YSTEMS CONTROL, INC.
!80 SHERIDAN AVENUE
P LO ALTO, CALIFORNIA 94306

F Report

May 1972

DUAL CONTROL AND IDENTIFICATION METHODS FOR AVIONIC SYSTEMS — PART I, DUAL CONTROL

Prepared for:
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1400 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22209

CONTRACT F44620-71-C-0077

SCI Project 5971

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Systems Control, Inc. 260 Sheridan Avenue Palo Alto, California 94306	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

DUAL CONTROL AND IDENTIFICATION METHODS FOR AVIONIC SYSTEMS - PART I, DUAL CONTROL

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Final

5. AUTHOR(S) (First name, middle initial, last name)

Edison T.S. Tse, Anthony J. Tether, Yaakov Bar-Shalom, and Lewis Meier

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 1972	95	36

8a. CONTRACT OR GRANT NO. F44620-71-C-0077	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 9769	
c. 61102F	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. 681304	AFOSR-TR-72-1186

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research (NM) 1400 Wilson Blvd. Arlington, Virginia 22209

13. ABSTRACT This report constitutes Part I of a one-year study on the dual control method for nonlinear systems.

The difficulties associated with the general stochastic control problem are discussed, and a new approach to the problem is developed using the concepts of dual control and the principle of optimality from dynamic programming. An adaptive dual control strategy is derived which has the characteristic of actively regulating learning which achieves the control objective. The theoretical development is specialized to a class of problems of controlling a time-varying linear system with random parameters. A specific algorithm is derived for this class of problems, and example problems are presented to demonstrate the computational feasibility of the new algorithm, the performance level of the new algorithm, and to give insight into the present theory.

Finally, several areas of applications for the present results are outlined, and abstracts of technical papers supported by this contract are given.

DD FORM 1473

UNCLASSIFIED

SYSTEMS CONTROL, INC.
260 SHERIDAN AVENUE
PALO ALTO, CALIFORNIA 94306

Final Report *May 1972*

DUAL CONTROL AND IDENTIFICATION METHODS FOR AVIONIC SYSTEMS — PART I, DUAL CONTROL

Principal Investigators: *Edison T. S. Tse*
Anthony J. Tether

Other Contributors: *Yaakov Bar-Shalom*
Lewis Meier

Prepared for:
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1400 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22209

CONTRACT F44620-71-C-0077

SCI Project 5971

Approved: P. E. MERRITT, President
R. E. LARSON, Vice President

Copy No. 4

SYSTEMS CONTROL, INC.
260 SHERIDAN AVENUE
PALO ALTO, CALIFORNIA 94306

TELEPHONE (415)
327-9333

Final Report May 1972

DUAL CONTROL AND IDENTIFICATION METHODS

FOR AVIONIC SYSTEMS - PART I, DUAL CONTROL

Prepared for:

Air Force Office of Scientific Research
1400 Wilson Boulevard
Arlington, Virginia 22209

Contract F44620-71-C-0077

SCI Project 5971

II

SYSTEMS CONTROL, INC.
260 SHERIDAN AVENUE
PALO ALTO, CALIFORNIA 94306

TELEPHONE (415)
327-9333

Final Report

May 1972

DUAL CONTROL AND IDENTIFICATION METHODS FOR AVIONIC SYSTEMS - PART I, DUAL CONTROL

Principal Investigators: Edison T. S. Tse
Anthony J. Tether

Other Contributors: Yaakov Bar-Shalom
Lewis Meier

Prepared for:

Air Force Office of Scientific Research
1400 Wilson Boulevard
Arlington Virginia 22209

Contract F44620-71-C-0077

SCI Project 5971

Approved: P. E. MERRITT, President
R. E. LARSON, Vice President

Copy No. ________________

III

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

In optimal deterministic control theory, the basic assumption is made that the effect of any future control action can be deduced exactly from the present state and the dynamical equation. In many situations, the necessity for control arises from the fact that there are disturbances and/or component failures in the physical system. These random phenomena prevent exact determination of the effect of all future actions, and therefore deterministic theory is not strictly applicable. If the effect of these random phenomena is small, one can still use optimal control theory to obtain a feedback control law based on deterministic considerations. The feedback nature of the control would tend to reduce the sensitivity to uncertainties but would require the state of the system to be measured exactly. Again, this assumption is good only when the measurement error is small in comparison with the signal being measured.

In many cases, the phenomena of uncertainty (including measurement error) can be appropriately modelled as stochastic processes, allowing them to be considered via stochastic optimal control theory. Using the Principle of Optimality one can reduce the stochastic optimal control to that of solving a stochastic Dynamic Programming equation [A1, B1]. Unfortunately this equation cannot be solved numerically in most situations. In this report, a new approach toward a practical solution for stochastic control problems is described. This report represents Part I of a one-year study supported by the Air Force Office of Scientific Research (AFOSR Project No. F44620-71-C-0077): Development of Dual Control and Identification Methods for Avionic Systems. Part II of the study: Input Design for Identification, is discussed in a separate report [M5].

1.1 Main Purposes

In 1960, Feldbaum, in a series of three papers, introduced dual control theory [F2]. His approach is a combination of statistical decision theory and dynamic programming. He pointed out abstractly that the control signal has two purposes that might be conflicting: one is to learn about any unknown parameters and/or the state of the system; the other is to achieve the control objective. Thus the best control must have the characteristic of appropriately distributing its energy for learning and control purposes. However, no further development or algorithms that implement these ideas appear in the literature. Feldbaum used a static example to demonstrate his dual control theory, but it is difficult to visualize how a dual control will work in a dynamic situation. One of the main purposes of this study is to provide a deeper understanding of dual control theory for dynamical systems. Another objective is to develop an approach toward obtaining (or approximating) a near-optimal dual control that can be implemented, with the objective of indicating the potential applications of the results to Air Force problems.

1.2 Outline of the Report

In Section II, optimal stochastic control theory is reviewed and the practical difficulties in computing and realizing the optimal control law are pointed out, both serving as a motivation for the development of the later sections.

In Section III, the stochastic control problem is reformulated in light of the dual nature of the control and a one-step optimal dual control strategy which possesses an active learning characteristic is obtained. This result is new, and in fact is entirely different from the other suboptimal approaches reported in the literature.

In Section IV, the results are specialized to a very important class of problems of controlling a time-varying linear system with random parameters, and a specific algorithm is developed for this class of problems. Since the derived algorithm is rather complicated, illustrative examples are presented to provide understanding of the dual nature of the resulting control strategy.

In Section V, three example problems, described in detail, are intended to demonstrate (1) the computational feasibility of the new algorithm, (2) the performance level of the new algorithm, and (3) to provide more insight into the dual control theory.

In Section VI, potential applications of the results obtained during this research are indicated and recommendations are made for areas for future research.

1.3 Summary of Contributions

A new formulation and a new stochastic control algorithm for general nonlinear stochastic systems has been developed. The algorithm possesses an active learning characteristic that is lacking in the existing suboptimal stochastic control algorithms described in the literature. Simulation studies demonstrate that this algorithm is potentially feasible for large classes of Air Force problems. Sizable improvement over the widely used certainty equivalence suboptimal control policy is demonstrated in the examples being considered. The important class of problems of controlling a linear time-varying system with random parameters is treated in detail, and a specific algorithm for this class of problems is obtained. Simulation studies on some example problems provide certain insights into the dual nature of the control. Also, these examples represent the only complete simulation studies on dual control in the literature.

1.4 Notations

Throughout the report, lower case underscored letters stand for vectors (e.g., $\underline{x}$, $\underline{y}$); upper case underscored letters stand for matrices (e.g., $\underline{A}$, $\underline{B}$). Noise disturbances are denoted by lower case underscored Greek letters (e.g., $\underline{\xi}$, $\underline{\eta}$).

The transpose of a matrix $\underline{A}$ is denoted by $\underline{A}'$. The transpose of a column vector, $\underline{x}$, is a row vector and is denoted by $\underline{x}'$.

Let $\underline{A}$ be an $n \times n$ square matrix; the trace of $\underline{A}$ is defined as

$$\mathrm{tr}\,\underline{A} = \sum_{i=1}^{n} a_{ii} \quad . \tag{1.1}$$

Using the convention that a vector is always in column form one has the gradient operator

$$\nabla_{\underline{\theta}} = \begin{bmatrix} \partial/\partial\theta_1 \\ \vdots \\ \partial/\partial\theta_s \end{bmatrix} \quad ; \quad \nabla_{\underline{x}} = \begin{bmatrix} \partial/\partial x_1 \\ \vdots \\ \partial/\partial x_n \end{bmatrix} \quad . \tag{1.2}$$

The gradient of the scalar function $H(\underline{x},\underline{\theta})$, a column vector, is written as

$$H_{\underline{x}} = \nabla_{\underline{x}} H \; ; \quad H_{\underline{\theta}} = \nabla_{\underline{\theta}} H \quad . \tag{1.3}$$

The Jacobian of the m-vector $\underline{f}$ is the matrix

$$\underline{f}_{\underline{x}} \triangleq \frac{\partial \underline{f}}{\partial \underline{x}} \triangleq \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = (\nabla_{\underline{x}} \underline{f}')' \quad . \tag{1.4}$$

Accordingly,

$$H_{\underline{x}\underline{\theta}} = [\nabla_{\underline{\theta}}(H_{\underline{x}})']' = [\nabla_{\underline{\theta}} \nabla'_{\underline{x}} H]' = \nabla_{\underline{x}} \nabla'_{\underline{\theta}} H = H'_{\underline{\theta}\underline{x}} \tag{1.5}$$

$$H_{\underline{x}\underline{x}} = \nabla_{\underline{x}} \nabla'_{\underline{x}} H \quad . \tag{1.6}$$

The natural base in R^n is denoted by $\{\underline{e}_i\}_{i=1}^n$ where

$$\underline{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow i^{th} \text{ component.} \tag{1.7}$$

II. OPTIMAL STOCHASTIC CONTROL

In this section, the formulation and solution for the optimal stochastic control problem for discrete time systems is discussed, as are the difficulties associated with the solution procedures. These difficulties motivate the specific dual control approach presented in Section III.

2.1 Problem Statement

Consider a discrete-time nonlinear stochastic system described by

$$\underline{x}(k+1) = \underline{f}[k,\underline{x}(k),\underline{u}(k)] + \underline{\xi}(k) \quad ;$$

$$\underline{y}(k) = \underline{h}[k,\underline{x}(k)] + \underline{\eta}(k) \quad , \quad k = 0,1,\ldots,N-1 \tag{2.1}$$

where $\underline{x}(k) \in R^n$, $\underline{u}(k) \in R^r$, and $\underline{y}(k) \in R^m$. It is assumed that $\underline{x}(0)$, $\{\underline{\xi}(k), \underline{\eta}(k+1)\}_{k=0}^{N-1}$ are independent Gaussian vectors with statistics:

$$E\{\underline{x}(0)\} = \hat{\underline{x}}(0|0) \; ; \; Cov\{\underline{x}(0)\} = \underline{\Sigma}(0|0) \tag{2.2}$$

$$E\{\underline{\xi}(k)\} = \underline{0} \quad ; \quad Cov\{\underline{\xi}(k)\} = \underline{Q}(k) \tag{2.3}$$

$$E\{\underline{\eta}(k+1)\} = \underline{0} \quad ; \quad Cov\{\underline{\eta}(k+1)\} = \underline{R}(k+1) \quad . \tag{2.4}$$

Consider further the performance measure

$$J = E\{\psi[\underline{x}(N)] + \sum_{k=0}^{N-1} \mathscr{L}[\underline{x}(k),\underline{u}(k),k]\} \tag{2.5}$$

where the expectation $E\{\cdot\}$ is taken over all underlying random quantities. Finally, consider admissible controls of the feedback type:

$$\underline{u}(k) = \underline{u}(k,Y^k,U^{k-1}); \; Y^k \triangleq \{\underline{y}(1),\ldots,\underline{y}(k)\}; \; U^{k-1} \triangleq \{\underline{u}(0),\ldots,\underline{u}(k-1)\} \tag{2.6}$$

The goal is to find the optimal control sequence $\{\underline{u}^*(k)\}_{k=0}^{N-1}$ that is of the form

(2.6) and minimizes the cost (2.5) subject to the dynamic constraint (2.1).

2.2 Optimal Stochastic Control Solution Method

To solve the optimal control problem stated in Section 2.1, Bayes' rule and dynamic programming are used. A complete derivation for the optimal solution is given by Meier [M1]; therefore, we shall only outline the derivation and summarize the results below.

An important concept is the information state. This can be viewed as a quantity which is equivalent† to the observation process Y^k and all a priori knowledge of the system and U^{k-1} in describing the future evolution of the system. Thus, an information state will summarize all the sufficient information content conveyed by the observation process Y^k, and past control sequence U^{k-1}. Clearly, the combined sequence (Y^k, U^{k-1}) is an information state. If we denote this information state by

$$\mathscr{P}_k^1 = (Y^k,\ U^{k-1}) \tag{2.7}$$

then a recursion relation for $\mathscr{P}_k^1$ is

$$\mathscr{P}_{k+1}^1\left[\mathscr{P}_k^1,\ \underline{u}(k)\right] = \left\{\underline{h}\{k+1, \underline{f}[k,\underline{x}(k),\ \underline{u}(k)] + \underline{\xi}(k) + \underline{\eta}(k+1)\}, \mathscr{P}_k^1, \underline{u}(k)\right\} \tag{2.8}$$

where $\underline{\xi}(k)$, $\underline{\eta}(k+1)$ and $\underline{x}(k)$ are random vectors. Another such information state is the conditional density $\mathscr{P}_k^2 \overset{\Delta}{=} p[\underline{x}(k)|Y^k, U^{k-1}]$. Using Bayes' rule, a recursive equation for the conditional density is given by [A1], [M1]

$$\mathscr{P}_{k+1}^2\left[\mathscr{P}_k^2,\ \underline{u}(k)\right] = \frac{1}{C_k} p[\underline{y}(k+1)|\underline{x}(k+1)] \int p[\underline{x}(k+1)|\underline{x}(k),\ \underline{u}(k)]\mathscr{P}_k^2\, d\underline{x}(k) \tag{2.9}$$

where C_k is a normalizing constant. Next, we can use the principle of optimality in the "information state" space, which gives us

---

† A precise definition of equivalent statistics is given by Streibel [S3].

the stochastic dynamic programming equation (see also Meier [M1]):

$$I^*\{\mathscr{P}_k, k\} = \min_{\underline{u}(k)} E \Big\{ \mathscr{L}[\underline{x}(k), \underline{u}(k), k] + I^*\{\mathscr{P}_{k+1}[\mathscr{P}_k, \underline{u}(k)],\ k+1\} \,|\, Y^k,\ U^{k-1} \Big\} \tag{2.10}$$

where $\underline{u}(k)$ is a <u>deterministic</u> quantity, $\mathscr{P}_k$ is an information state (can be either $\mathscr{P}_k^1$ or $\mathscr{P}_k^2$), and $I^*\{\cdot,k\}$ denotes the optimal cost-to-go associated with the information state at time k. If we use $\mathscr{P}_k^1$ as an information state, then the optimal control can be obtained by solving (2.8) where an optimal feedback table, $\{\underline{u}^*(Y^k, U^{k-1})\}_{k=1}^{N-1}$, is constructed for <u>all possible</u> pairs (Y^k, U^{k-1}), k=1, 1, ..., N-1. On the other hand, if we use $\mathscr{P}_k^2$ as an information state, then the optimal control can be solved by the following separate procedures:

A. <u>Control</u> - The optimum control law is found as a function of the conditional density $p[\underline{x}(k)|Y^k, U^{k-1}]$ by solving the stochastic dynamic programming equation (2.10). In general, this can be an off-line procedure.

B. <u>Estimation</u> - The conditional density is updated by use of the recursion relation (2.9), and the optimum input is obtained from the optimum control law. The updating of the conditional density must be done in real time.

2.3 Difficulties Associated with the Optimal Solution Procedure

Theoretically, the optimal control problem has been solved when equations (2.9) and (2.10) are derived; however, in practice, the problem only begins with these equations. In the following, we discuss the difficulties associated with the solution procedures using either $\mathscr{P}_k^1$ or $\mathscr{P}_k^2$ as the information state. This will motivate our development in the next section.

From (2.7), we note that the dimension of $\mathscr{P}_k^1$ grows linearly in k. Thus even with appropriate quantizing, the number of quantization points, which grows in time will soon become too large to be handled by a computer of any size. Note that the expectation in (2.10) requires the availability of the conditional density, $p[\underline{x}(k)|Y^k, U^{k-1}]$, which is usually infinite dimensional. This adds one more "dimension" of difficulty in carrying out the dynamic procedure. In general, the optimal cost-to-go-function, $I^*[.,.]$, <u>cannot</u> be expressed as an analytical function of the information state. Thus, direct solution of (2.10) becomes practically impossible for any computer. We face the similar kind of difficulty even if we use $\mathscr{P}_k^2$ as the information state. In this case the information state, $\mathscr{P}_k^2$, is usually of infinite dimension for all $k \geq 1$. One may attempt to approximate the solution for (2.10). However, even if this can be done, it still does not solve the dimensionality problem, since in general, the approximate optimal control law is nonlinear in the information state, and can only be expressed as a table look-up type of function of the information state. This prohibits functional realization of the optimal control law, and thus real-time generation of the optimum control value is practically impossible for most problems.

Note that the basic difficulty is in the control rather than in the estimation procedure. The updating of density although a difficult problem in itself, can be reasonably approximated efficiently by using parallel estimation procedures. Some recent results [B3],[T3],[A3],[L1] indicate the feasibility of parallel estimation. We should emphasize the fact that the capability of approximating the conditional density does not solve half the problem because the difficulty in obtaining the optimal control in real time is not so much due to the estimation procedure as to the growth in dimensionality and to the fact that even if an optimal control law is obtained, the extremely large (perhaps infinite) number of possible information states will prevent it from being realizable.

In the special case where the system (2.1) is linear, the conditional density $p[\underline{x}(k)|Y^k,U^{k-1}]$ is equivalent to the conditional mean estimate $\hat{\underline{x}}(k|k)$ (see Streibel [S3], Meier [M1], Tse [T2]), which is a finite dimensional vector generated by the Kalman filter. If in addition, the cost is quadratic, then the optimal cost-to-go $I^*[.,k]$ can be expressed <u>analytically</u> as a function of $\hat{\underline{x}}(k|k)$, so that equation (2.10) can be solved exactly to yield a realizable linear feedback law (Joseph and Tou [J1], Meier, Larson and Tether [M2], Streibel [S3], Tse [T4]). This result is known as the Separation Theorem or Certainty Equivalence Principle.

2.4 Previous Suboptimal Approaches

In the literature, the most popular approximation method used for combined estimation and control is linearization of the plant about the deterministic optimal trajectory and application of the well-known separation theorem to the resulting perturbation equations. However, this may not give good performance if the system is very nonlinear and the noise level is high. This is because with the linearization approach the control action is corrected only after it has been discovered that the trajectory has deviated from the nominal. But, in fact, if it is known that a disturbance will occur in the future, the control should be modified before as well as after the disturbance occurs in order to minimize its effects. Therefore, if linearization is to be used, some nominal trajectory other than the deterministic optimal trajectory should be used. Denham[D1], Meier[M3], and Vander Stoep[V1] considered the problem of choosing a nominal path to minimize a certain cost criteria on using second order analysis of the perturbed system along the nominal path. The advantage of these approaches is the simplicity of the resulting control law. The main drawback is the validity of assuming a nominal trajectory. This assumption is unjustified if there are uncontrollable unknown parameters in the system. A much more "adaptive" type of controller would be desirable.

The open-loop feedback optimal approach suggested by Dreyfus[D2], and applied to specific problems by Tse and Athans[T1], Bar-Shalom and Sivan[B2], Curry[C1], Aoki[A1] and Spang[S2], suffers from the drawback that the resulting control is passive in learning -- the decision of the control action does not anticipate the fact that future learning is possible. An extension of this approach -- the m-measurement feedback control suggested by Curry[C1] -- is only slightly less complicated than the optimal approach. To the authors' knowledge, no successful application of this method has been reported in the literature.

All these approaches take into account the past observation information but ignore the future observation program. In the next section, we shall describe a new method which is based on the Principle of Optimality on the information state $\mathscr{P}_k^2$ and the concept of dual control. In contrast to the previous approaches, this method will not only take explicitly into account the past observation information but also the future observation program.

III. DUAL CONTROL FOR STOCHASTIC SYSTEMS—AN ACTIVE LEARNING PROCEDURE

In Section II, it was noted that the main difficulties in implementing the optimal control law are:

(1) The information state is either infinite dimentional or finite but grows with time,

(2) The optimal cost-to-go associated with the information state is generally a non-analytic function,

(3) Storage of the control value associated with each information state at time k, $k=0,\cdots,N-1$ is practically impossible due to the large dimensionality.

Thus a reasonable suboptimal approach would be to

(1) Reduce the dimension of the information state space so that it stays a constant dimension for all time,

(2) Approximate the optimal cost-to-go associated with each information state at time k, $k=0,\cdots,N-1$,

(3) Compute the control value on-line rather than obtain the feedback law off-line and store the whole "feedback table."

Each of these procedures are discussed in detail in the following subsections. To simplify the discussion, assume that the cost is of the form:

$$\mathscr{L}[\underline{x}(k), \underline{u}(k),k] = L[\underline{x}(k),k] + \phi[\underline{u}(k),k] \quad .$$

The extension to the more general cost is straightforward.

3.1 Wide-Sense Adaptive Control

As discussed in Section II, the information state, $p[\underline{x}(k)|Y^k, U^{k-1}]$ is generally of infinite dimension. One approach to reduce this dimension is to use the "wide-sense" property[D2]; in this approach the controller is restricted to the form

$$\underline{u}(k) = \underline{u}[k,\hat{\underline{x}}(k|k),\underline{\Sigma}(k|k)] \tag{3.1}$$

where

$$\hat{\underline{x}}(k|k) = E\{\underline{x}(k)|Y^k, U^{k-1}\} \tag{3.2}$$

$$\underline{\Sigma}(k|k) = \text{Cov}\{\underline{x}(k)|Y^k, U^{k-1}\} \quad . \tag{3.3}$$

We shall call such a control scheme the wide-sense adaptive control law. The computation of $\hat{\underline{x}}(k|k)$, $\underline{\Sigma}(k|k)$ can be obtained by any one of the following methods:

(1) Extended Kalman Filter[S4],[J2]

(2) Adaptive Filter with Tuning[T5]

(3) Second Order Filter[A2]

(4) Parallel Estimator[B2],[T3],[A3].

Depending on the specific problem under investigation, one of these methods may be more appropriate than the others.

3.2 Perturbation Control and the Dual Cost

Before going into the new approximation procedure, consider first the perturbation control problem and obtain a cost that exhibits the dual property of the control.

The present time is indexed by k. Let us assume that U^{k-1} has been applied to the system, and that the observation sequence Y^k has been obtained. The conditional mean, $\hat{\underline{x}}(k|k)$, and covariance, $\underline{\Sigma}(k|k)$, are assumed available from a learning device, an estimator. Consider a nominal open-loop control sequence $U_o(k,N-1) \triangleq \{\underline{u}_o(j)\}_{j=k}^{N-1}$ and the associated nominal path

$$\underline{x}_o(j+1) = \underline{f}[j,\underline{x}_o(j),\underline{u}_o(j)] \quad ; \qquad j=k,\ldots,N-1 \tag{3.4}$$

with initial condition $\underline{x}_o(k) = \hat{\underline{x}}(k|k)$

Let $\delta\underline{x}(j)$ be a small perturbation about the nominal path due to the disturbance $\underline{\xi}(j)$ and a perturbation control $\delta\underline{u}(j)$. The true trajectory and control are given by

$$\underline{x}(j) = \underline{x}_o(j) + \delta\underline{x}(j) \quad ;$$

$$\underline{u}(j) = \underline{u}_o(j) + \delta\underline{u}(j) \tag{3.5}$$

with $\underline{x}(j)$, $\underline{u}(j)$ satisfying (2.1). Since $\delta\underline{x}(j)$, $\delta\underline{u}(j)$ are assumed to be small, we can approximate the cost-to-go by expanding it up to second order:

$$J(k) \triangleq E\{\psi[\underline{x}(N)] + \sum_{j=k}^{N-1}\left[L[\underline{x}(j),j] + \phi[\underline{u}(j),j]\right]|Y^k\}$$

$$\approx J_o(k) + E\{\psi'_{o,\underline{x}}\delta\underline{x}(N) + \frac{1}{2}\delta\underline{x}'(N)\psi_{o,\underline{x}\underline{x}}\delta\underline{x}(N) + \sum_{j=k}^{N-1}[L'_{o,\underline{x}}(j)\delta\underline{x}(j)$$

$$+\frac{1}{2}\delta\underline{x}'(j)L_{o,\underline{x}\underline{x}}(j)\delta\underline{x}(j) + \phi'_{o,\underline{u}}(j)\delta\underline{u}(j) + \frac{1}{2}\delta\underline{u}'(j)\phi_{o,\underline{u}\underline{u}}(j)\delta\underline{u}(j)]|Y^k\} \quad ; \tag{3.6}$$

where

$$J_o(k) = \psi[\underline{x}_o(N)] + \sum_{j=k}^{N-1} L[\underline{x}_o(j),j] + \phi[\underline{u}_o(j),j] \quad , \tag{3.7}$$

The quantities $\psi_{o,\underline{x}}$ and $\psi_{o,\underline{x}\underline{x}}$ are, respectively, the gradient and Hessian of $\psi(\cdot)$ with respect to $\underline{x}$ evaluated along the nominal trajectory. For a fixed nominal, choosing $\delta\underline{u}_o(j)$, $j=k,\ldots,N-1$ to minimize the incremental cost $\Delta J(k) \triangleq J(k)-J_o(k)$, one obtains a cost $J^*[k,U_o(k,N-1)]$ associated with the nominal control $U_o(k,N-1)$.

Let us consider the perturbation control problem. From (3.6), we have

$$\Delta J(k) \triangleq J(k) - J_o(k) = E\Big\{\psi'_{o,\underline{x}}\delta\hat{\underline{x}}(N|N) + \tfrac{1}{2}\delta\hat{\underline{x}}'(N|N)\psi_{o,\underline{x}\underline{x}}\delta\hat{\underline{x}}(N|N)$$

$$+ \sum_{j=k}^{N-1}[(L'_{o,\underline{x}}(j)\delta\hat{\underline{x}}(j|j) + \tfrac{1}{2}\delta\hat{\underline{x}}'(j|j)L_{o,\underline{x}\underline{x}}\delta\hat{\underline{x}}(j|j) + \phi'_{o,\underline{u}}(j)\delta\underline{u}(j)$$

$$+ \tfrac{1}{2}\delta\underline{u}'(j)\phi_{o,\underline{u}\underline{u}}(j)\delta\underline{u}))]|Y^k\Big\} + \tfrac{1}{2}\mathrm{tr}\Big\{\psi_{o,\underline{x}\underline{x}}\underline{\Sigma}_o(N|N) + \sum_{j=k}^{N-1}L_{o,\underline{x}\underline{x}}(j)\underline{\Sigma}_o(j|j)\Big\}$$

(3.8)

where $\delta\hat{x}(j|j) \triangleq E\{\delta\underline{x}(j)|Y^j\}$ and $\Sigma_o(j|j) \triangleq \mathrm{Cov}\{\delta\underline{x}(j)|Y^j\}$. The problem is to minimize $\Delta J(k)$, subject to the dynamic constraints of the second order incremental process.

Application of dynamic programming with retention of up to second order terms yields the following (the derivations of (3.9)-(3.17) can be found in Appendix A):

$$\delta\underline{u}_o^*(j) = -[H_{o,\underline{u}\underline{u}}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1}\{[\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j)$$

$$+ H_{o,\underline{u}\underline{x}}(j)]\delta\hat{\underline{x}}(j|j) + H_{o,\underline{u}}(j)\} \quad , \qquad (3.9)$$

where

$$H_o(j) \triangleq L[\underline{x}_o(j),j] + \phi[u_o(j),j] + \underline{p}'_o(j+1)\underline{f}_o(j); \underline{f}_o(j) \triangleq f[j,\underline{x}_o(j),\underline{u}_o(j)] \qquad (3.10)$$

$$\underline{p}_o(j) = H_{o,\underline{x}}(j) - [\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) + H_{o,\underline{u}\underline{x}}(j)]'$$

$$[H_{o,\underline{u}\underline{u}}(j) + f'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1}H_{o,\underline{u}}(j) \quad ; \quad \underline{p}_o(N) = \psi_{o,\underline{x}}$$

(3.11)

$$\underline{K}_o(j) = \underline{f}'_{o,x}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) - [\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) + H_{o,\underline{u}\,\underline{x}}(j)]'$$

$$\cdot\ [H_{o,\underline{u}\,\underline{u}}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1}$$

$$\cdot\ [\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) + H_{o,\underline{u}\,\underline{x}}(j)] + H_{o,\underline{x}\,\underline{x}}(j); \quad \underline{K}_o(N) = \psi_{o,\underline{x}\,\underline{x}}$$

(3.12)

and the optimal cost associated with the nominal $U_o(k,N-1)$ is given by

$$J^*[k,U_o(k,N-1)] = J_o(k) + g_o(k) + \frac{1}{2}tr\left\{\psi_{o,\underline{x}\,\underline{x}}\underline{\Sigma}_o(N|N) + \sum_{j=k}^{N-1}\{H_{o,\underline{x}\,\underline{x}}(j)\underline{\Sigma}_o(j|j) + [\underline{\Sigma}_o(j+1|j) - \underline{\Sigma}_o(j+1|j+1)]\underline{K}_o(j+1)\}\right\} \tag{3.13}$$

with $g_o(j)$ satisfying

$$g_o(j) = g_o(j+1) - \frac{1}{2}H_{o,\underline{u}}(j)[H_{o,\underline{u}\,\underline{u}}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1}\underline{H}_{o,\underline{u}}(j) \ ;$$

$$g_o(N) = 0 \tag{3.14}$$

and $\underline{\Sigma}_o(j|j)$ is the future error covariance which is assumed to be generated by the extended Kalman filter:

$$\underline{\Sigma}_o(j+1|j+1) = [\underline{I} - \underline{V}_o(j+1)\underline{h}_{o,\underline{x}}(j+1)]\underline{\Sigma}_o(j+1|j) \quad ; \quad j=k,\ldots,N-1$$

$$\underline{\Sigma}_o(k|k) = \underline{\Sigma}(k|k) \tag{3.15}$$

$$\underline{V}_o(j+1) = \underline{\Sigma}_o(j+1|j)\underline{h}'_{o,\underline{x}}(j+1) \cdot [\underline{h}_{o,\underline{x}}(j+1)\underline{\Sigma}_o(j+1|j)\underline{h}'_{o,\underline{x}}(j+1) + \underline{R}(j+1)]^{-1} \tag{3.16}$$

$$\underline{\Sigma}_o(j+1|j) = \underline{f}_{o,\underline{x}}(j)\underline{\Sigma}_o(j|j)\underline{f}'_{o,\underline{x}}(j) + \underline{Q}(j) \quad . \tag{3.17}$$

Note that the updated and one-step prediction error covariances, $\underline{\Sigma}_o(j+1|j+1)$

and $\underline{\Sigma}_o(j+1|j)$, are dependent on the choice of the nominal control $U_o(k,N-1)$. The cost $J^*[k,U_o(k,N-1)]$ associated with $U_o(k,N-1)$ involves

- Control cost $J_o(k)$
- Estimation cost--the remaining terms involving nonnegative weightings of error covariances

For this reason, $J^*[k,U_o(k,N-1)]$ will be called the dual cost associated with the nominal $U_o(k,N-1)$. We shall comment on the existence of $\underline{K}_o(j)$ and $\underline{p}_o(j)$ in Section 3.4, item 5.

3.3 One-Step Optimal Dual Control

The outline of the one-step optimal dual control procedure, which is the main result of this report, is as follows. It is assumed that at the present time k, one can apply an arbitrary control $\underline{u}(k)$. From time k to k+1 a second order extrapolation is performed and for $j\geq k+1$, the future time, only perturbation analysis about some nominal trajectory is carried out. By assuming that perturbation control will be applied in addition to a nominal from time k+1 to the end of the process, one obtains the expression of the cost (3.13), which includes the future estimation performance. Since this performance depends on the present control $\underline{u}(k)$, the method is to choose the control such as to minimize (3.13) which includes both control performance and estimation performance. It has to be pointed out that the use of the (fictitious) nominal trajectories and perturbations between k+1 and N is with the sole purpose of obtaining the value of the cost-to-go. The procedure is repeated at every step to obtain the value of the control to be used next.

Let $\{\underline{x}_\nu(k+1)\}_{\nu=1}^{\ell}$ be a set of points in the state space that are selected on the basis of past estimation performance. Associated with each $\underline{x}_\nu(k+1)$ is a sequence of nominal controls $\{\underline{u}_\nu(j)\}_{j+k+1}^{N-1}$. The ν^{th} nominal trajectory is obtained by

$$\underline{x}_\nu(j+1) = \underline{f}[j,\underline{x}_\nu(j),\ \underline{u}_\nu(j)] \quad , \quad j=k+1,\ldots,N-1 \quad . \tag{3.18}$$

Next, consider a control $\underline{u}(k)$ to be applied at time k. Expanding the function $\underline{f}[k,\cdot,\underline{u}(k)]$ about $\hat{\underline{x}}(k|k)$ up to second order terms, we have the predicted state and covariance given by

$$\hat{\underline{x}}(k+1|k) = \underline{f}[k,\hat{\underline{x}}(k|k),\underline{u}(k)] + \frac{1}{2}\sum_{i=1}^{n} \underline{e}_i \, tr\{f^i_{\underline{x}\underline{x}}[\hat{\underline{x}}(k|k),\underline{u}(k)]\underline{\Sigma}(k|k)\} \tag{3.19}$$

$$\underline{\Sigma}(k+1|k) = \underline{f}_{\underline{x}}[\hat{\underline{x}}(k|k),u(k)]\underline{\Sigma}(k|k)\underline{f}'_{\underline{x}}[\hat{\underline{x}}(k|k),\underline{u}(k)] + \underline{Q}(k)$$

$$+ \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{n}\underline{e}_i\underline{e}_j' \, tr\{f^i_{\underline{x}\underline{x}}[\hat{\underline{x}}(k|k),\underline{u}(k)]$$

$$\cdot\underline{\Sigma}(k|k)f^j_{\underline{x}\underline{x}}[\hat{\underline{x}}(k|k),\underline{u}(k)]\underline{\Sigma}(k|k)\} \tag{3.20}$$

where $f^i_{\underline{x}\underline{x}}$ denotes the Hessian of the i^{th} component of $\underline{f}$ with respect to $\underline{x}$ and $\{\underline{e}_i\}^n_{i=1}$ is the natural base in R^n. The updated error covariance for the incremental state estimate is:

$$\underline{\Sigma}(k+1|k+1) = \{\underline{I} - \underline{V}(k+1)\underline{h}_{\underline{x}}[k+1,\hat{\underline{x}}(k+1|k)]\}\underline{\Sigma}(k+1|k) \tag{3.21}$$

$$\underline{V}(k+1) = \underline{\Sigma}(k+1|k)\underline{h}_{\underline{x}}[k+1,\hat{\underline{x}}(k+1|k)]'\{\underline{h}_{\underline{x}}[k+1,\hat{\underline{x}}(k+1|k)]\underline{\Sigma}(k+1|k)$$

$$\underline{h}'_{\underline{x}}[k+1,\hat{\underline{x}}(k+1|k)] + \underline{R}(k+1)\}^{-1} \quad . \tag{3.22}$$

If the predicted state, $\hat{\underline{x}}(k+1|k)$, caused by the control $\underline{u}(k)$ is closest to $\underline{x}_\nu(k+1)$, i.e.,

$$\| \hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1) \| \le \| \hat{\underline{x}}(k+1|k) - \underline{x}_{\nu'}(k+1) \| \quad ; \quad \nu' = 1,\ldots,\ell \tag{3.23}$$

then the future analysis will be based on perturbation about the ν^{th} nominal as derived in the previous section. Note that for all admissible $\underline{u}(k)$, there corresponds a nearest nominal such that (3.23) is satisfied.[†] The error covariance $\underline{\Sigma}_\nu(j+1|j+1)$ is given by (3.15)-(3.17) with initial condition $\underline{\Sigma}(k+1|k+1)$, where $\underline{\Sigma}(k+1|k+1)$ is given by (3.21) and (3.22).

Since we assume that for $j \geq k+1$, only perturbation analysis will be carried out along the ν^{th} nominal if (3.23) is satisfied, the optimal cost-to-go time k+1 can be written, on the basis of the results of the previous subsection, as follows (see Appendix A for the derivation):

$$\begin{aligned} I^*[\hat{\underline{x}}(k+1|k+1),\underline{\Sigma}(k+1|k+1),k+1] = & J_\nu(k+1) + g_\nu(k+1) + \frac{1}{2}\mathrm{tr}\Big\{\psi_{\nu,\underline{x}\underline{x}}\underline{\Sigma}_\nu(N|N) \\ & + \sum_{j=k}^{N-1}\{H_{\nu,\underline{x}\underline{x}}(j)\underline{\Sigma}_\nu(j|j) + [\underline{\Sigma}_\nu(j+1|j) \\ & - \underline{\Sigma}_\nu(j+1|j+1)]\underline{K}_\nu(j+1)\}\Big\} + \underline{p}'_\nu(k+1)\tilde{\underline{x}}_\nu(k+1|k+1) \\ & + \frac{1}{2}\tilde{\underline{x}}'_\nu(k+1|k+1)\underline{K}_\nu(k+1)\tilde{\underline{x}}_\nu(k+1|k+1) \end{aligned} \tag{3.24}$$

where

$$\tilde{\underline{x}}_\nu(k+1|k+1) \triangleq \hat{\underline{x}}(k+1|k+1) - \underline{x}_\nu(k+1) \quad . \tag{3.25}$$

Therefore, the cost of applying $\underline{u}(k)$ can be approximated as follows:

$$\begin{aligned} I_1[\underline{u}(k)] &= E\{\phi[\underline{u}(k),k] + L[\underline{x}(k),k] + I^*[\hat{\underline{x}}(k+1|k+1),\underline{\Sigma}(k+1|k+1),k+1]|Y^k\} \\ &= \phi[\underline{u}(k),k] + E\{L[\underline{x}(k),k]|Y^k\} + J_\nu(k+1) + g_\nu(k+1) + \frac{1}{2}\mathrm{tr}\Big\{\psi_{\nu,\underline{x}\underline{x}}\underline{\Sigma}_\nu(N|N) \\ &\quad + \sum_{j=k}^{N-1}\{H_{\nu,\underline{x}\underline{x}}(j)\underline{\Sigma}_\nu(j|j) + [\underline{\Sigma}_\nu(j+1|j) - \underline{\Sigma}_\nu(j+1|j+1)]\underline{K}_\nu(j+1)\Big\} \\ &\quad + \underline{p}'_\nu(k+1)\tilde{\underline{x}}_\nu(k+1|k) + \frac{1}{2}\tilde{\underline{x}}'_\nu(k+1|k)\underline{K}_\nu(k+1)\tilde{\underline{x}}_\nu(k+1|k) \\ &\quad + \frac{1}{2}\mathrm{tr}\{[\underline{\Sigma}(k+1|k) - \underline{\Sigma}(k+1|k+1)]\underline{K}_\nu(k+1)\} \end{aligned} \tag{3.26}$$

[†] If there is more than one ν satisfying (3.23), we may choose any one of them.

where

$$\tilde{\underline{x}}_\nu(k+1|k) = \hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1) \quad . \tag{3.27}$$

Since $E\{L[\underline{x}(k),k]\ Y^k\}$ is independent of $\underline{u}(k)$, minimizing the cost (3.26) is equivalent to minimizing

$$\begin{aligned} J_d[\underline{u}(k)] = {} & J_\nu(k+1) + \phi[\underline{u}(k),k] + g_\nu(k+1) + \underline{p}_\nu'(k+1)[\hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1)] \\ & + [\hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1)]'\underline{K}_\nu(k+1)[\hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1)] \\ & + \frac{1}{2}tr\{[\underline{\Sigma}(k+_{|}k) - \underline{\Sigma}(k+1|k+1)]\underline{K}_\nu(k+1) + \psi_{\nu,\underline{x}\underline{x}}\underline{\Sigma}_\nu(N|N) \\ & + H_{\nu,\underline{x}\,\underline{x}}(k+1)\underline{\Sigma}(k+1|k+1) + \sum_{j=k+2}^{N-1} H_{\nu,\underline{x}\,\underline{x}}(j)\underline{\Sigma}_\nu(j|j) \\ & + \sum_{j=k+1}^{N-1}[\underline{\Sigma}_\nu(j+1|j) - \underline{\Sigma}_\nu(j+1|j+1)]\underline{K}_\nu(j+1)\} \end{aligned} \tag{3.28}$$

subject to the constraints (3.19), (3.20), (3.21), (3.22) and (3.23) and where $\underline{\Sigma}_\nu(j+1|j)$, $\underline{\Sigma}_\nu(j+1|j+1)$ are given by (3.17) and (3.15), respectively, with initial condition

$$\underline{\Sigma}_\nu(k+1|k+1) = \underline{\Sigma}(k+1|k+1) \quad . \tag{3.29}$$

The procedure for computing $J_d[\underline{u}(k)]$ is also described in Fig. 3.1.

One can extend this to the situation in which a nominal control sequence $\{\underline{u}_o[j;\hat{\underline{x}}(k+1|k)]\}_{j=k+1}^{N-1}$ is associated with each predicted state $\hat{\underline{x}}(k+1|k)$. Thus if a control $\underline{u}(k)$ yields $\hat{\underline{x}}(k+1|k)$, future analysis will be carried out around the nominal control $\{\underline{u}_o[j;\hat{\underline{x}}(k+1|k)]\}_{j=k+1}^{N-1}$ and the nominal trajectory

$$\underline{x}_o[j+1;\hat{\underline{x}}(k+1|k)] = \underline{f}\{j,\underline{x}_o[j;\hat{\underline{x}}(k+1|k)],\underline{u}_o[j;\hat{\underline{x}}(k+1|k)]\} \quad ;$$

$$j = k+1,\ldots,N-1 \quad ;$$

$$\underline{x}_o[k+1;\hat{\underline{x}}(k+1|k)] \triangleq \hat{\underline{x}}(k+1|k) \quad . \tag{3.30}$$

In this case, $J_d[\underline{u}(k)]$ becomes

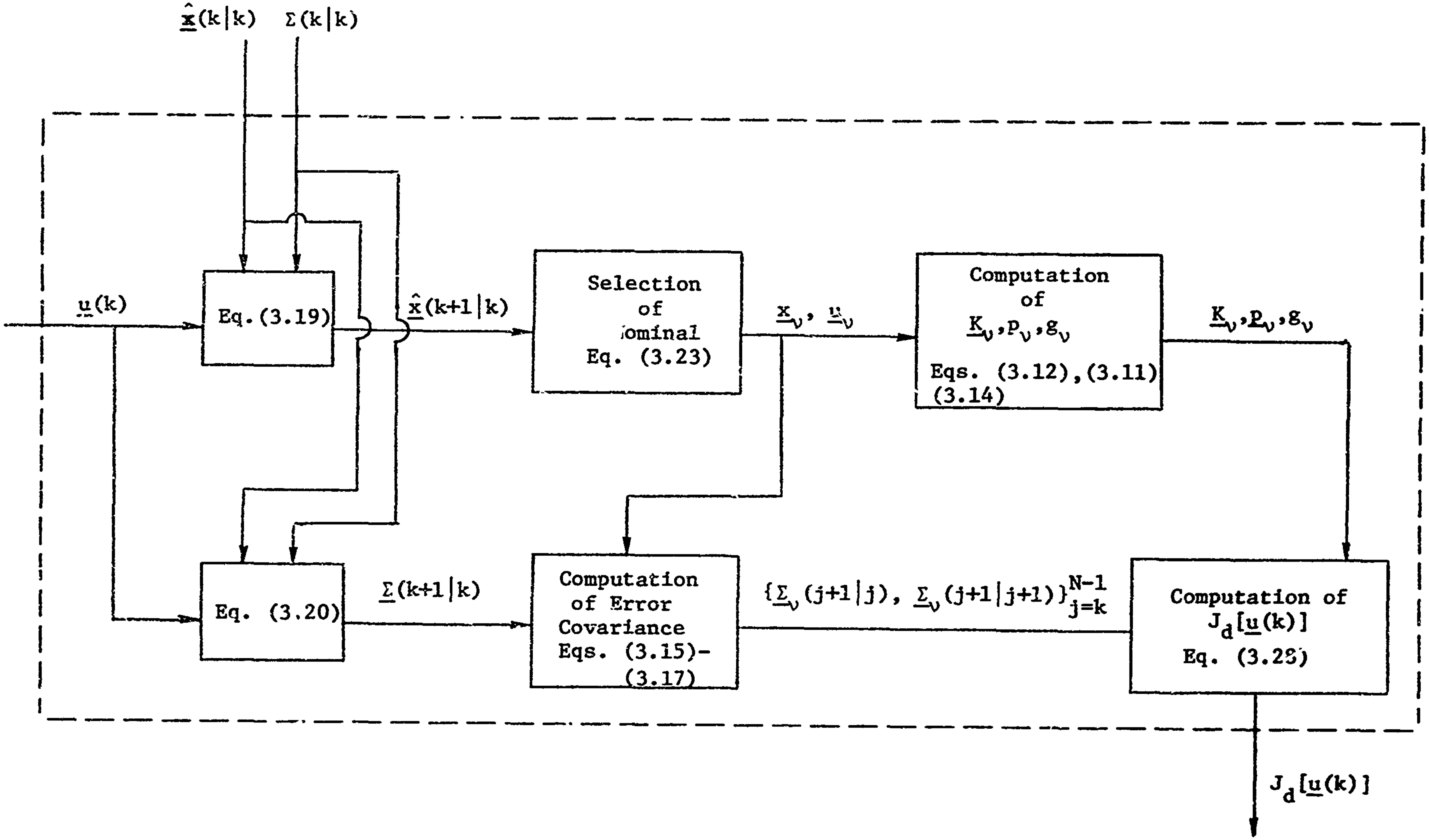

FIGURE 3.1 COMPUTATION OF THE ONE-STEP DUAL COST $J_d[\underline{u}(k)]$

$$J_d[\underline{u}(k)] = J_o(k+1) + \phi[\underline{u}(k),k] + g_o(k+1) + \frac{1}{2}\mathrm{tr}\Big\{[\underline{\Sigma}(k+1|k) - \underline{\Sigma}(k+1|k+1)]\underline{K}_o(k+1)$$

$$+ \psi_{o,\underline{x}\underline{x}}\underline{\Sigma}_o(N|N) + H_{o,\underline{x}\underline{x}}(k+1)\underline{\Sigma}(k+1|k+1) + \sum_{j=k+2}^{N-1} H_{o,\underline{x}\underline{x}}(j)\underline{\Sigma}_o(j|j)$$

$$+ \sum_{j=k+1}^{N-1}[\underline{\Sigma}_o(j+1|j) - \underline{\Sigma}_o(j+1|j+1)]\underline{K}_o(j+1)\Big\} \quad . \tag{3.31}$$

Depending on the problem under consideration, one may or may not want to discretize the state space for the predicted state.

Denote the optimal solution for the above one-step optimization problem by $\underline{u}^*(k)$. When $\underline{u}^*(k)$ is applied to the system and a new observation $\underline{y}(k+1)$ is obtained, the estimate of $\underline{x}(k+1)$ and its error covariance are updated and the same procedure is repeated to obtain $\underline{u}^*(k+1)$. Starting with k=0 to k=N-1, we obtain a sequence of controls $\{\underline{u}^*(k)\}_{k=0}^{N-1}$ which is called the <u>one-step optimum dual control</u>.

Note that in the above development, the choice of future nominal control is "fictitious"; it is only used to approximate the optimal cost-to-go function. Therefore, its choice is quite flexible and is dependent upon the problem under consideration. In Section IV, we indicate how these nominals can be selected for a special class of problems.

3.4 Remarks

1. Note that in most cases, J_d given in (3.28) or (3.31) cannot be expressed explicitly as a function of $\underline{u}(k)$; therefore, straightforward minimization techniques, such as taking the derivative with respect to $\underline{u}(k)$ and setting it to zero, would be of no use. Because of the rather complicated dependence of J_d on $\underline{u}(k)$, one has to search to find the minimizing $\underline{u}(k)$ which will be applied to the system. Search methods appropriate for finding $\underline{u}(k)$ are those of local variations or, if the control is a scalar, then a line search, e.g., Fibonacci. To obtain $\underline{u}^*(k)$, start the search at $\underline{u}_{ce}(k)$, the first of the sequence of controls obtained by assuming certainty equivalence (i.e., the separation theroem)

to be valid. Then determine which direction J_d decreases, next the "box" in which the minimum lies, and then narrow it down to a certain predetermined size, and finally make a quadratic interpolation from the last three points; the result is taken as $\underline{u}^*(k)$. A search procedure is described in Appendix B.

2. The approach described in this section requires appropriate selection of nominal controls, an essential in approximating the future optimal cost-to-go. Note that these nominals are not applied in the future, but only to give a rough idea of the optimum cost corresponding to future learning and control. This flexibility is a distinguished feature of our approach. One may consider this to be an advantage or disadvantage, depending on one's viewpoint. Clearly, such an approach will not be of use to a designer who knows nothing about the system he is controlling, since he is unable to select a set of appropriate nominal controls. However, an engineer who is familiar with the system he is controlling can use his heuristic knowledge to select the nominal controls. For him, this approach is of great value, because it makes use of his knowledge to come up with a good control strategy in a systematic manner. Thus, in some sense, the approach bears some characteristic of heuristic programming methods,[N1] where use is made of knowledge of the system to reduce the dimensionality of the program.

3. Let us comment on the dual nature of the control. The estimation purpose of the control is reflected by the covariances appearing in (3.20)-(3.22). If the predicted and updated error covariances are independent of the control, the dual property will disappear. This would be the case if the system is linear (with known parameters). In general, this dual property of the control is important.

4. We shall also distinguish two different types of learning procedures. Note that if the function $\underline{f}_{\underline{x}}(\underline{x},\underline{u})$ is not a function of the control $\underline{u}$ (e.g., when $\underline{f}(\underline{x},\underline{u}) = \underline{f}(\underline{x}) + \underline{g}(\underline{u})$ and the measurements are linear

then the error covariances $\underline{\Sigma}(k+1|k)$ and $\underline{\Sigma}(k+1|k+1)$ will be independent of the control action at time k; (see (3.20) and (3.21)). The control does not influence the estimation performance in one step, but the effect of the control in future estimation will appear n steps (n>1) after the time it is applied; (note the dependence on the nominal in (3.15)-(3.17)). In this case, the control has the capability of exciting certain modes of the system that will, in the future, enhance the estimation. A typical example is the problem of controlling a linear system with known zeroes but unknown poles. In the second case, if $\underline{f}_{\underline{x}}(\underline{x},\underline{u})$ is a function of $\underline{u}$, then the error covariances $\underline{\Sigma}(k+1|k)$ and $\underline{\Sigma}(k+1|k+1)$ will both be dependent on the control action. Besides exciting certain modes of the system, the control also has the capability of directly regulating the signal-to-noise ratio and isolating the effects of different parameters. A typical example is the problem of controlling a system where the control multiplies the state and/or some unknown parameters of the system. Thus, we see that the control is "actively adaptive" since it regulates its learning in an optimal manner.

5. A sufficient condition for $\underline{K}_o(j)$, $\underline{p}_o(j)$ and $g_o(j)$ ((3.12), (3.11), (3.14)) to exist is

$$H_{o,\underline{u}\,\underline{u}}(j) > 0 \quad . \tag{3.32}$$

Let us consider the deterministic control problem of minimizing the performance $J_o(k)$ given by (3.7) for the system described by (3.4). The Hamiltonian for this problem is given by (3.10); therefore, from (3.11)-(3.14), $\underline{p}_o(j)$ is the adjoint variable, $\underline{K}_o(j)$ is the return matrix for the linear quadratic control problem whose state equation is (3.4) linearized about the nominal and whose cost matrices are the second derivatives of the Hamiltonian evaluated along the nominal, and the quantity $J_o(k)$ is the deterministic performance when the initial state is $\underline{x}_o(k)$ and the nominal control is used. Thus condition (3.32) is equivalent to the existence of neighboring stationary paths about the nominal trajectory.[B4] In general, if the deterministic control problem has a solution and if the nominal control trajectory is that solution, then $\underline{H}_{o,uu}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)$ will be positive semi-definite. Where this deterministic optimal control is non-singular, this matrix is positive definite and thus invertible; and where it is singular the inverse should be replaced by the pseudo-inverse. For a general nominal trajectory no such statements can be made; however, for nominal trajectories near the deterministic optimum, one would expect similar properties to hold. Thus, one reasonable choice of the set $\{U_\nu(k+1,N-1)\}_{\nu=1}^{\ell}$ would be the deterministic optimal controls associated with $\{\underline{x}_\nu(k+1)\}_{\nu=1}^{\ell}$. In the special case where $\underline{f}[k,\underline{x}(k),\underline{u}(k)]$ is first order in the control and $\phi(u,j)$ is strictly convex in u, then $H_{o,\underline{u}\,\underline{u}}(j) = \phi_{o,\underline{u}\,\underline{u}}(j)$ which is positive definite by the convexity of ϕ; therefore in this special case the matrix can be inverted.

6. The results hold even when the cost has the more general form (2.5). The only change one needs to make is to replace (3.10) by

$$H_o(j) \triangleq \mathscr{L}[\underline{x}_o(j),u_o(j)] + \underline{p}'_o(j+1)\underline{f}[f,\underline{x}_o(j),\underline{u}_o(j)] \qquad (3.10)'$$

and the term $\phi[\underline{u}(k)]$ in (3.28) by

$$E\{\mathscr{L}[\underline{x}(k),\underline{u}(k),k]\,|\,Y^k\} \approx \mathscr{L}[\hat{\underline{x}}(k|k),\underline{u}(k),k] + \mathrm{tr}\{\mathscr{L}_{\underline{x}\,\underline{x}}[\hat{\underline{x}}(k|k),\underline{u}(k),k]\underline{\Sigma}(k|k) \quad . \qquad (3.33)$$

IV. ACTIVELY ADAPTIVE CONTROL FOR STOCHASTIC LINEAR SYSTEMS WITH RANDOM PARAMETERS VIA DUAL CONTROL

In this section, we consider the control of linear systems with unknown parameters, a class of problem of major theoretical and practical importance. A control strategy that regulates its speed of learning (i.e., the adaptivity is not passive but active) is obtained for this class of problems by specializing the results of section III.

4.1 Problem Statement

Consider a discrete-time linear system described by

$$\underline{x}(k+1) = \underline{A}[k,\underline{\theta}(k)]\underline{x}(k) + \underline{b}[k,\theta(k)]u(k) + \underline{\xi}(k)$$

$$\underline{y}(k) = \underline{C}[k,\theta(k)]\underline{x}(k) + \underline{n}(k) \qquad k=0,1,\ldots \tag{4.1}$$

where $\underline{x}(k)\varepsilon R^n$, $\underline{y}(k)\varepsilon R^m$, $\underline{\theta}(k)\varepsilon R^s$ and $u(k)$ is a scalar control.[†] It is assumed that $\underline{\theta}(k)$ is a Markov process satisfying

$$\underline{\theta}(k+1) = \underline{D}(k)\underline{\theta}(k) + \underline{\gamma}(k) \qquad k=0,1,\ldots \tag{4.2}$$

where $\underline{D}(k)$ is a known matrix. The vectors $\{\underline{x}(0),\ \underline{\theta}(0),\ \underline{\xi}(k),\ \underline{n}(k+1),\ \underline{\gamma}(k),\ k=0,1,\ldots\}$ are assumed to be mutually independent Gaussian random variables with known statistical laws:

$$\underline{x}(0)\sim\mathscr{G}[\hat{\underline{x}}(0),\underline{\Sigma}^{xx}(0)];\ \underline{\theta}(0)\sim\mathscr{G}[\hat{\underline{\theta}}(0),\underline{\Sigma}^{\theta\theta}(0)];\ \underline{\xi}(k)\sim\mathscr{G}[\underline{0},\underline{Q}(k)]$$

$$\underline{n}(k)\sim\mathscr{G}[\underline{0},\underline{R}(k)];\ \underline{\gamma}(k)\sim\mathscr{G}[\underline{0},\underline{G}(k)] \tag{4.3}$$

with $\underline{\Sigma}^{xx}(0)>\underline{0}$, $\underline{\Sigma}^{\theta\theta}(0)>\underline{0}$, $\underline{R}(k)>\underline{0}$, $\underline{Q}(k)\geq\underline{0}$, $\underline{G}(k)\geq\underline{0}$. The notation $\underline{v}\sim\mathscr{G}(\underline{a},\underline{B})$ is used to denote that the random vector $\underline{v}$ is Gaussian with mean $\underline{a}$ and covariance $\underline{B}$. Furthermore, we assume that the unknown parameter $\underline{\theta}(k)$ enters linearly in $\underline{A}(k,\cdot)$, $\underline{b}(k,\cdot)$ and $\underline{C}(k,\cdot)$.

[†]For simplicity, we shall discuss only the scalar input case. The results can be readily extended to the multi-input case. See also Section 4.5.

A control is admissible if it is non-anticipative; i.e.,

$$u(k) = u(k, Y^k, U^{k-1}) \; ; \; Y^k \triangleq \{\underline{y}(1), \cdots, \underline{y}(k)\}; \; U^{k-1} \triangleq \{u(1), \cdots u(k-1)\} \quad (4.4)$$

Our objective is to find an admissible control sequence U^{N-1} such that the cost functional

$$J(U) = \frac{1}{2} E \left\{ [\underline{x}(N) - \underline{\rho}(N)]' \underline{W}(N) [\underline{x}(N) - \underline{\rho}(N)] + \sum_{k=0}^{N-1} [\underline{x}(k) - \underline{\rho}(k)]' \right.$$

$$\left. \cdot \underline{W}(k) [\underline{x}(k) - \underline{\rho}(k)] + \lambda(k) u^2(k) \right\} \quad (4.5)$$

is minimized subject to the dynamic constraints (4.1) and (4.2). The expectation in (4.5) is over all the underlying random quantities $\underline{x}(0)$, $\underline{\theta}(0)$, $\{\underline{\xi}(k), \underline{\eta}(k+1), \underline{\gamma}(k), k=0,1,\cdots,N-1\}$. Assume the following:

1. $\underline{W}(k) \geq \underline{0}$ and $\lambda(k) > 0$,

2. $\{\underline{\rho}(k), k=0,1,\cdots,N\}$ is given a priori.

Note that if $\underline{\rho}(k) = \underline{0}$, $k=0,1,\cdots,N$, we have a regulator problem; if $\{\underline{\rho}(k)\}_{k=0}^{N}$ is a given trajectory, we have a tracking problem; and if $\underline{W}(k) = \underline{0}$, $k=0,\cdots,N-1$ but $\underline{W}(N) \neq \underline{0}$, we have an interception problem.

4.2 Previous Approaches

Before describing our new approach to this class of problem it is appropriate to summarize some of the past approaches and indicate how this work fits into the whole development.

This problem can be solved exactly if one can solve the stochastic dynamic programming equation (2.8) associated with the problem; unfortunately, a numerical solution for this is prohibited by the "curse of dimensionality" (see Section II). Thus different approaches have been suggested in treating this class of problem.

One popular approach is the certainty equivalence[A1]. If at a time instant, the estimates of the unknown parameters are available, a control law can be obtained by assuming the estimated parameters to be the true ones and solving the control problem accordingly. In this manner, we obtain a control law which is adaptive to the estimates. The problem now is reduced to that of closed-loop parameter estimation. Such an approach has been considered by Farison, et al.,[F1] Saridis and Lobbia.[S1] The question now is not "how to control the system," but rather "how well can we estimate the parameters." The advantage of this approach is the simplicity of the control law. The major drawback to the approach is that we are ignoring the confidence level on the parameter estimates in deriving the adaptive control scheme; one would expect that such a control scheme will result in a control system which is extremely sensitive to stochastic variations, which turns out to be the case.

If the design of adaptive systems takes not only the instantaneous parameter estimates but also the associated confidence levels into account, it would surely result in a "better" system. One such method is the open-loop feedback approach[D1]. Typical papers along this line are those by Bar-Shalom and Sivan[B2], Curry[C1], Aoki[A1], Spang[S2], and Tse and Athans[T1]. In the last-mentioned, it was demonstrated that in the case where only the input gain vector is unknown, the adaptive feedback gains of the control system depend upon the parameter error covariance matrix. In this open-loop feedback approach, the fact that the estimated parameter may not be exact is therefore taken into consideration, but the knowledge of future observation programs is completely ignored. The problem when the system is linear with unknown parameters that belong to a finite set has been studies by Stein and Saridis[S5] and Lainiotis, et al.[L2] Their solution was also of the open-loop feedback type because it did not take into account the effect of the control on the future estimation performance.

Yet another approach is to approximate the dynamic programming equation. Murphy[M1], Gorman and Zaborsky[G1] used this approach in considering the situation where the gain vector is unknown. To the aughors' knowledge, the extension of this approach to more general situations is not found in the literature.

The approach described in this section is based on the one-step optimal dual control theory developed in Section III. As we have noted, such a control scheme has the characteristic of appropriately distributing its energy for learning and control purposes. In view of this, it is obvious that the open-loop beedback control is, from the estimation point of view, passive. In contrast, the one-step optimal dual control is active, not only for the control purpose but also for the estimation purpose, because the performance depends also on the "quality" of the estimates. Therefore, the one-step optimal dual control can be called "actively adaptive" since it regulates its adaptation (learning) in a systematic manner.

4.3 The Optimal Cost-to-Go and the Dual Effect

In this subsection, the results developed in Section III will be specialized to the class of problems being considered here to obtain the approximate optimal cost-to-go.

Let the present time be denoted by k. Given a point represented by the augmented state $\underline{z}_o(k+1) \triangleq [\underline{x}_o'(k+1), \underline{\theta}_o'(k+1)]'$ in the augmented state space, one associates with it a nominal control sequence denoted by $\{u_o[j;\underline{z}_o(k+1)]\}_{j=k+1}^{N-1}$. A nominal trajectory originating from $\underline{z}_o(k+1)$ is generated by applying the above control sequence. Consider a control u(k) applied at time k and the resulting predicted state and covariance, denoted as $\hat{\underline{z}}(k+1|k)$ and $\underline{\Sigma}(k+1|k)$, respectively. In order to bring out the dual effect of the control, assume that for time $j\geq k+1$, a second order perturbation analysis will be carried out about the nominal trajectory originating at $\underline{z}_o(k+1) \triangleq \hat{\underline{z}}(k+1|k)$ with a certain nominal control sequence. The details on how this nominal is obtained are given in Section 4.4. The subscript "o" is used to denote both "nominal" control $\{u_o[j; \hat{\underline{z}}(k+1|k)]\}_{j=k+1}^{N-1}$ and the associated nominal trajectory $\{z_o[j; \hat{\underline{z}}(k+1|k)]\}_{j=k+1}^{N-1}$. In this manner, one obtains an approximate optimal "cost-to-go" $I^*[\hat{\underline{z}}(k+1|k), \underline{\Sigma}(k+1|k),k+1]$ associated with $\hat{\underline{z}}(k+1|k)$ and $\underline{\Sigma}(k+1|k)$, which is a function of u(k). This cost reflects both the future <u>estimation performance</u> and <u>control performance</u>. The minimization of this cost yields u*(k) and the procedure is repeated at every step.

Assume that the one-step prediction $\hat{\underline{z}}(k+1|k)$ and the associated error covariance

$$\underline{\Sigma}(k+1|k) \triangleq \text{cov}\{\underline{z}(k+1)|k\} \tag{4.6}$$

have been obtained (using, e.g., a second order filter) when a certain control u(k) is applied to the system. Let $\{\underline{z}_o(j)\}_{j=k+1}^{N}$ be the nominal trajectory obtained by applying the nominal control sequence $\{u_o[j;\ \hat{\underline{z}}(k+1|k)]\}_{j=k+1}^{N-1}$ to the deterministic part of the system (4.1), i.e.,

$$\underline{z}_o(j+1) \triangleq \begin{bmatrix} \underline{x}_o(j+1) \\ \underline{\theta}_o(j+1) \end{bmatrix} = \underline{f}_o(j) \triangleq \begin{bmatrix} \underline{f}_o^x(j) \\ \underline{f}_o^\theta(j) \end{bmatrix} \triangleq \begin{bmatrix} \underline{A}_o(j)\underline{x}_o(j)+\underline{b}_o(j)u_o(j) \\ \underline{D}(j)\underline{\theta}_o(j) \end{bmatrix} \tag{4.7}$$

where superscripts denote matrix partitions and

$$\begin{aligned} \underline{A}_o(j) &\triangleq \underline{A}[j,\underline{\theta}_o(j)] \\ \underline{b}_o(j) &\triangleq \underline{b}[j,\underline{\theta}_o(j)] \end{aligned} \tag{4.8}$$

with initial condition $\underline{z}_o(k+1) \triangleq \hat{\underline{z}}(k+1|k)$. For simplicity, the dependence on $\hat{\underline{z}}(k+1|k)$ will be suppressed and $u_o[j,\ \hat{\underline{z}}(k+1|k)]$ denoted by $u_o(j)$.

Define the Jacobian

$$\underline{f}_{o,\underline{z}}(j) \triangleq (\nabla_{\underline{z}}\ \underline{f}')'\Big|_{\underline{z}=\underline{z}_o} \triangleq \left\{ \begin{bmatrix} \nabla_{\underline{x}} \\ \nabla_{\underline{\theta}} \end{bmatrix} [\underline{f}^{x'}\quad \underline{f}^{\theta'}] \right\}'\Bigg|_{z=z_o}$$

$$= \begin{bmatrix} \underline{f}^x_{o,\underline{x}}(j) & \underline{f}^x_{o,\underline{\theta}}(j) \\ \underline{f}^\theta_{o,\underline{x}}(j) & \underline{f}^\theta_{o,\underline{\theta}}(j) \end{bmatrix} = \begin{bmatrix} \underline{A}_o(j) & \sum_{i=1}^{n} \underline{e}_i \underline{x}'\ \underline{a}^i_{\underline{\theta}}(j) + u_o(j)\ \underline{b}_{\underline{\theta}}(j) \\ 0 & \underline{D}(j) \end{bmatrix}. \tag{4.9*}$$

---

* Since $\underline{\theta}$ enters linearly in $\underline{A}$, $\underline{C}$ and $\underline{b}$, their partials with respect to $\underline{\theta}$ are constants.

The measurement vector in (4.1) can be written in terms of the augmented state

$$\underline{h}(j) = [\underline{C}(j,\theta(j)) \mid \underline{0}]\ \underline{z}(j) \qquad (4.9a)$$

and its Jacobian evaluated along the nominal is

$$\underline{h}_{o,\underline{z}} = [\underline{C}_o(j) \mid \sum_{i=1}^{m} \underline{e}_i\ \underline{x}'\ \underline{c}^i_{\underline{\theta}}\ ]\ ;\ \underline{C}_o(j) \triangleq \underline{C}(j;\ \underline{\theta}_o(j)) \qquad (4.9b)$$

where $\underline{a}^{i'}$ $i=1,\cdots,n$ and $\underline{c}^{i'}$ $i=1,\cdots,m$ are the corresponding rows of $\underline{A}$ and $\underline{C}$, respectively. Similarly,

$$\underline{f}_{o,u}(j) = \begin{bmatrix} \underline{b}_o(j) \\ \underline{0} \end{bmatrix} \quad . \qquad (4.10)$$

and b^i will denote the corresponding component of $\underline{b}$.

Using the results in Section III, the approximate optimal cost-to-go I^* is given by (see also Appendix C)

$$\begin{aligned} I^*[\hat{\underline{z}}(k+1|k),\ \underline{\Sigma}(k+1|k),\ k+1] &= J_o(k+1) + g_o(k+1) \\ &+ \frac{1}{2} \operatorname{tr}\{[\underline{\Sigma}(k+1|k) - \underline{\Sigma}_o(k+1|k+1)]\ \underline{K}_o(k+1) \\ &+ \sum_{j=k+1}^{N} \mathcal{U}(j)\ \underline{\Sigma}_o(j|j) + \sum_{j=k+1}^{N-1} [\underline{\Sigma}_o(j+1|j) - \underline{\Sigma}_o(j+1|j+1)]\underline{K}_o(j+1)\} \end{aligned} \qquad (4.11)$$

where

$$\begin{aligned} J_o(k+1) &= \frac{1}{2}\ [\underline{x}_o(N) - \underline{\rho}(N)]'\ \underline{W}(N)[\underline{x}_o(N) - \underline{\rho}(N)] \\ &+ \frac{1}{2} \sum_{j=k+1}^{N-1} [\underline{x}_o(j) - \underline{\rho}(j)]'\underline{W}(j)\ [\underline{x}_o(j) - \underline{\rho}(j)] \\ &+ \frac{1}{2}\ \lambda(j)\{u_o[j;\ \hat{\underline{z}}(k+1|k)]\}^2 \quad . \end{aligned} \qquad (4.12)$$

The cost matrices corresponding to the augmented state $\underline{z} = [\underline{x}'\underline{\theta}']$ are denoted by

$$\mathcal{W}(j) = \begin{bmatrix} \underline{W}(j) & \underline{0}^{ns} \\ \underline{0}^{sn} & \underline{0}^{ss} \end{bmatrix} \tag{4.13}$$

where $\underline{0}^{ns}$ denotes an nXs zero matrix and $\underline{K}_o(j)$, $g_o(j)$ satisfy the backward equations

$$\underline{K}_o(j) \triangleq \begin{bmatrix} \underline{K}_o^{xx}(j) & \underline{K}_o^{\theta x'}(j) \\ \underline{K}_o^{\theta x}(j) & \underline{K}_o^{\theta\theta}(j) \end{bmatrix} \tag{4.14}$$

$$\begin{aligned} \underline{K}_o^{xx}(j) = {} & \underline{A}_o'(j)[I - \mu_o(j)\underline{K}_o^{xx}(j+1)\underline{b}_o(j)\underline{b}_o'(j)]\underline{K}_o^{xx}(j+1)\underline{A}_o(j) \\ & + \underline{W}(j) \quad ; \qquad \underline{K}_o^{xx}(N) = \underline{W}(N) \end{aligned} \tag{4.15}$$

$$\begin{aligned} \underline{K}_o^{\theta x}(j) = {} & [\underline{f}_{o,\underline{\theta}}^{x'}(j)\ \underline{K}_o^{xx}(j+1) + \underline{D}'(j)\ \underline{K}_o^{\theta x}(j+1)]\underline{A}_o(j) \\ & - \mu_o(j) \Bigg\{ [\underline{f}_{o,\underline{\theta}}^{x'}(j)\ \underline{K}_o^{xx}(j+1) + \underline{D}'(j)\ \underline{K}_o^{\theta x}(j+1)]\underline{b}_o(j) \\ & + \left[\sum_{i=1}^{n} \underline{e}_i'\ \underline{p}_o^{x}(j+1)\ b_{\underline{\theta}}^{i}(j)\right]'\Bigg\} \Big\{\underline{b}_o'(j)\ \underline{K}_o^{xx}(j+1)\ \underline{A}_o(j)\Big\} \quad ; \quad \underline{K}_o^{\theta x}(j) = \underline{0} \end{aligned} \tag{4.16}$$

$$\begin{aligned} \underline{K}_o^{\theta\theta}(j) = {} & \underline{f}_{o,\underline{\theta}}^{x'}(j)\ \underline{K}_o^{xx}(j+1)\ \underline{f}_{o,\underline{\theta}}^{x}(j) + \underline{D}'(j)\ \underline{K}_o^{\theta x}(j+1)\ \underline{f}_{o,\theta}^{x}(j) \\ & + \underline{f}_{o,\underline{\theta}}^{x'}(j)\ \underline{K}_o^{x\theta}(j+1)\ \underline{D}(j) + \underline{D}'(j)\ \underline{K}_o^{\theta\theta}(j+1)\ \underline{D}(j) - \\ & -\mu_o(j) \Bigg\{ \underline{b}'(j)[\underline{K}_o^{xx}(j+1)\underline{f}_{o,\underline{\theta}}^{x}(j) + \underline{K}_o^{x\theta}(j+1)\underline{D}(j)] \\ & + \sum_{i=1}^{n} \underline{e}_i'\ \underline{p}_o^{x}(j+1)\ b_{\underline{\theta}}^{i}(j)\Bigg\}\Bigg\{\underline{b}'(j)[\underline{K}_o^{xx}(j+1)\underline{f}_{o,\underline{\theta}}^{x}(j) + \underline{K}_o^{x\theta}(j+1)\underline{D}(j)] \\ & + \sum_{i=1}^{n} \underline{e}_i'\ \underline{p}_o^{x}(j+1)\ b_{\underline{\theta}}^{i}(j)\Bigg\} \quad ; \quad \underline{K}_o^{\theta\theta}(N) = \underline{0} \end{aligned} \tag{4.17}$$

$$\mu_o(j) \triangleq [\lambda(j) + \underline{b}_o'(j)\ \underline{K}_o^{xx}(j+1)\ \underline{b}_o(j)]^{-1} \tag{4.18}$$

$$g_o(j) = g_o(j+1) - \mu_o(j)[\lambda(j)\ u_o(j) + \underline{p}_o^{x'}(j+1)\ \underline{b}_o(j)]^2; \quad g_o(N) = 0 \tag{4.19}$$

$$\underline{p}_o^x(j) = \underline{A}_o'(j)\ \underline{p}_o^x(j+1) + \underline{W}(j)[\underline{x}_o(j) - \underline{\rho}(j)] - \mu_o(j)\ \underline{A}_o'(j) \cdot \underline{K}_o^{xx}(j+1)\ \underline{b}_o(j)\ [\lambda(j)\ u_o(j) + \underline{p}_o^{x'}(j+1)\ \underline{b}_o(j)];$$

$$\underline{p}_o^x(N) = \underline{W}(N)[\underline{x}_o(N) - \underline{\rho}(N)] \tag{4.20}$$

and $\underline{\Sigma}_o(j+1|j)$, $\underline{\Sigma}_o(j+1|j+1)$, the predicted and updated error covariances of the augmented state satisfy the forward equations: $(j=k+1,\cdots,N)$

$$\underline{V}_o(j+1) = \underline{\Sigma}_o(j+1|j)\ \underline{h}_{o,\underline{z}}'(j+1)[\underline{h}_{o,\underline{z}}(j+1)\ \underline{\Sigma}_o(j+1|j)\ \underline{h}_{o,z}'(j+1) + \underline{R}(j+1)]^{-1}; \quad j=k,\ldots,N-1 \tag{4.21}$$

$$\underline{\Sigma}_o(j+1|j+1) = [I - \underline{V}_o(j+1)\underline{h}_{o,\underline{z}}(j+1)]\underline{\Sigma}_o(j+1|j); \quad j=k,\ldots,N-1 \tag{4.22}$$

$$\underline{\Sigma}_o(j+1|j) = \underline{f}_{o,\underline{z}}(j)\underline{\Sigma}_o(j|j)\underline{f}_{o,\underline{z}}'(j) + \underline{\mathcal{Q}}(j) \quad j=k+1,\ldots,N-1 \tag{4.23}$$

where

$$\underline{\mathcal{Q}}(j) = \begin{bmatrix} \underline{Q}(j) & 0 \\ 0 & \underline{G}(j) \end{bmatrix} . \tag{4.24}$$

The initial condition in (4.22) is $\underline{\Sigma}_o(k+1|k) = \underline{\Sigma}(k+1|k)$, the extrapolation covariance obtained after applying $u(k)$ to the system.

4.4 Nominal Selection and the Computation of the One-Step Dual Cost

In this section, the appropriate selection of the nominal control sequence associated with each predicted state will be discussed along with the detailed computation of the one-step dual cost.

One reasonable choice of the nominal control sequence $u_o[j; \underline{z}(k+1|k)]_{j=k+1}^{N-1}$ would be the certainty equivalence control; i.e., this sequence is obtained by solving the problem of minimizing

$$J_o(k+1) = \frac{1}{2}[\underline{x}_o(N) - \underline{\rho}(N)]' \underline{W}(N)[\underline{x}_o(N) - \underline{\rho}(N)]$$

$$+ \frac{1}{2}\sum_{j=k+1}^{N-1} \{[\underline{x}_o(j) - \underline{\rho}(j)]' \underline{W}(j)[\underline{x}_o(j) - \underline{\rho}(j)] + \lambda(j)[\underline{u}_o(j)]^2\} \quad (4.25)$$

subject to the constraints:

$$\underline{x}_o(j+1) = \underline{A}[j;\underline{\theta}_o(j)]\, \underline{x}_o(j) + \underline{b}[j; \underline{\theta}_o(j)]\, u_o(j); \; \underline{x}_o(k+1) = \hat{\underline{x}}(k+1|k) \quad (4.26)$$

$$\underline{\theta}_o(j+1) = \underline{D}(j)\, \underline{\theta}_o(j) \; ; \; \underline{\theta}_o(k+1) = \hat{\underline{\theta}}(k+1|k) \quad (4.27)$$

Note that $\underline{\theta}_o(j)$, $j=k+1,\ldots,N$ can be computed independently of how the control $u_o(j)$ is selected. The solution for this optimization problem can be obtained easily.[M2] The optimal control $u_o^*(j)$ is given by

$$u_o^*(j) = -\tilde{\mu}_o(j)\, \underline{b}_o'(j)\, [\tilde{\underline{K}}_o(j+1)\, \underline{A}_o(j)\, \underline{x}_o(j) + \tilde{\underline{p}}_o(j+1)] \quad (4.28)$$

where

$$\tilde{\mu}_o(j) \triangleq [\lambda(j) + \underline{b}_o'(j)\, \tilde{\underline{K}}_o(j+1)\, \underline{b}_o(j)]^{-1} \quad (4.29)$$

and $\tilde{\underline{K}}_o(j+1), \tilde{\underline{p}}_o(j+1)$ satisfy†

$$\tilde{\underline{K}}_o(j) = \underline{A}_o'(j)[I - \tilde{\mu}_o(j)\, \tilde{\underline{K}}_o(j+1)\, \underline{b}_o(j)\, \underline{b}_o'(j)]\, \tilde{\underline{K}}_o(j+1)\underline{A}_o(j)$$

$$+ \underline{W}(j) \; ; \quad \tilde{\underline{K}}_o(N) = \underline{W}(N) \quad (4.30)$$

---

† The squiggle here denotes quantities related to the certainty equivalence control which determines the nominal trajectory from k+1 to N.

$$\tilde{\underline{p}}_o(j) = \underline{A}_o'(j)[I - \tilde{\mu}_o(j)\, \tilde{\underline{K}}_o(j+1)\, \underline{b}_o(j)\, \underline{b}_o'(j)]\tilde{\underline{p}}_o(j+1)$$

$$- \underline{W}(j)\, \underline{\rho}(j) \quad ; \quad \tilde{\underline{p}}_o(N) = - \underline{W}(N)\, \underline{\rho}(n) \quad . \tag{4.31}$$

The corresponding minimum cost is

$$J_o(k+1) = \frac{1}{2}\, \hat{\underline{x}}'(k+1|k)\, \tilde{\underline{K}}_o(k+1)\, \hat{\underline{x}}(k+1|k) + \tilde{\underline{p}}_o'(k+1)\, \hat{\underline{x}}(k+1|k) + \tilde{g}_o(k+1) \tag{4.32}$$

where $\tilde{g}_o(j)$ satisfies

$$\tilde{g}_o(j) = \tilde{g}_o(j+1) - \frac{1}{2}\, \tilde{\mu}_o(j)\, \tilde{\underline{p}}_o'(j+1)\, \underline{b}_o(j)\, \underline{b}'(j)\, \tilde{\underline{p}}_o(j+1)$$

$$+ \frac{1}{2}\, \underline{\rho}'(j)\, \underline{W}(j)\, \underline{\rho}(j) \quad ; \quad \tilde{g}_o(N) = \frac{1}{2}\, \underline{\rho}'(N)\, \underline{W}(N)\, \underline{\rho}(N) \quad . \tag{4.33}$$

By comparing (4.30) with (4.15), we see that

$$\underline{K}_o^{xx}(j) = \tilde{\underline{K}}_o(j) \qquad j = k+1,\ldots,N \tag{4.34}$$

and hence from (4.18) and (4.29)

$$\mu_o(j) = \tilde{\mu}_o(j) \quad . \tag{4.35}$$

It is shown in Appendix D that

$$\underline{p}_o^x(j) = \tilde{\underline{K}}_o(j)\, \underline{x}_o(j) + \tilde{\underline{p}}_o(j) \quad . \tag{4.36}$$

From (4.35) and (4.28), we have

$$\lambda(j)\, u_o^*(j) + \underline{p}_o^{x'}(j+1)\, \underline{b}_o(j) = \lambda(j)\, u_o^*(j) + \underline{x}_o'(j)\, \underline{A}_o'(j)\, \underline{K}_o(j+1)\, \underline{b}_o(j)$$

$$+ u_o^*(j)\, \underline{b}_o'(j)\, \underline{K}_o(j+1)\, \underline{b}_o(j) + \tilde{\underline{p}}_o'(j+1)\, \underline{b}_o(j) = \underline{0} \quad . \tag{4.37}$$

Therefore (4.19) becomes

$$g_o(j) = 0 \ ; \quad j=k+1,\ldots,N \quad . \tag{4.38}$$

If we do not discretize the predicted state, the one-step cost can be computed by the following procedures:

1. Obtain $\hat{\underline{x}}(k+1|k)$, $\hat{\underline{\theta}}(k+1|k)$, and $\underline{\Sigma}(k+1|k)$ by

$$\hat{\underline{x}}(k+1|k) = \underline{A}[k;\hat{\underline{\theta}}(k|k)]\,\hat{\underline{x}}(k|k) + \underline{b}[k;\hat{\underline{\theta}}(k|k)]\,u(k) + \frac{1}{2}\sum_{i=1}^{n}\underline{e}_i\,\mathrm{tr}\{f^i_{\underline{x}\,\underline{x}}\,[\hat{\underline{x}}(k|k),u(k)]\,\underline{\Sigma}^{xx}(k|k)\} \tag{4.39}$$

$$\hat{\underline{\theta}}(k+1|k) = \underline{D}(k)\,\hat{\underline{\theta}}(k|k) \tag{4.40}$$

$$\underline{\Sigma}_o(k+1|k) = \underline{f}_{\underline{z}}(k)\,\underline{\Sigma}(k|k)\,\underline{f}'_{\underline{z}}(k) + \underline{\mathcal{Q}}(k) + \frac{1}{2}\sum_{i=1}^{n+s}\sum_{j=1}^{n+s}\underline{e}_i\,\underline{e}'_j\,\mathrm{tr}[f^i_{\underline{z}\,\underline{z}}(k)\,\underline{\Sigma}(k|k)\,f^j_{\underline{z}\,\underline{z}}(k)\,\underline{\Sigma}(k|k)]\,. \tag{4.41}$$

2. Generate $\underline{\theta}_o(j)$, $j \geq k+1$ via the equation (4.27).

3. Compute $\tilde{\underline{K}}_o(j)$, $\tilde{\underline{P}}_o(j)$, $j=k+1,\ldots,N$ using (4.30) and (4.31). Note that these equations are a function of $\hat{\underline{\theta}}(k|k)$ only, and are independent of $u(k)$.

4. Generate $\underline{x}_o(j)$, $j=k+1,\ldots,N$ using (4.26) (with $u_o(j) = u_o^*(j)$) and (4.28).

5. Compute $\underline{K}_o^{\theta x}(j)$, $\underline{K}_o^{\theta\theta}(j)$, $j=k+1,\ldots,N$ by (4.16) and (4.17). These are backward equations.

6. Form the matrix $\underline{K}_o(j)$, $k+1,\ldots N$, using (4.14) and (4.34).

7. Compute $\underline{\Sigma}_o(j+1|j+1)$, $j=k,\ldots,N-1$, $\underline{\Sigma}_o(j+1|j)$, $j=k+1,\ldots,N-1$, using (4.21)-(4.24).

8. Obtain the one-step dual cost by

$$J_d[u(k)] = \frac{1}{2}\lambda(k)\ u^2(k) + \frac{1}{2}\hat{\underline{x}}'(k+1|k)\ \tilde{\underline{K}}_o(k+1)\ \hat{\underline{x}}(k+1|k)$$

$$+ \tilde{\underline{p}}_o'(k+1)\ \hat{\underline{x}}(k+1|k)$$

$$+ \frac{1}{2}\mathrm{tr}\left\{ \sum_{j=k+1}^{N-1} \underline{W}(j)\ \underline{\Sigma}_o^{xx}(j|j) \right.$$

$$+ [\underline{\Sigma}(k+1|k) - \underline{\Sigma}_o(k+1|k+1)]\,\underline{K}_o(k+1)$$

$$\left. + \sum_{j=k+1}^{N-1} [\underline{\Sigma}_o(j+1|j) - \underline{\Sigma}_o(j+1|j+1)]\ \underline{K}_o(j+1) \right\} \quad . \qquad (4.42)$$

4.5 Remarks

1. The minimization of (4.42) is done by performing a search for $u^*(k)$. Since u is scalar, we can use the quadratic fit optimization method described in Appendix B.

2. The dual property of the control is revealed in (4.42) where the one-step cost to be minimized includes both control and estimation cost.

3. Let us partition the error covariance $\underline{\Sigma}_o$

$$\underline{\Sigma}_o = \begin{bmatrix} \underline{\Sigma}_o^{xx} & \Sigma_o^{x\theta} \\ \underline{\Sigma}_o^{\theta x} & \underline{\Sigma}_o^{\theta\theta} \end{bmatrix} \quad . \qquad (4.43)$$

Then if $\underline{\Sigma}_o^{\theta\theta}(k+1|k+1) = \underline{0}$, we must also have $\underline{\Sigma}_o^{x\theta}(k+1|k+1) = \underline{0}$ and for large k the one-step dual cost becomes

$$J_d[u(k)] = \frac{1}{2} h(k)\ u^2(k) + \frac{1}{2} \underline{\hat{x}}'(k+1|k)\ \underline{\tilde{K}}_o(k+1)\ \underline{\hat{x}}(k+1|k)$$

$$+ \underline{\tilde{p}}_o'(k+1)\ \underline{\hat{x}}(k+1|k) + \frac{1}{2}\,\mathrm{tr}\left\{ \sum_{j=k+1}^{N} \underline{W}(j)\ \underline{\Sigma}_o^{xx}(j|j) \right.$$

$$+ \underline{K}_o^{xx}(k+1)[\underline{\Sigma}^{xx}(k+1|k) - \underline{\Sigma}_o^{xx}(k+1|k+1)]$$

$$\left. + \sum_{j=k+1}^{N-1} \underline{K}_o^{xx}(j+1)[\underline{\Sigma}_o^{xx}(j+1|j) - \underline{\Sigma}_o^{xx}(j+1|j+1)] \right\} \quad . \quad (4.44)$$

Since if $\underline{\Sigma}_o^{\theta\theta}(k+1|k+1) = \underline{0}$,

$$f^{i}_{\underline{z}\,\underline{z}}[\underline{\hat{x}}(k|k),\ u(k)]\ \underline{\Sigma}(k|k) = \underline{0} \qquad (4.45)$$

we have from (4.39)

$$\underline{\hat{x}}(k+1|k) = \underline{A}[k,\ \underline{\hat{\theta}}(k|k)]\ \underline{\hat{x}}(k|k) + \underline{b}[k,\ \underline{\hat{\theta}}(k|k)]\ u(k) \quad . \quad (4.46)$$

Also, one can easily show from (4.21)-(4.24) that $\underline{\Sigma}_o^{xx}(j+1|j+1)$, $j=k,\ldots,N-1$, satisfy the minimum error equation of the linear system with known parameters $\underline{\theta}(j) = \underline{\hat{\theta}}(j|k)$. These imply that if we have high confidence on the parameter estimate, we can assume separation to hold.

4. The one-step dual cost reflects also the effect of the future observation program. For example, if it is known a priori that during the interval $\ell<j\le N$, $\ell\ge k$, no observations will be made, then we would have $\underline{\Sigma}_o(j|j-1) = \underline{\Sigma}_o(j|j)$. In this case $J_d[u(k)]$ becomes

$$J_d[u(k)] = \frac{1}{2}\lambda(k)\ u^2(k) + \frac{1}{2}\hat{\underline{x}}(k+1|k)\ \tilde{\underline{K}}_o(k+1)\ \hat{\underline{x}}(k+1|k)$$

$$+ \tilde{\underline{p}}_o'(k+1)\ \hat{\underline{x}}(k+1|k) + \frac{1}{2}\ \mathrm{tr}\Big\{ \sum_{j=k+1}^{\ell-1} \underline{W}(j)\ \underline{\Sigma}_o(j+1|j+1)$$

$$+ \sum_{j=\ell}^{N} \underline{W}(j)\ \underline{\Sigma}_o(j|\ell)$$

$$+ \underline{K}_o(k+1)[\underline{\Sigma}(k+1|k) - \underline{\Sigma}_o(k+1|k+1)]$$

$$+ \sum_{j=k+2}^{\ell-1} \underline{K}_o(j+1)[\underline{\Sigma}_o(j+1|j) - \underline{\Sigma}_o(j+1|j+1)]\Big\} \qquad (4.47)$$

Therefore, the knowledge that future observations will or will not be taken would change the present control strategy. If future learning will not take place, the present control tries to minimize the average control performance, whereas if future observation will take place, the present control will invest some of its energy to help the future learning. It is in this way that the dual control regulates its future learning under some control objective. Because of this "active learning" characteristic we call this control strategy an <u>actively adaptive control</u>.

5. The estimation cost of (4.42) is also a function of time-to-go. In the beginning of the control interval, the estimation cost is relatively high. The one-step optimal dual control must therefore be selected so that it compromises between control and estimation purposes. When k is approaching $N-1$, the estimation cost becomes smaller, and thus the one-step optimal dual control will give less weight to the estimation part and will finally concentrate on the control purpose.

6. For the case where the control $\underline{u}(k)$ is a vector rather than a scalar value, one can obtain exactly the same equations as above except that now $\tilde{\mu}_o(j)$, $\mu_o(j)$ are matrices and care is required in their placement in Equations (4.30)-(4.33) and (4.15)-(4.20). In the vector control case, the search for the one-step optimal is more complicated since we are searching over a volume rather than a line. Conceptually this does not create any new difficulty.

V. SIMULATION STUDIES

In this section, three example problems pertaining to dual control are considered, the purposes of which are:

(1) To investigate the computational feasibility of the one-step optimal dual control algorithm.

(2) To compare this algorithm with another widely used suboptimal algorithm -- the certainty equivalence.

(3) To understand the dual nature of the proposed algorithm; in particular, to understand the learning purpose of the control.

The first -- the scalar case example -- will be a simple one, so that we may understand the implications more clearly. The other two -- on interception and soft landing -- will be more complicated and will give additional insight into the dual control and some indication as to the computation feasibility of the proposed algorithm.

5.1 Scalar Case Example

Consider a scalar linear system

$$x(k+1) = ax(k) + bu(k) + \xi(k)$$

$$y(k) = x(k) + \eta(k) \tag{5.1}$$

where a,b are unknown constants and w(k), v(k) are independent zero-mean white noises with covariances q and r, respectively. The problem is to find a control sequence $\{u^*(k)\}_{k=0}^{N-1}$ such that the performance

$$J = E\left\{\frac{c}{2}[x(N) - \rho]^2 + \frac{1}{2}\sum_{k=0}^{N-1} u^2(k)\right\} \quad ; \qquad c > 0 \tag{5.2}$$

is minimized subject to the constraint (5.1) and

$$u^*(k) = u^*(Y^k, U^{k-1}, k) \quad . \tag{5.3}$$

A comparison of the certainty equivalence (C.E.) control strategy, and the actively adaptive dual control strategy as described in Section IV will be illustrated.

Two cases are considered. For both cases, $N = 20$, $\rho = 5$, $C = 100$, $x(0) = 0$, $a = 0.8$, $b = 0.5$, $q = 0.25$, $r = 0.04$; the initial guesses are $\hat{x}(0|0) = 0.13$, $\hat{a}(0|0) = 1.2$, $\hat{b}(0|0) = 0.3$ with initial error covariance

$$\mathrm{Cov}\left\{\begin{bmatrix} x(0) \\ a \\ b \end{bmatrix}\right\} = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.04 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \quad . \tag{5.4}$$

In case 1, the observations are available for all $k = 1, 2, \cdots, 19$; in case 2, the observations are available at $k = 1, \cdots, 14$; for $k \geq 15$, no observation is available. It is impossible to see how close the dual control strategy performance is to that of the truly optimum control strategy, since the truly optimum control strategy is very difficult to obtain. To give an idea about the performance level of the dual control strategy, we shall include the results for the optimal control when the parameters are all known. The performance for this will serve as a lower bound. It must be kept in mind that this lower bound is <u>not</u> achievable even by the truly optimal stochastic control for our problem. Ten Monte Carlo runs were performed for both cases, the results of which are shown in Tables 5.1 and 5.2. The first column shows the results for the optimum control when the parameters are <u>known</u>.

From Table 5.1, we see that, on the average, the dual control is better than is the C.E. control. An important fact here is that the dual control performance has a relatively small deviation from its average performance compared with that of the C.E. control. This property indicates that the dual control is more reliable than is C.E. control under stochastic effects.

TABLE 5.1

COMPARISON OF DUAL CONTROL WITH C.E.
CONTROL FOR THE SCALAR EXAMPLE (CASE 1)

	AVERAGE MISS DISTANCE SQUARED	AVERAGE TERMINAL ERROR SQUARED IN a	AVERAGE TERMINAL ERROR SQUARED IN b	AVERAGE PERFORMANCE	RANGE OF PERFORMANCE	STANDARD DEVIATION OF PERFORMANCE
OPTIMUM	0.0653	0	0	20.7	15.71-36.34	6
C.E.	0.311	0.126	0.233	34.7	22.11-70.43	17
DUAL	0.219	0.125	0.228	32.0	22.04-48.40	10

TABLE 5.2

COMPARISON OF DUAL CONTROL WITH C.E.
CONTROL FOR THE SCALAR EXAMPLE (CASE 2)

	AVERAGE MISS DISTANCE SQUARED	AVERAGE TERMINAL ERROR SQUARED IN a	AVERAGE TERMINAL ERROR SQUARED IN b	AVERAGE PERFORMANCE	RANGE OF PERFORMANCE	STANDARD DEVIATION OF PERFORMANCE
OPTIMUM	0.097	0	0	22.3	17.66-43.26	7
C.E.	7.353	1.609	.650	308.6	39.31-1882.5	561
DUAL	.143	.282	.258	74.5	66.6-113.36	13

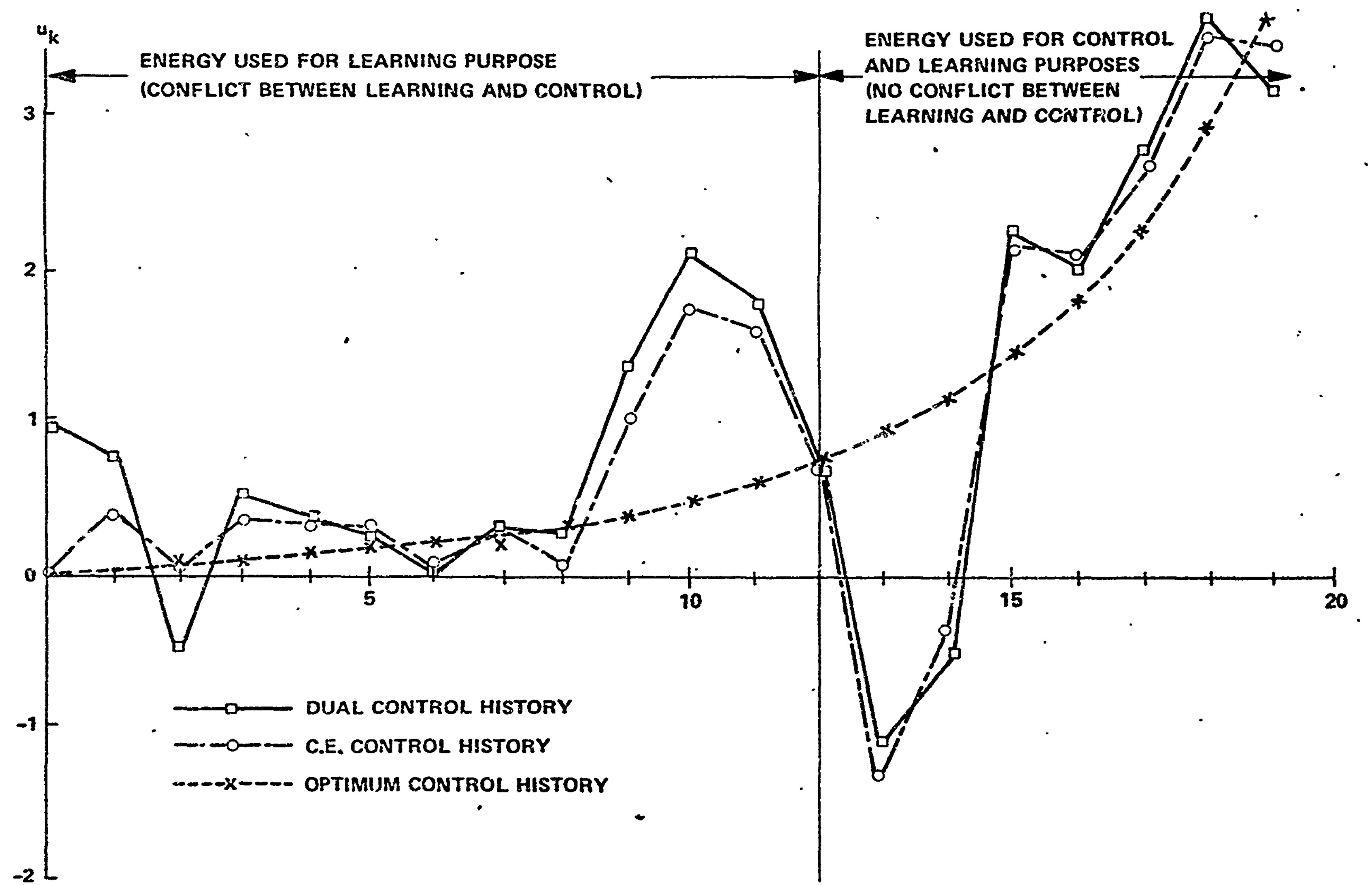

FIGURE 5.1 CONTROL HISTORY; ONE SAMPLE RUN FOR CASE 1

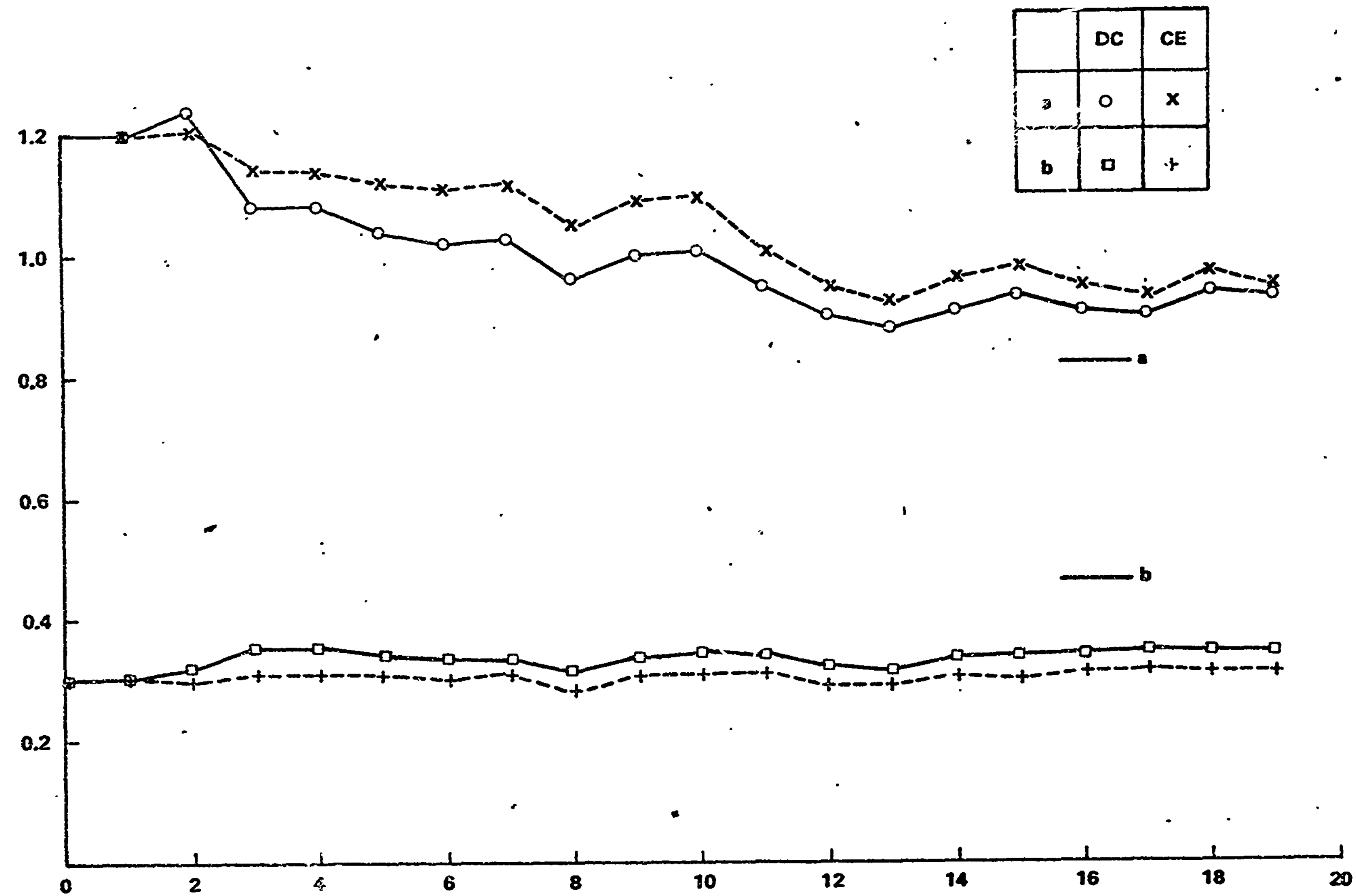

FIGURE 5.2 LEARNING HISTORY FOR DUAL CONTROL AND C.E. CONTROL (ONE SAMPLE RUN FOR CASE 1)

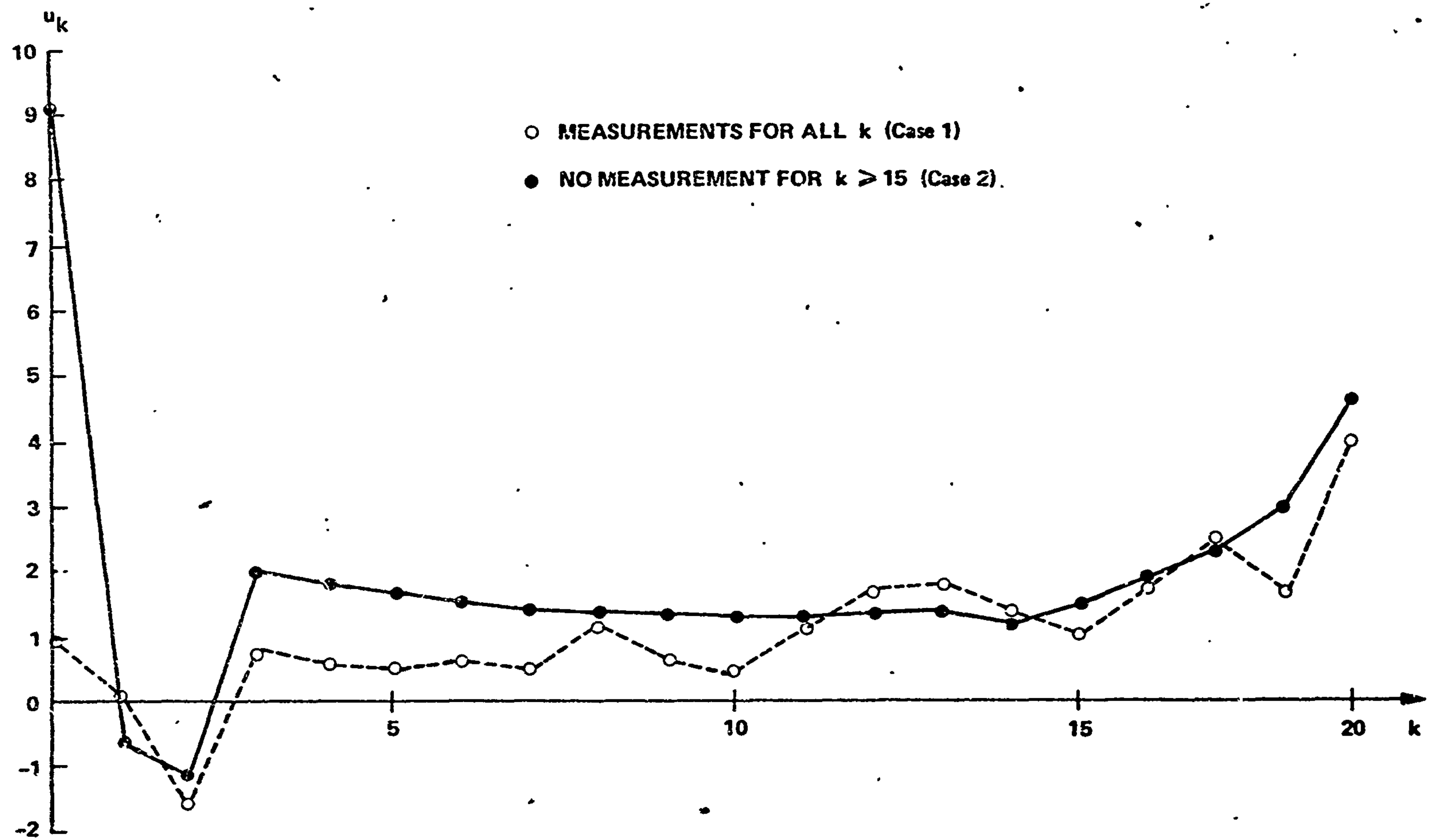

FIGURE 5.3 DUAL CONTROL HISTORY FOR CASE 1 AND CASE 2 (ONE SAMPLE RUN)

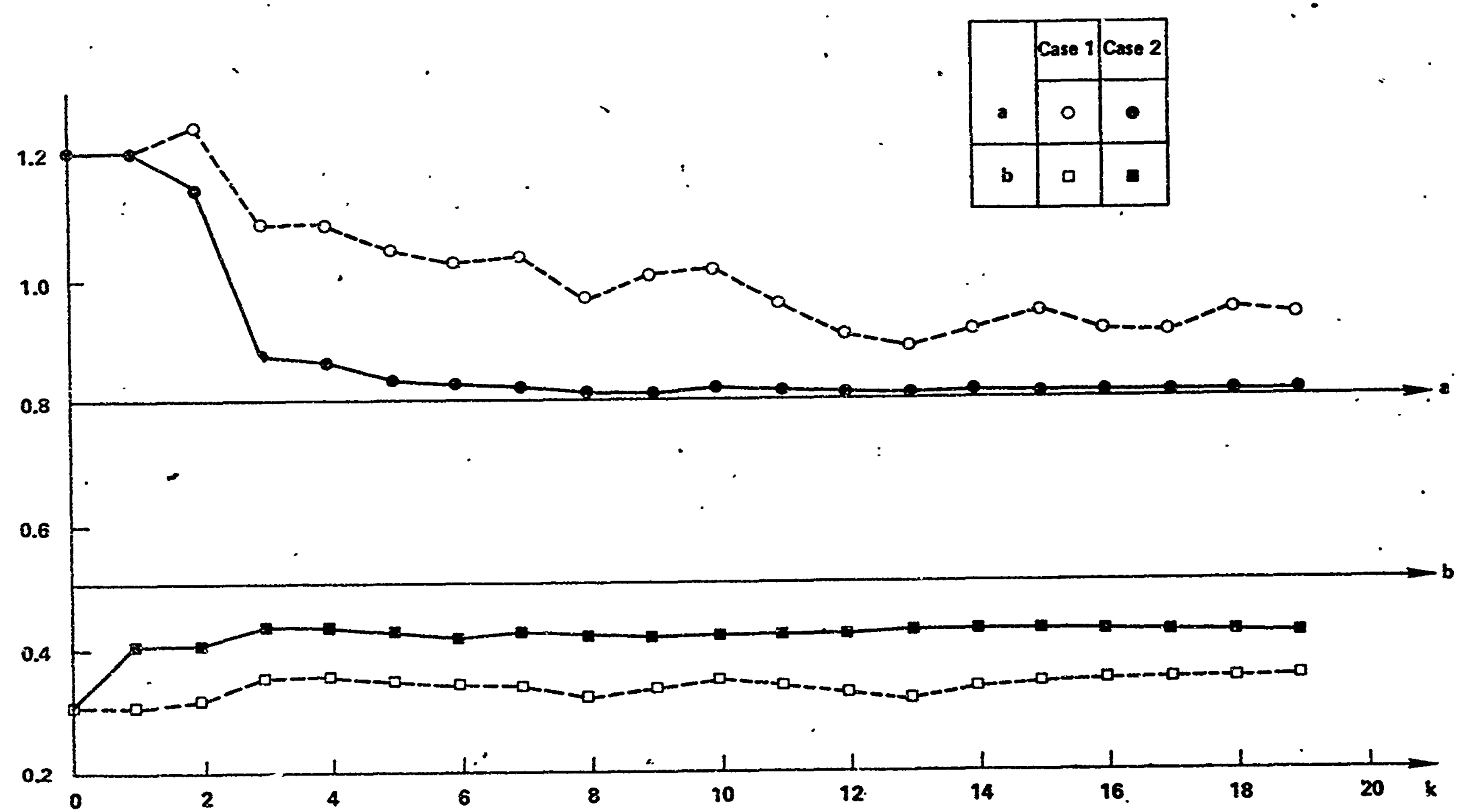

FIGURE 5.4 EVOLUTION OF THE PARAMETER ESTIMATES OBTAINED BY APPLYING THE ONE-STEP DUAL CONTROL (CASE 1 AND CASE 2; ONE SAMPLE RUN)

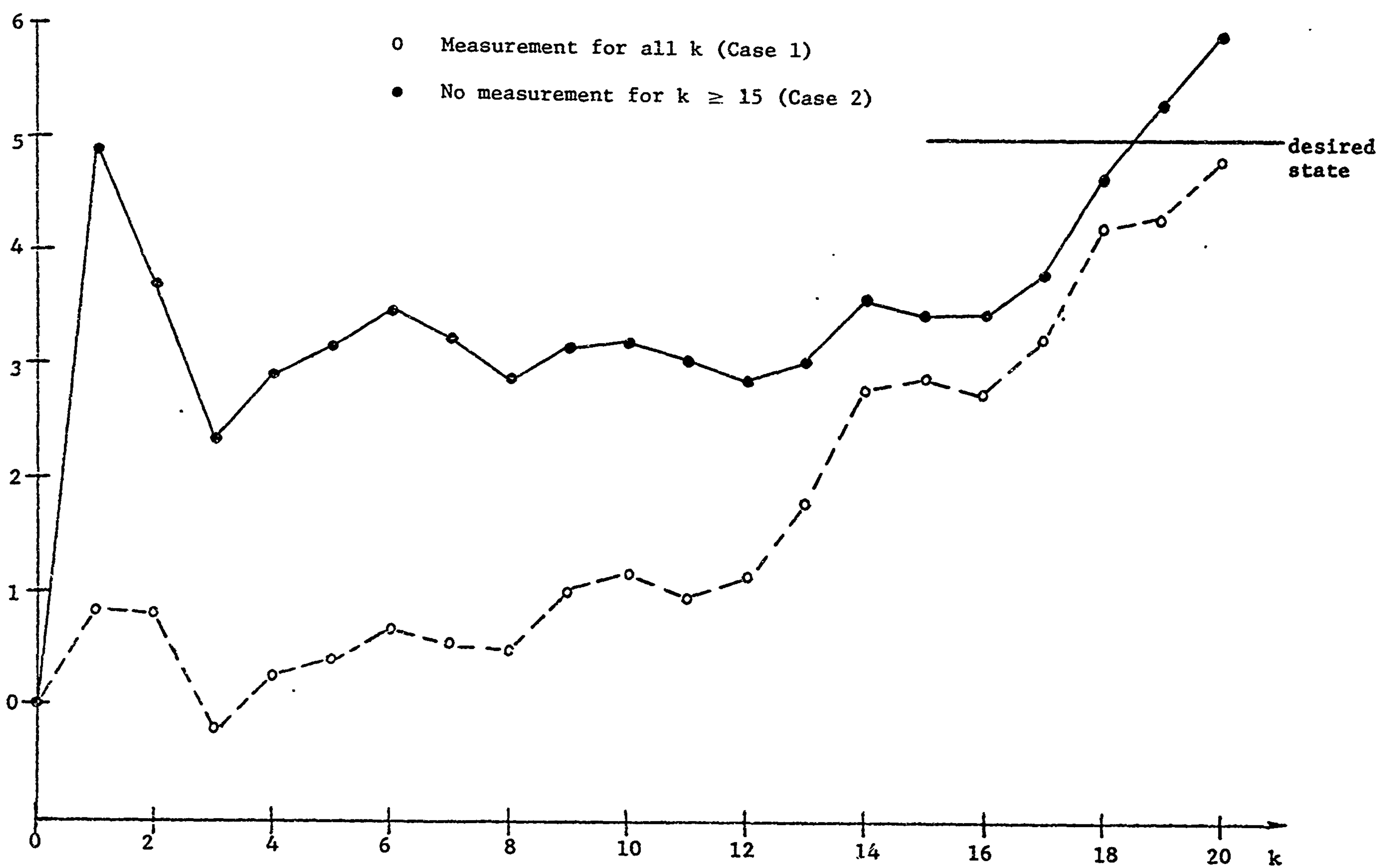

FIGURE 5.5 EVOLUTION OF THE STATE OBTAINED BY APPLYING ONE-STEP DUAL CONTROL

From the estimation performance, both the dual control and C.E. control perform well in terms of estimation at the terminal time, as shown in Figs. 5.1 and 5.2. Thus the fact that dual control performs better than C.E. control must be related to how fast learning is being performed. In Fig. 5.1, the control histories are plotted for one particular sample run. In this sample run, the noise sequences are the same for the optimal with known parameters, C.E., and dual controls. If the parameters are known exactly, learning is obviously not required and thus the control action will have only the control objective. Notice that to achieve this objective, the control energy should be kept small in the beginning and become larger toward the terminal time. In general, the C.E. control has this characteristic (the overshoots at about 10 and 13 are due to stochastic effects). However, the dual control acts quite differently, namely, at the initial time, the control value is quite far from zero. Thus, the dual control allocates some energy which is not directly intended for the control objective in the beginning.

In Fig. 5.2, the evolution of the parameters estimated is plotted for one sample run. As we notice in the figure, this energy is utilized for the learning purpose, which indicates that in the initial period, achieving the control objective and learning are in conflict. For $k \geq 12$, the control energy is building up in order to achieve the control objective. Since large control energy will excite the modes and improve the signal to noise ratio, it will promote learning. Thus for $k \geq 12$, learning and controlling are not in conflict. This explains why the C.E. control does have good estimates at the terminal time. In this case, learning is "accidental." To illustrate this point, we will see what happens if learning is not possible in the final period. This is shown in case 2. Table 5.2 shows the results of artificially terminating the final learning period. The C.E. control does very poorly in estimation, and, consequently, very poorly in achieving the control objective. This is reflected by the very large average cost and its standard deviation. The dual control, on the other hand, still performs reasonably well due to anticipation of the open-loop period at the end.

In Fig. 5.3, the control histories for another sample run with identical noise sequence for dual control case 1 and case 2 are plotted. Note that at the initial time, more energy is allocated to learning in case 2 than in case 1. This is so because in case 2, the derivation of the dual control takes into account that no learning is possible for $k \geq 15$, and thus any large control at the end will not help in learning; therefore, in order to achieve good control performance, a large amount of energy must be invested for pure learning purposes during the initial period to excite the system and to improve the signal-to-noise ratio. This is illustrated by Figs. 5.3, 5.4 and 5.5. As a result, the dual control achieves a much lower average cost and, at the same time, a much more reliable control strategy than does the C.E. control.

This active learning characteristic is a distinguishing feature of the dual control strategy, which depends not only on past observation information but also on the future observation program; therefore, the control value will differ depending on whether or not future observations will be made. Note that such a feature is not possessed by any of the existing suboptimal schemes suggested in the literature.

In the scalar example, a second order filter is used for on-line estimation of state and parameters. This estimation scheme is quite effective. Clearly, one may expect better performance if one uses a more sophisticated estimation algorithm; e.g., via parallel filters[B3],[T3],[A3].

One important point to be stressed is that the dual control strategy tries to improve the performance by considering what should be done <u>before</u> as well as <u>after</u> the parameters are identified, whereas the C.E. control strategy only tells what should be done after the parameters are identified.

5.2 Interception Example

In this subsection, the interception problem will be investigated. Consider a third order system

$$\underline{x}(k+1) = \underline{A}(\theta_1, \theta_2, \theta_3)\underline{x}(k) + \underline{B}(\theta_4, \theta_5, \theta_6)u(k) + \underline{\xi}(k) \quad (5.5)$$

$$y(k) = [0 \quad 0 \quad 1]\underline{x}(k) + \eta(k)$$

where

$$\underline{A}(\theta_1,\theta_2,\theta_3) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \theta_1 & \theta_2 & \theta_3 \end{bmatrix} \quad ; \qquad \underline{B}(\theta_4,\theta_5,\theta_6) = \begin{bmatrix} \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix} \tag{5.6}$$

and $\{\theta_i\}_{i=1}^{6}$ are unknown constant parameters with normal a priori statistics having mean and variance

$$\hat{\underline{\theta}}(o|o) = [1., -.6, .3, .1, .7, 1.5]' \quad ,$$

$$\underline{\Sigma}^{\theta\theta}(o|o) = \text{diag}(.1, .1, .01, .01, .01, .1) \quad .$$

The true parameters are

$$\underline{\theta} = [1.8 \;\; -1.01, .58, .3, .5, 1.]' \quad . \tag{5.7}$$

The initial state is assumed to be known:

$$\hat{\underline{x}}(0|0) = \underline{x}(0) = \underline{0} \quad . \tag{5.8}$$

The objective is to bring the third component of the state to a desired value. This is expressed by the cost

$$J = \frac{1}{2} E\{[x_3(N) - \rho]^2 + \sum_{i=0}^{N-1} \lambda u^2(i)\} \tag{5.9}$$

where ρ is some value and λ is chosen to be small. In our example ρ=20 and λ is chosen to be 10^{-3}. The noises $\{\xi_i(k)\}_{i=1}^{3}$ and $\eta(k+1)$ are assumed to be independent and are normally distributed with zero mean and unit variance. If we interpret x_3 as the position of an object, then this example corresponds to an interception problem: the guidance of an object to reach a certain point, without constraints on the velocity and acceleration of the object when it reaches that point. The difficulty lies in the fact that the poles and zeroes of the system are both unknown. The initial condition (5.7) represents the fact that the system is initially at rest.

Twenty Monte Carlo runs were performed on the interception example and average performances are summarized in Table 5.3 and Figs. 5.6-5.8. The performance for the optimal control when all the parameters are known is included to serve as a lower bound for the truly optimum performance for this problem. Again, it should be emphasized that this lower bound is unachievable even by the optimum stochastic controller for the system with unknown parameters.

As shown in Table 5.3, the dual control performance is an order of magnitude better than the C.E. control. The second and third rows indicate that the dual control performance is highly predictable, compared with the C.E. control. Note that the dual control uses only about twice the energy of the C.E. control, at the same time achieving a dramatic improvement in the miss distance squared over the C.E. control. This indicates that the dual control does use control energy at approrpiate times to improve learning, and thus achieves a satisfactory control objective.

Note that in Fig. 5.8, the dual control invested at the beginning considerable energy in learning. The effect of this is revealed in Figs. 5.6 and 5.7, where the average error squared for the parameters' estimates are displayed. Note that the learning in θ_4, θ_5, and θ_6 is much faster than the learning in θ_1, θ_2, and θ_3. As discussed in Section 3.4, the learning of θ_4, θ_5, and θ_6 results from the fact that large control will improve the signal-to-noise ratio for these parameters and thus the control can help in learning them in one step; on the other hand, the learning of θ_1, θ_2, and θ_3 is accomplished by exciting the modes of the system, and thus learning would be delayed until the system is properly excited.

Note that the C.E. control provides fairly good learning in θ_1, θ_2, and θ_3, but practically no learning in θ_4, θ_5, and θ_6. Note also that the C.E. control builds up energy very quickly after the tenth step. As observed in Fig. 5.7, some learning is θ_4, θ_5, and θ_6 is performed for $k \geq 10$, but

TABLE 5.3

SUMMARY OF RESULTS FOR THE INTERCEPTION EXAMPLE

CONTROL POLICY	OPTIMAL CONTROL WITH KNOWN PARAMETERS	C.E. CONTROL WITH UNKNOWN PARAMETERS	DUAL CONTROL WITH UNKNOWN PARAMETERS
AVERAGE COST	6	114	14
MAXIMUM COST IN A SAMPLE OF TWENTY RUNS	20	458	53
STANDARD DEVIATION OF THE COST	6	140	16
EXPECTED MISS DISTANCE SQUARED	12	225	22
WEIGHTED CUMULATIVE CONTROL ENERGY PRIOR TO FINAL STAGE	.1	1.4	3.2

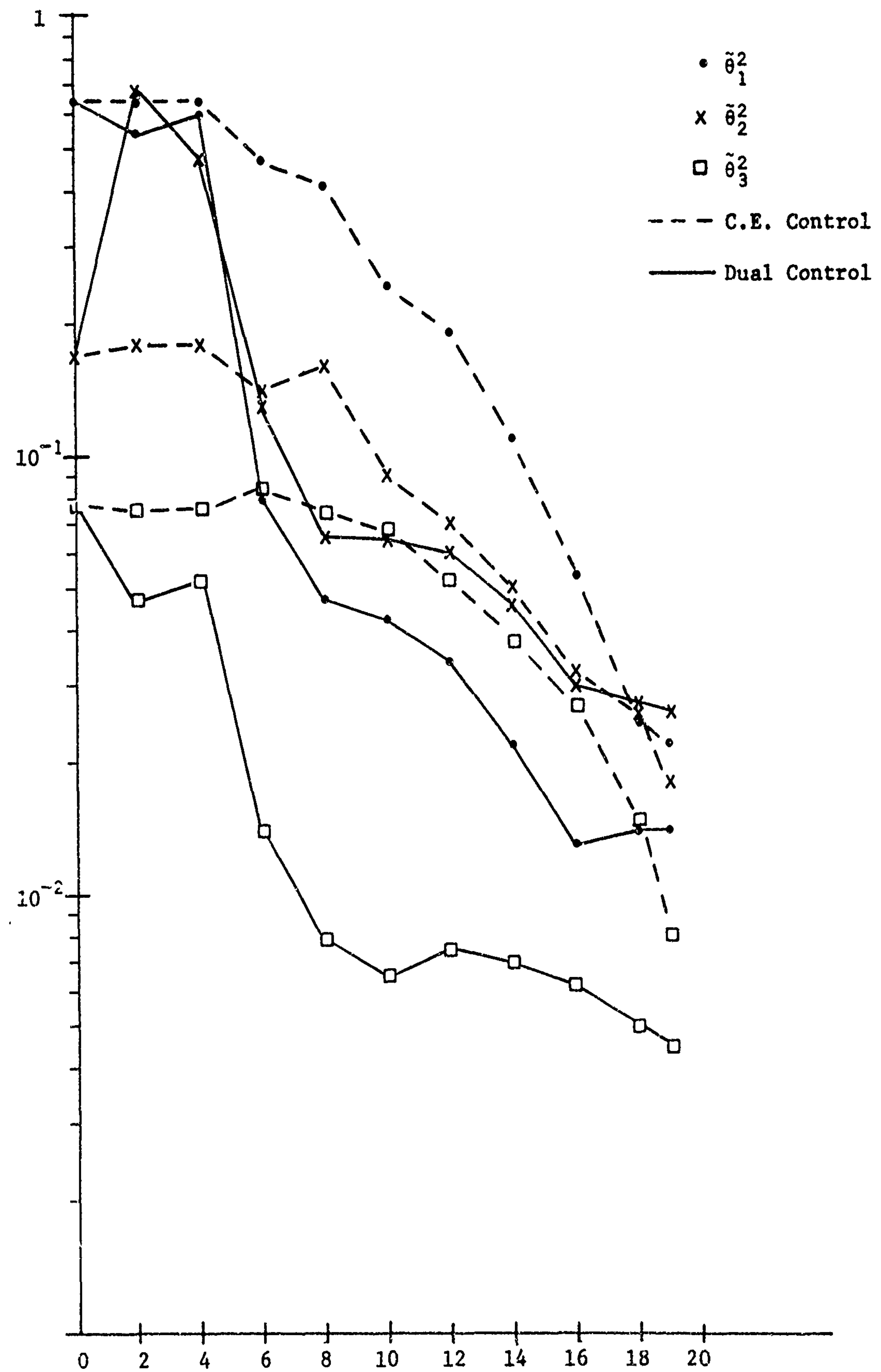

FIGURE 5.6 AVERAGE ESTIMATION ERROR SQUARED IN θ_1, θ_2, θ_3 FOR THE INTERCEPTION EXAMPLE

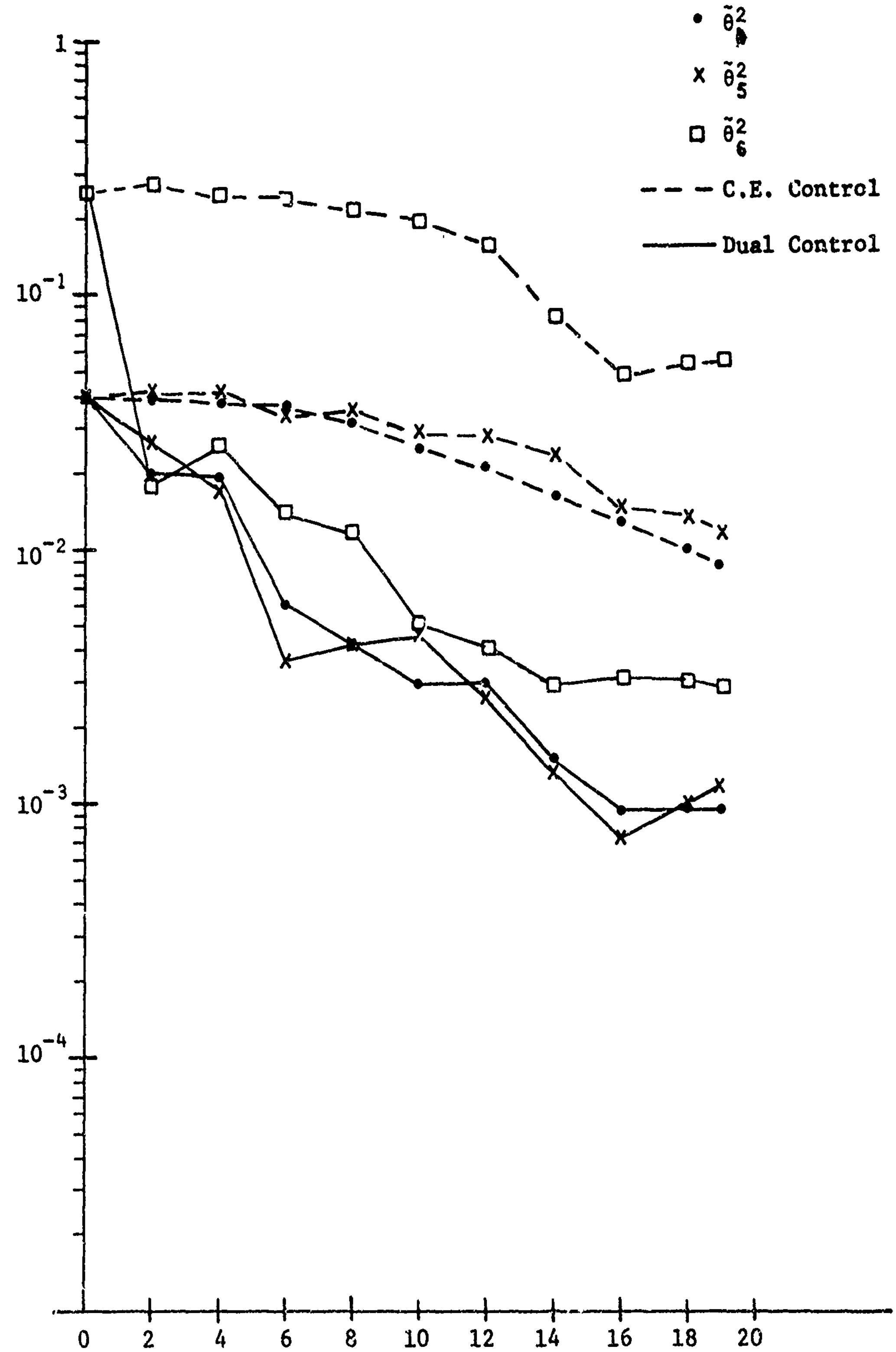

FIGURE 5.7 AVERAGE ESTIMATION ERROR SQUARED IN θ_4, θ_5, θ_6 FOR THE INTERCEPTION EXAMPLE

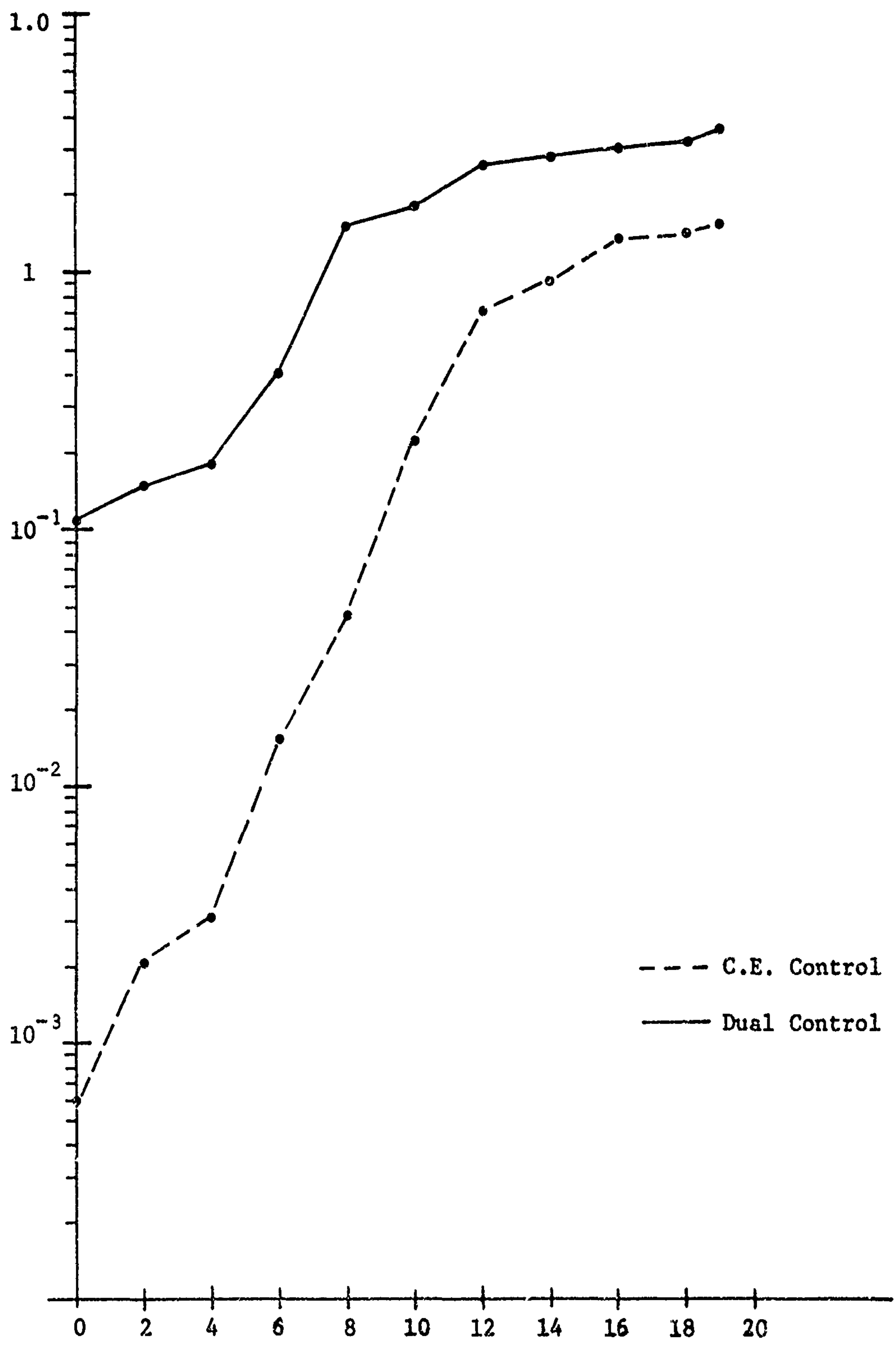

FIGURE 5.8 AVERAGE CUMULATIVE CONTROL ENERGY FOR THE INTERCEPTION EXAMPLE

prior to $k=10$, practically no control is applied and thus in θ_4, θ_5, and θ_6 no learning is done. The learning in θ_1, θ_2, and θ_3 before $k=10$ is due to the process noise, which serves as a random input that excites the modes of the system. Thus in this case, the learning is θ_1, θ_2, and θ_3 is quite accidental; also because this learning is too slow, it is of little use in achieving the control objective.

5.3 Soft Landing Example

Consider the same system with the same a priori conditions as discussed in Section 5.2. The only difference is that instead of bringing only the third component of the state to a desired value, the objective is to bring the final state to a certain point in the state space. This is expressed by

$$J = \frac{1}{2} E\{[\underline{x}(N) - \underline{\rho}]' \; [\underline{x}(N) - \underline{\rho}] + \sum_{i=0}^{N-1} \lambda \; u^2(i)\} \tag{5.10}$$

where $\underline{\rho}$ is a point in R^3 and λ is chosen to be small. This may be interpreted as a soft landing problem by selecting the $\underline{\rho}$ vector to be

$$\underline{\rho} = \begin{bmatrix} 0 \\ 0 \\ 20 \end{bmatrix} \tag{5.11}$$

and $\lambda = 10^{-3}$. Comparing the results of this problem to those obtained in Section 5.2 will provide more insight into the dual nature of the control. Twenty Monte Carlo runs were carried out for the C.E. control, the dual control, and the optimal control with known parameters. Again, the last-mentioned serves as an unachievable lower bound to the optimum performance. The results are summarized in Table 5.4 and Figs. 5.9-5.11.

Conceptually, the soft landing is a "harder" problem than the one considered in Section 5.2. Here, we want to "hit" a point in the state space, while in Section 5.2 we wanted to "hit" a surface. Therefore, it should be expected that the average cost is higher than in the previous example.

TABLE 5.4

SUMMARY OF RESULTS FOR THE SOFT LANDING EXAMPLE

CONTROL POLICY	OPTIMAL CONTROL WITH KNOWN PARAMETERS	C.E. CONTROL WITH UNKNOWN PARAMETERS	DUAL CONTROL WITH UNKNOWN PARAMETERS
AVERAGE COST	15	104	28
MAXIMUM COST IN A SAMPLE OF TWENTY RUNS	35	445	62
STANDARD DEVIATION OF THE COST	9	114	11
EXPECTED MISS DISTANCE SQUARED	28	192	32
WEIGHTED CUMULATIVE CONTROL ENERGY PRIOR TO FINAL STAGE	1	7	12

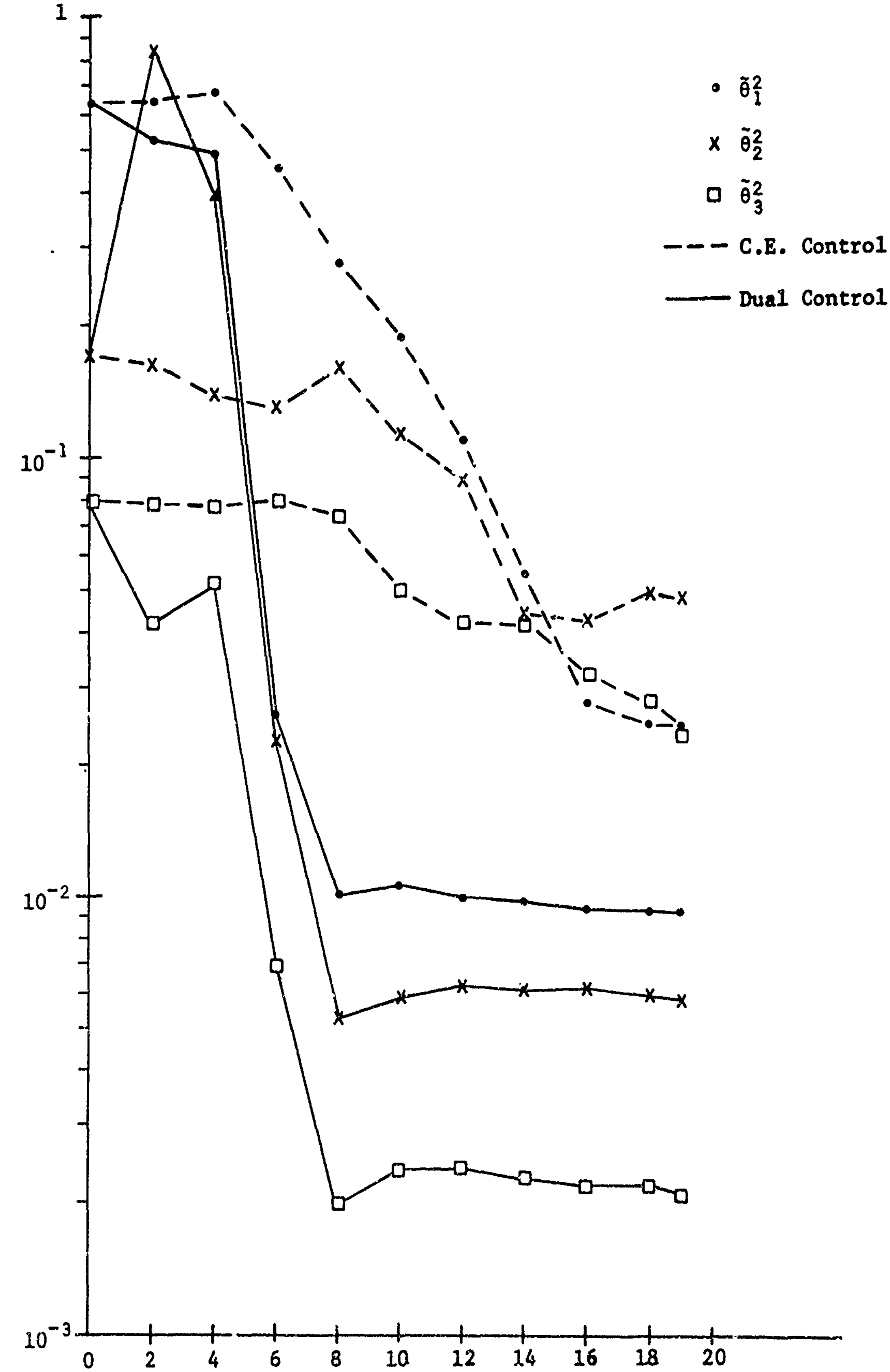

FIGURE 5.9 AVERAGE ESTIMATION ERROR SQUARED IN θ_1, θ_2, θ_3 FOR THE SOFT LANDING EXAMPLE

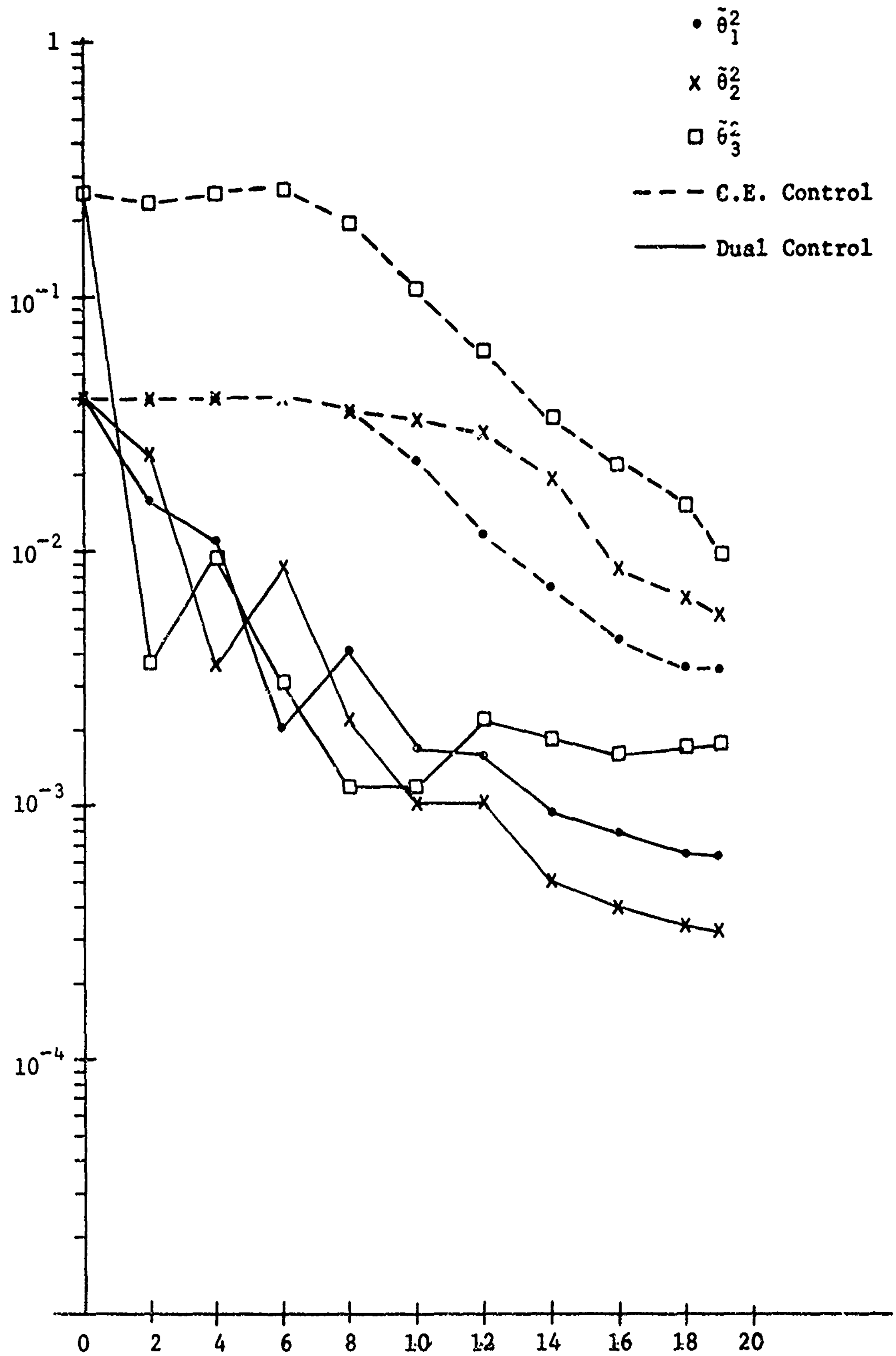

FIGURE 5.10 AVERAGE ESTIMATION ERROR SQUARED IN θ_4, θ_5, θ_6 FOR THE SOFT LANDING EXAMPLE

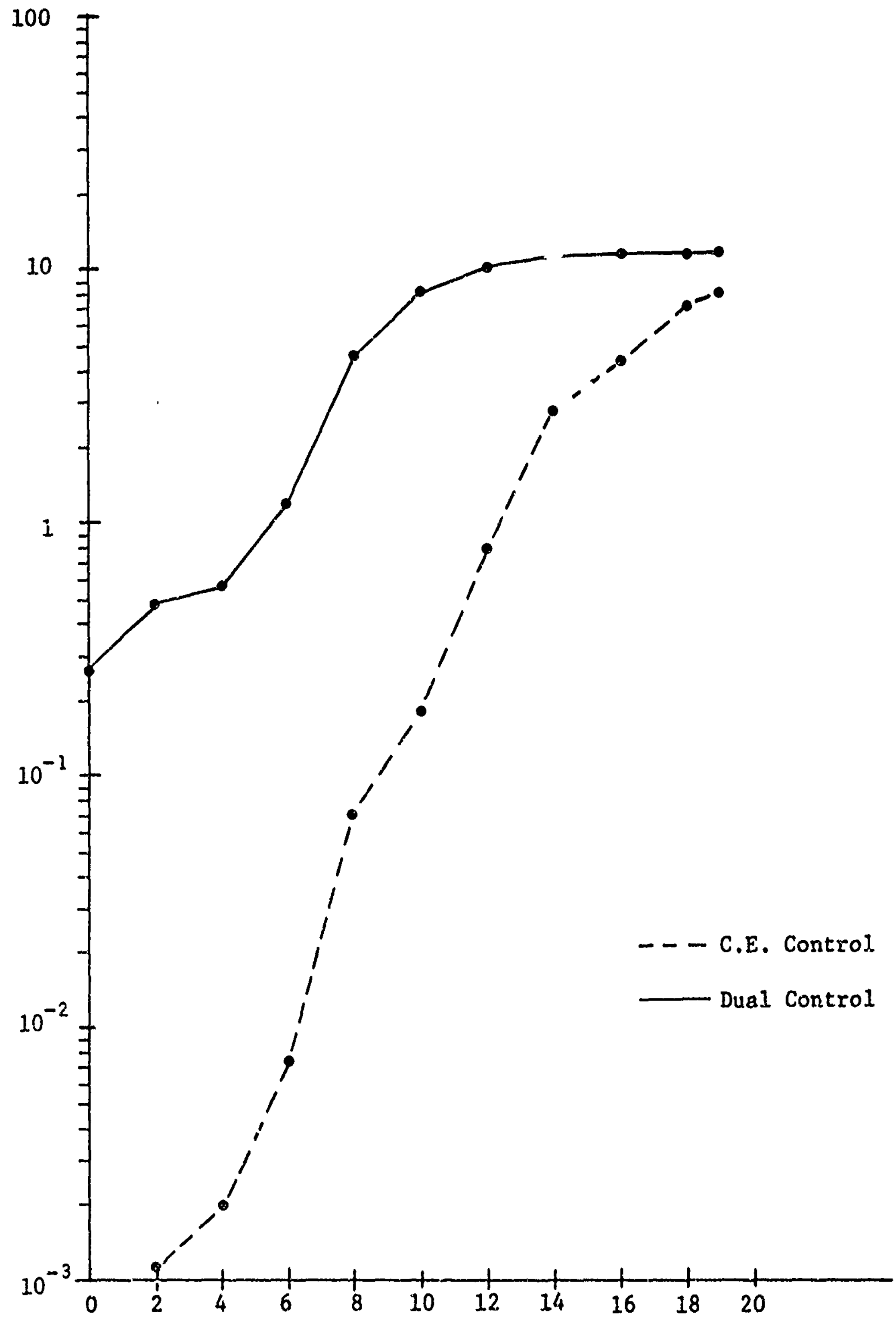

FIGURE 5.11 AVERAGE CUMULATIVE CONTROL ENERGY FOR THE SOFT LANDING EXAMPLE

This is seen to hold true, as shown in Tables 5.3 and 5.4, for the dual control and the optimal control with known parameters. However, for C.E. control, it does not hold true. This may look strange at the first sight, but careful analysis of the simulation results will offer an explanation for this.

In the following, the results of this example are examined in more detail, later, the comparisons of this example and that described in Section 5.2 are made.

Table 5.4 indicates the improvement of dual control over C.E. control, both in average performance and reliability. The terminal miss distance squared for the dual control is very close to the unachievable lower bound given by the optimal control with known parameters. To achieve this small miss distance, the dual control invests considerable energy for learning purposes. This can been seen in Fig. 5.11 where it is shown that a large amount of energy is invested at the initial time to promote future learning. As a result, the parameters are estimated very quickly (in about eight steps). After the parameters are adequately learned, the dual control smoothly hits the final point $\underline{\rho}$ (see Fig. 5.11). Again, note the delay in learning the parameters θ_1, θ_2, and θ_3.

The C.E. control, on the other hand, being only passive in learning, learns much slower, with the result that the terminal error is an order of magnitude higher than that of the dual control. As a consequence, the miss distance squared is substantially larger than that of the dual control. The C.E. control learning in θ_1, θ_2, and θ_3 is enhanced by the process noise, whereas the learning in θ_4, θ_5, and θ_6 is regulated by the control. In the C.E. case, this is very small in the initial period and builds up very quickly after time eight. Notice in Fig. 5.10 that the C.E. control did quite a bit of learning after time eight, but this learning is passive.

To understand the passive and active learning of the C.E. and dual control, the results of the soft landing example and the previous example will be compared. First compare the two C.E. controls. Note that the C.E.

control energy used in the soft landing example (we shall call this the second example) is much more than that used in the interception example (we shall call this the first example). Note from Figs. 5.9 and 5.11 that up to about k=12, the C.E. control uses about the same cumulative energy for the two examples. The fact that the final mission is different has not yet become important enough to change the control strategy. As a consequence, the learning for both cases is almost the same up to this time. In the first example, since the final destination is a surface, the controller can wait almost until the final time to apply a control to achieve the control objective, and therefore the C.E. control is still applying little energy after time twelve. The learning of the parameters θ_4, θ_5, and θ_6 is only slightly improved. However, for the second example, since the final destination is a point in the state space, the control must work "harder" to achieve its objective (transferring from one point to another arbitrary point requires three time units). Therefore, the control energy after time twelve increases very quickly for the second example. This results in a much better estimation on the gain parameters. Since the learning in the first example is poorer than in the second example for the C.E. control, a higher cost is accrued in the first example than in the second. Note that even though the second example is a "harder" problem, a better performance value is obtained. This is primarily because "accidental" learning is enhanced by the difficulty of achieving the final mission.

For dual control, quite a different control strategy at the beginning rather than at the end of the control interval can be noticed. The fact that a different end condition has to be fulfilled is propagated from the final time to the initial time. For the second example, the dual controller, realizing that the final mission is much more difficult to achieve, decides to invest more energy in the beginning, because learning is very important in this case to achieve a satisfactory final objective. Note the "speed" of learning in the second example compared with the first example (see Figs. 5.6, 5.7, 5.9, 5.10). The dual control regulates its energy in learning: in the first example where learning is less important, it does not insist in learning by applying large controls in the beginning; in the second example, the learning is much more important and thus more energy is utilized for the learning purpose. For both examples, the expected miss distances squared are comparable, thus, the increase in cost in the interception example is primarily due to the increase in accumulative input energy. This demonstrates the active learning characteristic of the dual control.

5.4 Remarks

(1) A comparison of the computation time required by the dual control with that for C.E. control gives some idea of the computation feasibility of the proposed algorithm. For the scalar example, the dual control requires, on the average, about twice as much time as the C.E. control per time unit. Note that in this example, we actually have a 3-dimensional problem. For the other two examples, it was found that the computation time for the dual control is on the average, approximately seven to eight times that of the C.E. control. Here, we actually have a 9-dimensional problem. However, judging from the improvement over the C.E. control, the extra computation time is worthwhile.

Note that the relative time between the dual control and the C.E. control increases as we have a higher dimensional problem. This is due to the fact that with higher dimension, the computation of the approximate optimal cost-to-go is relatively more time consuming. Thus for applications to classes of problems with high dimension, some improvement of the present algorithm is needed.

(2) The C.E. control is actually a very crude suboptimal method. More sophisticated algorithms have been suggested in the literature[L2]. One suggested approach is to have weighted C.E. control. This control is obtained having a bank of Kalman filters tuned at different parameters which adequately cover the parameter set, and an "optimal" (which is actually a C.E.) control generated for each Kalman filter in the bank, and finally all these controls are combined in a weighted manner. We must stress that this strategy does not possess active learning, and thus we would expect behavior similar to that in the C.E. control examples, probably with some improvement.

As seen from the above examples, active learning is the main characteristic which will yield a satisfactory performance, and therefore one can predict, with high confidence, that assuming the on-line estimation algorithm to be the same, such a suboptimal algorithm will be inferior to the dual control presented above. Moreover, if a serial computer is used, the computation time for the weighted C.E. approach would be equal to L times the C.E. approach computation time (both assume using the same estimation algorithm), where L is the number of Kalman filters. If there are six parameters, and each is quantized into only two levels, we have a total of 2^6 Kalman filters, and thus the computation time is about 65 times the C.E. approach computation time; this is much more time-consuming than the dual control approach.

The use of parallel computers may reduce the time for the weighted C.E. control approach, since this control law is parallel in structure. On the other hand, careful study of the present algorithm may show that it also possesses a parallel structure, though not in as obvious a manner.

(3) The active learning feature of this algorithm distinguishes it from the other approaches in the literature. The examples not only demonstrate that the dual control gives good performance, but more importantly it illustrates <u>why</u> it gives good performance.

(4) The present algorithm can be modified and refined so that it can eventually become feasible for real-time computation for a large class of problems. This is discussed further in the next section.

(5) The present algorithm can be used as a base for evaluating and comparing the performing of different ad hoc suboptimal algorithms. Even though the algorithm is still suboptimal, because of its active learning characteristic, it is felt that the algorithm is quite close to yielding the optimum performance.

VI. POTENTIAL APPLICATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

In this section, different classes of problems that are potential fields of application for the dual control theory developed in this study are indicated, and suggestions are made for areas of future research that are direct extensions of this work.

6.1 Applications

The dual control theory is applicable to general adaptive control problems where learning of the unknown environment and/or the state and parameters of the system under control is important in obtaining a good control strategy. The concept of active learning, which is introduced and developed in this study, is most important for these problems. Some of the problems that might benefit from the application of the dual control theory are listed below.

(1) Automatic Landing System -- The objective here is to bring a plane, approaching a land base, to land safely as quickly as possible. This requires knowledge of the position, velocity, and acceleration of the plane, as well as some unknown parameters (perhaps due to battle damage, imperfect preflight adjustment of the autopilot sensitivities, component degradation) in order to perform a safe landing.

(2) Interplanetary Missions -- Here, learning of the unknown environmental parameters is needed for controlling the vehicle.

(3) Low Altitude Missions -- In the final stage of a low altitude mission, an aircraft might want to fly higher to gain information; on the other hand it will be more exposed to enemy detection. In this situation, a tradeoff between gaining information and safety of aircraft exists and must be regulated.

(4) Homing Interception -- A homing interceptor equipped with a conformal array of an on-board radar is described in Fig. 6.1. From the figure, it can be seen that there are larger measurement errors for the head-on line-of sight. After collecting information about the position, velocity, and acceleration of a target from the on-board radar, the homing interceptor is to guide itself attempting to intercept the target. Thus starting toward a target with uncertain position and velocity, the interceptor must follow some trajectory that will perform active learning in order to increase the probability of successful intercept.

6.2 Future Research

The results obtained in the initial effort toward the "practical" stochastic control theory open up new areas for future research where the concept of actively adaptive control should play a central role. These new areas are outlined below.

(1) Improvement of Present Solution Procedure -- The method developed in this study is still not practical for some classes of problems where the dimensionality of the state and control vectors is large. Efforts should be spent in modifying and improving the present method so that it becomes tractable for a much larger class of problems. The approach would be to study carefully the present method and use it as a reference in obtaining simpler algorithms which retain the active learning feature.

(2) Free End-Time Problems -- Only fixed end-time problems have been considered in the present study. But in many practical situations, e.g., interception and soft landing, the final time is not prespecified but is chosen in some optimum manner. Therefore, after having gained understanding on the fixed end-time, the free end-time problem should be studied. The concepts and tools developed in the present study can be easily extended to become applicable to the free end-time problem.

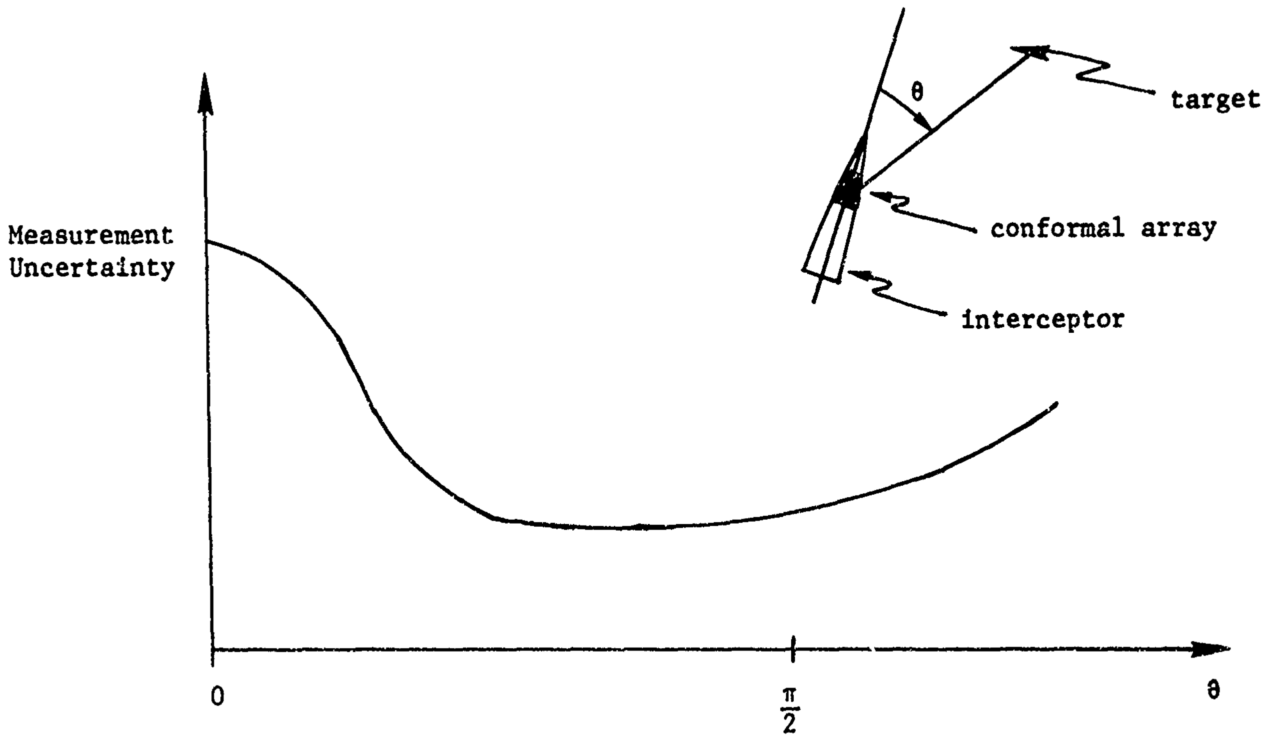

FIGURE 6.1 MEASUREMENT UNCERTAINTY MODEL

(3) Control and/or State Constraints Problems -- Throughout this study, no constraints on the control and the state were assumed. But, in actual applications, this assumption should be relaxed. Extension of present results to this class of problems is not straightforward but the concept of active learning will be helpful in both formulation and method of solution for this class of problems.

(4) Measurement Control Problems -- A large class of control problems, not directly covered by the classical theory of stochastic control, is the measurement control problem. This class of problems takes the general form of the block diagram shown in Fig. 6.2. The unique feature of the diagram is the measurement control which specifies how and when measurements are made. If the plant and the measurement systems are both linear and the cost is quadratic, Meier, et al.[M3], and Kramer[K1] showed that optimum measurement control affects only estimation and therefore, solved the problem via the separation principle. Such a result corresponds to the classical stochastic control of linear systems with quadratic criteria, where the optimal control has only a control purpose and the solution can be solved by the separation theorem. In the general nonlinear situation, both the measurement control and the plant control have the dual properties of trying to improve estimation and control. For this reason, the problem is called the dual measurement control.

Examples of dual measurement control problems arise in many Air Force problem applications. Three important cases occur when there are constraints on the total number of measurements allowed, when there are constraints on the types of measurements made, or when there are costs associated with making measurements. In the first situation, which occurs when there are only finite resources available to make measurements and each measurement uses up a given amount of resource, an optimal scheduling of measurements in real time is sought. The second situation is illustrated by a

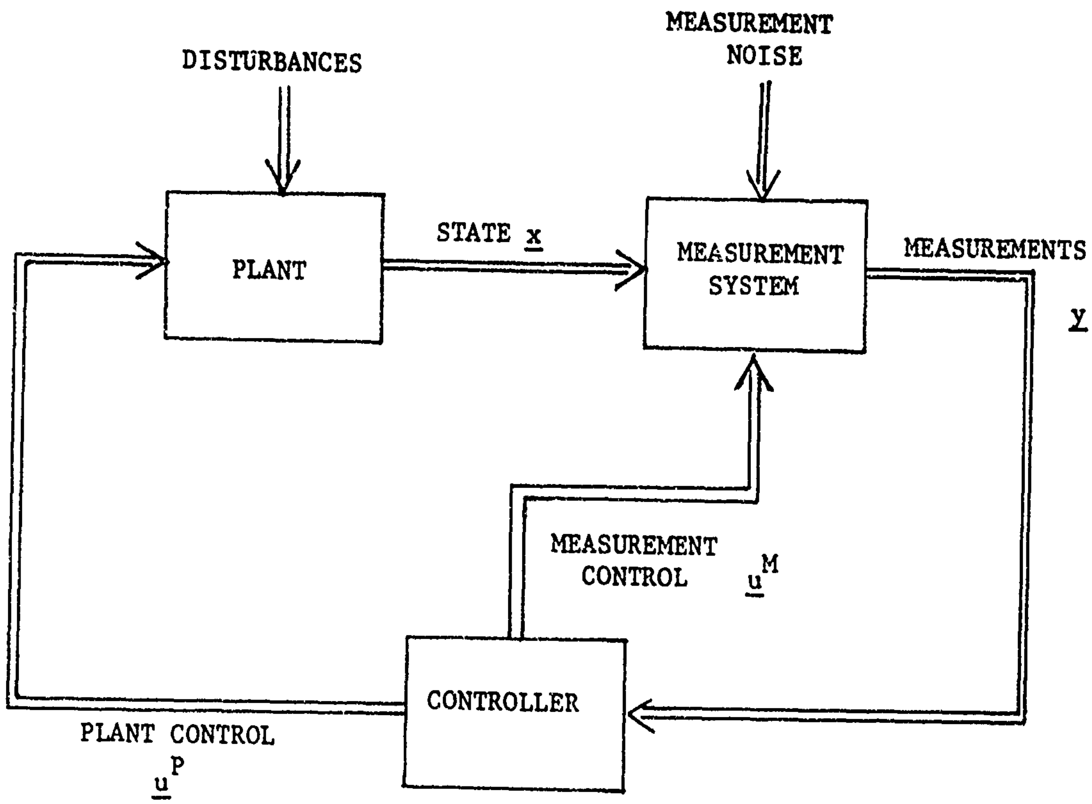

FIGURE 6.2 MEASUREMENT CONTROL PROBLEMS

radar with limited peak and average power. Within these constraints the quality of position (range) and velocity (doppler) measurements can be traded off by varying the radar pulse shape. In the third situation, a good example of measurement cost is when the use of a radar will gain information about an enemy but will also give the enemy information about the radar location. In this case the information given to the enemy may be represented as a cost of making the measurement that must be traded off with the benefits of making those measurements.

For this class of dual measurement problems, the concept of active learning is very important. The understanding gained in the present study will provide a fundamental framework for future study.

(5) Dual Control and Input Design for Identification--This study was concerned with the controlling of a system where learning of parameters is only an indirect objective. Part II of this contract,[M5] is concerned with learning unknown parameters where controlling is only an indirect objective. Therefore, these two separate problems are actually two faces of the same problem. It would be of interest to investigate the interrelation of these two problems.

VII. PUBLICATIONS UNDER THIS CONTRACT

The following publications are results from Part I of this contract.

(1) "Dual Control of Stochastic Nonlinear Systems," by E. Tse, L. Meier, and Y. Bar-Shalom (1971 IEEE Decision and Control Conference, Miami Beach, Florida).

Abstract

In stochastic control of nonlinear systems, estimation and control are dependent--the control, in addition to its effect on the state of the system, affects the estimation performance. A method for obtaining a dual control sequence is discussed that leads to a one-step optimization problem and a control strategy called the one-step dual control. An example problem is used to indicate the performance improvement when using the one-step dual control instead of the separation control policy.

(2) "On the Dual Control of Stochastic Discrete-Time Systems," by E. Tse, A. J. Tether, Y. Bar-Shalom, and L. Meier (Fifth International Hawaii Conference on Systems Science, Honolulu, January 1972).

Abstract

The dual nature of the control for stochastic nonlinear systems is stressed in formulating a stochastic control problem. Two methods for obtaining dual control sequence are discussed. The first method is the off-line optimal nominal selection, the second is called the one-step optimal dual control. An example is given which indicates that the one-step optimal dual control has great improvement over the control strategy obtained by imposing separation.

(3) "Wide-Sense Adaptive Dual Control of Stochastic Nonlinear Systems," by E. Tse, Y. Bar-Shalom and L. Meier (to appear in IEEE Trans. on Automatic Control).

Abstract

A new approach is presented for the problem of stochastic control of nonlinear systems. It is well known that, except for the Linear-Quadratic problem, the optimal stochastic controller cannot be obtained in practice. In general it is the curse of dimensionality which makes the strict application of the principle of optimality infeasible. The two subproblems of stochastic control, estimation and control property, are except for the Linear-Quadratic case intercoupled. As pointed out by Feldbaum, in addition to its effects on the state of the system, the control also affects the estimation performance. In this paper, the stochastic control problem is formulated such that this dual property of the control appears explicitly. The resulting control sequence exhibits the closed-loop property: it takes into account the past observations and also the future observation program. Thus in addition to being adaptive, this control also plans its future learning according to the control objective. Some preliminary simulation results illustrate these properties of the control.

(4) "An Actively Adaptive Control for Linear Systems with Random Parameters via the Dual Control Approach," by E. Tse and Y. Bar-Shalom (submitted to 1972 IEEE Decision and Control Conference; also to be reviewed for IEEE Transactions on Automatic Control)

Abstract

The problem of controlling a linear system with random parameters is being considered. An algorithm is obtained which seems to be appropriate in computational feasiblity for this class of problems. The algorithm possesses active learning characteristics in the sense that it regulates its adaptation (learning) in an optimum manner. Simulation studies are carried out in terms of two third-order examples. The example problems

provide additional insight into the active learning characteristic as compared to the passive learning possessed by certainty equivalence and many other suboptimal algorithms.

The following publications are supported partially by this contract.

(1) "Parallel Computation of the Conditional Mean State Estimate for Nonlinear Systems," by E. Tse (The Second Symposium on Nonlinear Estimation Theory, San Diego, 1971).

Abstract

This paper discusses an approach for approximating the conditional mean state estimate for nonlinear systems. The approach is motivated by realizing that some recent advances in computer organization, in particular parallel processing, could be used to reduce the computation time if the problem is appropriately formulated. It is shown how the estimation problem can be formulated properly so that this advantage can be utilized. Specific approximation methods are described in some detail.

(2) "Modal Trajectory Estimation and Parallel Computers," by R. E. Larson and E. Tse (The Second Symposium on Nonlinear Estimation Theory, San Diego, 1971).

Abstract

For nonlinear estimation, different estimation methods are appropriate depending on the estimation criterion being used; and different sufficient information statistics must be updated and stored in real time. For modal trajectory state estimation, i.e., estimation of the maximum likelihood trajectory in state space, the problem can be solved using the idea of dynamic programming; in this case the optimal return function serves as the sufficient statistic. Since there are a number of parallel operations that occur in the evaluation of the dynamic programming recursive formula, the the use of a parallel computer could greatly reduce the computer time and memory required for obtaining the modal trajectory

estimate. The purpose of this paper is to discuss the modal trajectory estimation method and how various algorithms for implementing dynamic programming in a parallel processor can be used to reduce the computational burden.

(3) "The Third Order Extended Kalman Filter," by L. Meier (The Second Symposium on Nonlinear Estimation Theory, San Diego, 1971)

Abstract

The Extended Kalman Filter accurate to the third order about a nominal is derived and compared to the extended Kalman filter accurate to second order. It is found that to be accurate to third order the covariance equation must be solved in real time; whereas for second order accuracy it may be solved a priori.

(4) "Parallel Computation of the Modal Trajectory Estimate," by R. E. Larson and E. Tse (Fifth International Hawaii Conference on Systems Science, Honolulu, January 1972).

Abstract

For modal trajectory state estimation, i.e., estimation of the maximum likelihood trajectory in state space, the problem can be solved using the idea of dynamic programming. The purpose of this paper is to discuss various algorithms for implementing the dynamic programming equation on a parallel computer. In particular, the following algorithms are examined: Parallel States Algorithm; Parallel Noises Algorithm, and Parallel States and Stages Algorithm.

(5) "Parallel Processing Algorithms for Modal Trajectory Estimation," by R. E. Larson and E. Tse (1972 JACC and to appear in IEEE Transactions on Automatic Control).

Abstract

For modal trajectory state estimation, i.e., estimation of the maximum likelihood trajectory in state space, the problem can be solved using the idea of dynamic programming. Since there are a number of parallel operations that occur in the evaluation of the dynamic programming recursive formula, the use of a parallel computer could greatly reduce the computer time and memory required for obtaining the modal trajectory estimate. The purpose of this paper is to discuss the modal trajectory estimation method and how various algorithms for implementing dynamic programming in a parallel processor can be used to reduce the computational burden. In particular, the following algorithms for implementing dynamic programming in parallel processors are examined: Parallel States Algorithm, Parallel Noises Algorithm, and Parallel States and Stages Algorithm.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Air Force Office of Scientific Research under AFOSR Project F44620-71-C-0077. We also wish to thank our Contract Monitor, Major A. M. Dayton, for many helpful conversations during the course of this project. Finally, we are grateful to our colleagues at Systems Control, Inc., for their criticism, comments, discussions and support; in particular, we wish to thank Dr. R. E. Larson who has greatly influenced our concept of stochastic dynamic programming.

Appendix A

THE OPTIMAL PERTURBATION CONTROL

Denote the optimal incremental cost-to-go by

$$\Delta J_o^*(Y^j,j) = \min_{\delta\underline{u}(j)} E\{ \min_{\delta\underline{u}(j+1)} E[\dots \min_{\delta\underline{u}(N-1)} E(\Delta J_o(Y^N,N)|Y^{N-1})|\dots|Y^{j+1}]Y^j\} \quad . \tag{A.1}$$

The alternating minimizations and expectations in the above reflect the closed-loop property of the control.[B2]. The principle of optimality leads to

$$\Delta J_o^*(Y^j,j) = \min_{\delta\underline{u}(j)} \Big\{ L'_{o,\underline{x}}(j)\delta\hat{\underline{x}}'(j|j) + \frac{1}{2}\delta\hat{\underline{x}}'(j|j)\, L_{o,\underline{x}\,\underline{x}}(j)\delta\hat{\underline{x}}(j|j)$$

$$+ \phi'_{o,\underline{u}}(j)\delta\underline{u}(j) + \frac{1}{2}\delta\underline{u}'(j)\,\phi_{o,\underline{u}\,\underline{u}}(j)\,\delta\underline{u}(j) + \frac{1}{2}\mathrm{tr}\,[L_{o,\underline{x}\,\underline{x}}(j)\underline{\Sigma}_o(j|j)]$$

$$+ E[\Delta J_o^*(Y^{j+1},j+1)|Y^j] \Big\} \quad . \tag{A.2}$$

The covariance $\underline{\Sigma}_o(j|j)$ is propagated, independently of the perturbation control, according to the extended Kalman filter equation. (See also Section 3.2.)

Take $\Delta J_o^*(Y^j,j)$ of the form

$$\Delta J_o^*(Y^j,j) = \hat{g}_o(j) + \underline{p}'_o(j)\delta\hat{\underline{x}}(j|j) + \frac{1}{2}\delta\hat{\underline{x}}'(j|j)\underline{K}_o(j)\delta\hat{\underline{x}}(j|j) \quad . \tag{A.3}$$

Substituting (A.3) into (A.2), the minimization of the right-hand side of (A.2) is obtained by letting

$$\frac{\partial}{\partial\underline{u}}[\phi'_{o,\underline{u}}(k) + \frac{1}{2}\delta\underline{u}'(k)\phi_{o,\underline{u}\,\underline{u}}(k)\delta\underline{u}(k) + \underline{p}'_o(k+1)\delta\hat{\underline{x}}(k+1|k)$$

$$+\frac{1}{2}\delta\hat{\underline{x}}'(k+1|k)\underline{K}_o(k+1)\delta\hat{\underline{x}}(k+1|k)] = \underline{0} \tag{A.4}$$

where to second order,

$$\delta\hat{\underline{x}}(j+1|j) = \underline{f}_{o,\underline{x}}(j)\delta\hat{\underline{x}}(j|j) + \underline{f}_{o,\underline{u}}(j)\delta\underline{u}(j) + \frac{1}{2}\sum_{i=1}^{n}\underline{e}_i\ \mathrm{tr}\{f^i_{o,\underline{x}\,\underline{x}}(j)[\underline{\Sigma}_o(j|j)$$

$$+ \delta\hat{\underline{x}}(j|j)\delta\hat{\underline{x}}'(j|j)]\} + \sum_{i=1}^{n}\underline{e}_i\ \delta\underline{u}'(j)\ f^i_{o,\underline{u}\,\underline{x}}(j)\delta\hat{\underline{x}}(j|j)$$

$$+ \frac{1}{2}\sum_{i=1}^{n}\underline{e}_i\ \delta\underline{u}'(j)\ f^i_{o,\underline{u}\,\underline{u}}(j)\ \delta\underline{u}(j) \quad . \tag{A.5}$$

Substitute (A.5) into (A.4) and retain terms only up to second order; the optimum $\delta\underline{u}_o^*(j)$ is given by

$$\delta\underline{u}_o^*(j) = -[\phi_{o,\underline{u}\,\underline{u}}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j) + \sum_{i=1}^{n}\underline{p}'_o(j+1)\underline{e}_i f^i_{o,\underline{u}\,\underline{u}}(j)]^{-1}$$

$$\{[\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) + \sum_{i=1}^{n}\underline{p}'_o(j+1)\underline{e}_i f^i_{o,\underline{u}\,\underline{x}}(j)]\delta\hat{\underline{x}}(j|j)$$

$$+ f'_{o,\underline{u}}(j)\underline{p}_o(j+1) + \phi_{o,\underline{u}}(j)\} \tag{A.6}$$

Substituting (A.6) and (A.3) into (A.2) (keeping only up to second order terms) and equating terms in zeroth, first and second order of $\delta\hat{\underline{x}}(k|k)$, one has, by using the definition of $H_o(j)$, (3.10), the equations for $\hat{g}_o(j)$, $\underline{p}_o(j)$, and $\underline{K}_o(j)$

$$\hat{g}_o(j) = g_o(j+1) - \frac{1}{2}H'_{o,\underline{u}}(j)\ [H_{o,\underline{u}\,\underline{u}}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1}$$

$$H_{o,\underline{u}}(j) + \frac{1}{2}\mathrm{tr}\,\{H_{o,\underline{x}\,\underline{x}}(j)\underline{\Sigma}_o(j|j) + [\underline{\Sigma}_o(j+1|j) - \underline{\Sigma}_o(j+1|j+1)]\underline{K}_o(j+1)\} \quad ;$$

$$\hat{g}_o(N) = \frac{1}{2}\mathrm{tr}\,[\psi_{o,\underline{x}\,\underline{x}}\underline{\Sigma}_o(N|N)] \tag{A.7}$$

$$\underline{p}_o(j) = H_{o,\underline{x}}(j) - [\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) + H_{o,\underline{u}\,\underline{x}}(j)]'$$

$$[H_{o,\underline{u}\,\underline{u}}(j) + f'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1} H_{o,\underline{u}}(j) \quad ; \quad \underline{p}_o(N) = \psi_{o,\underline{x}} \qquad \text{(A.8)}$$

$$\underline{K}_o(j) = \underline{f}'_{o,x}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) - [\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) + H_{o,\underline{u}\,\underline{x}}(j)]'$$

$$\cdot\ [H_{o,\underline{u}\,\underline{u}}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1}$$

$$\cdot\ [\underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{x}}(j) + H_{o,\underline{u}\,\underline{x}}(j)] + H_{o,\underline{x}\,\underline{x}}(j) \quad ; \quad \underline{K}_o(N) = \psi_{o,\underline{x}\,\underline{x}} \qquad \text{(A.9)}$$

The resulting optimum cost if $U_o(k,N-1)$ is selected is thus given by (note that $\delta\hat{\underline{x}}(k|k) = 0$)

$$J^*[k,U_o(k,N-1)] = J_o(k) + \Delta J_o^*(Y^k,k) = J_o(k) + \hat{g}_o(k) \quad . \qquad \text{(A.10)}$$

To stress the estimation performance reflected in $J^*[k,U_o(k,N-1)]$, define $g_o(j), j=k,k+1,\cdots,N$ according to

$$g_o(j) = g_o(j+1) - \frac{1}{2}H_{o,\underline{u}}(j)[H_{o,\underline{u}\,\underline{u}}(j) + \underline{f}'_{o,\underline{u}}(j)\underline{K}_o(j+1)\underline{f}_{o,\underline{u}}(j)]^{-1}\underline{H}_{o,\underline{u}}(j) \quad ;$$

$$g_o(N) = 0 \quad . \qquad \text{(A.11)}$$

Then by (A.7), (A.10), (A.11), $J^*[k,U_o(k,N-1)]$ can be expressed alternatively as

$$J^*[k,U_o(k,N-1)] = J_o(k) + g_o(k) + \frac{1}{2}\,\mathrm{tr}\{\psi_{o,\underline{x}\,\underline{x}}\underline{\Sigma}_o(N|N) + \sum_{j=k}^{N-1}\{H_{o,\underline{x}\,\underline{x}}(j)\underline{\Sigma}_o(j|j)$$

$$+ [\underline{\Sigma}_o(j+1|j) - \underline{\Sigma}_o(j+1|j+1)]K_o(j+1)\} \quad . \qquad \text{(A.12)}$$

In the one-step dual control consideration, for $j \geq k+1$, only perturbation analysis will be carried out along the νth nominal, thus the cost of applying $\underline{u}(k)$ can be approximated by

$$I_d[\underline{u}(k)] = E\{\phi[u(k),k] + L[x(k),k] + J_\nu(k+1) + g_\nu(k+1)$$

$$+ \underline{p}_\nu'(k+1)[\hat{\underline{x}}(k+1|k+1) - \underline{x}_\nu(k+1)]$$

$$+ \frac{1}{2}[\hat{\underline{x}}(k+1|k+1) - \underline{x}_\nu(k+1)]'\underline{K}_\nu(k+1)[\hat{\underline{x}}(k+1|k+1)$$

$$- \underline{x}_\nu(k+1)]|Y^k\} \quad . \tag{A.13}$$

Equation (3.28) can now be obtained by noting that

$$E\{\phi[\underline{u}(k),k] + J_\nu(k+1) + g_\nu(k+1)|Y^k\} = \phi[\underline{u}(k),k] + J_\nu(k+1) + g_\nu(k+1)$$

$$E\{\underline{p}_\nu'(k+1)[\hat{\underline{x}}(k+1|k+1) - \underline{x}_\nu(k+1)]|Y^k\} = \underline{p}_\nu'(k+1)[\hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1)]$$

$$E\{[\hat{\underline{x}}(k+1|k+1) - \underline{x}_\nu(k+1]'\underline{K}_\nu(k+1)[\hat{\underline{x}}(k+1|k+1) - \underline{x}_\nu(k+1)]|Y^k\}$$

$$= [\hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1)]\underline{K}_\nu(k+1)[\hat{\underline{x}}(k+1|k) - \underline{x}_\nu(k+1)]$$

$$+ \mathrm{tr}\{\underline{K}_\nu(k+1)[\underline{\Sigma}(k+1|k) - \underline{\Sigma}(k+1|k+1)]\} \quad . \tag{A.14}$$

Appendix B

THE QUADRATIC FIT OPTIMIZATION METHOD

The method to find the minimizing augument u^* of a convex function $f(u)$ with a quadratic fit is described below. At the kth iteration one has the function evaluated at three points $u_i^{(k)}$, $i=1,2,3$. The corresponding values are

$$f_i^{(k)} \triangleq f(u_i^{(k)}) \quad . \tag{B.1}$$

Assume these points are ordered such that

$$u_1^{(k)} < u_2^{(k)} < u_3^{(k)} \quad . \tag{B.2}$$

The convexity condition is

$$f_2 < \frac{(u_2-u_1)f_3 + (u_3-u_2)f_1}{u_3-u_1} \quad . \tag{B.3}$$

The quadratic fit has its minimum at

$$u_4^{(k)} = F[u_i^{(k)}, f_i^{(k)}, i = 1,2,3] \tag{B.4}$$

where

$$F[u_i, f_i, i=1,2,3] \triangleq \frac{1}{2} \frac{f_{12}\, a_{13} - f_{13}\, a_{12}}{f_{12}\, b_{13} - f_{13}\, b_{12}} \tag{B.5}$$

and

$$f_{ij} \triangleq f_i - f_j \tag{B.6}$$

$$a_{ij} \triangleq u_i^2 - u_j^2 \tag{B.7}$$

$$b_{ij} \triangleq u_i - u_j \ . \tag{B.8}$$

The three values $f_i^{(k)}$ will satisfy one and only one of the following sets of inequalities (this is a consequence of (B.3))

$$f_1 > f_2 \quad \text{and} \quad f_3 > f_2 \tag{B.9}$$

or

$$f_1 > f_2 > f_3 \tag{B.10}$$

or

$$f_1 < f_2 < f_3 \quad . \tag{B.11}$$

If (B.9) is satisfied then

$$u_1^{(k)} < u_4^{(k)} < u_3^{(k)} \quad . \tag{B.12}$$

The new point u_4 will in this case satisfy either

$$u_1^{(k)} < u_4^{(k)} < u_2^{(k)} \tag{B.13}$$

or

$$u_2^{(k)} < u_4^{(k)} < u_3^{(k)} \tag{B.14}$$

and the value of the function at the new point is

$$f_4 < f_2 \tag{B.15}$$

or

$$f_4 > f_2 \quad . \tag{B.16}$$

The procedure to choose the new set of three points is as follows, depending on which of Eq. (B.9) through (B.11) is satisfied.

I. Equation (B.9) is satisfied:
If {(B.13) and (B.15)} or {(B.14) and (B.16)} are satisfied, then the new set of three points is

$$U^{k+1} \triangleq \langle u_i^{(k+1)} \rangle_{i=1,2,3} = \text{ord}\{u_j^{(k)}\}_{j \neq 3} \tag{B.17}$$

where ord{·} stands for ordering in the sense of (B.2).

If {(B.13) and (B.16)} or {(B.14) and (B.15)} are satisfied, then

$$U^{k+1} = \text{ord}\{u_j^{(k)}\}_{j\neq 1} \quad . \tag{B.18}$$

II. Equation (B.10) is satisfied. The new set is given by (B.18).

III. Equation (B.11) is satisfied. The new set is given by (B.17).
The search will stop when

$$|u_4 - u_2| \le \varepsilon(u_4) \triangleq \max[c_1|u_4|, c_2] \quad . \tag{B.19}$$

The algorithm can be summarized as follows:

1. Given the first three points (ordered), evaluate the corresponding values of f.
2. a. Set $k = 0$

   b. Check for convexity (B.3). If convex, go to 3. Otherwise

   $$u_4 = u_2 + 2^k |u_3 - u_2| \,\text{sgn}\,(J_2 - J_3) \ .$$

   c. Set $k \leftarrow k+1$. Go to 2b.
3. Compute u_4 using (B.4).
4. If (B.9) is satisfied and also (B.19), set $u^* = u_4$ and exit. Otherwise evaluate f_4.
5. If (B.9) is satisfied use procedure I and then go to 2. Otherwise go to 6.
6. If (B.10) is satisfied use (B.18) and then go to 2. Otherwise go to 7.
7. Use (B.17) and then go to 2a.

Appendix C

THE APPROXIMATE OPTIMAL COST TO GO FOR LINEAR SYSTEM WITH RANDOM PARAMETERS

Define $H_o(j)$ by

$$H_o(j) \triangleq \frac{1}{2}\,[\underline{x}_o(j) - \underline{\rho}(j)]'\,\underline{W}(j)[\underline{x}_o(j) - \underline{\rho}(j)] + \frac{1}{2}\,\lambda(j)\,u^2(j)$$

$$+\ \underline{p}_o'(j+1)\ \underline{f}[j;\ \underline{x}_o(j),\ u_o(j)]\quad . \tag{C.1}$$

Partitioning $\underline{p}_o$ into two parts, $\underline{p}_o^x$ and $\underline{p}_o^\theta$, of dimensions n and s respectively, we have

$$H_o(j) \triangleq \frac{1}{2}\,[\underline{x}_o(j) - \underline{\rho}(j)]'\,\underline{W}(j)[\underline{x}_o(j) - \underline{\rho}(j)] + \frac{1}{2}\,\lambda(j)\,u^2(j)$$

$$+\ \underline{p}_o^{x'}(j+1)\ [\underline{A}_o(j)\ \underline{x}_o(j) + \underline{b}_o(j)\ u_o(j)]$$

$$+\ \underline{p}_o^{\theta'}(j+1)\ \underline{D}(j)\ \underline{\theta}_o(j)\quad . \tag{C.2}$$

Using the formulae in Section I, we have the following partial derivatives

$$H_{o,\underline{x}}(j) = \underline{W}(j)[\underline{x}_o(j) - \underline{\rho}(j)] + \underline{A}_o'(j)\ \underline{p}_o^x(j+1) \tag{C.3}$$

$$H_{o,\underline{\theta}}(j) = \sum_{i=1}^{n}\ [\underline{e}_i'\ \underline{p}_o^x(j+1)]\ [\underline{a}_{\underline{\theta}}^{i'}(j)\ \underline{x}_o^o(j) + b_{\underline{\theta}}^i(j)\quad u_o(j)]$$

$$+\ \underline{D}'(j)\ \underline{p}_o^\theta(j+1) \tag{C.4}$$

$$H_{o,u}(j) = \lambda(j)\ u_o(j) + \underline{p}_o^{x'}(j+1)\ \underline{b}_o(j) \tag{C.5}$$

$$H_{o,\underline{xx}}(j) = \underline{W}(j) \tag{C.6}$$

$$H_{o,\underline{x\theta}}(j) = \sum_{i=1}^{n}\ \underline{e}_i'\ \underline{p}_o^x(j+1)\ \underline{a}_{\underline{\theta}}^i(j) \tag{C.7}$$

$$H_{o,\underline{\theta x}}(j) = H_{o,\underline{x\theta}}'(j) \tag{C.8}$$

$$H_{o,\underline{\theta\theta}}(j) = \underline{0} \tag{C.9}$$

$$H_{o,u\underline{x}} = \underline{0} \tag{C.10}$$

$$H_{o,u\underline{\theta}} = \sum_{i=1}^{n} \underline{e}_i' \, \underline{p}_o^x(j+1) \, b_{\underline{\theta}}^i(j) \; . \tag{C.11}$$

Equations (4.11) - (4.20) are obtained by substituting (C.3) - (C.11) into (A.7) - (A.9). Note that $\underline{p}_o^{\theta}(j)$ does not appear in the computation of $\underline{K}_o(j)$, $\underline{p}_o^x(j)$ and $g_o(j)$, therefore its equation is not given in Section 4.3.

Appendix D

PROOF OF EQUATION (4.36)

It can be easily seen from the end conditions in (4.20) and (4.31) that (4.36) is satisfied for $j = N$. Now assuming that

$$\underline{p}_o^x(j+1) - \tilde{\underline{p}}_o(j+1) = \tilde{\underline{K}}_o(j+1)\underline{x}_o(j+1) \tag{D.1}$$

it will be shown that (4.36) holds. From (4.20) and (4.31), one has, making use of (4.34) and (4.35)

$$\underline{p}_o^x(j) - \tilde{\underline{p}}_o(j) = \underline{A}_o'(j)[\underline{p}_o^x(j+1) - \tilde{\underline{p}}_o(j+1)] + \underline{W}(j)\underline{x}_o(j)$$

$$- \mu_o(j)\underline{A}_o'(j)\tilde{\underline{K}}_o(j+1)\underline{b}_o(j)\{\lambda(j)u_o(j) + \underline{b}_o'(j)[\underline{p}_o^x(j+1) - \tilde{\underline{p}}_o(j+1)]\} \quad . \tag{D.2}$$

Inserting (D.1) into (D.2) and then using (4.26) yields

$$\underline{p}_o^x(j) - \tilde{\underline{p}}_o(j) = \underline{A}_o'(j)[I - \tilde{\mu}_o(j)\,\tilde{\underline{K}}_o(j+1)\underline{b}_o(j)\underline{b}_o'(j)]\tilde{\underline{K}}_o(j+1)\underline{x}_o(j+1)$$

$$+ \underline{W}(j)\underline{x}_o(j) - \tilde{\mu}_o(j)\underline{A}_o'(j)\tilde{\underline{K}}_o(j+1)\underline{b}_o(j)\lambda(j)u_o(j)$$

$$= \{\underline{A}_o'(j)[I - \tilde{\mu}_o(j)\tilde{\underline{K}}_o(j+1)\underline{b}_o(j)\underline{b}_o'(j)]\tilde{\underline{K}}_o(j+1)\underline{A}_o(j)+\underline{W}(j)\}\ \underline{x}_o(j)$$

$$+ \underline{A}_o'(j)[I - \tilde{\mu}_o(j)\tilde{\underline{K}}_o(j+1)\underline{b}_o(j)\underline{b}_o'(j) - \tilde{\mu}_o(j)\lambda(j)]\tilde{\underline{K}}_o(j+1)\underline{b}_o(j)u_o(j) \quad . \tag{D.3}$$

The last term above is equal to zero. This can be easily seen by rearranging it and using (4.29):

$$\underline{A}_o'(j)\tilde{\underline{K}}_o(j+1)\underline{b}_o(j)\ \{1 - \tilde{\mu}_o(j)[\underline{b}_o'(j)\tilde{\underline{K}}_o(j+1)\underline{b}_o(j) + \lambda(j)]\}\ u_o(j) = 0 \quad . \tag{D.4}$$

Now, using (D.4) and (4.30) in (D.3) one immediately obtains (4.36), the desired result. Notice that this result is independent of the control u_o.

REFERENCES

[A1] Aoki, A., Optimization of Stochastic Systems, Academic Press, 1967.

[A2] Athans, M., Wishner, R. P., and Bertolini, A., "Suboptimal State Estimation for Continuous Time Nonlinear Systems from Discrete Noisy Measurements," IEEE. Trans. on Automatic Control, Vol. AC-13, pp. 504-514, October 1968.

[A3] Alspach, D. L. and Sorenson, H. W., "Approximation of Density Function by a Sum of Gaussian for Nonlinear Baysian Estimation," Proceedings Symposium on Nonlinear Estimation Theory and Its Application, San Diego, California, Sept. 21-23, 1970.

[B1] Bellman, R., Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, New Jersey, 1961.

[B2] Bar-Shalom, Y. and Sivan, R., "On the Optimal Control of Discrete-Time Linear Systems with Random Parameters," IEEE Trans. on Automatic Control, Vol. AC-14, pp. 3-8, February 1969.

[B3] Bucy, R. S. and Senne, K. D., "Realization of Optimum Discrete-Time Nonlinear Estimators," Symposium on Nonlinear Estimation Theory and Its Application, San Diego, California, September 21-23, 1970.

[B4] Bryson, A. E. and Ho, Y. C., Applied Optimal Control, Blaisdell Publishing Company, 1969.

[C1] Curry, R. E., "A New Algorithm for Suboptimal Stochastic Control," IEEE Trans. on Automatic Control, Vol. AC-14, pp. 533-536, October 1969.

[D1] Denham, W., "Choosing the Nominal Path for a Dynamic System with Random Forcing Function to Optimize Statistical Performance," TR 449, Division of Eng. and App. Physics, Harvard University, 1964.

[D2] Doob, J. L., Stochastic Processes, John Wiley and Sons, Inc., 1953.

[D3] Dreyfus, S., "Some Types of Optimal Control of Stochastic Systems," SIAM J. on Control 2, No. 1, pp. 120-134, 1964.

[F1] Farison, J. B., Graham, R. E., and Shelton, R. C., "Identification and Control of Linear Discrete Systems," IEEE Trans. on Automatic Control, Vol. AC-12, pp. 438-442, August 1967.

[F2] Feldbaum, A. A., Optimal Control Systems, Academic Press, New York, 1965.

[G1] Gorman, D. and Zaborsky, J., "Stochastic Optimal Control of Continuous Time Systems with Unknown Gain," IEEE Trans. on Auto. Control, 13, pp. 630-638, December 1968.

[J1] Joseph, P. and Tou, J., "On Linear Control Theory," AIEE Trans. Applications and Industry, Vol. 80, pp. 193-196, September 1961.

[J2] Jazwinski, A., "Stochastic Processes and Filtering Theory," Academic Press, 1970.

[K1] Kramer, L. C., "On Stochastic Control and Optimal Measurement Strategies," Electronic Systems Lab., M.I.T., Report ESL-R-462, November 1971.

[L1] Larson, R. E. and Tse, E., "Modal Trajectory Estimation and Parallel Computers," 2nd Symposium on Nonlinear Estimation Theory, San Diego, California, September 1971.

[L2] Lainiotis, D. G., Upadhyay, T. N. and Deshpande, J. G., "Optimal Adaptive Control of Linear Systems," Proc. 1971 IEEE Conf. on Decision and Control, Miami Beach.

[M1] Meier, L., "Combined Optimal Control and Estimation," Proceedings of 3rd Allerton Conference on Systems and Circuits, 1965.

[M2] Meier, L., "Some Stochastic Control Theory," "Information Requirements for Supersonic Transport Operation," NASA Contractor Report CR-1570, pp. 91-95, July 1970.

[M3] Meier, L., Larson, R. E. and Tether, A. J., "Dynamic Programming for Stochastic Control of Discrete Systems," IEEE Trans. on Auto. Control, AC-16, pp. 767-775, December 1971.

[M4] Murphy, W. J., "Optimal Stochastic Control of Discrete Linear Systems with Unknown Gain," IEEE Trans. on Auto. Control, 13, pp. 338-344, August 1968.

[M5] Mehra, R. K., et al., "Dual Control and Identification Methods for Avionic Systems - Part II, Optimal Input for Linear Systems Identification with Applications to Aircraft Parameter Identification," Final Report, AFOSR Proj. F44620-71-C-0077, SCI Project 5971-02, Systems Control Inc., Palo Alto, California, April 1972.

[N1] Nilsson, N. J., Problem-Solving Methods in Artificial Intelligence, McGraw-Hill, 1971.

[S1] Saridis, G. and Lobbia, R. N., "Parameter Identification and Control of Linear Discrete-Time Systems," JACC, 1971.

[S2] Spang, H. A., "Optimum Control of an Unknown Linear Plant Using Bayesian Estimation of the Error," IEEE Trans. on Automatic Control, Vol. AC-10, pp. 80-83, January 1965.

[S3] Streibel, C., "Sufficient Statistics in the Optimum Control of Stochastic Systems," J. of Math. An. and Appl., 12, No. 3, pp. 576-592, 1965.

[S4] Sorenson, H. W., "Kalman Filtering Techniques," in Advances in Control Systems, Vol. 3, C. T. Leondes, ed. New York: Academic Press, 1966.

[S5] Stein, G. and Saridis, G. N., "A Parameter-Adaptive Control Technique," Automatica, Vol. 5, pp. 731-740, November 1969.

[T1] Tse, E. and Athans, M., "Adaptive Stochastic Control for Linear Systems," Part I and II, Proc. 9th IEEE Symp. on Adaptive Processes, Austin, Texas, December 1970.

[T2] Tse, E., "On the Optimal Control of Stochastic Linear Systems," IEEE Trans. on Automatic Control, Vol. AC-16, No. 6, December 1971.

[T3] Tse, E., "Parallel Computation of the Conditional Mean State Estimate for Nonlinear Systems," 2nd Symposium on Nonlinear Estimation Theory, San Diego, California, September 1971.

[T4] Tse, E., "On the Optimal Control of Linear Systems with Incomplete Information," Electron. Syst. Lab., Mass. Inst. Technol., Cambridge, Rep. ESL-R-412, January 1970.

[T5] Tse, E., Dressler, R. and Bar-Shalom, Y., "Application of Adaptive Tuning of Filter to Exoatmospheric Target Tracking," SCI Technical Memorandum, also submitted to 3rd Symposium on Nonlinear Estimation, San Diego, September 1972.

[V1] Vander Stoep, D. R., "Trajectory Shaping for the Minimization of State-Variable Estimation Errors," IEEE Trans. on Automatic Control, Vol. AC-13, No. 3, pp. 284-286, June 1968.