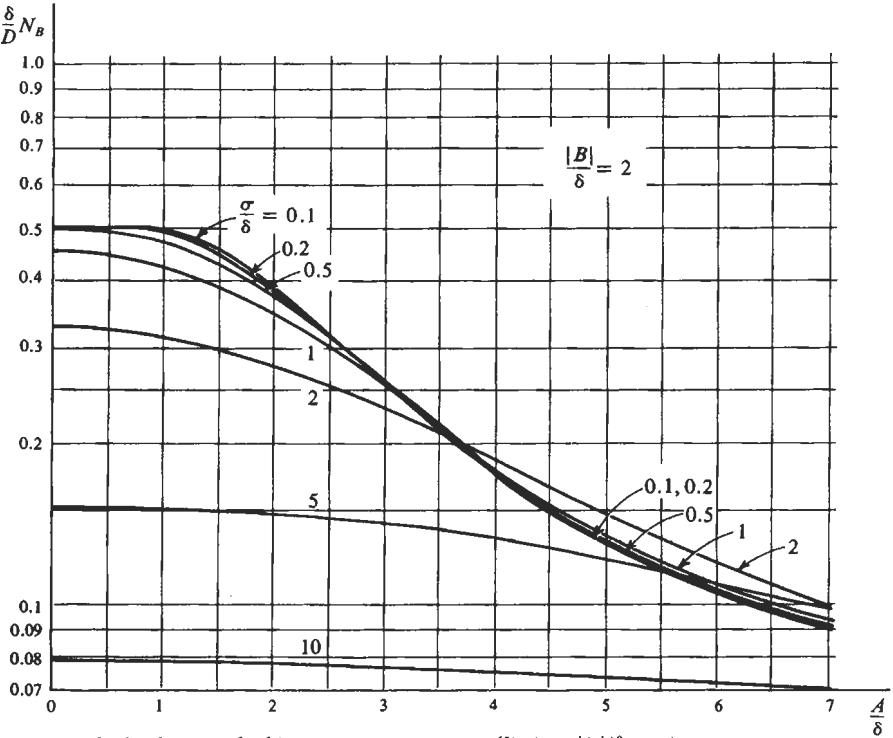


Figure E.3-2j Gain to the bias input component. (limiter, $|B|/\delta = 2$)

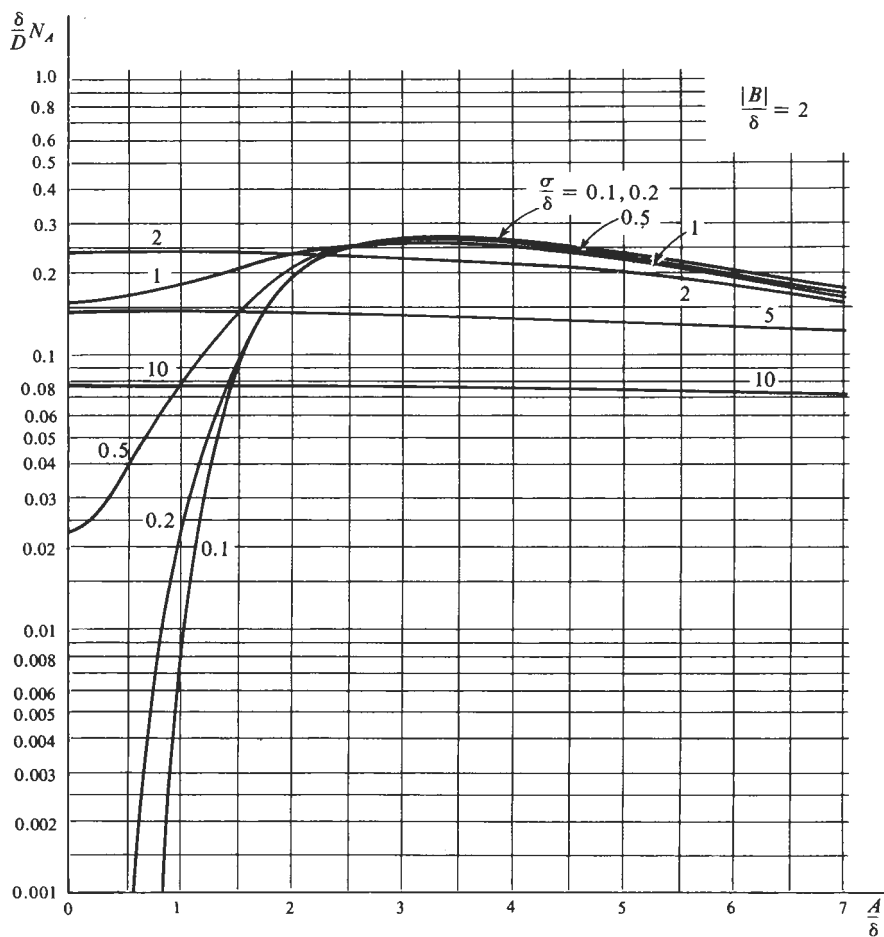


Figure E.3-2k Gain to the sinusoid input component. (limiter, $|B|/\delta = 2$)

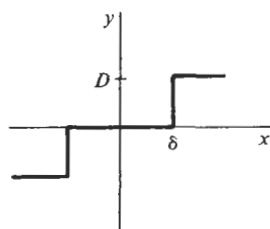


Figure E.3-3 Three-input RIDFs for the relay with dead zone nonlinearity.

In 11 parts

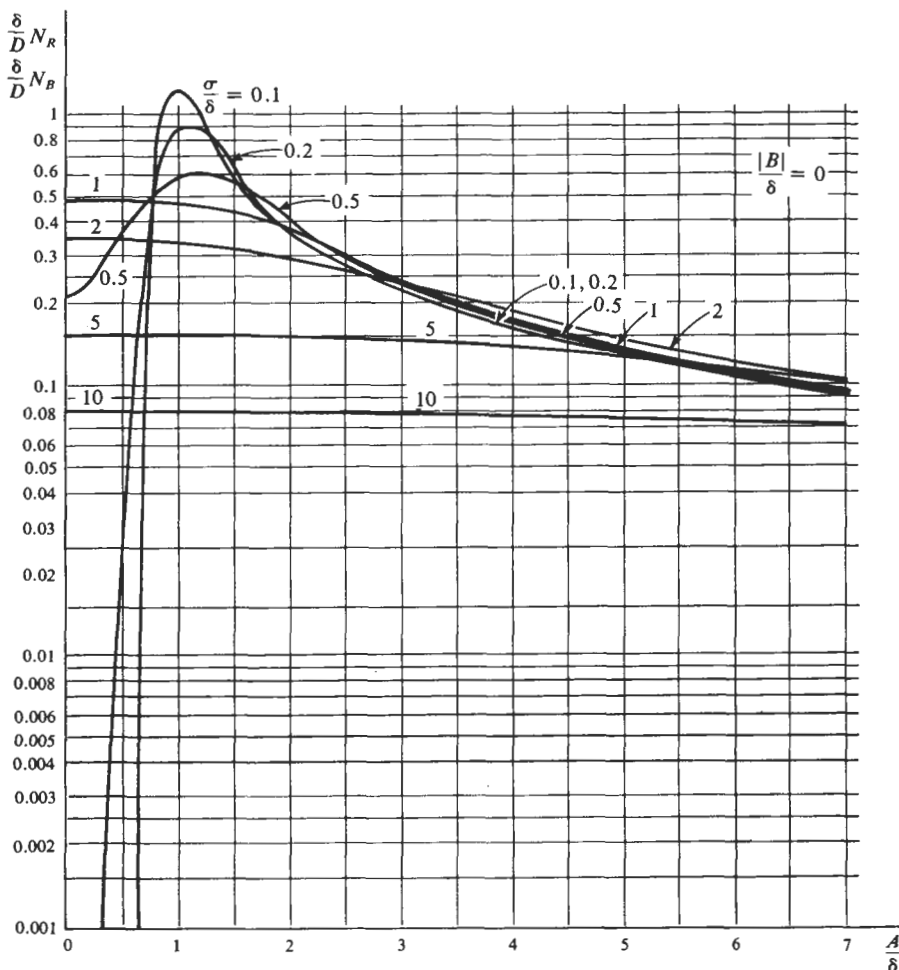


Figure E.3-3a Gain to the gaussian and bias input components. (relay with dead zone, $|B|/\delta = 0$)

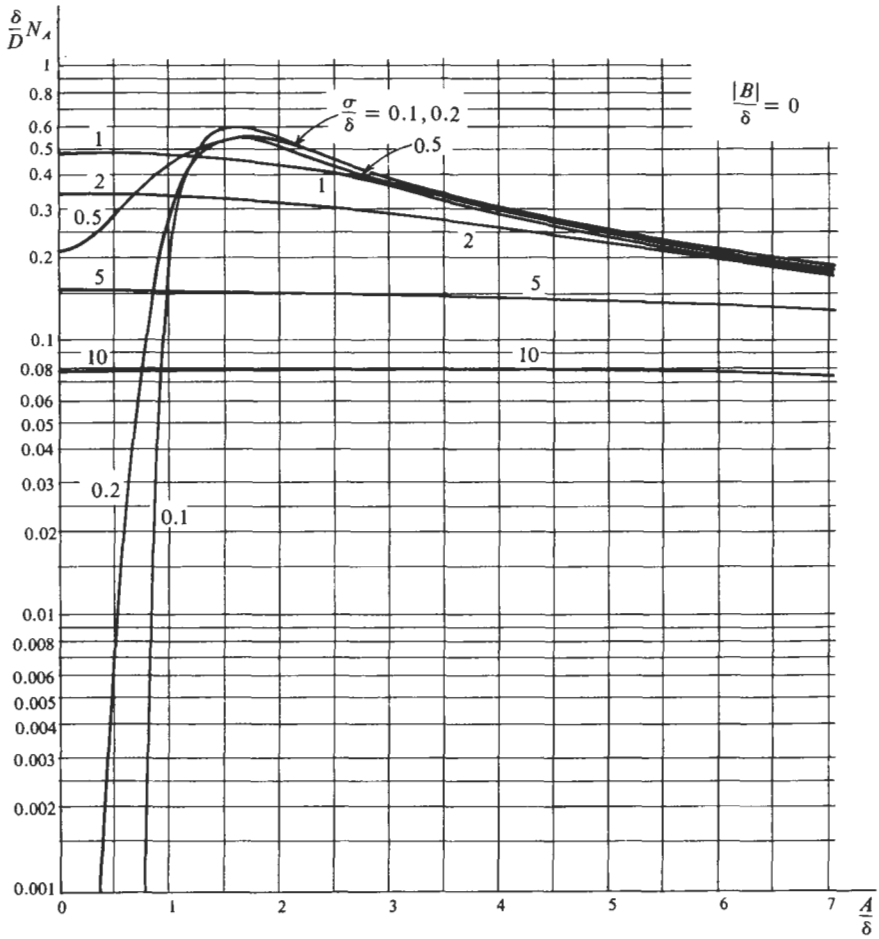


Figure E.3-3b Gain to the sinusoid input component. (relay with dead zone, $|B|/\delta = 0$)

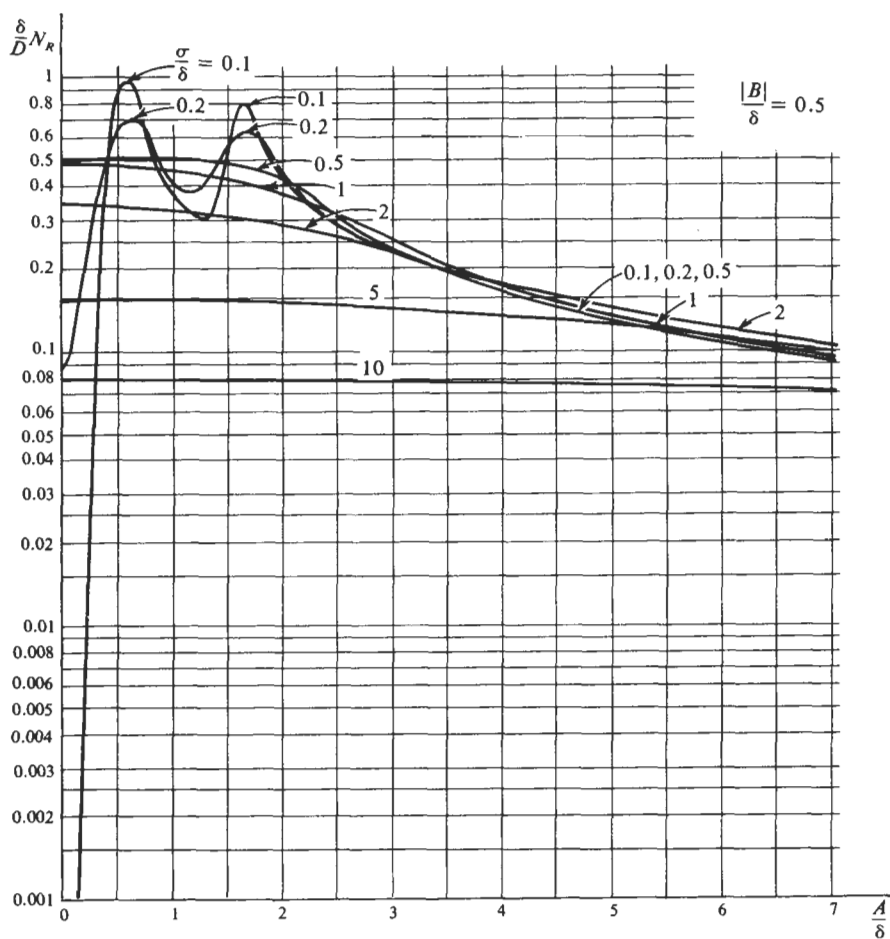


Figure E.3-3c Gain to the gaussian input component. (relay with dead zone, $|B|/\delta = 0.5$)

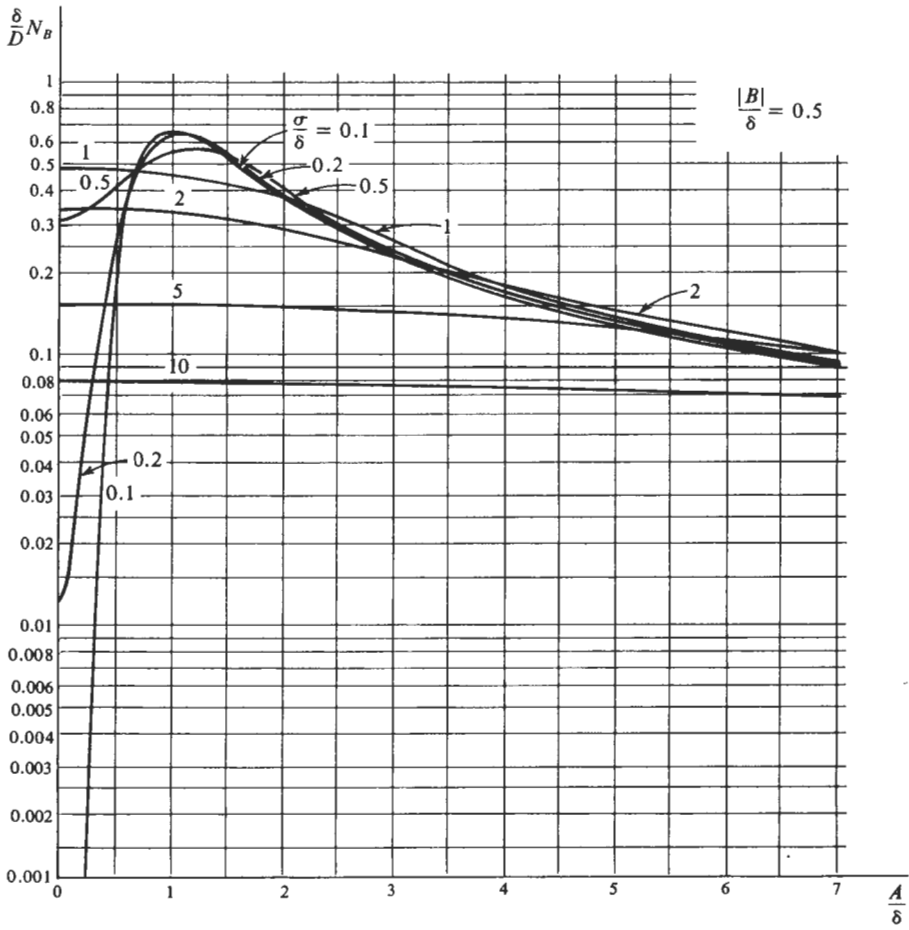


Figure E.3-3d Gain to the bias input component. (relay with dead zone, $|B|/\delta = 0.5$)

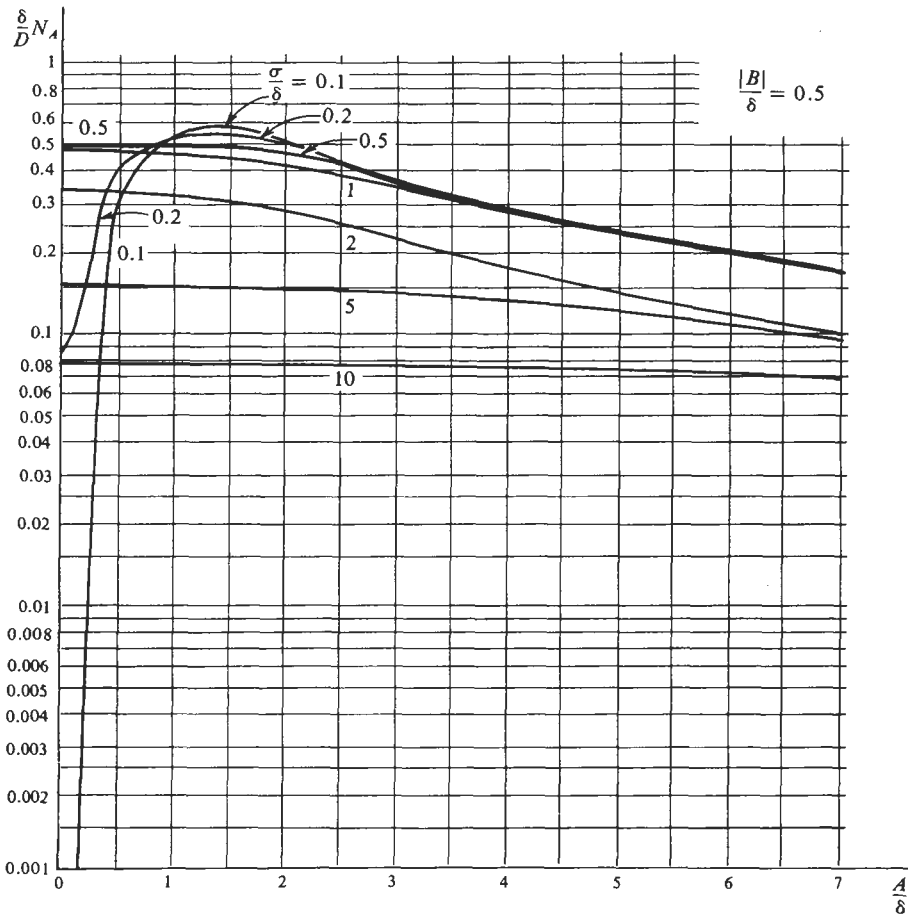


Figure E.3-3e Gain to the sinusoid input component. (relay with dead zone, $|B|/\delta = 0.5$)

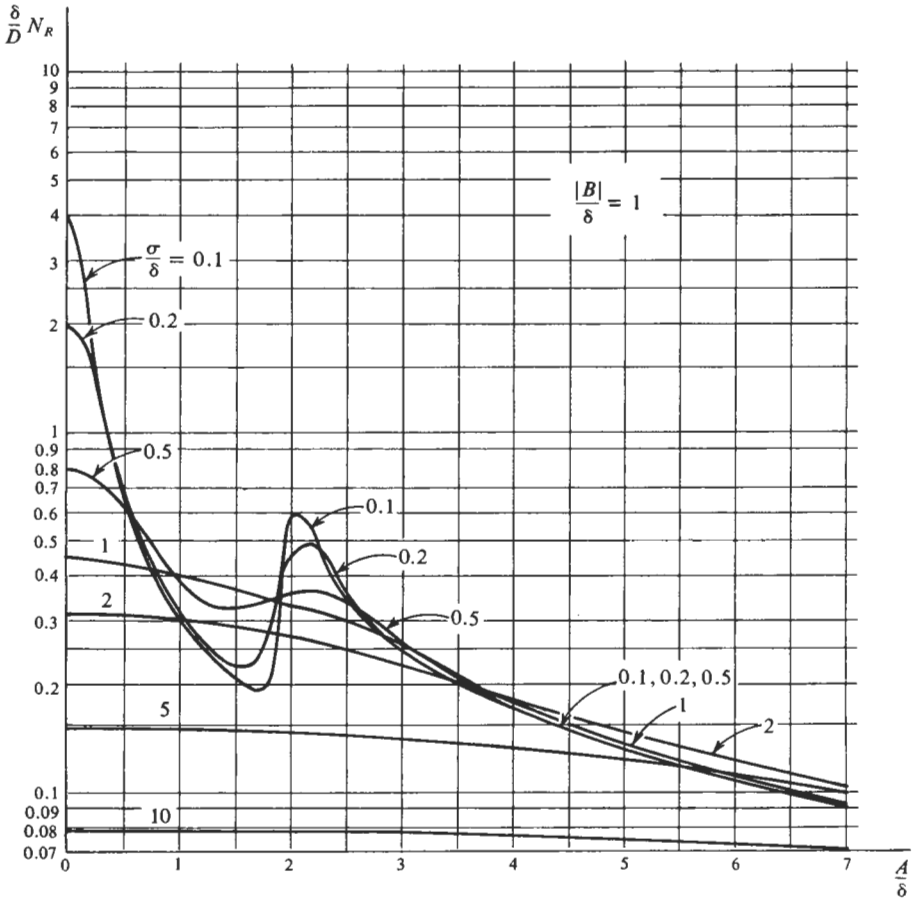


Figure E.3-3f Gain to the Gaussian input component. (relay with dead zone, $|B|/\delta = 1$)

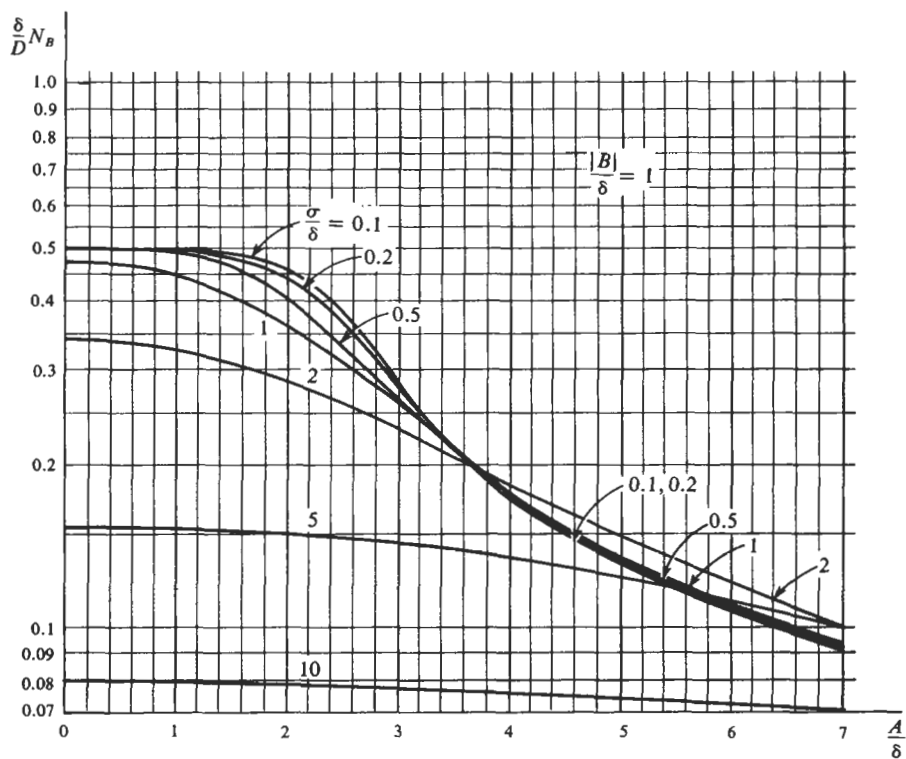


Figure E.3-3g Gain to the bias input component. (relay with dead zone, $|B|/\delta = 1$)

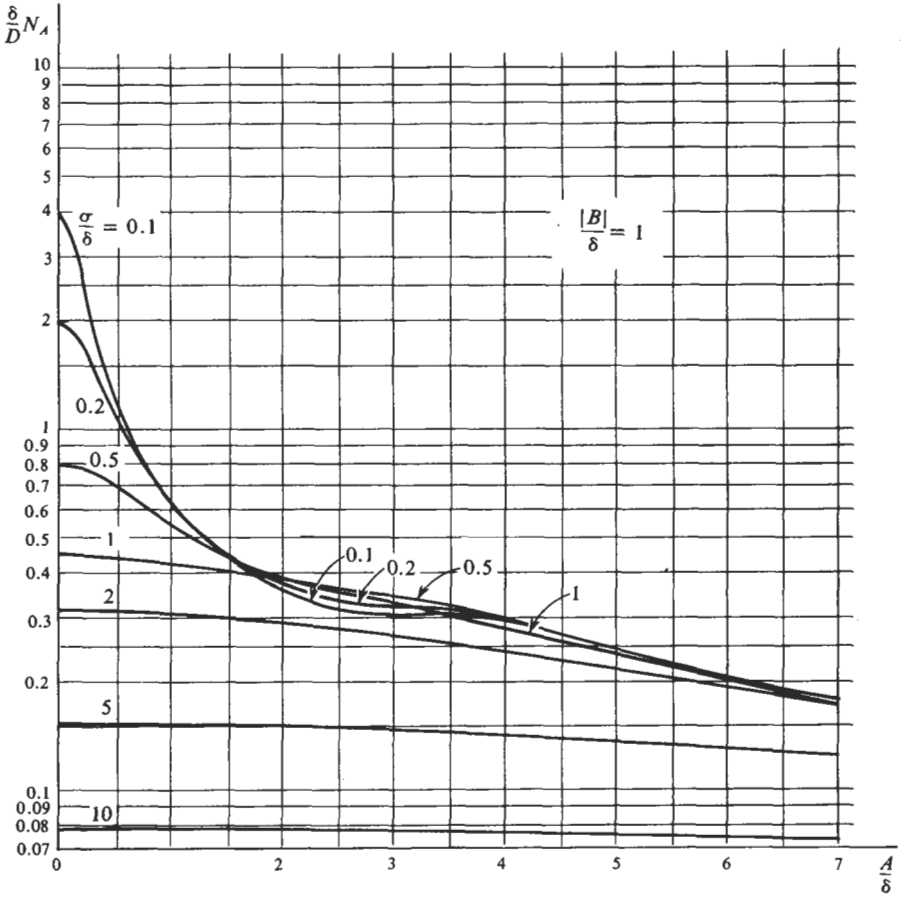


Figure E.3-3h Gain to the sinusoid input component. (relay with dead zone, $|B|/\delta = 1$)

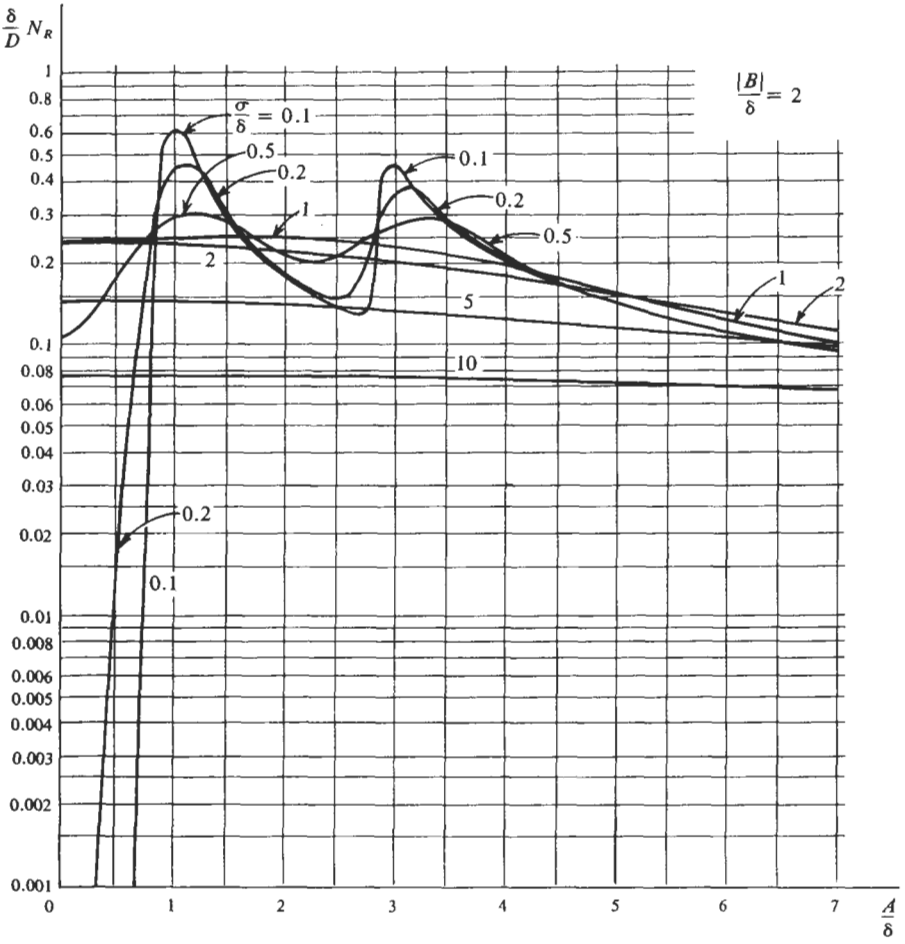


Figure E.3-3i Gain to the gaussian input component. (relay with dead zone, $|B|/\delta = 2$)

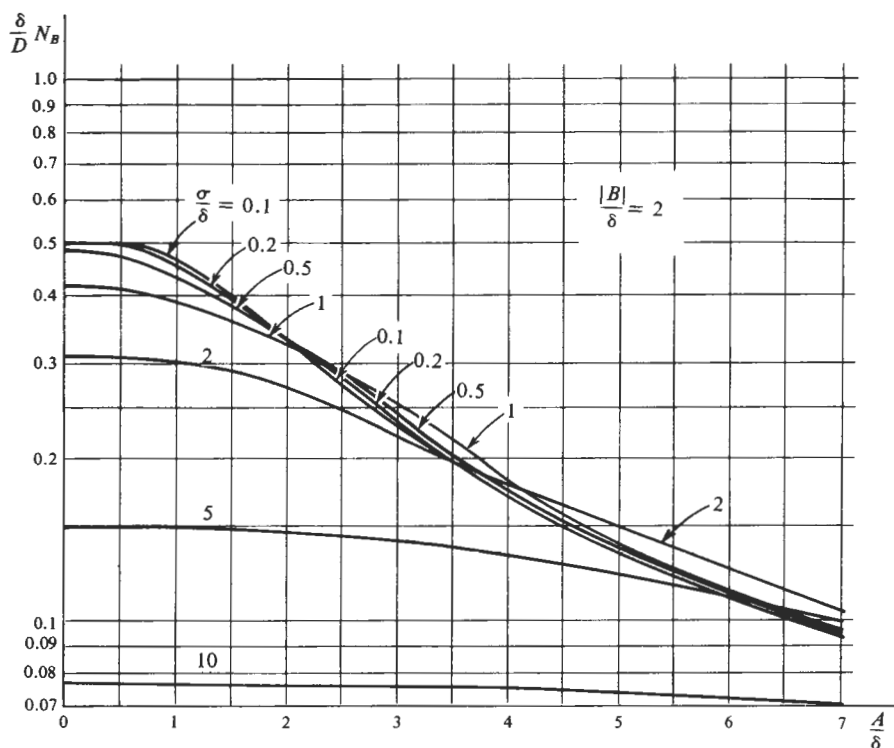


Figure E.3-3j Gain to the bias input component. (relay with dead zone, $|B|/\delta = 2$)

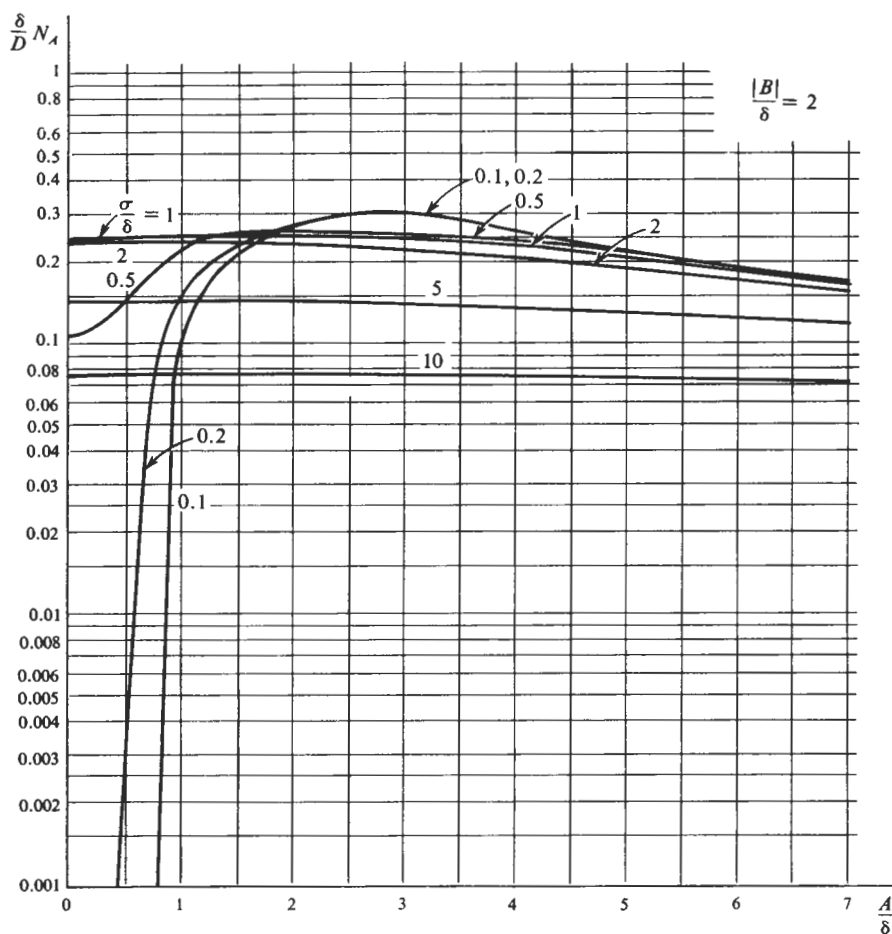


Figure E.3-3k Gain to the sinusoid input component. (relay with dead zone, $|B|/\delta = 2$)

APPENDIX F

TABLE OF SAMPLED DESCRIBING FUNCTIONS

$$x(t) = A \sin \left(\frac{1}{n} \omega_s t + \phi \right)$$

ω_s = sampling frequency

Note: The relay with or without hysteresis (two-level switch) can be treated simply by the method of Sec. 9.1. This table gives sampled describing functions for the relay with dead zone and the linear, lead pulse-width modulator.

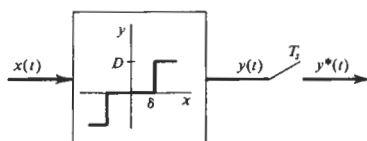


Figure F.1 Sampled describing function for the relay with dead zone nonlinearity (cf. Sec. 9.2).

$N(A, \phi)$ = sampled describing function

= amplitude and phase relation between the sinusoid $x(t)$ and the fundamental harmonic component of $y^*(t)$

$N^*(A, \phi)$ = z-transform describing function

$$= \left. \frac{\text{z transform of } y^*(t)}{\text{z transform of } x(t)} \right|_{z = \exp(j\omega_s T_s/n)}$$

$$N^*(A, \phi) = T_s N(A, \phi)$$

This figure, in 7 parts, consists of plots of extreme values of $-1/N(A, \phi)$ for $n = 2, 3, \dots, 8$. Only the unbiased output modes are included as they are the only modes of interest in the case of an unbiased input $x(t)$ (see Sec. 9.2).

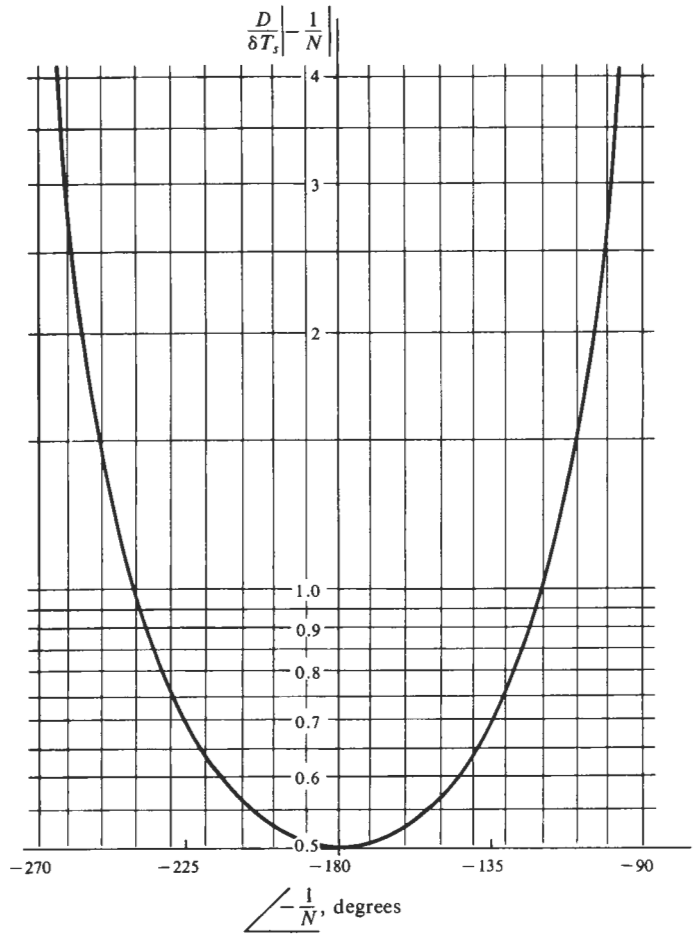


Figure F-1a $T = 2T_s$

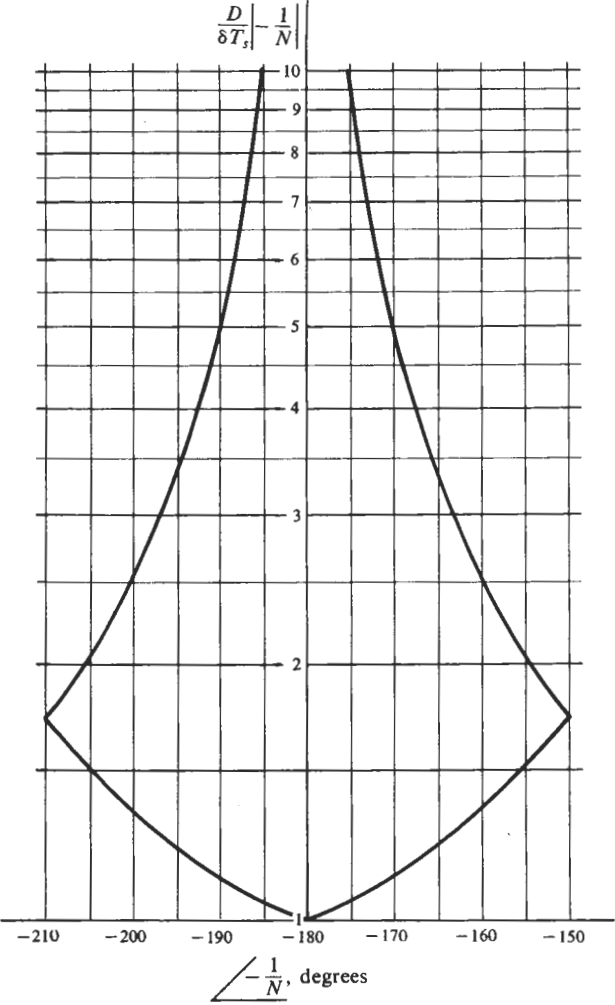


Figure F-1b $T = 3T_s$

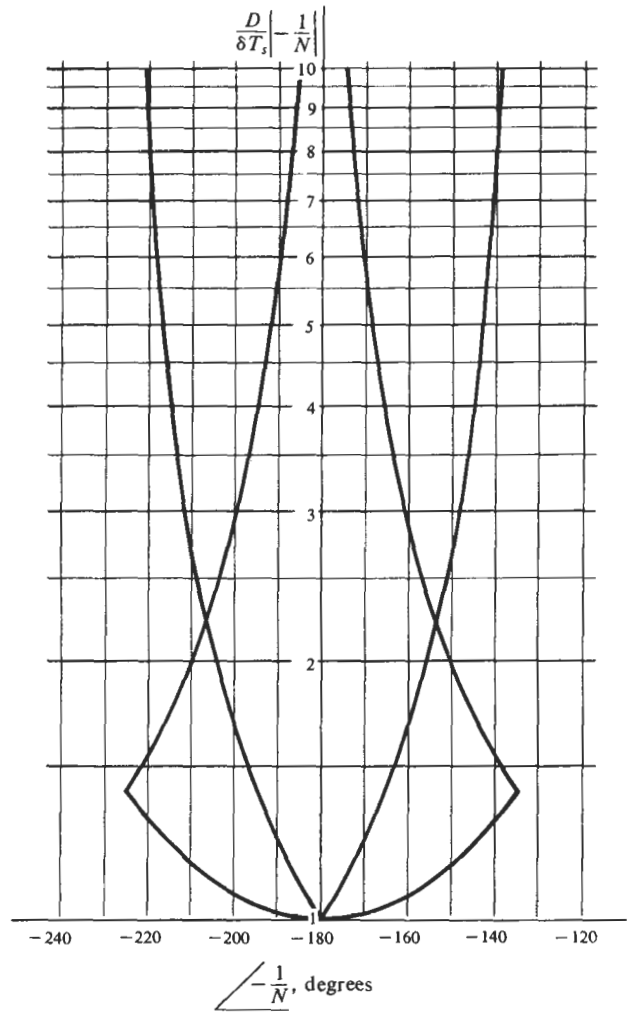


Figure F-1c $T = 4T_s$

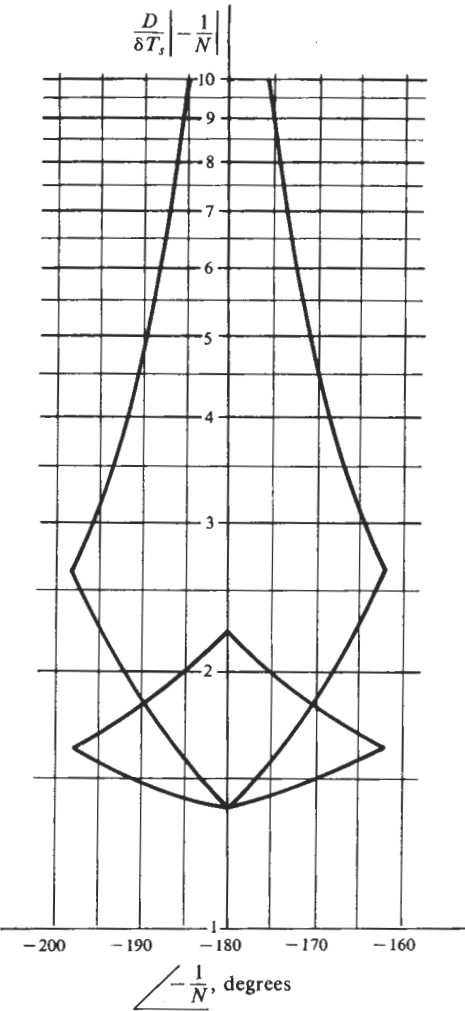


Figure F-1d $T = 5T_s$

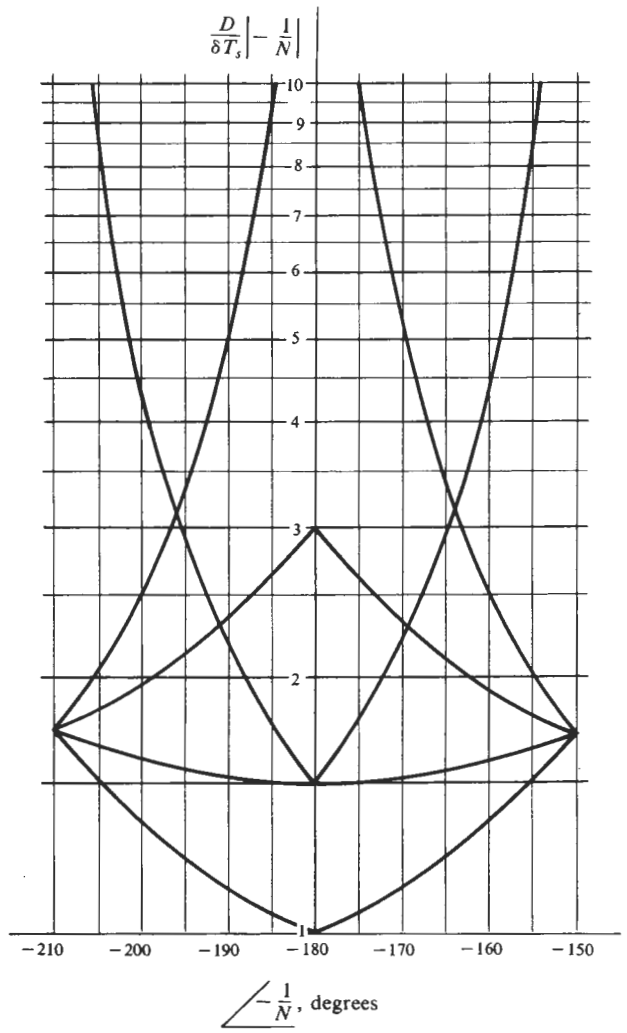


Figure F-1e $T = 6T_s$

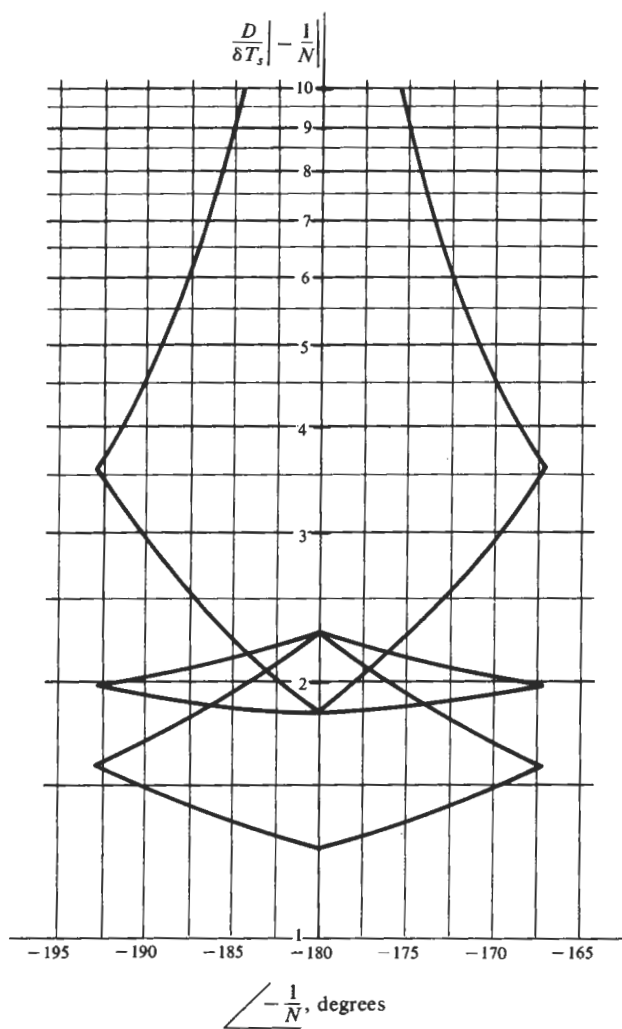


Figure F-1f $T = 7T_s$

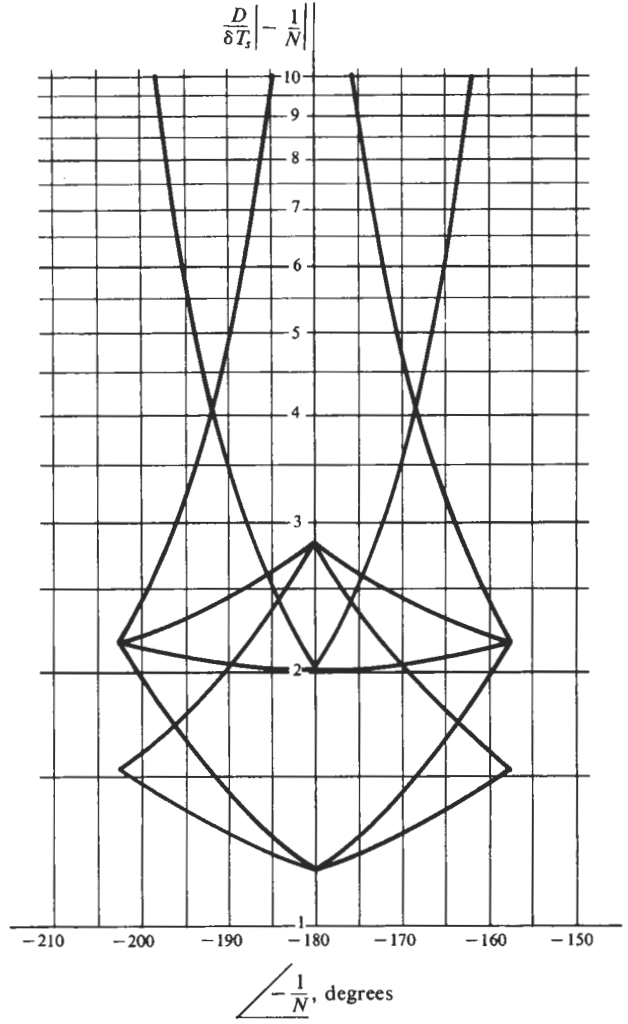


Figure F-1g $T = 8T_s$

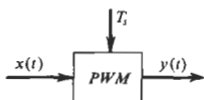


Figure F.2 Describing function for the linear, lead pulse-width modulator (cf. Sec. 9.5).

The operation of the *PWM* is defined by Eqs. (9.5-1,2), and is pictured in Fig. 9.5-1.

The describing function $N(A, \phi)$ is the usual amplitude and phase relation between the sinusoid $x(t)$ and the fundamental harmonic component of $y(t)$.

This figure, in 4 parts, consists of plots of $-1/N(A, \phi)$ for $n = 2, 4, 6, 8$ and a range of values of both A and ϕ . The solid curves correspond to constant values of A , and various values of ϕ are indicated. At other locations within the envelope of these curves, $-1/N(A, \phi)$ exists for other values of A and ϕ not shown.

The limiting values of these describing functions are

$$\begin{aligned} \lim_{A \rightarrow 0} \left(-\frac{Dk}{T_s} \frac{1}{N} \right) &= \frac{1}{2 \sin \phi} \angle \phi - 270^\circ & n = 2 \\ &= 1 \angle -180^\circ & n > 2 \\ \lim_{A \rightarrow \infty} \left(-\frac{Dk}{T_s} \frac{1}{N} \right) &= \frac{\pi k A}{4 T_s} \angle \phi - 180^\circ & \text{all } n \end{aligned}$$

where in each case ϕ may take any value in the range

$$0 < \phi < \frac{360}{n}$$

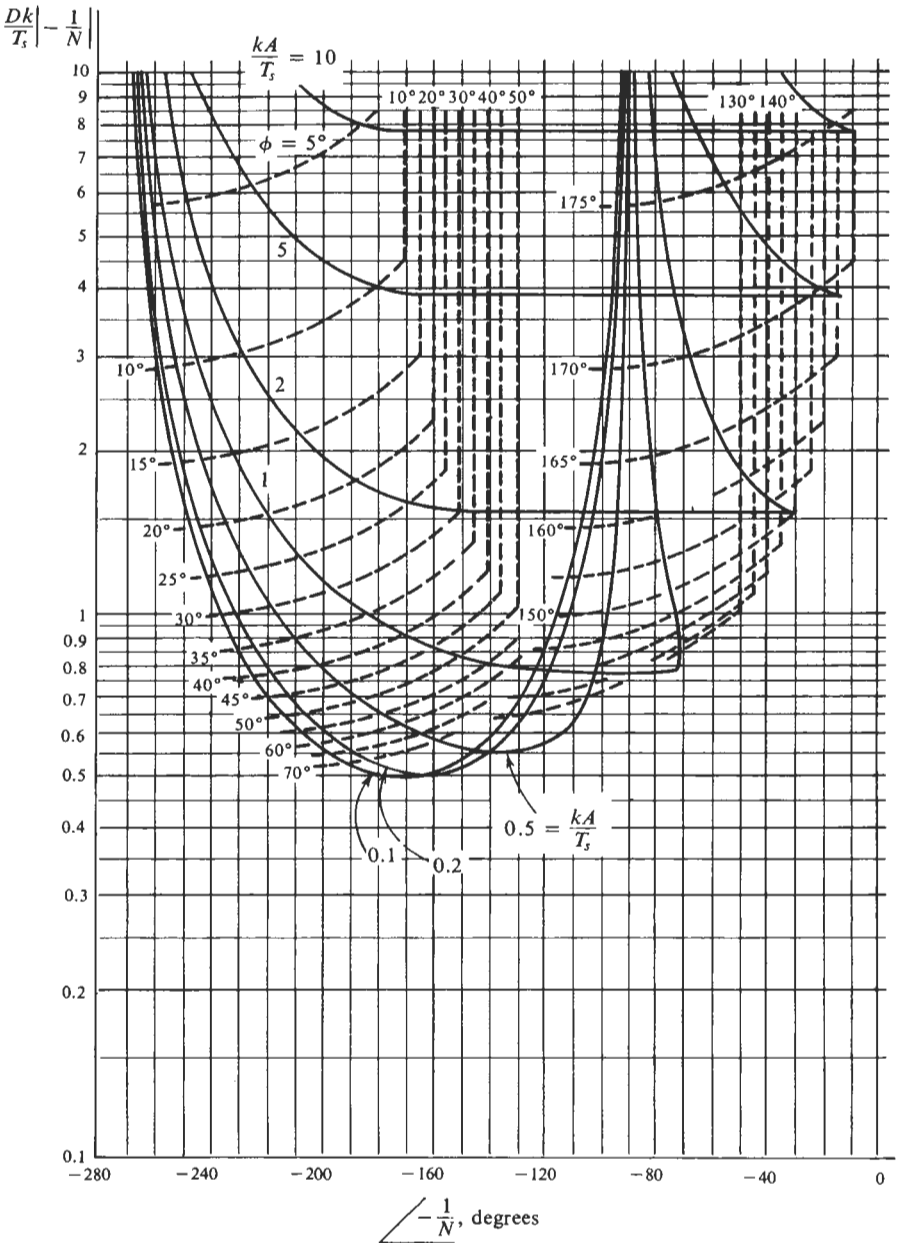


Figure F-2a $T = 2T_s$

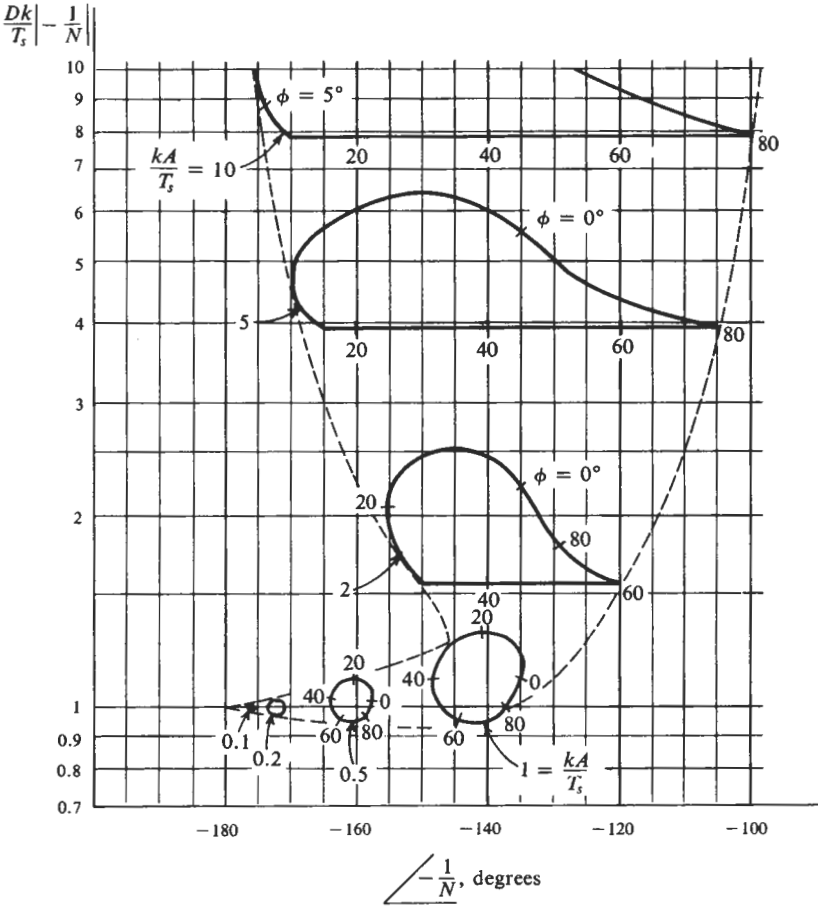


Figure F-2b $T = 4T_s$

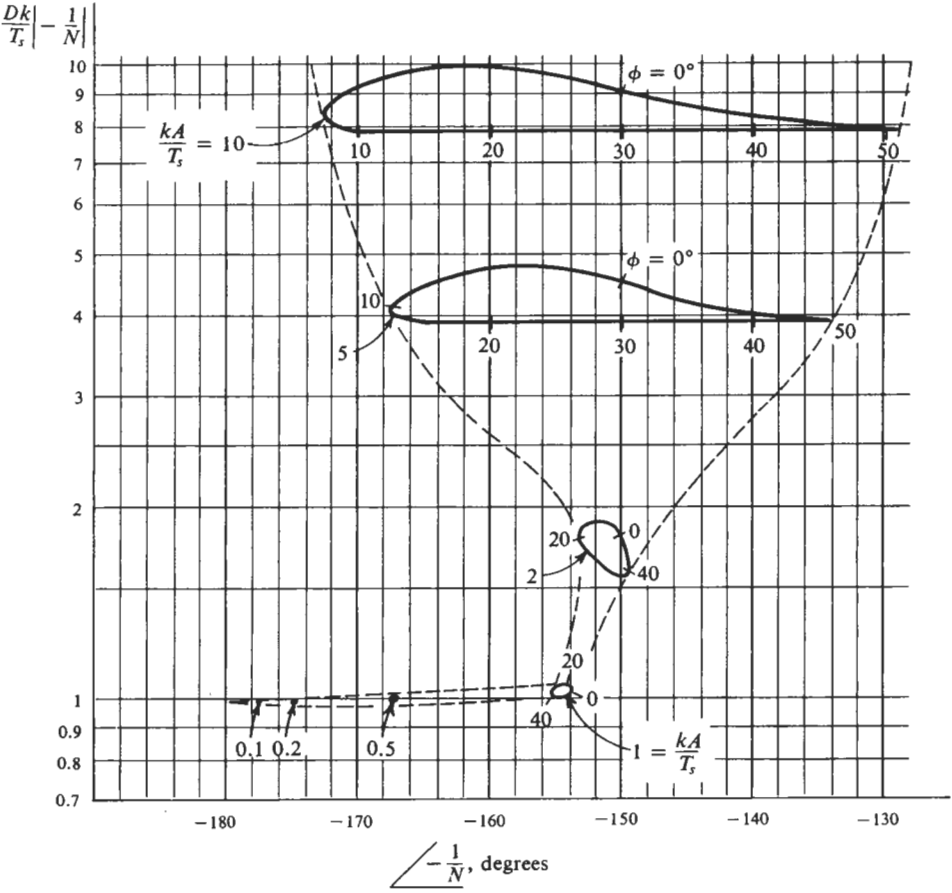


Figure F-2c $T = 6T_s$

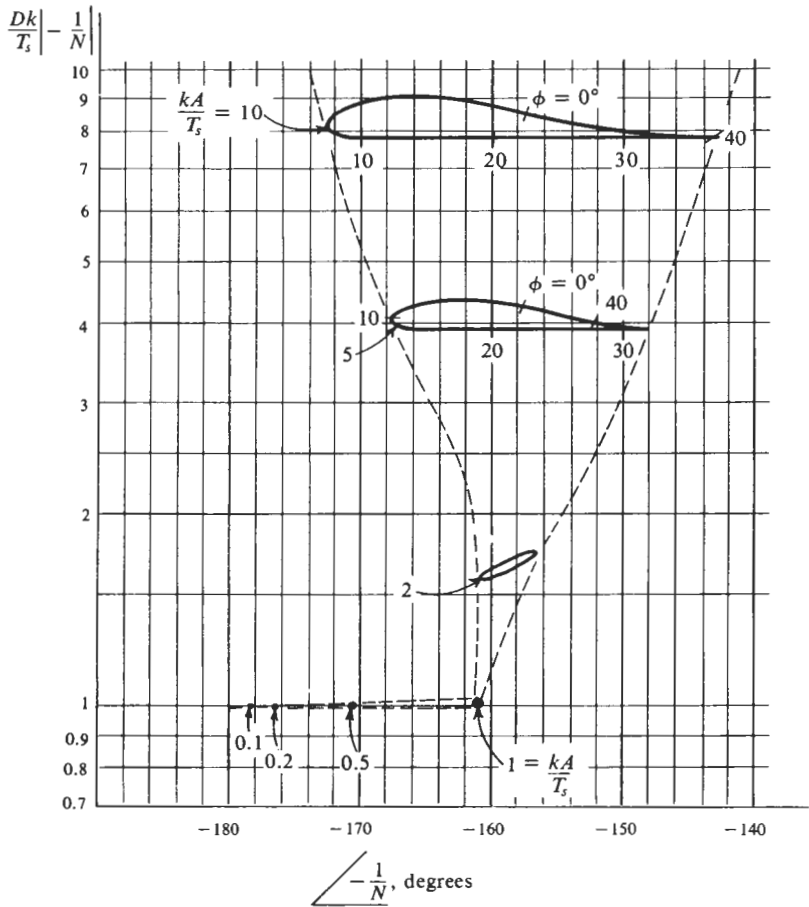


Figure F-2d $T = 8T_s$

ANALYTICAL JUSTIFICATION FOR
THE FILTER HYPOTHESIS¹

Presentation of the following development is motivated primarily by two considerations. First, it gives a mathematical setting for the *filter hypothesis*. This condition is arrived at directly from the system differential equation, as opposed to the physical or intuitive presentation in Chap. 3. Second, and quite importantly, this development points out that what one actually determines with the DF method is an *approximation to the first harmonic* of the true nonlinear oscillation, *not* the oscillation waveform itself.

Consider any autonomous nonlinear system which can be reduced to single-loop block-diagram form, including a single nonlinearity with input x and output $y(x, \dot{x})$, and a linear part $L(s)$, where

$$L(s) = \frac{P(s)}{Q(s)} \quad (\text{G-1})$$

$P(s)$ and $Q(s)$ are polynomials of arbitrary degree in s , the degree of $P(s)$ being lower than the degree of $Q(s)$. The equation defining the modes of behavior of the closed-loop system is

$$Q(s)x + P(s)y(x, \dot{x}) = 0 \quad (\text{G-2})$$

¹ This derivation closely follows that of Popov and Pal'tov (Chap. 6, Ref. 21, pp. 121–129, in the English translation.)

where $s = d/dt$. According to the DF point of view, the limit cycle behavior of this system is studied in terms of the existence of the solution

$$x = A \sin \omega t \quad (\text{G-3})$$

to the linearized equation [cf. Eq. (2.2-30)]

$$\left\{ Q(s) + P(s) \left[n_p(A, \omega) + \frac{n_q(A, \omega)}{\omega} s \right] \right\} x = 0 \quad (\text{G-4})$$

We now attempt to find the conditions which $P(s)$ and $Q(s)$ must satisfy in order that Eq. (G-2) will be sufficiently close to Eq. (G-4) in the presence of a strong nonlinearity, $y(x, \dot{x})$, when we seek a periodic solution $x(t)$ close to sinusoidal.

Let us therefore seek a solution to the given nonlinear differential equation (G-2) in the form

$$x(t) = x_1(t) + \epsilon x_r(t) \quad (\text{G-5})$$

where

$$x_1(t) = A_1 \sin \omega_1 t \quad (\text{G-6})$$

is the true first harmonic of the periodic solution, $x_r(t)$ is an arbitrary bounded time function, and ϵ is a small parameter, viz.,

$$\epsilon x_r(t) = \epsilon \sum_{k=2}^{\infty} A_k \sin(k\omega_1 t + \theta_k) \quad (\text{G-7})$$

Substituting for x from Eq. (G-5) into Eq. (G-2), we obtain

$$Q(s)(x_1 + \epsilon x_r) + P(s)y(x_1 + \epsilon x_r, \dot{x}_1 + \epsilon \dot{x}_r) = 0 \quad (\text{G-8})$$

Next, the nonlinearity output is expanded in the Taylor series (truncated at first order in ϵ):

$$y(x_1 + \epsilon x_r, \dot{x}_1 + \epsilon \dot{x}_r) \cong y(x_1, \dot{x}_1) + \epsilon \left[\frac{\partial y(x_1, \dot{x}_1)}{\partial x} x_r + \frac{\partial y(x_1, \dot{x}_1)}{\partial \dot{x}} \dot{x}_r \right] \quad (\text{G-9})$$

This expression states, in essence, that the periodic nonlinearity output in response to $x(t)$ is very close (i.e., to an order ϵ) to the response produced by the first harmonic, $x_1(t)$, alone. This will be true even for the discontinuous nonlinearities often encountered.

The first term in the right-hand member of Eq. (G-9) is the nonlinearity output when forced by a sinusoid of amplitude A_1 and frequency ω_1 . In DF notation, we have

$$y(x_1, \dot{x}_1) = F_0 + A_1 \left[n_p(A_1, \omega_1) + \frac{n_q(A_1, \omega_1)}{\omega_1} s \right] \sin \omega_1 t + \sum_{k=2}^{\infty} F_k \sin(k\omega_1 t + \varphi_k) \quad (\text{G-10})$$

where, as $k \rightarrow \infty$, we require $F_k \rightarrow 0$. The second term in the right-hand member of Eq. (G-9) is represented by its Fourier series, namely,

$$\epsilon \left[\frac{\partial y(x_1, \dot{x}_1)}{\partial x} x_r + \frac{\partial y(x_1, \dot{x}_1)}{\partial \dot{x}} \dot{x}_r \right] = \epsilon \sum_{k=0}^{\infty} G_k \sin(k\omega_1 t + \lambda_k) \quad (\text{G-11})$$

Substituting the expressions of Eqs. (G-6), (G-7), and (G-9) to (G-11) into (G-8) gives

$$\begin{aligned} Q(s)A_1 \sin \omega_1 t + \epsilon Q(s) \sum_{k=2}^{\infty} A_k \sin(k\omega_1 t + \theta_k) + P(s)F_0 \\ + P(s)A_1 \left[n_p(A_1, \omega_1) + \frac{n_q(A_1, \omega_1)}{\omega_1} s \right] \sin \omega_1 t + P(s) \sum_{k=2}^{\infty} F_k \sin(k\omega_1 t + \varphi_k) \\ + \epsilon P(s) \sum_{k=0}^{\infty} G_k \sin(k\omega_1 t + \lambda_k) = 0 \quad (\text{G-12}) \end{aligned}$$

In order to satisfy equality with zero, the coefficient of each harmonic in the left-hand member of Eq. (G-12) must be individually set to zero.

ZERO HARMONIC

From Eq. (G-12) we require

$$F_0 + \epsilon G_0 \sin \lambda_0 = 0 \quad (\text{G-13})$$

or to an order ϵ ,

$$F_0 = 0 \quad (\text{G-14})$$

Thus we have the first requirement imposed upon $y(x, \dot{x})$: the absence of a constant component. This is consistent with the content of Chap. 3 (although in Chap. 6 this restriction is removed).

FIRST HARMONIC

From Eq. (G-12) we require, upon division by $Q(s)$,

$$\begin{aligned} A_1 \sin \omega_1 t + L(s)A_1 \left[n_p(A_1, \omega_1) + \frac{n_q(A_1, \omega_1)}{\omega_1} s \right] \sin \omega_1 t \\ + \epsilon L(s)G_1 \sin(\omega_1 t + \lambda_1) = 0 \quad (\text{G-15}) \end{aligned}$$

or to an order ϵ ,

$$A \sin \omega t + L(s)A \left[n_p(A, \omega) + \frac{n_q(A, \omega)}{\omega} s \right] \sin \omega t = 0 \quad (\text{G-16})$$

This may be rewritten as

$$1 + L(j\omega)[n_p(A, \omega) + jn_q(A, \omega)] = 0 \quad (\text{G-17})$$

which is the familiar DF statement defining a system limit cycle. Hence, to an order ϵ , the DF method determines the *first harmonic* of the periodic solution of the nonlinear system [Eq. (G-2)], which is close to a sinusoidal solution (if this periodic solution exists).

HIGHER HARMONICS

For $k = 2, 3, \dots$, it follows from Eq. (G-12) that the k th harmonic terms must satisfy

$$\epsilon A_k \sin(k\omega_1 t + \theta_k) + L(jk\omega_1)F_k \sin(k\omega_1 t + \varphi_k) + \epsilon L(jk\omega_1)G_k \sin(k\omega_1 t + \lambda_k) = 0 \quad (\text{G-18})$$

In order to observe this equality, that is, in order that all the terms $\epsilon A_k \sin(k\omega_1 t + \theta_k)$ may in reality be small, we require small order of magnitude of the terms $|L(jk\omega_1)| F_k$. Each term must be at least of order ϵ ; that is, in comparison with A_1 , each term must be at least of the same order of magnitude as the quantity ϵx_r in comparison with x_1 . Thus, employing Eq. (G-17), we establish as a requirement the inequality

$$|L(jk\omega)| F_k \ll |L(j\omega)| A \sqrt{n_p^2(A, \omega) + n_q^2(A, \omega)} \quad (\text{G-19})$$

Since we are considering strong nonlinearities, it does not follow that the quantities F_k may be considered small in comparison with $A \sqrt{n_p^2 + n_q^2}$, particularly for lower values of k . Therefore we must require that

$$|L(jk\omega)| \ll |L(j\omega)| \quad \text{for } k = 2, 3, \dots \quad (\text{G-20})$$

and since the degree of $Q(s)$ exceeds the degree of $P(s)$,

$$|L(jk\omega)| \rightarrow 0 \quad \text{as } k \rightarrow \infty \quad (\text{G-21})$$

Equations (G-20) and (G-21) constitute the so-called "filter hypothesis."

APPENDIX H

INTRODUCTION TO PROBABILITY AND RANDOM PROCESSES

This appendix is not intended to be a definitive dissertation on the subject of random processes. The major concepts, definitions, and results which are employed in the text are stated here with little discussion and no proof. The reader who requires a more complete presentation of this material is referred to any one of several excellent books on the subject: among them Davenport and Root (Ref. 2), Laning and Battin (Ref. 3), and Lee (Ref. 4). Possibly the most important function served by this appendix is the definition of the notation and of certain conventions used in the text.

PROBABILITY

Consider an event E which is a possible outcome of a random experiment. We denote by $P(E)$ the *probability* of this event, and think of it intuitively as the limit, as the number of trials becomes large, of the ratio of the number of times E occurred to the number of times the experiment was tried. The joint event that A and B and C , etc., occurred is denoted by $ABC \cdots$, and the probability of this joint event, by $P(ABC \cdots)$. If these events A , B , C , etc., are mutually *independent*, which means that the occurrence of any one of them bears no relation to the occurrence of any other, the probability of the joint event is the product of the probabilities of the simple events. That is,

$$P(ABC \cdots) = P(A)P(B)P(C) \cdots \quad (\text{H-1})$$

if the events A, B, C , etc., are mutually independent. Actually, the mathematical definition of independence is the reverse of this statement, but the result of consequence is that independence of events and the multiplicative property of probabilities go together.

RANDOM VARIABLES

A random variable X is in simplest terms a variable which takes on values at random; it may be thought of as a function of the outcomes of some random experiment. The manner of specifying the probability with which different values are taken by the random variable is by the *probability distribution function* $F(x)$, which is defined by

$$F(x) = P(X \leq x) \quad (\text{H-2})$$

or by the *probability density function* $f(x)$, which is defined by

$$f(x) = \frac{dF(x)}{dx} \quad (\text{H-3})$$

The inverse of the defining relation for the probability density function is

$$F(x) = \int_{-\infty}^x f(u) du \quad (\text{H-4})$$

An evident characteristic of any probability distribution or density function is

$$F(\infty) = \int_{-\infty}^{\infty} f(u) du = 1 \quad (\text{H-5})$$

From the definition, the interpretation of $f(x)$ as the density of probability of the event that X takes a value in the vicinity of x is clear.

$$\begin{aligned} f(x) &= \lim_{dx \rightarrow 0} \frac{F(x + dx) - F(x)}{dx} \\ &= \lim_{dx \rightarrow 0} \frac{P(x < X \leq x + dx)}{dx} \end{aligned} \quad (\text{H-6})$$

This function is finite if the probability that X takes a value in the infinitesimal interval between x and $x + dx$ (the interval closed on the right) is an infinitesimal of order dx . This is usually true of random variables which take values over a continuous range. If, however, X takes a set of discrete values x_i with nonzero probabilities p_i , $f(x)$ is infinite at these values of x . This is accommodated by a set of delta functions weighted by the appropriate probabilities.

$$f(x) = \sum_i p_i \delta(x - x_i) \quad (\text{H-7})$$

A suitable definition of the *delta function*, $\delta(x)$, for the present purpose is a function which is zero everywhere except at $x = 0$, and infinite at that point in such a way that the integral of the function across the singularity is unity. An important property of the delta function which follows from this definition is

$$\int_{-\infty}^{\infty} G(x) \delta(x - x_0) dx = G(x_0) \quad (\text{H-8})$$

if $G(x)$ is a finite-valued function which is continuous at $x = x_0$.

A random variable may take values over a continuous range and, in addition, take a discrete set of values with nonzero probability. The resulting probability density function includes both a finite function of x and an additive set of probability-weighted delta functions; such a distribution is called *mixed*.

The simultaneous consideration of more than one random variable is often necessary or useful. In the case of two, the probability of the occurrence of pairs of values in a given range is prescribed by the *joint probability distribution function*.

$$F_2(x, y) = P(X \leq x \text{ and } Y \leq y) \quad (\text{H-9})$$

where X and Y are the two random variables under consideration. The corresponding *joint probability density function* is

$$f_2(x, y) = \frac{\partial^2 F_2(x, y)}{\partial x \partial y} \quad (\text{H-10})$$

It is clear that the individual probability distribution and density functions for X and Y can be derived from the joint distribution and density functions. For the distribution of X ,

$$F(x) = F_2(x, \infty) \quad (\text{H-11})$$

$$f(x) = \int_{-\infty}^{\infty} f_2(x, y) dy \quad (\text{H-12})$$

Corresponding relations give the distribution of Y . These concepts extend directly to the description of the joint characteristics of more than two random variables.

If X and Y are independent, the event $X \leq x$ is independent of the event $Y \leq y$; thus the probability for the joint occurrence of these events is the product of the probabilities for the individual events. Equation (H-9) then gives

$$\begin{aligned} F_2(x, y) &= P(X \leq x \text{ and } Y \leq y) \\ &= P(X \leq x)P(Y \leq y) \\ &= F_X(x)F_Y(y) \end{aligned} \quad (\text{H-13})$$

From Eq. (H-10) the joint probability density function is, then,

$$f_2(x, y) = f_X(x)f_Y(y) \quad (\text{H-14})$$

Expectations and statistics of random variables The *expectation* of a random variable is defined in words to be the sum of all values the random variable may take, each weighted by the probability with which the value is taken. For a random variable which takes values over a continuous range, this summation is done by integration. The probability, in the limit as $dx \rightarrow 0$, that X takes a value in the infinitesimal interval of width dx near x is given by Eq. (H-6) to be $f(x) dx$. Thus the expectation of X , which we denote by $\bar{X}$, is

$$\bar{X} = \int_{-\infty}^{\infty} xf(x) dx \quad (\text{H-15})$$

This is also called the *mean value* of X , or the mean of the distribution of X . This is a precisely defined number toward which the average of a number of observations of X tends, in the probabilistic sense, as the number of observations becomes large. Equation (H-15) is the analytic definition of the expectation, or mean, of a random variable. This expression is usable for random variables having a continuous, discrete, or mixed distribution if the set of discrete values which the random variable takes is represented by impulses in $f(x)$ according to Eq. (H-7).

It is of frequent importance to find the expectation of a function of a random variable. If Y is defined to be some function of the random variable X , say, $Y = g(X)$, then Y is itself a random variable with a distribution derivable from the distribution of X . The expectation of Y is defined by Eq. (H-15), where the probability density function for Y would be used in the integral. Fortunately, this procedure can be abbreviated. The expectation of any function of X can be calculated directly from the distribution of X by the integral

$$\overline{g(X)} = \int_{-\infty}^{\infty} g(x)f(x) dx \quad (\text{H-16})$$

An important statistical parameter descriptive of the distribution of X is its *mean-squared value*. Using Eq. (H-16), the expectation of the square of X is written

$$\overline{X^2} = \int_{-\infty}^{\infty} x^2 f(x) dx \quad (\text{H-17})$$

The *variance* of a random variable is the mean-squared deviation of the random variable from its mean; it is denoted by σ^2 .

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - \bar{X})^2 f(x) dx \\ &= \overline{X^2} - \bar{X}^2 \end{aligned} \quad (\text{H-18})$$

The square root of the variance, or σ , is called the *standard deviation* of the random variable.

Other functions whose expectations we shall wish to calculate are sums and products of random variables. It is easily shown that the expectation of the sum of random variables is equal to the sum of the expectations,

$$\overline{X_1 + X_2 + \cdots + X_n} = \bar{X}_1 + \bar{X}_2 + \cdots + \bar{X}_n \quad (\text{H-19})$$

whether or not the variables are independent, and that the expectation of the product of random variables is equal to the product of the expectations,

$$\overline{X_1 X_2 \cdots X_n} = \bar{X}_1 \bar{X}_2 \cdots \bar{X}_n \quad (\text{H-20})$$

if the variables are independent. It is also true that the variance of the sum of random variables is equal to the sum of the variances if the variables are independent.

A very important concept is that of statistical dependence between random variables. A partial indication of the degree to which one variable is related to another is given by the *covariance*, which is the expectation of the product of the deviations of two random variables from their means.

$$\begin{aligned} \overline{(X - \bar{X})(Y - \bar{Y})} &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy (x - \bar{X})(y - \bar{Y}) f_2(x, y) \\ &= \bar{X}\bar{Y} - \bar{X}\bar{Y} \end{aligned} \quad (\text{H-21})$$

This covariance, normalized by the standard deviations of X and Y , is called the *correlation coefficient*, and is denoted ρ .

$$\rho = \frac{\bar{X}\bar{Y} - \bar{X}\bar{Y}}{\sigma_X \sigma_Y} \quad (\text{H-22})$$

The correlation coefficient is a measure of the degree of linear dependence between X and Y . If X and Y are independent, ρ is zero; if Y is a linear function of X , ρ is ± 1 . If an attempt is made to approximate Y by some linear function of X , the minimum possible mean-squared error in the approximation is $\sigma_Y^2(1 - \rho^2)$. This provides another interpretation of ρ as a measure of the degree of linear dependence between random variables.

One additional function associated with the distribution of a random variable which should be introduced is the *characteristic function*. It is defined by

$$\begin{aligned} g(t) &= \overline{\exp(jtX)} \\ &= \int_{-\infty}^{\infty} \exp(jtx) f(x) dx \end{aligned} \quad (\text{H-23})$$

A property of the characteristic function which largely explains its value is that the characteristic function of a sum of independent random variables is

the product of the characteristic functions of the individual variables. If the characteristic function of a random variable is known, the probability density function can be determined from

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(t) \exp(-jtx) dt \quad (\text{H-24})$$

Notice that Eqs. (H-23) and (H-24) are in the form of a Fourier transform pair. Another useful relation is

$$\overline{X^n} = j^{-n} \left. \frac{d^n g(t)}{dt^n} \right|_{t=0} \quad (\text{H-25})$$

The uniform and normal probability distributions Two specific forms of probability distribution which are referred to in the text are the uniform distribution and the normal distribution. The *uniform distribution* is characterized by a uniform (constant) probability density over some finite interval. The magnitude of the density function in this interval is the reciprocal of the interval width as required to make the integral of the function unity. This function is pictured in Fig. H-1. The *normal probability density function*, shown in Fig. H-2, has the analytic form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x-m)^2}{2\sigma^2} \right] \quad (\text{H-26})$$

where the two parameters which define the distribution are m , the mean, and σ , the standard deviation. By calculating the characteristic function for a normally distributed random variable, one can immediately show that the distribution of the sum of independent normally distributed variables is also normal. Actually, this remarkable property of preservation of form of the distribution is true of the sum of normally distributed random variables whether they are independent or not. Even more remarkable is the fact that under certain circumstances the distribution of the sum of independent random variables, each having an *arbitrary* distribution, tends toward the normal distribution as the number of variables in the sum tends toward infinity. This statement, together with the conditions under which the result can be proved, is known as the *central limit theorem*. The conditions are rarely tested in practical situations, but the empirically observed fact is that a great many random variables—and especially those encountered by control-system engineers—display a distribution which closely approximates the normal. The reason for the common occurrence of normally distributed random variables is certainly stated in the central limit theorem.

Reference is made in the text to two random variables which possess a

bivariate normal distribution. The form of the joint probability density function for such zero-mean variables is

$$f_2(x_1, x_2) = \frac{1}{2\pi\sqrt{m_{11}m_{22} - m_{12}^2}} \exp \left[-\frac{m_{22}x_1^2 - 2m_{12}x_1x_2 + m_{11}x_2^2}{2(m_{11}m_{22} - m_{12}^2)} \right]$$

where $m_{ij} = \overline{X_i X_j}$.

This can also be written in terms of statistical parameters previously defined as

$$f_2(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp \left[-\frac{\frac{x_1^2}{\sigma_1^2} - 2\rho\frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2} + \frac{x_2^2}{\sigma_2^2}}{2(1 - \rho^2)} \right] \quad (\text{H-27})$$

RANDOM PROCESSES

A random process may be thought of as a collection, or *ensemble*, of functions of time, any one of which might be observed on any trial of an experiment. The ensemble may include a finite number, a countable infinity, or a noncountable infinity of such functions. We shall denote the ensemble of functions by $\{x(t)\}$, and any observed member of the ensemble by $x(t)$. The value of the observed member of the ensemble at a particular time, say, t_1 , as shown in Fig. H-3, is a random variable; on repeated trials of the experiment, $x(t_1)$ takes different values at random. The probability that $x(t_1)$ takes values in a certain range is given by the probability distribution function, as it is for any random variable. In this case we show explicitly in the notation the dependence on the time of observation.

$$F(x_1, t_1) = P[x(t_1) \leq x_1] \quad (\text{H-28})$$

The corresponding probability density function is

$$f(x_1, t_1) = \frac{dF(x_1, t_1)}{dx_1} \quad (\text{H-29})$$

These functions suffice to define, in a probabilistic sense, the range of amplitudes which the random process displays. To gain a sense of how quickly varying the members of the ensemble are likely to be, one has to observe the same member function at more than one time. The probability for the occurrence of a pair of values in certain ranges is given by the second-order joint probability distribution function

$$F_2(x_1, t_1; x_2, t_2) = P[x(t_1) \leq x_1 \text{ and } x(t_2) \leq x_2] \quad (\text{H-30})$$

and the corresponding joint probability density function

$$f_2(x_1, t_1; x_2, t_2) = \frac{\partial^2 F_2(x_1, t_1; x_2, t_2)}{\partial x_1 \partial x_2} \quad (\text{H-31})$$

Higher-ordered joint distribution and density functions can be defined following this pattern, but only rarely does one attempt to deal with more than the second-order statistics of random processes.

If two random processes are under consideration, the simplest distribution and density functions which give some indication of their joint statistical characteristics are the second-order functions

$$F_2(x, t_1; y, t_2) = P[x(t_1) \leq x \text{ and } y(t_2) \leq y] \quad (\text{H-32})$$

$$f_2(x, t_1; y, t_2) = \frac{\partial^2 F_2(x, t_1; y, t_2)}{\partial x \partial y} \quad (\text{H-33})$$

Actually, the characterization of random processes, in practice, is usually limited to even less information than that given by the second-order distribution or density functions. Only the first moments of these distributions are commonly measured. These moments are called auto- and cross-correlation functions. The *autocorrelation function* is defined as

$$\varphi_{xx}(t_1, t_2) = \overline{x(t_1)x(t_2)} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 x_1 x_2 f_2(x_1, t_1; x_2, t_2) \quad (\text{H-34})$$

and the *cross-correlation function* as

$$\varphi_{xy}(t_1, t_2) = \overline{x(t_1)y(t_2)} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy xy f_2(x, t_1; y, t_2) \quad (\text{H-35})$$

In the case where $\overline{x(t_1)}$, $\overline{x(t_2)}$, and $\overline{y(t_2)}$ are all zero, these correlation functions are the covariances of the indicated random variables. If they are then normalized by the corresponding standard deviations, according to Eq. (H-22), they become correlation coefficients which measure on a scale from -1 to $+1$ the degree of linear dependence between the variables.

A *stationary random process* is one whose statistical properties are invariant in time. This implies that the first probability density function for the process, $f(x_1, t_1)$, is independent of the time of observation t_1 . Then all the moments of this distribution, such as $\overline{x(t_1)}$ and $\overline{x(t_1)^2}$, are also independent of time; they are constants. The second probability density function is not in this case dependent on the absolute times of observation, t_1 and t_2 , but still depends on the difference between them. So if t_2 is written as

$$t_2 = t_1 + \tau \quad (\text{H-36})$$

$f_2(x_1, t_1; x_2, t_2)$ becomes $f_2(x_1, t_1; x_2, t_1 + \tau)$, which is independent of t_1 , but still a function of τ . The correlation functions are then functions only of the single variable τ .

$$\varphi_{xx}(\tau) = \overline{x(t_1)x(t_1 + \tau)} \quad (\text{H-37})$$

$$\varphi_{xy}(\tau) = \overline{x(t_1)y(t_1 + \tau)} \quad (\text{H-38})$$

Both of these are independent of t_1 if the random processes are stationary. We note the following properties of these correlation functions:

$$\varphi_{xx}(-\tau) = \varphi_{xx}(\tau) \quad (\text{H-39})$$

$$\varphi_{xy}(-\tau) = \varphi_{yx}(\tau) \quad (\text{H-40})$$

One further concept associated with stationary random processes is the *ergodic hypothesis*. This hypothesis claims that any statistic calculated by averaging over all members of an ergodic ensemble at a fixed time can also be calculated by averaging over all time on a single representative member of the ensemble. The key to this notion is the word “representative.” If a particular member of the ensemble is to be statistically representative of all, it must display at various points in time the full range of amplitude, rate of change of amplitude, etc., which are to be found among all the members of the ensemble. A classic example of a stationary ensemble which is not ergodic is the ensemble of constant functions. The failing in this case is that no member of the ensemble is representative of all. In practice, almost all empirical results for stationary processes are derived from tests on a single function under the assumption that the ergodic hypothesis holds. In this case the common statistics associated with a random process are written

$$\bar{x} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt \quad (\text{H-41})$$

$$\overline{x^2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)^2 dt \quad (\text{H-42})$$

$$\varphi_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t + \tau) dt \quad (\text{H-43})$$

$$\varphi_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)y(t + \tau) dt \quad (\text{H-44})$$

An example of a stationary ergodic random process is the ensemble of sinusoids of given amplitude and frequency with a uniform distribution of phase. The member functions of this ensemble are all of the form

$$x(t) = A \sin(\omega t + \theta) \quad (\text{H-45})$$

where θ is a random variable having the uniform distribution over the interval $(0, 2\pi)$ radians. Any average taken over the members of this ensemble at any fixed time would find all phase angles represented with equal probability density. But the same is true of an average over all time on any one member. For this process, then, all members of the ensemble qualify as "representative." Note that any distribution of the phase angle θ other than the uniform distribution over an integral number of cycles would define a nonstationary process.

Another random process which plays a central role in the text is the *gaussian process*, which is characterized by the property that its joint probability distribution functions of all orders are multidimensional normal distributions. For a gaussian process, then, the distribution of $x(t)$ for any t is the normal distribution, for which the density function is expressed by Eq. (H-26); the joint distribution of $x(t_1)$ and $x(t_2)$ for any t_1 and t_2 is the bivariate normal distribution of Eq. (H-27), and so on for the higher-ordered joint distributions. The n -dimensional normal distribution for zero-mean variables is specified by the elements of the n th-order covariance matrix, that is, by the $m_{ij} = \overline{X_i X_j}$ for $i, j = 1, 2, \dots, n$. But in this case

$$\begin{aligned} X_i &= x(t_i) \\ \text{and } m_{ij} &= \overline{x(t_i)x(t_j)} \\ &= \varphi_{xx}(t_i, t_j) \end{aligned}$$

Thus all the statistics of a gaussian process are defined by the autocorrelation function for the process. This property is clearly a great boon to analytic operations.

LINEAR SYSTEMS

The input-output relation for a linear system may be written

$$y(t) = \int_{-\infty}^t x(\tau) w(t, \tau) d\tau \quad (\text{H-46})$$

where $x(t)$ = input function

$y(t)$ = output

$w(t, \tau)$ = system *weighting function*, the response
at time t to a unit impulse input at time τ .

Using this relation, the statistics of the output process can be written in terms of those of the input.

$$\overline{y(t)} = \int_{-\infty}^t \overline{x(\tau)} w(t, \tau) d\tau \quad (\text{H-47})$$

$$\overline{y(t)^2} = \int_{-\infty}^t d\tau_1 w(t, \tau_1) \int_{-\infty}^t d\tau_2 w(t, \tau_2) \varphi_{xx}(\tau_1, \tau_2) \quad (\text{H-48})$$

$$\varphi_{yy}(t_1, t_2) = \int_{-\infty}^{t_1} d\tau_1 w(t_1, \tau_1) \int_{-\infty}^{t_2} d\tau_2 w(t_2, \tau_2) \varphi_{xx}(\tau_1, \tau_2) \quad (\text{H-49})$$

$$\varphi_{xy}(t_1, t_2) = \int_{-\infty}^{t_2} w(t_2, \tau) \varphi_{xx}(t_1, \tau) d\tau \quad (\text{H-50})$$

If the input process is stationary and the system time-invariant, the output process is also stationary in the steady state. These expressions then reduce to

$$y(t) = \int_0^\infty w(\tau) x(t - \tau) d\tau \quad (\text{H-51})$$

$$\bar{y} = \bar{x} \int_0^\infty w(t) dt \quad (\text{H-52})$$

$$\overline{y^2} = \int_0^\infty d\tau_1 w(\tau_1) \int_0^\infty d\tau_2 w(\tau_2) \varphi_{xx}(\tau_1 - \tau_2) \quad (\text{H-53})$$

$$\varphi_{yy}(\tau) = \int_0^\infty d\tau_1 w(\tau_1) \int_0^\infty d\tau_2 w(\tau_2) \varphi_{xx}(\tau + \tau_1 - \tau_2) \quad (\text{H-54})$$

$$\varphi_{xy}(\tau) = \int_0^\infty w(\tau_1) \varphi_{xx}(\tau - \tau_1) d\tau_1 \quad (\text{H-55})$$

Analytic operations on linear invariant systems are facilitated by the use of integral transforms which transform the convolution input-output relation of Eq. (H-51) into the algebraic operation of multiplication. Since members of stationary random ensembles must necessarily be visualized as existing for all negative and positive time, the two-sided Fourier transform is the appropriate transformation to employ in this instance. The Fourier transforms of the correlation functions defined above then appear quite naturally in analysis. The Fourier transform of the autocorrelation function

$$\Phi_{xx}(\omega) = \int_{-\infty}^\infty \varphi_{xx}(\tau) \exp(-j\omega\tau) d\tau \quad (\text{H-56})$$

is called the *power spectral density function*, or power density spectrum of the random process $\{x(t)\}$. The term "power" is here used in a generalized sense, indicating the expected squared value of the members of the ensemble. $\Phi_{xx}(\omega)$ is indeed the spectral distribution of power density for $\{x(t)\}$ in that integration of $\Phi_{xx}(\omega)$ over frequencies in the band from ω_1 to ω_2 yields the mean-squared value of the process which consists only of those harmonic components of $\{x(t)\}$ that lie between ω_1 and ω_2 . In particular, the mean-squared value of $\{x(t)\}$ itself is given by integration of the power density

spectrum for the random process over the full range of ω . This last result is seen as a specialization of the inverse transform relation corresponding to Eq. (H-56).

$$\varphi_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(\omega) \exp(j\omega\tau) d\omega \quad (\text{H-57})$$

$$\overline{x^2} = \varphi_{xx}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(\omega) d\omega \quad (\text{H-58})$$

The input-output relation for power spectral density functions is derived by calculating the Fourier transform of the autocorrelation function of the output of a linear invariant system as expressed by Eq. (H-54).

$$\begin{aligned} \Phi_{yy}(\omega) &= W(j\omega)W(-j\omega)\Phi_{xx}(\omega) \\ &= |W(j\omega)|^2 \Phi_{xx}(\omega) \end{aligned} \quad (\text{H-59})$$

where $W(j\omega)$ is the steady-state *sinusoidal response function* for the system, which is also the Fourier transform of the system weighting function.

$$W(j\omega) = \int_{-\infty}^{\infty} w(t) \exp(-j\omega t) dt \quad (\text{H-60})$$

A particularly simple form for the power density spectrum is a constant, $\Phi_{xx}(\omega) = \Phi_0$. This implies that power density is distributed uniformly over all frequency components in the full infinite range. By analogy with the corresponding situation in the case of white light, such a random process, usually a noise, is called *white noise*. The autocorrelation function for white noise is a delta function.

$$\begin{aligned} \varphi_{nn}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_0 \exp(j\omega\tau) d\omega \\ &= \Phi_0 \delta(\tau) \end{aligned} \quad (\text{H-61})$$

The mean-squared value of white noise, $\varphi_{nn}(0)$, is infinite, and so the process is not physically realizable. However, it does serve as a very useful approximation to situations in which the noise is wideband compared with the bandwidth of the system, and the concept is useful in analytic operations.

The Fourier transform of the cross-correlation function is called the *cross power spectral density function*.

$$\Phi_{xy}(\omega) = \int_{-\infty}^{\infty} \varphi_{xy}(\tau) \exp(-j\omega\tau) d\tau \quad (\text{H-62})$$

If $x(t)$ is the input to, and $y(t)$ the output from, a linear invariant system, so that $\varphi_{xy}(\tau)$ is given by Eq. (H-55), then the input-output cross power density spectrum is

$$\Phi_{xy}(\omega) = W(j\omega)\Phi_{xx}(\omega) \quad (\text{H-63})$$

The calculation of the distribution of amplitudes of the random processes which appear at various points in linear systems is in general a most complicated problem. The only case in which a simple result is known is that of a gaussian process applied to a linear system; in this case it can be proved that the processes appearing at all points in the system are also gaussian. The credibility of this property is perhaps indicated by noting that the input-output relation for a linear system [Eq. (H-46)] is the limit of a sum of the form

$$y(t) = \lim_{\substack{\Delta\tau \rightarrow 0 \\ N \rightarrow \infty \\ \tau_N = t}} \sum_{i=-\infty}^N x(\tau_i) w(t, \tau_i) \Delta\tau_i \quad (\text{H-64})$$

We have already noted that the sum of normally distributed random variables is normally distributed, and since any constant times a normally distributed variable is also normal, we may conclude that any linear combination of normally distributed random variables is normally distributed. Equation (H-64) expresses the output of a linear system as the limit of a linear combination of past values of the input. Thus, if the input at all past times has a normal distribution, the output at any time must also be normal. This property also holds for the higher-ordered joint distributions, with the result that if $x(t)$ is a gaussian process, so is $y(t)$.

Of even greater consequence is the empirically observed fact that *non-gaussian inputs tend to become more nearly gaussian as a result of linear filtering*. If the input were nongaussian white noise, one could refer to Eq. (H-64) and invoke the central limit theorem to argue that, as the filtering bandwidth is decreased, the number of terms in the sum for $y(t)$ which make a significant contribution increases, and thus the distribution of $y(t)$ should approach the normal, regardless of the distribution of the $x(\tau_i)$. In fact, this tendency is observed for nonwhite inputs as well; so the gaussian random process has the singular position of being that process toward which many others tend as a result of linear filtering. The most evident exception to this rule is a random ensemble in which every member contains a periodic component of the same period. Low-pass linear filtering tends to reduce these periodic signals to their fundamental sinusoidal components, and a sinusoid does not display the normal distribution of amplitudes. But if every member of a random ensemble has a periodic component of the same period, the process contains nonzero power at the fundamental frequency of these periodic components, and perhaps at some of the harmonics of this frequency as well. Nonzero power at any discrete frequency implies infinite power density at that frequency. Thus it might be said that if a random process has a finite power density spectrum, it may be expected to approach a gaussian process as a result of low-pass linear filtering. Unfortunately, it does not seem possible to phrase this general statement in quantitative terms.

INSTANTANEOUS NONLINEARITIES

The second-order statistics of the output of a static single-valued nonlinearity depend on the form of the nonlinearity

$$y = y(x) \quad (\text{H-65})$$

and the second-order joint probability density function for the input random process. For example, the autocorrelation function for the output process is

$$\varphi_{yy}(\tau) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 y(x_1) y(x_2) f_2(x_1, t_1; x_2, t_1 + \tau) \quad (\text{H-66})$$

which is independent of t_1 if $\{x(t)\}$ is stationary. If interest is centered on closed-loop systems which contain a nonlinear device, it is rarely possible, as a practical matter, to determine the second-order probability density function for the input to the nonlinearity, unless one can argue, on the basis of linear filtering, that the process should be approximately gaussian. For a stationary gaussian process, $f_2(x_1, t_1; x_2, t_1 + \tau)$ is given by Eq. (H-27) with

$$\sigma_1^2 = \overline{x(t_1)x(t_1)} = \varphi_{xx}(0) = \sigma^2 \quad (\text{H-67})$$

since we are considering zero-mean variables. Also,

$$\sigma_2^2 = \overline{x(t_1 + \tau)x(t_1 + \tau)} = \sigma^2 \quad (\text{H-68})$$

$$\rho = \frac{\overline{x(t_1)x(t_1 + \tau)}}{\sigma_1\sigma_2} = \frac{1}{\sigma^2} \varphi_{xx}(\tau) = \rho_{xx}(\tau) \quad (\text{H-69})$$

Equation (H-27) becomes

$$f_2(x_1, t_1; x_2, t_1 + \tau) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1-\rho^2)} \right] \quad (\text{H-70})$$

Thus, for a gaussian input, Eq. (H-66) is written

$$\begin{aligned} \varphi_{yy}(\tau) &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 y(x_1) y(x_2) \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \\ &\quad \times \exp \left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1-\rho^2)} \right] \\ &= \int_{-\infty}^{\infty} du_1 \int_{-\infty}^{\infty} du_2 y(\sigma u_1) y(\sigma u_2) \frac{1}{2\pi\sqrt{1-\rho^2}} \\ &\quad \times \exp \left[-\frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{2(1-\rho^2)} \right] \quad (\text{H-71}) \end{aligned}$$

The evaluation of this double integral can be systematized through the expansion of the gaussian joint probability density function into a double

series of Hermite polynomials. These functions are used in varying forms; perhaps the most convenient for the present purpose is

$$H_k(u) = (-1)^k \exp\left(\frac{u^2}{2}\right) \frac{d^k}{du^k} \left[\exp\left(-\frac{u^2}{2}\right) \right] \quad (\text{H-72})$$

The first few of these functions are

$$H_0(u) = 1 \quad (\text{H-73a})$$

$$H_1(u) = u \quad (\text{H-73b})$$

$$H_2(u) = u^2 - 1 \quad (\text{H-73c})$$

$$H_3(u) = u^3 - 3u \quad (\text{H-73d})$$

and in general,

$$H_{k+1}(u) = uH_k(u) - kH_{k-1}(u) \quad (\text{H-74})$$

These functions form a complete set and are orthogonal over the doubly infinite interval with respect to the weighting function $\exp(-u^2/2)$. The orthogonality conditions are

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H_k(u) H_j(u) \exp\left(-\frac{u^2}{2}\right) du = \begin{cases} k! & j = k \\ 0 & j \neq k \end{cases} \quad (\text{H-75})$$

Expansion of the gaussian probability density function in terms of these Hermite polynomials gives (Ref. 1)

$$\begin{aligned} \frac{1}{\sqrt{1-\rho^2}} \exp\left[-\frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{2(1-\rho^2)}\right] \\ = \exp\left(-\frac{u_1^2 + u_2^2}{2}\right) \sum_{k=0}^{\infty} H_k(u_1) H_k(u_2) \frac{\rho^k}{k!} \end{aligned} \quad (\text{H-76})$$

With this expansion, Eq. (H-71) becomes

$$\begin{aligned} \varphi_{vv}(\tau) &= \frac{1}{2\pi} \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \int_{-\infty}^{\infty} du_1 \int_{-\infty}^{\infty} du_2 y(\sigma u_1) y(\sigma u_2) \\ &\quad \times \exp\left(-\frac{u_1^2 + u_2^2}{2}\right) H_k(u_1) H_k(u_2) \\ &= \sum_{k=0}^{\infty} a_k^2 \rho_{xx}^k(\tau) \end{aligned} \quad (\text{H-77})$$

$$\text{where} \quad a_k = \frac{1}{\sqrt{2\pi k!}} \int_{-\infty}^{\infty} y(\sigma u) \exp\left(-\frac{u^2}{2}\right) H_k(u) du \quad (\text{H-78})$$

This expresses the output autocorrelation function as a power series in the normalized input autocorrelation function. Note that $H_k(u)$ is odd for k odd and is even for k even; thus, in the common case of an odd nonlinearity, $a_k = 0$ for k even.

It is clear that the output of the nonlinearity contains power at higher frequencies than the input. The normalized autocorrelation function $\rho_{xx}(\tau)$ has a magnitude less than or equal to unity everywhere; so the powers of this function appearing in Eq. (H-77) fall off to zero faster than the function itself. For example, if the input process has the familiar exponential autocorrelation function, $\rho_{xx}^3(\tau)$ will decay three times faster, as indicated in Fig. H-4. But this contribution to the output autocorrelation function has a power spectral density function which is three times wider than that of the input process. The higher-ordered terms in the expansion of the output autocorrelation function make still broader band contributions to the output power spectral density function. This is in a way analogous to the response of a nonlinear device to a sinusoidal input, which consists of a fundamental frequency component plus higher-frequency harmonics.

Other statistics of the output of the nonlinearity can also be expressed in terms of these a_k . The mean of the output is

$$\begin{aligned}\overline{y(t)} &= \int_{-\infty}^{\infty} y(x) f(x, t) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y(\sigma u) \exp\left(-\frac{u^2}{2}\right) du \\ &= a_0\end{aligned}\tag{H-79}$$

in the case of an unbiased stationary gaussian input. Also, in this case, the input-output cross-correlation function is

$$\begin{aligned}\varphi_{xy}(\tau) &= \frac{1}{2\pi} \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \int_{-\infty}^{\infty} du_1 \int_{-\infty}^{\infty} du_2 \sigma u_1 y(\sigma u_2) \exp\left(-\frac{u_1^2 + u_2^2}{2}\right) H_k(u_1) H_k(u_2) \\ &= \sum_{k=0}^{\infty} a_k \rho^k \left[\frac{1}{\sqrt{2\pi k!}} \int_{-\infty}^{\infty} \sigma u_1 \exp\left(-\frac{u_1^2}{2}\right) H_k(u_1) du_1 \right] \\ &= \sigma a_1 \rho_{xx}(\tau) \\ &= \frac{a_1}{\sigma} \varphi_{xx}(\tau)\end{aligned}\tag{H-80}$$

using Eq. (H-73b) and the orthogonality conditions of Eq. (H-75). The input-output cross-correlation function for a static single-valued nonlinearity with an unbiased gaussian input is thus found to be just a constant times the input autocorrelation function, as it is for a static linear gain. In the linear case, the constant is the gain, whereas in the nonlinear case, the constant depends both on the nonlinearity and on the rms value of the input.

REFERENCES

1. Cramer, H.: "Mathematical Methods of Statistics," Princeton University Press, Princeton, N.J., 1946, p. 133.
2. Davenport, W. B., Jr., and W. L. Root: "An Introduction to Random Signals and Noise," McGraw-Hill Book Company, New York, 1958.
3. Laning, J. H., Jr., and R. H. Battin: "Random Processes in Automatic Control," McGraw-Hill Book Company, New York, 1956.
4. Lee, Y. W.: "Statistical Theory of Communication," John Wiley & Sons, Inc., New York, 1960.

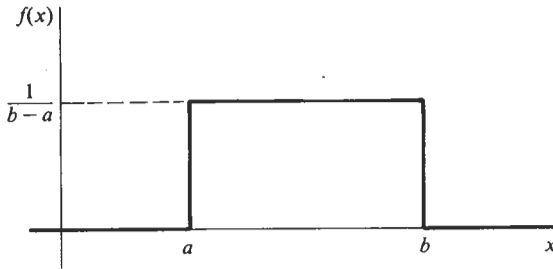


Figure H-1 The uniform probability density function.

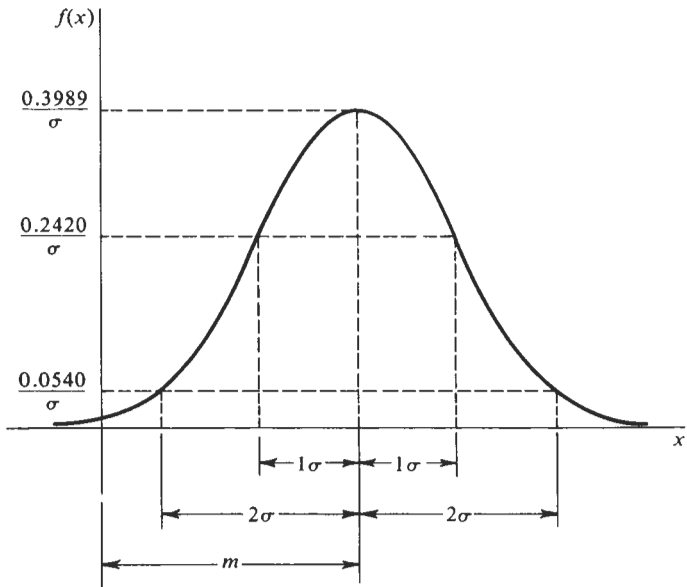


Figure H-2 The normal probability density function.

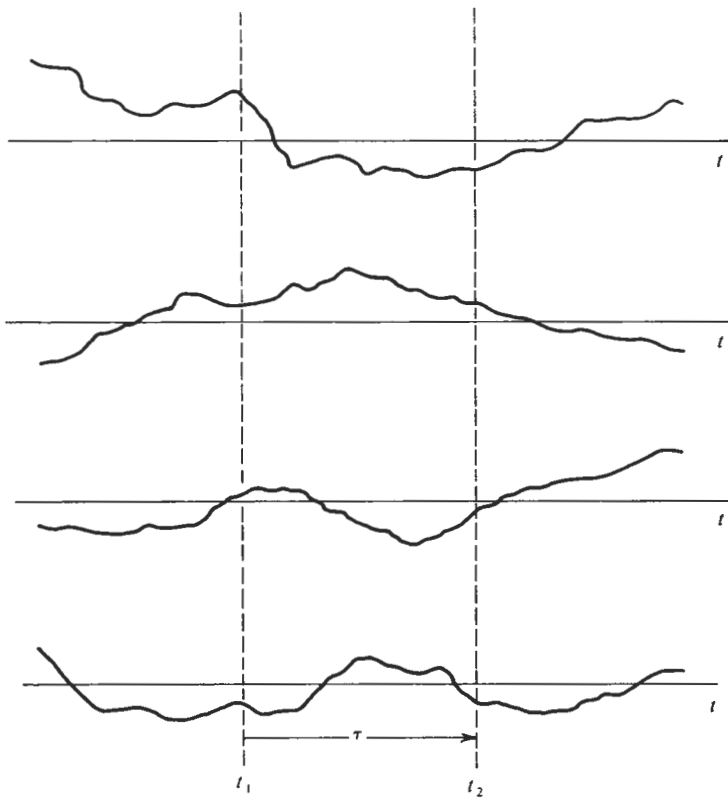
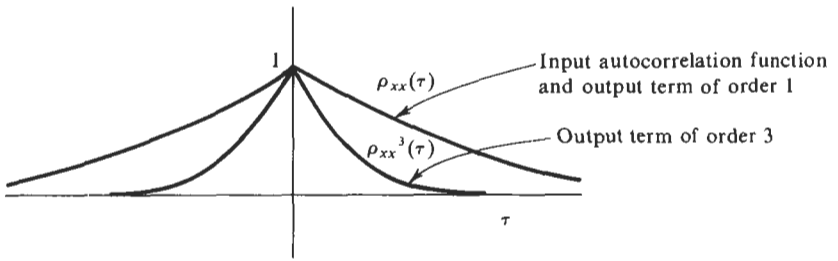
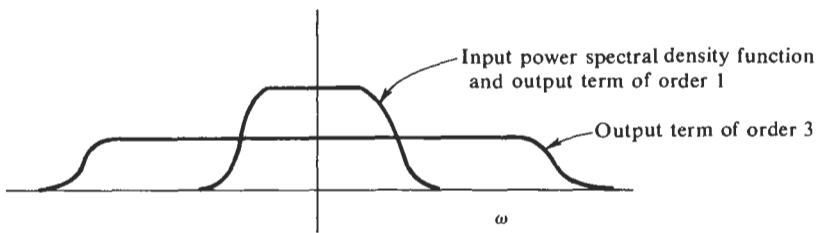


Figure H-3 Members of the ensemble $\{x(t)\}$.



(a) Autocorrelation functions



(b) Power spectral density functions

Figure H-4 Characteristics of the nonlinearity input and output random processes.

INDEX

- Absolute-value nonlinearity, 351
Accuracy (*see* Describing functions)
Adaptive control system, 323, 327–332
 missile roll control, 332–341
Aizerman, M. A., 162*n.*, 199, 206
Amplitude-phase plot (*see* Gain-phase plot)
Amplitude-ratio condition, 308, 319, 336, 338
Amplitude-ratio-decibel conversion
 table, 515–517
Amplitude-regulating loop (*see* Limit cycle)
Amplitude transfer function, 233–237
Amsler, B. E., 259
Approximator for a nonlinearity, 19
Artificial dither (*see* Dither)
Asymmetric nonlinearity, 300
 in limit cycle calculations, 340–345
 piecewise linear, 311
Asymptotic methods, 43
Atherton, D. P., 19, 30, 108, 368, 387, 396
Attenuation of harmonics, 172
Autocorrelation function, 636
Autoresonance hypothesis, 162*n.*
Averaging, method of, 45
Axelby, G. S., 368
Backlash, 67
 friction-controlled, 68–69
 inertia-controlled, 68, 70–71
 viscous friction plus inertia load, 77–79
Bang-bang nonlinearity (*see* Ideal relay nonlinearity)
Barker, H. A., 222
Barrett, J. F., 368
Bass, R. W., 181–183
Battin, R. H., 629
Bautin, N., 125

- Bennett, W. R., 255
 Bergen, A. R., 188, 498
 Bessel function, 65, 266, 378–380
 Biased-output ideal-relay, 343–344
 Biased subharmonic mode, 270
 Bode plot, 117
 Bogoliubov, N., 42, 45, 180–181, 220–222
 Bohn, E. V., 198
 Bonenn, Z., 132, 280
 Booton, R. C., Jr., 19, 368, 369, 399
 Boyer, R. C., 354*n*.
 Brockett, R. W., 185*n*.
 Bulgakov, B. V., 181
 Burrington, R. S., 375

 Cahen, G., 121
 Calculation of describing functions
 approximate, 89–96, 385–392
 DF, 54–85
 DIDF, 305–317
 EIDF, 454–456
 RIDF, 370–385
 sampled describing function, 475, 484–492
 TSIDF, 253–267
 Caron, J. Y., 367
 Central limit theorem, 634
 Characteristic equation, 113, 155, 323, 496
 Chatter (*see* Limit cycle)
 Chebyshev polynomial, 108
 Chen, K., 438, 442–449, 452–453
 Chow, C. K., 474, 490
 Chuang, K., 368
 Clauser, F. H., 222
 Clegg integrator, 79, 126–127, 151, 315–317
 Coales, J. F., 368
 Compensation techniques
 linear compensation, 145–148, 331–332, 407–411, 470
 nonlinear model feedback, 148–151
 other nonlinear networks, 151–154, 411–414
 Complete elliptic integrals, defined, 258
 Conditionally stable system, 286, 325
 Contacter (*see* Ideal relay nonlinearity)
 Control system, feedback, 3
 Corrected-conventional describing function, 85, 109

 Cosgriff, R. I., 431
 Coulomb friction, rotary, 105
 Covariance, 633
 Cross-correlation equivalence, 22, 24
 Cross-correlation function, 636
 Cross power spectral density function, 640
 Cubic nonlinearity, 65, 116, 136, 243, 253, 265–269, 296, 315, 357, 372, 402, 491
 Culver, W. J., 367
 Cunningham, W. J., 205

 Da-Chuan, S., 506
 Damping ratio, 47
 Davenport, W. B., Jr., 629
 D'Azzo, J. J., 85
 Dead zone, 4–6
 (*See also* Relay with dead zone)
 Deadband gain (*see* Dead zone)
 Decibel (*see* Amplitude-ratio-decibel conversion table)
 Delay time, 447
 Delfeld, F. R., 506
 Describing functions
 applicability, 15, 16
 limitations, 16
 model of the nonlinearity, 19
 multiple inputs of the same form, 33–34
 point of view, 14
 for specific signal forms, 25–33
 for small signals, 34–37, 303–304, 426–427
 a unified theory of, 18–37
 (*See also* specific types)
 DF (*see* Sinusoidal-input describing function)
 DIDF (*see* Dual-input describing function)
 di Tada, E. G., 97
 Dither, 339, 345–357
 linearizing effect of, 349
 random, 415–420
 sinusoidal, 348
 square-wave, 348
 triangular-wave, 346–348
 Dixon, M. V., 490
 Double Fourier series expansion, 255–260
 Douce, J. L., 253, 271, 272, 411

- Dominant behavioral modes, 447
- Drive capability, choice of, 327
- Dual-input describing function, 298–364
 - tabulation of, 539–553
- Dual-mode nonlinear integrator (*see* Clegg integrator)
- Dual-mode operation, 399
- Dutilh, J., 49

- Earth satellite, orbit period of, 298–299
- EIDF (*see* Exponential-input describing function)
- Elgerd, O. I., 267
- Elliptic describing function, 85
- Equivalent nonlinear element, 30, 52, 309, 348, 354, 358, 387, 427–429
- Error, mean-squared approximation, 20–23
- Error coefficients, 447
- Exact determination of limit cycles
 - relay control systems, 185–199
 - sampled nonlinear systems, 497–505
- Exponential-input describing function, 439, 454–459
 - formulation of, 454–455
 - tabulation of, 456

- $f(\gamma)$, 59, 317
- Feedback control system, 3
- Filter hypothesis, 111, 162, 183
 - analytical justification for, 625–628
 - first harmonic, 627
 - higher harmonics, 628
- Fitts, R. E., 296
- Flip-flop, DF for, 109
- Forced response
 - of limit cycling systems, 317–329
 - steady-state forced errors, 326
 - transient response behavior, 318–321
 - of randomly disturbed systems, 428–429
- Fourier series, 49, 255–260
- Fourier transform of nonlinearity, 260, 378–380
- Franklin, G. F., 462
- Frequency-dependent describing functions
 - DF, 75–84
 - DIDF, 316
- Frequency-ratio condition, 308

- Frequency response
 - calculation of, 125–143
 - counterexamples, 271–274
 - effect of additional random input, 418–420

- $g(\gamma)$, 313, 317
- Gain-changing nonlinearity, 55–60
- Gain-phase plot, 117–118, 121, 124, 146–148, 275, 328, 469
- Gamma function, 64*n*.
- Gaussian process, 638
- Gelb, A., 222, 298
- Generating solution, 181
- Gibson, J. E., 85, 97, 117*n*., 260, 272
- Gille, J. C., 60*n*.
- Goldfarb, L. C., 49
- Gorozdos, R. E., 259
- Gould, L. A., 409
- Graham, D., 60*n*.
- Grensted, P. E. W., 222
- Gronner, A. D., 88
- Guillemin, E. A., 195
- Gusev, M. I., 30

- Half-wave linear rectifier, 288
- Hamel, B., 198
- Harmonic balance, 299
- Harmonic linearization, 47, 49, 222, 366
- Harmonic nonlinearity, 65, 166, 266, 283, 288, 377
- Harmonically related input sinusoids, 254
- Harmonics
 - generation of, 71–75, 81, 260
 - importance of, 163–174
- Hatanaka, H., 134
- Hermite polynomials, 371–372, 643
- Holzman, J. M., 183
- Hysteresis, 4–5, 65, 170, 187
 - rectangular hysteresis, 67, 188, 195, 197, 309, 465–472
- Hysteretic relay (*see* Hysteresis)

- Ideal relay nonlinearity, 57, 176, 187, 224–229, 257, 260, 263, 266, 273, 306–309, 320, 343, 350, 403, 457, 477–484
- In-phase gain, 53, 302
- Incommensurate periods, 251

- Incremental-input describing function,
253, 276–286, 291, 303, 308, 311,
315, 344, 360
for nonsynchronous inputs, 278
for synchronous inputs, 276–278
Inherent adaptivity (*see* Adaptive control
system)
Integral representation of nonlinearity,
260, 378–380
Intentionally inexact model feedback,
150–151
Inverted pendulum control system, 363
- Jaffe, R. C., 267
Johnson, E. C., 181–182
Jump resonance, 132
effect of random dither, 419
treatment using incremental-input de-
scribing function, 278–280
Jump resonance curve, critical, 133–138
Jury, E. I., 462, 498, 506
- Kaiser, J. F., 409
Kalb, R. M., 255
Kalman filter, 411
Katkovnik, V. Y., 370
Kazakov, I. Y., 367, 368
Kazda, L. F., 368
King, R. E., 271, 411
Klotter, K., 84
Kochenburger, R. J., 49
Krylov, N., 42, 45, 220–222
Kuo, B. C., 462, 475, 490
Kwatny, H. G., 416
- Laning, J. H., Jr., 629
Laser gyroscope, 364
Lee, Y. W., 629
Leonhard, A., 291*n*.
Lepschy, A., 491
Levadi, V. S., 431
Li, Y. T., 298
Limit cycle, 7, 8, 111
amplitude regulation, 231, 338–340
build-up, 230
determination of, 111–120, 189–190,
323, 335–336, 340–344, 415–416,
468–470
- Limit cycle
dynamics, 231–241, 280–282
excitation, 8, 272–273, 415
frequency regulation, 241
quenching, 8, 273–274, 415
stability, 120–125, 241–246, 274–276,
492–497
Limit-cycling control systems, 321–341
Limit-cycling minor loop, 335
Limiter, 57, 60, 272, 455, 494–497
Linear system approximations, 443–448
for transient response, 446–448
Linearization
harmonic, 366
small-signal, 12
statistical, 366
Livesley, R. K., 253
Loeb, J. M., 121, 298
Lozier, J. C., 298, 327
Lubbock, J. K., 222
Lur'e, A. I., 199
Lyapunov's direct method, 10
- MacColl, L. A., 298
McRuer, D., 60*n*.
Magnus, K., 60*n*., 168*n*.
Mahalanabis, A. K., 431
Massachusetts Institute of Technology,
326
Mathieu equation, 291–295
Mean squared value, 632
Mean value, 632
Merchav, S. J., 368
Mesarovic, M. D., 367
Meserve, W. E., 499
Minorsky, N., 354
Mitropolsky, Y. A., 180–181
Modified nonlinearity (*see* Equivalent
nonlinear element)
Morosanov, I. S., 370
Multiple limit cycles, 274–276
Multiple nonlinearity systems
limit cycles in, 156–161
treatment of, 154–161
Murphy, G. J., 506
- Naslin, P., 350, 351*n*., 361
Nath, A. K., 431
Nease, R. F., 493, 506

- Neumann factor, use of, 261
 Newton, G. C., 409
 Nikiforuk, P. N., 405, 419
 Nishimura, T., 498, 506
 Non-harmonically related input sinusoids, 254
 Nonlinearity
 approximation for, 19
 dynamic, 4, 5, 83
 explicit, 4
 hard, 65
 implicit, 4, 83
 reduction of form, 5, 6
 soft, 65
 static, multiple-valued, 4
 single-valued, 4, 29, 31–33, 642
 Nonlinearity test inputs
 bias input, 15, 25–26
 exponential input, 455
 gaussian input, 15, 31–33, 370
 multiple inputs, 33–37, 250–263, 301, 370–384
 sinusoidal input, 15, 26–30, 49, 485, 491, 507
 transient input, 215, 439
 Nonoscillatory transients, 439–460
 Numerical integration, 93–96, 385–392
 Nuttall, A. H., 369
- Odd square-law nonlinearity, 360
 Oldenburger, R., 354, 357, 369, 420
 On-off element (*see* Ideal relay nonlinearity)
 Oppelt, W., 49
 Optical beam-riding missile, 284–286
 Optical tracker error characteristic, 105
 Optimal control (*see* Time-optimal computer control system; Wiener filter)
 Orbit period, calculation of, 298–300
 Oscillating control systems, 298
 (*See also* Limit-cycling control systems)
 Oscillations
 conservative free, 111, 164
 forced, 111, 190
 regional synchronous behavior, 192
 Oscillator, second-order, 47
- Padé approximants, 195, 445
 Pai, M. A., 499
- Pal'tov, N. P., 298
 Parallel-connected nonlinearities, 86–88
 Parameter-adaptive control (*see* Adaptive control system)
 Pastel, M. P., 222
 Pendulum, oscillations of, 165–166
 Perturbation method, 43–45
 Pervozvanskii, A. A., 370
 Phase plane, 10, 229
 Piecewise-linear nonlinearity, 60, 311, 380
 Piecewise-sinusoidal approximation, 222
 Poincaré, H., 43
 Point of apparent constraint, 328
 Polar plot, 117–118, 124, 243
 Polynomial nonlinearity, 63, 313, 371
 Popov, E. P., 30, 121, 174, 222, 298, 427
 Popov, V. M., 184–185
 Power-series expansion, 11, 264
 Power spectral density function, 639
 Prasanna-Kumar, K. S., 85
 Preload nonlinearity, 163, 281
 Probability, 629
 Probability density function, 630
 Probability distribution function, 630
 Probability function, 376, 377, 383, 393
 Probability integral, 375, 377, 394
 Pseudo nonlinearity (*see* Equivalent nonlinear element)
 Pulse-excited clock pendulum, 168–170
 Pulse-width-modulated systems, 505–509
 Pupkov, K. A., 367
 Pyshkin, I. V., 493, 506
- Quadrature gain, 53, 302
 calculation by area, 305
 Quantizer, 62
 Quasi-linearization, 13–14, 18–19, 439
- Raggazini, J. R., 462
 Ramp response
 limit-cycling system, 326
 randomly disturbed system, 429
 Random-input describing function
 accuracy of, 401–407
 approximate calculation of, 385–392
 tabulation of, 565–610

- Random process, 635
 - ergodic, 637
 - gaussian, 638
 - stationary, 636
- Random variable, 630
- Rankine, R. R., Jr., 85
- Rationally related periods, 252
- Rayleigh's equation, 208
- Rectangular hysteresis, 67, 465
- Recursive solution, 44, 265
- Regional division of the s plane, 321
- Relay (*see* Ideal relay nonlinearity)
 - with dead-zone, 57, 60, 114, 148, 187, 284, 384, 484–491
- Relay control systems, specific techniques
 - for continuous, 185–199, 369–370
 - for sampled, 465–472
- Relay-controlled split-field series motor, 351
- Replacement rules for transient oscillations, 217, 218
- Residual, 50, 71, 85, 162, 440
- Residual shaping network, 443–454
- Remnant (*see* Residual)
- Rice, S. O., 261
- RIDF (*see* Random-input describing function)
- RMS describing function, 85, 109
- Root locus plot, 124, 324–325, 337
- Root, W. L., 629
- Routh-Hurwitz stability test, 113–114
- Ruberti, A., 491
- Russell, F. A., 474, 490
- Sampled describing function, 475
 - accuracy of, 490
 - tabulation of, 611–624
- Sandberg, I. W., 183
- Saturation function, 59
- Saturation nonlinearity (*see* Limiter)
- Sawaragi, Y., 369
- Secular terms, 44
- Sensitivity function, 331, 360
- Series-connected nonlinearities, 88–89
- Series prefilter, 331
- Servomanometer, 352
- Shuck, O. H., 331
- Signal forms for describing function
 - analysis, 15
- Signal stabilization, 274, 354
- Simple pendulum, oscillations of, 165–168
- Simulation of system, 9
- Sinusoidal-input describing function
 - accuracy of, 162–185
 - approximate calculation of, 89–97
 - correction term for, 174–179
 - inversion of, 97–101
 - placing bounds on, 96–97
 - synthesis of, 86–89
 - tabulation of, 519–538
- Slowly varying amplitude and phase
 - method of, 43, 45–47
 - in transient oscillations, 212–214
- Slowly varying inputs, 304, 425–429
- Small parameters, method of, 43
- Smith, H. W., 369, 401, 407, 428
- Smith, O. J. M., 153
- Solution of nonlinear equations
 - closed form, 10
 - by series expansions, 11–12
- Somerville, M. J., 19, 30, 368, 387
- Sparacino, R. R., 138, 141–144
- Spoilers, 351
- Sridhar, R., 260, 272, 369, 420
- Stability curve, 279
- Standard deviation, 633
- Stoker, J. J., 43, 84, 132
- Stout, T. M., 83
- Subharmonic resonance, 15, 267–271
 - twenty-eight degree rule, 290
 - twenty-one degree rule, 272
- Superposition, 1
 - failure of, 2, 25
- Switching criterion, 139
- Systems
 - behavior of, 7–9
 - definitions of
 - linear, 1, 638
 - nonlinear, 2
 - time-invariant, 1, 639
 - time-varying, 2
- Takahashi, S., 369
- Temperature control system, 170
- Test of prototype system, 9
- Test inputs (*see* Nonlinearity test inputs)
- Thaler, G. J., 222
- Three-level relay (*see* Relay with dead zone)

- Three-sinusoid-input describing function, 295
- Threshold (*see* Dead zone)
- Time-delay, 195, 197–198
- Time-lag nonlinear differential equation, 116–117, 243–246
- Time-optimal computer control system, frequency response of, 138–144
- Torng, H. C., 499, 501
- Tou, J. T., 462
- Transient-input describing function, 439–454
- Transient oscillations, 212–249, 280–283, 457–458
 - quasi-static solution, 224
 - small- σ solution, 223–224
 - static solution, 224
- Transient response
 - nonoscillatory, 439–459
 - oscillatory, 212–231, 280–282
 - (*See also* Forced response)
- Transport lag (*see* Time-delay)
- TSIDF (*see* Two-sinusoid-input describing function)
- Tsyppkin, Ya. Z., 171, 188–198, 370
- Tustin, A., 49
- Two-sinusoid-input describing function, 250–296
 - accuracy of, 260
 - approximation of, 358
 - tabulation of, 557–564
- Unstable open-loop system, 325
- Van der Pol, B., 39, 42, 45, 104
- Vander Velde, W. E., 30, 222, 298
- Variance, 632
- Voronov, A. A., 222, 230
- Weighting function, 2, 638
- West, J. C., 253, 405, 419
- Wiener filter, 407
- Wiener-Hopf equation, 31
- z-transform describing function, 475

MIT OpenCourseWare
<https://ocw.mit.edu/>

16.30 Estimation and Control of Aerospace Systems
Spring 2004

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.