Time-optimal and suboptimal control of sampled-data systems
by John Leonard Pokoski

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in Electrical Engineering
Montana State University

Abstract:
The subject of this thesis is 'the study of both linear and nonlinear sampled-data systems with a minimum-time performance measure.

The content of the thesis may be divided into three sections: First, time-optimal control schemes for systems having input saturation are reviewed, and it is shown that for practical systems, there is generally a substantial cost advantage gained by relaxing the minimumtime requirement. Methods of measuring the amount of ”suboptimality" of such systems are then proposed. Several practical suboptimal strategies are analyzed and compared, and the analysis indicates a simple method of improving one of these strategies. It is also shown that these methods may be used to measure the sensitivity to parameter variations of systems which are designed to be either time-optimal or time-suboptimal. The proposed method of analysis is also used to show that several control systems which are claimed by their designers to be time-optimal are actually time-suboptimal. Second, a relationship called the time-loss hypothesis is indicated between the minimum times for continuous and sampled-data systems constrained by input signal saturation. This hypothesis states that for controllable systems containing plants having n real poles and no numerator dynamics, the minimum time required by a sampled-data controller is no greater than . tf + nT where tf is the time required by a continuous controller. The hypothesis is proven for first-order systems and several special cases of second-order systems, and evidence is presented for the validity of the relationship for other cases. Third, the theory of deadbeat response of linear systems is extended to parabolic inputs with minimum squared error restrictions on the step and ramp responses. It is found that minimizing only the sum of the squared ramp errors (subject to the parabolic deadbeat restrictions) also minimizes the maximum ramp error.

No such relationship exists if the sum of the squared step errors is minimized. Design methods are presented for digital controllers which provide trade-offs between the number of sample periods to deadbeat response to a parabolic input and either 1) minimum sum of squared ramp errors, 2) minimum sum of squared step errors, or 3) minimum-maximum step error.

221

TIME-OPTIMAL AND SUBOPTIMAL CONTROL OF SAMPLED-DATA SYSTEMS

By

John Leonard Pokoski

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

Approved:

Paul E. Ukleski
Head, Major Department

Donald G. Pierre
Chairman, Examining Committee

Louis J. Ashton
Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

August, 1967

ACKNOWLEDGEMENT

The author wishes to express his appreciation to the Electrical Engineering Department of Montana State University and to the National Aeronautics and Space Administration, for their financial support. Reproduction costs of this thesis were made available through contract number N00123-67-C-2408 between the United States Navy Electronics Laboratory, San Diego, and Montana State University.

He is also indebted to Professor Donald A. Pierre, who was always willing to sacrifice his time to give advice and encouragement.

He is grateful to his children, who were ever ready to "help". However, he is most grateful to his wife, Jane. Without her, this would not have been possible.

1098 m

TABLE OF CONTENTS

Page

CHAPTER 1: AN ANALYSIS SCHEME FOR SUBOPTIMAL, MINIMUM-TIME, SAMPLED-DATA SYSTEMS 1

1.1 Introduction 2

1.2 A Brief History of Minimum-Time Problems 3

1.3 Suboptimal Systems 8

1.4 Desoer and Wing's Time-Optimal Strategy 9

1.5 A Measure of Suboptimality 14

1.6 Some Examples of Measurement of Suboptimality . . 18

1.7 Sensitivity Analysis 24

1.8 Systems With Inputs 30

1.9 Counter Examples to Two "Time-Optimal" Systems . . 32

1.10 Conclusions 35

CHAPTER 2: SOME RELATIONS BETWEEN THE DISCRETE AND THE CONTINUOUS MINIMUM-TIME PROBLEMS 50

2.1 Introduction 51

2.2 First-Order Systems 53

2.3 Second-Order System With Real Eigenvalues 57

2.3.1 One Eigenvalue at Zero 58

2.3.2 Distinct, Nonzero Eigenvalues 68

2.3.3 Double Integrator 72

2.4 Complex Conjugate Eigenvalues 73

2.5 Pulse-Width-Modulated Systems 77

2.6 Conclusions 78

Page

CHAPTER 3: DEADBEAT RESPONSE TO PARABOLIC INPUTS WITH MINIMUM-SQUARED-ERROR RESTRICTIONS ON RAMP AND STEP INPUTS 89

3.1 Introduction 90

3.2 Derivation of the General Set of Equations 93

3.3 General Solution With Minimum Ramp-Error-Squared Performance Measure 98

3.4 General Solution With Minimum Step-Error-Squared Performance Measure 102

3.5 Conclusion 107

CHAPTER 4: SUMMARY AND SUGGESTED FUTURE RESEARCH 121

4.1 Summary of Essential Values of the Thesis Research 122

4.2 Aspects Meriting Additional Study 123

REFERENCES CITED 125

LIST OF TABLES

Page

Table 2.1. Data illustrating a method of finding the slope of the three-second isochrone at the first vertex of R_3 88

Table 3.1. Transfer function coefficients for squared-error performance measure with n = 4, 5 and 6 119

Table 3.2. Transfer function coefficients for minimizing the maximum step error with n = 4, 5 and 6 120

LIST OF FIGURES

		Page
Figure 1.1.	Continuous minimum time system	37
Figure 1.2.	Pulse-amplitude-modulated minimum time system	37
Figure 1.3.	r_1, r_2, r_3, r_4 and r_5 vectors for the plant characterized by $1/s(s+1)$ and a sample period of one-second	38
Figure 1.4.	R_N and R'_N regions, polygonal curve and critical curve for $N \leq 3$.	39
Figure 1.5.	$\theta^{-1}R_3$ for the plant characterized by $1/s(s+1)$ and a one-second sample period	40
Figure 1.6.	Minimum suboptimal regions in R'_3 for "saturation strategy" operating on the plant characterized by $1/s(s+1)$ and a one-second sample period	41
Figure 1.7.	Minimum suboptimal regions in R'_3 for "saturation strategy" operating on the plant characterized by $1/s(s+0.5)$ and a one-second sample period	42
Figure 1.8.	Portions of $\theta^{-1}R_k$ boundaries and suboptimal areas within R'_3 for Martens and Semmelhack strategy	43
Figure 1.9.	Suboptimal regions generated by application of the modified Martens and Semmelhack suboptimal strategy	44
Figure 1.10.	Suboptimal regions generated by application of Desoer and Wing's optimal strategy with a suboptimal critical curve.	45
Figure 1.11.	"Fixed" suboptimal regions for Desoer and Wing optimal strategy due to plant gain variations	46
Figure 1.12.	Saturating input system having a plant characterized by $1/s(s+a)$	47

Page

Figure 1.13. Saturating input system having a plant characterized by $1/(s+a)(s+b)$ 47

Figure 1.14. R_2' for ramp input with slope equal to 0.2 applied to a plant characterized by $1/s(s+a)$ and a one-second sample period 48

Figure 1.15. R_2' for ramp input with slope equal to 0.5 applied to a plant characterized by $1/s(s+a)$ and a one-second sample period 48

Figure 1.16. Region within R_3 for which + 1 or - 1 is sub-optimal for the plant characterized by $1/s(s+0.25)$ and a one-second sample period . . . 49

Figure 2.1. Continuous minimum-time system 80

Figure 2.2. Sampled-data minimum-time system 80

Figure 2.3. State space for plant characterized by $1/s(s+1)$ and a one-second sample period 81

Figure 2.4. R_1, R_2, R_3 and R_4 for plant characterized by $1/s(s-1)$ and one-second sample period 82

Figure 2.5. R_1, R_2, R_3 and R_4 for the plant characterized by $1/(s+0.693)(s+1)$ with a one-second sample period 83

Figure 2.6. R_1, R_2, R_3 and R_4 for the plant characterized by $1/(s-1)(s+1)$ and a one-second sample period 84

Figure 2.7. Switch curves and two typical trajectories corresponding to the plant characterized by $1/(s^2+1)$ 85

Page

Figure 2.8. R_1, R_2, R_3, R_4 and a segment of the boundary of R_5 for the plant characterized by $1/(s^2+1)$ . . 86

Figure 2.9. R_1, R_2 and R_3 for PWM control of the plant characterized by $1/s^2$ 87

Figure 3.1. Closed-loop system with digital controller 109

Figure 3.2. Unit step response for n = 5 with λ as parameter 110

Figure 3.3. Unit ramp response for n = 5, with λ as parameter 111

Figure 3.4. Unit step response for n = 4 and 6 with $\lambda = 0$ and ∞ 112

Figure 3.5. Unit ramp response for n = 4 and 6 with $\lambda = 0$ and ∞ 113

Figure 3.6. Unit step and ramp responses with n very large and T = 1 114

Figure 3.7. Unit step response with maximum magnitude of step error minimized 115

Figure 3.8. Unit ramp response with maximum magnitude of step error minimized 116

Figure 3.9. Unit step responses for n = 5 117

Figure 3.10. Unit ramp responses for n = 5 118

ABSTRACT

The subject of this thesis is the study of both linear and nonlinear sampled-data systems with a minimum-time performance measure.

The content of the thesis may be divided into three sections: First, time-optimal control schemes for systems having input saturation are reviewed, and it is shown that for practical systems, there is generally a substantial cost advantage gained by relaxing the minimum-time requirement. Methods of measuring the amount of "suboptimality" of such systems are then proposed. Several practical suboptimal strategies are analyzed and compared, and the analysis indicates a simple method of improving one of these strategies. It is also shown that these methods may be used to measure the sensitivity to parameter variations of systems which are designed to be either time-optimal or time-suboptimal. The proposed method of analysis is also used to show that several control systems which are claimed by their designers to be time-optimal are actually time-suboptimal. Second, a relationship called the time-loss hypothesis is indicated between the minimum times for continuous and sampled-data systems constrained by input signal saturation. This hypothesis states that for controllable systems containing plants having n real poles and no numerator dynamics, the minimum time required by a sampled-data controller is no greater than $t_f + nT$ where t_f is the time required by a continuous controller. The hypothesis is proven for first-order systems and several special cases of second-order systems, and evidence is presented for the validity of the relationship for other cases. Third, the theory of deadbeat response of linear systems is extended to parabolic inputs with minimum squared error restrictions on the step and ramp responses. It is found that minimizing only the sum of the squared ramp errors (subject to the parabolic deadbeat restrictions) also minimizes the maximum ramp error. No such relationship exists if the sum of the squared step errors is minimized. Design methods are presented for digital controllers which provide trade-offs between the number of sample periods to deadbeat response to a parabolic input and either 1) minimum sum of squared ramp errors, 2) minimum sum of squared step errors, or 3) minimum-maximum step error.

CHAPTER 1

AN ANALYSIS SCHEME FOR SUBOPTIMAL, MINIMUM-TIME, SAMPLED-DATA SYSTEMS

1.1 INTRODUCTION

As the title suggests, the principal purpose of this chapter is to propose and illustrate a method of measuring the amount of "suboptimality" for sampled-data systems which use more than the minimum time necessary to reach a zero-error condition.

Sections 1.3 through 1.7 are centered on this problem in relation to a regulator system; that is, a zero input system, the function of which is to keep its output at a rest state despite any unwanted system disturbances. In Section 1.8, it is shown how the various concepts which are introduced can be extended to systems having inputs; that is, positioning or tracking systems.

Section 1.2 contains a brief history of the time-optimum problem, for both continuous and sampled systems. In Section 1.3, suboptimal systems are defined; their advantages and their utility are considered. Section 1.4 contains a brief description of the time-optimal strategy proposed by Desoer and Wing [8, 9, 10], because many of their ideas are used throughout this chapter and the next. In Section 1.5, methods of analyzing suboptimal systems and measuring the amount of suboptimality are proposed, and in Section 1.6, examples are given to illustrate these methods. In Section 1.7, it is pointed out that a system which is designed to be time-optimal will always in practice be suboptimal due to plant parameter variations. The methods of analysis for suboptimal systems are then used for a sensitivity analysis of a suboptimal system and a system designed to be time-optimal.

In Section 1.9, an analysis is given on two systems [39, 26] which are claimed as time-optimum by their designers. These systems are shown to actually be suboptimal, thus clarifying some confusion which has existed in the literature on this matter.

State space methods and phase-plane models [39, 23] are used extensively throughout this chapter.

1.2 A BRIEF HISTORY OF MINIMUM-TIME PROBLEMS

The classic minimum-time problem is that of finding the control signal which will reduce the error of a linear system without numerator dynamics (no plant zeroes) to zero in the minimum possible time if the control signal is limited by a magnitude constraint (Figure 1.1). The system equation is

$$\dot{x}(t) = A\,x(t) + B\,u(t) \quad ; \quad -L < u < L \tag{1.1}$$

where x and B are n-dimensional vectors, A is an n by n constant matrix, u(t) is a scalar, and L is a constant. The problem is known by a variety of names, such as the time-optimal control problem, the bang-bang problem, on-off servo problem, et cetera. If the ideal output in Figure 1.1 is zero, the system is called a regulator. The following history of the problem is not intended to be complete, but only to indicate the principal developments.

The first paper published on this problem was by McDonald [27] in 1950. He was concerned with minimizing the transient of a second-

order system. McDonald's arguments were heuristic, as were those of Hopkin [15] in 1951, when he used phase-plane analysis on a servomotor subject to saturation and found that the control signal must always be at the saturation level for time-optimal control. He also found that the phase plane may be divided into two equal sections by a "switch curve." The time-optimal control for states lying on one side of this curve is positive saturation, while on the other side, it requires negative saturation. Bushaw [6] in 1952, attacked the problem more rigorously, and showed that some of McDonald's and Hopkin's intuitive arguments do not hold for a plant with complex conjugate poles. In 1953, La Salle [22] proved that the best (in the minimal-time sense) bang-bang system, i.e. a system whose input is always at a plus or minus saturation level, is the optimum of all systems subject to the same saturation limits. In 1954, Bogner and Kazda [5] found that for systems with real poles, results indicate that as the order of the servomechanism increases by one, so does the number of switchings. In 1956, Bellman, Glicksberg and Gross, [2] gave a general treatment of the minimum-time regulator problem. Possibly the greatest advance concerning this problem was made by Pontryagin [34] who introduced the maximum principle in 1957. The use of this principle made clear that the time-optimal control for an n'th-order plant having real poles consists of a signal always at the saturation level, but having no more than n-1 changes of sign (switchings). If the plant has complex conjugate poles, the control signal is always at the saturation

level, but there is no limit on the number of sign changes. Although the maximum principle is an important aid in determining the form of the control, it does not eliminate the problem of implementing the controller, which becomes quite complex, particularly for high-order systems.

With the advent of the electronic digital computer, minimum-time control of systems by use of discrete controllers (i.e. sampled-data controllers) began to be investigated. The most common sampled-data system may be represented by a sampler followed by a zero-order hold circuit which in turn is followed by the plant as in Figure 1.2. There is no loss in generality if the system is normalized to give saturation limits of plus one and minus one [8]. The output of the zero-order hold circuit is constant throughout the sampling period (T), and is equal to the value of the input to the sampler at the previous sampling instant. This system is said to be pulse-amplitude modulated (PAM). Another common system is called pulse-width modulated (PWM). In this case the input to the plant is always at a saturation level, but the length of the pulse varies, depending upon the magnitude of the control signal at the previous sampling instant.

In 1957, the PAM, sampled-data, bounded-input, minimum-time problem was investigated independently by Kalman [17] and Krasovskii [19, 20]. The difference equation is, in this case,

$$x[(k+1)T] = A\,x(kT) + B\,u(kT), \quad -L < u < L \tag{1.2}$$

with the dimensions the same as in (1.1). Krasovskii's method, which is essentially a variational approach, is summarized by Zadeh [44] who notes that the method, although conceptually simple, is computationally difficult. Kalman suggested the division of the state space into regions which are determined by the minimum possible number of sample periods required to reach the origin. It follows that if a control strategy, during each sample period, moves the state of the system from the region which is k sample periods from the origin to the region which is k-1 sample periods from the origin, the strategy is time-optimal. Kalman's work was extended by Desoer and Wing in a series of papers [8, 9, 10]. In the first [8] they analyzed a system having the plant $1/s(s+a)$ with "a" greater than zero and proposed a special computer to implement an optimal strategy. They also noted that for almost all initial states the optimal strategy is not unique. A brief summary of this paper is given in Section 1.4. In reference [9] they gave an optimal controller for the plant $1/\prod_{i=1}^{n}(s-\lambda_i)$ where the λ_i are real, distinct, and nonpositive, while in reference [10] they showed that the real and distinct restrictions are not necessary. In 1962, Zadeh and Whalen [44] noted that the problem can be solved by use of linear programming techniques. Torng [38] in 1964 detailed the linear programming formulation of the problem, and, since the time-optimal control sequence is, in general, not unique, he added an additional criterion (minimum fuel) which must be satisfied, thus enabling a unique choice of the minimum-time control sequence. In 1962,

Ho [14] made an initial guess of the optimal control and then used an iterative approach to "zero in" on an optimal control sequence. Tou and Vadhanaphuti [41, 39] proposed an optimal position controller which essentially uses the difference between the desired and actual positions as the input to the saturating amplifier. However, this system is shown to really be suboptimal in Section 1.9. Meksawan and Murphy [26] used a modification of Tou's approach, and their system is also shown to be suboptimal in Section 1.9. Tou [40, 39] also solved the minimum-time regulator problem by treating the forcing functions at the sampling instants as variables in the state equations, and increasing the number of sampling periods until an optimal sequence within the saturation limits can be found. This approach is basically linear programming, although Tou uses other terminology. In reference [39] Tou indicates that for an n'th-order system, if the minimal number of sample periods required is n + q, the first q driving signals may be set at the saturation limit. This is shown to be untrue in Section 1.9. Neustadt [29] claims to have solved the problem if the forcing function is allowed to be an r-dimensional vector. That is, in (1.2), u(kT) is an r vector and B is an n by r matrix. He also allows A and B to be time varying. His approach is similar to Krasovskii's [19, 20].

The important factor to be noted about all of the above time-optimal strategies is that they are all, to various degrees, unsuitable for on-line control. The inadequacies are discussed by Eaton [12], Koepcke [18], and Martens and Semmelhack [25]. In general, all of the

methods either require extensive computation times or the permanent storage of switching surfaces, which may require excessive memory space. Eaton [12] designed a special-purpose on-line computer for the job, but it is regarded as too expensive by Martens and Semmelhack [25]. Koepcke's on-line approach requires storage of pre-computed tables whose size depends upon the maximum possible number of sample periods required to reach the origin of the state space. Thus, if a large segment of the state space is to be included, the storage required may become excessive. Martens and Semmelhack [25] proposed a suboptimal approach which will be analyzed in Section 1.6.

The PWM minimum time problem has been studied principally by Polak [32] who used an approach somewhat similar to Kalman's [17].

1.3 SUB-OPTIMAL SYSTEMS

First, a definition of a suboptimal minimum-time system is in order. A suboptimal minimum-time system is simply one in which at least one of all the possible initial states of the system takes longer than the minimum possible time to reach the desired final state.

From this definition it immediately becomes evident that it is practically impossible to construct a time-optimal system, because any parameter variation in the time-optimal controller, or in the plant for which the controller was designed will, in general, cause some initial state to take longer than minimum time to reach the desired final state. (Henceforth such an initial state will be referred to as suboptimal.) Since plant parameter identification always involves

approximations anyway, it is obvious that it would be unwise to design such a minimum-time system in which it is extremely important that the minimum time be used in all cases.

Another practical consideration is cost. It is intuitively evident that a suboptimal controller should be less expensive than an optimal controller. In Section 1.6, it is shown that the suboptimal regulators which are discussed are much simpler, hence much cheaper than any of the optimal systems discussed in Section 1.2.

If trade-offs are to be made between system cost and system suboptimality, it is necessary to have some way of measuring the suboptimality of a system, and generally, it would also be important to know the worst-case time, that is the longest amount of time that can possibly occur between any initial and final state. For example, Martens and Semmelhack [25] propose a suboptimal strategy "which is on occasion suboptimal during saturated inputs but optimal for all linear operation." The vagueness of this statement leaves much to be desired.

1.4 DESOER AND WING'S TIME-OPTIMAL STRATEGY

Since various aspects of Desoer and Wing's work [8, 9, 10] will be used throughout this dissertation, some of their results will be briefly outlined. The second-order regulator system will be studied and, in particular, the plant $1/s(s+a)$ [8].

The differential equation corresponding to this system is

$$\ddot{c}(t) + a\,\dot{c}(t) = u(t) \qquad , \qquad 0 \leqslant t \leqslant T \tag{1.3}$$

where T is the sample period, u(t) is constant between sampling instants, and u(t) is bounded between plus and minus one (saturation limits). If the initial conditions are expressed as c(0) and c(0), the solution to (1.3) at t equals T may be written in the state equation format as

$$\begin{bmatrix} c(T) \\ \\ c(T) \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{1 - e^{-aT}}{a} \\ \\ 0 & e^{-aT} \end{bmatrix} \begin{bmatrix} c(0) \\ \\ c(0) \end{bmatrix} + \begin{bmatrix} \dfrac{e^{-aT} + aT - 1}{a^2} \\ \\ \dfrac{1 - e^{-aT}}{a} \end{bmatrix} u(0) \tag{1.4}$$

It simplifies matters somewhat to put equation (1.4) into canonical form by using the linear, normalized eigenvector transformation

$$\begin{bmatrix} \gamma_1(t) \\ \\ \gamma_2(t) \end{bmatrix} = \begin{bmatrix} 0 & (1+a^2)^{\frac{1}{2}}/a \\ \\ -1 & -1/a \end{bmatrix} \begin{bmatrix} c(t) \\ \\ \dot{c}(t) \end{bmatrix} \tag{1.5}$$

which results in the difference equation for the general interval between kT and (k+1)T, as follows:

$$\begin{bmatrix} \gamma_1[(k+1)T] \\ \gamma_2[(k+1)T] \end{bmatrix} = \begin{bmatrix} e^{-aT} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1(kT) \\ \gamma_2(kT) \end{bmatrix} + \begin{bmatrix} (1-e^{-aT})(1+a^2)^{\frac{1}{2}}/a^2 \\ -T/a \end{bmatrix} u(kT) \tag{1.6}$$

or, more concisely,

$$\gamma[(k+1)T] = \theta\gamma(kT) + d\, u(kT) \tag{1.7}$$

where the θ and d identities are obvious. Desoer and Wing then define

$$r_k = -\theta^{-k}d = \begin{bmatrix} -e^{kaT}(1 - e^{-aT})(1+a^2)^{\frac{1}{2}}/a^2 \\ T/a \end{bmatrix} \tag{1.8}$$

Several r_k vectors are illustrated in Figure 1.3 for "a" equals T equals one. (These values will be used in the running example.) Two other definitions are now made. R_N' is defined as the set of initial states that can be brought to equilibrium in N sample periods or less, while R_N is the set of all initial states that can be brought to the origin in N sample periods and no less. It is evident that R_N may be obtained by deleting from R_N' all those states belonging to R_{N-1}'. Desoer and Wing then prove that R_N' is the closed set whose boundary is the convex polygon π_N which has the following 2N

vertices:

$$
\begin{aligned}
OP_1 &= r_1 - r_2 - r_3 \dots - r_N \\
OP_2 &= r_1 + r_2 - r_3 \dots - r_N \\
&\vdots \\
OP_N &= r_1 + r_2 + r_3 \dots + r_N \\
OP_{-1} &= -r_1 + r_2 + r_3 \dots + r_N \\
OP_{-2} &= -r_1 - r_2 + r_3 \dots + r_N \\
&\vdots \\
OP_{-N} &= -r_1 - r_2 - r_3 \dots - r_N
\end{aligned}
\tag{1.9}
$$

Figure 1.4 illustrates R_1', R_2' and R_3' along with R_1, R_2 and R_3.

It is apparent that if a strategy can be found that, during each sample period, transfers the state of the system from R_N to R_{N-1}, the strategy is time-optimal. Desoer and Wing show that such a strategy implies that the following equation is satisfied.

$$R_N - ur_1 = \theta^{-1} R_{N-1} \tag{1.10}$$

The implication of (1.10) is that if $\gamma(0)$ is in R_N, there is a u with

magnitude less than or equal to one such that $\gamma(0) - ur_1$ is in $\theta^{-1}R_{N-1}$. This in turn implies that $\gamma(1)$ will be in R_{N-1}. At the beginning of each sample period, the system state can be called $\gamma(0)$ so that the above relations can be used repetitively. The boundaries of $\theta^{-1}R_{N-1}$ are found by adding $\pm r_1$ to the boundary of R_N in all inward directions. $\theta^{-1}R_3$ is shown in Figure 1.5. Thus if any $\gamma(0)$ state is in R_N and some percentage (m_1) of r_1 may be added to this state to bring it into $\theta^{-1}R_{N-1}$, the strategy is optimal. Several facts are now evident. First of all, the inner boundary of $\theta^{-1}R_{N-1}$ is inconsequential, since it can not be reached from R_N. Secondly, if $\gamma(0)$ is on the outer boundary of R_N, there is a unique strategy which will place $\gamma(1)$ in R_{N-1}. Thirdly, if $\gamma(0)$ is in R_N but not on its outer boundary, there are an infinite number of time-optimal strategies.

Desoer and Wing's optimal strategy is as follows. Draw a curve, called the critical curve, by adding $\pm r_1$, in an inward direction, to one outer edge of the R_N regions, as shown in Figure 1.4. The series of outer edges is called the polygonal curve, and is also shown in Figure 1.4. The critical curve is thus composed of the r_2, r_3, r_4, ... vectors placed "tail to head" with all having the same sign. The strategy is to compute m_1 such that $\gamma(0) - m_1 r_1$ is a point on the critical curve. If the calculated value of m_1 is greater than or equal to one, let m_1 equal one. If the calculated value of m_1 is less than or equal to minus one, let m_1 equal minus one. If the calculated value of m_1 is between minus one and plus one, use this value. Desoer

and Wing show a method of implementing this strategy in reference [8].

1.5 A MEASURE OF SUBOPTIMALITY

Despite the cost advantage of suboptimal controllers, very little has been done to provide a measure of suboptimality. However, in 1964, Polak [33] proposed the following figure of merit for suboptimal systems,

$$m_1 = \int_{x \in X} c(x)\, p(x)\, dV \tag{1.11}$$

where x is any initial state of the system, X is a bounded region of the state space, dV is an element of volume of the state space, p(x)dV is the probability that x lies in dV, and c(x) is some cost function. This general idea will be applied to the minimum-time problem in this section and methods of evaluating c(x) and m_1 will be proposed.

The assumption will first be made that the set of all possible initial states is bounded. This is a realistic assumption, and should cause no practical problems. It is also assumed that the probability of an initial state lying in a certain volume of the state space is governed by a probability distribution function which can be determined. If the system is a regulator, and something is known about the input noise statistics, the desired probability distribution can be found by a simulation. In some cases, this distribution can be found analytically. For example, if the input noise has a gaussian distribution such that the saturation effects can be taken into account, the multi-

variate probability distribution function for the output can be written explicitly [7]. Now consider a tracking system. Suppose the controller controls a rocket launcher which is in an integrated firing system. The state of the system as it is commanded to leave one target and begin tracking another would be strongly correlated to the structure of the integrated fire control system, which could then be used to help determine the probability distribution of initial states.

Throughout the remainder of this chapter, the problem will be discussed in terms of a second-order system. However, the various ideas can be extended to higher-order systems with the usual accompanying complexity. On the other hand, it is a common practice to use dominant-pole synthesis for controller design [16] and this approximation would, in many cases be adequate for suboptimal system analysis.

Once the suboptimal strategy is defined, and the initial state is specified, the number of sample periods required to reach the origin from this state is determined. The time-optimal number of sample periods is also known, since N is the minimum possible number of periods required for an initial state in R_N, and the R_N boundaries are known. In general, the regions of the state space containing initial states requiring more than the minimum number of sample periods will be grouped into regions whose boundaries are determined by the suboptimal strategy. The probability of the initial state falling into this "suboptimal area", A, is, of course, $\int_A p \, dA$, where p is the probability distribution of initial states. One obvious cost factor which could be used to weight

the suboptimality measure is the difference between the minimum possible number of sample periods required, and the number of sample periods required by the strategy being used. This gives, essentially, the "average" number of extra sample periods required by a typical initial state. Examples of this measurement for various strategies will be given in Section 1.6.

The problem with the above "exact" measure of suboptimality is that, although the trajectory from the initial state to the origin is theoretically known, the tracing of this trajectory may become quite tedious in practice. This becomes evident even for a relatively simple example considered in Section 1.6. Thus, to enable a quicker comparison of suboptimal strategies, a "rougher" suboptimality measure is often desirable. The basic idea behind the approximate measure of suboptimality which is proposed herein is that of considering each R_N region separately, and treating each state transition as a random process. As was pointed out above, the process is actually deterministic once the initial state is known, if the controlling action is not disrupted by noise. However, it is reasonable to assume that if the controller is forcing the state of the system into R_k, the probability of the R_k state being suboptimal is strongly related to the percentage of suboptimal states in R_k. The proposed approximate suboptimality measure is

$$M = \sum_{k=1}^{j} \sum_{i} (i + 1 - k)\, p_{ki} \tag{1.12}$$

where k is the index associated with R_k regions, and j is related to the outer bound of possible initial states; i.e., R_j. p_{ki} is the percentage of states in R_k which will be transferred by the control strategy to R_i. It will be shown in Section 1.6, that p_{ki} can be measured relatively easily by using a slight extension of Desoer and Wing's work [8]. $(i + 1 - k)$ is a weighting factor corresponding to the number of sample periods lost at each R_k. For example, if the strategy transfers the state from R_k to R_{k-1}, the strategy is optimal, i equals k-1, and the weighting factor is zero. Similarly, if the new state is also in R_k, one sample period is lost, and the weighting factor is one. If the strategy forces the state to R_{k+1}, the weighting factor is two.

It is reasonable to consider the probability of time loss separately for each R_k region, since any initial state in R_N must pass through R_{N-1}, R_{N-2}, ... R_1 during its trajectory. However, a close correlation between M and the average measure of time loss should not be expected (particularly if j is large) because of the broad assumption made. On the other hand, equation (1.12) is a useful tool for comparing several suboptimal strategies for the same system with the same set of possible initial states.

In the preceding analysis, the control strategy is essentially considered as a Markov process, since each new state is considered to be a probabilistic function of the present state. Markov probability transition-matrix notation [36] will be occasionally used in the

examples in Section 1.6.

1.6 SOME EXAMPLES OF MEASUREMENT OF SUBOPTIMALITY

Example 1.6.1

Several authors [39, 26, 25] have proposed strategies for second-order systems by which the forcing function is always at its maximum magnitude except for the final two sample periods. (Once the state is in R_2, operation is linear, and the optimal controlling actions are relatively easy to determine.) In the first examples the plants $1/s(s+1)$ and $1/s(s+0.5)$ will be analyzed, and it will be assumed that the initial state occurs in R_3'. Thus, only R_3 can contain suboptimal states. It will also be assumed that the forcing function for states in R_3 is the "better" of the two saturation levels, plus one or minus one. This is an important consideration, since, if the "better" forcing function is used, the state can do no worse than remain in R_3, while if the other level is chosen, the state could be driven into some R_k with k greater than three. This will be explained in more detail in a later example. It will also be assumed throughout this section that the probability distribution of initial state is uniform. The sample period is one second.

Now consider Figure 1.6. The shaded areas contain all possible suboptimal initial states. It can be easily shown that a plus or minus one forcing function applied to any state in the shaded areas will require one additional sample period in its trajectory to the origin.

The shaded region represents 19.3% the area of R_3, and 16.5% of R_3'. (These, and most other area measurements in this chapter were made in an analogue manner, with the associated inaccuracies.) Thus, the "exact" measurement of suboptimality is 0.165, while from (1.12), M equals 0.193. Similarly, Figure 1.7 shows that for $1/s(s + 0.5)$, the exact measure is 0.309 while M equals 0.371. It must be noted that the initial states which are considered in the second example do not include the same region of the state space as in the first example.

Example 1.6.2

In this example, a system governed by a different suboptimal strategy [25] is studied. A brief description of the strategy is appropriate. The original strategy was employed on a stochastic input system, but its modification to a regulator is straightforward. The plant $1/s(s+1)$ will be considered, along with a one-second sample period. From (1.6), the state equation is

$$\begin{bmatrix} \gamma_1(k+1) \\ \gamma_2(k+1) \end{bmatrix} = \begin{bmatrix} .368 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1(k) \\ \gamma_2(k) \end{bmatrix} + \begin{bmatrix} .894 \\ -1 \end{bmatrix} u(k) \tag{1.13}$$

or

$$\gamma(k+1) = \theta\gamma(k) + d\, u(k) \tag{1.14}$$

By applying this relation again,

$$\gamma(k+2) = \theta^2\gamma(k) + \theta du(k) + du(k+1) \tag{1.15}$$

The basic principle of the strategy is to assume that the origin can always be reached in two sample periods. Thus ideal values $\hat{u}(k)$ and $\hat{u}(k+1)$ of $u(k)$ and $u(k+1)$ can be calculated from (1.15). This gives

$$\begin{bmatrix} \hat{u}(k) \\ \hat{u}(k+1) \end{bmatrix} = \begin{bmatrix} .239\gamma_1(k) + 1.58\gamma_2(k) \\ -.239\gamma_1(k) - 0.583\gamma_2(k) \end{bmatrix} \tag{1.16}$$

Of course, for states external to R_2', either $\hat{u}(k)$ or $\hat{u}(k+1)$ or both will be beyond the saturation level. In these cases, the following rule is employed.

$$u(k) = \begin{cases} \mathrm{sgn}[\hat{u}(k)] \text{ if } \hat{u}(k) \geqslant 1 \\ \hat{u}(k) \text{ if } |\hat{u}(k)| < 1 \text{ and } |\hat{u}(k+1)| < 1 \\ \mathrm{sgn}[\hat{u}(k+1)] \text{ if } |\hat{u}(k)| < 1 \text{ and } |\hat{u}(k+1)| > 1 \end{cases} \tag{1.17}$$

A new set of control signals are computed during each sample period.

Reference will be made to Figure 1.8 in the following discussion. It will first be assumed that all possible initial states occur in R_3'. An analysis will show that it is advantageous to modify the strategy, after which another analysis will be made.

Several facts are quite evident: 1) The strategy is always optimal

within R_2', since R_2' is the linear region. 2) Because of symmetry, only the region $\gamma_1 < 0$ need be considered. (Any reasonable strategy for this type of problem will in general be symmetric.) 3) Inspection of equation (1.17) and Figure 1.8 indicates that the strategy is optimal in the third (and first) quadrant.

A slight extension of equation (1.10) can be used to measure how many sample periods are "lost" by the various suboptimal regions in the fourth quadrant. It can be shown that if

$$R_N - ur_1 = \theta^{-1}R_p \tag{1.18}$$

the state of the system will be driven from R_N to R_p. Since the $\theta^{-1}R_p$ areas are known throughout the state plane, the next sequential R_p regions can easily be determined from the present region and the strategy being employed.

Thus, for the running example, Figure 1.8 shows the new R_k regions which are entered by the suboptimal states in R_3. Thus 8.27% of the possible initial states in R_3' are forced to remain in R_3, 13.9% are forced into R_4, 10.35% are forced into R_5, and 2.64% are forced into R_6. It is apparent that an exact analysis of this strategy would be quite tedious. However, it is easy to apply (1.12) to give an approximate measure of M equal to 0.777.

Example 1.6.3

It will now be shown that the strategy of Example 1.6.3 may be

easily modified to give a smaller amount of suboptimality. The modified strategy is

$$u(k) = \begin{cases} \hat{u}(k) \text{ if } |\hat{u}(k)| \leqslant 1 \text{ and } |\hat{u}(k+1)| \leqslant 1 \\ \operatorname{sgn}[\hat{u}(k)] \text{ if } |\hat{u}(k)| > 1 \text{ and } |\hat{u}(k)| > |\hat{u}(k+1)| \\ \operatorname{sgn}[\hat{u}(k+1)] \text{ if } |\hat{u}(k+1)| > 1 \text{ and } |\hat{u}(k+1)| > |\hat{u}(k)| \end{cases} \tag{1.19}$$

It can be shown that this strategy is optimal in R_2' and in the first and third quadrants. It can also be shown that, external to R_2', the strategy implies that $u(k)$ equals $+1$ if $\gamma_1(k)$ is positive and $u(k)$ equals -1 if $\gamma_1(k)$ is negative. Figure 1.9 shows the suboptimal region in R_3'. An important factor is that each suboptimal state is forced to remain in R_3, rather than some states being forced to "higher" R_k regions. Thus, although 34.4% of the states in R_3 are suboptimal (as compared with 35.2% for the unmodified strategy), M is equal to 0.344 (as compared with 0.776 of the preceding example).

A further analysis will now be made, using the modified strategy, but assuming that initial states may occur anywhere in R_5'. The suboptimal regions are shown in Figure 1.9, and M is found to be 0.637. A probability transition matrix may be written for initial states as

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} & p_{05} \\ p_{10} & p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{30} & p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{40} & p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{50} & p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & .656 & .344 & 0 & 0 \\ 0 & 0 & 0 & .814 & .186 & 0 \\ 0 & 0 & 0 & 0 & .893 & .107 \end{bmatrix} \tag{1.20}$$

where the element in the i'th row and j'th column indicates the probability of a transition from R_i to R_k.

Example 1.6.4

The next example will be a brief analysis of the resulting suboptimal strategy if Desoer and Wing's critical curve is simplified, but their strategy is still used. The reason for choosing such a system is that in continuous systems, switching curves and surfaces have been similarly simplified [13] in order to gain cost advantages. The correct critical curve will be used within R_2', since, in this linear area, there is a unique choice of optimal controlling actions.

In the example, once again the plant is $1/s(s+1)$ and the sample period is one second. It is assumed that all initial states occur within R_5'. The suboptimal switching curve is shown in Figure 1.10. Although the section of the curve external to R_2' was chosen rather arbitrarily, it would certainly be feasible to pick this portion of

the curve with the objective of minimizing the suboptimality measure. Figure 1.10 shows the suboptimal regions. The probability transition matrix is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & .896 & .104 & 0 & 0 \\ 0 & 0 & 0 & .963 & .037 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{1.21}$$

M is found to equal 0.141 in this case. Note that this measure is smaller than for the modified Martens and Semmelhack strategy previously discussed. On the other hand, the modified Martens and Semmelhack strategy is easier to implement.

One final word will be said about all of the examples of this section. In each case, the outer boundary of possible initial states was the outer boundary of some R_k region. This was done to keep the examples as clear as possible. Of course the suboptimality measures which were proposed could be used for any area of the state plane, and would depend upon the area which was chosen.

1.7 SENSITIVITY ANALYSIS

The classical purpose of sensitivity analysis [4, 33] has been

to measure the change in the transmission characteristics of a system per change in some plant parameter. Since the study of optimal control systems has gained importance, some attention has been given to redefining the sensitivity function with respect to the performance measure being optimized [11, 37]; however, sensitivity associated with the minimum-time problem has not been studied in a general framework. In this section, a sensitivity analysis will be made on a minimum-time system, and on a suboptimal minimum-time system.

The suboptimality measures introduced in this chapter can be applied to the sensitivity analysis of minimum-time systems, because parameter variations will cause the minimum-time strategy to be suboptimal in practice. Of course, systems which are suboptimal initially can be similarly studied. In either case, one important factor must be considered: if any of the parameters used in the minimum-time design are disturbed, it will generally take an infinitely long time to reduce the state of the system to the origin so that it remains there when the controlling action is removed. Thus, to make the minimum-time sensitivity measure meaningful, the target must be considered to be some finite region about the origin, rather than the origin itself. In the examples given in this section, the finite region will be chosen to be R_2'. In addition to being convenient for the calculations, it is reasonable to choose this region because, for second-order plants, it is the only region of the state space in which the control action is linear, and perhaps a more conventional linear controller could be applied in this region, along with a more conventional sensitivity analysis.

Once again, the system to be analyzed for the sake of illustration contains the plant $1/s(s+1)$ and uses a sample period of one second. The first strategy to be considered is Desoer and Wing's optimal strategy. The parameter variation is assumed to occur in the gain of the saturating amplifier, and the variation is $\pm\delta$ about the nominal value of one where the positive δ is always assumed to be very small. Only the $\gamma_1 < 0$ region will be studied, since the $\gamma_1 > 0$ area is symmetric to it. It will be assumed that all initial states occur within R_5'.

First consider $+\delta$. In this case the gain is slightly larger than that used in the optimal design. Since the strategy is always to drive for the critical curve, the only suboptimal region in R_3 is immediately to the left of the critical curve, as is shown in Figure 1.11. The r_1 vector starting from any state occuring in this region will extend past the critical curve and out of $\theta^{-1}R_2$ by an amount δ. There is also a very small suboptimal region in the lower right-hand corner of R_3 which is negligible in comparison to the area previously considered.

It appears that similar suboptimal regions are located immediately to the left of the critical curve in R_4 and R_5. However, a closer analysis shows that this is not the case. First consider the possibility of overdriving the critical curve in R_4. Suppose the strategy is applied to a state in the parallelogram immediately to the left of the critical curve. If $\gamma(k)$ is in this region, for the nominal gain, a force between minus one and zero will place $\gamma(k+1)$ on the polygonal curve in R_3. However, considering the $+\delta$ variations, equation (1.13)

gives

$$\begin{bmatrix}\gamma_1(k+1)\\ \gamma_2(k+1)\end{bmatrix} = \begin{bmatrix}.368\gamma_1(k) - .894(1+\delta)u\\ \gamma_2(k) + (1+\delta)u\end{bmatrix} \quad , \quad 0 < u < 1 \tag{1.22}$$

The next force must be + 1 (nominally), since $\gamma(k+1)$ is to the right of the critical curve. Thus,

$$\begin{bmatrix}\gamma_1(k+2)\\ \gamma_2(k+2)\end{bmatrix} = \begin{bmatrix}[.368\gamma_1(k) - .894(1+\delta)u\ ]\ .368 + .894(1+\delta)\\ \gamma_2(k) + (1+\delta)u - (1+\delta)\end{bmatrix} \tag{1.23}$$

Note that the δ variation is advantageous in this case, since it tends to force $\gamma(k+2)$ closer to R_2. Equation (1.23) shows that $\gamma_1(k+2)$ is $.894\delta(1-.368u)$ above its nominal value, while $\gamma_2(k+2)$ is $\delta(1-u)$ to the left of its nominal value. The nominal value of $\gamma_2(k+2)$ is on the boundary of R_2. The question is whether or not the actual $\gamma_2(k+2)$ is internal to R_2. Note that the slope of r_2, which bounds R_2 in this vicinity is -6.61. If the slope of the line connecting the nominal and actual values of $\gamma(k+2)$ is more negative than -6.61, $\gamma(k+2)$ is in R_2, and $\gamma(k)$ was not a suboptimal state. Thus,

$$\frac{.894\delta(1-.368u)}{-\delta(1-u)} < -6.61 \tag{1.24}$$

which gives

$$u > .91 \tag{1.25}$$

Thus, there is a parallelogram in R_4 of "width" .09, which will cause an extra sample period to be taken. This set of states is shown in Figure 1.11.

Now consider the possibility of starting in R_5 immediately to the left of the critical curve, and thus immediately losing a sample period. An analysis similar to the above, shows that there is no region in R_5 for which this can occur. In other words, the "overdrive" corrects itself in this case.

In addition to the suboptimal regions, there are narrow regions of "width" external to each R_k whose states are actually improved by the $+\delta$ variation. However, these areas are relatively small and will be neglected in the calculation of M.

Thus, considering only the suboptimal regions which are independent of the magnitude of the $+\delta$ variation, the probability transition matrix may be written as

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} & p_{05} \\ p_{10} & p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ 0 & 0 & .681 & .319 & 0 & 0 \\ 0 & 0 & 0 & .989 & .012 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{1.26}$$

where the first three rows are not considered. M is found to be 0.331.

Gain variations of $-\delta$ produce suboptimal regions to the right of the critical curve in R_3, R_4 and R_5 (Figure 1.11). The transition matrix is

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} & p_{05} \\ p_{10} & p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ 0 & 0 & .681 & .319 & 0 & 0 \\ 0 & 0 & 0 & .872 & .128 & 0 \\ 0 & 0 & 0 & 0 & .857 & .143 \end{bmatrix} \tag{1.27}$$

and M equals 0.59. In addition, there is a narrow suboptimal strip just inside each R_k region whose area is proportional to δ and will be neglected.

The most interesting result of the above study is that the sensitivity involves a large constant term which is independent of δ. Of course, this is only true of the strategy being used, and not in general.

The sensitivity of the system could be similarly measured with respect to variations in the time constant of the plant which is also one in this case. These variations will, in turn, cause variations in r_1, and the resulting suboptimal effects can be found.

A cursory analysis will now be made of the sensitivity to gain

variations of the modified Martens and Semmelhack suboptimal strategy. Consider a $+\delta$ change in gain. In this case the nominal suboptimal states are increased about their inner perimeter by a strip whose area is directly proportional to δ. Corresponding to $-\delta$, the nominal suboptimal states are decreased by a similar strip about their inner perimeter. However, there is an additional suboptimal strip about the inside of the outer boundary of the R_k regions.

Thus, in this case, the sensitivity is a linear function of δ, in contrast with the previous example which involved a large constant plus a linear function of δ.

1.8 SYSTEMS WITH INPUTS

Test inputs used to study system characteristics usually include inputs of the form t^k, where k is a positive integer or zero. The most commonly used are steps, ramps and parabolas. It is easy to show that a second-order system can not track a parabolic (or higher) input since the forcing function will eventually have to be greater than the saturation limit on the input. Kalman [16] points out that to track a ramp with zero steady-state error, the second-order plant must be a double integrator. Similarly, a plant with one integrator can track a step with zero steady-state error. However, if the steady-state error need not be zero, it can be easily shown that the system of Figure 1.12 (one integrator) can be made to track a ramp whose slope is less than or equal to L/a. Similarly, the system of Figure 1.13

(no integrators) will track a step whose magnitude is less than L/ab. In this section, it will be shown how the R_k regions can be described, and Desoer and Wing's optimal strategy applied to the plant $1/s(s+a)$ for step and ramp inputs.

The differential equation for this plant is given by (1.3). However, since the $c(t)$ state variable is not being driven to zero, but is required to follow the input $r(t)$, it is convenient to use as the state variable the difference between the two. Thus,

$$e(t) = r(t) - c(t) \tag{1.28}$$

Combining (1.28) with (1.3) gives

$$\ddot{e}(t) + a\,\dot{e}(t) = \ddot{r}(t) + a\,\dot{r}(t) - u(t) \tag{1.29}$$

If $r(t)$ is a step initiated at $t = 0$, (1.29) reduces to

$$\ddot{e}(t) + a\,\dot{e}(t) = -\,u(t) \quad , \quad t > 0 \tag{1.30}$$

If $r(t)$ is a ramp with slope K initiated at $t = 0$, (1.29) reduces to

$$\ddot{e}(t) + a\,\dot{e}(t) = a\,K - u(t) \quad , \quad t > 0 \tag{1.31}$$

A comparison between (1.30) and (1.3) makes clear that the R_k regions

corresponding to each of the equations are identical. This will be further clarified in the next section when the system is analyzed for a step input. On the other hand, a comparison between (1.31) and (1.3) shows that, for a ramp input, the location of the R_k regions depends upon the slope. Figures 1.14 and 1.15 show R_1 and R_2 for slopes of 0.2 and 0.5. Although the Desoer and Wing strategy could be easily applied to this system for various step inputs, the application to ramp inputs would be relatively difficult due to the shifting of the critical curve which would be required. Thus, even though it is possible to track the ramp with a nonzero steady-state error, there are added difficulties in the implementation of such a system.

1.9 COUNTER EXAMPLES TO TWO "TIME-OPTIMAL" SYSTEMS

Consider the system of equation (1.30) with "a" equal to 0.25 and a sample period of one second. Figure 1.16 shows the R_k regions if the transformation of equation (1.5) is applied to e and $\dot{e}$. Also, assume a step input of magnitude R equals 2.5. Assuming that the system is initially at rest, the initial errors are $e(0^+)$ equals 2.5 and $\dot{e}(0^+)$ equals 0. This corresponds to $\gamma_1(0)$ equals 0 and $\gamma_2(0)$ equals -2.5. Inspection of Figure 1.16 shows that this initial state lies in R_4, thus involving a minimum time of four sample periods. One possible optimal control sequence is 1, 0.1, 0.27, -0.734.

Meksawan and Murphy [26] derive equations (1.32) and (1.33) for the first two signals of an optimal control sequence.

$$u(0) = \frac{a\ R}{K\ T(1 - e^{-aT})} \tag{1.32}$$

and

$$u(T) = -\frac{a\ R}{K\ T(e^{aT} - 1)} \tag{1.33}$$

For this example, these equations give a u(0) equal to 2.83 and u(T) equal to -2.2. Meksawan and Murphy say that if both of these control signals exceed the saturation values in magnitude (± 1), it has been found from experience that the time-optimal control requires that the first two control signals applied must be at the saturation level with the same sign as the step input. Hence, according to their strategy, u(0) equals u(T) equals + 1.

Equations (1.28, 1.30, 1.5 and 1.6) can be used to find $\gamma(1)$ and $\gamma(2)$.

$$\begin{bmatrix} \gamma_1(T) \\ \gamma_2(T) \end{bmatrix} = \begin{bmatrix} -3.65 \\ 1.50 \end{bmatrix}, \quad \begin{bmatrix} \gamma_1(2T) \\ \gamma_2(2T) \end{bmatrix} = \begin{bmatrix} -6.47 \\ 5.50 \end{bmatrix} \tag{1.34}$$

Thus, $\gamma(T)$ is in R_3, but $\gamma(2T)$ is also in R_3. Thus u(T) equal to + 1 is not an optimal input function. Figure 1.17 shows that $\gamma(T)$ lies in a region which is suboptimal to all inputs at the saturation level. It is not difficult to find other examples of plant-input combinations

for which the strategy of driving at the saturation level during the first two sample periods is suboptimal.

Tou's approach [39] is to always drive the system with e(kT). If e(kT) is greater than the saturation level, he says to use the closer saturation level. For the system described above, e(0) is 2.5. Thus u(0) equals 1 which, in turn, results in c(T) equal to 0.464. Thus e(T) equals 2.036 and u(T) equals 1. But the previous example shows that an initial sequence of + 1, + 1 is suboptimal; hence, Tou's strategy is suboptimal.

The reason that neither strategy worked in this case was simply that the trajectory entered a region external to R_2 where a saturated input could not be employed in an optimal manner. Now consider the plant 1/s(s+1) whose R_k regions are shown in Figure 1.7 along with the R_3 region for which saturated inputs can not be optimal. Now suppose any step input is applied while the system is in the rest state. $\dot{e}(0)$ is zero which means that $\gamma_1(0)$ is also zero. However, the maximum magnitude to which $\dot{c}$ (and thus $\dot{e}$) can be driven is ± 1. Thus the maximum possible value which γ_1 can attain is ± 1.414. Figure 1.6 shows, however, that the minimum magnitude reached by the region which is suboptimal for saturated inputs is approximately ± 10. Thus the problem encountered by the plant 1/s(s+0.25) can not occur for the plant 1/s(s+1), which is the plant usually employed for examples of time-optimal strategies, and, in particular, is used in both Meksawan and Murphy's and Tou's examples.

1.10 CONCLUSIONS

In this chapter a brief history was given of the minimum-time control problem. It was then shown that, in practice, it may be much less costly to use a suboptimal system. A method was then proposed to give an exact measure of how close a system is to being optimal. Since this exact measurement is usually quite tedious to evaluate, an approximate measure was also proposed. Several examples of measuring the suboptimality of practical strategies were given, and it was shown that the insight provided by the analysis of one of the strategies indicated a simple manner of improving it. It was then shown how the suboptimality measures could be applied to measure the sensitivity of a system to plant variations.

At this point, it is informative to make a comparison between the Desoer and Wing optimal strategy and the modified Martens and Semmelhack suboptimal strategy. It was shown in Section 1.7 that the Desoer and Wing strategy is much more sensitive to parameter variations than the modified Martens and Semmelhack strategy. Moreover, for negative variations in the plant gain, the "fixed" suboptimal regions for the optimal strategy are comparable to those for the suboptimal strategy. Thus, the "rough" measure indicates that there is little reason to choose one strategy over the other. If the amount of time-loss is critical, a much more exact analysis must be made. From a cost viewpoint, the suboptimal strategy is much easier to implement. However, this factor could become irrelevant if the available computing

device is large enough to generate the optimal strategy.

Finally, counter examples were given for two strategies which had been claimed as optimal by their designers, and it was shown that the analysis methods used gave added insight into the reasons for the suboptimality of these systems.

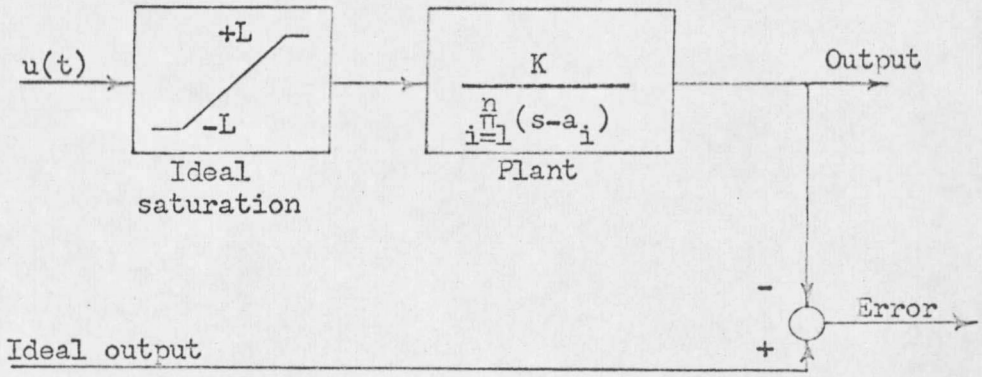

Figure 1.1. Continuous minimum time system.

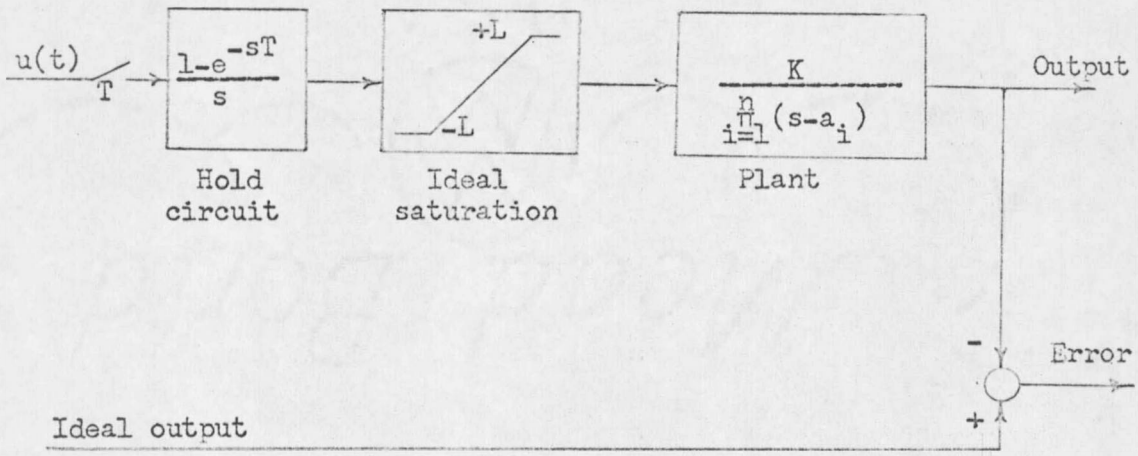

Figure 1.2. Pulse-amplitude-modulated minimum time system.

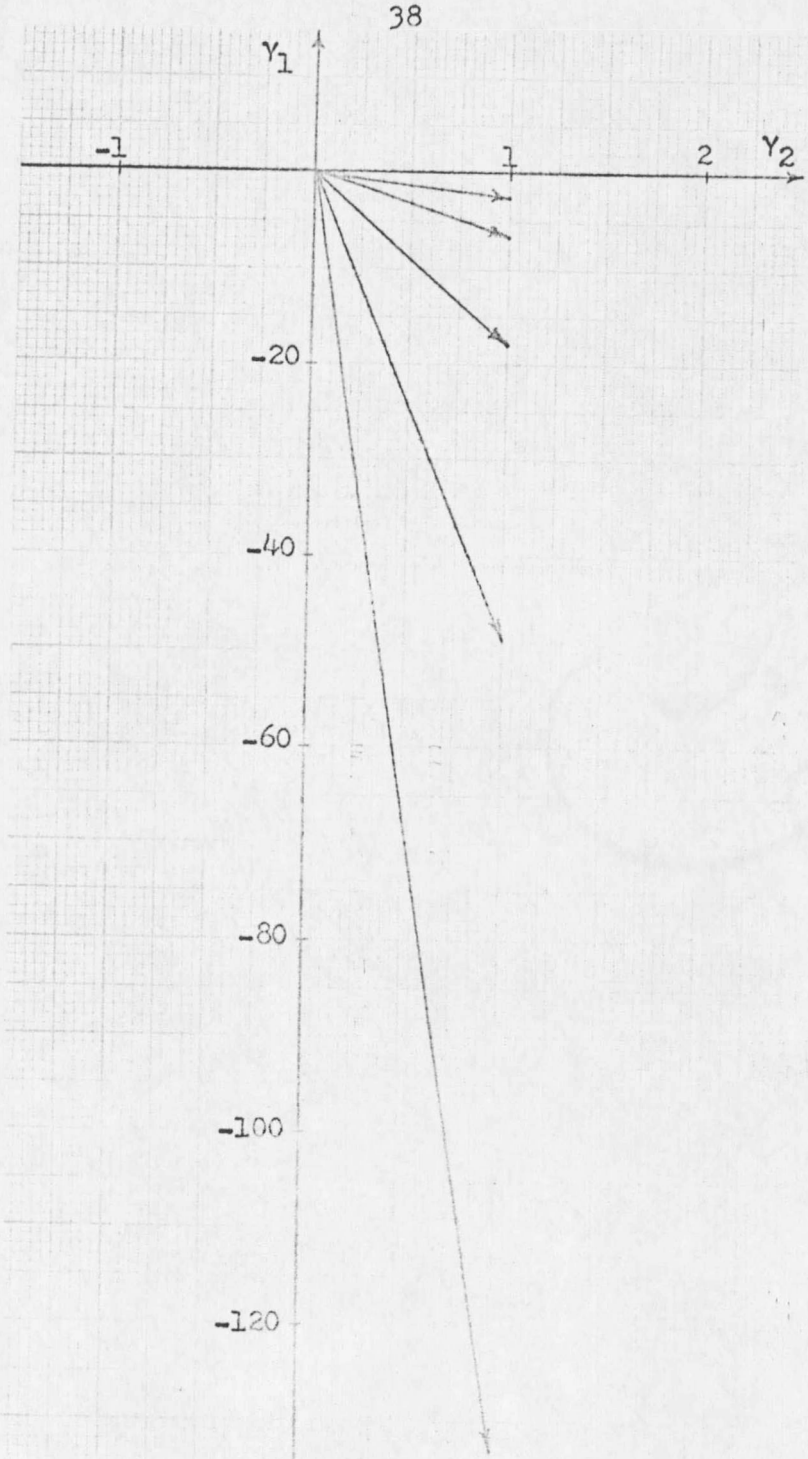

Figure 1.3. r_1, r_2, r_3, r_4 and r_5 vectors for the plant characterized by $1/s(s+1)$ and a sample period of one-second (see equation (1.8)).

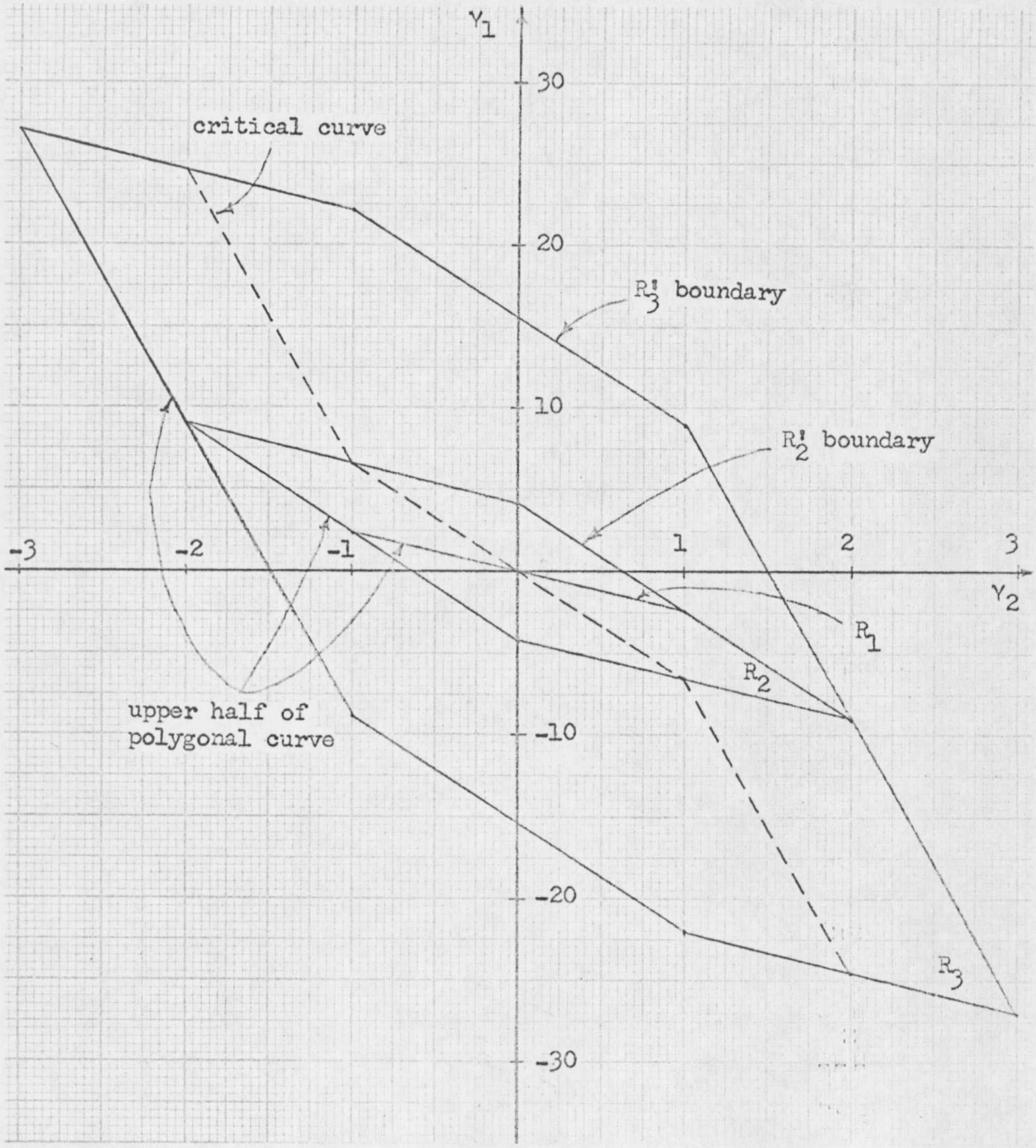

Figure 1.4. R_N and R_N' regions, polygonal curve and critical curve for $N \leq 3$. The plant is characterized by $1/s(s+1)$ and the sample period is one second (see Section 1.4).

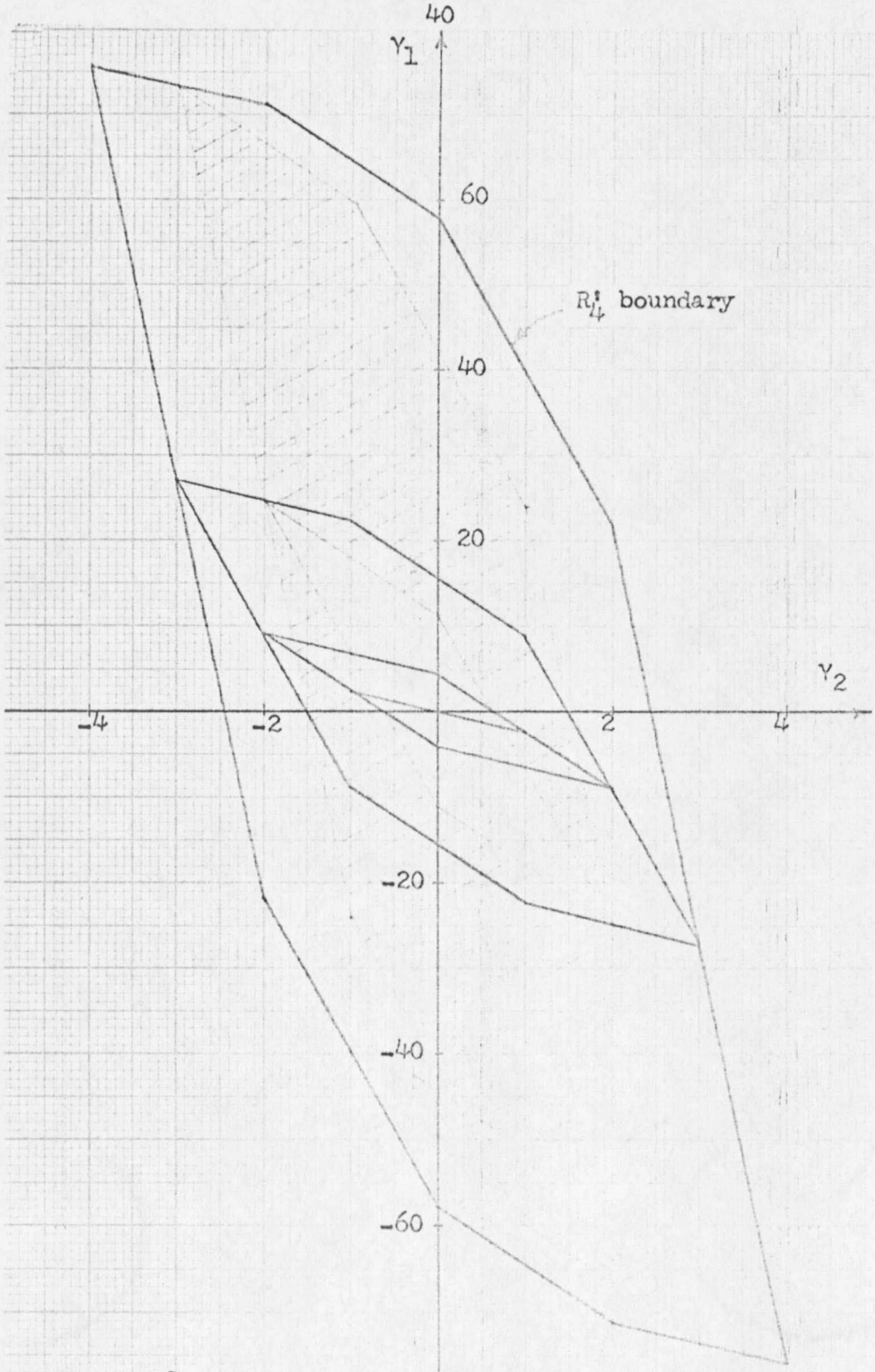

Figure 1.5. $\theta^{-1}R_3$ (crosshatched region) for the plant characterized by $1/s(s+1)$ and a one-second sample period (see equation (1.10)).

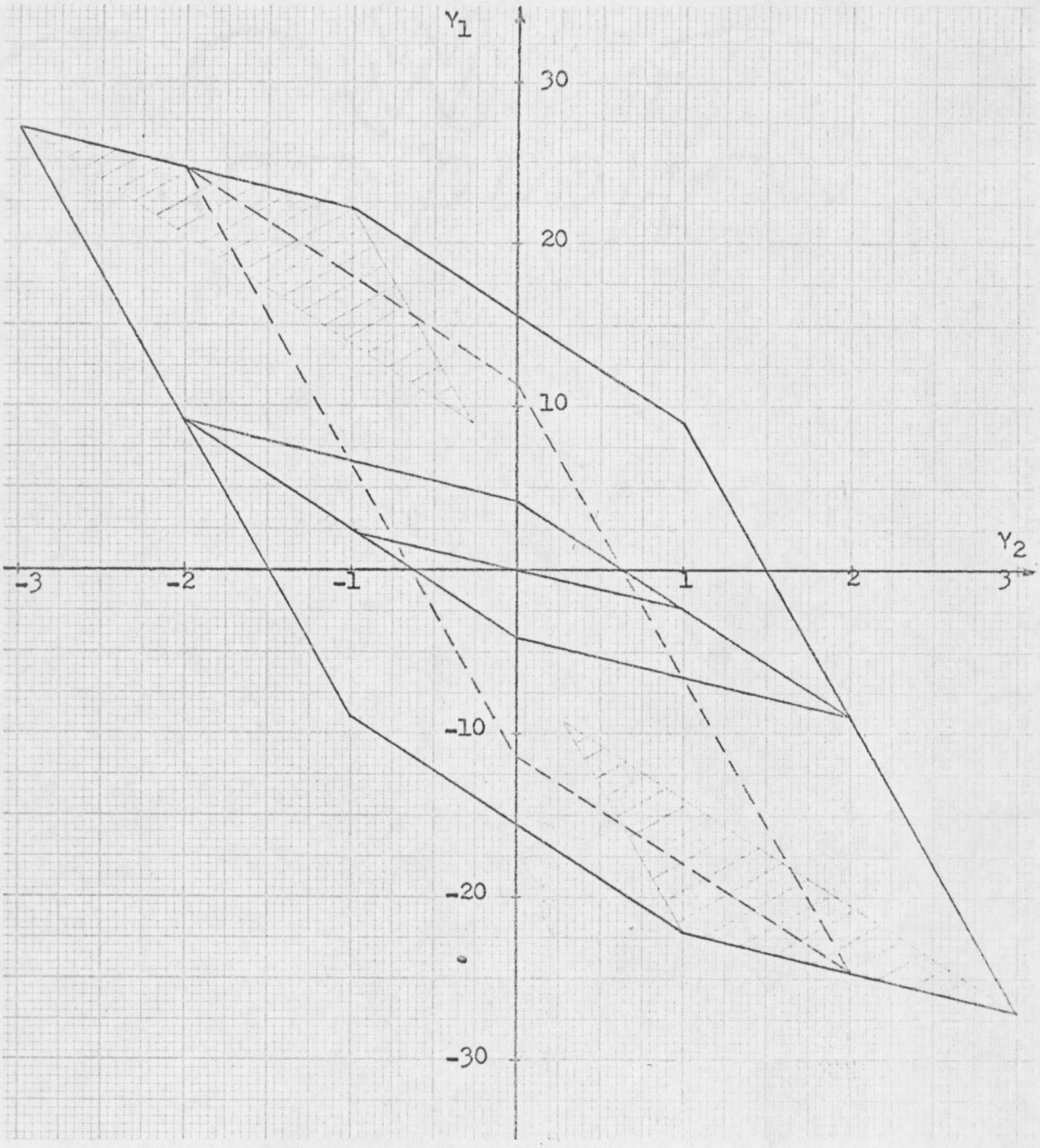

Figure 1.6. Minimum suboptimal regions (crosshatched) in R_3' for "saturation strategy" operating on the plant characterized by $1/s(s+1)$ and a one-second sample period (see Example 1.6.1.).

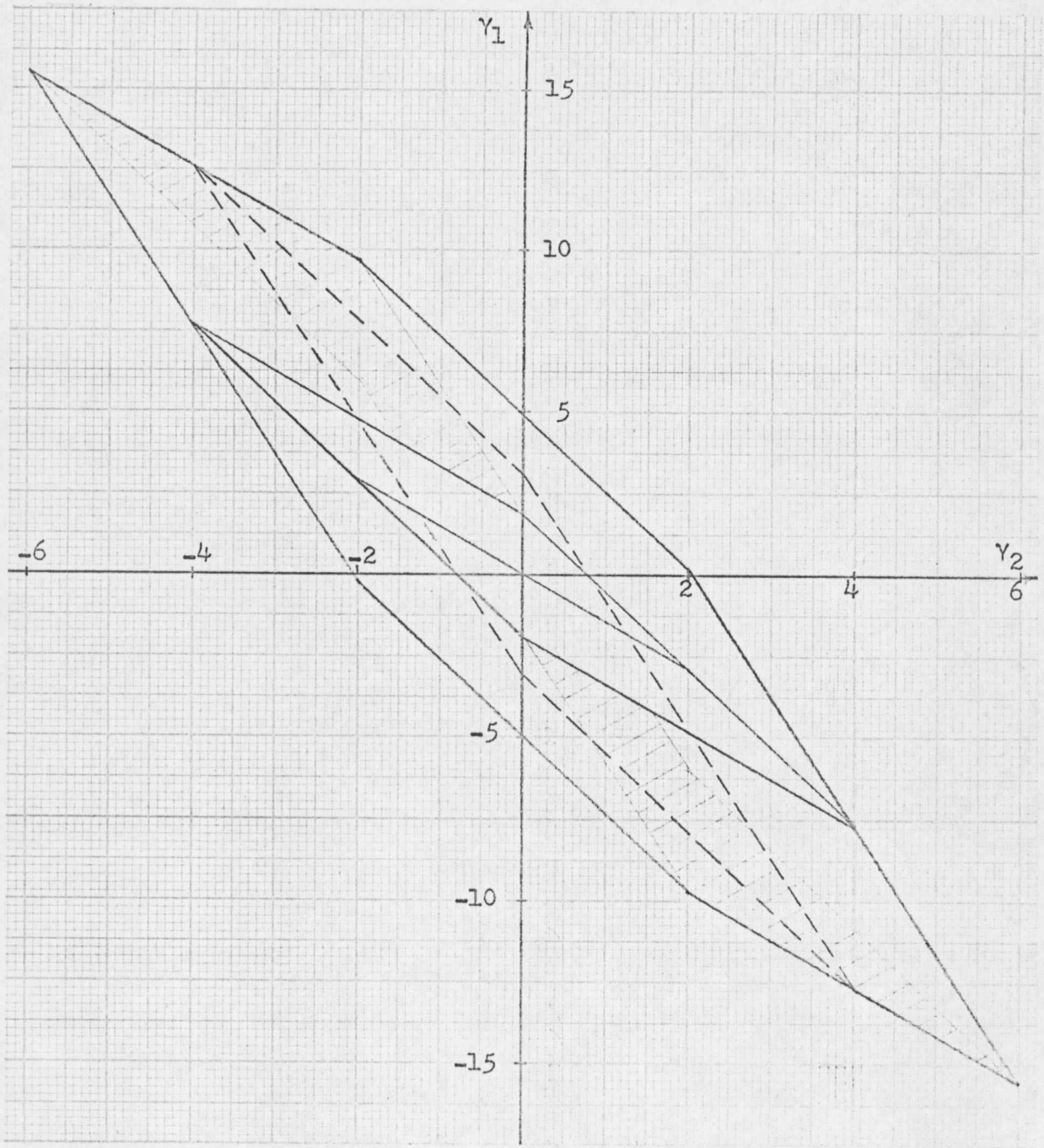

Figure 1.7. Minimum suboptimal regions (crosshatched) in R_3' for "saturation strategy" operating on the plant characterized by $1/s(s+0.5)$ and a one-second sample period (see Example 1.6.1.).

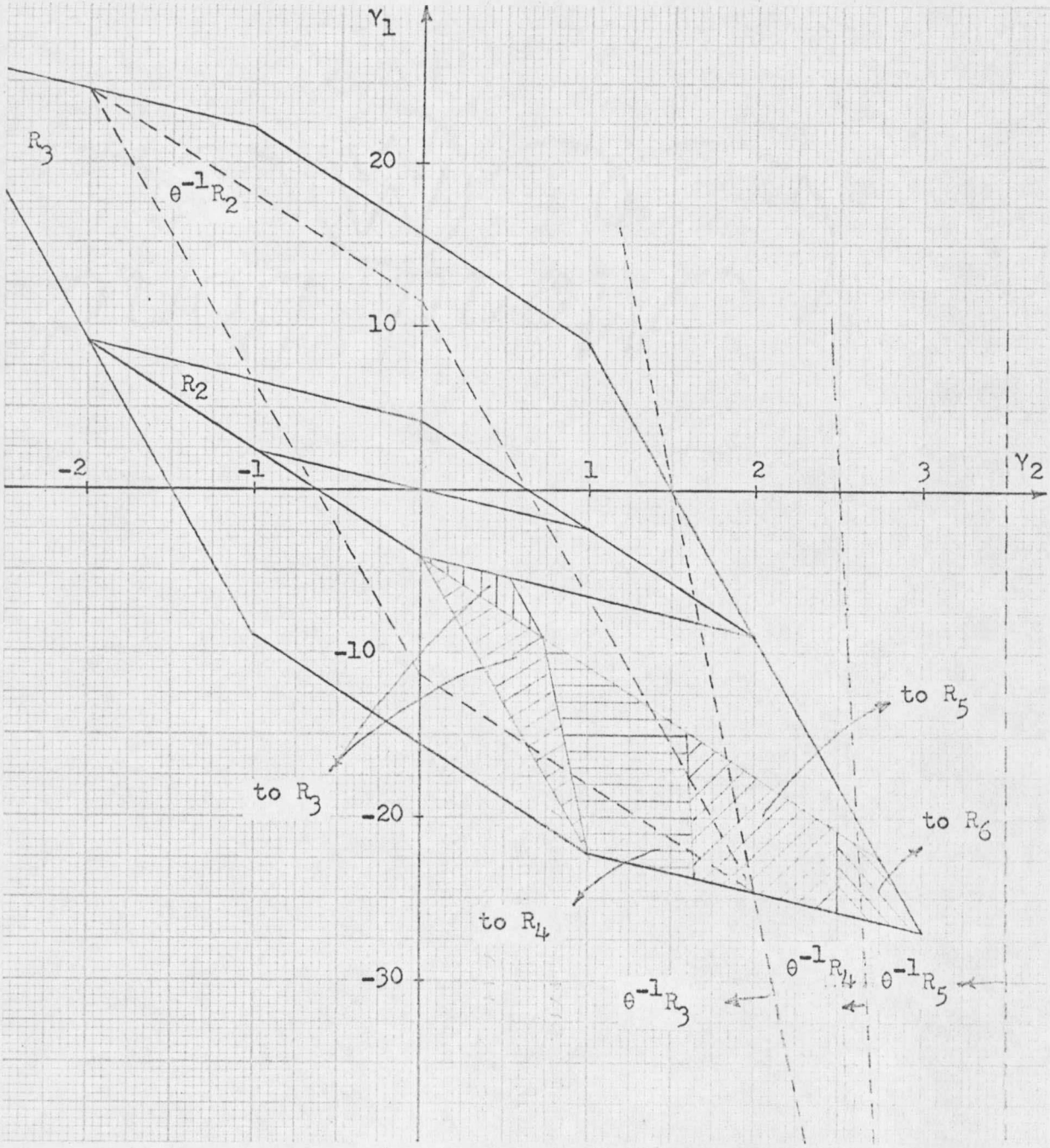

Figure 1.8. Portions of $\theta^{-1}R_k$ boundaries and suboptimal areas within R_3' for Martens and Semmelhack strategy (see Example 1.6.2.). The plant is characterized by $1/s(s+1)$, and the sample period is one-second.

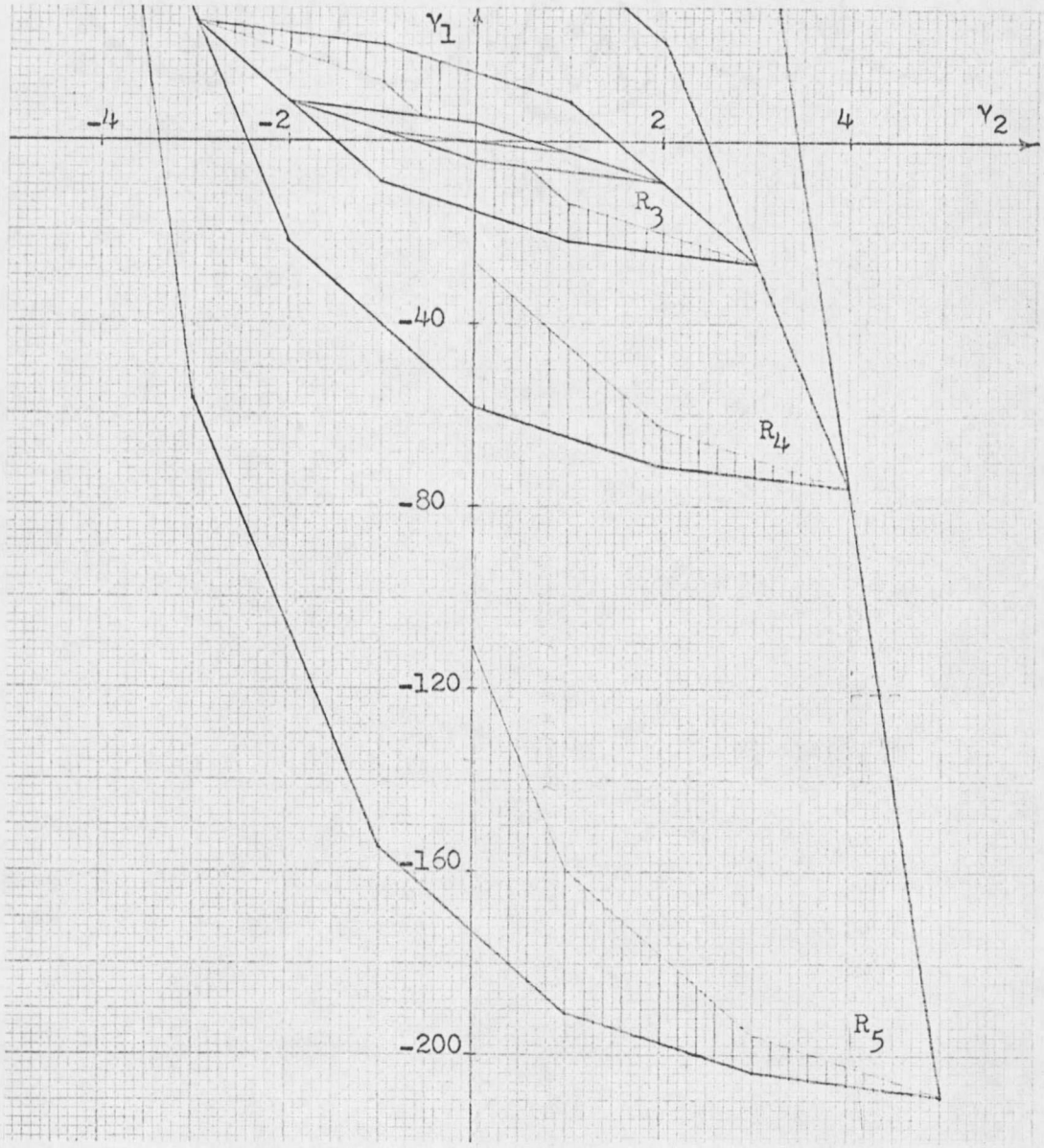

Figure 1.9. Suboptimal regions generated by application of the modified Martens and Semmelhack suboptimal strategy (see Example 1.6.3.). The plant is characterized by 1/s(s+1) and the sample period is one second.

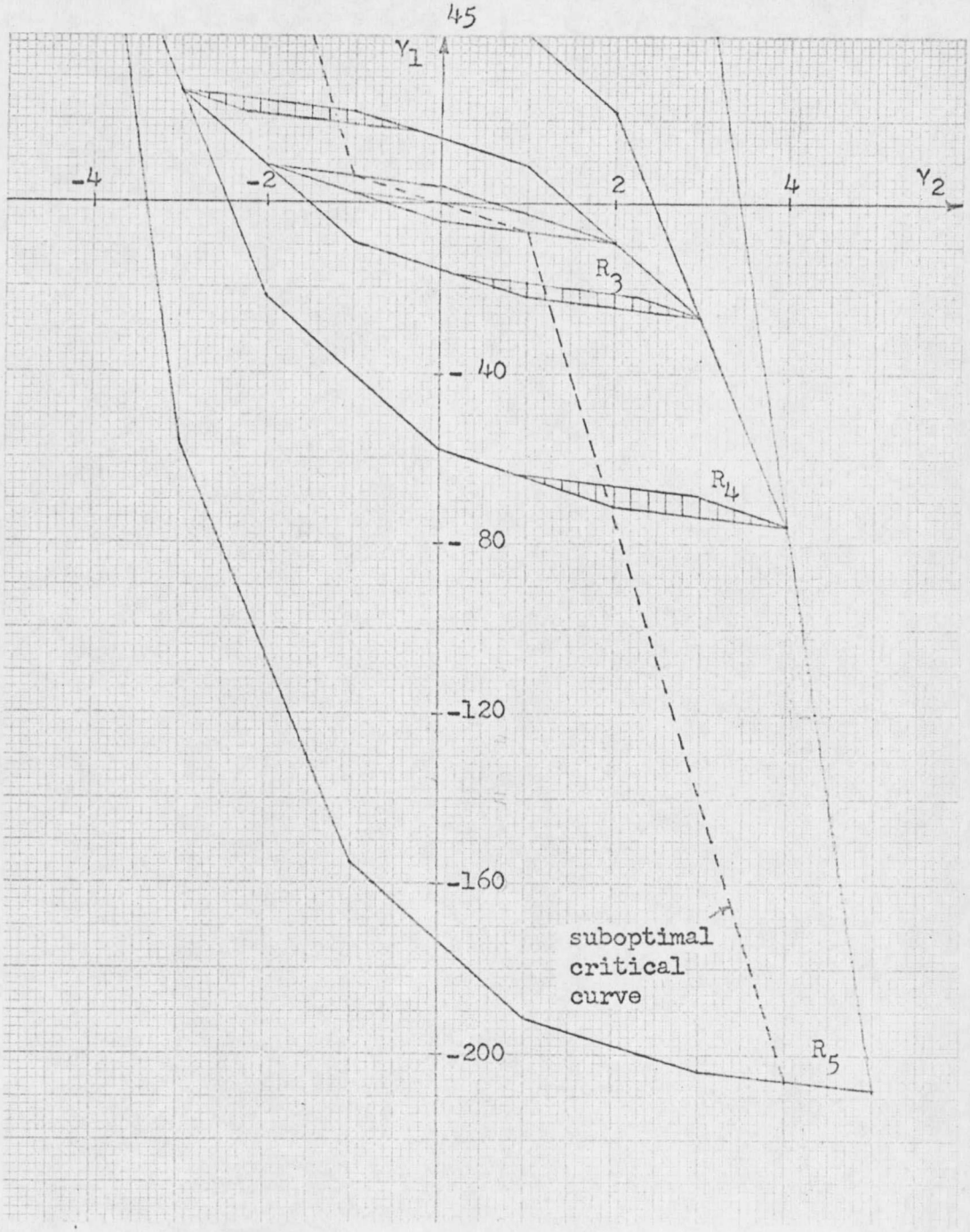

Figure 1.10. Suboptimal regions generated by application of Desoer and Wing's optimal strategy with a suboptimal critical curve (see Example 1.6.4.). The plant is characterized by 1/s(s+1) and the sample period is one second.

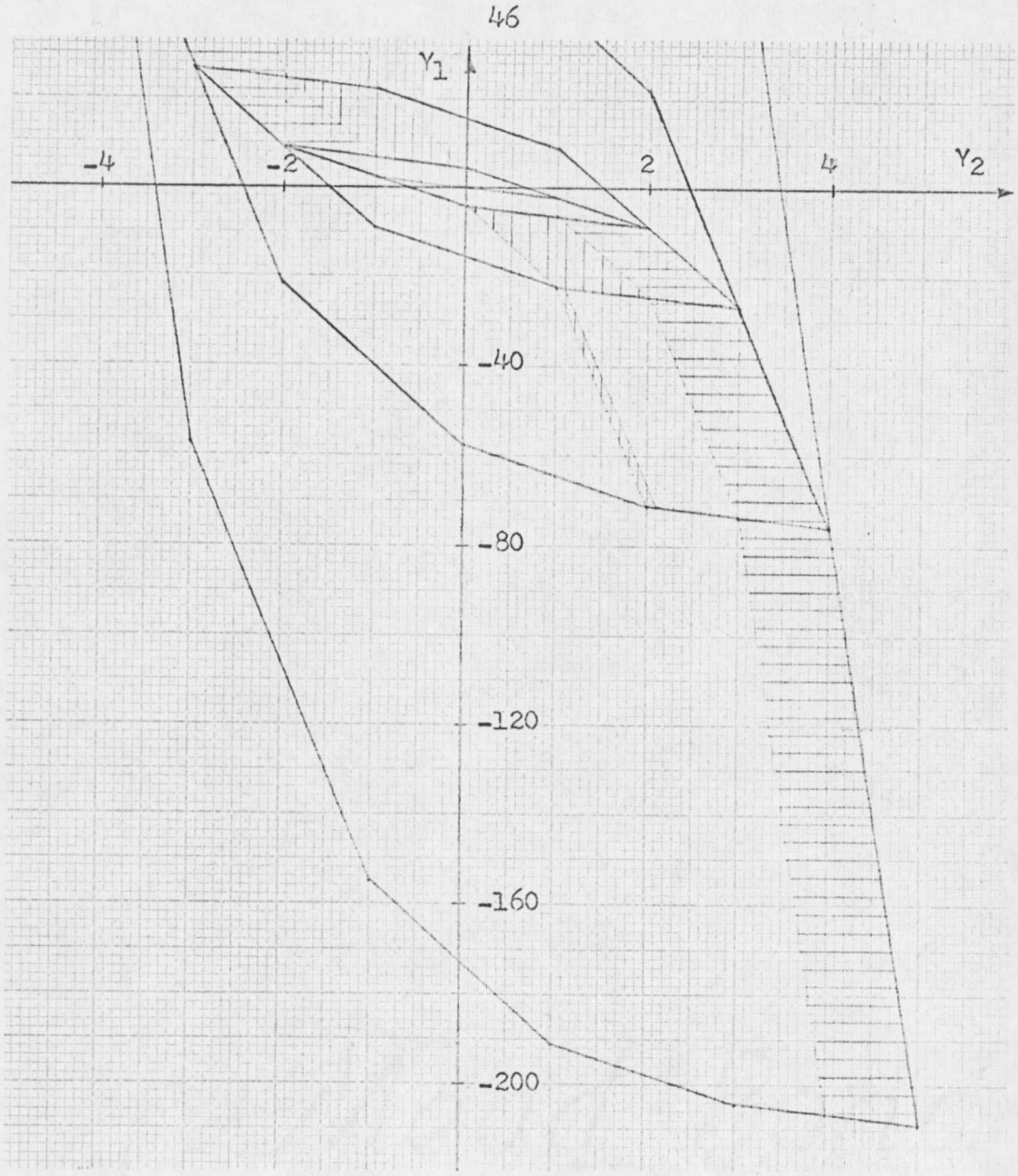

Figure 1.11. "Fixed" suboptimal regions for Desoer and Wing optimal strategy due to plant gain variations. Suboptimal regions due to negative variations are indicated by horizontal bars, while those due to positive variations are indicated by vertical bars. The plant is characterized by $1/s(s+1)$ and the sample period is one second (see Section 1.7).

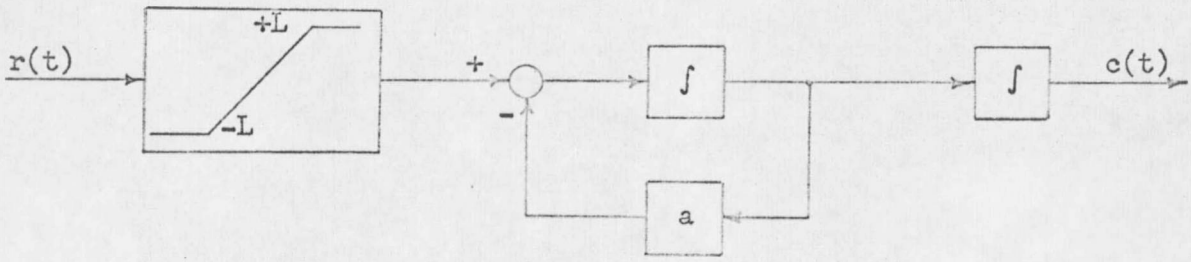

Figure 1.12. Saturating input system having a plant characterized by $1/s(s+a)$.

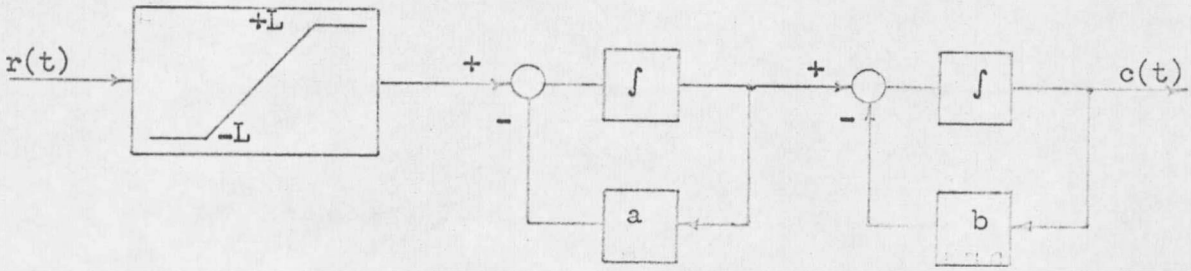

Figure 1.13. Saturating input system having a plant characterized by $1/(s+a)(s+b)$.

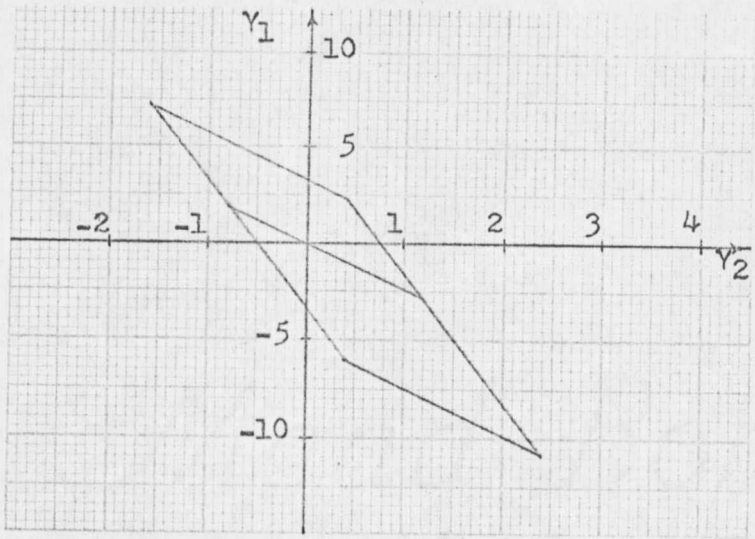

Figure 1.14. R_2' for ramp input with slope equal to 0.2 applied to a plant characterized by 1/s(s+a) and a one-second sample period.

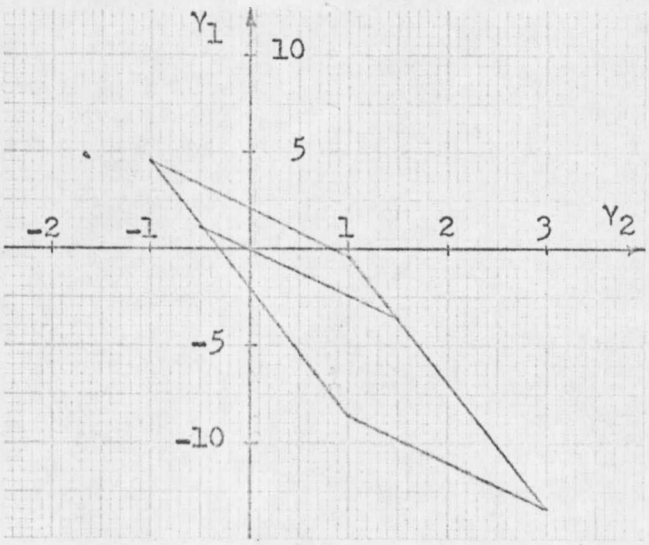

Figure 1.15. R_2' for ramp input with slope equal to 0.5 applied to a plant characterized by 1/s(s+a) and a one-second sample period.

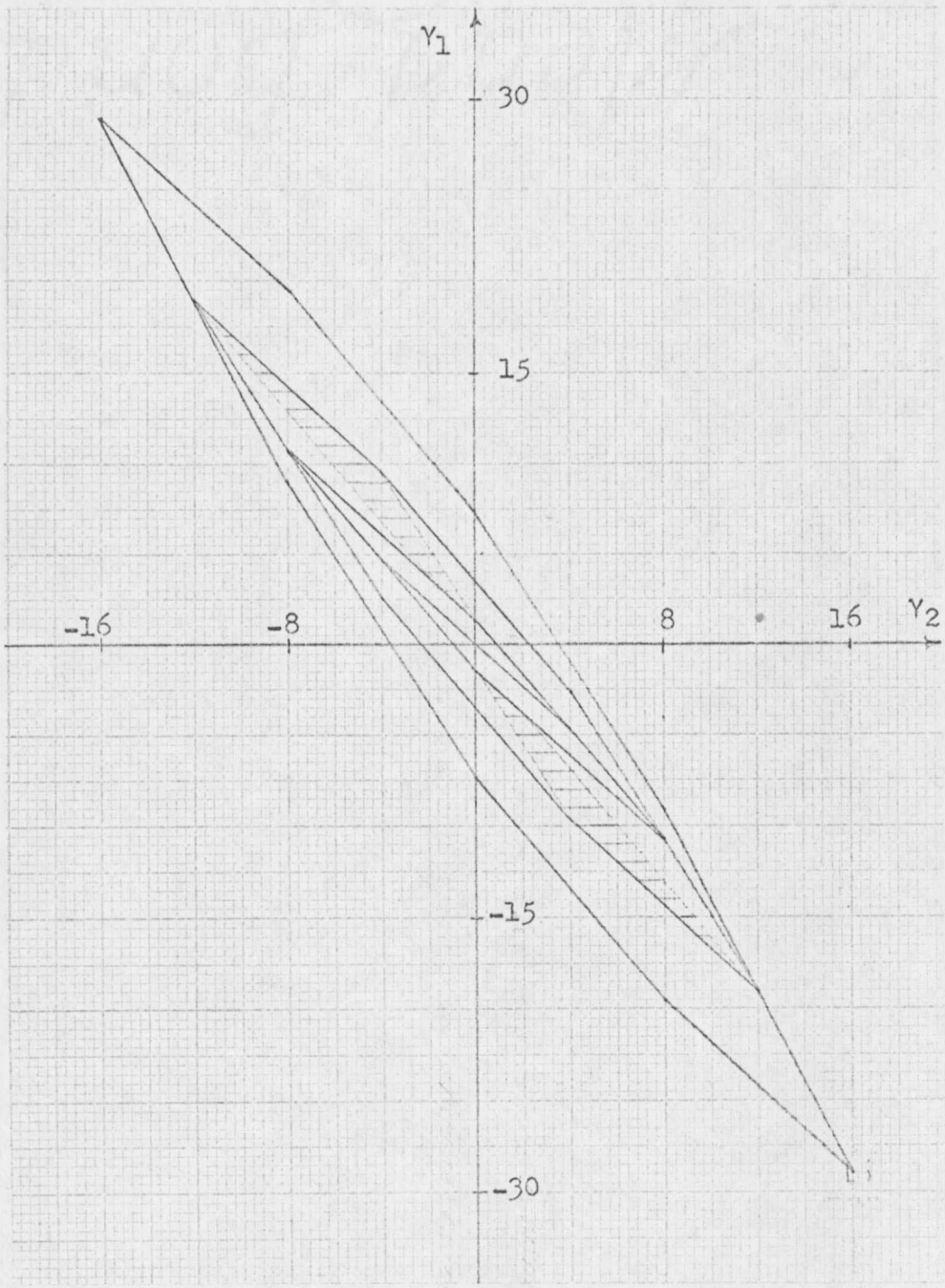

Figure 1.16. Region (horizontal bars) within R_3 for which + 1 or - 1 is suboptimal for the plant characterized by 1/s(s+0.25) and a one-second sample period (see Section 1.9).

CHAPTER 2

SOME RELATIONS BETWEEN THE DISCRETE AND THE CONTINUOUS MINIMUM-TIME PROBLEMS

2.1 INTRODUCTION

It is well known that if the input to a linear system is constrained by a magnitude bound, and it is desired that the state of the system be altered, the minimum-time control signal is of the bang-bang type. It follows that, if this control signal is additionally constrained so that it can change sign only at fixed intervals as in the sampled-data case, the time required to make the same change in system states will be at least as long as the bang-bang minimal time, and in general, will be longer.

It is intuitive that as the sample period approaches zero, the minimum time will decrease to that of the continuous-time controller. This intuition has been verified for a special case by Desoer and Wing [8] and in general by Neustadt [29]. It would certainly be of value if a general relationship could be found between the "continuous minimum-time" and the "discrete minimum-time". For example, the advantages of discrete control (such as several systems being controlled by one time-shared digital computer) would seem even more inviting if it were known that the time lost by the sampling process was known to be within a small, determinable bound.

Thus, the major effort of this chapter will be to justify the following <u>time-loss hypothesis</u>:

Consider a linear, second-order plant with real eigenvalues, no numerator dynamics, and a magnitude-limited input. If the state of this system can be driven from any initial state to the origin in a

minimum time of t_f seconds by a bang-bang controller, and if the plant is controllable in the sampled-data sense [10], this plant can be driven from the same initial state to the origin by a pulse-amplitude-modulated (PAM) controller with a fixed sample period T in no more than $t_f + 2T$ seconds.

Desoer and Wing's definition [10] of sampled-data controllability is that such a system is "controllable" if all possible states of the system can be forced to the origin by a realizable forcing function. With magnitude-limited control actions, Desoer and Wing's definition implies that the system's plant contains no poles in the right half of the s plane. However, even if the plant does have one or more poles with positive real parts, there is a bounded region containing states which can be forced to the origin. The theorems presented in this chapter will be proven for cases of both right- and left-half-plane poles, thus implying the veracity of the time-loss hypothesis for such bounded "controllable" regions.

The arguments used will be such that the above time-loss hypothesis is also credible for an n'th-order system; that is, the introduction of sampling can increase the minimum time to the origin by a maximum of nT seconds. The general continuous system to be considered is shown in Figure 2.1, while the general sampled-data system is in Figure 2.2. Desoer and Wing [8] show that there is no loss of generality if K and A in these figures are both normalized to one, since this merely alters the time scale. This normalization will be used throughout the

chapter.

In Section 2.2, it is proven that, for a first-order linear system, the introduction of sampling can increase the "bang-bang minimum time" by a maximum of T seconds. In Section 2.3, several theorems are derived relating the minimal isochrones of a second-order system to the R_k regions[1] for the same system. These theorems are shown to have some intrinsic value, and are also used to justify the time-loss hypothesis. The hypothesis is also proven for several special cases. In Section 2.4, a system with complex conjugate eigenvalues is studied, and it is shown why the above time-loss hypothesis is not true for this case. In Section 2.5, consideration is given to the possible application of the time-loss hypothesis to pulse-width-modulated systems.

2.2 FIRST-ORDER SYSTEMS

Suppose that a system is governed by the equation

$$\dot{c} = ac + u \quad , \quad |u| < 1 \tag{2.1}$$

where initially $c(0) = c_0$, and at the final time t_f, $c(t_f) = 0$.

First consider the continuous case. The bang-bang principle gives that the minimum-time control does not change sign. For one set of initial states, the optimal control is plus one for all t in the interval $[0, t_f]$. The solution to equation (2.1) is

1. The meanings of this terminology are given in Section 1.4.

$$c(t) = c(0)e^{at} + e^{at} \int_0^t e^{-a\tau} u(\tau)\, d\tau \tag{2.2}$$

Now assume that

$$t_f = kT + \theta T \quad , \quad 0 \leqslant \theta < 1 \tag{2.3}$$

where T is the sample period for the sampled-data controller to be considered later in this section and k is a positive integer or zero. Thus, for u = +1,

$$c(t_f) = 0 = c_0 e^{a(k+\theta)T} + e^{a(k+\theta)T} \int_0^{(k+\theta)T} e^{-a\tau}\, d\tau \tag{2.4}$$

Solving this equation in terms of the initial state c_0 gives

$$c_0 = \frac{e^{-a(k+\theta)T} - 1}{a} \tag{2.5}$$

Now consider the discrete time case, and suppose that a control signal of plus one is applied from t equals zero to t = kT seconds, and some control level u_k is applied from kT to (k+1)T seconds such that at the end of (k+1)T seconds, c equals zero. If it can be shown that the magnitude of u_k is less than one for this interval, the implication is that the discrete system will never take more than T seconds longer to regulate than does the continuous system. The so-

lution for the discrete case is

$$c(t_f) = 0 = c_0 e^{a(k+1)T} + e^{a(k+1)T} \left[\int_0^{kT} e^{-a\tau} d\tau + \int_{kT}^{(k+1)T} u_k e^{-a\tau} \, d\tau \right] \tag{2.6}$$

This equation can be solved for u_k, and c_0 can be replaced by the right-hand member of equation (2.5) since the initial states are assumed to be identical. This gives, after simplification,

$$u_k = \frac{e^{(1-\theta)aT} - e^{aT}}{1 - e^{aT}} \tag{2.7}$$

Equation (2.7) will be broken down into three cases.

Case 1: $a < 0$.

This implies that

$$1 - e^{aT} > 0 \tag{2.8}$$

for all values of T. Similarly,

$$e^{(1-\theta)aT} - e^{aT} > 0 \tag{2.9}$$

for all values of T. This implies that

$$u_k > 0 \tag{2.10}$$

Also,

$$0 < e^{(1-\theta)aT} < 1 \tag{2.11}$$

This means that the numerator of equation (2.7) is always less than the denominator. Thus

$$0 < u_k < 1 \tag{2.12}$$

Case 2: $a > 0$.

$$1 - e^{aT} < 0 \tag{2.13}$$

$$e^{(1-\theta)aT} - e^{aT} < 0 \tag{2.14}$$

Thus, both numerator and denominator of equation (2.7) are negative and

$$u_k > 0 \tag{2.15}$$

Also,

$$e^{(1-\theta)aT} > 1 \tag{2.16}$$

Therefore the numerator of (2.7) is less negative than the denominator. Thus,

$$0 < u_k < 1 \tag{2.17}$$

Case 3: $a = 0$.

Take the limit of equation (2.7) as "a" approaches zero. This gives

$$u_k = \theta \tag{2.18}$$

where, of course, θ is between zero and one.

It is obvious that an initial choice of minus one for the optimal control would give similar results. This concludes the proof of the time-loss hypothesis for first-order systems.

2.3 SECOND-ORDER SYSTEM WITH REAL EIGENVALUES

The time-loss hypothesis will be justified by relating the R_k regions of the discrete-time system to the kT-minimum isochrones of the continuous system. A t-minimum isochrone is the locus of all points in state space from which the origin can be reached in a minimum time of t seconds. In the remainder of this chapter, the word minimum will often be omitted, as all isochrones will be assumed to be of minimum time. In Section 2.3.1, several theorems regarding

the isochrones, R_k regions and the switching curve will be stated and proven for a plant having one pole at the origin in the s plane; that is, for the plant $1/s(s+a)$, with "a" nonzero. The time-loss hypothesis will then be proven for the special case, e^{aT} greater than two. In Section 2.3.2, the theorems will be extended to the case of two nonzero poles which are real and distinct; that is, for the plant $1/(s+\lambda_1)(s+\lambda_2)$, the time-loss hypothesis will be proven for $e^{\lambda_1 T}$ and $e^{\lambda_2 T}$ both greater than two. In Section 2.3.3, the special case of a double integrator will be considered, and the time-loss hypothesis will be proven for the special case when the sampling period is less than one second.

2.3.1 ONE EIGENVALUE AT ZERO

Suppose that for values of t between 0 and T, the system is governed by the differential equation

$$\ddot{c}(t) + a\dot{c}(t) = u \quad , \quad |u| \leq 1 \tag{2.19}$$

which is the same as equation (1.3). As in Chapter 1, a change of variables will be made, so that the working state equation is (1.6). The resulting r_k, as defined in Chapter 1, will be repeated here for convenience.

$$r_k = \begin{bmatrix} -e^{kaT}(1-e^{-aT})(1+a^2)^{\frac{1}{2}}/a^2 \\ T/a \end{bmatrix} \tag{2.20}$$

The following theorems may be more easily visualized by considering Figure 2.3, which portrays a finite region of the state space containing the switch curve, kT isochrones and R_k regions for equation (1.6), with "a" equal to plus one. Similar curves are displayed in Figure 2.4 for the plant 1/s(s-1), i.e. for "a" equals minus one. Also shown in Figure 2.4 are the boundaries of the controllable region for this plant. In this case, all states for which $|\gamma_1| < \sqrt{2}$ are controllable.

THEOREM 1: The slope of the switch curve at kT seconds from the origin is $-(1+a^2)^{\frac{1}{2}}e^{akT}$.

PROOF: Because the polygonal curve approaches the switch curve as T approaches zero [8], and because the r_k vector (2.20) forms the polygonal curve at $t_0 \equiv kT$, the slope of the switch curve may be found by holding kT fixed at t_0, and then taking the limit of the slope of the r_k vector as T approaches zero. Thus, letting γ_1 denote the upper element of r_k and γ_2 the lower element gives

$$\left.\frac{d\gamma_1}{d\gamma_2}\right|_{t_0} = \lim_{\Delta\gamma_2 \to 0} \left.\frac{\Delta\gamma_1}{\Delta\gamma_2}\right|_{t_0} = \lim_{T \to 0} \text{slope of } r_k\Big|_{t_0}$$

$$= \lim_{T \to 0} \frac{-(1+a^2)^{\frac{1}{2}}(1-e^{-aT})}{aT} e^{at_0}$$

$$= -(1+a^2)^{\frac{1}{2}}e^{at_0} = -(1+a^2)^{\frac{1}{2}}e^{akT} \qquad (2.21)$$

THEOREM 2: The kT isochrone passes through all the vertices of R_k.

PROOF: Neustadt [29] proves that for a linear n'th-order autonomous sampled-data system with real distinct eigenvalues, the time-optimal control sequence consists of n or fewer alternating "blocks" of + 1's and - 1's, which may be connected by pulses having values between minus one and plus one. However, as is evident from the method of construction of the R_k regions, the control sequence necessary to drive a state at an R_k vertex to the origin consists only of the alternating blocks of plus and minus one, with no intermediate pulses. For example, for a second-order system, the optimal control sequence for the two R_1 vertices are +1 and -1. Similarly, for R_2, they are +1, +1; +1, -1; -1, -1; and -1, +1. For the six vertices of R_3, they are +1, +1, +1; +1, +1, -1; +1, -1, -1; -1, -1, -1; -1, -1, +1; and -1, +1, +1. Thus, the sampled-data control sequence for the R_k vertices is identical to the bang-bang control of the minimum-time continuous controller. This, in turn, implies that each vertex of R_k lies on the kT minimum isochrone.

THEOREM 3: The outer boundary of R_k approaches the kT isochrone if T is allowed to approach zero while kT remains fixed.

PROOF: This is an immediate consequence of Theorem 2. Because kT is held fixed, the r_k vectors which form R_k become smaller and more numerous, until, in the limit, each point on the R_k boundary is

a vertex.

THEOREM 4: The slope of the switch curve is equal to the slope of one side of the isochrone at their intersection.

PROOF: One side of the R_k region intersecting the switch curve at kT seconds from the origin is bounded by r_k. The corresponding section of the polygonal curve is also bounded by r_k. Theorem 4 thus follows from Theorem 3 and the proof of Theorem 1.

THEOREM 5: All of the isochrones for a system have sides with an identical slope at corners. This corner slope is equal to the slope of r_1 as T approaches zero.

PROOF: This follows directly from Theorem 3 and from the fact that one side of each "corner" of an R_k region begins with r_1. For the running example, this slope is $-(1+a^2)^{\frac{1}{2}}$.

THEOREM 6: Let the corner of the kT isochrone correspond to the zeroth vertex of the R_k region. Then the slope of the kT isochrone at R_k's i'th vertex is $-e^{ai}(1+a^2)^{\frac{1}{2}}$, where the i'th vertex is counted counter-clockwise from the zeroth vertex if "a" is positive, and clockwise if "a" is negative.

PROOF: The proof may best be illustrated by considering the limiting case of the first vertex as T approaches zero. The slope of the continuous isochrone at that point can be found by finding the

slope of r_k as T approaches zero while holding kT equal to one. To clarify this, consider the running example, where T equals one, "a" equals one, and k equals three. Thus the slope of the three-second isochrone is being found at the point of its intersection with the first vertex of R_3. Table 2.1 gives the slopes of the r_k vectors intersecting this point as T decreases. The last entry in the table is found by use of equation (2.20), as follows:

$$\text{slope of } r_k = \frac{-e^{akT}(1 - e^{-aT})(1+a^2)^{\frac{1}{2}}}{aT} = \frac{-e^{kT}(1 - e^{-T})2^{\frac{1}{2}}}{T} \tag{2.22}$$

from which

$$\text{slope of isochrone at first vertex} = \lim_{T \to 0} \frac{-e^{1}(1 - e^{-T})2^{\frac{1}{2}}}{T}$$

$$= - 2^{\frac{1}{2}}\, e = -3.85 \tag{2.23}$$

The generalization from the first vertex to the i'th vertex is quite apparent. Note that the final vertex always corresponds to "i equals k" in Theorem 6. This, in turn, corresponds to the slope of the switch curve, as is indicated by Theorems 1 and 4.

THEOREM 7: The slope of one side of the kT isochrone at its corner has a value between the slope values of the adjoining R_k and R_{k+1} sides.

PROOF: Consider two cases: "a" positive, and "a" negative.

Figure 2.3 depicts a system with "a" positive, while Figure 2.4 depicts the state space for a system with negative "a".

Case 1: Positive "a".

First it is shown that the slope of the kT isochrone is less negative than the slope of r_{k+1}, which forms the adjoining side of R_{k+1}. That is,

$$-(1+a^2)^{\frac{1}{2}} \frac{(1 - e^{-aT})}{aT} \; e^{(k+1)aT} < -(1+a^2)^{\frac{1}{2}} e^{kaT} \tag{2.24}$$

which gives

$$\frac{e^{aT} - 1}{aT} > 1 \tag{2.25}$$

The left side of (2.25) may be written in series form as $1 + \frac{aT}{2!} + \frac{(aT)^2}{3!} + \ldots$, which is obviously greater than one.

Next it is shown that the slope of the kT isochrone is more negative than the slope of r_k, which forms the adjoining side of R_k. That is

$$-(1+a^2)^{\frac{1}{2}} \frac{(1 - e^{-aT})}{aT} \; e^{kaT} > -(1+a^2)^{\frac{1}{2}} \; e^{kaT} \tag{2.26}$$

or

$$\frac{1 - e^{-aT}}{aT} < 1 \tag{2.27}$$

$(1 - e^{-aT})/aT$ has no critical points. At aT equals zero, its value is + 1. As aT becomes greater than zero, the function is monotonically decreasing, approaching zero in the limit.

Case 2: Negative "a".

This case is proven by simply replacing "a" by minus "a" in (2.25) and (2.27) and also reversing the direction of the inequalities.

The preceding theorems (and their generalizations, which are given in Section 2.3.2) will be used to justify the time-loss hypothesis. However, they also have a certain amount of intrinsic value, since they shed some light on isochrone configurations and their relation to the switch curve and the R_k regions. In addition, Theorems 2, 3, and 6 provide a handy method of sketching an isochrone to any desired degree of accuracy. It is much easier to approximate the kT isochrone with the R_k region and the proper slope relationships than it is to find and use an exact analytic expression for the isochrone.

Two other theorems which will be needed are proven by Athans and Falb [1]. Simply stated, they are 1) the minimum isochrones are convex; 2) the minimum isochrones increase their "distance" from the origin in a "smooth" manner with increasing time.

The general justification of the time-loss hypothesis proceeds as follows for the second-order case. Figure 2.3 will be used for reference. First, make the assumption that any kT isochrone does not extend into region R_{k+2}. (For example, in Figure 2.3, the two-second

isochrone touches, but does not enter R_4.) Note that any point in the state space "external" to the kT isochrone takes more than kT seconds to reach the origin if controlled by a continuous controller. However, those points within R'_{k+2} can reach the origin in (k + 2)T seconds or less, by discrete control. Thus, if the original state location is in R'_{k+2} but external to the kT isochrone, the hypothesis is proven. Now consider the states between the kT and (k-1)T isochrones. These states all take more than (k-1)T seconds to reach the origin by continuous time control. This region is all within R'_{k+1}; that is, less than (k+1)T seconds are required by the discrete controller. Hence the time-loss hypothesis still holds. Since the same arguments may be made about the (k-1)T isochrone, the hypothesis is proven, subject to the validity of the initial assumption in this paragraph.

The principal argument for the validity of this assumption is Theorem 7. It is obvious that the corners of the kT isochrone will always touch R_{k+2}. The geometry of the isochrone in this vicinity thus seems to be the most critical. However, Theorem 7 gives that the kT isochrone at its corners will always be sloping away from the outer boundary of R_{k+1} (which is also the inner boundary of R_{k+2}). For points on the kT isochrone away from the corners, Theorem 2 strongly indicates that they will be "drawn away from" R_{k+1}. Another factor is that the magnitude of T should not matter in any general proof of the time-loss hypothesis since varying T doesn't really alter the various geometrical shapes in the state space, only their scale. Inspection of the

construction of the R_k regions for higher-order systems (see Figures 10, 11, 12 and 13 in Desoer and Wing [9]) indicates that similar arguments can be used to justify the extended time-loss hypothesis for those systems.

The verbal statements of the preceding proof will now be examined in detail for the running example of this section. The approach is to divide the lower side of the kT isochrone into segments, and to show that each segment does not intersect R_{k+2}. The isochrone segments are formed by moving away from each R_k vertex by half the γ_2 distance (T/a) to the next R_k vertex. To put it another way, start at a corner of the isochrone and mark off a new segment each time a horizontal distance of T/a is reached. Now consider those segments proceeding out of the R_k vertices to the right. It is easily seen from a slight generalization of Theorem 7, and from the convexity property of isochrones, that the slope of the isochrone segment is such that it moves away from the R_{k+1} segment directly below it. Thus, these segments will never intersect. Now consider the isochrone segments going to the left of the R_k vertices. In this case a linear approximation of the isochrone segment will be used to show nonintersection. The i'th vertex of R_k will be assigned a reference value of zero. Thus,

$$\gamma_1 \text{ (kT isochrone segment to left of i'th vertex)} > -e^{ai}(1+a^2)^{\frac{1}{2}}\left(-\frac{T}{a}\right) \tag{2.28}$$

for i = 0, 1, 2, ... , k-1, and it must be shown that this value is

greater than the value of the R_{k+1} boundary segment directly below (see Figure 2.3). Thus,

$$\gamma_1 \text{ (kT isochrone segment to left of i'th vertex)} > \gamma_1(r_k) - \gamma_1(r_i) - \alpha \tag{2.29}$$

where α is the portion of $\gamma_i(r_i)$ which is directly below the isochrone segment. Thus,

$$p\, e^{ai}\, (1+a^2)^{\frac{1}{2}} \frac{T}{a} > -e^{kaT}(1 - e^{-aT}) \frac{(1+a^2)^{\frac{1}{2}}}{a^2} + (1+p)\, e^{iaT}(1 - e^{-aT}) \frac{(1+a^2)^{\frac{1}{2}}}{a^2}, \quad 0 < p < 1 \tag{2.30}$$

or

$$-p\, e^{ai} < [\, e^{kaT} - (1 + p)\, e^{iaT}\,](\frac{1 - e^{-aT}}{aT}) \tag{2.31}$$

In the worst case, k equals i + 1. Thus,

$$-p\, e^{ai} < e^{iaT}\, [\, e^{aT} - (1 + p)](\frac{1 - e^{-aT}}{aT}) \tag{2.32}$$

Again, for the worst case, let p equal one on the right-hand side of the inequality while p equals zero on the left-hand side. Thus

$$0 < e^{iaT}(e^{aT} - 2)\left(\frac{1 - e^{-aT}}{aT}\right) \tag{2.33}$$

Note that e^{iaT} is always positive; $(1 - e^{-aT})/aT$ is also positive if "a" is greater than zero. Thus, to satisfy inequality (2.33), e^{aT} must be greater than two. This means that the hypothesis is proven for products of aT greater than .693. However, it does not mean that the time-loss hypothesis is invalid for other values of aT, only that the approximations made in the above development are not accurate enough.

2.3.2 DISTINCT, NONZERO EIGENVALUES

The plant of interest is $1/(s-\lambda_1)(s-\lambda_2)$, where neither λ_1 nor λ_2 are zero, and they have different values between plus and minus one. The corresponding state equation is

$$\begin{bmatrix} \gamma_1(T) \\ \\ \gamma_2(T) \end{bmatrix} = \begin{bmatrix} e^{\lambda_1 T}\gamma_1(0) \\ \\ e^{\lambda_2 T}\gamma_2(0) \end{bmatrix} + \begin{bmatrix} \dfrac{(1 - e^{\lambda_1 T})}{\lambda_1(\lambda_1-\lambda_2)} \\ \\ \dfrac{(1 - e^{\lambda_2 T})}{\lambda_2(\lambda_2-\lambda_1)} \end{bmatrix} u(0) \tag{2.34}$$

The eigenvectors have not been normalized in this case. The r_k vector is found to be

$$r_k = \begin{bmatrix} \dfrac{e^{-k\lambda_1 T}(e^{\lambda_1 T} - 1)}{\lambda_1(\lambda_1 - \lambda_2)} \\ \\ \dfrac{e^{-k\lambda_2 T}(e^{\lambda_2 T} - 1)}{\lambda_2(\lambda_2 - \lambda_1)} \end{bmatrix} \tag{2.35}$$

Theorems 2, 3, 4 and 5, and their proofs, hold for the above system. Theorem 1 may be derived to give the slope of the switch curve at t_0 as

$$\left.\frac{d\gamma_1}{d\gamma_2}\right|_{t_0} = -e^{(\lambda_2-\lambda_1)t_0} = -e^{(\lambda_2-\lambda_1)kT} \tag{2.36}$$

Similarly, Theorem 6 may be modified to show that the slope of the kT isochrone at the i'th vertex of R_k is $-e^{(\lambda_2-\lambda_1)i}$. The i'th vertex is counted clockwise if λ_1 is less than λ_2. The statement of Theorem 7 still holds, but, because of its importance, it will be proven for this case. It is necessary to prove either the upper or lower set of inequalities in (2.37).

$$\text{slope of } r_k \lessgtr \text{ slope of kT isochrone } \lessgtr \text{ slope of } r_{k+1} \tag{2.37}$$

In this case,

$$\frac{\lambda_2(\lambda_2-\lambda_1)e^{-\lambda_1 kT}(e^{\lambda_1 T}-1)}{\lambda_1(\lambda_1-\lambda_2)e^{-\lambda_2 kT}(e^{\lambda_2 T}-1)} \lesseqgtr - e^{(\lambda_2-\lambda_1)kT} \lesseqgtr \frac{\lambda_2(\lambda_2-\lambda_1)e^{-\lambda_1 kT}(1-e^{-\lambda_1 T})}{\lambda_1(\lambda_1-\lambda_2)e^{-\lambda_2 kT}(1-e^{-\lambda_2 T})} \tag{2.38}$$

which reduces to

$$\frac{\lambda_2(e^{\lambda_1 T}-1)}{\lambda_1(e^{\lambda_2 T}-1)} \lesseqgtr 1 \lesseqgtr \frac{\lambda_2(1-e^{-\lambda_1 T})}{\lambda_1(1-e^{-\lambda_2 T})} \tag{2.39}$$

It can be assumed without any loss in generality that λ_2 is greater than λ_1. The proof is broken down into three cases.

Case 1: Both eigenvalues positive.

The right side of (2.39) can be written as

$$\frac{(1 - e^{-\lambda_1 T})/\lambda_1 T}{(1 - e^{-\lambda_2 T})/\lambda_2 T} \tag{2.40}$$

which is greater than one since the function $(1-e^{-x})/x$ is monotonically decreasing.

Also, the left side of (2.39) may be written as

$$\frac{\sum_{k=1}^{\infty} \lambda_1^{k-1} T^k/k!}{\sum_{k=1}^{\infty} \lambda_2^{k-1} T^k/k!} \tag{2.41}$$

which is obviously less than one.

Case 2: Both eigenvalues negative.

For this case, change the sign of λ_i, and assume λ_i is positive. (2.39) may thus be written as

$$\frac{\lambda_2(1-e^{-\lambda_1 T})}{\lambda_1(1-e^{-\lambda_2 T})} \lesseqgtr 1 \lesseqgtr \frac{\lambda_2(e^{\lambda_1 T}-1)}{\lambda_1(e^{\lambda_2 T}-1)} \tag{2.42}$$

which is really the same as Case 1.

Case 3: λ_2 positive, λ_1 negative.

Change the sign on λ_1 and assume λ_1 is positive. (2.39) may be written

$$\frac{(1-e^{-\lambda_1 T})/\lambda_1 T}{(e^{\lambda_2 T}-1)/\lambda_2 T} \lesseqgtr 1 \lesseqgtr \frac{(e^{\lambda_1 T}-1)/\lambda_1 T}{(1-e^{-\lambda_2 T})/\lambda_2 T} \tag{2.43}$$

The statements following inequality (2.25) show that $(e^{\lambda_i T}-1)/\lambda_i T$ is greater than one, while those following inequality (2.27) show that $(1-e^{-\lambda_i T})/\lambda_i T$ is less than one. Thus, the upper set of inequality signs holds in (2.43).

This concludes the proof of Theorem 7 for two nonzero, distinct eigenvalues. If the eigenvalues are not distinct, a similar proof can be developed.

The time-loss hypothesis is easy to prove for the special case of both $\lambda_1 T$ and $\lambda_2 T$ less than -.692. In this case, each of the r_k components in equation (2.35) doubles as k is incremented by one. As Figure 2.5 illustrates, the kT isochrone to the right of the R_k vertices will always be sloping away from the segment of R_{k+1} directly below it. This is an immediate consequence of Theorems 2, 6 and 7. Once again, it must be emphasized that the proof of the hypothesis for this special case does not indicate that the hypothesis fails for other cases -- for example, see Figure 2.6 which illustrates the state space for the plant $1/(s+1)(s-1)$.

2.3.3 DOUBLE INTEGRATOR

Here the plant is $1/s^2$. The state equation is

$$\begin{bmatrix} c(T) \\ \dot{c}(T) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c(0) \\ \dot{c}(0) \end{bmatrix} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} u(0) \tag{2.44}$$

and the r_k vector is

$$r_k = \begin{bmatrix} (k-\frac{1}{2})T^2 \\ -T \end{bmatrix} \tag{2.45}$$

All of the theorems previously developed, or variations of them, hold for this system. In particular, the slope of the kT isochrone at the i'th vertex of R_k can be shown to be -i. This leads to a proof (sim-

ilar to that in Section 2.3.1) of the hypothesis for another special case. The inequality (corresponding to (2.29)) which must be satisfied is

$$piT > -(k-\tfrac{1}{2})T^2 + (1+p)(i-\tfrac{1}{2})T^2,$$

$$i = 0, 1, 2, \ldots, k-1; \; 0 < p \leq 1 \tag{2.46}$$

Once again, for the worst case, let $k = i+1$. Thus,

$$-pi < \left[1 - p(i-\tfrac{1}{2})\right] T$$

$$-i < \left(\frac{1}{p} - i + \tfrac{1}{2}\right) T$$

$$i(T - 1) < \left(\frac{1}{p} + \tfrac{1}{2}\right) T \tag{2.47}$$

which is satisfied for all i and p if T is less than or equal to one.

2.4 COMPLEX CONJUGATE EIGENVALUES

In this section it will be shown that the time-loss hypothesis is not true for a plant with complex conjugate poles. The simplest plant of this type, the harmonic oscillator, will be analyzed. The transfer function is $1/(s^2+w^2)$, and in particular, w will be set equal to one in the running example. The differential equation for this system is

$$\ddot{c}(t) + c(t) = u(t) \quad , \quad -1 \leqslant u(t) \leqslant 1 \tag{2.48}$$

It will be advantageous to summarize initially certain properties of the continuous-time optimal control for the plant $1/(s^2+1)$. This material is covered in detail by Athans and Falb [1].

The first factor of interest is that the time-optimal control is piecewise constant, and must switch between the values plus one and minus one. This is also a property of the plants with real poles, which were considered earlier. The second factor of interest is that there is no upper bound on the number of switchings of the time-optimal control. A third property is that the time-optimal control can remain constant for no more than π/w seconds. Properties two and three stand in marked contrast to the second-order system with real eigenvalues whose time-optimal control has, at most, one switching, but has no limit on the time spent at either signal level.

The switch curve for this system is also quite different from the real eigenvalue system. It is formed by a series of semicircles centered along the $c(t)$ axis, as is indicated in Figure 2.7. Some typical optimal trajectories are also illustrated. Note that only the two semicircles closest to the origin form portions of optimal trajectories.

Within the circle with radius $2/w$ about the origin, the minimum isochrones are composed of <u>two</u> circular arcs. (This circle is, itself,

the π/w isochrone.) These arcs meet at "corners", thus resembling the isochrones previously discussed. External to this circle however, the isochrones are formed by four circular arcs and are differentiable everywhere, i.e. they do not have corners. An additional property is exhibited by the $i\pi/w$ isochrones, where i is a positive integer: in this case, the isochrones are circles about the origin of radius 2i.

With this background, it becomes fairly evident that the time-loss hypothesis discussed previously for real eigenvalue systems should not be expected to hold for the harmonic oscillator, or for any system having similar time-optimal controls; i.e. those with complex conjugate eigenvalues. The reasoning behind this is as follows. As the initial state of the system is moved farther from the origin, the continuous minimum-time control will require more switchings, with the length of time between switchings being fixed. However, the sampled-data controller can only allow switchings to occur at the sampling instants. Since the sample period will not, in general, be the same as the time between continuous-time switchings, it is reasonable that, as more switchings occur, it is less likely that the discrete minimum time will be close to the continuous minimum time.

Definite counter examples to the hypothesis are illustrated in Figure 2.8, in which the 1, 2, 3 and 4 second isochrones are drawn along with R_1, R_2, R_3, R_4 and part of R_5 for a sample period of one second. The r_k vector in this case is

$$r_k = \begin{bmatrix} -.46 \cos k + .842 \sin k \\ -.46 \sin k - .842 \cos k \end{bmatrix} \tag{2.49}$$

The sampled-data minimum-time control sequences for the R_k vertices are also listed, with + indicating a plus one signal, and - indicating a minus one signal. Note that the first counter examples occur near those vertices of R_4 at which the control sequences ++++ and ---- are optimal. The areas external to R_4 but internal to the three-second isochrone contain states which take five seconds to reach the origin with discrete control and less than three seconds to reach the origin with continuous control. Thus, for a second-order system, the time-optimal discrete controller takes more than two seconds longer than the time-optimal continuous controller for certain initial states. An important point to be considered is that the R_4 vertices mentioned above do not lie on the four-second isochrone. This is, of course, in contrast with real-eigenvalue systems, in which all R_k vertices lie on the kT isochrones. Note that the vertices of R_1, R_2 and R_3, have no more than one switching and never correspond to control sequences staying plus one or minus one for more than π seconds. Thus they correspond to continuous optimal control sequences. On the other hand, the ++++ and ---- vertices of R_4 have the same sign for four seconds and are thus suboptimal in the continuous sense. It is apparent that as k increases, fewer of the R_k vertices will lie on the minimum isochrones.

In fact, for k larger than six (and T equal to one second) none of the R_k vertices lie on the kT isochrones, but gradually "shrink away" from them. This indicates that perhaps a relationship exists between continuous optimal time, the number of switchings, and the discrete optimal time for a system with no restrictions in the eigenvalues.

2.5 PULSE-WIDTH-MODULATED SYSTEMS

Polak [32] approaches minimum-time control of second-order pulse-width-modulated sampled-data systems from the same general point of view as Desoer and Wing. The thing of interest here is that the state space is divided into R_k regions which have the same significance as in the amplitude-modulation case.

Basically, the method of construction of the R_k regions is as follows. First, the switch curve for the continuous-time controller is divided into segments whose ends correspond to specific minimum times from the origin. Each of these times must be an integral multiple of the sample period. R_1 is the section of the switch curve (on each side of the origin) between minimum times zero and T. The outer boundary of R_2 is traced by the 2T point of the switch curve as the segment of the switch curve from T to 2T is "slid" around the perimeter of R_1. Similarly, R_k is formed by sliding the section of the switch curve between (k-1)T and kT around the perimeter of R_{k-1}. Figure 2.9 shows several R_k regions and kT isochrones for a double-integrator plant. Note that, for the region of the state space which

is shown, the time-loss hypothesis stated in Section 2.1 holds. Of course, no general conclusions can be drawn, but one important factor is evident: as in the PAM case, the most critical areas of the state space are in the regions where R_k touches R_{k+1}. However, it is obvious that the kT isochrone lies between R_k and R_{k+1} in these regions. This follows immediately from Theorem 4, which gives that one side of each isochrone is tangent to the switch curve, and from the symmetry of the switch curve about the origin. In other words, if the kT isochrone is "convex" at its tangent point to the switch curve, the critical R_{k+1} boundary is "concave" at that point. This relationship is, of course, similar to Theorem 7, which was the principal argument for justifying the time-loss hypothesis for the PAM controller. Of course, for first-order systems, the continuous minimum time is identical with the PWM minimum time.

2.6 CONCLUSIONS

In this chapter, a time-loss hypothesis was proposed which related the minimum control time for a real eigenvalue system with a continuous-time controller to the minimum control time for the same system with a discrete-time controller. The hypothesis was investigated in detail for PAM control of first- and second-order systems, resulting in a proof of the hypothesis for first-order systems and certain special cases of second-order systems. In addition, substantial evidence is presented supporting the hypothesis for any controllable second-order system with real eigenvalues. In the process, seven theorems are stated and pro-

ven which relate the switch curve, minimum isochrones, and R_k regions. A system with complex conjugate eigenvalues is also studied to provide further insight into the continuous versus discrete controller relationships. Finally, a brief introduction to minimum-time PWM control is given, and the feasibility of extending the hypothesis to PWM systems is indicated.

The need for future research in this area is evident. The information presented in this chapter indicates the possibility of a significant relationship between the minimum control times taken by a continuous controller and a PAM or PWM controller, and it would certainly be an important result if such a relationship could be proven in general.

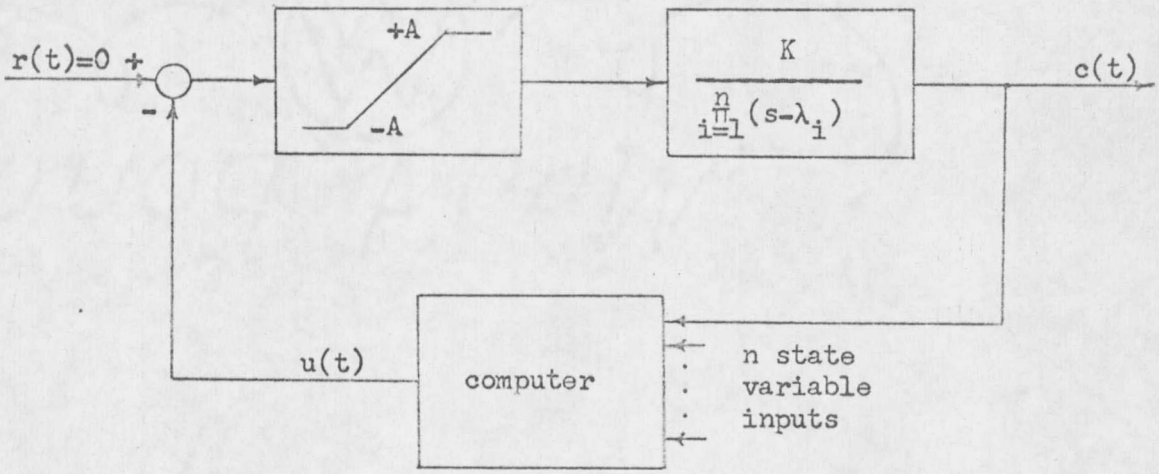

Figure 2.1. Continuous minimum-time system.

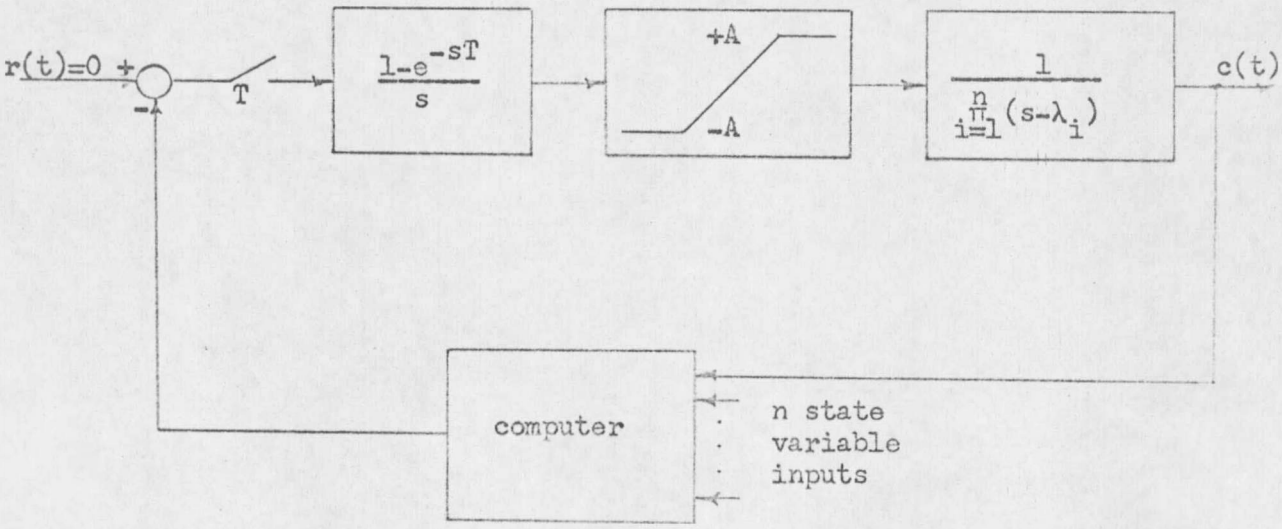

Figure 2.2. Sampled-data minimum-time system.

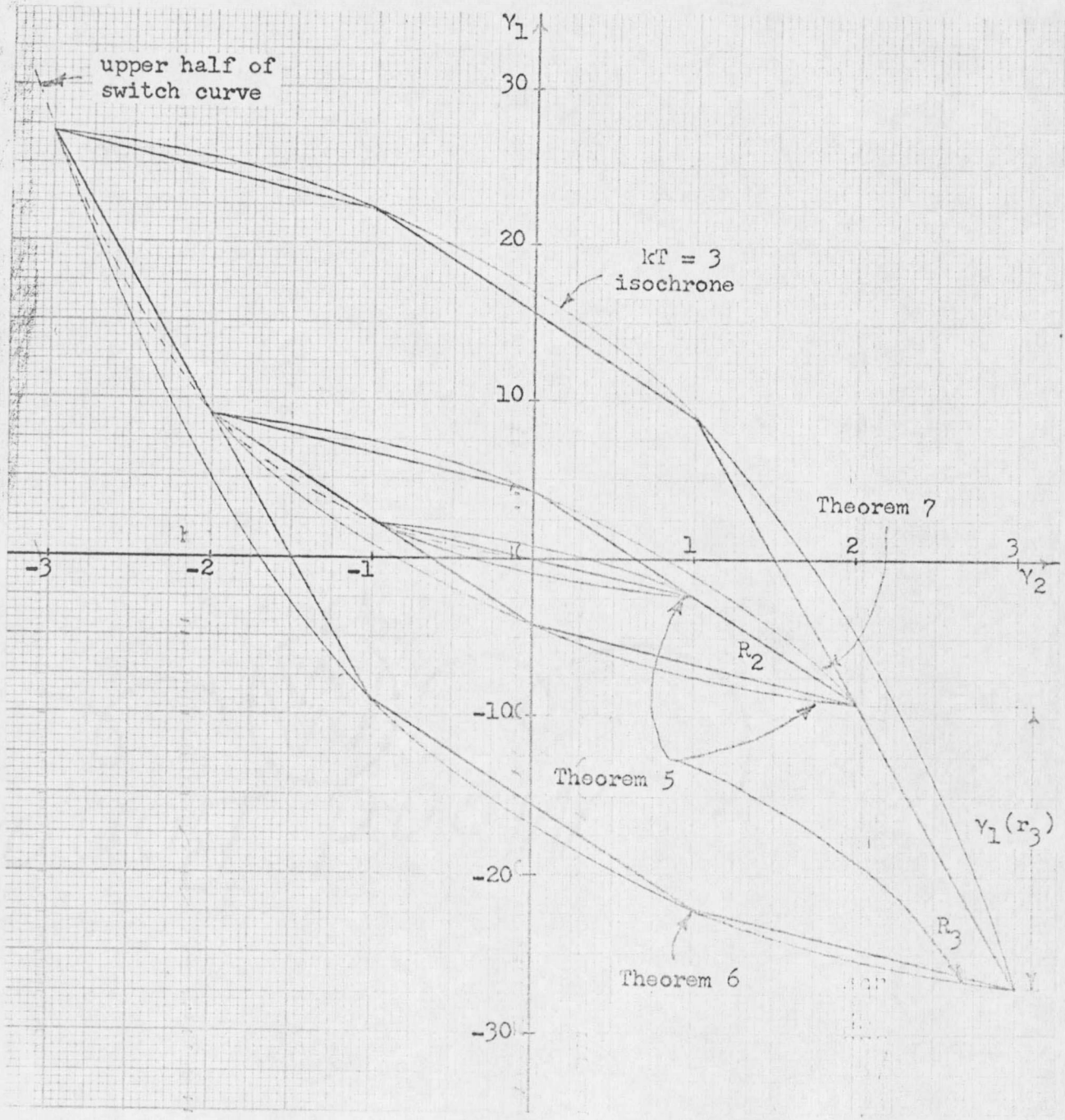

Figure 2.3. State space for plant characterized by $1/s(s+1)$ and one-second sample period (see equation (1.6)).

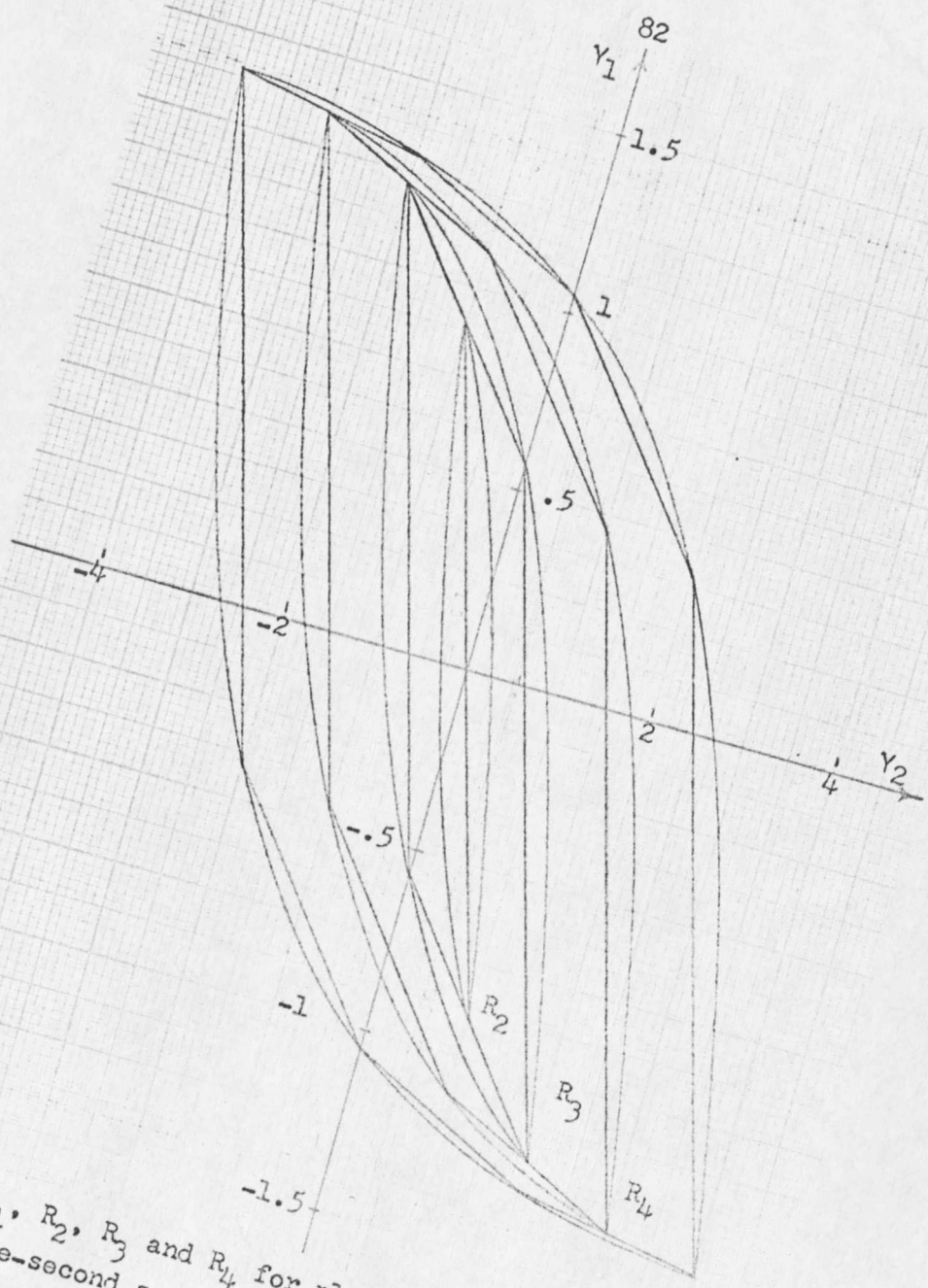

Figure 2.4. R_1, R_2, R_3 and R_4 for plant characterized by $1/s(s-1)$ and one-second sample period. Also, minimum isochrones for one, two, three and four seconds are shown. All initial states for which $|Y_1| > \sqrt{2}$ are uncontrollable.

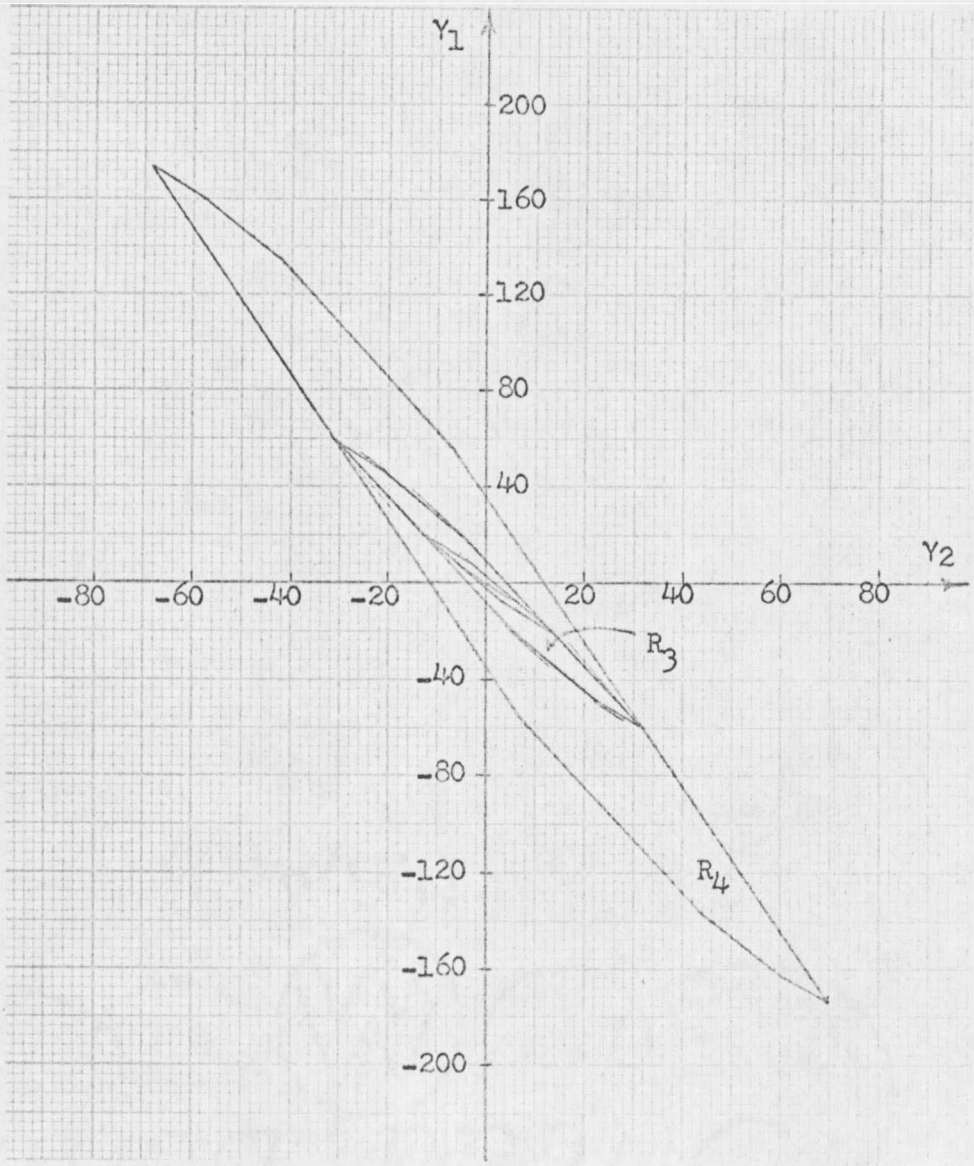

Figure 2.5. R_1, R_2, R_3 and R_4 for the plant characterized by $1/(s+0.693)(s+1)$ with a one-second sample period. Slopes of the three-second isochrone are also shown (see Section 2.3.2).

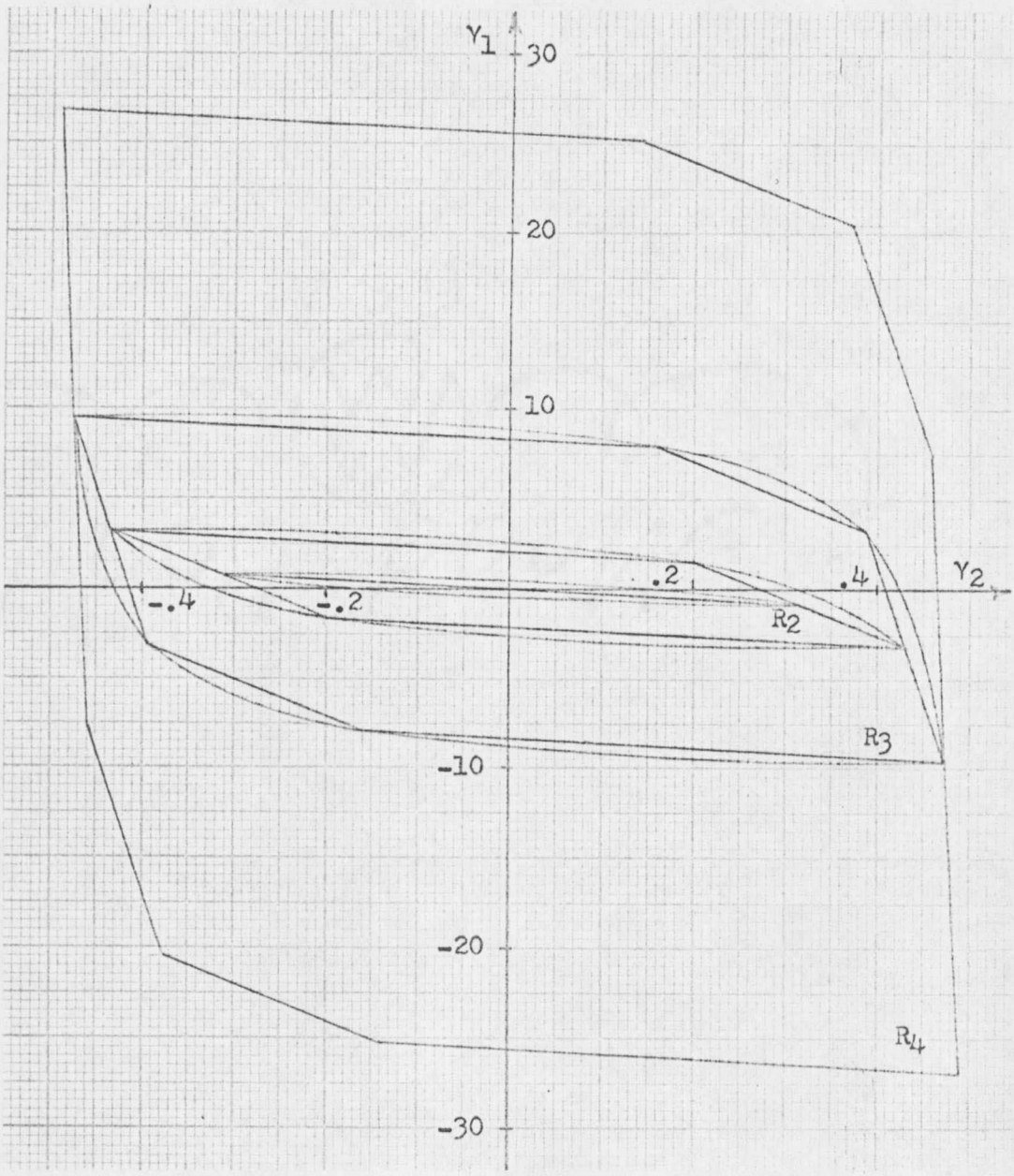

Figure 2.6. R_1, R_2, R_3 and R_4 for the plant characterized by $1/(s-1)(s+1)$ and a one-second sample period. Also, minimum isochrones for one, two and three seconds are shown.

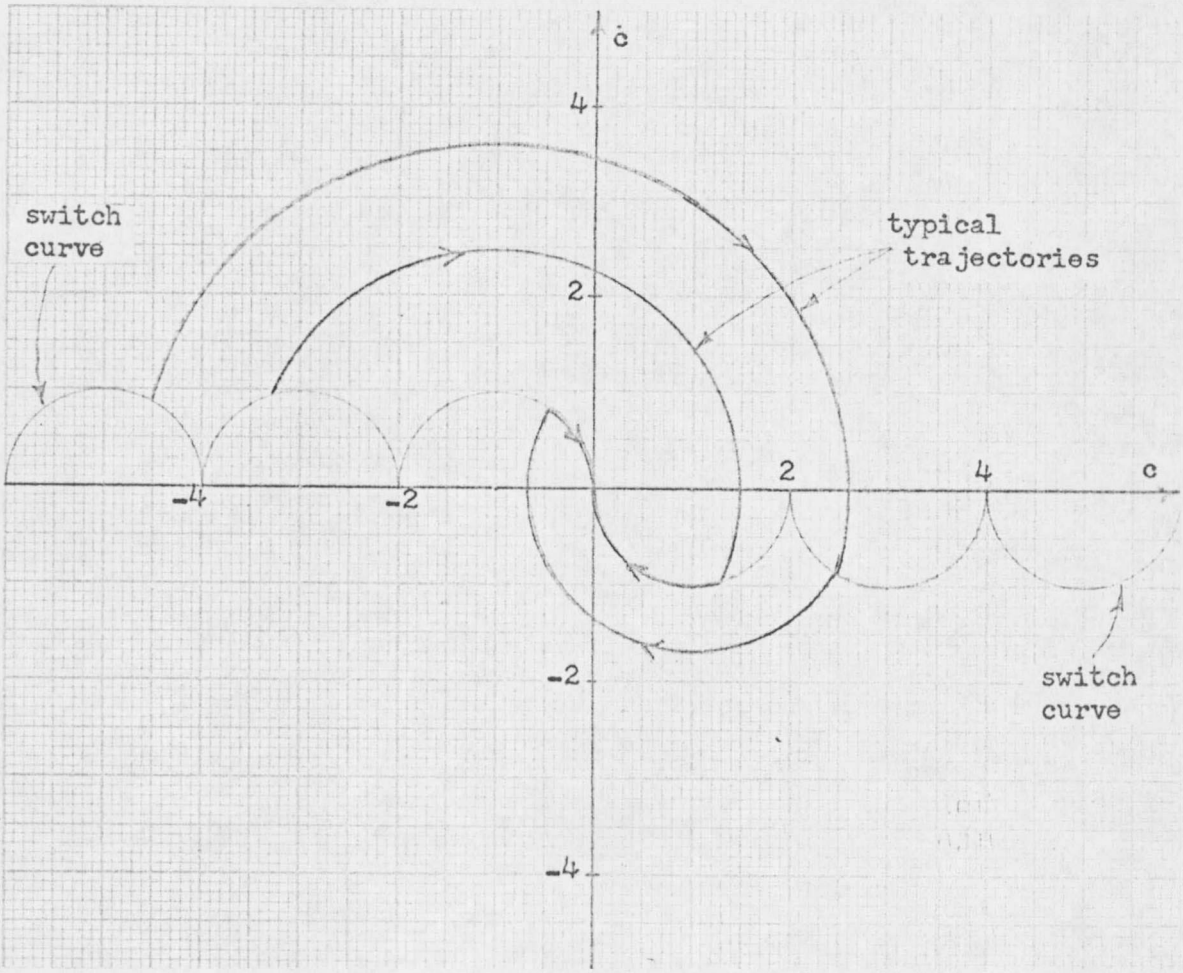

Figure 2.7. Switch curves and two typical trajectories corresponding to the plant characterized by $1/(s^2+1)$.

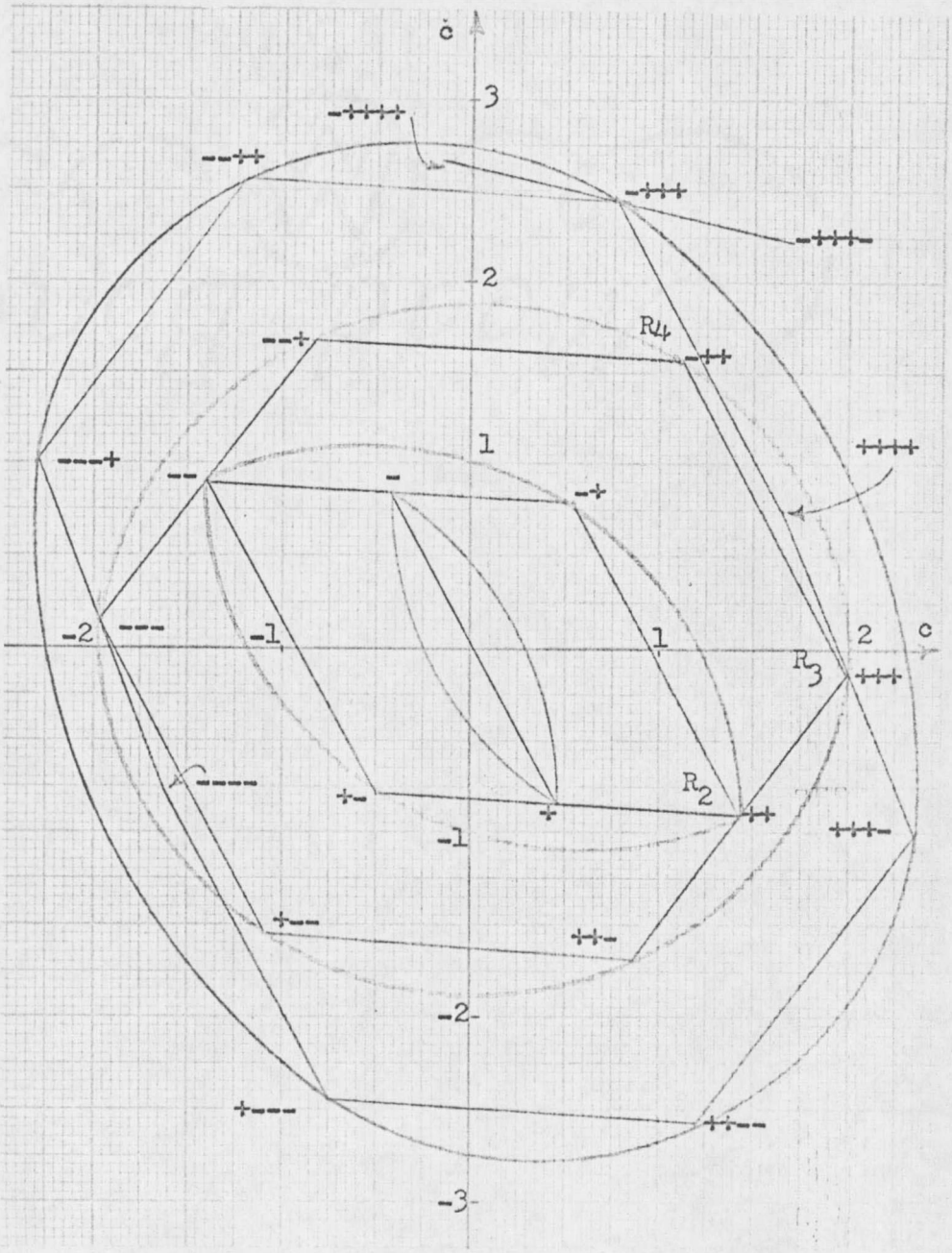

Figure 2.8. R_1, R_2, R_3, R_4 and a segment of the boundary of R_5 for the plant characterized by $1/(s^2+1)$. The sample period is one second. The vertices are labeled with their proper minimum-time control sequence. Also shown are the minimum isochrones for one, two, three and four seconds (see Section 2.4).

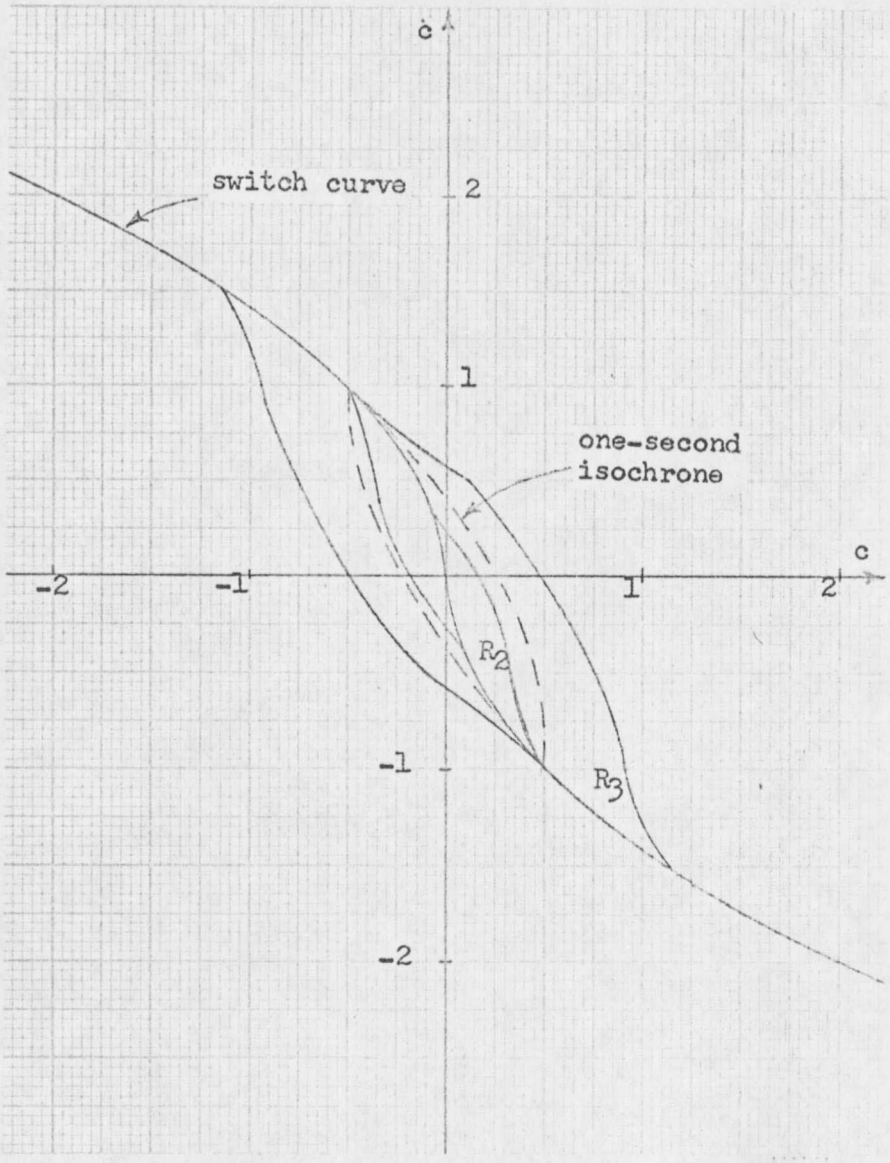

Figure 2.9. R_1, R_2 and R_3 for PWM control of the plant characterized by $1/s^2$. The sample period is 0.5 seconds. Also shown is the one-second minimum isochrone.

Table 2.1. Data illustrating a method of finding the slope of the three-second isochrone at the first vertex of R_3 (see Theorem 6).

T	R_k	adjoining r_k	slope of adjoining r_k	adjoining r_{k+1}	slope of adjoining r_{k+1}
1	R_3	r_1	-2.43	r_2	-6.61
0.5	R_6	r_2	-3.03	r_3	-4.98
0.25	R_{12}	r_4	-3.40	r_5	-4.36
0.125	R_{24}	r_8	-3.62	r_9	-4.10
0	R_∞	r_∞	-3.85	$r_{\infty+1}$	-3.85

CHAPTER 3

DEADBEAT RESPONSE TO PARABOLIC INPUTS WITH MINIMUM-SQUARED-ERROR RESTRICTIONS ON RAMP AND STEP INPUTS

3.1 INTRODUCTION

It is well known [35, 21] that a digital controller (D(z), Figure 3.1) can be designed for any physically realizable plant (G(s), Figure 3.1) such that, for certain prototype inputs, the output will follow the input exactly after some minimum number of sample periods; that is, the system has a deadbeat response. As in continuous control theory, the prototype inputs are usually chosen to be of the form t^k, k = 0, 1, 2; i.e. step, ramp, or parabolic. It is known that a minimum-time controller designed for a certain input such as 1) parabolic, or 2) ramp, will also provide a deadbeat response for a lower order input such as 1) ramp or step, or 2) step, but the overshoot resulting from the lower order input is usually much greater than can be tolerated in practice. In 1956, Bertram [3] considered the case of step response to a deadbeat system designed for a ramp input. He modified the digital controller design such that the sum I of the sampled step-response errors squared was minimized at the cost of a specified increase in response time, where

$$I = \sum_{k=0}^{\infty} [e(kT)]^2 \tag{3.1}$$

In 1965 and 1967, Bertram's work was extended by Pierre, Lorchirachoonkul, and Ross [30, 31]. Two significant results which they obtained are: 1) the overshoot obtained by Bertram's approach is shown to be the minimum possible under the assumed deadbeat response

constraints; 2) a limit on the performance of such a system is

$$(n-1)B \geqslant 100 \tag{3.2}$$

where B is the percentage of overshoot resulting from a step input, and n is the number of sampling periods used for deadbeat response to a ramp input.

In 1967, Lorchirachoonkul [24] extended this work to account for a pure time delay in the system. He shows that the limit on the performance of such a system is

$$(n-m)B \geqslant 100m \tag{3.3}$$

where B and n are defined as before, and

$$\frac{D}{T} + 1 \leqslant m < \frac{D}{T} + 2 \tag{3.4}$$

where D is the length of the pure time delay, T is the length of the sample period, and m is an integer.

In this chapter, the above mentioned works are extended to parabolic inputs. The theory of z-transforms [35, 21] is used throughout the chapter. It is assumed that the plant transfer function, G(z), contains no poles or zeros outside the unit circle of the z plane. Intersample ripple, caused by possible zeros within the unit circle,

is not considered. The performance measure in this case is stipulated here as

$$I = \sum_{k=0}^{\infty} [e_s(kT)]^2 + h[e_r(kT)]^2 \tag{3.5}$$

where $e_s(kT)$ is the unit step error at the k'th sampling instant, $e_r(kT)$ is the unit ramp error at the k'th sampling instant, and h is a weighting factor. In Section 3.2, equations are derived to determine the coefficients of the closed-loop transfer function that yields a minimum of I in (3.5) but subject to satisfying deadbeat criteria. The equations are solved for the special cases where n = 4, 5 and 6, with h = 0, 1 and ∞, and various step and ramp responses are plotted and compared. In Section 3.3, the equations are solved for n general as h becomes arbitrarily large; i.e. only the sum of the ramp errors squared is minimized. For this use, the performance limit

$$\max |e_r| \geq \frac{T}{n-2} \tag{3.6}$$

is shown to hold, and this maximum error is shown to be less than that achievable by any other system (of the form in Figure 3.1) with the same deadbeat constraints. In Section 3.4, the equations of Section 3.2 are solved for n general with h = 0; i.e. only the step error squared is minimized. It is found that the step response exhibits both an overshoot and an undershoot, with the maximum value of the overshoot

being $(n-1)/(n-2)$ and that of the undershoot being $1/(n-2)$. However, contrary to previous results, these maximum errors are not the minimum possible. A linear programming method is given to calculate the closed-loop transfer-function coefficients while minimizing the maximum step error, and the step and ramp responses are plotted for $n = 4$, 5 and 6.

3.2 DERIVATION OF THE GENERAL SET OF EQUATIONS

Since the error for a ramp or step input to a system designed for deadbeat response to a parabolic input is zero at the k'th sampling instant for $k \geq n$, where nT is the deadbeat response time, the performance measure can be written as

$$I = \sum_{k=0}^{n-1} [e_s(kT)]^2 + h[e_r(kT)]^2 \tag{3.7}$$

For the system of Figure 3.1 to have a deadbeat response, it is well-known that the closed-loop transfer function $M(z)$ of a physically realizable system must be of the form

$$M(z) = b_1 z^{-1} + b_2 z^{-2} + \ldots + b_n z^{-n} \tag{3.8}$$

where the b_i's are constants. It is equally well-known that for the system to exhibit deadbeat response to a parabolic input, $1 - M(z)$ must be of the form

$$1 - M(z) = (1 - z^{-1})^3 (1 + a_1 z^{-1} + \ldots + a_{n-3} z^{-n+3}) \tag{3.9}$$

The right-hand member of equation (3.8) can be substituted for M(z) in equation (3.9), and the coefficients of like powers of z can be equated to yield the following relationships.

$$
\begin{aligned}
b_1 &= 3 - a_1 \\
b_2 &= -3 + 3a_1 - a_2 \\
b_3 &= 1 - 3a_1 + 3a_2 - a_3 \\
b_4 &= a_1 - 3a_2 + 3a_3 - a_4 \\
&\;\;\vdots \\
b_k &= a_{k-3} - 3a_{k-2} + 3a_{k-1} - a_k \qquad (3.10) \\
&\;\;\vdots \\
b_{n-3} &= a_{n-6} - 3a_{n-5} + 3a_{n-4} - a_{n-3} \\
b_{n-2} &= a_{n-5} - 3a_{n-4} + 3a_{n-3} \\
b_{n-1} &= a_{n-4} - 3a_{n-3} \\
b_n &= a_{n-3}
\end{aligned}
$$

The ramp error at the sampling instants, $e_r(kT) = r_r(kT) - c_r(kT)$, can be found by noting that $C_r(z) = R_r(z)M(z)$, where $C_r(z)$ is the z-transform of $c_r(t)$ and $R_r(z)$ is the z-transform of $r_r(t)$. M(z) is given by (3.8) and $R_r(z)$ equals $Tz^{-1}/(1 - z^{-1})^2$. The step error is found sim-

ilarly, except that $R_s(z)$ equals $1/(1 - z^{-1})$. Thus, the step and ramp errors are found to be

k	$e_s(kT)$	$e_r(kT)$
0	1	0
1	$1 - b_1$	T
2	$1 - b_1 - b_2$	$T[2 - b_1]$
3	$1 - b_1 - b_2 - b_3$	$T[3 - b_2 - 2b_1]$
4	$1 - b_1 - b_2 - b_3 - b_4$	$T[4 - b_3 - 2b_2 - 3b_1]$
.	.	.
.	.	.
.	.	.
k	$1 - b_1 - b_2 \dots - b_k$	$T[k - b_{k-1} - 2b_{k-2} \dots -(k-1)b_1]$

(3.11)

By use of (3.11), the performance measure (3.7) can be written as

$$I = 1 + \sum_{k=1}^{n-1} [1-b_1-b_2\dots-b_k]^2 + hT^2[k-b_{k-1}-2b_{k-2}\dots-(k-2)b_2-(k-1)b_1]^2 \tag{3.12}$$

Equations (3.10) can be used to write the performance measure in terms of the a_i's. This yields, after simplification,

$$I = \sum_{k=1}^{n-3} (a_k-2a_{k-1}+a_{k-2})^2 + hT^2(a_k-a_{k-1})^2 + a_{n-4}^2 - 4a_{n-3}a_{n-4} + (5+hT^2)a_{n-3}^2 + hT^2 + 1 \tag{3.13}$$

where $a_0 = 1$, $a_i = 0$ for negative i, and $n \geqslant 4$.

The equations resulting in the a_k's which minimize I in equation (3.13) may be found by the classical approach. Thus,

$$\frac{\partial I}{\partial a_1} = 2(a_1-2) + hT^2 2(a_1-1) - 4(a_2-2a_1+1) - hT^2 2(a_2-a_1) + 2(a_3-2a_2+a_1) = 0$$

$$\frac{\partial I}{\partial a_2} = 2(a_2-2a_1+1) + hT^2 2(a_2-a_1) - 4(a_3-2a_2+a_1) - hT^2 2(a_3-a_2) + 2(a_4-2a_3+a_2) = 0$$

$$\frac{\partial I}{\partial a_3} = 2(a_3-2a_2+a_1) + hT^2 2(a_3-a_2) - 4(a_4-2a_3+a_2) - hT^2 2(a_4-a_3) + 2(a_5-2a_4+a_3) = 0$$

$$\begin{matrix} \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \end{matrix} \qquad (3.14)$$

$$\frac{\partial I}{\partial a_{n-4}} = 2(a_{n-4}-2a_{n-5}+a_{n-6}) + hT^2 2(a_{n-4}-a_{n-5}) - 4(a_{n-3}-2a_{n-4}+a_{n-5}) - hT^2 2(a_{n-3}-a_{n-4}) + 2a_{n-4} - 4a_{n-3} = 0$$

$$\frac{\partial I}{\partial a_{n-3}} = 2(a_{n-3}-2a_{n-4}+a_{n-5}) + hT^2 2(a_{n-3}-a_{n-4}) - 4a_{n-4} + 2(5+hT^2)a_{n-3} = 0$$

Equations (3.14) may be simplified to give,

$$
\begin{aligned}
0 &= (6 + 2\lambda)a_1 + (-4-\lambda)a_2 + a_3 - \lambda - 4 \\
0 &= (-4-\lambda)a_1 + (6+2\lambda)a_2 + (-4-\lambda)a_3 + a_4 + 1 \\
0 &= a_1 + (-4-\lambda)a_2 + (6+2\lambda)a_3 + (-4-\lambda)a_4 + a_5 \\
&\cdot \qquad \cdot \qquad \cdot \\
&\cdot \qquad \cdot \qquad \cdot \\
&\cdot \qquad \cdot \qquad \cdot \\
0 &= a_{n-6} + (-4-\lambda)a_{n-5} + (6+2\lambda)a_{n-4} + (-4-\lambda)a_{n-3} \\
0 &= a_{n-5} + (-4-\lambda)a_{n-4} + (6+2\lambda)a_{n-3}
\end{aligned}
\tag{3.15}
$$

where λ equals hT^2.

Because of computational difficulties, the above set of equations will not be solved for the a_i's as functions of both n and λ. Instead, they are solved for n = 4, 5 and 6, with λ as a parameter. Various responses will then be plotted for $\lambda = 0$, 1 and ∞. Since $\lambda = hT^2$, the implication is that the sample period, T, is held fixed, while h = 0, $1/T^2$ and ∞. In Section 3.3, the a_i's are calculated for a general n if only the ramp error is considered in the performance measure, while in Section 3.4, the a_i's are found with only the step error considered.

Table 3.1 gives the a_i's for n = 4, 5 and 6. The b_i's are calculated from equation (3.10) and are also listed. Figure 3.2 shows the step response for n = 5, with $\lambda = 0$, 1 and ∞, and Figure 3.3 gives the corresponding ramp response. (Figures 1.2 through 1.10 depict various responses at the sampling instants. The data points are connected by

straight line segments for clarity.) Since the responses are quite similar for $\lambda = 0$ and ∞, and the response for $\lambda = 1$ generally lies in between, only the parameter values $\lambda = 0$ and $\lambda = \infty$ are considered in subsequent plots. Figure 3.2 evidences the fact that a smaller maximum step overshoot results if the step error squared is minimized rather than the ramp error squared. This is not surprising in view of previously published results [30, 31]. However, it should be noted that an undershoot also occurs which is smaller if the ramp error squared is minimized. Another somewhat surprising result is that the transient response for $\lambda = 0$ is really not much different quantitatively than for $\lambda = \infty$. For example, the percent step overshoot is 100% if $\lambda = 0$ and 133% if $\lambda = \infty$, while the percent undershoot is 50% if $\lambda = 0$ and 33% if $\lambda = \infty$. Similarly, the maximum percent ramp error is 50% with $\lambda = 0$ and 33% if $\lambda = \infty$. Figures 3.4 and 3.5 give the step and ramp responses for $n = 4$ and $n = 6$. These show that as n is increased, the peak transient errors decrease. This will be examined in more detail in the next two sections.

3.3 GENERAL SOLUTION WITH MINIMUM RAMP-ERROR-SQUARED PERFORMANCE MEASURE

The set of equations for a minimum ramp-error-squared criteria can easily be found by letting λ become arbitrarily large in equations (3.15). This gives

$$
\begin{aligned}
0 &= 2a_1 - a_2 - 1 \\
0 &= 2a_2 - a_3 - a_1 \\
&\cdot \\
&\cdot \\
&\cdot \\
0 &= 2a_{n-4} - a_{n-3} - a_{n-5} \\
0 &= 2a_{n-3} - a_{n-4}
\end{aligned}
\tag{3.16}
$$

These equations may be solved simultaneously to give

$$a_k = \frac{n-k-2}{n-2} \quad , \qquad k = 1, 2, \ldots , n-3 \tag{3.17}$$

It will now be proven that these a_k's result in the minimum of the maximum ramp response error. Equations (3.10) and (3.11) indicate that the error in the ramp response at the k'th sampling instant is

$$e_k = (a_{k-1} - a_{k-2})T, \qquad 1 < k < n \tag{3.18}$$

Substitution of (3.17) into (3.18) gives

$$e_k = -\frac{T}{n-2} \quad , \qquad 1 < k < n \tag{3.19}$$

Now suppose any a_{k-1} is varied by $-\delta$, when δ is any positive number.

Then

$$e_k = \left(\frac{n-k-1}{n-2} - \delta - \frac{n-k}{n-2} \right)T = \left(- \frac{1}{n-2} - \delta \right)T \tag{3.20}$$

which is larger in magnitude than any previous error. Now let any a_{k-1} be varied by $+\delta$ where δ is any positive number. Then

$$e_{k+1} = \left(\frac{n-k-2}{n-2} - \frac{n-k-1}{n-2} - \delta \right)T = \left(- \frac{1}{n-2} - \delta \right)T \tag{3.21}$$

which is also larger in magnitude than any previous error; i.e.,

$$\max \left| e_r \right| \geq \frac{T}{n-2} \tag{3.22}$$

Thus, any set of a_i's other than those given by the minimum ramp-error-squared criterion will result in a larger error magnitude at some sampling instant. It should be pointed out that this proof differs somewhat from that used by Pierre, et al. [30, 31]. They use only the "b_i equations" in their proof that the step overshoot is minimized when the closed-loop transfer function fulfills a minimum step-error-squared criterion subject to satisfying the deadbeat constraint at $t = nT$. A generalization of their stated result is that the step overshoot for the minimum step-error-squared criterion is less than that for any system (of the type in Figure 3.1) which gives a ramp response passing through nT at $t = nT$.

It follows from Figure 3.1 that

$$D(z)G_p(z) = \frac{M(z)}{1-M(z)} = \frac{b_1z^{-1}+b_2z^{-2}+\ldots+b_nz^{-n}}{1-(b_1z^{-1}+b_2z^{-2}+\ldots+b_nz^{-n})} \tag{3.23}$$

Using equations (3.10) and (3.17), the b_i's are found to be

$$\begin{aligned} b_1 &= \frac{2n-3}{n-2} \\ b_2 &= -\frac{n-1}{n-2} \\ b_k &= 0, \qquad 2 < k < n-1 \\ b_{n-1} &= -\frac{1}{n-2} \\ b_n &= \frac{1}{n-2} \end{aligned} \tag{3.24}$$

Thus, the equation for the digital controller is found to be

$$D(z) = \frac{1}{G_p(z)} \; \frac{(2n-3)z^{-1}-(n-1)z^{-2}-z^{-n+1}+z^{-n}}{(n-2)-[(2n-3)z^{-1}-(n-1)z^{-2}-z^{-n+1}+z^{-n}]} \tag{3.25}$$

It is also of interest to determine what the step and ramp responses will be if the number of samples to parabolic deadbeat is allowed to become arbitrarily large. This can be done by taking $\lim_{n\to\infty} b_i$ in equations (3.24) and using the result in equations (3.11). Thus,

$$
\begin{aligned}
b_1 &= 2 \\
b_2 &= -1 \\
b_k &= 0 \quad , \qquad k > 2
\end{aligned}
\tag{3.26}
$$

This gives a step error of -1 (100% overshoot) and the unavoidable ramp error of T at k=1, while both responses have zero error for $k > 1$. (see Figure 3.6).

3.4 GENERAL SOLUTION WITH MINIMUM STEP-ERROR-SQUARED PERFORMANCE MEASURE

The set of equations for a minimum step-error-squared criteria can be found by setting λ equal to zero in equations (3.15). This gives

$$
\begin{aligned}
4 &= 6a_1 - 4a_2 + a_3 \\
-1 &= -4a_1 + 6a_2 - 4a_3 + a_4 \\
0 &= a_1 - 4a_2 + 6a_3 - 4a_4 + a_5 \\
&\quad \vdots \\
0 &= a_{n-7} - 4a_{n-6} + 6a_{n-5} - 4a_{n-4} + a_{n-3} \\
0 &= a_{n-6} - 4a_{n-5} + 6a_{n-4} - 4a_{n-3} \\
0 &= a_{n-5} - 4a_{n-4} + 6a_{n-3}
\end{aligned}
\tag{3.27}
$$

After some tedious manipulations, this set of algebraic equations can be solved to give

$$a_k = \frac{(k+1)(n-k-1)(n-k-2)}{(n-1)(n-2)} \quad , \qquad k = 1, 2, \ldots , n-3 \tag{3.28}$$

Equations (3.10) can now be solved to give

$$\begin{aligned} b_1 &= \frac{n+3}{n-1} \\ b_k &= \frac{-6}{(n-1)(n-2)} \qquad\qquad 1 < k < n \\ b_n &= \frac{2}{n-1} \end{aligned} \tag{3.29}$$

Note that, in this case, as $n \to \infty$, b_1 equals 1, while b_k equals 0 for $k > 1$. This implies (equations (3.11)) that the step response error is zero after the $k = 0$ sampling instant. On the other hand, the ramp response has a constant error of T for all $k > 0$ (see Figure 3.6).

Once again, the corresponding digital compensator may be determined.

$$D(z) = \frac{1}{G_p(z)} \frac{(n-2)(n+3)z^{-1}(1-z^{-1})-6(z^{-2}-z^{-n})+2(n-2)z^{-n}(1-z^{-n})}{(n-1)(n-2)(1-z^{-1})-(n-2)(n+3)z^{-1}(1-z^{-1})+6(z^{-2}-z^{-n})-2(n-2)z^{-n}(1-z^{-n})} \tag{3.30}$$

Note the increased complexity (more memory elements) of this controller.

It might be guessed that minimization of the sum of the step errors squared would also minimize the maximum step overshoot, as it does with a deadbeat ramp constraint. However, this is not the case. This is best demonstrated by a simple counter example. Let $n = 4$. From equation (3.27), a_1 is found to be 2/3. This, in turn, gives step errors of $e_1 = -4/3$, $e_2 = -1/3$, and $e_3 = 2/3$. Now suppose a_1 has the value 1. The resulting errors are $e_1 = -1$, $e_2 = -1$, and $e_3 = 1$, which have a smaller maximum error magnitude ($|e_1| = 1$) than for the previous case ($|e_1| = 4/3$). It can be shown by reasoning equivalent to that used to derive equation (3.22), that $a_1 = 1$ indeed minimizes the maximum step error. Similarly, the a_i's and b_i's which minimize the maximum step errors for $n = 5$ and $n = 6$ can be calculated, and are listed in Table 3.2. The corresponding step and ramp responses are plotted in Figures 3.7 and 3.8. A comparison of the step and ramp responses for the three different performance measures is made in Figures 3.9 and 3.10 for $n = 5$.

As n becomes large, it becomes quite tedious to calculate the a_i's for the "min-max" performance measure discussed above. For large n, linear programming techniques [42] are quite useful. In this problem the performance measure to be minimized is $\max |e_i|$, $i = 1, 2, \ldots, n-1$, which is nonlinear. However, it may be linearized by introducing an auxiliary variable g such that

$$g + e_i \geq 0$$
$$g - e_i \geq 0 \qquad i = 1, 2, \ldots, n-1 \tag{3.31}$$

This implies that

$$g \geq |e_i| \quad , \qquad i = 1, 2, \ldots, n-1 \tag{3.32}$$

which in turn implies that minimizing g minimizes the maximum value of $|e_i|$ for $i = 1, 2, \ldots, n-1$.

As an example, the standard linear programming equations will be set up for the case $n = 4$.

The error equations in terms of the a_i's are

$$\begin{aligned} e_1 &= a_1 - 2 \\ e_2 &= -2a_1 + 1 \\ e_3 &= a_1 \end{aligned} \tag{3.33}$$

Combining equations (3.31) and (3.33) gives

$$
\begin{aligned}
g + a_1 - 2 &\geqslant 0 \\
g - a_1 + 2 &\geqslant 0 \\
g - 2a_1 + 1 &\geqslant 0 \\
g + 2a_1 - 1 &\geqslant 0 \\
g + a_1 \quad &\geqslant 0 \\
g - a_1 \quad &\geqslant 0
\end{aligned}
\tag{3.34}
$$

Since a_1 is unrestricted in sign, the relationship

$$a_1 = a_+ - a_- \tag{3.35}$$

is introduced, with $a_+ \geqslant 0$ and $a_- \geqslant 0$. Also, the inequality constraints may be replaced by equality constraints if slack variables, $x_i \geq 0$, are introduced. Thus, the final set of constraint equations is

$$
\begin{aligned}
g + a_+ - a_- - x_1 &= 2 \\
-g + a_+ - a_- + x_2 &= 2 \\
-g + 2a_+ - 2a_- + x_3 &= 1 \\
g + 2a_+ - 2a_- - x_4 &= 1 \\
g + a_+ - a_- - x_5 &= 0 \\
g - a_+ + a_- - x_6 &= 0
\end{aligned}
\tag{3.36}
$$

and the performance measure to be maximized is

$$P = -g \tag{3.37}$$

The problem is now in standard linear programming form.

3.5 CONCLUSION

In this chapter, the works by Bertram [3] and by Pierre, Lorchirachoonkul, and Ross [30, 31] are extended to prototype inputs of the parabolic type. The coefficients of the closed-loop transfer function of a system which is designed to give deadbeat response to a parabolic input are determined by minimizing either the sum of the square of the errors at the sampling instants for a step input, or a ramp input, or a weighted combination of the two.

In Section 3.2, it is shown that the step and ramp responses for the various performance measures mentioned above are quite similar, quantitatively, for the special cases of 4, 5 or 6 sample periods to deadbeat. In Section 3.3, only the ramp error is considered in the performance measure, and a maximum error performance limit is derived. In Section 3.4, only the step error is considered: a formula for calculating the maximum step overshoot and undershoot is derived in terms of the number of sample periods to deadbeat, but, in this case, it is shown that these are not the minimum possible under the deadbeat constraint. A different performance criterion, that of minimizing the maximum absolute value of step response error, is introduced; and a linear programming scheme to determine the coefficients of the closed-

loop transfer function which satisfy this criterion is presented.

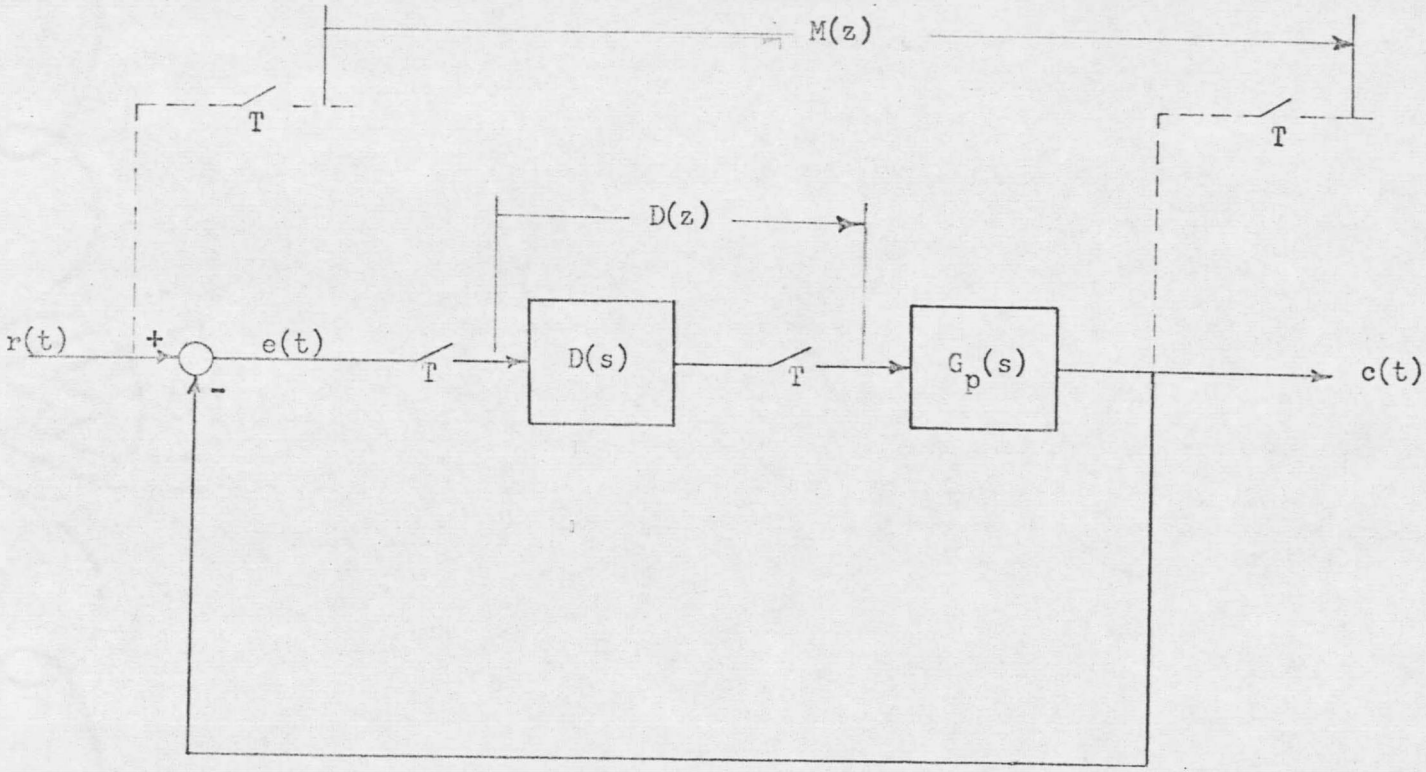

Figure 3.1. Closed-loop system with digital controller.

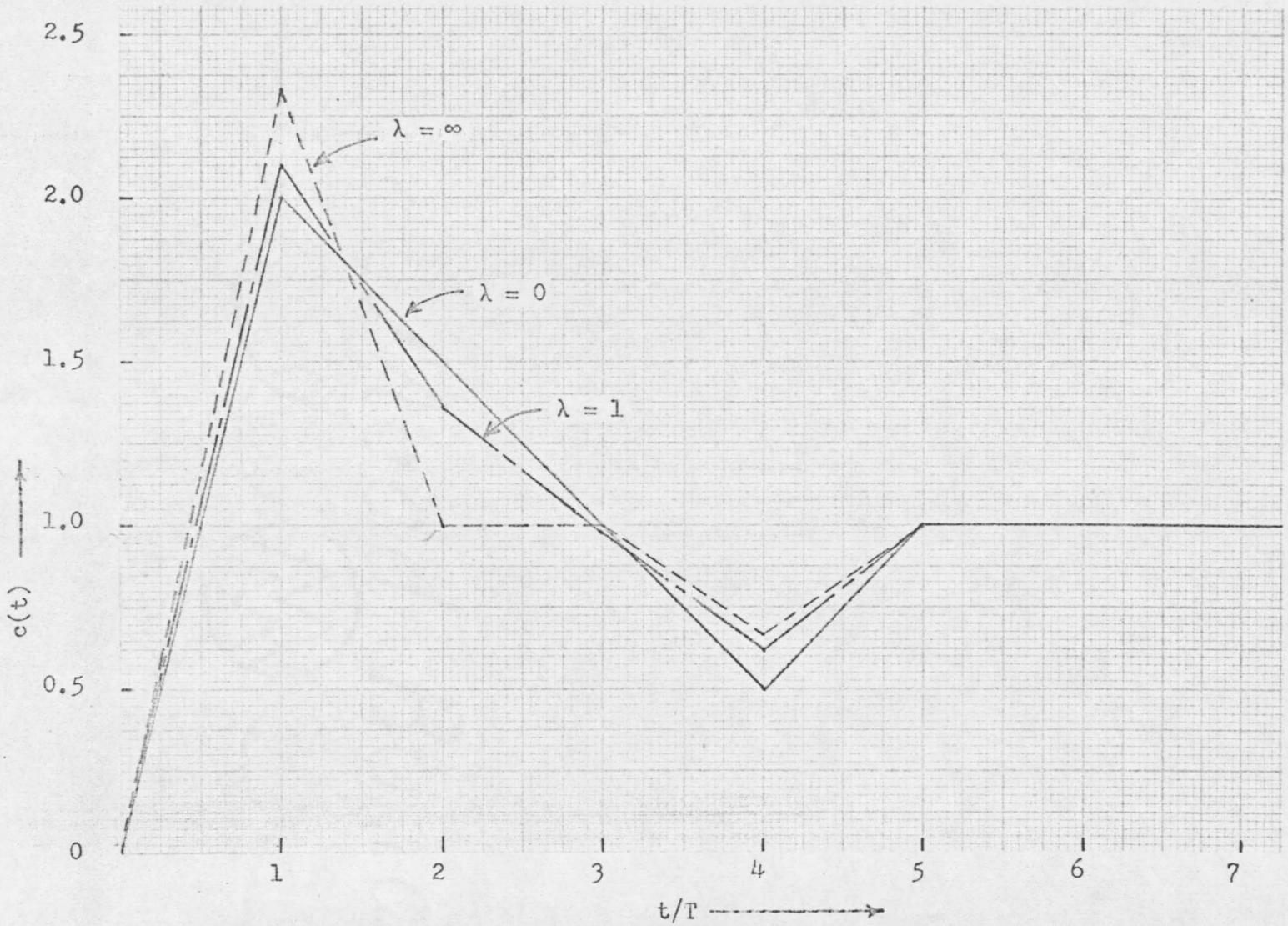

Figure 3.2. Unit step response for n = 5 with λ as parameter.

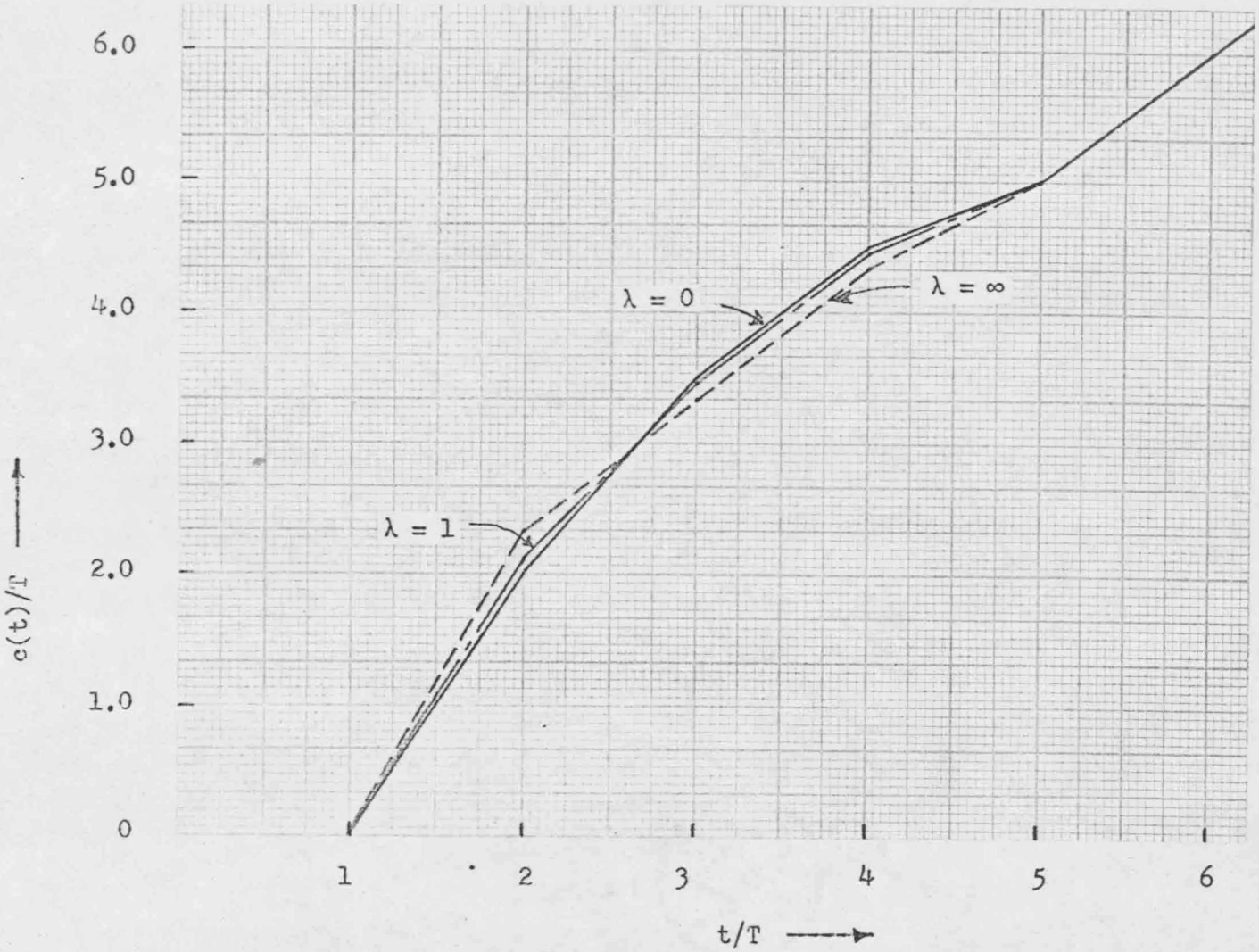

Figure 3.3. Unit ramp response for n = 5, with λ as a parameter.

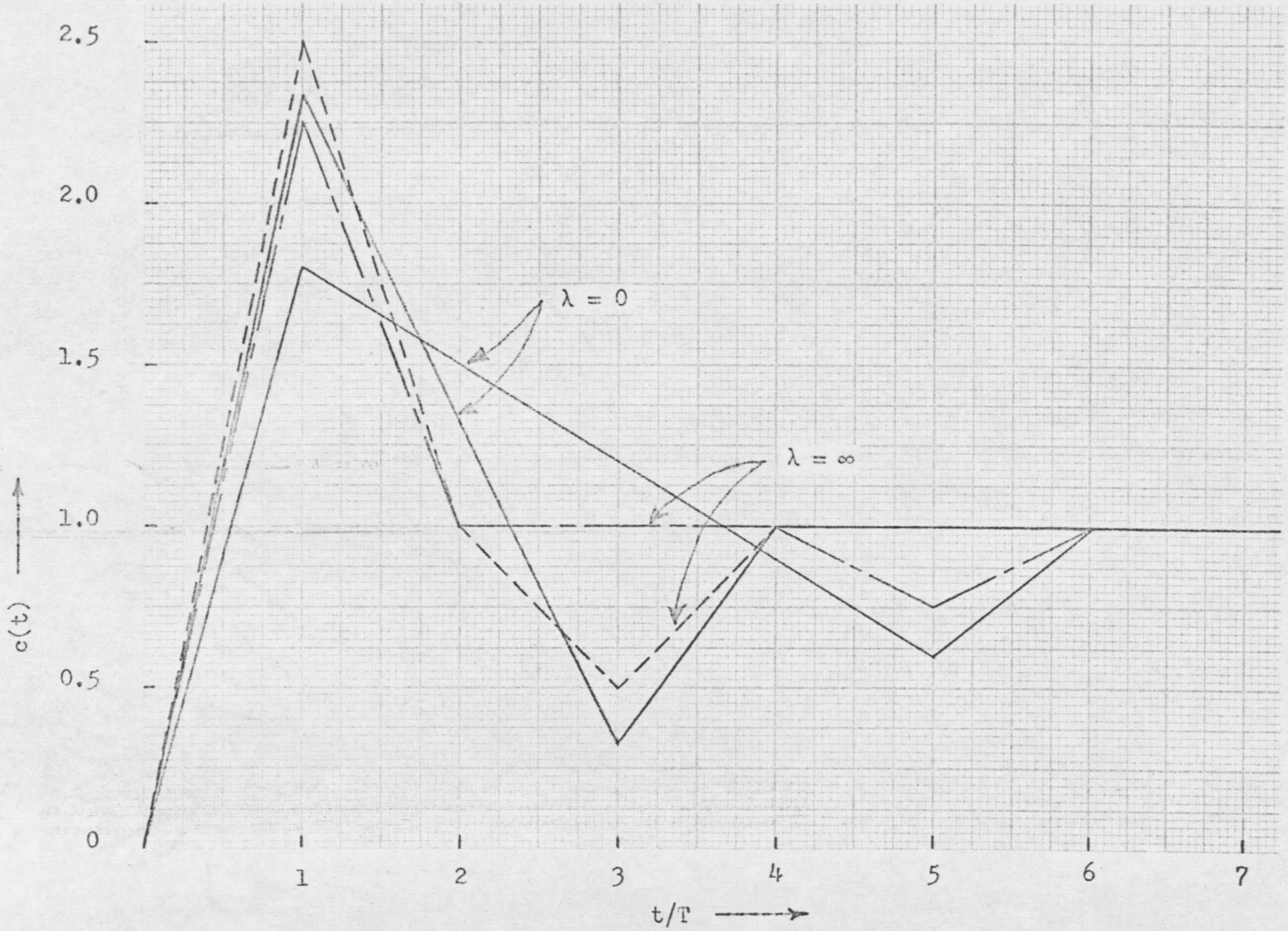

Figure 3.4. Unit step response for n = 4 and 6 with λ = 0 and ∞.

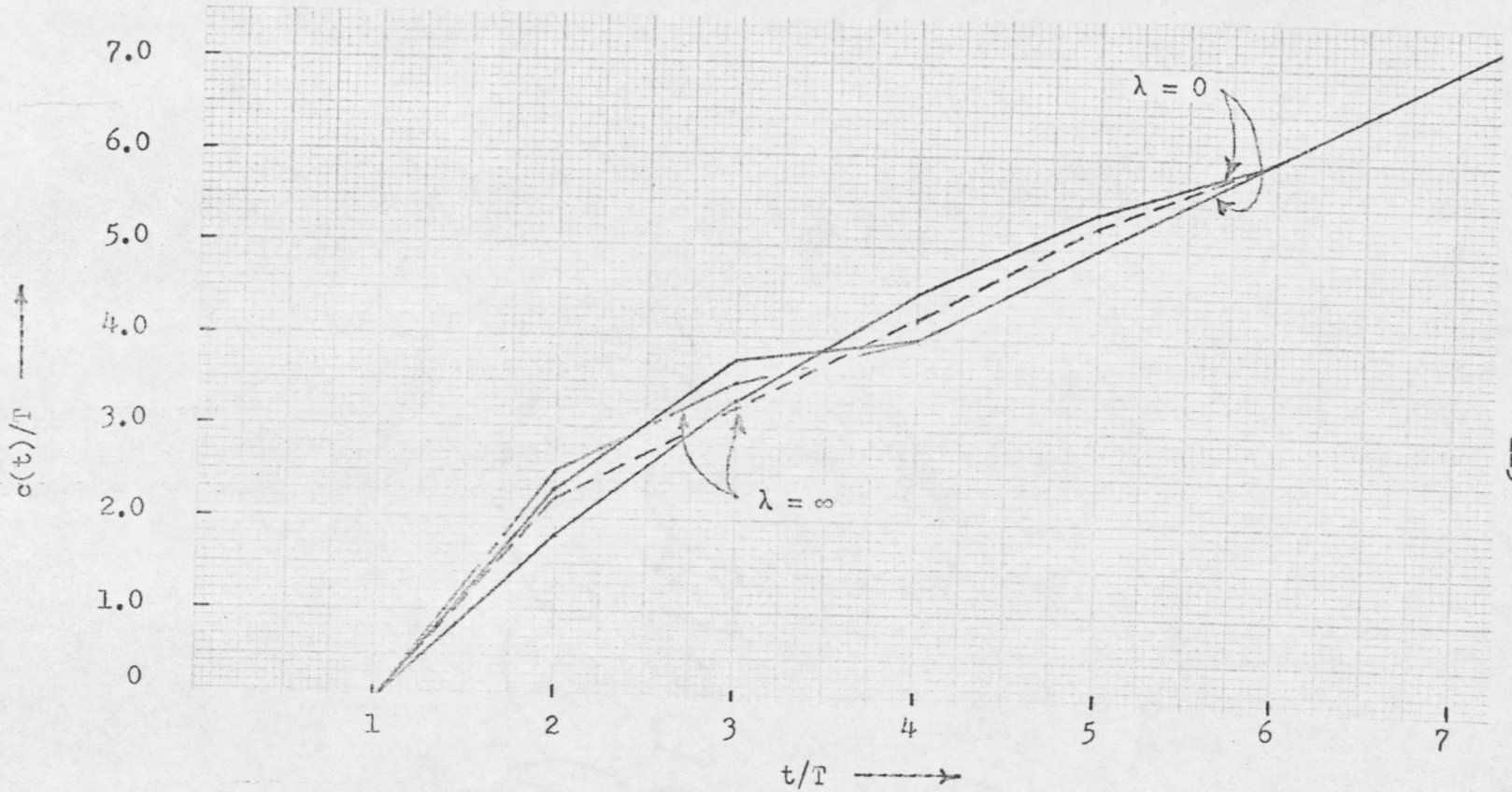

Figure 3.5. Unit ramp response for n = 4 and 6 with λ = 0 and ∞.

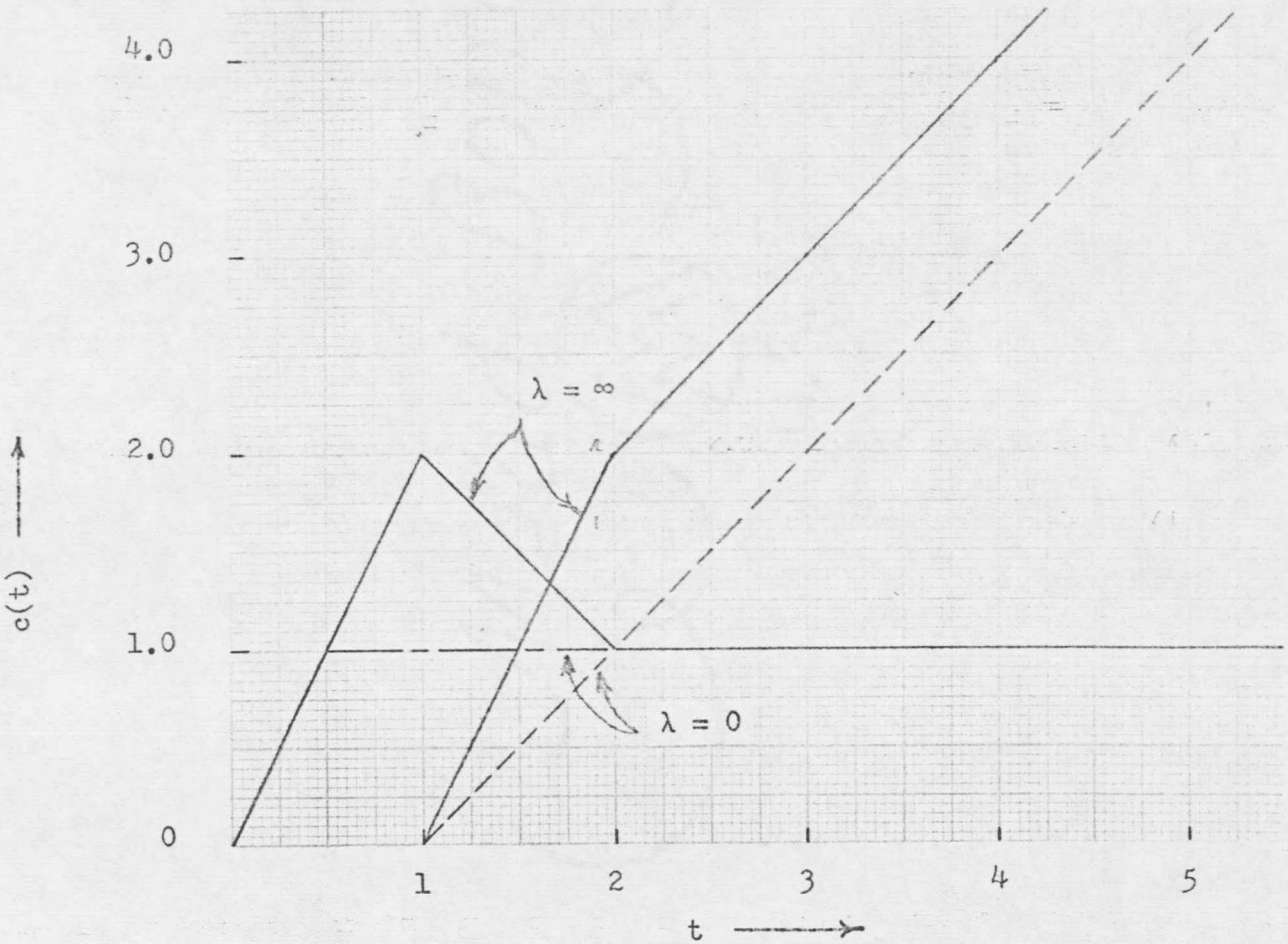

Figure 3.6. Unit step and ramp responses with n very large and T = 1.

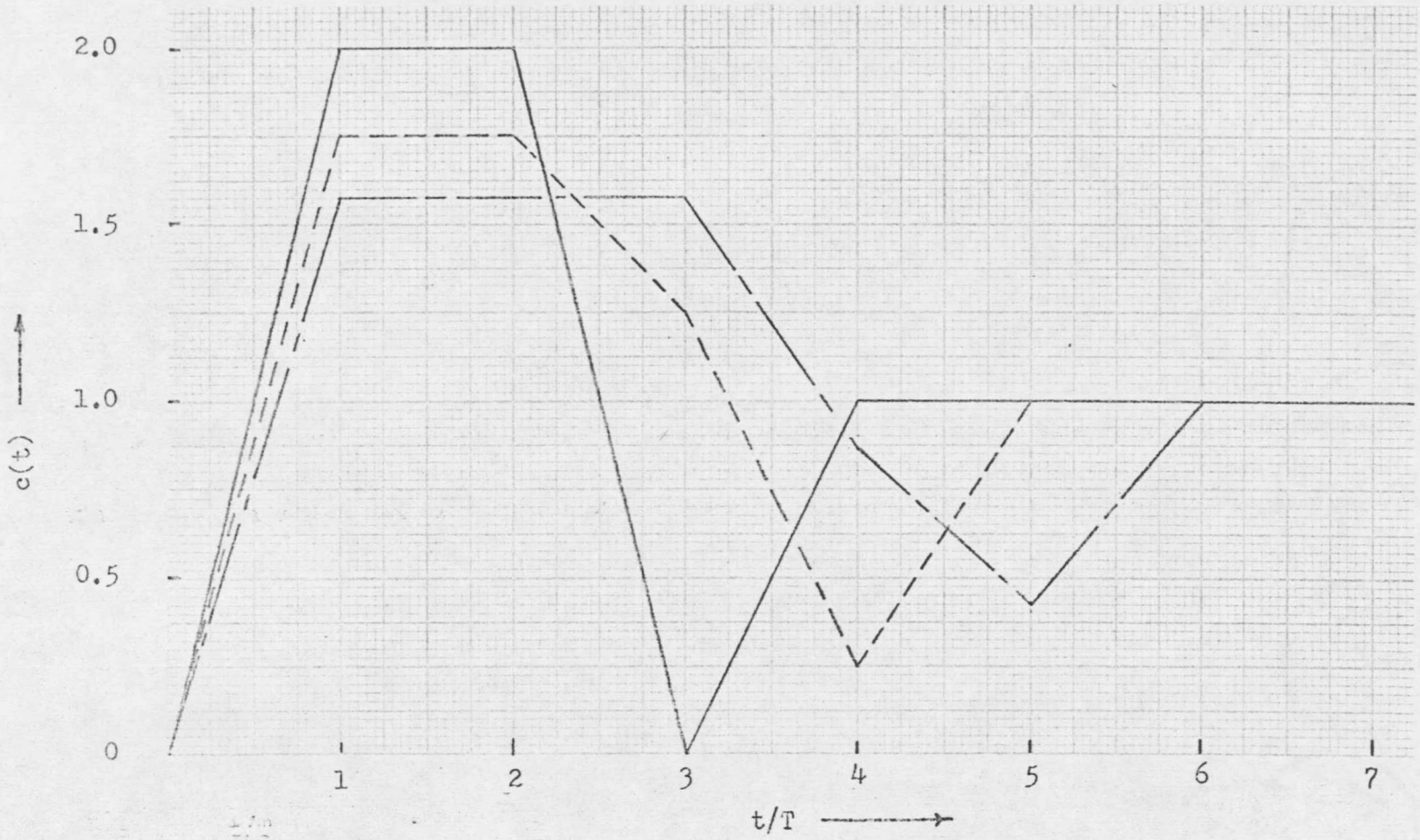

Figure 3.7. Unit step response with maximum magnitude of step error minimized.

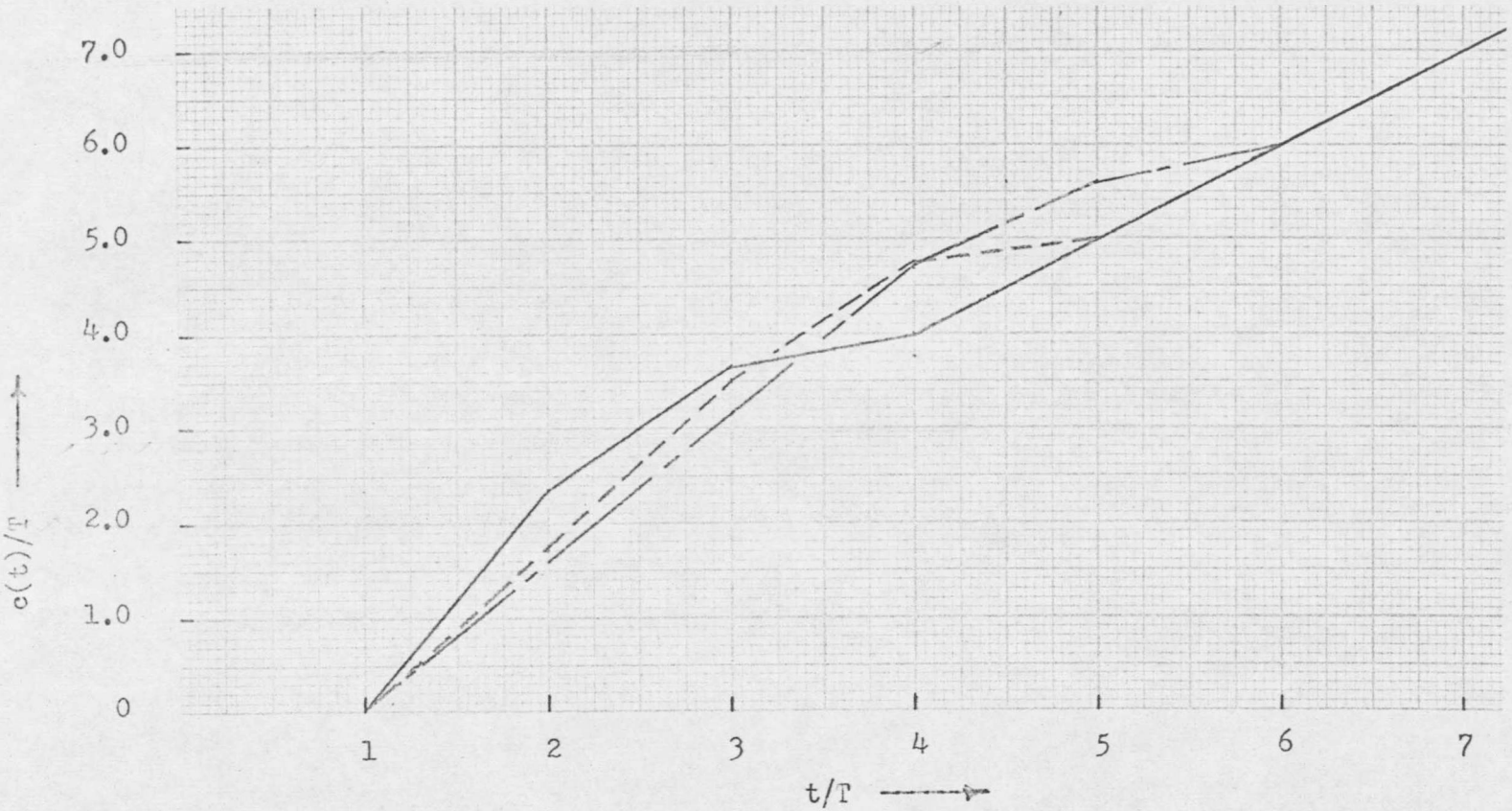

Figure 3.8. Unit ramp response with maximum magnitude of step error minimized.

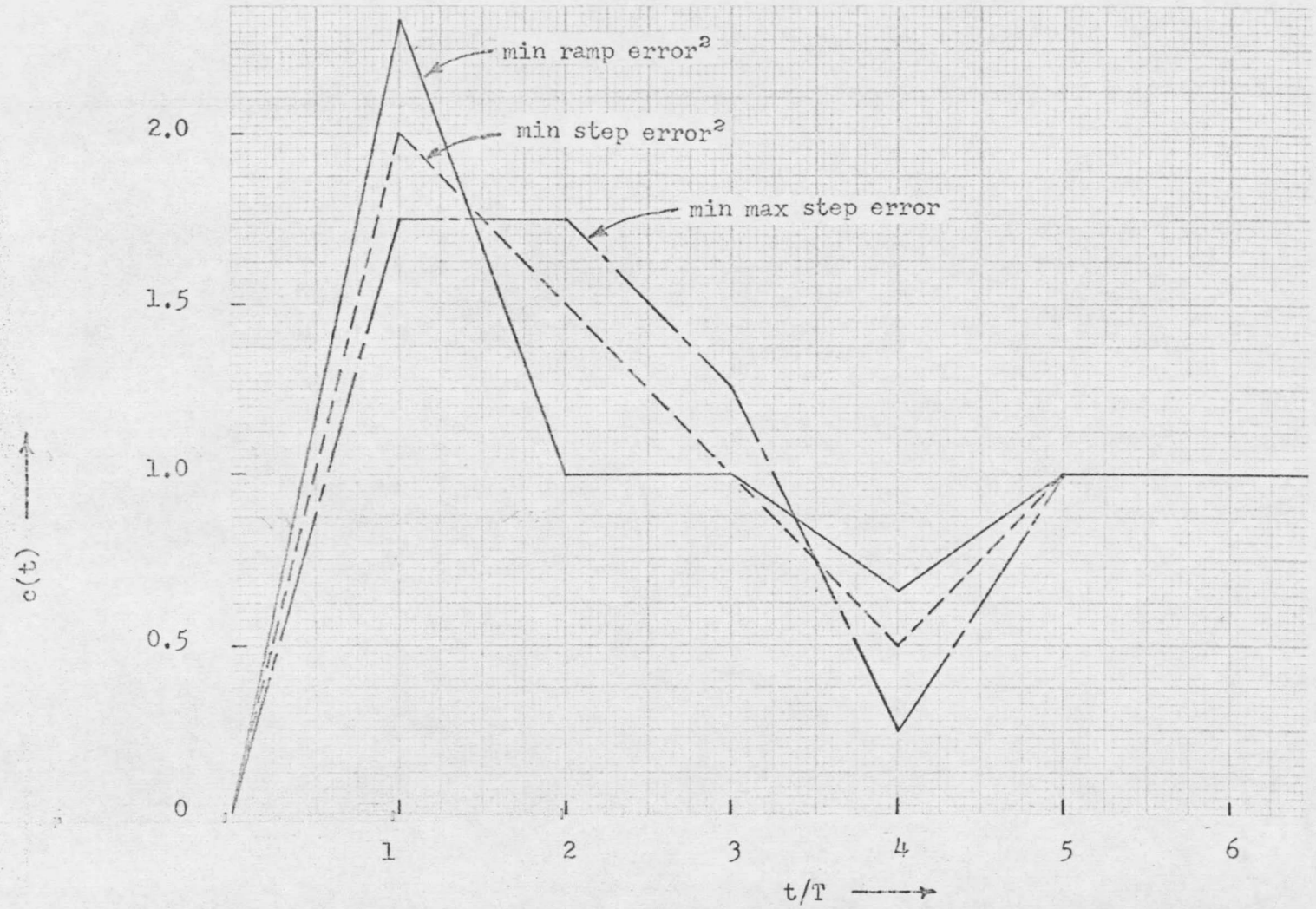

Figure 3.9. Unit step responses for n = 5.

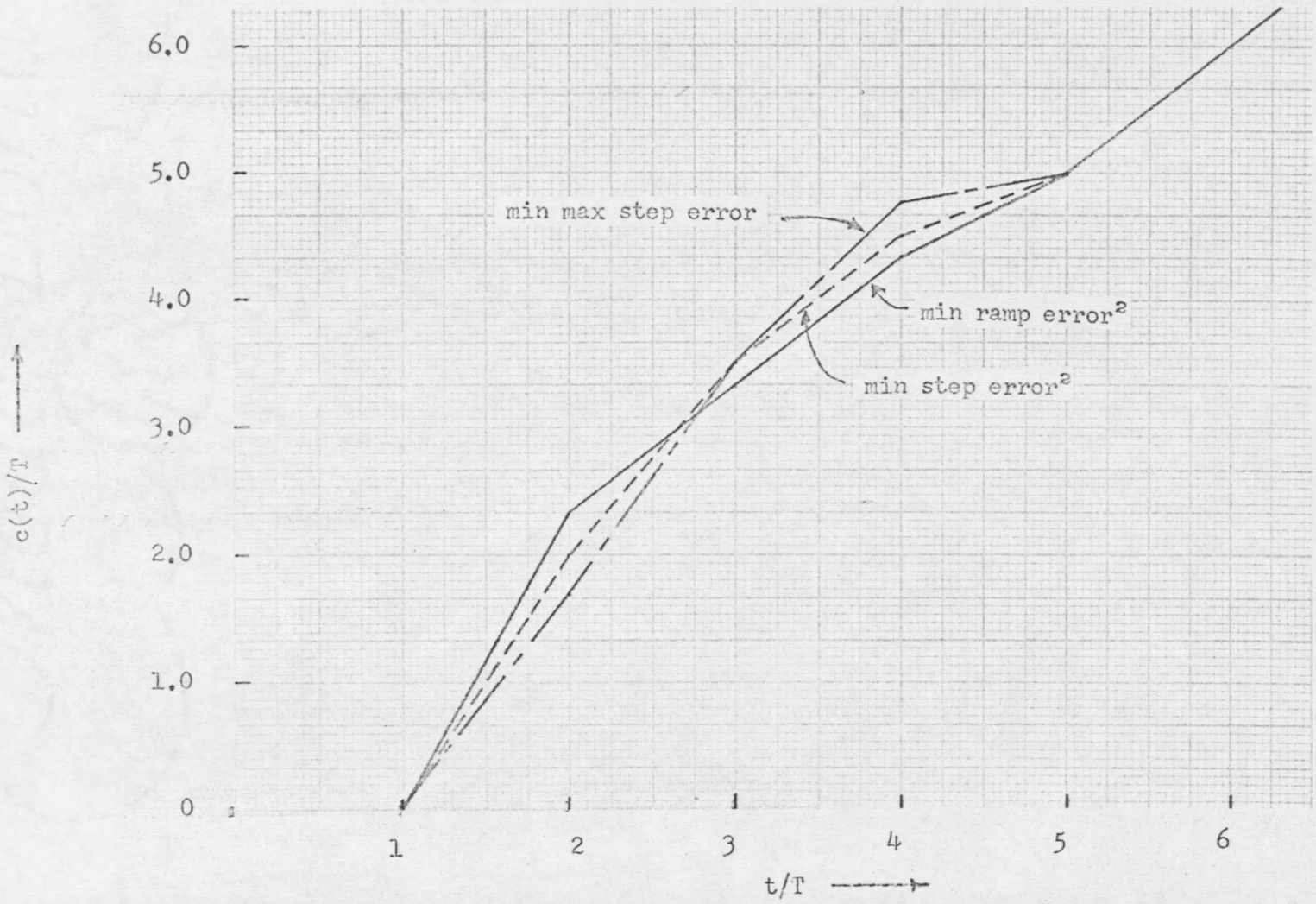

Figure 3.10. Unit ramp responses for n = 5.

Table 3.1. Transfer Function Coefficients for Squared Error Performance Measure with n = 4, 5 and 6.

n	a_i	b_i
4	$a_1 = \frac{4 + \lambda}{6 + 2\lambda}$	$b_1 = \frac{14 + 5\lambda}{6 + 2\lambda}$
		$b_2 = \frac{-6 - 3\lambda}{6 + 2\lambda}$
		$b_3 = \frac{-6 - \lambda}{6 + 2\lambda}$
		$b_4 = \frac{4 + \lambda}{6 + 2\lambda}$
5	$a_1 = \frac{2\lambda^2+13\lambda+20}{3\lambda^2+16\lambda+20}$	$b_1 = \frac{7\lambda^2+35\lambda+40}{3\lambda^2+16\lambda+20}$
	$a_2 = \frac{\lambda^2+ 6\lambda+10}{3\lambda^2+16\lambda+20}$	$b_2 = \frac{-4\lambda^2-15\lambda-10}{3\lambda^2+16\lambda+20}$
		$b_3 = \frac{- 5\lambda-10}{3\lambda^2+16\lambda+20}$
		$b_4 = \frac{-\lambda^2- 5\lambda-10}{3\lambda^2+16\lambda+20}$
		$b_5 = \frac{\lambda^2+ 6\lambda+10}{3\lambda^2+16\lambda+20}$
6	$a_1 = \frac{3\lambda^3+26\lambda^2+71\lambda+60}{4\lambda^3+30\lambda^2+70\lambda+50}$	$b_1 = \frac{9\lambda^3+64\lambda^2+139\lambda+90}{4\lambda^3+30\lambda^2+ 70\lambda+50}$
	$a_2 = \frac{2\lambda^3+17\lambda^2+48\lambda+45}{4\lambda^3+30\lambda^2+70\lambda+50}$	$b_2 = \frac{-5\lambda^3-29\lambda^2-45\lambda-15}{4\lambda^3+30\lambda^2+70\lambda+50}$
	$a_3 = \frac{\lambda^3+ 8\lambda^2+21\lambda+20}{4\lambda^3+30\lambda^2+70\lambda+50}$	$b_3 = \frac{- 5\lambda^2-20\lambda-15}{4\lambda^3+30\lambda^2+70\lambda+50}$
		$b_4 = \frac{- \lambda^2-10\lambda-15}{4\lambda^3+30\lambda^2+70\lambda+50}$
		$b_5 = \frac{-\lambda^3- 7\lambda^2-15\lambda-15}{4\lambda^3+30\lambda^2+70\lambda+50}$
		$b_6 = \frac{\lambda^3+ 8\lambda^2+21\lambda+20}{4\lambda^3+30\lambda^2+70\lambda+50}$

Table 3.2. Transfer Function Coefficients for Minimizing the Maximum Step Error with n = 4, 5 and 6.

n	a_i	b_i
4	$a_1 = 1$	$b_1 = 2$
		$b_2 = 0$
		$b_3 = -2$
		$b_4 = 1$
5	$a_1 = 5/4$	$b_1 = 7/4$
	$a_2 = 3/4$	$b_2 = 0$
		$b_3 = -1/2$
		$b_4 = -1$
		$b_5 = 3/4$
6	$a_1 = 10/7$	$b_1 = 11/7$
	$a_2 = 9/7$	$b_2 = 0$
	$a_3 = 4/7$	$b_3 = 0$
		$b_4 = -5/7$
		$b_5 = -3/7$
		$b_6 = 4/7$

CHAPTER 4

SUMMARY AND SUGGESTED FUTURE RESEARCH

4.1 SUMMARY OF ESSENTIAL VALUES OF THE THESIS RESEARCH

Time-optimal control and suboptimal control of sampled-data systems constrained by a saturation nonlinearity are examined in this thesis. Also, the response of linear sample-data systems to various prototype inputs is investigated.

In Chapter 1, a brief review of the time-optimal control schemes presented in the literature is given. It is then shown that it can be advantageous to relax the minimum time requirement somewhat, in order to reduce the complexity and cost of the controller. An analysis method is then presented to measure the amount of "suboptimality" for such a controller. A practical suboptimal system is then analyzed, and the analysis indicates a simple method of reducing the suboptimality of the system. Two systems which are claimed as time-optimal by their designers are also analyzed, and found to actually be suboptimal. In addition, it is shown that the analysis method is useful for computing the sensitivity to plant parameter variations for both optimal and suboptimal systems.

A hypothesis relating the minimum times of sampled-data and continuous systems is presented in Chapter 2. This time-loss hypothesis is proven for first-order pulse-amplitude-modulated and pulse-width-modulated systems, and for several special cases of second-order systems. In addition, evidence is presented which indicates the general validity of the hypothesis for systems with real eigenvalues.

The theory of deadbeat response of linear sampled-data systems

is extended to apply to parabolic inputs with minimum-squared-error constraints on the step and ramp responses. It is found that if only the sum of the ramp errors squared is minimized, the maximum ramp error is also minimized. On the other hand, minimizing the sum of the step errors squared does not minimize the maximum step error. However, a linear programming method to calculate the numerator and denominator coefficients of the digital controller which causes minimization of the maximum step error is presented.

4.2 ASPECTS MERITING ADDITIONAL STUDY

Since the material presented concerning measurement of suboptimality is meant to be a practical design guide, a worthy research effort could be directed at the extension of this material to systems with multiple nonlinearities. In particular, Nagata, Kodama, and Kumagai [28] derive R_k regions, similar to those described in this paper, for systems with more than one bounded state variable. It appears that the suboptimality of such systems could be analyzed by methods similar to those presented in Chapter 1.

The importance of further research in the area investigated in Chapter 2, is evident. Although the truth of the time-loss hypothesis of Chapter 2 is fairly evident, a worthy research effort could be devoted to the general proof of this hypothesis and to generalizations of it for n'th-order systems with either real or complex eigenvalues and either real or complex eigenvalues and either PAM or PWM control.

Since, in practical applications, a steady state error <u>exactly</u>

equal to zero is not necessary, a useful extension of the material presented in Chapter 3 would be an investigation of possible cost advantages of a controller which only requires the response of a system to be within some small neighborhood of deadbeat.

REFERENCES CITED

1 Athans, M., and P. L. Falb, Optimal Control: An Introduction to the Theory and Its Applications, New York: McGraw-Hill Book Company, Inc., 1966, 879 pp.

2 Bellman, R., I. Glicksberg, and O. Gross, "On the bang-bang control problem," Quart. Appl. Mathematics, vol. 14, no. 1, April 1956, pp. 11-18.

3 Bertram, J. E., "Factors in the design of digital controllers for sampled-data feedback systems," AIEE Transactions (Applications and Industry), vol. 75, part II, July 1956, pp. 151-159.

4 Bode, H. W., Network Analysis and Feedback Amplifier Design, New Jersey: Van Nostrand, 1945.

5 Bogner, I., and L. F. Kazda, "An investigation of the switching criteria for higher order servomechanisms," AIEE Transactions (Applications and Industry), vol. 73, part II, no. 13, July 1954, pp. 118-127.

6 Bushaw, D. W., "Differential equations with a discontinuous forcing term," Ph.D. dissertation, Princeton University, Princeton, New Jersey, 1952.

7 Davenport, W. B., and W. L. Root, An Introduction to the Theory of Random Signals and Noise, New York: McGraw-Hill Book Company, Inc., 1958, 393 pp.

8 Desoer, C. A., and J. Wing, "An optimal strategy for a saturating sampled-data system," IRE Transactions on Automatic Control, vol. AC-6, February 1961, pp. 5-15.

9 Desoer, C. A., and J. Wing, "A minimal time discrete system," IRE Transactions on Automatic Control, vol. AC-7, May 1961, pp. 111-125.

10 Desoer, C. A., and J. Wing, "The minimal time regulator problem for linear sampled-data systems: general theory," Journal of the Franklin Institute, vol. 271, September 1961, pp. 208-226.

11 Dorato, P., "On sensitivity in optimal control systems," IEEE Transactions on Automatic Control, vol. AC-8, no. 3, July 1963, pp. 256-257.

12 Eaton, J. H., "An on-line solution to sampled-data time optimal control," Journal of Electronics and Control, vol. 15, no. 4, October 1963, pp. 333-341.

13 Flugge-Lotz, I., and H. A. Titus, "Optimum and quasi-optimum control of third- and fourth-order systems," (abstract), Preprint Volume of the Fourth Joint Automatic Control Conference, June 1963, p. 231.

14 Ho, Y. C., "A successive approximation technique for an optimal control subject to input saturation," Transactions of the ASME, Journal of Basic Engineering, vol. 84, series D, no. 1, March 1962, pp. 101-110.

15 Hopkin, A. M., "A phase plane approach to the compensation of saturating servomechanisms," AIEE Transactions (Communications and Electronics), vol. 70, part I, 1951, pp. 631-639.

16 Kalman, R. E., "Analysis and design of second and higher order saturating servomechanisms," AIEE Transactions (Applications and Industry), vol. 74, part II, 1955, pp. 294-310.

17 Kalman, R. E., "Optimal nonlinear control of saturating systems by intermittent action," IRE WESCON Convention Record, part IV, 1957, pp. 130-135.

18 Koepcke, R. W., "A solution to the sampled minimum-time problem," Preprint Volume of the Fourth Joint Automatic Control Conference, June 1963, pp. 94-100.

19 Krasovskii, N. N., "On an optimal control problem," Priklad. Mat. i. Mekh., vol. 21, no. 5, 1957, pp. 670-677.

20 Krasovskii, N. N., "On the theory of optimal regulation," Avtomat. i. Telemekh., vol. 18, no. 11, 1957, pp. 960-970.

21 Kuo, B. C., Analysis and Synthesis of Sampled-Data Control Systems, Prentice-Hall, Inc., 1964, 528 pp.

22 La Salle, J. P., "Study of the basic principle underlying the 'bang-bang' servo," Goodyear Aircraft Corporation Report, CER-5518, July 1953.

23 Lindorff, D. P., Theory of Sampled-Data Control Systems, New York: John Wiley and Sons, Inc., 1965, 305 pp.

24 Lorchirachoonkul, V., "Optimal control of sampled-data and stochastic distributed-parameter systems," Ph.D. dissertation, Montana State University, Bozeman, Montana, March 1967, 154 pp.

25 Martens, H. R., and H. P. Semmelhack, "Optimum digital control of shipboard plants with stochastic inputs," Recent Advances in Optimization Techniques, edited by A. Lavi and T. P. Vogl, New York: John Wiley and Sons, Inc., 1966, pp. 419-448.

26 Meksawan, J., and G. J. Murphy, "Optimum design of nonlinear sampled-data control systems," Regelungstechnik, vol. 11, no. 7, July 1963, pp. 295-299.

27 McDonald, D., "Nonlinear techniques for improving servo performance," National Electronics Conference, vol. 6, 1950, pp. 400-421.

28 Nagata, A., S. Kodama, and S. Kumagai, "Time optimal discrete control systems with bounded state variables," IEEE Transactions on Automatic Control, vol. AC-10, no. 2, April 1965, pp. 155-164.

29 Neustadt, L. W., "Discrete time optimal control systems," Nonlinear Differential Equations and Nonlinear Mechanics, edited by J. P. La Salle and S. Lefshetz, New York: Academic Press, 1963, pp. 267-283.

30 Pierre, D. A., V. Lorchirachoonkul, and M. E. Ross, "Deadbeat response with minimal overshoot compromise," Preprint no. 3.2-3-65 of the Twentieth Annual ISA Conference, October 1965.

31 Pierre, D. A., V. Lorchirachoonkul, and M. E. Ross, "A performance limit for a class of linear sampled-data control systems," IEEE Transactions on Automatic Control, vol. AC-12, February 1967, pp. 112-113.

32 Polak, E., "Minimum time control of second order pulse width modulated sampled-data systems," Transactions of the ASME, Journal of Basic Engineering, vol. 84, series D, no. 1, March 1962, pp. 107-110.

33 Polak, E., "On the evaluation of optimal and nonoptimal control strategies," IEEE Transactions on Automatic Control, vol. AC-9, no. 2, April 1964, pp. 175-176.

34 Pontryagin, L. S., "Some mathematical problems arising in connection with the theory of optimal automatic control systems. Basic problems of automatic regulation and control," Proceedings of the session of the AN SSSR on the scientific problems of the automation of production, Izvestiya, AN SSSR, 1957.

35 Ragazzini, J. R., and G. E. Franklin, Sampled-Data Control Systems, McGraw-Hill Book Company, Inc., 1958, 331 pp.

36 Reza, F. M., An Introduction to Information Theory, New York: McGraw-Hill Book Company, Inc., 1961, 496 pp.

37 Rohrer, R. A., and M. Sobral, "Sensitivity considerations in optimal systems design," IEEE Transactions on Automatic Control, vol. AC-10, no. 1, January 1965, pp. 43-48.

38 Torng, H. C., "Optimization of discrete control systems through linear programming," Journal of the Franklin Institute, vol. 277, no. 7, July 1964, pp. 28-44.

39 Tou, J. T., Modern Control Theory, New York: McGraw-Hill Book Company, Inc., 1964, 427 pp.

40 Tou, J. T., "Synthesis of discrete systems subject to control signal saturation," Journal of the Franklin Institute, vol. 277, no. 5, May 1964, pp. 401-413.

41 Tou, J. T., and B. Vadhanaphuti, "Optimum control of nonlinear discrete-data systems," AIEE Transactions (Applications and Industry), vol. 80, part II, 1961, pp. 166-171.

42 Whalen, V. H., "Linear programming for optimal control," Ph.D. dissertation, University of California, Berkeley, California, 1963, 69 pp.

43 Zadeh, L. A., "Optimal control problems in discrete-time systems," Computer Control Systems Technology, C. T. Leondes, editor, New York: McGraw-Hill Book Company, Inc., 1961, pp. 389-414.

44 Zadeh, L. A., and B. H. Whalen, "On optimal control and linear programming," IRE Transactions on Automatic Control, vol. AC-7, July 1962, pp. 45-46.