CONTROL OF UNCERTAIN SYSTEMS WITH A
SET-MEMBERSHIP DESCRIPTION OF THE UNCERTAINTY

by

Dimitri Panteli Bertsekas

Electronic Systems Laboratory
Department of Electrical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

CONTROL OF UNCERTAIN SYSTEMS WITH A
SET-MEMBERSHIP DESCRIPTION OF THE UNCERTAINTY

by

DIMITRI PANTELI BERTSEKAS

Diploma of Mechanical and Electrical Engineering,
National Technical University of Athens, Greece
1965

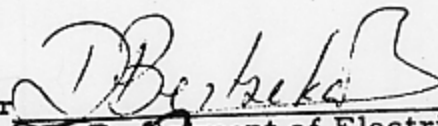M.S., George Washington University, Washington D.C.
1969

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May, 1971

Signature of Author_____
Department of Electrical Engineering, May 7, 1971

Certified by_____
Thesis Supervisor

Accepted by_____
Chairman, Department Committee on Graduate Students

CONTROL OF UNCERTAIN SYSTEMS WITH A
SET-MEMBERSHIP DESCRIPTION OF THE UNCERTAINTY

by

Dimitri P. Bertsekas

ABSTRACT

The problem of optimal feedback control of uncertain discrete-time dynamic systems is considered, where the uncertain quantities do not have a stochastic description but instead they are known to belong to given sets. The problem is converted to a sequential minimax problem and dynamic programming is suggested as a general method for its solution. The notion of a sufficiently informative function, which parallels the notion of a sufficient statistic of stochastic optimal control, is introduced, and the possible decomposition of the optimal controller into an estimator and an actuator is demonstrated.

Some special cases involving a linear system are further examined. A problem involving a convex cost functional and perfect state information for the controller is considered in detail. Particular attention is given to a special case, the problem of reachability of a target tube, and an ellipsoidal approximation algorithm is obtained which leads to linear control laws. State estimation problems are also examined, and some algorithms are derived which offer distinct advantages over existing estimation schemes. These algorithms are subsequently used in the solution of some reachability problems with imperfect state information for the controller.

THESIS SUPERVISOR: Ian B. Rhodes
TITLE: Assistant Professor, Electrical Engineering

ACKNOWLEDGMENT

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

CHAPTER 1

INTRODUCTION

1. General Remarks

The problem of optimal control of uncertain systems has tradi-
tionally been treated in a stochastic framework in the sence that the un-
certain quantities are modeled as random vectors and random processes
with statistical properties which are assumed known. The controller
selected is the one for which the expected value of a suitable cost func-
tional is minimized. In this framework some mathematically elegant
results have been obtained, notable cases being the separation theorem
for a linear system, linear measurements and quadratic cost functional, [J1],
[G1], [Su1] and the separation theorem for a linear system, linear mea-
surements, Gaussian disturbances and nonquadratic cost functional. [St1],
[Wo3] Specification of the a priori statistics of all the uncertain quantities
involved must be made in any such problem. In many practical situations
however these statistics are not available, and cannot be obtained either
because of physical constraints or due to prohibitive cost. In such cases
however the designer may have information of less detailed structure con-
cerning the uncertain quantities, such as for instance bounds on the mag-
nitude or energy of the uncertain quantities. In other words the designer
may be given a set where the uncertain quantities are known to belong.
A possible design approach under these circumstances would then be to
select the controller from some admissible class which performs best
when the uncertain quantities assume their worst possible values within

the given set. In its simplest form the corresponding decision problem is described by a triplet (U, Q, J), where U is the set of controllers under consideration, Q is the set in which the uncertain quantities are known to belong and $J: U \times Q \rightarrow [-\infty, +\infty]$ is a given cost function. The objective is to find

$$\bar{J} = \inf_{u \in U} \sup_{q \in Q} J(u, q) \tag{1.1}$$

and, if it exists, the minimizing controller $\bar{u}$ in U.

Problems of the general form of equation (1.1) can also arise in the context of other situations. In some cases the nature of the problem calls for a pessimistic or worst case approach such as when specified tolerances must be met with certainty. For example in a chemical process control problem it may be necessary to guarantee that the state will stay in a specified region of the state space, or equivalently avoid a critical region of the state space where process instability may occur. In other cases a worst case analysis is performed in order to provide a comparison with the performance of a design adopted on the basis of other considerations.

Optimal uncertain control problems that can be reduced to the form of equation (1.1) are referred to as <u>Minimax Control Problems</u> and are the object of study of this thesis.

The modelling of uncertainties as quantities that are unknown except that they belong to prescribed sets has received attention before, dating to Wald's statistical decision theory.[Wal] In the context of Wald's theory the decision problem (U, Q, J) mentioned earlier is viewed as a game against Nature and a saddle point of this game in (possibly) randomized strategies is sought. Whenever a saddle point in pure strategies exists, i.e., whenever

$$\inf_{u \in U} \sup_{q \in Q} J(u, q) = \sup_{q \in Q} \inf_{u \in U} J(u, q) \qquad (1.2)$$

Wald's approach is equivalent to the worst case approach. When however the equality (1.2) does not hold Wald's theory recommends randomization in the spaces of strategies U and Q, and the worst case viewpoint is lost. Wald's theory was applied by Sworder[Sw1] to discrete-time control systems with limited success since randomization within the admissible set of controllers was not considered appealing from the practical viewpoint of an engineer.

The consideration of the minimax approach to the optimal control of discrete-time uncertain systems without the randomization suggested by Wald's theory was recommended by Feldbaum[F1], [F2] and systematically studied by Witsenhausen.[W1], [W2]

Problems of system state estimation for the case where the uncertain quantities are described by their membership in given sets have also been considered by Witsenhausen,[W1], [W3] Schweppe[S1], [S2], [S3] and others.[Sc1], [H1] Such problems, though important in their own right, arise in connection with minimax control problems for which the controller has available only a noise-corrupted measurement of an output of the system rather than an exact measurement of the system state. Although the emphasis in this thesis is in the feedback control of uncertain systems, some state estimation problems will also be considered which have a direct relation to feedback control problems. In the next section we shall state the basic problem considered in the thesis and outline the general approach which we will adopt towards its solution.

2.   The Basic Problem

The objective of this thesis is the study of the following problem:

Problem 1.1: Given is the discrete-time dynamic system

$$x_{k+1} = f_k(x_k, u_k, w_k), \qquad k = 0, 1, \ldots, N-1 \qquad (1.3)$$

where $x_k \epsilon R^n$, $k = 0, 1, \ldots, N$ is the state vector, $u_k \epsilon R^m$, $k = 0, 1, \ldots, N-1$, is the control vector, $w_k \epsilon R^r$, $k = 0, 1, \ldots, N-1$, is the input disturbance vector, and $f_k : R^n \times R^m \times R^r \rightarrow R^n$ are known functions.

Available to the controller are measurements of the form

$$z_k = h_k(x_k, v_k)^*, \qquad k = 1, 2, \ldots, N-1 \qquad (1.4)$$

where, for all $k = 1, 2, \ldots, N-1$, $z_k \epsilon R^s$ is the measurement vector, $v_k \epsilon R^p$ is the measurement noise vector, and $h_k : R^n \times R^p \rightarrow R^s$ are known functions. The uncertain quantities lumped in a vector $q \epsilon R^{n+Nr+(N-1)p}$

$$q = (x_o', w_o', w_1', \ldots, w_{N-1}', v_1', v_2', \ldots, v_{N-1}')' \qquad (1.5)$$

are known to belong to a given subset Q of $R^{n+Nr+(N-1)p}$

$$q \epsilon Q \qquad (1.6)$$

---

* We restrict attention to this form of measurement equation without loss of generality. A measurement equation containing the control vector explicitly

$$z_k = g_k(x_k, u_{k-1}, v_k)$$

can be reduced to the form (1.4) by introducing additional state variables through the equation

$$\tilde{x}_k = u_{k-1}$$

Attention is restricted to control laws of the form

$$\mu_k : R^{k(s+m)} \to R^m, \qquad k = 0, 1, \ldots, N-1$$

taking values

$$u_k = \mu_k(z_1, z_2, \ldots, z_k, u_o, u_1, \ldots, u_{k-1}), \qquad k = 0, 1, \ldots, N-1$$

where μ_o is interpreted as a constant vector $(\mu_o = u_o)$. It is required to find (if it exists) the control law in this class for which the cost functional

$$J(\mu_o, \mu_1, \ldots, \mu_{N-1}) = \sup_{q \in Q} F[x_1, x_2, \ldots, x_N, \mu_o, \mu_1(z_1, u_o), \ldots,$$

$$\mu_{N-1}(z_1, \ldots, u_{N-2})] \qquad (1.7)$$

is minimized, subject to the system and measurement equation constraints (1.3), (1.4) and where the function $F : R^{N(n+m)} \to (-\infty, \infty]$ is given.

It should be noted that in the statement of the above problem we take into account implicitly the presence of state and control constraints, since we allow the function F in the cost functional (1.7) to take the value ∞. We simply specify that the function F takes the value ∞ whenever some constraint is violated. Thus, for example, state and control constraints of the form $x_k \in X_k$, $u_{k-1} \in U_{k-1}$, where X_k, U_{k-1}, $k = 1, 2, \ldots, N$, are given sets, are accounted for by adding to the function F the function

$$\sum_{i=1}^{N} \{\delta(x_i | X_i) + \delta[\mu_{i-1}(z_1, \ldots, u_{i-2}) | U_{i-1}]\}$$

where $\delta(y|Y)$ denotes the indicator function of a set Y $(\delta(y|Y) = 0$ if $y \in Y$, $\delta(y|Y) = \infty$ if $y \notin Y)$.

The Problem 1.1 can, in principle, be solved by dynamic programming, and the appropriate algorithm will be presented in this thesis. However it is in general very difficult from this algorithm to characterize efficiently the optimal controller which solves Problem 1.1. Thus special cases with increased structure will be considered in order to obtain additional results related to the characterization of the optimal controller and in order to gain increased understanding into the structure of the solution.

One of the major difficulties in solving the general Problem 1.1 results from the fact that the value of the current state of the system (1.3) is not available to the controller but instead only partial information is known about it via the measurements (1.4). This fact results in that, in general, the optimal control law will be a function of all the prior measurements, i.e., in general the controller will need to store all the prior measurements or, possibly, the value of a complicated function of these measurements. However, as in the corresponding stochastic situation, whenever an exact measurement of the current state is available to the controller, i.e., in equation (1.4) we have

$$h_k(x_k, v_k) = x_k \tag{1.8}$$

and in addition the input disturbances w_k are individually constrained at each time

$$w_k \in W_k \subset R^r$$

and the function F in equation (1.7) is of the additive form

$$F(x_1, x_2, \ldots, x_N, u_o, u_1, \ldots, u_{N-1}) = \sum_{k=1}^{N} g_k(x_k, u_{k-1})$$

then it can be shown that the optimal control law is of the form $u_k = \mu_k(x_k)$. In other words the control law need only be a function of the current state, with a substantial simplification resulting. Alternatively expressed, under the circumstances described above, the value of the current state contains all information about the past history of the system which is necessary for the specification of the optimal control.

The special case of Problem 1.1 where equation (1.8) holds is referred to as the minimax control problem with perfect state information and receives considerable attention in this thesis. A large part of the thesis, Chapters 2 and 3, are devoted to problems with perfect state information. This serves a double purpose. In addition to studying a class of problems which is of interest in its own right, we obtain results which are useful for deriving optimal or suboptimal solutions for some minimax control problems with imperfect state information. This is true in particular for the problem of the reachability of a target tube which will be considered extensively in the thesis.

3.   Contributions and Organization of the Thesis

The problems considered in this thesis can be divided into three broad categories. Minimax control problems with perfect state information are considered in Chapters 2 and 3, minimax control problems with imperfect state information are considered in Chapters 5 and 6, and state estimation problems are examined in Chapter 4.

In Chapter 2 a minimax control problem with perfect state information is considered for the case of a linear system and a cost functional with some

convexity properties. This problem in a somewhat less general form was considered first by Witsenhausen.[W1],[W2] Some new results concerning existence of optimal control laws are obtained, and the investigation of necessary conditions for optimality is carried out in depth. A minimax principle is derived for this problem which holds however only under some restrictive assumptions. When specialized to the case of a deterministic optimal control problem this minimax principle yields a minimum principle for which the cost functional is not required to be differentiable.

In Chapter 3 the problem of reachability of a target tube is considered for the perfect state information case. This problem can be viewed as a special case of the problem considered in Chapter 2. Necessary and sufficient conditions for the existence of a solution are obtained. These conditions can also be derived with little effort from Witsenhausen's results.[W1],[W2] In addition a new ellipsoidal approximation algorithm, which appears to have some potential for practical applications, is derived and its properties are investigated.

In Chapter 4 the problem of the system state estimation is examined for a set-membership description of the uncertainty. Attention is restricted to linear systems and two different set-membership descriptions of the uncertainty, the cases of energy constraints and instantaneous ellipsoidal constraints on the uncertain quantities. Some new estimation algorithms are obtained for both cases. In particular, for the case of instantaneous ellipsoidal constraints for the uncertain quantities, an estimator is obtained which offers distinct advantages over the estimator proposed by Schweppe.[S1] Furthermore we use a new approach towards the solution of the problem which allows us to treat some problems not considered as yet in the literature including the smoothing problem.

In Chapter 5 the general case of Problem 1.1 is examined and the dynamic programming algorithm for its solution is developed. This algorithm differs in its form and is more general than the algorithm of Witsenhausen[W1] although the same basic ideas are involved. Subsequently the notion of a sufficiently informative function, which parallels the notion of a sufficient statistic of stochastic optimal control, is formulated for the first time. Some results are then derived which illustrate the dual function of the optimal controller as an estimator and an actuator. This parallels the dual estimation-actuation interpretation of the function of the optimal controller in the analogous problem when the uncertainties are modeled as random vectors or stochastic processes.

Finally in Chapter 6 the problem of the reachability of a target tube with imperfect state information is considered for the case of a linear system. The material in this chapter is new. For the special case of energy constraints on the uncertain quantities the optimal controller is completely characterized, and its separation in an estimator and an actuator is explicitly demonstrated. The case of instantaneous ellipsoidal constraints on the uncertain quantities is also considered, and a suboptimal algorithm is derived which offers some practical implementation advantages.

For the development of some of the results of Chapter 2 it is necessary to appeal in a nontrivial way to the theory of convex functions.[R1] Since portions of this theory are comparatively recent and not very widely known, the required results have been summarized in Appendix I. It should be noted that this theory is used only in Chapter 2, and is not necessary for the developments in the remainder of the thesis.

CHAPTER 2

LINEAR MINIMAX CONTROL PROBLEMS
WITH PERFECT STATE INFORMATION

1.   General Remarks

In this chapter we consider a minimax control problem with perfect state information. As was mentioned in the previous chapter the fact that the controller has available at each time a perfect measurement of the system state results in a substantial simplification in the solution of the problem. For example the dynamic programming algorithm, which is the basic method for solving minimax control problems, becomes greatly simplified for this case. Furthermore, in this chapter we make some additional assumptions which enable us to obtain some deeper analytical results. We assume that the dynamical system involved is linear, and that the cost functional has some convexity properties. This will allow us to consider in detail questions of existence of solutions and necessary conditions for optimality. In addition it will be shown for this case that if the sets where the input disturbances are known to belong are polyhedra, the computational requirements of the dynamic programming algorithm can be further significantly reduced. The results mentioned above rely heavily on the additional structure of linearity for the system and convexity for the cost functional, and do not appear to be available without them. In this way the problem of this chapter should be considered as the special case of the minimax control problem 1.1 which is most amenable to somewhat deeper analysis, and for which the obtained results are considerably stronger than in the general case. Yet this special case is sufficiently general to be of interest in its own right, and the cor-

responding results provide insights into the solution of other more general minimax control problems.

For the development of some of the results of this chapter we will need to draw heavily on some comparatively recent and not very widely known results of the theory of convex functions.[R1] The related theory has been outlined in Appendix I and will be used mainly after Section 3 of this chapter. This theory will not be needed later in the thesis. The reader who is interested in subsequent chapters can proceed to those chapters after section 3 without loss of continuity.

In the next section the minimax control problem of this chapter will be formulated and its solution by dynamic programming will be shown subsequently in Section 3. In Section 4 the properties of the dynamic programming algorithm will be investigated and sufficient conditions for existence of optimal control laws will be derived. In Section 5 necessary conditions for optimality will be obtained. In particular a minimax principle is proved which however holds under somewhat restrictive assumptions. When specialized to deterministic optimal control problems this minimax principle yields a minimum principle for which the cost functional is not assumed differentiable.

2.   Problem Formulation

The object of study in this chapter is the following problem.

Problem 2.1: Consider the linear discrete-time dynamic system:

$$x_{k+1} = A_k x_k + B_k u_k + G_k w_k, \qquad k = 0, 1, \ldots, N-1 \qquad (2.1)$$

where $x_k \in R^n$, $k = 0, 1, .., N$, is the state vector, $u_k \in R^m$, $k = 0, 1, .., N-1$, is the control vector, $w_k \in R^r$, $k = 0, 1, ..., N-1$, is the disturbance vector, and A_k, B_k, G_k, $k = 0, 1, ..., N-1$ are given matrices.

It is assumed that the initial state x_o is known and that the disturbance vectors w_k belong to given nonempty sets $W_k \subseteq R^r$

$$w_k \in W_k, \qquad k = 0, 1, ..., N-1 \qquad (2.2)$$

Attention is restricted to control laws of the form

$$\mu_k : R^n \to R^m, \qquad k = 0, 1, ..., N-1 \qquad (2.3)$$

taking values

$$u_k = \mu_k(x_k), \qquad k = 0, 1, ..., N-1 \qquad (2.4)$$

It is required to find (if it exists) the control law in this class for which the cost functional

$$J(\mu_o, \mu_1, ..., \mu_{N-1}) = \sup_{\substack{w_k \in W_k \\ k=0, 1, .., N-1}} \sum_{k=1}^{N} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\} \qquad (2.5)$$

is minimized, subject to the system equation constraints (2.1), and where the functions $f_k : R^n \to (-\infty, +\infty]$, $g_{k-1} : R^m \to (-\infty, +\infty]$, $k = 1, 2, ..., N$, are given closed proper convex functions.

In Definition A.4 of Appendix I, a closed proper convex function $f : R^n \to (-\infty, +\infty]$ is defined to be an extended real valued convex function which is lower semicontinuous and such that $-\infty < f(x)$ for all $x \in R^n$ and with $f(x) < +\infty$ for at least one $x \in R^n$. Closed proper convex functions are reviewed in more detail in Appendix I. One of the advantages of using extended real

valued functions in the cost functional (2.5) is that state constraints and
control constraints of the form $x_k \epsilon X_k$, $u_k \epsilon U_k$ where X_k, U_k are given
convex sets can be conveniently incorporated in the cost functional rather
than stated explicitly. This is accomplished by adding under the summation
sign in the right hand side of equation (2.5) the indicator functions

$$\delta(x_k | X_k) = \begin{cases} 0 & \text{if } x_k \epsilon X_k \\ +\infty & \text{if } x_k \notin X_k \end{cases}$$

$$\delta(u_k | U_k) = \begin{cases} 0 & \text{if } u_k \epsilon U_k \\ +\infty & \text{if } u_k \notin U_k \end{cases}$$

Since the theory of extended real valued convex functions is well estab-
lished,[R1] introduction of the extended real line does not create difficulties
as long as one is careful to avoid the meaningless sums $\infty - \infty$ and $-\infty + \infty$.

The optimal controller in Problem 2.1 is required to be in feedback
form. As a consequence, local variational analysis is very difficult for this
problem and dynamic programming remains the only method to proceed for
solution. The development of the dynamic programming algorithm for
Problem 2.1 is the object of the next section.

3.   Solution by Dynamic Programming

Let us denote by $\bar{J}_{x_o}$ the optimal value of the cost functional (2.5)

$$\bar{J}_{x_o} = \inf_{\substack{\mu_k \\ k=0,1,\ldots,N-1}} J(\mu_o, \mu_1, \ldots, \mu_{N-1}) \qquad (2.6)$$

The dynamic programming algorithm to be described in the following pro-
position provides the optimal value $\bar{J}_{x_o}$ at the last step of a recursive se-

quence of minimization and maximization steps. Furthermore the optimal
control law (if it exists) can be obtained from the sequence of the minimi-
zation steps in a much simpler way than directly from the equation (2.6).

Proposition 2.1: Assume that for the functions H_k defined below we have
$-\infty < H_k(x_k)$ for all $x_k \epsilon R^n$ and $k = 0, 1, \ldots, N-1$. Then the optimal value
$\bar{J}_{x_o}$ of the cost functional (2.5) is given by

$$\bar{J}_{x_o} = J_o(x_o) \tag{2.7}$$

where the function $J_o : R^n \to (-\infty, +\infty]$ is given by the last step of the re-
cursive algorithm

$$J_N(x_N) = f_N(x_N) \tag{2.8}$$

$$E_{k+1}(x) = \sup_{w_k \epsilon W_k} J_{k+1}(x + G_k w_k), \quad k = 0, 1, \ldots, N-1 \tag{2.9}$$

$$H_k(x_k) = \inf_{u_k} \{E_{k+1}(A_k x_k + B_k u_k) + g_k(u_k)\}, \quad k = 0, 1, \ldots, N-1 \tag{2.10}$$

$$J_k(x_k) = H_k(x_k) + f_k(x_k), \quad k = 1, 2, \ldots, N-1 \tag{2.11}$$

$$J_o(x_o) = H_o(x_o) \tag{2.12}$$

Proof: Since $-\infty < H_{N-1}(x_{N-1})$ for all $x_{N-1} \epsilon R^n$, we have that for every
$\epsilon > 0$ there exists a function $\mu_{N-1, \epsilon} : R^n \to R^m$ such that

$$E_N[A_{N-1}x_{N-1} + B_{N-1}\mu_{N-1, \epsilon}(x_{N-1})] + g_{N-1}[\mu_{N-1, \epsilon}(x_{N-1})]$$

$$\leq \inf_{\mu_{N-1}} \{E_N[A_{N-1}x_{N-1} + B_{N-1}\mu_{N-1}(x_{N-1})] + g_{N-1}[\mu_{N-1}(x_{N-1})]\} + \epsilon$$

$$= H_{N-1}(x_{N-1}) + \epsilon \tag{2.13}$$

By using equations (2.6) and (2.9) we have

$$\bar{J}_{x_o} = \inf_{\substack{\mu_k \\ k=0,1,..,N-1}} \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-1}} \sum_{k=1}^{N} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

$$= \inf_{\substack{\mu_k \\ k=0,1,..,N-2}} \inf_{\mu_{N-1}} \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-2}} \sum_{k=1}^{N-1} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

$$+ E_N[A_{N-1}x_{N-1} + B_{N-1}\mu_{N-1}(x_{N-1})] + g_{N-1}[\mu_{N-1}(x_{N-1})]$$

$$\leq \inf_{\substack{\mu_k \\ k=0,1,..,N-2}} \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-2}} \sum_{k=1}^{N-1} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

$$+ E_N[A_{N-1}x_{N-1} + B_{N-1}\mu_{N-1,\epsilon}(x_{N-1})] + g_{N-1}[\mu_{N-1,\epsilon}(x_{N-1})]$$

Using (2.13) to strenghthen the above inequality we obtain

$$\bar{J}_{x_o} \leq \inf_{\substack{\mu_k \\ k=0,1,..,N-2}} \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-2}} \sum_{k=1}^{N-1} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

$$+ H_{N-1}(x_{N-1}) + \epsilon$$

$$= \inf_{\substack{\mu_k \\ k=0,1,..,N-2}} \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-2}} \inf_{\mu_{N-1}} \sum_{k=1}^{N-1} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

$$+ E_N[A_{N-1}x_{N-1} + B_{N-1}\mu_{N-1}(x_{N-1})] + g_{N-1}[\mu_{N-1}(x_{N-1})] + \epsilon$$

(by using the minimax inequality)

$$\leq \inf_{\substack{\mu_k \\ k=0,1,..,N-2}} \inf_{\mu_{N-1}} \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-2}} \sum_{k=1}^{N-1} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

$$+ E_N[A_{N-1} x_{N-1} + B_{N-1} \mu_{N-1}(x_{N-1})] + g_{N-1}[\mu_{N-1}(x_{N-1})] + \epsilon$$

$$= \bar{J}_{x_0} + \epsilon$$

Since the above relations hold for every $\epsilon > 0$ we conclude that

$$\bar{J}_{x_0} = \inf_{\substack{\mu_k \\ k=0,1,..,N-2}} \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-2}} \left\{ \sum_{k=1}^{N-1} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\} + H_{N-1}(x_{N-1})] \right\}$$

By repeating the above procedure we eventually obtain

$$\bar{J}_{x_0} = H_0(x_0) = J_0(x_0) \qquad \text{Q.E.D.}$$

We remark that the value of the function H_k at a point x_k has the usual interpretation of the "cost-to-go" from the point x_k at time k. This value can be a real number or ∞ but by the assumption $-\infty < H_k(x_k)$, for all $x_k \in R^n$ it cannot be $-\infty$.

The occurance of the equality $H_k(x_k) = -\infty$ for some $x_k \in R^n$ indicates a degeneracy in the problem statement and in particular in the cost functional chosen. It implies the existence of control laws which result in a value of "cost-to-go" which is arbitrarily small starting at state x_k at time k, thus indicating that the optimization problem is not well posed. The assumption $-\infty < H_k(x_k)$ for all $x_k \in R^n$ and all k can be guaranteed to hold under quite general assumptions which will be stated in the next section.

The occurance of a value $H_k(x_k) = \infty$ for some $x_k \in R^n$ has an interesting interpretation in the case where the constraint sets W_k for the disturbance vectors are bounded. It should be recalled that the extended real valued functions f_k and g_k in the cost functional (2.5) specify constraint sets for

the state and the control vectors. It must be $x_k \in X_k$ and $u_k \in U_k$ for all k where the sets X_k and U_k are given for all k by

$$X_k = \{x_k \mid f_k(x_k) < \infty\}$$

$$U_k = \{u_k \mid g_k(u_k) < \infty\}$$

A value $H_k(x_k) = \infty$ implies that, starting from the state x_k at time k, for every control law that the controller uses subject to the control constraints, there exist disturbance vectors within the given sets W_k which will cause a violation of a state constraint at some later stage. It should be noted that if there do not exist any state constraints, i.e., the functions f_k are real valued, and in addition the sets W_k are bounded then we will have $H_k(x_k) < \infty$ for all $x_k \in R^n$ and all k.

The value of the optimal control law $\bar{\mu}_k$ at a point x_k can be obtained from the dynamic programming algorithm as

$$\bar{\mu}_k(x_k) = \bar{u}_k \tag{2.14}$$

where $\bar{u}_k$ is a point (assuming it exists) where the infinum in equation (2.10) is attained for the fixed point x_k. In the case where for the fixed point x_k the infimum in (2.10) is attained at more than one point the equation (2.14) still holds with $\bar{u}_k$ being any one of those points.

Aside from the dimensionality problem, common to every algorithm of this nature, an additional drawback of the dynamic programming algorithm is the maximization indicated in equation (2.9). It will be shown later that, under some quite general assumptions, the functions J_k are convex. Therefore if the set W_k is a compact polyhedron the search for the supremum in

equation (2.9) can be confined to the finite set of the vertices ((R1), Th. 32.2) of W_k thus partly alleviating the computational requirements. In some cases however the sets W_k are only indirectly known via their support functions. This will often occur, for example if the discrete time system (2.1) results from sampling a continuous time linear system.[W2] In this case approximation of the sets W_k by a polyhedron is possible with any desired degree of accuracy. If however this approximation is considered undesirable, use of a dual algorithm [W2] based on equations which will be presented in the next section may be advantageous.

In any case the DP algorithm provides a good starting point for obtaining existence results and necessary conditions for optimality. In the following section its properties will be investigated. In particular properties of the functions E_{k+1}, H_k, J_k, $k = 0, 1, \ldots, N-1$, of equations (2.9) through (2.12) will be deduced. In addition the question of existence of optimal control laws will be considered.

4.   ## Properties of the Dynamic Programming Algorithm and Existence of Optimal Control Laws

Properties of the dynamic programming algorithm will be investigated under assumptions which cover most special cases of the general Problem 2.1 which are of practical interest. Under these assumptions, the question of existence of an optimal control law will be answered satisfactorily. It should be noted that the statement "an optimal control law exists", as we will use it here, means that for every point $x_k \epsilon R^n$ and for every k, k = 0, 1, ..., N-1, there exists a vector $u_k \epsilon R^n$ such that the infimum in equation (2.10) is attained. This does not exclude the possibility that this infimum is ∞. With this inter-

pretation if for some $x_k \epsilon R^n$ we have $E_{k+1}(A_k x_k + B_k u_k) + g_k(u_k) = \infty$, for all $u_k \epsilon R^n$ then, from equation (2.10), $H_k(x_k) = \infty$ and the infimum in (2.10) is attained for every $u_k \epsilon R^n$. This in turn according to our terminology implies existence of an optimal control law in as much as the point x_k is concerned.

The point of view that we adopt concerning the existence of an optimal control law coincides with the usual point of view whenever the given initial condition x_0 is such that the optimal value of the cost functional $\overline{J}_{x_0}$ is finite. As explained in the previous section, whenever the sets W_k are bounded, a value $\overline{J}_{x_0} = \infty$ may occur due to the presence of state constraints $x_k \epsilon X_k$ implied by the functions f_k in (2.5) where

$$X_k = \{x_k | f_k(x_k) < \infty\}$$

There may also exist control constraints $u_k \epsilon U_k$ implied by the functions g_k in (2.5)

$$U_k = \{u_k | g_k(u_k) < \infty\}$$

A value $\overline{J}_{x_0} = \infty$ indicates that for every control law $\mu_k(x_k)$, $k = 0, 1, \ldots, N-1$, there exist disturbance vectors $w_k \epsilon W_k$, $k = 0, 1, \ldots, N-1$, which will cause either a violation of a control constraint or a violation of a state constraint at some stage during the operation of the closed-loop system. In other words a value $\overline{J}_{x_0} = \infty$ indicates that there does not exist a control law which can guarantee the satisfaction of all the constraints of the problem. The question of the existence of such a control law will not be considered in this chapter. This question however is central in the problem of the reachability of a target tube and will be answered in the context of that problem in the next chapter

In order to avoid some rather uninteresting but analytically irritating situations we will make the following assumption which will hold throughout the remainder of this chapter.

Assumption 2.1:

(a) Each of the functions J_k, E_k, H_k of equations (2.8)

through (2.12) is not the constant $+\infty$ function.

(b) The sets W_k, $k = 0, 1, \ldots, N-1$, are compact.

Before we proceed with stating the assumptions under which we will examine the problem of existence of an optimal control law, let us consider the circumstances under which the minimum of a convex function $f : R^n \rightarrow (-\infty, \infty]$ may not be attained. If f is lower semicontinuous the only such situation arises when f decreases monotonically along some direction with the result that either the function is not bounded below or the infimum of the function is finite but "attained at infinity." Typical examples are the functions $f(x) = x$ and $f(x) = e^{-x}$ with $x \in R$. Thus in order to prove the existence of a minimizing vector it is necessary to impose some conditions which will guarantee that the function will not decrease monotonically (recede) along some direction. Such conditions involve the notion of a direction of recession of a closed proper convex function. This notion is introduced in Definition A.10 of Appendix I and its importance in providing existence results for optimization problems is stressed in Proposition A.23 of Appendix I. The assumptions concerning the cost functional (2.5) which we will make involve this notion. We shall consider the following special cases.

Special Case R: In the cost functional (2.5) every direction of recession of each of the functions f_k, $k = 1, 2, \ldots, N$, and g_k, $k = 0, 1, \ldots, N-1$, is a direction in which this function is constant.

Notice that a closed proper convex function f with no direction of recession is characterized by the fact that its nonempty level sets

$$F_a = \{x \mid f(x) \le a\}, \qquad a: \text{real number}$$

are compact, a requirement satisfied by the functions f_k, g_k of many cost functionals of the form (2.5) which are of interest in practice. However we allow the functions f_k and g_k to have directions in which they are constant in order to retain the possibility to weight in the cost functional only certain components of the state and control vectors. The case $f_k(x_k) = x_k' Q x_k$ where Q is only positive semidefinite symmetric matrix is a typical example of such a situation. The basic property of a function belonging to the special case R is that there does not exist any halfline $\{z \mid z = x + \lambda y, \lambda \ge 0\}$ originating at some point $x \in R^n$ and pointing in the direction of some vector $y \in R^n$ along which the function is monotonically decreasing. This excludes the possibility that either the value $f_k(x_k)$ or the value $g_k(u_k)$ decreases monotonically as either x_k or u_k become arbitrarily large (in norm) along some direction.

A second special case which we will consider is the following:

Special Case C: The functions g_k, $k = 0, 1, \ldots, N-1$, of the cost functional (2.5) have a recession function of the form

$$(g_k 0^+)(z) = +\infty \quad \text{for } z \neq 0, \qquad (g_k 0^+)(0) = 0, \quad k = 0, 1, \ldots, N-1$$

The notion of the recession function of a closed proper convex function is introduced in Definition A.9 of Appendix I. Essentially the condition

$$(g_k 0^+)(z) \neq +\infty \quad \text{for } z \neq 0, \qquad (g_k 0^+)(0) = 0$$

requires that the penalty to the controller for using control vectors large in norm is sufficiently great. For example a function g_k does not satisfy this condition if it is uniformly Lipschitz continuous. On the other hand the requirement of the special case C is satisfied if the set $U_k = \{u_k \mid g_k(u_k) < \infty\}$ is compact or if for instance $g_k(u_k) = u_k^! R u_k$ where R is a positive definite symmetric matrix.

Throughout this chapter we shall use the assumption:

Assumption 2.2: The cost functional (2.5) satisfies the requirements of either Special Case R or Special Case C.

We are ready now to prove the following proposition which states that under our assumptions convexity and lower semicontinuity are preserved in the dynamic programming algorithm and that optimal control laws exist.

Proposition 2.2:

(a) Under the Assumptions 2.1 and 2.2 the functions J_k, E_k, H_k of equations (2.8) through (2.12) are closed proper convex functions for all k. This implies in particular that the assumption $-\infty < H_k(x_k)$, for all $x_k \epsilon R^n$ in Proposition 2.1 holds for all k.

(b) The supremum in equation (2.9) is attained.

(c) An optimal control law exists.

Proof: Consider first the function

$$E_N(x) = \sup_{w_{N-1} \epsilon W_{N-1}} f_N(x + G_{N-1} w_{N-1})$$

By Proposition A.10 of Appendix I and using also Assumption 2.1 the function E_N is a closed proper convex function. The supremum is attained for every x since the function $\overline{J}_N(w_{N-1}) = f_N(x + G_{N-1}w_{N-1})$ is lower semicontinuous by Proposition A.12 of Appendix I and the set W_{N-1} is compact. Notice also that for all $w_{N-1} \in W_{N-1}$ the function of x $E_{N, w_{N-1}}(x) = f_N(x + G_{N-1}w_{N-1})$ has the same recession function $E_{N, w_{N-1}}0^+ = f_N0^+$, and thus by Proposition A.10 of Appendix I $E_N0^+ = f_N0^+$. Thus if every direction of recession of f_N is a direction in which it is constant the same is true for the function E_N.

Consider now the function H_{N-1}

$$H_{N-1}(x_{N-1}) = \inf_{u_{N-1}} \{E_N(A_{N-1}x_{N-1} + B_{N-1}u_{N-1}) + g_{N-1}(u_{N-1})\}$$

$$(2.14)$$

By equation (A.2) of Appendix I the function H_{N-1} is given by

$$H_{N-1} = [E_N \square (-B_{N-1})g_{N-1}]A_{N-1} \qquad (2.15)$$

where the notation in the above equation is introduced in Propositions A.4 and A.6 of Appendix I. By using Proposition A.13 of Appendix I we have that both in the special case C and in the special case R the function H_{N-1} is a closed proper convex function and that the infimum is attained in equation (2.14). Also in the special case R every direction of recession of the function H_{N-1} is a direction in which H_{N-1} is constant. The same is true for the function $J_{N-1} = H_{N-1} + f_{N-1}$ since by Proposition A.9 we have $J_{N-1}0^+ = H_{N-1}0^+ + f_{N-1}0^+$ and every direction of recession of the function f_{N-1} is a direction in which f_{N-1} is constant. Thus the proposition is proved for k = N and all the necessary facts have been established so that we can

proceed in exactly the same manner to prove the proposition for k = N-1
and recursively for all k. Q.E.D.

The equation (2.15) shows that the "cost-to-go" function H_{N-1} can
be obtained from the functions E_N and g_{N-1} through operations that have
been extensively studied in the literature.[R1] This fact is very helpful in
the search for sufficient conditions for existence of optimal control laws.
For example stronger sufficient conditions can be derived in the case where
the functions f_k and g_k of the cost functional (2.5) are polyhedral. By making
use of the results of Section 19 in (R1) it can be readily proved that the Pro-
position 2.2 holds for this case under assumptions that are weaker than
Assumption 2.2.

The equation (2.15) can be used also for calculating the conjugate
functions H_k^* and J_k^* via Propositions A.14 and A.15 of Appendix I. We have:

$$H_k^*(x^*) = cl\{A_k'[E_{k+1}^*(x^*) + g_k^*(-B_k'x^*)]\} \qquad (2.16)$$

$$J_k^*(x^*) = cl\{H_k^*(x^*) \,\Box\, f_k^*(x^*)\} \qquad (2.17)$$

where the closure operation $cl\{\cdot\}$ and the infinal convolution operation $\Box$
are introduced in Definition A.5 and Proposition A.4 of Appendix I. The
conjugate E_{k+1}^* of the function E_{k+1} is given by

$$E_{k+1}^*(x^*) = conv\{J_{k+1}^*(x^*) - \sigma(G_k'x^*|W_k)\} \qquad (2.18)$$

where $\sigma(\cdot\,|W_k)$ is the support function of the set W_k and the convex hull
operation is as in Definition A.3 of Appendix I. The equation (2.18) follows
directly from Proposition 25 in (W2).

The equations (2.16), (2.17) and (2.18) can form the basis for a dual algorithm for calculating the optimal cost similar to the one proposed in (W2). The implementation of this algorithm will not be discussed in this thesis. A special case has been analyzed in detail in (W2).

5.   Necessary Conditions for Optimality

In dynamic optimization problems necessary conditions for optimality are usually expressed in terms of the costate vector and the related adjoint equation. This is true for the case of the Pontryagin Minimum Principle[P1], [At1] as well as the Minimax Principle of Zero Sum Differential Games as described by Isaacs.[Is1] In both these cases at points of an optimal trajectory where the "cost-to-go" function is differentiable, the costate vector is equal to the gradient of the "cost-to-go" function. In light of this fact it is not surprising that the necessary conditions for optimality which we derive for the minimax problem of this chapter involve vectors in the subdifferentials (generalized gradients) of the "cost-to-go" functions J_k, H_k of the equations (2.8) through (2.12). The notion of the subdifferential $\partial f(x)$ of a convex function of at a point x is introduced in Definition A.12 of Appendix I and some of the pertinent facts are summarized in subsequent propositions. It should be noted that the use of subdifferential theory in the analysis is necessitated by the fact that the "cost-to-go" functions J_k and H_k will in most cases be nondifferentiable even if the functions f_k and g_k in the cost functional (2.5) are real valued and differentiable. This is mainly due to the maximization indicated in equation (2.9) as will be shown later.

We now prove the following necessary conditions in order that the supremum and infimum in the equations

$$E_{k+1}(x) = \sup_{w_k \in W_k} J_{k+1}(x + G_k w_k), \qquad k = 0, 1, .., N-1 \qquad (2.9)$$

$$H_k(x_k) = \inf_{u_k} \{E_{k+1}(A_k x_k + B_k u_k) + g_k(u_k)\}, \qquad k = 0, 1, .., N-1 \qquad (2.10)$$

are attained at given points.

Proposition 2.3: For a fixed point $x \in R^n$ let $\overline{w}_k \in W_k$ be a point where the supremum is attained in equation (2.9). Then for all vectors $x_{k+1}^* \in \partial J_{k+1}(x + G_k \overline{w}_k)$ we have

$$\langle x_{k+1}^*, G_k \overline{w}_k \rangle = \max_{w_k \in W_k} \langle x_{k+1}^*, G_k w_k \rangle$$

where $\partial J_{k+1}(x + G_k \overline{w}_k)$ denotes the subdifferential of the function J_{k+1} at the point $(x + G_k \overline{w}_k)$.

Proof: Let $x_{k+1}^* \in \partial J_{k+1}(x + G_k \overline{w}_k)$. By Proposition A.18 of Appendix I we have

$$J_{k+1}(x + G_k \overline{w}_k) = \langle x_{k+1}^*, x + G_k \overline{w}_k \rangle - J_{k+1}^*(x_{k+1}^*) \qquad (2.19)$$

By equation (2.18) we have

$$E_{k+1}^*(x_{k+1}^*) = \text{conv}\{J_{k+1}^*(x_{k+1}^*) - \sigma(G_k' x_{k+1}^* | W_k)\}$$

$$\leq J_{k+1}^*(x_{k+1}^*) - \sigma(G_k' x_{k+1}^* | W_k)$$

Using the above inequality in equation (2.19)

$$J_{k+1}(x + G_k \overline{w}_k) \leq \langle x + G_k \overline{w}_k, x_{k+1}^* \rangle - E_{k+1}^*(x_{k+1}^*) - \sigma(G_k' x_{k+1}^* | W_k)$$

On the other hand

$$\langle x, x_{k+1}^* \rangle - E_{k+1}^*(x_{k+1}^*) \leq \sup_{x_{k+1}^*} \{\langle x, x_{k+1}^* \rangle - E_{k+1}^*(x_{k+1}^*)\}$$

$$= E_{k+1}(x) = J_{k+1}(x + G_k \overline{w}_k) \qquad (2.21)$$

Combining the inequalities (2.20) and (2.21) we obtain

$$\langle x_{k+1}^*, G_k \overline{w}_k \rangle \geq \sigma(G_k' x_{k+1}^* | W_k) = \max_{w_k \in W_k} \langle x_{k+1}^*, G_k w_k \rangle$$

which proves the desired equation. Q.E.D.

Consider now the function $\widetilde{H}_k$ defined by

$$\widetilde{H}_k(x) = \inf_{u_k} \{E_{k+1}(x + B_k u_k) + g_k(u_k)\} \qquad (2.22)$$

It is clear that if for a fixed point $x_k \in R^n$ the infimum in equation (2.10) is attained at a point $\overline{u}_k$ then the infimum in equation (2.22) is attained at the same point $\overline{u}_k$ when $x = A_k x_k$. Notice that for all $x_k \in R^n$ we have $H_k(x_k) = \widetilde{H}(A_k x_k)$ and that $\widetilde{H}_k = E_{k+1} \square (-B_k) g_k$, a relation which is proved in the same way as equation (A.2) in Appendix I. From Propositions A.14 and A.15 it follows then that the conjugate convex function $\widetilde{H}_k^*$ of the function $\widetilde{H}_k$ is given by

$$\widetilde{H}_k^*(x^*) = E_{k+1}^*(x^*) + g_k^*(-B_k' x^*) \qquad (2.23)$$

We now have:

Proposition 2.4: For a fixed point $x_k \in R^n$ let $\overline{u}_k$ be a point where the infimum is attained in equation (2.10). Then for all vectors $x^* \in \partial \widetilde{H}_k(A_k x_k)$ we have

$$\langle x^*, B_k \overline{u}_k \rangle + g_k(\overline{u}_k) = \min_{u_k} \{\langle x^*, B_k u_k \rangle + g_k(u_k)\}$$

where $\tilde{H}_k$ is the function defined in equation (2.22).

<u>Proof</u>: Let $x^* \epsilon \partial \tilde{H}_k(A_k x_k)$. By Proposition A.18 we have

$$H_k(x_k) = \tilde{H}_k(A_k x_k) = \langle A_k x_k, x^* \rangle - \tilde{H}_k^*(x^*)$$

or by equation (2.23)

$$H_k(x_k) = \langle A_k x_k, x^* \rangle - E_{k+1}^*(x^*) - g_k^*(-B_k' x^*) \qquad (2.24)$$

On the other hand by the optimality of $\bar{u}_k$

$$H_k(x_k) = E_{k+1}(A_k x_k + B_k \bar{u}_k) + g_k(\bar{u}_k)$$

$$= \sup_{x^*} \{\langle A_k x_k + B_k \bar{u}_k, x^* \rangle - E_{k+1}^*(x^*)\} + g_k(\bar{u}_k)$$

$$\geq \langle A_k x_k, x^* \rangle - E_{k+1}^*(x^*) + \langle B_k \bar{u}_k, x^* \rangle + g_k(\bar{u}_k) \qquad (2.25)$$

Combining relations (2.24) and (2.25)

$$\langle x^*, B_k \bar{u}_k \rangle + g_k(\bar{u}_k) \leq -g_k^*(-B_k' x^*)$$

$$= \inf_{u_k} \{\langle x^*, B_k u_k \rangle + g_k(u_k)\}$$

which proves the desired equation. Q.E.D.

Notice that if the matrix A_k is invertible then by Proposition A.20 of the Appendix

$$\partial \tilde{H}_k(A_k x_k) = A_k'^{-1} \partial H_k(x_k)$$

and the necessary condition of Proposition 2.4 becomes

$$\langle A_k'^{-1} x_k^*, B_k \bar{u}_k \rangle + g_k(\bar{u}_k) = \min_{u_k} \{\langle A_k'^{-1} x_k^*, B_k u_k \rangle + g_k(u_k)\}, \forall x_k^* \epsilon \partial H_k(x_k)$$

A sufficient condition for the infimum to be attained at a given point in equation (2.10) is given by the following proposition:

Proposition 2.5: Assume that for some vector x^*_{k+1} and some vector z we have $x^*_{k+1} \epsilon \partial E_{k+1}(z)$ and that for a vector $\bar{u}_k$

$$\langle x^*_{k+1}, B_k \bar{u}_k \rangle + g_k(\bar{u}_k) = \min_{u_k} \{ \langle x^*_{k+1}, B_k u_k \rangle + g_k(u_k) \}$$

Then we have

$$\widetilde{H}_k(z - B_k \bar{u}_k) = E_{k+1}(z) + g_k(\bar{u}_k) \qquad (2.26)$$

i.e., the infimum in equation (2.22) is attained at the point $\bar{u}_k$ when $x = z - B_k \bar{u}_k$. In addition we have $x^*_{k+1} \epsilon \partial \widetilde{H}_k(z - B_k \bar{u}_k)$.

Proof: We have $\widetilde{H}_k(z - B_k \bar{u}_k) \leq E_{k+1}(z) + g_k(\bar{u}_k)$ and by using equations (2.22) and (2.23)

$$\langle x^*_{k+1}, z - B_k \bar{u}_k \rangle - \widetilde{H}_k(z - B_k \bar{u}_k)$$

$$\geq \langle x^*_{k+1}, z \rangle - E_{k+1}(z) + \langle -B'_k x^*_{k+1}, \bar{u}_k \rangle - g_k(\bar{u}_k)$$

$$= E^*_{k+1}(x^*_{k+1}) + g^*_k(-B'_k x^*_{k+1}) = \widetilde{H}^*_{k+1}(x^*_{k+1})$$

$$= \sup_z \{ \langle x^*_{k+1}, z - B_k \bar{u}_k \rangle - \widetilde{H}_k(z - B_k \bar{u}_k) \}$$

Hence equality holds in the above algebra implying equation (2.26) and that $x^*_{k+1} \epsilon \partial \widetilde{H}_k(z - B_k \bar{u}_k)$) Q.E.D.

It is interesting to make the following observation in Propositions 2.3 and 2.4. Consider a fixed point $\bar{x}_k \in R^n$, and let $\bar{u}_k$ be a vector where the infimum is attained in equation (2.10). Let also $\overline{w}_k$ be a point where the supremum is attained in equation (2.9) for $x = A_k \bar{x}_k + B_k \bar{u}_k$. Then for any vector x^* such that

$$x^* \epsilon \partial H_k(A_k \bar{x}_k) \bigcap \partial J_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k + G_k \overline{w}_k) \tag{2.27}$$

we have from Propositions 2.3 and 2.4 that

$$\langle x^*, B_k \bar{u}_k + G_k \overline{w}_k \rangle + g_k(\hat{u}_k)$$

$$= \min_{u_k} \max_{w_k \epsilon W_k} \{ \langle x^*, B_k u_k + G_k w_k \rangle + g_k(u_k) \} \tag{2.28}$$

or equivalently

$$\langle x^*, A_k \bar{x}_k + B_k \bar{u}_k + G_k \overline{w}_k \rangle + g_k(\bar{u}_k)$$

$$= \min_{u_k} \max_{w_k \epsilon W_k} \{ \langle x^*, A_k \bar{x}_k + B_k u_k + G_k w_k \rangle + g_k(u_k) \}$$

Notice that the expression within braces in the above relation is the familiar Hamiltonian. It is evident that if along an optimal trajectory one could guarantee for every k the existence of vectors x^* such that the relation (2.27) holds and find a law for propagation of these vectors (i.e., an adjoint equation) then the Proposition 2.3 and 2.4 would be pieced together into a Minimax Principle. The remainder of this section will be devoted to an effort in this direction.

We first give the definition of a minimax sequence and a minimax trajectory:

<u>Definition 2.1</u>: A sequence of control and disturbance vectors

$$\{\bar{u}_o, \bar{w}_o, \bar{u}_1, \bar{w}_1, \ldots, \bar{u}_{N-1}, \bar{w}_{N-1}\}$$

is called a <u>minimax sequence</u> and the corresponding trajectory $\{\bar{x}_o, \bar{x}_1, \ldots, \bar{x}_N\}$ given by

$$\bar{x}_{k+1} = A_k \bar{x}_k + B_k \bar{u}_k + G_k \bar{w}_k, \qquad k = 0, 1, \ldots, N-1$$

$$\bar{x}_o = x_o$$

is called a <u>minimax trajectory</u> if for all k

$$H_k(\bar{x}_k) = \tilde{H}_k(A_k\bar{x}_k) = E_{k+1}(A_k\bar{x}_k + B_k\bar{u}_k) + g_k(\bar{u}_k)$$

$$= \inf_{u_k} \{E_{k+1}(A_k\bar{x}_k + B_k u_k) + g_k(u_k)\} \tag{2.10}$$

$$E_{k+1}(A_k\bar{x}_k + B_k\bar{u}_k) = J_{k+1}(\bar{x}_{k+1})$$

$$= \sup_{w_k \in W_k} J_{k+1}(A_k\bar{x}_k + B_k\bar{u}_k + G_k w_k)$$

A minimax trajectory results during operation of the system (2.1) when an optimal control law is used and when disturbances are selected (by Nature) in an optimal fashion. It is evident from the Definition 2.1 and the dynamic programming algorithm of Proposition 2.1 that if a minimax sequence and a corresponding minimax trajectory could be found then the optimal cost for Problem 2.1 would be obtained as

$$\bar{J}_{x_o} = J_o(x_o) = \sum_{k=1}^{N} \{f_k(\bar{x}_k) + g_{k-1}(\bar{u}_{k-1})\}$$

and the problem would be at least partially solved. In what follows we obtain

a necessary condition in order for a control and disturbance sequence to be a minimax sequence under some special assumptions.

We shall make the following assumption concerning the convex functions H_k, $\tilde{H}_k$ defined by equations (2.8) through (2.12) and equation (2.22)

Assumption 2.3:

    (a) For all k the range of the matrix A_k contains a vector in $ri(dom \tilde{H}_k)$

    (b) For all k we have

$$ri(dom f_k) \cap ri(dom H_k) \neq \phi$$

where the relative interior of the effective domain $ri(dom \cdot)$ of a convex function is defined in Definition A.7 of Appendix I.

The assumption (a) above is needed in order to guarantee by Proposition A.20 of Appendix I that

$$\partial H_k(x_k) = A_k' \partial \tilde{H}_k(A_k x_k)$$

where $\tilde{H}_k$ is the function defined in equation (2.22), a relation essential for the proof of Proposition 2.8. This assumption will hold for most problems. In particular it will hold if the matrix A_k is invertible or if the functions f_k are real valued in which case it can be easily seen that the functions H_k will also be real valued and hence $dom H_k = ri(dom H_k) = R^n$. The reason for introducing assumption (b) above is to guarantee by Proposition A.19 of the Appendix I that

$$\partial J_k(x_k) = \partial H_k(x_k) + \partial f_k(x_k)$$

a relation also essential in the proof of Proposition 2.8. This assumption will again hold for most problems and in particular it will hold if the functions f_k are real valued.

Assume now that $\{\bar{u}_o, \bar{w}_o, \bar{u}_1, \bar{w}_1, \ldots, \bar{u}_{N-1}, \bar{w}_{N-1}\}$ is a minimax sequence and $\{\bar{x}_o, \bar{x}_1, \ldots, \bar{x}_N\}$ is the corresponding minimax trajectory. The necessary conditions of Propositions 2.3 and 2.4 hold for the vectors $\bar{u}_k$ and $\bar{w}_k$. Some preliminary facts concerning the subdifferentials $\partial \tilde{H}_k(A_k \bar{x}_k)$, $\partial J_{k+1}(\bar{x}_{k+1})$ will be proved now in the following two Propositions:

<u>Proposition 2.6</u>: For all $k = 0, 1, \ldots, N-1$ we have

$$\partial \tilde{H}_k(A_k \bar{x}_k) \subset \partial E_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k)$$

<u>Proof</u>: Let $x^* \epsilon \partial \tilde{H}_k(A_k \bar{x}_k)$. By the subgradient inequality (A.4) in Appendix I we have

$$\tilde{H}_k(z) \geq \tilde{H}_k(A_k \bar{x}_k) + <z - A_k \bar{x}_k, x^*>, \qquad \forall z \epsilon R^n \qquad (2.29)$$

Since from equation (2.22)

$$\tilde{H}_k(A_k \bar{x}_k) = E_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k) + g_k(\bar{u}_k)$$

$$\tilde{H}_k(z) \leq E_{k+1}(z + B_k \bar{u}_k) + g_k(\bar{u}_k)$$

using the above relations in (2.29) we obtain

$$E_{k+1}(z + B_k \bar{u}_k) \geq E_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k) + <z - A_k \bar{x}_k, x^*>, \forall z \epsilon R^n$$

which implies that $x^* \epsilon \partial E_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k)$. Q.E.D.

Proposition 2.7: For a fixed point x let $\overline{W}_k(x)$ be the set of points where the supremum is attained in equation (2.9). Then

$$\partial E_{k+1}(x) \supset \mathrm{conv}\{x^* \epsilon \partial J_{k+1}(x + G_k \overline{w}_k) \mid \overline{w}_k \epsilon \overline{W}_k(x)\} \qquad (2.30)$$

and if $x \epsilon \mathrm{int}(\mathrm{dom}\ E_{k+1})$ we have

$$\partial E_{k+1}(x) = \mathrm{conv}\{x^* \epsilon \partial J_{k+1}(x + G_k \overline{w}_k) \mid \overline{w}_k \epsilon \overline{W}_k(x)\} \qquad (2.31)$$

where $\mathrm{conv}\{\cdot\}$ denotes convex hull of the set within braces.

Proof: Let $\overline{w}_k \epsilon \overline{W}_k(x)$ and $x^* \epsilon \partial J_{k+1}(z + G_k \overline{w}_k)$ then

$$J_{k+1}(z + G_k \overline{w}_k) \geq J_{k+1}(x + G_k \overline{w}_k) + <z-x, x^*>, \quad \forall z \epsilon R^n \qquad (2.32)$$

Since $J_{k+1}(x + G_k \overline{w}_k) = E_{k+1}(x)$ and $J_{k+1}(z + G_k \overline{w}_k) \leq E_{k+1}(z)$ from the relation (2.32) we obtain

$$E_{k+1}(z) \geq E_{k+1}(x) + <z-x, x^*>, \qquad \forall z \epsilon R^n$$

implying that $x^* \epsilon \partial E_{k+1}(x)$ and therefore

$$\partial E_{k+1}(x) \supset \partial J_{k+1}(x + G_k \overline{w}_k), \qquad \forall\ \overline{w}_k \epsilon \overline{W}_k(x)$$

Since $\partial E_{k+1}(x)$ is a closed convex set

$$\partial E_{k+1}(x) \supset \mathrm{conv}\{x^* \epsilon \partial J_{k+1}(x + G_k \overline{w}_k) \mid \overline{w}_k \epsilon \overline{W}_k(x)\}$$

To prove the equality (2.31) observe that the function $\mathscr{G}(x, w_k) = J_{k+1}(x + G_k w_k)$ satisfies all the assumptions of Proposition A.22 of Appendix I. The equality (2.31) follows directly from the conclusion of this proposition. Q.E.D.

From Propositions 2.6 and 2.7 it cannot be guaranteed that the set intersection indicated in relation (2.27) is a nonempty set, and in fact

examples can be found where

$$\partial \tilde{H}(A_k \bar{x}_k) \cap \partial J_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k + G_k \bar{w}_k) = \phi$$

If however the equality

$$\partial E_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k) = \partial J_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k + G_k \bar{w}_k) \qquad (2.32)$$

holds, then from Proposition 2.6 we obtain

$$\partial \tilde{H}(A_k \bar{x}_k) \subset \partial J_{k+1}(A_k \bar{x}_k + B_k \bar{u}_k + G_k \bar{w}_k)$$

in which case, assuming $\partial \tilde{H}_k(A_k \bar{x}_k) \neq \phi$, the minimax condition of equation (2.28) would hold for every $x^* \epsilon \partial \tilde{H}_k(A_k \bar{x}_k)$.

By Proposition 2.7 the equation (2.32) is satisfied for every point $x = (A_k \bar{x}_k + B_k \bar{u}_k) \epsilon \text{int}(\text{dom } E_{k+1})$ for which the supremum in equation (2.9) is attained at a single point. It may be satisfied also for other points on the boundary of dom E_{k+1}. Points $(A_k \bar{x}_k + B_k \bar{u}_k)$ for which equation (2.32) is satisfied will be called nonsingular according to the following definition.

<u>Definition 2.1</u>: For fixed k, k = 1, 2, .., N, a point x is called <u>nonsingular</u> if for every vector $\bar{w}_k$ such that the supremum in equation (2.9) is attained, we have

$$\partial E_{k+1}(x) = \partial J_{k+1}(x + G_k \bar{w}_k)$$

A point x which is not nonsingular will be called <u>singular</u>.

It should be noted that in view of Proposition 2.7, every point x at which E_{k+1} is differentiable is by necessity nonsingular, assumming that $\partial J_{k+1}(x + G_k \bar{w}_k) \neq \phi$.

We shall also need to distinguish initial states x_o for which the subdifferential $\partial J_o(x_o)$ is nonempty by the following definition.

Definition 2.2: The initial state x_o will be called regular if the subdifferential $\partial J_o(x_o)$ is nonempty.

Notice that by Proposition A.17 of the Appendix all initial states $x_o \in ri(dom\ J_o)$ are regular whereas all initial states for which $J_o(x_o) = \infty$ are not regular.

We are now ready to prove the following necessary condition for a minimax sequence.

Proposition 2.8: Let $\{\bar{u}_o, \bar{w}_o, \bar{u}_1, \bar{w}_1, \ldots, \bar{u}_{N-1}, \bar{w}_{N-1}\}$ be a minimax sequence and let $\{x_o, \bar{x}_1, \ldots, \bar{x}_N\}$ be the corresponding minimax trajectory. Assume that the initial state x_o is regular and that the points $(A_k\bar{x}_k + B_k\bar{u}_k)$, $k = 0, 1, \ldots,$ N-1 are nonsingular. Then there exist vectors $x_1^*, x_2^*, \ldots, x_N^*, p_1^*, p_2^*, \ldots, p_{N-1}^*$ satisfying the adjoint equation

$$x_k^* = A_k' x_{k+1}^* + p_k^*, \qquad k = 1, \ldots, N-1 \qquad (2.33)$$

with

$$x_N^* \in \partial f_N(\bar{x}_N)$$

$$p_k^* \in \partial f_k(\bar{x}_k), \qquad k = 1, 2, \ldots, N-1$$

and such that

$$\langle x_{k+1}^*, B_k\bar{u}_k + G_k\bar{w}_k \rangle + g_k(\bar{u}_k)$$

$$= \min_{u_k} \max_{w_k \in W_k} \{\langle x_{k+1}^*, B_k u_k + G_k w_k \rangle + g_k(u_k)\}, \quad k = 0, 1, \ldots, N-1$$

$$(2.34)$$

<u>Proof</u>: By the fact that x_o is a regular point we have $\partial J_o(x_o) \neq \phi$. Since by the Assumption 2.3 we have $\partial J_o(x_o) = A_o' \partial H_o(A_o x_o)$ we conclude that $\partial H_o(A_o x_o) \neq \phi$. Take x_1^* to be any vector in $\partial H_o(A_o x_o)$. By Proposition 2.5 and the fact that the point $(A_o x_o + B_o \bar{u}_o)$ is nonsingular we have

$$\partial H_o(A_o x_o) \subset \partial E_1(A_o x_o + B_o \bar{u}_o) = \partial J_1(\bar{x}_1)$$

and therefore by Propositions 2.3 and 2.4 the minimax condition of equation (2.34) is satisfied for $k = 0$. By the Assumption 2.3 we have

$$\partial J_1(\bar{x}_1) = A_1' \partial H_1(A_1 \bar{x}_1) + \partial f_1(\bar{x}_1)$$

and thus we can find vectors x_2^* and p_1^* such that

$$x_1^* = A_1' x_2^* + p_2^*$$

and $x_2^* \epsilon \partial H_1(A_1 \bar{x}_1)$, $p_2^* \epsilon \partial f_1(\bar{x}_1)$. Again by Proposition 2.5 and the fact that the point $(A_1 \bar{x}_1 + B_1 \bar{u}_1)$ is nonsingular we have

$$\partial H_1(A_1 \bar{x}_1) \subset \partial E_1(A_o \bar{x}_o + B_o \bar{u}_o) = \partial J_2(\bar{x}_2)$$

and therefore by Propositions 2.3 and 2.4 the minimax condition of equation (2.34) is satisfied for $k = 1$. By proceeding in a similar manner we construct the sequence $x_1^*, x_2^*, \ldots, x_N^*, p_1^*, p_2^*, \ldots, p_{N-1}^*$. For these vectors the adjoint equation (2.33) as well as the minimax condition (2.34) is satisfied. Q.E.D.

The Proposition 2.8 states that a minimax principle holds along a minimax trajectory provided this trajectory does not go through singular points and the initial condition is regular. This is reminiscent of the minimax principle of differential games which holds provided the optimal trajectory does not go through singular surfaces.

Except for the assumption concerning nonsingular points every other assumption used in the proof of Proposition 2.8 is required in order to rule out rather pathological cases which seldom occur in practice. However the assumption that the trajectory does not go through singular points is a formidable one. Except for particularly well behaved problems, singular points are a common occurance and invariably minimax trajectories corresponding to some initial states will go through these points. One can prove in fact that if x_o is an initial state which is such that there exists a minimax trajectory starting from x_o which does not go through singular points then we must have

$$J_L(x_o) = J_o(x_o) = \bar{J}_{x_o} \tag{2.35}$$

where $\bar{J}_{x_o} = J_o(x_o)$ is the optimal cost of Problem 2.1 corresponding to x_o and $J_L(x_o)$ is given by

$$J_L(x_o) = \sup_{\substack{w_k \in W_k \\ k=0,1,..,N-1}} \inf_{\substack{\mu_k \\ k=0,1,..,N-1}} \sum_{k=1}^{N} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

The equation (2.35) implies the existence of a saddle point in the zero-sum game where the players are the controller and nature and the payoff function is

$$\sum_{k=1}^{N} \{f_k(x_k) + g_{k-1}[\mu_{k-1}(x_{k-1})]\}$$

Since we have $J_L(x_o) \leq J_o(x_o)$ with strict inequality holding in general for a "large" set of initial states, the equation (2.35) illustrates the limitations of Proposition 2.8.

It appears that, in general, it is a formidable task to determine the set of initial states which are such that their corresponding minimax trajectories do not go through singular points. The same is true for determining whether a particular point is singular or not. Thus even if a candidate for a minimax sequence is found through Proposition 2.8 it may be very difficult to verify whether in fact it is a minimax sequence.

In conclusion the necessary conditions presented in this section should be expected to provide a complete solution to only a limited class of problems. The class of problems for which the singular points either do not exist at all or can be detected by graphical or analytical methods.

One class of problems where singular points do not occur is the case where the functions f_k of the cost functional (2.5) are linear

$$f_k(x_k) = <x_k, c_k>, \qquad k = 1, 2, \ldots, N-1$$

where $<\cdot, \cdot>$ denotes inner product and c_k are given vectors in R^n. If the functions g_k, $k = 0, 1, \ldots, N-1$ satisfy the requirements of special case C then it can be easily proved that the functions E_k, H_k, J_k of the DP algorithm are linear functions. In particular the functions E_k are differentiable, and therefore singular points do not appear. A minimax sequence for this problem can be obtained by making use of the minimax condition of Proposition 2.8.

The minimax principle of Proposition 2.8 however can be used in still a different way. Assume that a sequence $\{\bar{u}_o, \bar{w}_o, \bar{u}_1, \bar{w}_1, \ldots, \bar{u}_{N-1}, \bar{w}_{N-1}\}$ with a corresponding trajectory $\{x_o, \bar{x}_1, \ldots, \bar{x}_N\}$ has been found via the minimax condition (2.34), and that one cannot verify whether indeed this sequence is minimax sequence. Let

$$\tilde{J}_{x_o} = \sum_{k=1}^{N} [f_k(\tilde{x}_k) + g_{k-1}(\bar{u}_{k-1})]$$

be the value of the cost corresponding to the sequence. Then one can easily prove by making use of Propositions 2.3, 2.4, and 2.5 that the inequality

$$\tilde{J}_{x_o} \leq \bar{J}_{x_o} \tag{2.36}$$

holds, where $\bar{J}_{x_o}$ is the optimal value of the cost functional (2.5). In some cases now minimax problems are solved in order to determine the optimal value $\bar{J}_{x_o}$ and compare it with the worst-case performance, say J_{x_o}, of a controller selected on the basis of other considerations. The reasoning used is that if the difference $(J_{x_o} - \bar{J}_{x_o})$ is relatively small then it can be concluded that the worst-case performance of this suboptimal controller is not unduly poor. Since, by using the relation (2.36), we have

$$0 \leq J_{x_o} - \bar{J}_{x_o} \leq J_{x_o} - \tilde{J}_{x_o}$$

a "small" value of $(J_{x_o} - \tilde{J}_{x_o})$ can guarantee that the worst-case performance of the controller under consideration is acceptable.

6. Discussion and Sources

The basic approach towards the solution of the problem of this chapter is dynamic programming. The computational requirements for this algorithm depend on the dimension of the system and the nature of the sets W_k in which the disturbance is known to belong. If the sets W_k are compact polyhedra with a relatively small number of vertices the computational requirements are only slightly greater than those for a deterministic optimal control problem with the same state and control vector dimensions.

In some cases a dual algorithm involving the conjugate convex functions of the "cost-to-go" functions can offer computational advantages, particularly if the sets W_k are only indirectly known via their support functions.

New results in this chapter are the existence results of Proposition 2.2 and some of the necessary conditions in Section 4. The minimax principle of Proposition 2.8 should not be considered as a powerful tool for solving a wide variety of problems. It can be useful however in some cases and it is of theoretical interest since, together with the developments preliminary to its proof, it provides insight into the mechanism of optimality for the problem of this chapter.

Two special cases of Problem 2.1 are of interest in deterministic optimal control theory. In the first case the sets W_k consist of a single point $\bar{w}_k$, $W_k = \{\bar{w}_k\}$, $k = 0, 1, \ldots, N-1$. For this optimal control problem the Proposition 2.2 yields existence results that to the author's knowledge, are stronger than those available in the literature. Some of the results on existence of optimal controls in Lee and Markus[L1] are along the same lines. For the same case the Proposition 2.8 yields a Minimum Principle which holds for a linear discrete-time system and a convex but <u>not differentiable</u> cost functional. This Minimum Principle is a sufficient as well as necessary condition for optimality as can be easily verified by using Proposition 2.5. Notice that for this case there exist no singular points due to the nature of the sets W_k. A similar Minimum Principle for a linear continuous-time system and a convex but not differentiable cost functional has already appeared in (Hel). Necessary conditions along similar lines can also be found in (R2), (Lu2), (B3). A second special case of interest in

deterministic optimal control theory is the case where the system is de-
scribed by the equation

$$x_{k+1} = A_k x_k + G_k w_k, \qquad k = 0, 1, \ldots, N-1 \qquad (2.36)$$

and it is required to find

$$\overline{J}_{x_o} = \sup_{\substack{w_k \in W_k \\ k=0,1,\ldots,N-1}} \sum_{k=1}^{N} f_k(x_k) \qquad (2.37)$$

where f_k, $k = 1, 2, \ldots, N$ are real valued convex functions. This problem
can be recognized as the special case of Problem 2.1 with the functions
g_k defined as

$$g_k(u_k) = \infty \quad \text{for } u_k \neq 0, \ g_k(0) = 0, \qquad k = 0, 1, \ldots, N-1$$

For this problem Proposition 2.8 yields a Maximum Principle which can be
proved without the assumption that the optimal trajectory does not go through
singular points, and that the initial state is a regular point. The proof is
based on Propositions 2.3 and 2.7 and an argument similar to the one used
for the proof of Proposition 2.8. This Maximum Principle holds for a linear
system and a nondifferentiable convex functional and provides a necessary,
but not sufficient, condition for optimality for the problem of equations (2.36)
and (2.37). It can be easily generalized for the case of system (2.36) where
it is required to find

$$\overline{J}_{x_o} = \sup_{\substack{w_k \in W_k \\ k=0,1,\ldots,N-1}} \sum_{k=1}^{N} \{ f_k(x_k) + h_{k-1}(w_{k-1}) \}$$

where h_k, $k = 0, 1, \ldots, N-1$, are any continuous real valued functions on R^r.

Open-loop discrete time minimax control problems can be viewed as single-stage feedback problems ($N = 1$), and therefore for the case of a linear system and a convex cost functional they can be considered as a special case of Problem 2.1. In addition to the results of this chapter the necessary conditions in (Da1), (Da2), (Bra1), (D1), (D2), (D3) and the computational algorithms in (D1), (Sa1), (Ps1), (B3) can be used for the solution of such problems. Some of the material in these references is applicable to nonlinear and nonconvex open-loop problems as well.

Linear discrete-time minimax control problems with perfect state information were considered first by Witsenhausen[W1], [W2] who developed the dynamic programming algorithm and its dual for the case of the cost functional $J(\mu_o, \mu_1, \ldots, \mu_{N-1}) = \sup_{\substack{w_k \in W_k \\ k=0,1,\ldots,N-1}} f_N(x_N)$. He considered in detail the implementation of the dual algorithm for this case and gave a necessary condition which parallels Proposition 2.4 of this chapter. He also observed that a minimax principle in general does not hold due to the presence of singular points.

The dynamic programming algorithm of this chapter can be extended to much more general minimax control problems as will be seen in Chapter 5. All the other results of this chapter rely on linearity and convexity. Their extension however to continuous-time linear systems appears to involve great technical difficulties.

CHAPTER 3

REACHABILITY OF A TARGET TUBE WITH PERFECT STATE INFORMATION

1. General Remarks

In this chapter we consider a special case of the problem of the previous chapter which will be referred to as the problem of the Reachability of a Target Tube by the state of the system when the controller has available at each time a perfect measurement of the system state. The motivation for considering this problem arises from two basic problems of deterministic control theory, the controllability problem, and the tracking (servomechanism) problem. The controllability problem is concerned with transfering the state of a system from an initial state-time pair to a final state-time pair. The tracking problem is concerned with keeping the state trajectory of the system "sufficiently close" to a prescribed trajectory.

The problems considered in this chapter can be viewed as the analogs of these two problems when there are disturbances driving the system. In accordance with the general approach of this thesis we assume that these disturbances are unknown except for the fact that they belong to given sets. Under these circumstances, the most natural analog of the deterministic controllability problem is that of steering the system state at the final time into a desired target set under all possible combinations of disturbances. In other words, we would like to design a feedback controller in such a way as to guarantee that the final state of the system will always lie in a prescribed target set despite the presence of uncertainties. In a similar vein, a natural analog of the

tracking problem under these same conditions is to keep the entire state trajectory in a "tube" containing the desired trajectory under all possible disturbances. We refer to these two problems as those of "Reachability of a Target Set" and "Reachability of a Target Tube". Possible applications of these two problems can be expected in the control of systems under uncertainty when either a set-membership description of the uncertain quantities is more readily available than a probabilistic one, or where specified tolerances must be met with certainty.

In the next section we formulate the problem of Reachability of a Target Tube which involves a linear discrete-time system. The problem of Reachability of a Target Set under the same circumstances can be viewed as a special case of the problem of Reachability of a Target Tube and will not be considered explicitly. The solution of the problem by dynamic programming will be given in Section 3 by making direct use of the results obtained in Chapter 2. In Sections 4 and 5 we shift the emphasis to the development of algorithms which have potential for practical applications. We consider the case where all the given sets are ellipsoids in the appropriate Euclidean spaces and we develop an ellipsoidal approximation algorithm which results in a control law which is a linear function of the system state, thus offering attractive implementation advantages.

2.   Problem Formulation

We will consider the following problem:

Problem 3.1: Given is the linear discrete-time dynamic system:

$$x_{k+1} = A_k x_k + B_k u_k + G_k w_k, \quad k = 0, 1, \ldots, N-1 \tag{3.1}$$

where $x_k \in R^n$, $k = 0, 1, .., N$, are the state vectors, $u_k \in R^m$, $k = 0, 1, .., N-1$, are the control vectors, $w_k \in R^r$, $k = 0, .., N-1$, are disturbance vectors, and A_k, B_k, G_k, $k = 0, 1, .., N-1$ are given matrices of appropriate dimension.

The initial state x_o is known and the disturbance vectors w_k belong to given compact sets $W_k \subset R^r$, $w_k \in W_k$, $k = 0, 1, ..., N-1$.

Attention is restricted to control laws of the form

$$\mu_k : R^n \rightarrow U_k, \qquad k = 0, 1, ..., N-1$$

taking values

$$u_k = \mu_k(x_k), \qquad k = 0, 1, ..., N-1$$

where $U_k \subset R^m$, $k = 0, 1, .., N-1$, are given closed convex sets. It is required to find (if it exists) a control law in this class such that for all k the state x_{k+1} of the closed-loop system

$$x_{k+1} = A_k x_k + B_k \mu_k(x_k) + G_k w_k \qquad (3.2)$$

is contained in given closed convex sets X_{k+1}, $k = 0, 1, .., N-1$, for all possible values of the disturbance vectors w_k.

We shall say that the target tube $\{X_1, X_2, ..., X_N\}$ is <u>reachable</u> if there exists such a control law.

It is easy to see that the Problem 3.1 is a special case of the Problem 2.1 of the previous chapter with the cost functional defined as

$$J(\mu_o, \mu_1, ..., \mu_{N-1}) = \sup_{\substack{w_k \in W_k \\ k = 0, 1, .., N-1}} \sum_{k=1}^{N} \{\delta(x_k | X_k) + \delta[\mu_{k-1}(x_{k-1}) | U_{k-1}]\} \qquad (3.3)$$

where $\delta(x | X)$ denotes the indicator function of the set X $\left(\delta(x|X) = 0 \text{ if } x \in X, \delta(x|X) = \infty \text{ if } x \notin X\right)$

With this definition the target tube $\{X_1, X_2, ..., X_N\}$ is reachable if the optimal value $\bar{J}_{x_o}$ of the cost functional (3.3) is 0. It is not reachable if $\bar{J}_{x_o} = \infty$.

It should be noted that the problem of this section has also been considered in a somewhat more general form in (Bl). The approach used in this reference is purely geometrical and does not rely on the solution of the Problem 2.1.

3. The Dynamic Programming Algorithm

Application of the dynamic programming algorithm of Proposition 2.1 of the previous chapter yields the optimal cost

$$\bar{J}_{x_0} = J_0(x_0) = \delta(x_0 | X_0^*) \tag{3.4}$$

from the recursive equations

$$E_{k+1}(x) = \delta(x | T_{k+1}), \quad k = 0, 2, \ldots, N-1$$

$$J_k(x_k) = \delta(x_k | X_k^*), \quad k = 0, 1, \ldots, N$$

where the sets T_k and X_k^* are given by the relations

$$X_N^* = X_N \tag{3.5}$$

$$T_{k+1} = \{x | (x + G_k W_k) \subset X_{k+1}^* \}, \quad k = 0, 1, \ldots, N-1 \tag{3.6}$$

$$X_k^* = \{x_k | (A_k x_k + B_k u_k) \epsilon T_{k+1}, \text{ for some } u_k \epsilon U_k \} \cap X_k, \quad k = 1, 2, \ldots, N-1 \tag{3.7}$$

$$X_0^* = \{x_0 | (A_0 x_0 + B_k u_k) \epsilon T_1, \text{ for some } u_0 \epsilon U_0 \} \tag{3.8}$$

If $x_0 \epsilon X_0^*$, by equation (3.4), the optimal cost is 0 implying the existence of a control law that achieves reachability of the target tube $\{X_1, X_2, \ldots, X_N\}$. If $x_0 \notin X_0^*$ then the target tube is not reachable.

Some of the properties of the sets T_k and X_k^* of equations (3.5) through

(3.8) can be obtained by making use of propositions derived earlier in Chapter 2. Thus by Proposition 2.2 these sets are convex whenever nonempty, and if the sets U_k are compact they are also closed. If in addition the matrices A_k, $k = 0, 1, \ldots, N-1$, are invertible then it can be proved that the sets T_k and X_k^* are compact. Also since the support function and the indicator function of a closed convex set are conjugate to each other the support functions of the sets T_k and X_k^* can be obtained by making use of the equations (2.16), (2.17) and (2.18) of the previous chapter.

A control law that achieves reachability can be obtained as follows. To every $x_k \epsilon X_k^*$ associate a vector $\mu_k(x_k) = u_k \epsilon U_k$ such that

$$(A_k x_k + B_k u_k) \epsilon T_{k+1}$$

By definition of the set X_k^* such a vector exists. It can be seen that if the target tube $\{X_1, X_2, \ldots, X_N\}$ is reachable from the initial state of the system, then if we use a control law defined as above the state x_k will belong to the set X_k^* for all k, and thus definition of the control law for vectors outside the set X_k^* is redundant.

For purposes of future reference the tube $\{T_1, T_2, \ldots, T_N\}$ will be called _effective target tube_. The tube $\{X_o^*, X_1^*, \ldots, X_N^*\}$ will be called _modified target tube_ and in fact it specifies the region of state space where the state will lie when a control law that achieves reachability is used.

For practical applications it is important that the sets T_k and X_k^* of the effective and modified target tube can be characterized by a finite set of numbers. This is possible when the given sets X_k and U_k are convex polyhedra. The sets T_k and X_k^* are in this case polyhedra and thus can be characterized by a finite number of bounding hyperplanes. The corres-

ponding algorithm[B1] is however beset by the fact that the number of bounding hyperplanes increases at every step of the algorithm. In addition the implementation of a control law that achieves reachability can be quite cumbersome.

In the case where the given sets are not polyhedra, characterization of the sets T_k, X_k^* of the effective and modified target tubes by a finite set of numbers is in general infeasible. One can however conceive of constructing sets that internally approximate the sets T_k, X_k^* and which can be characterized by a finite set of numbers. One such possibility is to approximate the sets T_k and X_k^* for each k by ellipsoids $\overline{T}_k \subset T_k$, $\overline{X}_k^* \subset X_k^*$ since an ellipsoid is completely characterized by its center and a weighting matrix. Then in order for the original target tube $\{X_1, X_2, \ldots, X_N\}$ to be reachable from the initial state x_o, it is sufficient (but not necessary) that $x_o \epsilon \overline{X}_o^*$. This approximation approach is the basis for an ellipsoidal approximation algorithm given in the next section, where results on the optimal control of linear systems with quadratic cost criteria are used not only to obtain ellipsoidal approximating sets but also to derive control laws which are <u>linear</u>.

4.   An Ellipsoidal Approximation Algorithm

Consider the special case of Problem 3.1 for which the constraint sets are the ellipsoids described by:

$$X_k = \{x_k | x_k' C_k' C_k x_k \leq 1\}, \quad k = 1, 2, \ldots, N-1 \tag{3.9}$$

$$X_N = \{x_N | x_N' \Psi x_N \leq 1\} \tag{3.10}$$

$$U_k = \{u_k | u_k' R_k u_k \leq 1\}, \quad k = 0, 1, \ldots, N-1 \tag{3.11}$$

$$W_k = \{w_k | w_k' Q_k w_k \leq 1\}, \quad k = 0, 1, \ldots, N-1 \tag{3.12}$$

where the matrices Ψ, R_k, Q_k are given positive definite symmetric matrices for all $k = 0, 1, \ldots, N-1$, and the matrices C_k are given. We also assume that the matrices A_k in the system (3.1) are invertible.[*]

We first internally approximate the set T_N of equation (3.6) by an ellipsoid. To this end, we state the following lemma the proof of which can be found in (S1).

Lemma 3.1: Consider two ellipsoids E_1, E_2 in R^n with support functions $\sigma(x|E_1) = (x'Q_1 x)^{1/2}$, $\sigma(x|E_2) = (x'Q_2 x)^{1/2}$. Their vector sum $E_1 + E_2$ is contained in the ellipsoid E with support function

$$\sigma(x|E) = \{x'[\beta^{-1}Q_1 + (1-\beta)^{-1}Q_2]\}^{1/2}$$

where β is a free scalar parameter with $0 < \beta < 1$.

For ellipsoid $\overline{T}_N$ to be contained in the set T_N of equation (3.6) it is sufficient that $\overline{T}_N + G_{N-1}W_{N-1} \subset X_N$. The support functions of the ellipsoids $G_{N-1}W_{N-1}$ and X_N for the case considered in this section are, $\sigma(x|G_{N-1}W_{N-1}) = (x'G_{N-1}Q_{N-1}^{-1}G_{N-1}'x)^{1/2}$ and $\sigma(x|X_N) = (x'\Psi^{-1}x)^{1/2}$. By Lemma 3.1 if follows that the relation $\overline{T}_N + G_{N-1}W_{N-1} \subset X_N$ is satisfied if the support function of $\overline{T}_N$ is given by

---

[*] Notice that if the discrete-time system (3.1) results from sampling of a continuous-time linear system the matrices A_k will always be invertible. However, it is easy to see that in what follows invertibility of the matrices A_k is not necessary if the matrices $C_k'C_k$ are positive definite for all k.

$$\sigma(x|\overline{T}_N) = (x'F_N^{-1}x)^{1/2}$$

where the matrix F_N^{-1} is given by

$$F_N^{-1} = (1-\beta_N)(V^{-1} - \beta_N^{-1}G_{N-1}Q_{N-1}^{-1}G_{N-1}') \qquad (3.13)$$

and β_N is a free parameter such that $0 < \beta_N < 1$. If the given constraint sets are such that the set T_N has a nonempty interior, there exists a scalar β_N with $0 < \beta_N < 1$ such that the matrix F_N of equation (3.13) is positive definite and the ellipsoid

$$\overline{T}_N = \{x|x'F_N x \leq 1\}$$

is contained in the set T_N.

By using the ellipsoid $\overline{T}_N$ a set contained in the set X_{N-1}^* of equation (3.7) is now defined as the set of points x_{N-1} with the property that both

$$x_{N-1}'C_{N-1}'C_{N-1}x_{N-1} \leq 1 \qquad (3.14)$$

and

$$(A_{N-1}x_{N-1} + B_{N-1}u_{N-1}) \in \overline{T}_N \text{ for some } u_{N-1} \in U_{N-1}.$$

The second requirement becomes in this case that

$$x'F_N x \leq 1 \text{ for some } u_{N-1} \text{ with } u_{N-1}'R_{N-1}u_{N-1} \leq 1 \qquad (3.15)$$

and with

$$x = A_{N-1}x_{N-1} + B_{N-1}u_{N-1} \qquad (3.16)$$

The set of points x_{N-1} satisfying relations (3.14), (3.15) and (3.16) clearly contains the set of points x_{N-1} with the property that for some $u_{N-1} \in R^m$

$$x'_{N-1} C'_{N-1} C_{N-1} x_{N-1} + u'_{N-1} R_{N-1} u_{N-1} + x' F_N x \leq 1 \qquad (3.17)$$

$$x = A_{N-1} x_{N-1} + B_{N-1} u_{N-1} \qquad (3.18)$$

By well known results on the linear quadratic regulator problem of optimal control[K1] the set of points x_{N-1} satisfying the equations (3.17) and (3.18) for some $u_{N-1} \epsilon R^n$ is given by

$$\overline{X}^*_{N-1} = \{x_{N-1} | x'_{N-1} K_{N-1} x_{N-1} \leq 1\} \qquad (3.19)$$

where the positive definite matrix K_{N-1} is given by the discrete Riccati equation

$$K_{N-1} = A'_{N-1} (F_N^{-1} + B_{N-1} R_{N-1}^{-1} B'_{N-1})^{-1} A_{N-1} + C'_{N-1} C_{N-1} \qquad (3.20)$$

Furthermore a control law which achieves reachability is given by

$$u_{N-1} = \mu_{N-1}(x_{N-1}) = -(R_{N-1} + B'_{N-1} F_N B_{N-1})^{-1} B'_{N-1} F_N A_{N-1} x_{N-1} \qquad (3.21)$$

By proceeding with similar approximations we define sets $\overline{T}_{N-1}, \overline{X}^*_{N-2}, \ldots, \overline{T}_1, \overline{X}^*_0$. If some ellipsoid $\overline{T}_k$ is empty, then the algorithm breaks down. This of course does not imply that the original target tube is not reachable, since approximations were involved in obtaining $\overline{T}_k$. In this case if we wish to proceed with the ellipsoidal algorithm we will have to start with a "larger" target tube or "larger" control constraint sets. We summarize the algorithm below:

An internally approximate modified target tube $\{\overline{X}^*_0, \overline{X}^*_1, \ldots, \overline{X}^*_N\}$ and effective target tube $\{\overline{T}_1, \overline{T}_2, \ldots, \overline{T}_N\}$ are given recursively by the equations:

$$\overline{X}^*_k = \{x_k | x'_k K_k x_k \leq 1\}, \quad k = 1, 2, \ldots, N \qquad (3.22)$$

$$\overline{T}_k = \{x | x'F_k x \leq 1\}, \quad k = 1, 2, \ldots, N \tag{3.23}$$

where

$$F_k^{-1} = (1 - \beta_k)(K_k^{-1} - \beta_k^{-1} G_{k-1} Q_{k-1}^{-1} G_{k-1}') \tag{3.24}$$

$$K_{k-1} = A_{k-1}'(F_k^{-1} + B_{k-1} R_{k-1}^{-1} B_{k-1}')^{-1} A_{k-1} + C_{k-1}' C_{k-1} \tag{3.25}$$

$$K_N = \Psi \tag{3.26}$$

and the free parameters β_k, $k = 1, 2, \ldots, N$, are such that $0 < \beta_k < 1$ and the matrices F_k are positive definite for all k.

A sufficient condition for reachability is that the set

$$\overline{X}_o^* = \{x_o | x_o' K_o x_o \leq 1\} \tag{3.27}$$

contains the initial state x_o, where

$$K_o = A_o'(F_1^{-1} + B_o' R_o^{-1} B_o)^{-1} A_o \tag{3.28}$$

Furthermore a control law that achieves reachability is given by the equation

$$\mu_k(x_k) = -(R_k + B_k' F_{k+1} B_k)^{-1} B_k' F_{k+1} A_k x_k, \quad k = 0, 1, \ldots, N-1 \tag{3.29}$$

We remark that another control law that achieves reachability is the control law with a dead zone given by equation (3.29) if $x_k' A_k' F_{k+1} A_k x_k > 1$ (i.e., if $A_k x_k \notin \overline{T}_{k+1}$), and $\mu_k(x_k) = 0$ otherwise. In certain applications the use of a dead zone can be particularly beneficial.

It should be mentioned that a similar ellipsoidal algorithm can be obtained for the case where the given sets X_{k+1}, U_k, W_k, $k = 0, 1, \ldots, N-1$, are ellipsoids which are centered at given points other than the origin.

The case where the system (3.1) is time-invariant, the given constraint sets are constant and the final time N approaches infinity is highly interesting. The behaviour and the convergence properties of the ellipsoidal algorithm under these circumstances will be examined in the next section.

5.   Infinite Time Behaviour of the Ellipsoidal Algorithm

Consider now the case where the system (3.1) is time-invariant

$$x_{k+1} = Ax_k + Bu_k + Gw_k \tag{3.30}$$

the given constraint sets are constant, i.e., $R_k = R$, $Q_k = Q$, $C_k'C_k = \Psi$, for all k, and the final time N approaches infinity.

The ellipsoidal algorithm of equations (3.22) through (3.26) under these circumstances, and assuming a constant scalar β with $0 < \beta < 1$, i.e., $\beta_k = \beta$ for all k, is given by the equations:

$$\overline{X}_k^* = \{x_k | x_k' K_k x_k \leq 1\} \tag{3.31}$$

$$\overline{T}_k = \{x | x' F_k x \leq 1\} \tag{3.32}$$

where

$$F_k^{-1} = (1 - \beta)(K_k^{-1} - \beta^{-1} GQ^{-1} G') \tag{3.33}$$

$$K_{k-1} = A'(F_k^{-1} + BR^{-1}B')^{-1}A + \Psi \tag{3.34}$$

$$K_N = \Psi \tag{3.35}$$

under the assumption that the matrix F_k^{-1} is positive definite for all k.

Assume that for some scalar β with $0 < \beta < 1$ the algorithm of equation (3.31) through (3.35) possesses a positive definite steady state

solution $K_{-\infty}$ given by

$$K_{-\infty} = A'(F_{-\infty}^{-1} + BR^{-1}B')^{-1}A + \Psi \qquad (3.36)$$

for which the matrix

$$F_{-\infty}^{-1} = (1 - \beta)(K_{-\infty}^{-1} - \beta^{-1}GQ^{-1}G') \qquad (3.37)$$

is positive definite. Then if the initial state belongs to the set $\overline{X}^* = \{x \mid x'K_{-\infty}x \leq 1\}$ the state of the system (3.30) can be made to stay indefinitely in the tube $\{\overline{X}^*, \overline{X}^*, \ldots\}$ by application of the linear time invariant control law

$$\mu(x) = -(R + B'F_{-\infty}B)^{-1}BF_{-\infty}Ax \qquad (3.38)$$

Since we will have $\overline{X}^* \subset X = \{x \mid x'\Psi x \leq 1\}$ infinite time reachability of the given target tube $\{X, X, \ldots\}$ is achieved.

It should be noted that in the actual operation of the closed-loop system the initially given tube $\{X, X, \ldots\}$ loses its significance since the system state will always remain in the internal tube $\{\overline{X}^*, \overline{X}^*, \ldots\}$ the sets $\overline{X}^*$ of which will differ significantly from the sets X of the initial tube both in "size" and "orientation". Thus in any infinite time design procedure the given set X and the corresponding matrix Ψ take the role of a design parameter which can be adjusted to obtain different steady state solutions $K_{-\infty}$ of the algorithm.

A question of importance is under what circumstances the algorithm of equations (3.33) through (3.35) will converge to a steady state solution $K_{-\infty}$ satisfying the equations (3.36) and (3.37). Clearly given the system (3.30) and the matrix Q specifying the disturbance constraint set W, the

matrices R and Ψ must specify a sufficiently "large" control constraint set and a sufficiently "large" target tube relative to the size of the disturbance set and the nature of the matrices A, B and G of the system. Thus if the matrices R and Ψ specify relatively small constraint sets the algorithm of equations (3.38) through (3.35) should not be expected to converge to a steady state and guarantee reachability from some initial states. Now in any practical situation the designer is given the system (3.30) and the matrix Q specifying the set W where the input disturbance belongs, and usually there is a certain degree of freedom in adjusting the control constraints, and particularly the matrix Ψ specifying the target tube which in view of the comment of the previous paragraph plays the role of a design parameter. In this sense a possible design procedure is to initially select the matrices R and Ψ and in case the algorithm does not converge to a steady state solution, to decrease these matrices by multiplication by factors less than one and repeat the procedure until convergence and satisfaction of the designer. It is important however to know under what circumstances there exist matrices R and Ψ such that the algorithm converges to a steady state, and furthermore under what conditions such matrices can be obtained by repeatedly multiplying any initially selected matrices R_1 and Ψ_1 by factors of less than one. This is the object of the next proposition which states that the design procedure outlined above is successful provided the system (3.30) is stabilizable, i.e., if there exists a matrix L such that the matrix (A - BL) is stable (has eigenvalues within the unit disk of the complex plane). Notice that the system (3.30) is stabilizable provided the pair (A, B) is controllable (but not conversely)[Wol].

Proposition 3.1: Assume that the system (3.30) and the positive definite symmetric matrix Q are given and that the system (3.30) is stabilizable. Then given any positive definite symmetric matrices Ψ_1 and R_1 of appropriate dimension, there exists a scalar β_1, $0 < \beta_1 < 1$ such that for every scalar β, $0 < \beta \leq \beta_1$ there exist scalars a_1, b_1 depending on β such that for all matrices $\Psi = a\Psi_1$, $R = bR_1$ with $0 < a \leq a_1$, $0 < b \leq b_1$, the algorithm of equations (3.33) through (3.35) converges to a positive definite symmetric steady state solution $K_{-\infty}$ satisfying equations (3.36) and (3.37).

The proof of the above proposition follows similar, yet a little more complicated, arguments with a proof of convergence of usual Riccati equations to a steady state solution[Wo2]. Due to its length this proof will be presented in Appendix II.

Another important question concerning the infinite time ellipsoidal algorithm is whether the resulting linear time-invariant control law makes the closed-loop system asymptotically stable. This question is answered in the affirmative in the following proposition.

Proposition 3.2: Assume that the algorithm of equations (3.33) through (3.35) converges to a steady state solution $K_{-\infty}$, where $K_{-\infty}$ is a positive definite symmetric matrix for which the matrix $F_{-\infty}$ of equation (3.37) is also positive definite. Then the closed-loop system resulting from application of the linear time-invariant control law of equation (3.38) is asymptotically stable.

The proof of the above proposition will also be given in Appendix II.

An immediate consequence of the above proposition is that transients due to initial states will vanish eventually during the operation of the closed-loop system. More accurately for any $\epsilon > 0$ it can be guaranteed that after a sufficient number of steps the state will be confined in the set $\overline{X}^* + \epsilon B$, where B is the unit ball in R^n, and this will occur for any initial state x_o in R^n.

6.   <u>Discussion and Sources</u>

The problem of the reachability of a target tube was examined in this chapter with emphasis in the development of an ellipsoidal approximation algorithm that appears to have potential for practical applications.

The attractive feature of the ellipsoidal algorithm is that it provides a linear control law which in the infinite time case makes the closed-loop system asymptotically stable. Furthermore for the infinite-time case the existence result of Proposition 3.1 guarantees that the algorithm is applicable to every linear time-invariant system which is stabilizable. Thus the ellipsoidal algorithm appears to offer practical advantages as a design method for many regulation and tracking problems which involve a linear system, and for which the statistics of the uncertain quantities are unknown and difficult to measure, or for which specified tolerances must be met with certainty.

A number of questions concerning the performance of the algorithm remain as yet unresolved. One such question concerns the quality of the approximation involved in the algorithm. If appears to be very difficult to obtain precise estimates of the approximation which are applicable to large classes of systems. Thus some further research and simulations

are required in this area. Another question, and in the author's opinion the most important, touches upon the merits of the whole minimax design philosophy. Minimax designs are in general concervative, optimal against the worst case. In the particular case of the ellipsoidal algorithm the result is that the feedback gains of the controller tend to be large in magnitude, a feature which in some cases may be undesirable. Furthermore this situation is adversely effected by the approximations involved. Only simulations and practical experience can give some answers to this question.

Many of the results of this chapter have been reported in (B1). The approach used in this reference is purely geometrical and is applicable to a large class of problems. It not required that the system is linear and that the given sets are closed, convex or compact. In fact not even the linear vector space structure of the space of definition of the system is necessary. However the ellipsoidal algorithm is applicable only to the class of problems considered in this chapter. The Propositions 3.1 and 3.2 have not appeared earlier. It is interesting to note that the equations of the ellipsoidal algorithm are very similar to Riccati equations related to linear multistage games with quadratic cost functional[Rh1]. In fact the proofs of Propositions 3.1 and 3.2 were to a large extent motivated by this similarity.

The problem of the reachability of a target set is the special case of the Problem 3.1 where the sets X_k for all k except k = N are equal to the whole space R^n. This problem for a linear discrete-time system, closed-loop control and perfect state information was first considered by Witsenhausen[W1],[W2] in the framework of a more general minimax

control problem. The same problem for a continuous-time system but an open-loop controller was also considered by Delfour and Mitter in (Del).

Problems related to those of this chapter that require attention are the case of a nonlinear system and the case of a continuous, linear or nonlinear, system. The results in (Bl) are applicable to nonlinear discrete-time systems however no practical algorithms applicable to nontrivial systems are available at this moment. The case of a continuous-time linear system is considerably more difficult to handle than the case of a discrete-time system. Some results obtained by the author in this area are not as yet conclusive.

Finally we note that the problem of the reachability of a target tube with imperfect state information, including the case where instead of the entire state only a linear output of the system is measured exactly, will be considered in Chapter 6.

CHAPTER 4

STATE ESTIMATION PROBLEMS FOR A SET
DESCRIPTION OF THE UNCERTAINTY

1. ### General Remarks

In this chapter we digress from minimax control problems in order to consider some state estimation problems which involve a set-membership description of the uncertainty. Such problems, though important in their own right, are essential for the solution of minimax control problems with imperfect state information. Although the concepts to be presented are applicable to much more general situations we will be concerned exclusively with the case of a linear discrete-time dynamic system

$$x_{k+1} = A_k x_k + B_k w_k , \qquad k = 0, 1, \ldots, N-1 \qquad (4.1)$$

to which there are available noise-corrupted measurements

$$z_k = C_k x_k + v_k \qquad (4.2)$$

where $x_k \in R^n$ is the system state, $w_k \in R^r$ is an input disturbance vector and $v_k \in R^p$ is the measurement noise vector. We assume that there is no control input to the system. The algorithms that we derive however can be trivially modified to take into account the effect of any known deterministic input by virtue of the linearity of the system and measurement equations.

In a stochastic estimation problem involving the system (4.1) with the measurements (4.2) the uncertain quantities, i.e., the initial state and the input and measurement noise vectors, are modelled as mutually independent random vectors with known probability density functions. In this case all information about the system state at any time that is provided by the measurements is contained in the probability density function of the state

conditioned on these measurements. This conditional probability density function is then used, explicitly or implicitly, to determine an estimate of the system state which is best in some prescribed sence.

In the case considered here, the uncertain quantities are not modelled as random vectors, but are considered instead to be unknown except that they belong to given subsets of appropriate vector spaces. Under these circumstances, all information about the system state at any time that is provided by the measurements may be summarized in the set of all states consistent with both the measurements received and the constraints on the uncertain quantities. Once this set of possible states is characterized a point estimate can be selected using some criterion such as the minimax error criterion for example. In what follows however we will be concerned exclusively with the characterization of the set of possible states or some approximation thereof. Since for the special cases that we will consider this set will be an ellipsoid, if a point estimate is desired the center of the ellipsoid is the natural candidate.

Two distinct types of constraints on the uncertain quantities will be considered. The first is the energy-type constraint

$$x_o' \Psi^{-1} x_o + \sum_{k=1}^{N} (w_{k-1}' Q_{k-1}^{-1} w_{k-1} + v_k' R_k^{-1} v_k) \leq 1$$

where Ψ, Q_k, R_k are given positive definite symmetric matrices for all k. For this constraint we show that the set of possible states at any time consistent with the output measurements is an ellipsoid whose center and weighting matrix are generated by equations identical to those associated with the best linear estimator (Kalman filter) for a certain stochastic esti-

mation problem. We shall demonstrate a one-one correspondence be-
tween estimation problems where the uncertain quantities satisfy an energy
constraint and linear minimum variance stochastic estimation problems.
Once this correspondence is established we will be able to use available
results in stochastic estimation theory to derive estimators for the energy
constraint case for a number of problems of interest including the filtering,
prediction and smoothing problems.

The second type of constraint that we consider is the more practically
important case where the uncertain quantities are constrained at each instant
of time to lie in ellipsoids, i.e.,

$$x_o' \Psi^{-1} x_o \leq 1, \quad w_{k-1}' Q_{k-1}^{-1} w_{k-1} \leq 1, \quad v_k' R_k^{-1} v_k \leq 1, \quad k = 1, 2, \ldots, N$$

In this case the set of states consistent with the measurements is not an
ellipsoid and it is not, in general, characterized by a finite set of numbers.
However, an ellipsoid bound to it can be determined by bounding the instant-
aneous constraints by an energy constraint and using the results derived for
that constraint. The resulting estimator for the case of the filtering and the
prediction problem is similar to that proposed by Schweppe[S1] but it has
two important advantages: first the gain matrices do not depend on the
particular measurements received and are therefore precomputable and,
second, it reduces to a constant system as the final time becomes infinite.
In all other respects it is comparable to that proposed by Schweppe. Further-
more our approach permits the derivation of an estimator for the smoothing
problem which has not been previously considered in the literature.

2. Formulation of the Problem with an Energy Constraint

In this section we formulate a general estimation problem involving a linear discrete-time dynamic system and a combined energy constraint on the uncertain quantities. This problem includes as special cases the filtering, prediction, and smoothing problems.

Problem 4.1: Consider the linear discrete-time dynamic system

$$x_{k+1} = A_k x_k + B_k w_k , \qquad k = 0, 1, \ldots, N-1$$

to which there are available noise-corrupted measurements

$$z_k = C_k x_k + v_k$$

where $x_k \epsilon R^n$ is the system state, $w_k \epsilon R^r$ is the input disturbance vector, $v_k \epsilon R^p$ is the measurement noise vector, and the matrices A_k, B_k, C_k have the appropriate dimensions. The initial state x_o and the disturbances w_k, v_k are assumed unknown except that they satisfy the energy constraint

$$x_o' \Psi^{-1} x_o + \sum_{k=1}^{N} (w_{k-1}' Q_{k-1}^{-1} w_{k-1} + v_k' R_k^{-1} v_k) \leq 1 \qquad (4.3)$$

where Ψ, Q_{k-1}, R_k, $k = 1, 2, \ldots, N$, are given positive definite symmetric matrices. Let i, k be arbitrary integers, $0 \leq i \leq N$, $0 \leq k \leq N$. It is required to find the set $X_{i|k}$ of possible system states x_i at time i which are consistent with the constraint (4.3) and the measurements $z_1, z_2, \ldots, z_k$ up to time k.

We remark that if i = k this problem is usually called the filtering problem, if i > k it is called the prediction problem, and if i < k it is called the smoothing problem.

In the next section we will obtain a general solution to the above problem by associating it with a stochastic estimation problem, the solution of which is well known. We will then use this general solution to obtain estimators for the special cases of filtering, prediction, and smoothing.

3. A General Solution to the Problem with an Energy Constraint

Given any estimation problem where the uncertain quantities are unknown except that they lie in some given set it is possible to give a precise characterization of the set $X_{i|k}$ of possible states x_i at time i consistent with the measurements $z_1, z_2, \ldots, z_k$ in terms of the given constraint set and the system and measurement equations. This characterization is usually quite elaborate but for the Problem 4.1 it takes a particularly useful form. A great deal of insight can be obtained through it, and most importantly it leads to a direct correspondence with linear minimum variance stochastic estimation problems. We will first introduce some notation.

Let $u \in R^{n+N(r+p)}$ be the vector consisting of all the uncertain quantities according to the relation

$$u = (x_o', w_o', w_1', \ldots, w_{N-1}', v_1', v_2', \ldots, v_N')' \tag{4.4}$$

Let us also combine all measurements received up to time k into one vector

$$\zeta_k = (z_1', z_2', \ldots, z_k')' \tag{4.5}$$

Both the state of the system x_i at any time i, and the vector ζ_k can be obtained from the vector u of equation (4.4) by a linear transformation

$$x_i = L_i u \tag{4.6}$$

$$\zeta_k = D_k u \tag{4.7}$$

where the $n \times [n + N(r+p)]$ matrix L_i and the $kp \times [n + N(r+p)]$ matrix D_k are given for all i and k by

$$L_i = [\Phi(i, 0), \Phi(i, 1)B_o, \ldots, \Phi(i, i-1)B_{i-2}, B_{i-1}, 0, \ldots, 0] \tag{4.8}$$

$$D_k = \begin{bmatrix} C_1\Phi(1, 0), & C_1B_o, & 0, 0, \ldots, 0 & 0, \ldots, 0 \; I, 0, \ldots, 0 \; 0, \ldots, 0 \\ C_2\Phi(2, 0), & C_2\Phi(2, 1)B_o, & C_2B_1, 0, \ldots, 0 & 0, \ldots, 0 \; 0, I, 0, \ldots, 0 \; 0, \ldots, 0 \\ \hline C_k\Phi(k, 0) & C_k\Phi(k, 1)B_o, & C_k\Phi(k, 2)B_1, \ldots, C_kB_{k-1}, \; 0, \ldots, 0 \; 0, \ldots, 0, I \; 0, \ldots, 0 \end{bmatrix} \tag{4.9}$$

where the transition matrix $\Phi(i, j)$ is given by

$$\Phi(i, j) = A_{i-1}A_{i-2}\cdots A_j \qquad \text{for } j < i$$

$$\Phi(i, i) = I$$

and where the dimensions of the zero and identity matrices in the above equations are consistent with the multiplications indicated in equations (4.6) and (4.7).

The energy constraint (4.3) implies that the vector u of equation (4.4) belongs to the ellipsoid

$$U = \{u \mid u'M^{-1}u \leq 1\} \tag{4.10}$$

where the positive definite matrix M is defined as

$$M = \begin{bmatrix} \Psi_{Q_0} & & & & & 0 \\ & Q_1 & & & & \\ & & Q_{N-1} & & & \\ & & & R_1 & & \\ & & & & R_2 & \\ 0 & & & & & R_N \end{bmatrix} \tag{4.11}$$

Now, for fixed measurements ζ_k, the set $X_{i|k}$ which solves the Problem 4.1 can be conveniently characterized as

$$X_{i|k} = \{x \mid x = L_i u, \ \zeta_k = D_k u, \ u \in U\} \tag{4.12}$$

By defining the set $\hat{U}_k$ of all possible vectors u consistent with the measurement vector ζ_k

$$\hat{U}_k = \{u \mid \zeta_k = D_k u, \ u \in U\} \tag{4.13}$$

we have from equation (4.12) that the set $X_{i|k}$ is given by the equation

$$X_{i|k} = L_i \hat{U}_k \tag{4.14}$$

Thus the set $X_{i|k}$ can be obtained by a linear transformation on the set $\hat{U}_k$ which is the set intersection of the set U with the linear variety (manifold) $\{u \mid \zeta_k = D_k u\}$ defined by the measurements. Since the set U is an ellipsoid in the space $R^{n+N(r+p)}$ the set intersection $\hat{U}_k$ is also an ellipsoid and the set $X_{i|k}$ is also an ellipsoid since by equation (4.14) it is obtained through a linear transformation on an ellipsoid. We proceed to characterize the center and weighting matrix of the ellipsoid $X_{i|k}$ in the following proposition.

<u>Proposition 4.1</u>: The ellipsoid $X_{i|k}$ which solves Problem 4.1 is given for all i, k, $0 \leq i \leq N$, $0 \leq k \leq N$ by

$$X_{i|k} = \{x \mid (x - \hat{x}_{i|k})' \Sigma_{i|k}^{-1} (x - \hat{x}_{i|k}) \leq 1 - \delta^2(k)\} \tag{4.15}$$

where the matrix $\Sigma_{i|k}$ is given by

$$\Sigma_{i|k} = L_i [M - M D_k' (D_k M D_k')^{-1} D_k M] L_i' \tag{4.16}$$

the n-vector $\hat{x}_{i|k}$ is given by

$$\hat{\Sigma}_{i|k} = L_i MD_k'(D_k MD_k')^{-1} \zeta_k \qquad (4.17)$$

and the nonnegative scalar $\delta^2(k)$ is given by

$$\delta^2(k) = \zeta_k'(D_k MD_k')^{-1} \zeta_k \qquad (4.18)$$

In the case where the matrix $\Sigma_{i|k}$ is only positive semidefinite but not positive definite the ellipsoid $X_{i|k}$ is characterized by its support function

$$\sigma(x^* | X_{i|k}) = <x^*, \hat{x}_{i|k}> + [1 - \delta^2(k)]^{1/2}(x^{*_!}\Sigma_{i|k}x^*)^{1/2} \qquad (4.19)$$

It should be mentioned that, as will be shown later, the matrix $\Sigma_{i|k}$ is invertible provided the matrices A_k in the system equation (4.1) are invertible.

Proof: Since the equation (4.19) implies the equation (4.15) whenever the matrix $\Sigma_{i|k}$ is invertible, it will be sufficient to prove

$$\sigma(x^* | X_{i|k}) = \sup_{x \in X_{i|k}} <x^*, x> = <x^*, \hat{x}_{i|k}> + [1 - \delta^2(k)]^{1/2}(x^{*_!}\Sigma_{i|k}x^*)^{1/2}$$

We will first characterize the support function of the ellipsoid $\hat{U}_k$ of equation (4.14)

$$\hat{U}_k = \{u | \zeta_k = D_k u, \, u'M^{-1}u \le 1\} \qquad (4.14)$$

Consider the space $R^{n+N(r+p)}$ with the norm

$$||u|| = (u'M^{-1}u)^{1/2}$$

With this norm the set $U = \{u | u'M^{-1}u \le 1\}$ becomes the unit ball in $R^{n+N(r+p)}$ and the set $\hat{U}_k$ of equation (4.14) is the intersection of the unit ball with the linear variety $\{u | \zeta_k = D_k u\}$. Let $\hat{u}_k$ be the (unique) vector of minimum norm on this linear variety given by the projection theorem[Lu1]

$$\hat{u}_k = MD_k'(D_k MD_k')^{-1} \zeta_k \qquad (4.20)$$

It can be seen from equation (4.9) that the matrix D_k has full rank and therefore the matrix $(D_k M D_k')$ is invertible thus justifying the notation used above.

The set $\hat{U}_k$ is now given by:

$$\hat{U}_k = \{u| \; \|u - \hat{u}_k\|^2 \leq 1 - \|\hat{u}_k\|^2 \;, \qquad \zeta_k = D_k u\}$$

and can be also characterized as

$$\hat{U}_k = \hat{u}_k + (1 - \|\hat{u}_k\|^2)^{1/2} \bar{U}_k$$

where $\bar{U}_k$ is the intersection of the unit ball with the nullspace $N(D_k)$ of the matrix D_k. From the above equation we have for the support function of $\hat{U}_k$

$$\sigma(u^* | \hat{U}_k) = <u^*, \hat{u}_k> + (1 - \|\hat{u}_k\|^2)^{1/2} \sup_{\substack{\|u\|^2 \leq 1 \\ u \in N(\bar{D}_k)}} <u^*, u> \qquad (4.21)$$

By using Theorem 5.8.1 in (Lul) we have

$$\sup_{\substack{\|u\|^2 < 1 \\ u \in N(\bar{D}_k)}} <u^*, u> = \|\hat{u}^*\|$$

where $\hat{u}^*$ is the projection of the vector u^* on the subspace $N(D_k)$. Using again the projection theorem we obtain

$$\|\hat{u}^*\| = \{u^{*'}[M - MD_k'(D_k M D_k')^{-1} D_k M] u^*\}^{1/2}$$

Using this relation in (4.21) we have

$$\sigma(u^* | \hat{U}_k) = <u^*, \hat{u}_k> + (1 - \|\hat{u}_k\|^2)^{1/2} \{u^{*'}[M - MD_k'(D_k M D_k')^{-1} D_k M] u^*\}^{1/2}$$

Using the fact that $X_{i|k} = L_i \hat{U}_k$ in the above equation, we have

$$\sigma(x^* | X_{i|k}) = \sigma(L_i' x^* | \hat{U}_k)$$

$$= <x^*, L_i \hat{u}_k> + (1 - ||\hat{u}_k||^2)^{1/2} \{x^{*'} L_i [M - MD_k'(D_k MD_k')^{-1} D_k M] L_i' x^*\}^{1/2}$$

Now from equations (4.16), (4.17), (4.18) and (4.20) we obtain

$$\sigma(x^* | X_{i|k}) = <x^*, \hat{x}_{i|k}> + [1 - \delta^2(k)]^{1/2} (x^{*'} \Sigma_{i|k} x^*)^{1/2}$$

which was to be proved. Q.E.D.

The equations (4.17) and (4.16) for the center $\hat{x}_{i|k}$ and the weighting matrix $\Sigma_{i|k}$ of the ellipsoid $X_{i|k}$ appear to quite formidable in view of the complicated expressions (4.8) and (4.9) for the matrices L_i and D_k. Yet we will be able to obtain efficient recursive algorithms for the computation of $\hat{x}_{i|k}$ and $\Sigma_{i|k}$ by associating the Problem 4.1 with the following linear minimum variance estimation problem.

Problem 4.1': Consider the Problem 4.1 where the vectors $x_0, w_0, w_1, \ldots, w_{N-1}$, $v_1, v_2, \ldots, v_N$ instead of satisfying the energy constraint (4.3), are independent random vectors with zero mean and covariances

$$E\{x_0, x_0'\} = \Psi, \quad E\{w_{i-1} w_{i-1}'\} = Q_{i-1}, \quad E\{v_i v_i'\} = R_i, \quad i = 1, 2, \ldots, N$$

Find the linear minimum variance estimate $\hat{x}_{i|k}$ of the system state x_i at time i given the measurements $z_1, z_2, \ldots, z_k$ and also find the covariance of the estimation error

$$\Sigma_{i|k} = E\{(x_i - \hat{x}_{i|k})(x_i - \hat{x}_{i|k})'\}$$

By using the relations (4.6) and (4.7) it can be seen that the above problem is equivalent to finding the linear minimum variance estimate

$\hat{x}_{i|k}$ of the vector x_i

$$x_i = L_i u \tag{4.6}$$

given the measurement

$$\zeta_k = D_k u \tag{4.7}$$

where u is a zero mean random vector with covariance M given by equation (4.11). The solution of this problem is well known and given in many sources (Lu1), (Br1). The estimate $\hat{x}_{i|k}$ is given by equation (4.17), i.e., by the same expression as the center of the ellipsoid $X_{i|k}$ in Proposition 4.1. The covariance matrix $\Sigma_{i|k}$ is given by equation (4.16), the same expression which gives the weighting matrix of the ellipsoid $X_{i|k}$ in Proposition 4.1. Thus there is a one-one correspondence between Problem 4.1 and the stochastic estimation Problem 4.1' which is reflected in identical expressions for the center $\hat{x}_{i|k}$ and weighting matrix $\Sigma_{i|k}$ of the ellipsoid $X_{i|k}$ on one hand, and the linear minimum variance estimate $\hat{x}_{i|k}$ and error covariance $\Sigma_{i|k}$ in Problem 4.1' on the other. Now from the well known results in stochastic estimation theory the estimate $\hat{x}_{i|k}$ and error covariance $\Sigma_{i|k}$ are computed by efficient recursive algorithms (Kalman estimators) which do not require storage of the measurements. The same algorithms are applicable and can be used for obtaining the center $\hat{x}_{i|k}$ and weighting matrix $\Sigma_{i|k}$ of the ellipsoid $X_{i|k}$-solution of Problem 4.1.

Concerning the scalar $\delta^2(k)$ of equation (4.18) the following recursive relation can be proved for all k

$$\delta^2(k) = \delta^2(k-1) + (z_k - C_k A_{k-1} \hat{x}_{k-1|k-1})'(C_k \Sigma_{k|k-1} C_k' + R_k)^{-1}(z_k - C_k A_{k-1} \hat{x}_{k-1|k-1}) \tag{4.22}$$

This relation can be used to calculate the scalar $\delta^2(k)$ recursively without requiring storage of all the measurements up to time k. The equation (4.22), the continuous counterpart of which has been proved in (B2), can be proved in a number of ways. One possible method is by direct manipulation from the equation (4.18) using the equations (4.8), (4.9) and (4.11). This method is straightforward but too lengthy and tedious to be profitably displayed here. Another method to prove the equation (4.22) is by considering the filtering case of Problem 4.1 and by casting it as an optimal tracking problem as was done in (B2). The equation (4.22) follows directly from the solution of this tracking problem.

The preceding discussions have demonstrated that the ellipsoid $X_{i|k}$-solution of Problem 4.1 can be characterized from Proposition 4.1 by using results of stochastic estimation theory (Kalman estimators) for the recursive computation of the center $\hat{x}_{i|k}$ and the weighting matrix $\Sigma_{i|k}$ and by using equation (4.22) for the computation of the scalar $\delta^2(k)$. In the next section we shall explicitly characterize the solution of the Problem 4.1 for the case of the filtering problem.

We finally note that the correspondence between the Problems 4.1 and 4.1' can be extended to some related problems not explicitly considered here. Such is the problem where there is no error in the measurement equation (4.2), i.e., $z_k = C_k x_k$. In this case the energy constraint (4.3) becomes

$$x_o' \Psi^{-1} x_o + \sum_{i=1}^{N} w_{i-1}' Q_{i-1}^{-1} w_{i-1} \leq 1$$

The Proposition 4.1 can be easily shown to hold with the matrices L_i, D_k, M

appropriately modified. Under these circumstances however the matrix D_k need not have full rank and consequently the matrix $(D_k M D_k')$ may not be invertible. In this case it can be proved similarly as in Proposition 4.1 that the center $\hat{x}_{i|k}$ of the ellipsoid $X_{i|k}$, the weighting matrix $\Sigma_{i|k}$ and the scalar $\delta^2(k)$ in equation (4.19) are given by

$$\Sigma_{i|k} = L_i (M - S_k D_k M) L_i'$$

where the matrix S_k is any solution of the equation

$$S_k D_k M D_k' = M D_k'$$

and

$$\hat{x}_{i|k} = L_i M D_k' y_k$$

$$\delta^2(k) = y_k' D_k M D_k' y_k$$

where the vector y_k is any solution of the equation

$$D_k M D_k' y_k = \zeta_k$$

The correspondence with a stochastic estimation problem similar to Problem 4.1' which involves no measurement noise can still be established and the results for this problem (T1), (T2) can be used for the solution of the estimation problem with an energy constraint but no measurement noise.

4. **Filtering for the Case of Energy Constraints**

In this section we will utilize the general solution of Problem 4.1 as given by Proposition 4.1 and the one-one correspondence with the stochastic estimation Problem 4.1' that was demonstrated in the previous section to write down explicitly and in recursive form the solution for the

filtering case. Entirely similar equations can be written for the prediction and smoothing cases. [B2], [Fr1], [Ra1] Although it is possible to write the solution for the general case [A1] for simplicity we will assume in this and subsequent sections that the matrices A_k in the system (4.1) are invertible for all k. This assumption will guarantee the existence of all the inverses that will appear in the expressions that follow.

<u>Proposition 4.2</u>: The solution of Problem 4.1 in the filtering case is the ellipsoid $X_{k|k}$ given for all k, $0 \le k \le N$, by the equation

$$X_{k|k} = \{x | (x - \hat{x}_k)' \Sigma_{k|k}^{-1} (x - \hat{x}_k) \le 1 - \delta^2(k)\} \qquad (4.23)$$

where the positive definite symmetric matrix $\Sigma_{k|k}$ is given recursively by the Riccati equation

$$\Sigma_{i|i} = (\Sigma_{i|i-1}^{-1} + C_i' R_i^{-1} C_i)^{-1} \qquad (4.24)$$

$$\Sigma_{i|i-1} = A_{i-1} \Sigma_{i-1|i-1} A_{i-1}' + B_{i-1} Q_{i-1} B_{i-1}' \qquad (4.25)$$

$$\Sigma_{o|o} = \Psi \qquad (4.26)$$

the vector $\hat{x}_k$ is the solution of the equation

$$\hat{x}_{i+1} = A_i \hat{x}_i + \Sigma_{i+1|i+1} C_{i+1}' R_{i+1}^{-1} (z_{i+1} - C_{i+1} A_i \hat{x}_i) \qquad (4.27)$$

$$\hat{x}_o = 0 \qquad (4.28)$$

and the nonnegative scalar $\delta^2(k)$ is given by the equation

$$\delta^2(k) = \sum_{i=1}^{k} (z_i - C_i A_{i-1} \hat{x}_{i-1})' (C_i \Sigma_{i|i-1} C_i' + R_i)^{-1} (z_i - C_i A_{i-1} \hat{x}_{i-1}) \qquad (4.29)$$

Proof: The proof follows directly from Proposition 4.1 for i = k, by utilizing the correspondence with the stochastic estimation Problem 4.1' demonstrated in the previous section, and by using also equation (4.22).

5.   Formulation of the Problem with Instantaneous Constraints

While the preceding sections show it to be of theoretical interest, the model for the uncertainty described by the energy constraint (4.3) is of limited use as far as practical applications are concerned. A situation which appears more often in practice is that in which the uncertain quantities are individually constrained at each point in time. In this section we formulate such a problem which is then solved in Sections 6 and 7 using the results of the preceding sections. In particular, we bound the instantaneous constraints by a single combined energy constraint and apply the results of Section 4. We concentrate our attention to the filtering case. Similar estimators can be derived for the prediction and smoothing problems by using the same approach. The resulting estimator is shown to be simpler but otherwise comparable to the one proposed by Schweppe[S1] with the additional advantage that it possesses a steady-state structure.

Problem 4.2: Consider Problem 4.1 in which the single energy constraint (4.3) on the uncertain quantities is replaced by the three individual instantaneous constraints

$$x_o' \Psi^{-1} x_o \le 1 \tag{4.30a}$$

$$w_k' Q_k^{-1} w_k \le 1, \qquad k = 0, 1, .., N-1 \tag{4.30b}$$

$$v_k' R_k^{-1} v_k \le 1, \qquad k = 1, 2, .., N \tag{4.30c}$$

where Ψ, Q_k, R_k are positive definite symmetric matrices. As in Problem 4.1, find the set $X_{i|k}$ of system states at time i that are consistent with both the measurements $z_I, z_2, \ldots, z_k$ up to time k and the constraints (4.30).

6.   ## The Filtering Problem with Instantaneous Constraints

Contrary to the case of energy constraints, it is very difficult to obtain the exact solution of Problem 4.2. As mentioned earlier the energy constraint (4.3) defines an ellipsoid in the space $R^{n+N(r+p)}$. Since the measurements $z_1, z_2, \ldots, z_k$ define a linear variety in this space and since the intersection of an ellipsoid with a linear variety is also an ellipsoid, the set of possible system states $X_{i|k}$, obtained by a linear transformation on this ellipsoid intersection, is also an ellipsoid, as found in Sections 3 and 4. The individual instantaneous constraints (4.30) do not, on the other hand, define an ellipsoid, and thus the intersection of the linear variety defined by the observed measurements with the subset of $R^{n+N(r+p)}$ satisfying (4.30) is not in general an ellipsoid. Consequently, the set of system states at time i consistent with the measurements $z_1, z_2, \ldots, z_k$ is not in general an ellipsoid: it is a convex set that, in contrast to the ellipsoidal case, cannot in general be characterized by a finite set of numbers.

Thus one is forced to seek approximate solutions to Problem 4.2. The approach taken by Schweppe[S1] is to compute a bounding ellipsoid to the set $X_{i|k}$. Since an ellipsoid in R^n is completely characterized by an n-vector (its center), and an $n \times n$ weighting matrix, the storage problem is reduced to more manageable proportions. Schweppe considered the filtering and prediction problems for a discrete-time system in (S1), and gave a

recursive algorithm for the center and weighting matrix of a bounding ellipsoid to the set of possible states. The approach used was to bound recursively the set of possible states at each time instant by an ellipsoid. This algorithm was later extended to a continuous-time system using a discrete-to-continuous limiting argument.[S2] The following lemma gives the filtering algorithm that is presented by Schweppe in (S1).

Lemma 4.1: A bounding ellipsoid to the set of system states $X_{k|k}$ of Problem 4.2, is given for all k, $0 \leq k \leq N$, by:

$$X^*_{k|k} = \{x \mid (x - \hat{x}_k)' \Sigma^{-1}_{k|k} (x - \hat{x}_k) \leq 1\}$$

where the positive definite matrix $\Sigma_{k|k}$ is given recursively by the equations

$$\Sigma_{i|i} = (1 - \delta^2_i)[(1 - \rho_i)\Sigma^{-1}_{i|i-1} + \rho_i C'_i R^{-1}_i C_i]^{-1} \tag{4.32}$$

$$\Sigma_{i|i-1} = (1 - \beta_{i-1})^{-1} A_{i-1} \Sigma_{i-1|i-1} A'_{i-1} + \beta^{-1}_{i-1} B_{i-1} Q_{i-1} B'_{i-1} \tag{4.33}$$

$$\Sigma_{o|o} = \Psi \tag{4.34}$$

the vector $\hat{x}_k$ is the solution of the equation

$$\hat{x}_{i+1} = A_i \hat{x}_i + \rho_{i+1}(1 - \delta^2_{i+1})^{-1} \Sigma_{i+1|i+1} C'_{i+1} R^{-1}_{i+1}(z_{i+1} - C_{i+1} A_i \hat{x}_i) \tag{4.35}$$

with the initial condition

$$\hat{x}_o = 0 \tag{4.36}$$

and the nonegative scalar δ^2_i is given for all i by

$$\delta^2_i = (z_i - C_i A_{i-1} \hat{x}_{i-1})'[(1-\rho_i)^{-1} C_i \Sigma_{i|i-1} C'_i + \rho^{-1}_i R_i]^{-1}(z_i - C_i A_{i-1} \hat{x}_{i-1}) \tag{4.37}$$

where the scalars β_{i-1}, ρ_i, are free parameters with $0 < \beta_{i-1} < 1$, $0 < \rho_i < 1$, $i = 1, 2, \ldots, N$.

The estimator of the above lemma has the same basic structure as the stochastic Kalman filter. It should be noted, however, that the gain matrix $\rho_{i+1}(1 - \delta_{i+1}^2)^{-1} \Sigma_{i+1|i+1} C'_{i+1} R_{i+1}^{-1}$ depends on the measurements received at a particular run and must be calculated from the equations (4.32) through (4.34) on-line. Furthermore, even for a time-invariant system, this estimator does not possess a steady state structure due to the fact that the solution of equations (4.32) through (4.34) does not converge to a steady state as time increases.

These disadvantages are avoided in the estimator we now derive. The approach is again to bound the set of possible states consistent with the observations by an ellipsoid. In contrast to (S1), we do this indirectly by bounding the instantaneous constraints (4.30) with an energy constraint of the form (4.3) and then using the results of Section 4 to produce an ellipsoidal bound on $X_{k|k}$. We will restrict our attention to the filtering problem. Entirely similar arguments can be used to derive estimators for the prediction and smoothing problems.

An energy bound for the instantaneous constraints (4.30) is given in the following lemma:

<u>Lemma 4.2</u>: The set $U_k \subseteq R^{n+k(r+p)}$ where

$$U_k = \{x_o, w_o, \ldots, w_{k-1}, v_1, \ldots, v_k \mid x'_o \Psi^1 x_o \leq 1, \; w'_{i-1} Q_{i-1}^{-1} w_{i-1} \leq 1,$$

$$v'_i R_i^{-1} v_i \leq 1, \; i = 1, 2, \ldots, k\}$$

is contained in the set

$$U_k^* = \{x_o, w_o, \ldots, w_{k-1}, v_1, \ldots, v_k \mid a_1 x_o' \Psi^{-1} x_o + \sum_{i=1}^{k} (a_{2,i-1} w_{i-1}' Q_{i-1}^{-1} w_{i-1}$$
$$+ a_{3,i} v_i' R_i^{-1} v_i \leq 1\} \qquad (4.39)$$

where a_1, $a_{2,i-1}$, $a_{3,i}$, $i = 1, 2, \ldots, k$, are any nonnegative real numbers with

$$a_1 + \sum_{i=1}^{k} (a_{2,i-1} + a_{3,i}) = 1 \qquad (4.40)$$

<u>Proof</u>: Multiply (4.30a, b, c) by a_1, $a_{2,i-1}$, $a_{3,i}$, respectively, sum the last two from $i = 1$ to $i = k$ and use (4.40). Q.E.D.

Having bounded the instantaneous constraints (4.38) by the energy constraint (4.39), we are now in a position to apply the results of Proposition 4.2 to give a bounding ellipsoid to the set $X_{k|k}$. The equations that result by application of Proposition 4.2 become simpler if we write a_1, $a_{2,i-1}$, and $a_{3,i}$ in the following form:

$$a_1 = (1-\beta_o)(1-\rho_1)(1-\beta_1)(1-\rho_2) \cdots (1-\beta_{k-1})(1-\rho_k)$$

$$a_{2,0} = \beta_o(1-\rho_1)(1-\beta_1)(1-\rho_2) \cdots (1-\beta_{k-1})(1-\rho_k)$$

$$a_{3,1} = \rho_1(1-\beta_1)(1-\rho_2) \cdots (1-\beta_{k-1})(1-\rho_k) \qquad (4.41)$$

$$- -$$

$$a_{2,k-1} = \beta_{k-1}(1-\rho_k)$$

$$a_{3,k} = \rho_k$$

where β_{i-1}, ρ_i, $i = 1, 2, \ldots, k$ are any real numbers with $0 < \beta_{i-1} < 1$, $0 < \rho_i < 1$.

It is easy to see that for the parameters a_1, $a_{2,i-1}$, $a_{3,i}$ as defined by equations (4.41) we have $a_1 + \sum_{i=1}^{k}(a_{2,i-1} + a_{3,i}) = 1$.

By combining now Lemma 4.2 under the identifications (4.41) with Proposition 4.2 we have after straightforward manipulation the following solution to Problem 4.2 for the filtering case.

Proposition 4.3: A bounding ellipsoid to the set of system states $X_{k|k}$ of Problem 4.2 is given for all k, $0 \leq k \leq N$, by the equation

$$X^*_{k|k} = \{x | (x - \hat{x}_k)'\Sigma^{-1}_{k|k}(x - \hat{x}_k) \leq 1 - \delta^2(k)\} \tag{4.42}$$

where the positive definite symmetric matrix $\Sigma_{k|k}$ is given recursively by the equations

$$\Sigma_{i|i} = [(1-\rho_i)\Sigma^{-1}_{i|i-1} + \rho_i C'_i R^{-1}_i C_i]^{-1} \tag{4.43}$$

$$\Sigma_{i|i-1} = (1-\beta_{i-1})^{-1}A_{i-1}\Sigma_{i-1|i-1}A'_{i-1} + \beta^{-1}_{i-1}B_{i-1}Q_{i-1}B'_{i-1} \tag{4.44}$$

$$\Sigma_{o|o} = \Psi \tag{4.45}$$

the vector $\hat{x}_k$ is the solution of the equation

$$\hat{x}_{i+1} = A_i\hat{x}_i + \rho_{i+1}\Sigma_{i+1|i+1}C'_{i+1}R^{-1}_{i+1}(z_{i+1} - C_{i+1}A_i\hat{x}_i) \tag{4.46}$$

with the initial condition

$$\hat{x}_o = 0 \tag{4.47}$$

and the nonnegative scalar $\delta^2(k)$ is the solution of the equation

$$\delta^2(i) = (1-\beta_{i-1})(1-\rho_i)\delta^2(i-1)$$
$$+ (z_i - C_i A_{i-1}\hat{x}_{i-1})'[(1-\rho_i)^{-1}C_i\Sigma_{i|i-1}C'_i + \rho^{-1}_i R_i]^{-1}(z_i - C_i A_{i-1}\hat{x}_{i-1}) \tag{4.48}$$

with the initial condition

$$\delta^2(0) = 0 \tag{4.49}$$

and β_{i-1}, ρ_i, $i = 1, 2, \ldots, N$, are any real numbers with $0 < \beta_{i-1} < 1$, $0 < \rho_i < 1$.

It can be seen that the estimator of the above proposition has a similar structure with the stochastic Kalman filter as well as with the estimator of Lemma 4.1. However, it has the important advantage over the latter that the gain matrix $\{\rho_{i+1} \Sigma_{i+1|i+1} C'_{i+1} R^{-1}_{i+1}\}$ is precomputable once the parameters β_{i-1}, ρ_i are selected. Furthermore, as will be discussed in the next section, for a time-invariant system the estimator of Proposition 4.3 can be implemented as a time-invariant system if the final time N approaches infinity. In practical applications this last advantage can be of extreme importance.

A vital question concerns the comparison of the quality of approximation to the set of possible states provided by the two estimators. It turns out that the approximation is comparable in the following sense. Let $\{\beta'_o, \rho'_1, \ldots, \beta'_{N-1}, \rho'_N\}$ be a set of parameters used in the estimator of Lemma 4.1 and $\{\beta_o, \rho_1, \ldots, \beta_{N-1}, \rho_N\}$ be a set of parameters used in the estimator of Proposition 4.3. Then if we select[*] for $i = 1, 2, \ldots, N$

$$\beta'_{i-1} = \frac{\beta_{i-1}}{[1-\delta^2(i-1)](1-\beta_{i-1}) + \beta_{i-1}} \tag{4.50}$$

$$\rho'_i = \frac{\rho_i}{\{[1-\delta^2(i-1)](1-\beta_{i-1}) + \beta_{i-1}\}(1-\rho_i) + \rho_i} \tag{4.51}$$

---

[*] This fact was brought to the author's attention by F. Schlaepfer.

where $\delta^2(i-1)$ is the measurement dependent term of equation (4.48) in Proposition 4.3, the estimate ellipsoids $X^*_{k|k}$ provided by the two estimators are identical for all k and for all sets of received measurements for which equations (4.50) and (4.51) hold.

Another important question concerns the quality of the approximation of the bounding ellipsoid $X^*_{k|k}$ produced by the estimator of Proposition 4.3 to the exact set of possible states $X_{k|k}$. This is a question largely unresolved to this date. It appears to be very difficult to obtain estimates of the approximation involved which will be applicable to a large class of problems. For any given problem however to is possible to estimate exactly the approximation in any direction as it will be discussed in Section 8. A related problem which will also be discussed in Section 8 is the question of the optimal selection of the parameters β_{i-1} and ρ_i.

7. Constant Systems and Infinite Time Intervals

In this section we consider the special case of Problem 4.2 where the system and the disturbance ellipsoid sets are constant, i.e., $A_k = A$, $B_k = B$, $C_k = C$, $Q_k = Q$, $R_k = R$ for all k. If we select the parameters β_k, ρ_k to be also constant (i.e., $\beta_k = \beta$, $\rho_k = \rho$ for all k), the equations (4.43), (4.44) for the matrix $\Sigma_{k|k}$ in Proposition 4.3 become

$$\Sigma_{k|k} = [(1-\rho)\Sigma^{-1}_{k|k-1} + \rho C'R^{-1}C]^{-1} \qquad (4.52)$$

$$\Sigma_{k|k-1} = (1-\beta)^{-1}A\Sigma_{k-1|k-1}A' + \beta^{-1}BQB' \qquad (4.53)$$

with initial condition $\Sigma_{o|o} = \Psi$. These equations can be put into the usual discrete-time Riccati equation form

$$\Sigma_{k|k} = (\Sigma_{k|k-1}^{-1} + C'R^{*-1}C)^{-1} \tag{4.54}$$

$$\Sigma_{k|k-1} = A^*\Sigma_{k-1|k-1}A^{*\prime} + BQ^*B' \tag{4.55}$$

by defining the matrices A^*, Q^*, R^* as

$$A^* = (1-\beta)^{-1/2}(1-\rho)^{-1/2}A, \quad Q^* = \beta^{-1}(1-\rho)^{-1}Q, \quad R^* = \rho^{-1}R \tag{4.56}$$

It is well known[T1] that the solution $\Sigma_{k|k}$ of equations (4.54), (4.55) converges to a positive definite symmetric matrix Σ_∞ as $k \to \infty$ if the pair (A^*, C) is completely observable and the pair (A^*, B) is completely controllable. The pair (A^*, B) is completely controllable if and only if the pair (A, B) is completely controllable i.e., the constant system (4.1) is completely controllable. This can be seen by the fact that the matrix A^* is a scalar multiple of the matrix A and therefore the subspace spanned by the column vectors of the matrix $A^m B$ is the same as the subspace spanned by the column vectors of the matrix $A^{*m}B$ for all $m = 0, 1, \ldots, n-1$. Similarly, the pair (A^*, C) is completely observable if and only if the pair (A, C) is completely observable. Thus, for a completely controllable and observable time-invariant system, the gain $\{\Sigma_{k|k}C'R^{*-1}\}$ in the estimator of Proposition 4.3 after an initial transient will converge to a steady state constant gain $\{\Sigma_\infty C'R^{*-1}\}$. For practical reasons, one would like to implement the estimator as a time-invariant system using the steady-state gain for the whole time interval, i.e., starting at the initial time $k = 0$. This is possible since, as we will prove below, the approximation that results by neglecting the initial transient vanishes as time goes to infinity.

Using the identifications (4.56), the estimator of Proposition 4.3 for a time-invariant system gives the estimate ellipsoid

$$X^*_{k|k} = \{x \mid (x - \hat{x}_k)' \Sigma^{-1}_{k|k}(x - \hat{x}_k) \leq 1 - \delta^2(k)\} \tag{4.57}$$

where

$$\Sigma_{k|k} = (\Sigma_{k|k-1} + C'R^{*-1}C)^{-1} \tag{4.58}$$

$$\Sigma_{k|k-1} = A^*\Sigma_{k-1|k-1}A^{*'} + BQ^*B' \tag{4.59}$$

$$\hat{x}_{k+1} = A\hat{x}_k + \Sigma_{k+1|k+1}C'R^{*-1}(z_{k+1} - CA\hat{x}_k) \tag{4.60}$$

$$\delta^2(k) = (1-\beta_{k-1})(1-\rho_k)\delta^2(k-1)$$
$$+ (z_k - CA\hat{x}_{k-1})'(C\Sigma_{k|k-1}C' + R^*)^{-1}(z_k - CA\hat{x}_{k-1}) \tag{4.61}$$

with

$$\Sigma_{0|0} = \Psi, \quad \hat{x}_0 = 0, \quad \delta^2(0) = 0 \tag{4.62}$$

If $\Sigma_{k|k} \to \Sigma_\infty$ and $\Sigma_{k|k-1} \to \tilde{\Sigma}_\infty$ as $k \to \infty$ and we implement the estimator as a time-invariant system using the steady-state gain $\{\Sigma_\infty C'R^{*-1}\}$ the resulting estimate ellipsoid will be given by

$$Y_{k|k} = \{x \mid (x - \hat{y}_k)' \Sigma^{-1}_\infty (x - \hat{y}_k) \leq 1 - \delta^2(k)\} \tag{4.63}$$

where

$$\hat{y}_{k+1} = A\hat{y}_k + \Sigma_\infty C'R^{*-1}(z_{k+1} - CA\hat{y}_k) \tag{4.64}$$

$$\tilde{\delta}^2(k) = (1-\beta)(1-\rho)\tilde{\delta}^2(k-1) + (z_k - CA\hat{y}_{k-1})'(C\tilde{\Sigma}_\infty C' + R^*)^{-1}(z_k - CA\hat{y}_{k-1}) \tag{4.65}$$

with

$$\hat{y}_0 = 0, \qquad \tilde{\delta}^2(0) = 0 \tag{4.66}$$

Using the fact that $\Sigma_{k|k} \to \Sigma_\infty$ and $\Sigma_{k|k-1} \to \tilde{\Sigma}_\infty$ as $k \to \infty$, it will now be proved that $\hat{y}_k \to \hat{x}_k$ and $\tilde{\delta}^2(k) \to \delta^2(k)$ as $k \to \infty$, i.e., that the estimate

ellipsoid $Y_{k|k}$ of equation (4.63) "converges" to the set $X^*_{k|k}$ of equation
(4.57) as $k \to \infty$. To this end let $\Sigma_{k|k} = \Sigma_\infty + H_k$ where $H_k \to 0$ as $k \to \infty$.
Then from equations (4.60) and (4.64) we have

$$\hat{x}_{k+1} - \hat{y}_{k+1} = (A - \Sigma_\infty C'R^{*-1}CA)(\hat{x}_k - \hat{y}_k) + H_k C'R^{*-1}(z_{k+1} - CA\hat{x}_k) \quad (4.67)$$

Now note that the matrix $(A - \Sigma_\infty C'R^{*-1}CA)$ is stable (has eigenvalues within
the unit disk), since by equation (4.56)

$$A - \Sigma_\infty C'R^{*-1}CA = (1-\beta)^{1/2}(1-\rho)^{1/2}(A^* - \Sigma_\infty C'R^{*-1}CA^*)$$

and the matrix $(A^* - \Sigma_\infty C'R^{*-1}CA^*)$ is stable by a well-known property of the
Riccati equation. Furthermore, the driving term $H_k C'R^{*-1}(z_{k+1} - CA\hat{x}_k)$
goes to zero as $k \to \infty$ since $H_k \to 0$ as $k \to \infty$ and $(z_{k+1} - CA\hat{x}_k)$ is bounded.
Therefore, the solution of equation (4.67) goes asymptotically to zero as
$k \to \infty$ and hence $\hat{y}_k \to \hat{x}_k$ as $k \to \infty$.

Also from equations (4.61) and (4.65)

$$\delta^2(k+1) - \tilde{\delta}^2(k+1) = (1-\beta)(1-\rho)[\delta^2(k) - \tilde{\delta}^2(k)] + \epsilon_{k+1} \quad (4.68)$$

where

$$\epsilon_{k+1} = (z_{k+1} - CA\hat{x}_k)'(C\Sigma_{k+1|k}C' + R^*)^{-1}(z_{k+1} - CA\hat{x}_k)$$
$$- (z_{k+1} - CA\hat{y}_k)'(C\Sigma_\infty C' + R^*)^{-1}(z_{k+1} - CA\hat{y}_k)$$

Since $\hat{y}_k \to \hat{x}_k$ and $\Sigma_{k+1|k} \to \tilde{\Sigma}_\infty$ as $k \to \infty$ we have $\epsilon_{k+1} \to 0$ as $k \to \infty$ and
since $0 < (1-\beta)(1-\rho) < 1$ the solution of the equation (4.68) goes to zero as
$k \to \infty$. Hence $\tilde{\delta}^2(k) \to \delta^2(k)$ as $k \to \infty$.

Thus, in applications where the system is constant and the final time
approaches infinity, one can use the steady-state time-invariant estimator
and be assured that the error that results from neglecting the initial trans-
ient of the solution of the Riccati equation vanishes as time increases.

8. Discussion and Sources

Two state estimation problems which involve a linear system and a set-membership description of the uncertainty were examined in this chapter. For the case of an energy constraint on the uncertain quantities the set of possible states consistent with the measurements was shown to be an ellipsoid which was characterized by recursive estimators similar to Kalman estimators used in stochastic estimation problems. The results for the energy constraint case were then used to obtain bounding ellipsoid estimators for the, more often appearing in practice, case of instantaneous ellipsoidal constraints on the uncertain quantities. These estimators have the same basic structure as the Kalman estimators and offer distinct advantages over existing schemes[S1], [S2].

The basic practical advantage of the estimators proposed in this chapter is that they provide intelligent designs with a minimal amount of information. Instead of requiring precise statistics of the uncertain quantities only bounds on the magnitude or energy of the uncertain quantities are necessary. Since the estimators have the same basic structure as Kalman estimators the approach used here in effect suggests an intelligent way of selecting the gain matrices of the estimator with a minimal amount of information.

One of the questions yet largely unresolved concerns the quality of the approximation involved in the algorithms for the instantaneous constraint case. Related to this question is the problem of optimal selection of the free parameters β_i and ρ_i that appear in the algorithms. There are two difficulties related to this problem. First a criterion for optimization must

be chosen. Second an optimization algorithm must be devised based on this criterion. Even choosing a good criterion is a difficult question. For instance a method which appears at first sight to be reasonable is to find the parameters β_i, ρ_i for which the trace of the weighting matrix $\Sigma_{N|N}$ at the final time is minimized. An algorithm for selection of the parameters so as to optimize this criterion was derived by the author yet for some simple examples the resulting selection of the parameters led to an indeed poor design. Presently there exists no optimization algorithm for selecting the parameters β_i, ρ_i, and some trial and error must be used for their selection. For the case of a time-invariant system and an infinite time interval this is not very troublesome since in this case only two parameters β, ρ must be selected with $0 < \beta < 1$, $0 < \rho < 1$.

Given now any bounding ellipsoid estimator of the form appearing in Proposition 4.3, and any set of measurements $z_1, z_2, \ldots, z_k$ a comparison of the bounding ellipsoid $X_{k|k}^*$ with the exact set of possible states $X_{k|k}$ can be made in any direction x^* by comparing the value of the support function

$$\sigma(x^* | X_{k|k}^*) = \langle x^*, \hat{x}_k \rangle + [1 - \delta^2(k)]^{1/2} (x^{*\prime} \Sigma_{k|k} x^*)^{1/2}$$

with the value of the support function $\sigma(x^* | X_{k|k})$. This latter value can be calculated from

$$\sigma(x^* | X_{k|k}) = \sup_{x_k \in X_{k|k}} \langle x^*, x_k \rangle$$

subject to the constraints

$$x_{i+1} = A_i x_i + B_i w_i , \qquad i = 0, 1, \ldots, k-1$$

$$z_i = C_i x_i + v_i , \qquad i = 1, 2, \ldots, k$$

$$x_o' \Psi^{-1} x_o \leq 1$$

$$w_i' Q_i^{-1} w_i \leq 1, \qquad i = 0, 1, \ldots, k-1$$

$$v_i' R_i^{-1} v_i \leq 1, \qquad i = 1, 2, \ldots, k$$

a linear program with linear and quadratic constraints. A comparison of these values for a number of directions of interest and for a variety of sets of measurements can be informative concerning the quality of the approximation of the estimates provided by the given bounding ellipsoid algorithm. We mention that the question of parameter selection and of the quality of approximation have been discussed by Schlapfer.[Sc1] Some simulations can also be found in the same reference.

Similar results to those obtained in this chapter can be derived for a variety of problems not explicitly considered here. One such problem was briefly discussed in Section 3 and concerns the case where there is no measurement noise in equation (4.2) and the energy constraint is of the form

$$x_o' \Psi^{-1} x_o + \sum_{i=0}^{N-1} w_i' Q_i^{-1} w_i \leq 1$$

The estimator for this problem is very similar to the corresponding stochastic estimator[T1] and can be used to obtain a bounding ellipsoid algorithm for the related instantaneous constraint case where there is no measurement noise by using a similar bounding operation to the one in Section 6. Another problem that can be treated similarly is the static estimation problem[S2] which does not involve a dynamic system.

The continuous time counterparts of the estimators of this chapter have already appeared in (B2). The approach used in this reference was

to associate the estimation problem for the energy constraint case with the standard tracking problem of optimal control theory in which time is reversed. This approach can also be used for most of the problems considered here, and has the advantage that it demonstrates in a direct way the duality between linear estimation problems and linear quadratic optimal control problems. However the approach used here is more efficient in that it is applicable to more general cases. In particular it is applicable to those problems for which the estimate ellipsoid is degenerate (has a weighting matrix which is positive semidefinite but not positive definite). Furthermore it proves explicitly the one-one correspondence between estimation problems with an energy constraint description of the uncertainty and linear minimum variance stochastic estimation problem. The reader familiar with the Hilbert space formulation of stochastic estimation problems[Lu1] will have no difficulty observing from the proof of Proposition 4.1 that the solutions of Problem 4.1 and Problem 4.1' involve dual applications of the projection theorem which result in identical equations.

Estimation problems involving a set-membership description of the uncertainty were first considered by Witsenhausen[W1] in the framework of minimax control problems with imperfect state information. The set description approach towards the estimation problem gained attention following the work of Schweppe[S1], [S2] who demonstrated that by using ellipsoidal approximations, algorithms with potential for practical applications could be devised. The results of this chapter were in fact largely motivated by Schweppe's work. Extensions of Schweppe's algorithms to distributed parameter systems were obtained by Schlaepfer[Sc1] Such extensions should be possible for the results of this chapter as well. An estimation problem

which does not involve ellipsoids is the one which involves a linear discrete-time system and instantaneous polyhedral constraints for the uncertain quantities. Such constraints are interesting because the resulting set of possible states consistent with the measurements can be characterized precisely by a finite set of bounding hyperplanes. However the number of these bounding hyperplanes increases with time thus possibly creating a serious storage as well as computational problem. A method for obtaining polyhedral approximations to the set of possible states using only a fixed number of bounding hyperplanes is discussed by Hnyilidza[H1].

Finally it should be noted that the results presented in this chapter rely heavily on the linearity of the system, and it appears to be quite difficult to obtain extensions to nonlinear estimation problems. However such problems have not been sufficiently explored up to now and are worthy of attention.

CHAPTER 5

MINIMAX CONTROL PROBLEMS WITH IMPERFECT STATE INFORMATION

1. General Remarks

We now turn our attention to minimax control problems with imperfect state information. We will consider the general Problem 1.1 which was introduced in Chapter 1. The special case of this problem where the system is linear, the cost functional has some convexity properties, and the controller has available an exact measurement of the system state has been examined in Chapter 2. The additional structure of this special case allowed us to obtain results that are considerably stronger than those that can be deduced for the general Problem 1.1. For this latter problem if is very difficult to obtain results concerning existence of solutions or necessary conditions for optimality. Furthermore the solution of the problem by dynamic programming, which will be presented in Section 3, becomes extremely complicated in general. This is due mainly to the fact that, as will be demonstrated, the optimal controller performs the dual function of state identification and system actuation. The complexities of this situation are well known from stochastic optimal control problems.[F1],[A1] We will be able to obtain insight into the dual function of the optimal controller through the notion of a sufficiently informative function which parallels the familiar notion of a sufficient statistic[St1] of stochastic optimal control. The notion of a sufficiently informative function will be introduced in Section 4, and it will be used for demonstrating the separation of the optimal controller into an estimator and an actuator. The special case of a linear system where

the uncertain quantities satisfy an energy constraint will be further investigated in Section 5. For this case it will be shown that the estimator part of the optimal controller can be easily and efficiently characterized. Still for this case the actuator part of the optimal controller cannot in general be easily characterized although we shall demonstrate the characterization of this actuator for the special case of a reachability problem in the next chapter.

In the next section we restate and briefly discuss the Problem 1.1 which is the object of study of this chapter.

2.   Problem Formulation

We shall consider the following problem:

<u>Problem 5.1</u>: Given is the discrete-time dynamic system

$$x_{k+1} = f_k(x_k, u_k, w_k) , \qquad k = 0, 1, \ldots, N-1 \qquad (5.1)$$

where $x_k \epsilon R^n$, $k = 0, 1, \ldots, N$ is the state vector, $u_k \epsilon R^m$, $k = 0, 1, \ldots, N-1$, is the control vector, $w_k \epsilon R^r$, $k = 0, 1, \ldots, N-1$, is the input disturbance vector, and $f_k : R^n x R^m x R^r \rightarrow R^n$ are known functions.

Available to the controller are measurements of the form

$$z_k = h_k(x_k, v_k) , \qquad k = 0, 1, \ldots, N-1 \qquad (5.2)$$

where $z_k \epsilon R^s$, $k = 1, 2, \ldots, N-1$, is the measurement vector, $v_k \epsilon R^p$, $k = 1, 2, \ldots, N-1$, is the measurement noise vector, and $h_k : R^n x R^p \rightarrow R^s$ are known functions.

The uncertain quantities lumped in a vector $q \epsilon R^{n+Nr+(N-1)p}$

$$q = (x_o', w_o', w_1', \ldots, w_{N-1}', v_1', v_2', \ldots, v_{N-1}')' \qquad (5.3)$$

are known to belong to a given subset Q of $R^{n+Nr+(N-1)p}$

$$q \in Q \tag{5.4}$$

Attention is restricted to control laws of the form

$$\mu_k : R^{k(s+m)} \to R^m, \qquad k = 0, 1, \ldots, N-1 \tag{5.5}$$

taking values

$$u_k = \mu_k(z_1, z_2, \ldots, z_k, u_o, u_1, \ldots, u_{k-1}), \quad k = 0, 1, \ldots, N-1 \tag{5.6}$$

where μ_o is interpreted as a constant vector.

It is required to find (if it exists) the control law in this class for which the cost functional

$$J(\mu_o, \mu_1, \ldots, \mu_{N-1}) = \sup_{q \in Q} F[x_1, x_2, \ldots, x_N, \mu_o, \mu_1(z_1, u_o), \ldots, \mu_{N-1}(z_1, \ldots, u_{N-1})] \tag{5.7}$$

is minimized subject to the system equation constraints (5.1), and where $F : R^{N(n+m)} \to (-\infty, +\infty]$ is a given function.

As in Problem 2.1, the use of the (semiclosed) extended real line as the range of the function F permits the incorporation of state and control constraints in the cost functional by adding to the function F the indicator functions of the state and control constraint sets.

In the next section we will present a dynamic programming algorithm for solution of the Problem 5.1. Using this algorithm we will then be able to reach some conclusions concerning the structure of the optimal control law.

3. Solution by Dynamic Programming

Consider the optimal value of the cost function (5.7)

$$\bar{J} = \inf_{\substack{\mu_k \\ k = 0, 1, \ldots, N-1}} \sup_{q \in Q} F(x_1, x_2, \ldots, x_N, u_0, u_1, \ldots, u_{N-1}) \tag{5.8}$$

The purpose of the dynamic programming algorithm is to convert the minimization problem indicated in the above equation to a sequence of simpler minimization problems by taking advantage of the sequential evolution of the system state, and the information available to the controller according to equations (5.1) and (5.2). However matters are somewhat complicated in the above problem by the presence of uncertainty since in the process of generating the state and measurement vectors the disturbances are intermediately selected by, say, Nature with the objective to maximize the value of the cost. For this reason the development of the dynamic programming algorithm will require a somewhat elaborate construction. We give first the following preliminary definitions.

Let $P(R^s)$ be the power set (set of all subsets) of R^s and consider the following function

$$\hat{Z}_1 : R^m \rightarrow P(R^p)$$

which assigns to the vector $u_0 \in R^m$ the set $\hat{Z}_1(u_0) \subset R^p$ consisting of all possible measurement vectors z_1 given by equation (5.2) which are consistent with the constraint set Q, the system equation (5.1) and the control vector u_0. In other words we have $z_1 \in \hat{Z}_1(u_0)$ if and only if there exists a vector $q = (x_0', w_0', w_1', \ldots, w_{N-1}', v_1', v_2', \ldots, v_{N-1}')' \in Q$ such that

$$z_1 = h_1[f_0(x_0, u_0, w_0), v_1]$$

Similarly we define for $k = 2, 3, \ldots, N-1$ the function

$$\hat{Z}_k : R^{(k-1)s+km} \to P(R^s) \tag{5.9}$$

which assigns to the vectors $z_1, z_2, \ldots, z_{k-1}, u_0, u_1, \ldots, u_{k-1}$ the set $\hat{Z}_k(z_1, z_2, \ldots, z_{k-1}, u_0, u_1, \ldots, u_{k-1}) \subset R^s$ of all measurement vectors z_k given by equation (5.2) which are consistent with the constraint set Q, the previous measurement vectors $z_1, z_2, \ldots, z_{k-1}$, and the previous control vectors $u_0, u_1, \ldots, u_{k-1}$.

We also define the function

$$\hat{Q} : R^{(N-1)s+Nm} \to P(R^{n+Nr+(N-1)p}) \tag{5.10}$$

which assigns to the vectors $z_1, z_2, \ldots, z_{N-1}, u_0, u_1, \ldots, u_{N-1}$ the set $\hat{Q}(z_1, z_2, \ldots, z_{N-1}, u_0, u_1, \ldots, u_{N-1}) \subset R^{n+Nr+(N-1)p}$ of all vectors $q = (x_0', w_0', w_1', \ldots, w_{N-1}', v_1', v_2', \ldots, v_{N-1}')$ which belong to the set Q and are consistent with the measurements $z_1, z_2, \ldots, z_{N-1}$, and the control vectors $u_0, u_1, \ldots, u_{N-1}$. In other words a vector $q = (x_0', w_0', w_1', \ldots, w_{N-1}', v_1', v_2', \ldots, v_{N-1}')'$ belongs to the set $\hat{Q}(z_1, z_2, \ldots, z_{N-1}, u_0, u_1, \ldots, u_{N-1})$ if and only if $q \epsilon Q$ and the vectors $x_0', w_0', \ldots, w_{N-1}', v_1', \ldots, v_{N-1}', z_1, \ldots, z_{N-1}, u_0, \ldots, u_{N-1}$ together satisfy the system and measurement equation (5.1) and (5.2) for all k.

In order to simplify the notation we will make use of the following vector ζ_k, $k = 1, 2, \ldots, N-1$, which consists of all the information available to the controller at time k

$$\zeta_k = \{z_1, z_2, \ldots, z_k, u_0, u_1, \ldots, u_{k-1}\} \tag{5.11}$$

Using this notation we write for the control law μ_k and the functions $\hat{Z}_k, \hat{Q}$ of equations (5.9) and (5.10)

$$\mu_k(z_1, z_2, \ldots, z_k, u_o, u_1, \ldots, u_{k-1}) = \mu_k(\zeta_k) = u_k \tag{5.12}$$

$$\hat{Z}_k(z_1, z_2, \ldots, z_{k-1}, u_o, u_1, \ldots, u_{k-1}) = \hat{Z}_k(\zeta_{k-1}, u_{k-1}) \tag{5.13}$$

$$\hat{Q}(z_1, z_2, \ldots, z_{N-1}, u_o, u_1, \ldots, u_{N-1}) = \hat{Q}(\zeta_{N-1}, u_{N-1}) \tag{5.14}$$

It should be noted that for some vectors ζ_{k-1} it is possible that the set $\hat{Z}_k(\zeta_{k-1}, u_{k-1})$ or the set $\hat{Q}(\zeta_{N-1}, u_{N-1})$ is empty for all $u_{k-1} \in R^m$ implying that the vector $\zeta_{k-1} = (z_1', z_2', \ldots, z_{k-1}', u_o', u_1', \ldots, u_{k-2}')'$ is inconsistent with the constraint set Q and the system and measurement equations. Notice also that whether the set $\hat{Z}_k(\zeta_{k-1}, u_{k-1})$ is empty or nonempty depends on the vector ζ_{k-1} alone and is entirely independent of u_{k-1}. In equations to follow in which empty sets appear we will adopt the convention stated in Appendix I that the supremum of the empty set is $-\infty$ ($\sup \phi = -\infty$). Another possible approach would be to restrict the domain of definition of the functions $\hat{Z}_k$, $\hat{Q}$ to include only those vectors ζ_{k-1} for which the sets $\hat{Z}_k(\zeta_{k-1}, u_{k-1})$, $\hat{Q}(\zeta_{N-1}, u_{N-1})$ are nonempty. Since in any actual operation of the system these sets will always be nonempty this restriction results in no loss of generality.

We are now ready to state and prove the following dynamic programming algorithm.

Proposition 5.1: Assume that for the functions H_k defined below we have $-\infty < H_k(\zeta_k)$, $k = 1, 2, \ldots, N-2$, for all vectors ζ_k such that the set $\hat{Z}_{k+1}(\zeta_k, u_k)$ is nonempty (for all $u_k \in R^m$), and $-\infty < H_{N-1}(\zeta_{N-1})$ for all vectors ζ_{N-1} such that the set $\hat{Q}(\zeta_{N-1}, u_{N-1})$ is nonempty. Then the optimal value $\bar{J}$ of the cost functional (5.7) is given by

$$\bar{J} = \inf_{u_o} E_1(u_o) \tag{5.15}$$

where the function $E_1 : R^m \to (-\infty, +\infty]$ is given by the last step of the recursive algorithm

$$E_N(\zeta_{N-1}, u_{N-1}) = \sup_{q \in \hat{Q}(\zeta_{N-1}, u_{N-1})} F(x_1, x_2, \ldots, x_N, u_0, u_1, \ldots, u_{N-1}) \tag{5.16}$$

$$H_k(\zeta_k) = \inf_{u_k} E_{k+1}(\zeta_k, u_k), \qquad k = 1, 2, \ldots, N-1 \tag{5.17}$$

$$E_{k+1}(\zeta_k, u_k) = \sup_{z_{k+1} \in \hat{Z}_{k+1}(\zeta_k, u_k)} H_{k+1}(\zeta_k, u_k, z_{k+1}) \tag{5.18}$$

$$= \sup_{z_{k+1} \in \hat{Z}_{k+1}(\zeta_k, u_k)} H_{k+1}(\zeta_{k+1}), \quad k = 0, 1, \ldots, N-2$$

Proof: Consider the cost functional (5.7)

$$J(\mu_0, \mu_1, \ldots, \mu_{N-1}) = \sup_{q \in Q} F[x_1, x_2, \ldots, x_N, \mu_0, \mu_1(\zeta_1), \ldots, \mu_{N-1}(\zeta_{N-1})] \tag{5.7}$$

and the functions

$$J_{N-1}(\mu_0, \mu_1, \ldots, \mu_{N-2}) = \inf_{\mu_{N-1}} J(\mu_0, \mu_1, \ldots, \mu_{N-1}) \tag{5.19}$$

$$J_k(\mu_0, \mu_1, \ldots, \mu_{k-1}) = \inf_{\mu_k} J_{k+1}(\mu_1 \mu_2, \ldots, \mu_k), \quad k = 1, 2, \ldots, N-2 \tag{5.20}$$

We have for the optimal value of the cost functional

$$\bar{J} = \inf_{\mu_0} J_1(\mu_0) = \inf_{u_0} J_1(u_0) \tag{5.21}$$

To prove the proposition we will recursively show the equations

$$J_{N-1}(\mu_0, \mu_1, \ldots, \mu_{N-2}) = \sup_{z_1 \in \hat{Z}_1(u_0)} \cdots \sup_{z_{N-1} \in \hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2})} H_{N-1}(\zeta_{N-1}) \tag{5.22}$$

$$J_k(\mu_0, \mu_1, \ldots, \mu_{k-1}) = \sup_{z_1 \in \hat{Z}_1(\mu_0)} \cdots \sup_{z_k \in \hat{Z}_k(\zeta_{k-1}, u_{k-1})} H_k(\zeta_k) \qquad (5.23)$$

$$J_1(\mu_0) = J_1(u_0) = \sup_{z_1 \in \hat{Z}_1(u_0)} H_1(z_1, u_0) = E_1(u_0) \qquad (5.24)$$

where in the above equations u_k denotes, for all k, the value of the function μ_k at the point ζ_k.

The equation (5.15) which is to proved follows then by comparing equations (5.21) and (5.24).

We begin by proving equation (5.22). Consider the function E_N of equation (5.16)

$$E_N(\zeta_{N-1}, u_{N-1}) = \sup_{q \in \hat{Q}(\zeta_{N-1}, u_{N-1})} F(x_1, x_2, \ldots, x_N, u_0, u_1, \ldots, u_{N-1})$$

We have $E_N(\zeta_{N-1}, u_{N-1}) = -\infty$ for all vectors ζ_{N-1} such that $\hat{Q}(\zeta_{N-1}, u_{N-1}) = \phi$ (for all $u_{N-1} \in R^m$) and we have $E_N(\zeta_{N-1}, u_{N-1}) > -\infty$ otherwise since the function F does not take the value $-\infty$. By the assumption that $H_{N-1}(\zeta_{N-1}) > -\infty$ for all ζ_{N-1} for which $\hat{Q}(\zeta_{N-1}, u_{N-1}) \neq \phi$ we have

$$H_{N-1}(\zeta_{N-1}) = \inf_{u_{N-1}} E_N(\zeta_{N-1}, u_{N-1}) = -\infty, \text{ for all } \zeta_{N-1} \text{ such that}$$

$$\hat{Q}(\zeta_{N-1}, u_{N-1}) = \phi$$

$$H_{N-1}(\zeta_{N-1}) = \inf_{u_{N-1}} E_N(\zeta_{N-1}, u_{N-1}) > -\infty, \text{ for all } \zeta_{N-1} \text{ such that}$$

$$\hat{Q}(\zeta_{N-1}, u_{N-1}) \neq \phi$$

Thus for every $\epsilon > 0$ there exists a function

$$\mu_{N-1, \epsilon} : R^{(N-1)(s+m)} \to R^m \text{ such that}$$

$$E_N[\zeta_{N-1}, \mu_{N-1,\epsilon}(\zeta_{N-1})] \leq \inf_{\mu_{N-1}} E_N[\zeta_{N-1}, \mu_{N-1}(\zeta_{N-1})] +$$

$$\tag{5.25}$$

$$= \inf_{u_{N-1}} E_N(\zeta_{N-1}, u_{N-1}) + \epsilon = H_{N-1}(\zeta_{N-1}) + \epsilon$$

for all $\zeta_{N-1} \epsilon R^{(N-1)(s+m)}$

We have now from equation (5.22) for any fixed functions $\mu_o, \mu_1, \ldots,$ μ_{N-2} with $u_o = \mu_o, u_1 = \mu_1(\zeta_1), \ldots, u_{N-2} = \mu_{N-2}(\zeta_{N-2})$

$$J_{N-1}(\mu_o, \mu_1, \ldots, \mu_{N-2}) = \inf_{\mu_{N-1}} \sup_{q \epsilon Q} F(x_1, u_2, \ldots, x_N, u_o, u_1, \ldots, u_{N-1})$$

$$= \inf_{\mu_{N-1}} \sup_{z_1 \epsilon Z_1(u_o)} \cdots \sup_{z_{N-1} \epsilon Z_{N-1}(\zeta_{N-2}, u_{N-2})} \sup_{q \epsilon Q(\zeta_{N-1}, u_{N-1})} F(x_1, u_2, \ldots, x_N, u_o, u_1, \ldots, u_{N-1})$$

$$= \inf_{\mu_{N-1}} \sup_{z_1 \epsilon Z_1(u_o)} \cdots \sup_{z_{N-1} \epsilon Z_{N-1}(\zeta_{N-2}, u_{N-2})} E_N[\zeta_{N-1}, \mu_{N-1}(\zeta_{N-1})]$$

$$\leq \sup_{z_1 \epsilon Z_1(u_o)} \cdots \sup_{z_{N-1} \epsilon Z_{N-1}(\zeta_{N-2}, u_{N-2})} E_N[\zeta_{N-1}, \mu_{N-1,\epsilon}(\zeta_{N-1})]$$

(by using relation (5.25))

$$\leq \sup_{z_1 \epsilon Z_1(u_o)} \cdots \sup_{z_{N-1} \epsilon Z_{N-1}(\zeta_{N-2}, u_{N-2})} \inf_{\mu_{N-1}} E_N[\zeta_{N-1}, \mu_{N-1}(\zeta_{N-1})] + \epsilon$$

(by using the minimax inequality)

$$\leq \inf_{\mu_{N-1}} \sup_{z_1 \epsilon Z_1(u_o)} \cdots \sup_{z_{N-1} \epsilon Z_{N-1}(\zeta_{N-2}, u_{N-2})} E_N[\zeta_{N-1}, \mu_{N-1}(\zeta_{N-1})] + \epsilon$$

$$= J_{N-1}(\mu_o, \mu_1, \ldots, \mu_{N-2}) + \epsilon$$

Since these relations hold for any $\epsilon > 0$ we have equality throughout the above algebra proving equation (5.22)

$$J_{N-1}(\mu_o, \mu_1, \ldots, \mu_{N-2}) = \sup_{z_1 \epsilon Z_1(u_o)} \cdots \sup_{z_{N-1} \epsilon Z_{N-1}(\zeta_{N-2}, u_{N-2})} \inf_{\mu_{N-1}} E_N[\zeta_{N-1}, \mu_{N-1}(\zeta_{N-1})]$$

$$= \sup_{z_1 \epsilon \hat{Z}_1(u_o)} \cdots \sup_{z_{N-1} \epsilon \hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2})} H_{N-1}(\zeta_{N-1})$$

By using now equation (5.22) the equation (5.23) can be proved recursively for all k by using identical arguments as the ones used above to prove equation (5.22). The conclusion of the proposition then follows from equations (5.21) and (5.24). Q.E.D.

We remark that the value $H_k(\zeta_k)$ of the function H_k of equation (5.17) has the interpretation of the "cost-to-go" at time k from the point of view of the controller on the basis of the current information vector ζ_k. If the vector ζ_k is consistent with the constraint set Q and the system and measurement equations (as it will always be in any actual operation of the system) the value $H_k(\zeta_k)$ is either a real number or $+\infty$. In the case where the set Q is bounded a value of $+\infty$ indicates similarly to the related case discussed in Chapter 2 that there exists a disturbance selection policy (on the part of Nature) that can cause a violation of a state constraint regardless of the control law that the controller uses subject to the control constraints that the extended real valued function F of the cost functional (5.7) implies.

The optimal control law if it exists can be obtained from the algorithm as

$$\bar{\mu}_k(\zeta_k) = \bar{u}_k, \qquad k = 0, 1, \ldots, N-1$$

where $\bar{u}_k$ is the point where the infimum is attained in equation (5.17) for the fixed point ζ_k.

The dynamic programming algorithm of Proposition 5.1 can be profitably interpreted in terms of game theory, and in particular in terms

of multistage games of perfect information.[B11] The optimal value of the cost $\bar{J}$ can be viewed as the upper value (or min-max) of a game played by two opponents, the controller selecting the control law $\mu_o, \mu_1, \ldots, \mu_{N-1}$, and Nature selecting the uncertain quantities q from the set Q. The information based on which the decision of the controller is made, is fixed by the form of the functions μ_k, i.e., by the information vectors ζ_k. Since however only the upper value of the game is of interest here a variety of equivalent methods of selections of the vector q and corresponding information patterns can be assigned to Nature. One such information pattern and method for selection of the components of the vector q corresponds to the following sequence of events

(1) Controller selects u_o

(2) Nature selects z_1 from the set $Z_1(u_o)$

(3) Controller selects u_1

(4) Nature selects z_2 from the set $Z_2(z_1, u_o, u_1)$

- -

(2N-1) Controller selects u_{N-1}

(2N) Nature selects all the uncertain quantities q = $(x_o', w_o', w_1', \ldots, w_{N-1}', v_1', v_2', \ldots, v_{N-1}')'$ from the set $Q(\zeta_{N-1}, u_{N-1})$.

Each selection by either Controller or Nature is made with full knowledge of the outcomes of previous selections.

This sequence of events is fictitious, however it accurately reflects the sequence of events as viewed by the controller whose only information concerning the course of the game at time k is the information vector ζ_k, i.e., all measurements, and all control selections up to that time.

A moment's reflection shows that in fact the dynamic programming algorithm determines the (pure) value $\bar{J}$ of the game of perfect information described above. This value is the same as the optimal cost $\bar{J}$ of the Problem 5.1.

Finding the optimal cost $\bar{J}$ and the optimal control law from the dynamic programming algorithm of Proposition 5.1 is in general a very difficult task. Part of the difficulty stems from the fact that, loosely speaking, the objective of the controller is dual in nature; first to actuate the system in a favorable fashion and second to try to improve the quality of his estimate of the uncertainty in the system. This is a familiar situation from stochastic optimal control, known as dual control problem, [F1] the formidable complexities of which have been widely discussed in the literature. In stochastic optimal control insight into the structure of the optimal controller, and its dual function, can be obtained through the notion of a sufficient statistic.[St1] Similar insight will be obtained for the minimax controller of this chapter by introducing in the next section the related concept of a sufficiently informative function.

4. <u>Sufficiently Informative Functions</u>

Let us consider the following definition:

<u>Definition 5.1</u>: A function $S_k : R^{k(s+m)} \rightarrow \Sigma_k$ where Σ_k is some space will be called <u>sufficiently informative</u> with respect to Problem 5.1 if there exists a function $\bar{E}_{k+1} : \Sigma_k \times R^m \rightarrow [-\infty, +\infty]$ such that

$$\bar{E}_{k+1}[S_k(\zeta_k), u_k] = E_{k+1}(\zeta_k, u_k) \tag{5.26}$$

where E_{k+1} is the function defined in equations (5.16), (5.18) for k = 0, 1, ..., N-1.

The value of a sufficiently informative function at any point will be called sufficient information.

The clear consequence of the above definition is that if S_k is a sufficiently informative function and an optimal control law $\bar{\mu}_k$ exists, then this optimal control law can be implemented as the composition

$$\bar{\mu}_k(\zeta_k) = \bar{\mu}_k^* \cdot S_k(\zeta_k) \qquad (5.27)$$

where $\bar{\mu}_k^*$ is a suitable function which can be determined by minimizing the function $\bar{E}_{k+1}$ of equation (5.26) with respect to u_k. As a result the control at any time need only depend on the sufficient information $S_k(\zeta_k)$, and therefore if this sufficient information can be more easily generated or stored than the information vector ζ_k it may be advantageous to implement the control law in the form of equation (5.27).

Factorizations of the optimal control law into the composition of two functions as in equation (5.27) have been widely considered in stochastic optimal control theory, and are commonly referred to as separation theorems. In such problems the function S_k or its value is usually called a sufficient statistic. Particularly simple sufficient statistics have been found for problems involving a linear system, linear measurements and Gaussian white input and measurement noises.[St1] In other problems sufficient statistics of interest take the form of conditional probability density functions conditioned on the information available.[St1] Such sufficient statistics imply the factorization of the optimal control law into an estimator S_k computing the conditional probability density function of some

quantities, which may differ depending on the problem given, and an actuator $\bar{\mu}_k^*$ applying a control input to the system. In Chapter 4 it was demonstrated that in estimation problems which involve a set-membership description of the uncertainty the set of possible states consistent with the measurements received plays a role analogous to that of conditional probability density functions in stochastic estimation problems. Thus it should not come as a surprise that for the Problem 5.1 we will be able to derive sufficiently informative functions that involve sets of possible system states (or other quantities) consistent with the measurements received. In what follows we derive such sufficiently informative functions and further concentrate in the well behaved case of a linear system and an energy constraint on the uncertain quantities for which, as was demonstrated in Chapter 4, the set of possible states can be characterized by a finite set of numbers. We first introduce the following notation.

We denote for all k by

$$S_k(x_1, \ldots, x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1} | \zeta_k) \qquad (5.27)$$

the subset of $R^{kn+(N-k)r+(N-k-1)p}$ which consists of all vectors $(x_1, \ldots, x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1})$ that are consistent with the measurements $z_1, z_2, \ldots, z_k$, the control vectors $u_o, u_1, \ldots, u_{k-1}$, the system and measurement equations (5.1), (5.2) and the constraint set $q \in Q$.

We denote similarly by

$$S_k(x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1} | \zeta_k) \qquad (5.28)$$

$$S_k(x_1, x_2, \ldots, x_k | \zeta_k) \qquad (5.29)$$

$$S_k(x_k \mid \zeta_k) \tag{5.30}$$

the respective sets of all possible quantities within the parentheses that are consistent with the information vector ζ_k, the system and measurement equations, and the constraint $q \in Q$.

With the above notation we have the following proposition.

<u>Proposition 5.2</u>: A sufficiently informative function with respect to Problem 5.1 is the function

$$S_k : R^{k(s+m)} \rightarrow P(R^{kn+(N-k)r+(N-k-1)p}) \times R^{km}$$

given for all k by

$$S_k(\zeta_k) = [S_k(x_1, \ldots, x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1} \mid \zeta_k), u_o, u_1, \ldots, u_{k-1}] \tag{5.31}$$

<u>Proof</u>: Consider the function E_N of equation (5.18)

$$E_N(\zeta_{N-1}, u_{N-1}) = \sup_{q \in \hat{Q}(\zeta_{N-1}, u_{N-1})} F(x_1, x_2, \ldots, x_N, u_o, u_1, \ldots, u_{N-1})$$

This equation can also be written as

$$E_N(\zeta_{N-1}, u_{N-1})$$

$$= \sup_{(x_1, x_2, \ldots, x_{N-1}, w_{N-1}) \in S_{N-1}(x_1, \ldots, x_{N-1}, w_{N-1} \mid \zeta_{N-1})}$$

$$F[x_1, x_2, \ldots, x_{N-1}, \zeta_{N-1}(x_{N-1}, u_{N-1}, w_{N-1}), u_o, u_1, \ldots, u_{N-1}]$$

$$= \overline{E}_N[S_{N-1}(x_1, \ldots, x_{N-1}, w_{N-1} \mid \zeta_{N-1}), u_o, u_1, \ldots, u_{N-2}, u_{N-1}]$$

for a suitable function $\overline{E}_N$, proving that the function $S_{N-1}(\zeta_{N-1})$ of equation (5.31) is sufficiently informative according to Definition 5.1.

The function H_{N-1} of equation (5.18) can now be written as

$$H_{N-1}(\zeta_{N-1}) = \inf_{u_{N-1}} \overline{E}_N[S_{N-1}(x_1, \ldots, x_{N-1}, w_{N-1}|\zeta_{N-1}), u_0, u_1, \ldots, u_{N-2}, u_{N-1}]$$

$$= \overline{H}_{N-1}[S_{N-1}(x_1, \ldots, x_{N-1}, w_{N-1}|\zeta_{N-1}), u_0, u_1, \ldots, u_{N-2}]$$

for a suitable function $\overline{H}_{N-1}$.

Now for the function E_{N-1} of equation (5.16) we have

$$E_{N-1}(\zeta_{N-2}, u_{N-2}) = \sup_{z_{N-1} \in \hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2})}$$

$$\overline{H}_{N-1}[S_{N-1}(x_1, \ldots, x_{N-1}, w_{N-1}|\zeta_{N-1}), u_0, u_1, \ldots, u_{N-2}]$$
$$\tag{5.32}$$

The set $\hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2})$ can be described as

$$\hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2}) = \{z_{N-1}|z_{N-1} = h_{N-1}[f_{N-2}(x_{N-2}, u_{N-2}, w_{N-2}), v_{N-1}],$$

$$(x_{N-2}, w_{N-2}, v_{N-1}) \in S_{N-2}(x_{N-2}, w_{N-2}, v_{N-1}|\zeta_{N-2})\}$$
$$\tag{5.33}$$

where the set $S_{N-2}(x_{N-2}, w_{N-2}, v_{N-1}|\zeta_{N-2})$ is the set of all possible vectors $(x_{N-2}, w_{N-2}, v_{N-1})$ which are consistent with the measurements $z_1, z_2, \ldots, z_{N-2}$ and the control vectors $u_0, u_1, \ldots, u_{N-3}$ according to the system and measurement equations, and the constraint set Q.

The set $S_{N-2}(x_{N-2}, w_{N-2}, v_{N-1}|\zeta_{N-2})$ can be obtained as the projection of the set $S_{N-2}(x_1, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1}|\zeta_{N-2})$ on the space of vectors $(x_{N-2}, w_{N-2}, v_{N-1})$. Therefore the equation (5.33) can be written in the form

$$\hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2}) = \hat{Z}^*_{N-1}[S_{N-2}(x_1, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1}|\zeta_{N-2}), u_{N-2}]$$
$$\tag{5.34}$$

for a suitable function $\hat{Z}^*_{N-1}(\cdot, \cdot)$.

Also the set $S_{N-1}(x_1, \ldots, x_{N-1}, w_{N-1}|\zeta_{N-1})$ can be written as

$$S_{N-1}(x_1, \ldots, x_{N-1}, w_{N-1} | \zeta_{N-1}) = \{x_1, x_2, \ldots, x_{N-1}, w_{N-1} | \qquad (5.35)$$

$$x_{N-1} = f_{N-2}(x_{N-2}, u_{N-2}, w_{N-2}),$$

$$z_{N-1} = h_{N-1}(x_{N-1}, v_{N-1}), (x_1, x_2, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1})$$

$$\epsilon S_{N-2}(x_1, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1} | \zeta_{N-2})\}$$

$$= S_{N-1}^*[S_{N-2}(x_1, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1} | \zeta_{N-2}), z_{N-1}, u_{N-2}]$$

for a suitable function $S_{N-1}^*(\cdot, \cdot, \cdot)$.

By substitution of equations (5.34), (5.35) in equation (5.32) we obtain

$$E_{N-1}(\zeta_{N-2}, u_{N-2}) = \sup_{z_{N-1} \epsilon Z_{N-1}^* S_{N-2}(x_1, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1} | \zeta_{N-2}), u_{N-2}]}$$

$$\overline{H}_{N-1}[S_{N-1}^*[S_{N-2}(x_1, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1} | \zeta_{N-2}), z_{N-1}, u_{N-2}], u_0, \ldots, u_{N-2}]$$

$$= \overline{E}_{N-1}[S_{N-2}(x_1, \ldots, x_{N-2}, w_{N-2}, w_{N-1}, v_{N-1} | \zeta_{N-2}), u_0, u_1, \ldots, u_{N-2}]$$

for a suitable function $\overline{E}_{N-1}$, proving that the function $S_{N-2}(\zeta_{N-2})$ of equation (5.31) is sufficiently informative according to Definition 5.1.

By using identical arguments as above the function $S_k(\zeta_k)$ of equation (5.31) is proved to be sufficiently informative for all k. Q.E.D.

From the proof of the above proposition it can be easily seen that a simpler sufficiently informative function can be derived if some of the states x_i, i = 1, 2, .., N do not appear explicitly in the cost functional (5.7). Thus if, for example, the function F in equation (5.7) is of the form

$$F(x_1, x_2, \ldots, x_N, u_0, u_1, \ldots, u_{N-1}) = f(x_N, u_0, u_1, \ldots, u_{N-1}) \qquad (5.36)$$

then the function

$$S_k(\zeta_k) = [S_k(x_k, w_k, w_{k+1}, \ldots, w_{N-1}, v_{k+1}, v_{k+2}, \ldots, v_{N-1} | \zeta_k), u_o, u_1, \ldots, u_{k-1}]$$

(5.37)

is sufficiently informative.

Further simplifications result if the constraint set Q for the uncertain quantities has a property implying that the set of values that any particular uncertain quantity can take is independent of the values of the other uncertain quantities. We will consider the case where the set Q has the form

$$Q = \{x_o, w_o, w_1, \ldots, w_{N-1}, v_1, v_2, \ldots, v_{N-1} | \ x_o \epsilon X_o, w_i \epsilon W_i, \ i = 0, 1, \ldots, N-1,$$
$$v_k \epsilon V_k, \ k = 1, 2, \ldots, N-1\}$$

(5.38)

where X_o, W_i, V_k are given subsets of the corresponding Euclidean spaces. The case where the constraint Q is of the form (5.38) should be considered analogous to the case of white input and measurement noises in the corresponding stochastic problem. We have the following proposition:

Proposition 5.3: Assume that the constraint set Q has the form of equation (5.38). Then the function S_k given for all k by

$$S_k(\zeta_k) = [S_k(x_1, x_2, \ldots, x_k | \zeta_k), u_o, u_1, \ldots, u_{N-1}]$$

(5.39)

is sufficiently informative.

If the function F in equation (5.7) has the form of equation (5.36) then the function S_k given for all k by

$$S_k(\zeta_k) = [S_k(x_k | \zeta_k), u_o, u_1, \ldots, u_{N-1}]$$

(5.40)

is sufficiently informative.

Proof: The proof follows by trivial modifications of the proof of Proposition 5.2 to take into account the special structure of the set Q in equation (5.38). Q.E.D.

The above propositions clearly illustrate the dual function of the optimal controller. By equation (5.27) the optimal control law is of the form

$$\bar{\mu}_k = \bar{\mu}_k^* \cdot S_k \tag{5.41}$$

i.e., it is the composition of the sufficiently informative function S_k and the function $\bar{\mu}_k^*$. The function S_k in the case of Propositions 5.2 and 5.3 represents an estimator and the function $\bar{\mu}_k^*$ represents the actuator. Alternatively the optimal controller can be viewed as being composed of two cascaded parts. The first part produces an estimate set and the second part accepts as input this estimate set and produces a control vector. This control vector is stored and recalled in the future by the controller.

It should be noted that there is an important difference between the sufficiently informative functions derived for the Problem 5.1, and sufficient statistics for the corresponding stochastic optimal control problem in that the possible additivity of the cost functional (5.7) results in no simplification for the sufficiently informative function. Thus the function S_k of equation (5.31) cannot in general be simplified if the function F in the cost functional (5.7) is of the form

$$F(x_1, x_2, \ldots, x_N, u_o, u_1, \ldots, u_{N-1}) = \sum_{i=1}^{N} f_i(x_i, u_{i-1})$$

where f_i and g_{i-1} are given functions. In the corresponding stochastic problem however important simplifications in the sufficient statistic result.[St1]

This difference is due to the fact that whereas the expectation operation is linear and distributes over addition the maximization operation is not.

The equation (5.41) which demonstrates the structure of the optimal control law as the composition of an estimator and an actuator can provide insight concerning the complexity and the implementation of this optimal control law. For example as was illustrated in the previous chapter a case where the estimator has convenient structure is the case of a linear system and an energy constraint for the uncertain quantities. The sufficiently informative functions of Propositions 5.2, 5.3 for this case can be characterized by a small set of numbers and can be computed recursively. As an illustration we state the following proposition.

<u>Proposition 5.4</u>: Consider the special case of Problem 5.1 where the system is linear

$$x_{k+1} = A_k x_k + B_k u_k + G_k w_k, \qquad k = 0, 1, .., N-1 \qquad (5.42)$$

with linear measurements

$$z_k = C_k x_k + v_k, \qquad k = 1, 2, .., N-1 \qquad (5.43)$$

and the set Q for the uncertain quantities is specified by the energy constraint

$$x_o' \Psi^{-1} x_o + \sum_{i=0}^{N-1} w_i' Q_i^{-1} w_i + \sum_{i=1}^{N-1} v_i' R_i^{-1} v_i \leq 1 \qquad (5.44)$$

where Ψ, Q_i, R_i are positive definite symmetric matrices for all i.

Assume also that the cost functional is of the form of equation (5.36)

$$F(x_1, x_2, ..., x_N, u_o, u_1, ..., u_{N-1}) = f(x_N, u_o, u_1, ..., u_{N-1}) \qquad (5.36)$$

Then the function $S_k : R^{k(s+m)} \rightarrow R^n \times [0, 1] \times R^{km}$ given by

$$S_k(\zeta_k) = [\hat{x}_k, \delta^2(k), u_0, u_1, \ldots, u_{k-1}] \tag{5.45}$$

is sufficiently informative for all k, where the n-vector $\hat{x}_k$ and the scalar $\delta^2(k)$ are generated recursively by the estimator equations

$$\hat{x}_{i+1} = A_i\hat{x}_i + B_iu_i + \Sigma_{i+1|i+1}C'_{i+1}R^{-1}_{i+1}(z_{i+1} - C_{i+1}A_i\hat{x}_i - C_{i+1}B_iu_i)$$

$$i = 0, 1, \ldots, k-1 \tag{5.46}$$

$$\hat{x}_o = 0 \tag{5.47}$$

$$\Sigma_{i|i} = [\Sigma^{-1}_{i|i-1} + C'_iR^{-1}_iC_i]^{-1} \tag{5.48}$$

$$\Sigma_{i|i-1} = A_{i-1}\Sigma_{i-1|i-1}A'_{i-1} + G_{i-1}Q_{i-1}G'_{i-1} \tag{5.49}$$

$$\Sigma_{o|o} = \Psi \tag{5.50}$$

$$\delta^2(i+1) = \delta^2(i) + (z_{i+1} - C_{i+1}A_i\hat{x}_i - C_{i+1}B_iu_i)'(C_{i+1}\Sigma_{i+1|i}C'_{i+1} + R_{i+1})^{-1}$$

$$(z_{i+1} - C_{i+1}A_i\hat{x}_i - C_{i+1}B_iu_i) \tag{5.51}$$

$$\delta^2(0) = 0 \tag{5.52}$$

Proof: From the results in Chapter 4 we obtain that the set $S_k(x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1} | \zeta_k)$ is given by

$$S_k(x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1} | \zeta_k)$$

$$= \{x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1} | (x_k - \hat{x}_k)'\Sigma^{-1}_{k|k}(x_k - \hat{x}_k)$$

$$+ \sum_{i=k}^{N-1} w'_iQ^{-1}_iw_i + \sum_{i=k+1}^{N-1} v'_iR^{-1}_iv_i \leq 1 - \delta^2(k)\} \tag{5.53}$$

Since the matrix $\Sigma_{k|k}$ is precomputable and the matrices Q_i and R_i are given the set $S_k(x_k, w_k, \ldots, w_{N-1}, v_{k+1}, \ldots, v_{N-1} | \zeta_k)$ is completely determined from

the vector $\hat{x}_k$ and the scalar $\delta^2(k)$. By combining the equation (5.37) and the Definition 5.1 the result follows. Q.E.D.

Thus for the problem of the above proposition the estimator part of the optimal controller can be completely and efficiently characterized. For the reachability case of this problem, i.e., the case where

$$f(x_N, u_0, u_1, \ldots, u_{N-1}) = \delta(x_N | X_N) + \sum_{i=0}^{N-1} \delta(u_i | U_i)$$

with X_N, U_i given sets, the actuator part of the optimal controller can also be completely characterized as we will demonstrate in the next chapter. Similar results with the Proposition 5.4 can be obtained for a general cost functional of the form of equation (5.7). However in this case, as can be seen from the Proposition 5.2 and the results in Chapter 4, the sufficiently informative function will be of the form

$$S_k(\zeta_k) = [\hat{x}_{1|k}, \hat{x}_{2|k}, \ldots, \hat{x}_{k|k}, \delta^2(k), u_0, u_1, \ldots, u_{N-1}]$$

where $\hat{x}_{i|k}$ will be given by smoothing equations for all $i < k$.

For the case of a linear system with instantaneous ellipsoidal constraints on the uncertain quantities the sufficiently informative functions of the Propositions 5.2 and 5.3 cannot be characterized by a finite set of numbers neither can they be easily generated by an estimator as was demonstrated in the previous chapter. This indicates that for such problems the characterization of the optimal control law should be in general very difficult. However for the problem of the reachability of a target tube which involves such constraints on the uncertain quantities a method for obtaining suboptimal controllers that can be more easily implemented will be developed in the next chapter.

5.   <u>Discussion and Sources</u>

The basic method for the solution of minimax control problems
with imperfect state information is the dynamic programming algorithm
of Proposition 5.1. It should be noted that both in the development and
the proof of this algorithm we did not make use of the fact that the state
space, control space and disturbance spaces are Euclidean spaces and in
fact the Proposition 5.1 can be generalized for the case where these spaces
are arbitrary sets.

In general the solution of the problem by dynamic programming is
a very difficult task, and only for simple systems and simple constraint
sets such a solution can be practical.

It appears that the most well behaved special case of Problem 5.1
is the case where the system is linear and the set of the uncertain quantities
is specified by an energy constraint. For this case a sufficiently informative
function can be recursively generated by estimators developed earlier in
Chapter 4. Even for this case the actuator part of the optimal controller
may not be easily characterized. In the next chapter we will give a pre-
cise characterization of the actuator for the special case of a reachability
problem. However it appears that there is no special case of the Problem
5.1 with a solution as elegant as that provided by the separation theorem
of stochastic optimal control for linear systems and quadratic criteria.[J1],[G1]

The only source for minimax control problems with imperfect state
information appears to be the original work of Witsenhausen,[W1] who dem-
onstrated the application of dynamic programming to this problem. The
use of dynamic programming in sequential games has been known at least

since the proof that finite games of perfect information have a saddle point.[V1] The concept and the development of the Proposition 5.1 is based on game theory considerations, and it involves the construction of a sequential game of perfect information in its extensive form.[Kul] The algorithm of Proposition 5.1 differs in its form and is more general than Witsenhausen's algorithm however the same basic ideas are involved. The notion of a sufficiently informative function is introduced for the first time here in analogy with the notion of a sufficient statistic of stochastic control. It has mainly theoretical value in that it forms the basis for demonstrating the decomposition of the optimal controller into an estimator and an actuator. This decomposition provides insight into the structure of the optimal controller, and in some cases it can serve as a starting point for developing suboptimal control schemes.

CHAPTER 6

SOME REACHABILITY PROBLEMS WITH IMPERFECT STATE INFORMATION

1.   General Remarks

As was demonstrated in the previous chapter it is in general very difficult to characterize completely the optimal controller in minimax control problems with imperfect state information. For the case of a linear system with an energy constraint for the uncertain quantities it was shown, however, that the optimal controller may be realized as an estimator followed by an actuator and that the estimator can be easily and efficiently characterized using the results of Chapter 4. We shall demonstrate in the next section that for the case of the problem of the reachability of the target set the actuator part of the optimal controller can also be precisely characterized, and thus we shall give a complete solution to this problem. However the implementation of the optimal controller will still be quite difficult despite the simplification achieved.

In Section 3 we will consider a problem of reachability of a target tube which involves a linear system and instantaneous ellipsoidal constraints for the uncertain quantities. For this problem it appears that, in general, a practical implementation of the optimal control law is indeed very difficult. For this reason we present a suboptimal control scheme for this problem by making use of the bounding ellipsoid estimation algorithm presented earlier in Chapter 4.

2.   Reachability of a Target Set for the Case of Energy Constraints

We first formulate the reachability problem that we will consider in this section.

Problem 6.1: Consider the linear discrete-time dynamic system

$$x_{k+1} = A_k x_k + B_k u_k + G_k w_k, \qquad k = 0, 1, .., N-1 \qquad (6.1)$$

with the linear measurements

$$z_k = C_k x_k + v_k, \qquad k = 1, 2, ..., N-1 \qquad (6.2)$$

where $x_k \epsilon R^n$, $k = 0, 1, .., N$, is the state vector, $u_k \epsilon R^m$, $k = 0, 1, .., N-1$, is the control vector, $w_k \epsilon R^r$, $k = 0, 1, .., N-1$, is the input disturbance vector, $z_k \epsilon R^s$, $k = 1, 2, .., N-1$, is the measurement vector, $v_k \epsilon R^p$, $k = 1, 2, .., N-1$ is the measurement disturbance vector, and A_k, B_k, G_k, C_k are given matrices of appropriate dimension.

The initial state x_o, and the disturbances w_k, v_k are assumed unknown except that they satisfy the energy constraint

$$x_o' \Psi^{-1} x_o + \sum_{k=1}^{N} (w_{k-1}' Q_{k-1}^{-1} w_{k-1} + v_k' R_k^{-1} v_k) \leq 1 \qquad (6.3)$$

where Ψ, Q_{k-1}, R_k, $k = 1, 2, .., N$ are given positive definite matrices.

Attention is restricted to control laws of the form

$$\mu_k : R^{k(s+m)} \rightarrow U_k, \qquad k = 0, 1, ..., N-1$$

taking values

$$u_k = \mu_k(z_1, z_2, ..., z_k, u_o, u_1, ..., u_{k-1}), \quad k = 0, 1, .., N-1$$

where μ_o is interpreted as a constant vector and where $U_k \subset R^m$, $k = 0$, $1, \ldots, N-1$, are given sets. It is required to find a control law in this class such that the final state x_N of the resulting closed-loop system belongs to a given set $X_N \subset R^n$ for all possible values of the uncertain quantities.

We will say that the target set X_N is reachable if there exists such a control law.

The above problem can be recognized as the special case of Problem 5.1 of the previous chapter where the system and measurements are linear, the constraint set Q is specified by the energy constraint (6.3) and the cost functional is

$$J(\mu_o, \mu_1, \ldots, \mu_{N-1}) = \sup_{q \in Q} [\delta(x_N | X_N) + \sum_{i=0}^{N-1} \delta(u_i | U_i)] \qquad (6.4)$$

where $\delta(y | Y)$ denotes the indicator function of the set Y. Consequently it follows by Proposition 5.4 that for this problem the optimal control law is of the form

$$\mu_k(z_1, z_2, \ldots, z_k, u_o, u_1, \ldots, u_{k-1}) = \mu_k^*[\hat{x}_k, \delta^2(k), u_o, u_1, \ldots, u_{k-1}] \qquad (6.5)$$

where $\hat{x}_k$ and $\delta^2(k)$ are given for all k by the estimator equations (5.46) through (5.52). We will now characterize the optimal control law (6.5) in the following proposition.

Proposition 6.1: Consider the sets $\hat{X}_k \subset R^n \times [0,1]$, $k = 0, 1, \ldots, N-1$ defined recursively by the relations

$$\hat{X}_{N-1} = \{\hat{x}_{N-1}, \delta^2(N-1) | \quad \exists \quad u_{N-1} \epsilon U_{N-1} \quad \text{such that}$$

$$(A_{N-1}\hat{x}_{N-1} + B_{N-1}u_{N-1} + G_{N-1}w_{N-1}) \epsilon X_N, \quad \forall x_{N-1}, w_{N-1},$$

$$(x_{N-1} - \hat{x}_{N-1})' \Sigma_{N-1|N-1}^{-1}(x_{N-1} - \hat{x}_{N-1}) + w_{N-1}' Q_{N-1}^{-1} w_{N-1} \le 1 - \delta^2(N-1)\}$$

$$(6.6)$$

$$\hat{X}_k = \{\hat{x}_k, \delta^2(k) | \quad \exists \quad u_k \epsilon U_k \quad \text{such that}$$

$$[(A_k\hat{x}_k + B_k u_k + \Sigma_{k+1|k+1} C_{k+1}' R_{k+1}^{-1} d_k),$$

$$(\delta^2(k) + d_k'(C_{k+1}\Sigma_{k+1|k}C_{k+1}' + R_{k+1})^{-1}d_k)] \epsilon \hat{X}_{k+1},$$

$$\forall d_k, \quad d_k'(C_{k+1}\Sigma_{k+1|k}C_{k+1}' + R_{k+1})^{-1}d_k \le 1 - \delta^2(k)\} \quad (6.7)$$

$$k = 0, 1, \ldots, N-2$$

where the matrices $\Sigma_{k+1|k+1}$, $\Sigma_{k+1|k}$ are given for all k by the equations (5.48) through (5.50).

Then the target set X_N is reachable if and only if

$$(0, 0) \epsilon \hat{X}_o \qquad (6.8)$$

Under these circumstances a control law $\{\bar{\mu}_o, \bar{\mu}_1, \ldots, \bar{\mu}_{N-1}\}$ that achieves reachability can be obtained as

$$\bar{u}_k = \bar{\mu}_k(z_1, \ldots, z_k, u_o, \ldots, u_{k-1}) = \bar{\mu}_k^*[\hat{x}_k, \delta^2(k)]$$

where for each pair $[\hat{x}_k, \delta^2(k)] \epsilon \hat{X}_k$ the vector $\bar{u}_k$ is such that $\bar{u}_k \epsilon U_k$ and if k = N-1, we have $(A_{N-1}x_{N-1} + B_{N-1}\bar{u}_{N-1} + G_{N-1}w_{N-1}) \epsilon X_N$ for all x_{N-1}, w_{N-1} with $(x_{N-1} - \hat{x}_{N-1})' \Sigma_{N-1|N-1}^{-1}(x_{N-1} - \hat{x}_{N-1}) + w_{N-1}' Q_{N-1}^{-1} w_{N-1} \le 1 - \delta^2(N-1)$, if k = 0, 1, .., N-2, we have $[(A_k\hat{x}_k + B_k\bar{u}_k + \Sigma_{k+1|k+1}C_{k+1}'R_{k+1}^{-1}d_k),$

$(\delta^2(k) + d_k'(C_{k+1}\Sigma_{k+1|k}C_{k+1}' + R_{k+1})^{-1}d_k)] \epsilon \hat{X}_{k+1}$ for all d_k with

$d_k'(C_{k+1}\Sigma_{k+1|k}C_{k+1}' + R_{k+1})^{-1}d_k \le 1 - \delta^2(k)$

<u>Proof:</u> We shall use the dynamic programming algorithm of Proposition 5.1 for the cost functional (6.4), and the equations (5.46) through (5.53) in Proposition 5.4. We have from (5.17)

$$
H_{N-1}(\zeta_{N-1}) = \inf_{u_{N-1}} \sup_{q \in \hat{Q}(\zeta_{N-1}, u_{N-1})} [\delta(A_{N-1}x_{N-1} + B_{N-1}u_{N-1}
$$

$$
+ G_{N-1}w_{N-1} | X_N) + \sum_{i=0}^{N-1} \delta(u_i | U_i)]
$$

$$
= \inf_{u_{N-1}} \sup_{(x_{N-1}, w_{N-1}) \in S_{N-1}(x_{N-1}, w_{N-1} | \zeta_{N-1})} [\delta(A_{N-1}x_{N-1} + B_{N-1}u_{N-1}
$$

$$
+ G_{N-1}w_{N-1} | X_N) + \sum_{i=0}^{N-1} \delta(u_i | U_i)]
$$

Since by equation (5.53)

$$
S_{N-1}(x_{N-1}, w_{N-1} | \zeta_{N-1}) = \{x_{N-1}, w_{N-1} | (x_{N-1} - \hat{x}_{N-1})' \Sigma_{N-1|N-1}^{-1} (x_{N-1} - \hat{x}_{N-1})
$$

$$
+ w_{N-1}' Q_{N-1}^{-1} w_{N-1} \leq 1 - \delta^2(N-1)\}
$$

we have that for every ζ_{N-1} such that $S_{N-1}(x_{N-1}, w_{N-1} | \zeta_{N-1}) \neq \phi$

$$
H_{N-1}(\zeta_{N-1}) = \delta[\hat{\hat{x}}_{N-1}, \delta^2(N-1) | \hat{X}_{N-1}] + \sum_{i=0}^{N-2} \delta(u_i | U_i)
$$

where $\hat{\hat{X}}_{N-1}$ is the set defined in equation (6.6).

Using now equations (5.17), (5.18) we have

$$
H_{N-2}(\zeta_{N-2}) = \inf_{u_{N-2}} \sup_{z_{N-1} \in \hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2})} \{\delta[\hat{\hat{x}}_{N-1}, \delta^2(N-1) | \hat{X}_{N-1}]
$$

$$
+ \sum_{i=0}^{N-2} \delta(u_i | U_i)\} \tag{6.9}
$$

for every ζ_{N-2} such that $\hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2}) \neq \phi$. According to the equations (5.46), (5.51) we write

$$\hat{x}_{N-1} = A_{N-2}\hat{x}_{N-2} + B_{N-2}u_{N-2} + \Sigma_{N-1|N-1}C'_{N-1}R^{-1}_{N-1}d_{N-2} \quad (6.10)$$

$$\delta^2(N-1) = \delta^2(N-2) + d'_{N-2}(C_{N-1}\Sigma_{N-1|N-2}C'_{N-1} + R_{N-1})^{-1}d_{N-2} \quad (6.11)$$

where

$$d_{N-2} = z_{N-1} - C_{N-1}A_{N-2}\hat{x}_{N-2} - C_{N-1}B_{N-2}u_{N-2} \qquad (6.12)$$

Also it can be easily proved using the results in Chapter 4 that the set $\hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2})$ is given by

$$\hat{Z}_{N-1}(\zeta_{N-2}, u_{N-2}) = \{z_{N-1} \mid d'_{N-2}(C_{N-1}\Sigma_{N-1|N-2}C_{N-1} + R_{N-1})^{-1}d_{N-2} \leq 1-\delta^2(N-2)\}$$

where d_{N-2} is given by (6.12).

From equations (6.9) through (6.12) it follows that

$$H_{N-2}(\zeta_{N-2}) = \delta[\hat{x}_{N-2}, \delta^2(N-2)|\hat{X}_{N-2}] + \sum_{i=0}^{N-3} \delta(u_i|U_i)$$

where $\hat{X}_{N-2}$ is defined in equation (6.7).

By proceeding similarly we obtain that the optimal value of the cost functional (6.4) is

$$\bar{J} = \delta[\hat{x}_o, \delta^2(0)|\hat{X}_o]$$

Since we have $x_o = 0$, $\delta^2(0) = 0$ we obtain

$$\bar{J} = 0 \quad \Longleftrightarrow \quad (0, 0) \in \hat{X}_o$$

$$\bar{J} = \infty \quad \Longleftrightarrow \quad (0, 0) \notin \hat{X}_o$$

Since we have $\bar{J} = 0$ if and only if the target set X_N is reachable the condition (6.8) is proved.

The fact that the optimal control law is of the form indicated in the proposition can be easily seen from the preceding arguments. Q.E.D.

A closer examination of the Proposition 6.1 reveals the following mechanism for the optimal controller. First the estimator of equations (5.46) through (5.52)

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + \Sigma_{k+1|k+1} C'_{k+1} R^{-1}_{k+1} d_k \tag{6.13}$$

$$\delta^2(k+1) = \delta^2(k) + d'_k (C_{k+1} \Sigma_{k+1|k} C'_{k+1} + R_k)^{-1} d_k \tag{6.14}$$

$$d_k = z_{k+1} - C_{k+1} A_k \hat{x}_k - C_{k+1} B_k u_k \tag{6.15}$$

is used to generate the sufficiently informative function which in this case is $S_k(\zeta_k) = [\hat{x}_k, \delta^2(k)]$. Then for any given $[\hat{x}_k, \delta^2(k)] \in \hat{X}_k$ the controller selects the control vector u_k in a way that the "state" $[\hat{x}_{k+1}, \delta^2(k+1)]$ of the estimator (6.13), (6.14) at the next time instant $(k+1)$ will belong to the set $\hat{X}_{k+1}$ for all possible values of the error residual vector d_k

$$d_k = z_{k+1} - C_{k+1} A_k \hat{x}_k - C_{k+1} B_k u_k$$

Thus in effect the optimal controller operates in a way that achieves reachability of the set $\hat{X}_{k+1}$ by the sufficient information $[\hat{x}_{k+1}, \delta^2(k+1)]$ and eventually reachability of the set $\hat{X}_{N-1}$ by the sufficient information $[\hat{x}_{N-1}, \delta^2(N-1)]$. The sufficient information $[\hat{x}_k, \delta^2(k)]$ can be viewed as the state of the (n+1)-dimensional system defined by the equations (6.13), (6.14) with the initial state

$$[\hat{x}_o, \delta^2(0)] = (0, 0)$$

This system is driven by the control u_k, and perturbed by the disturbance

vector d_k of equation (6.15), and it is linear in the state and the control but nonlinear in the disturbance. Furthermore the disturbance d_k satisfies at each time the constraint

$$d_k'(C_{k+1}\Sigma_{k+1|k}C_{k+1}' + R_{k+1})^{-1}d_k \leq 1 - \delta^2(k) \tag{6.16}$$

which is a <u>state-dependent</u> constraint.

We can conclude from the above discussion that in effect the solution of the Problem 6.1 involves the solution of a target set reachability problem with perfect state information. This reachability problem however does not involve the original system (6.1) but instead it involves the (n+1)-dimensional estimator described by the equations (6.13), (6.14) the state of which is the sufficient information $[\hat{x}_k, \delta^2(k)]$. The objective of the controller is to achieve reachability of the target set $\hat{X}_{N-1}$ by the final state $[\hat{x}_{N-1}, \delta^2(N-1)]$ of this estimator since if $[\hat{x}_{N-1}, \delta^2(N-1)]\epsilon \hat{X}_{N-1}$ the reachability of the target set X_N can be guaranteed by equation (6.6). Since the controller can use the estimator which produces at each time k the sufficient information $[\hat{x}_k, \delta^2(k)]$ this is a reachability problem with perfect state information. However this problem is more complicated than the reachability problems that we considered in Chapter 3 since the disturbance d_k of equation (6.15) enters nonlinearly in the equation (6.14), and the constraint (6.16) is state dependent. For this reason the construction of the sets $\hat{X}_k$, $k = 0, 1, .., N-1$, is considerably more complicated than the construction of the effective and modified target sets that we considered in Chapter 3. As a result the implementation of the optimal controller of Proposition 6.1 is very difficult in general. By using however internal approximations to the sets $\hat{X}_k$ it is possible to derive suboptimal control

schemes that achieve reachability, and can be more easily implemented. We shall not pursue this matter here since our primary objective in this section has been to demonstrate the interesting fact that, for the case that we consider, the problem of state reachability with imperfect information is equivalent to an estimate reachability problem with perfect information.

3. Reachability of a Target Tube with Instantaneous Ellipsoidal Constraints

In this section we consider the natural extension of the problem of the reachability of a target tube that was considered in Chapter 3 to the case where, instead of having perfect knowledge of the system state, the controller has access only to noise-corrupted measurements of the system output. We will examine the case of a linear system and instantaneous ellipsoidal constraints for the uncertain quantities. As was explained in the previous chapter the implementation of the optimal controller for this problem is in general very difficult. Our objective in this section will be to develop suboptimal control schemes that achieve reachability of the target tube, and that can be more practically implemented.

We will consider the linear system of equation (6.1)

$$x_{k+1} = A_k x_k + B_k u_k + G_k w_k , \qquad k = 0, 1, .., N-1 \qquad (6.1)$$

with the linear measurements

$$z_k = C_k x_k + v_k , \qquad k = 1, 2, .., N-1 \qquad (6.2)$$

We assume that the initial state x_o and the disturbance vectors w_k, v_k satisfy the constraints

$$x_o' \Psi^{-1} x_o \leq 1 \tag{6.17a}$$

$$w_k' Q_k^{-1} w_k \leq 1, \qquad k = 0, 1, .., N-1 \tag{6.17b}$$

$$v_k' R_k^{-1} v_k \leq 1, \qquad k = 1, 2, .., N-1 \tag{6.17c}$$

where Ψ, Q_k, R_k are given positive definite matrices.

We are seeking a control law

$$\mu_k = R^{k(s+m)} \to U_k \qquad k = 1, 2, .., N-1$$

$$u_k = \mu_k(z_1, z_2, .., z_k, u_o, u_1, .., u_{k-1})$$

$$u_o = \mu_o \epsilon U_o$$

where U_k, $k = 0, 1, .., N-1$, are given sets, which is such that the state x_k, of the resulting closed-loop system (6.1) is contained in the given sets X_k, $k = 1, 2, .., N$, for each k, and for all possible values of the initial state x_o and the input and measurement distrubances w_k, v_k which satisfy the constraints (6.17).

We shall say that the target tube $\{X_1, X_2, .., X_N\}$ is reachable if there exists such a control law.

Given at time k the measurements $z_1, z_2, .., z_k$ and the prior controls $u_o, u_1, .., u_{k-1}$, the controller can calculate a bounding ellipsoid $X_{k|k}^*$ to the set $X_{k|k}$ of all possible states x_k consistent with these measurements and controls by using the bounding ellipsoid estimator of Proposition 4.3. By taking into account the presence of control vectors in Proposition 4.3 we have for all k:

$$X_{k|k}^* = \{x_k | (x_k - \hat{x}_k)' \Sigma_{k|k}^{-1} (x_k - \hat{x}_k) \leq 1 - \delta^2(k)\} \tag{6.18}$$

where $\Sigma_{k|k}$, $\hat{x}_k$, $\delta^2(k)$ are given recursively by:

$$\Sigma_{i|i} = [(1-\rho_i)\Sigma_{i|i-1}^{-1} + \rho_i C_i' R_i^{-1} C_i]^{-1} \tag{6.19}$$

$$\Sigma_{i|i-1} = (1-\beta_{i-1})^{-1} A_{i-1} \Sigma_{i-1|i-1} A_{i-1}' + \beta_{i-1}^{-1} B_{i-1} Q_{i-1} B_{i-1}' \tag{6.20}$$

$$\Sigma_{o|o} = \Psi \tag{6.21}$$

$$\hat{x}_{i+1} = A_i \hat{x}_i + B_i u_i + \rho_{i+1} \Sigma_{i+1|i+1} C_{i+1}' R_{i+1}^{-1}(z_{i+1} - C_{i+1} A_i \hat{x}_i - C_{i+1} B_i u_i) \tag{6.22}$$

$$\hat{x}_o = 0 \tag{6.23}$$

$$\delta^2(i) = (1-\beta_{i-1})(1-\rho_i)\delta^2(i-1) + (z_i - C_i A_{i-1}\hat{x}_{i-1} - C_i B_{i-1} u_{i-1})'$$
$$[(1-\rho_i)^{-1} C_i \Sigma_{i|i-1} C_i' + \rho_i^{-1} R_i]^{-1}(z_i - C_i A_{i-1}\hat{x}_{i-1} - C_i B_{i-1} u_{i-1}) \tag{6.24}$$

$$\delta^2(0) = 0 \tag{6.25}$$

and β_{i-1}, ρ_i, $i = 1, 2, .., N$, are any real numbers with $0 < \beta_{i-1} < 1$, $0 < \rho_i < 1$.

Consider now the ellipsoid

$$S_k = \{x | x'\Sigma_{k|k}^{-1} x \leq 1\} , \qquad k = 1, 2, .., N \tag{6.26}$$

The ellipsoid S_k is precomputable, since the matrix $\Sigma_{k|k}$ does not depend on the measurements, and expresses the maximum possible ($\delta^2(k) = 0$) amount of uncertainty about the state x_k when the estimate $\hat{x}_k$ is known. Since from equations (6.18), (6.26) we have

$$X_{k|k} \subset X_{k|k}^* \subset \hat{x}_k + S_k \tag{6.27}$$

it is clear that in order for the state x_k to belong to the set X_k it is sufficient that the state estimate $\hat{x}_k$ belongs to the set

$$\hat{X}_k = \{\hat{x}_k | \hat{x}_k + S_k \subseteq X_k\} \tag{6.28}$$

Thus for the purpose of obtaining sufficient conditions for reachability, we can shift emphasis from the problem of the reachability of the target tube $\{X_1, X_2, \ldots, X_N\}$ by the system state x_k, to the problem of the reachability of the target tube $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N\}$ by the state estimate $\hat{x}_k$. This latter problem will be shown to be a reachability problem with perfect state information of the form that we have already considered in Chapter 3.

By substituting equations (6.1), (6.2) into (6.22) we have that the estimate x_k is generated by the equation

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + L_{k+1} d_k \tag{6.29}$$

where the lumped disturbance d_k is given by

$$d_k = C_{k+1} A_k (x_k - \hat{x}_k) + C_{k+1} G_k w_k + v_{k+1} \tag{6.30}$$

and the (precomputable) gain matrix L_{k+1} is given by

$$L_{k+1} = \rho_{k+1} \Sigma_{k+1|k+1} C'_{k+1} R_{k+1}^{-1} \tag{6.31}$$

Furthermore it follows immediately from equation (6.30) that d_k belongs to the known set

$$D_k = C_{k+1} A_k S_k + C_{k+1} G_k W_k + V_{k+1} \tag{6.32}$$

where S_k is defined by (6.26) and the ellipsoids W_k, V_{k+1} are specified by the constraints (6.17b) and (6.17c)

$$W_k = \{w_k \,|\, w_k' Q_k^{-1} w_k \leq 1\}$$

$$V_{k+1} = \{v_{k+1} \,|\, v_{k+1}' R_{k+1}^{-1} v_{k+1} \leq 1\}$$

Thus a sufficient condition for the reachability of the given target tube $\{X_1, X_2, \ldots, X_N\}$ by the system state x_k in the presence of imperfect information is that the target tube $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N\}$ defined by (6.28) be reachable by the state $\hat{x}_k$ of the estimator (6.29) in the presence of the disturbances d_k which belong to the known set D_k of equation (6.32). Since the estimate $\hat{x}_k$ is generated by the controller and known to him at each time k, we are faced with a target tube reachability problem with perfect state information of the form that was examined in Chapter 3.

The solution of this estimate reachability problem can be given using the results of Chapter 3. Define analogously to equations (3.5), through (3.8) the effective target sets $\hat{T}_{k+1}$, and the modified target sets $\hat{X}_k^*$ by the equations

$$\hat{X}_N^* = \hat{X}_N \tag{6.33}$$

$$\hat{T}_{k+1} = \{\hat{x} \,|\, (\hat{x} + L_{k+1} D_k) \subset \hat{X}_{k+1}^*\}, \qquad k = 0, 1, \ldots, N-1 \tag{6.34}$$

$$\hat{X}_k^* = \{\hat{x}_k \,|\, (A_k \hat{x}_k + B_k u_k) \in \hat{T}_{k+1}, \text{ for some } u_k \in U_k\} \cap \hat{X}_k \tag{6.35}$$
$$k = 1, 2, \ldots, N$$

$$\hat{X}_o^* = \{\hat{x}_o \,|\, (A_o \hat{x}_o + B_o u_o) \in \hat{T}_1, \text{ for some } u_o \in U_o\} \tag{6.36}$$

Then the target tube $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N\}$ is reachable by the state $\hat{x}_k$ of the estimator (6.29) if and only if

$$\hat{x}_o = 0 \in \hat{X}_o^*$$

Since the reachability of the target tube $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N\}$ is a sufficient condition for the reachability of the target tube $\{X_1, X_2, \ldots, X_N\}$ by the state x_k of the system (6.1) in the presence of imperfect information we have the following proposition.

Proposition 6.2: A sufficient condition for reachability of the target tube $\{X_1, X_2, \ldots, X_N\}$ is that

$$0 \in \hat{X}_o^*$$

where the set $\hat{X}_o^*$ is given recursively by equations (6.33) through (6.36).

The control law that achieves reachability of the target tube $\{X_1, X_2, \ldots, X_N\}$ is the one that achieves reachability of the target tube $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N\}$ by the estimate $\hat{x}_k$ of the estimator (6.29), and it is of the form

$$u_k = \mu_k(\hat{x}_k) , \qquad k = 0, 1, \ldots, N-1$$

A possible method for obtaining such a control law which in addition is linear, is to make use of the ellipsoidal algorithm of Sections 3, 4 in Chapter 3 assuming that the sets $\hat{X}_k$ and U_k are, or can be approximated by, ellipsoids. In this case we can use the following ellipsoid D_k^* which bounds the disturbance set D_k of equation (6.32)

$$D_k^* = \{d_k \mid d_k'[(1-\rho_{k+1})^{-1}C_{k+1}\Sigma_{k+1|k}C_{k+1}' + \rho_{k+1}^{-1}R_{k+1}]^{-1}d_k \leq 1\}$$

The inclusion $D_k \subset D_k^*$ can be easily verified from equation (6.19), (6.20), (6.31), (6.32). The infinite time version of the ellipsoidal algorithm can also be used for constant systems with constant constraint sets when the final time N approaches infinity. In this case the infinite time bounding ellipsoid estimator will be used to generate the estimate $\hat{x}_k$.

It should be noted that in the derivation of the sufficient condition of Proposition 6.2 we have made several weakening assumptions. We assumed that the controller uses the bounding ellipsoid estimator to produce the estimate set $X^*_{k|k}$ whereas the controller could conceivably calculate the exact set of possible states $X_{k|k}$. Furthermore we have not taken advantage of the possibility to obtain a smaller estimate set by using the term $\delta^2(k)$ in equation (6.18). A stronger sufficient condition for reachability can be obtained by making use of this term. However the additional complexity which would be introduced would make the resulting control scheme impractical. On the other hand our approach to reduce the state reachability problem with imperfect information to an estimate reachability problem with perfect information results in a control scheme the implementation of which presents no more difficulty than the one considered in Chapter 3.

4.   Discussion and Sources

In this chapter two reachability problems with imperfect state information were examined, both involving a linear system. The first problem is a target set reachability problem where the constraint set for the uncertain quantities is specified by an energy constraint. For this problem we characterized completely the optimal control law. This control law is in general quite difficult to implement, although suboptimal schemes can be derived which can be implemented more practically. However our primary objective has been to demonstrate that, for this problem, the state reachability problem with imperfect information is equivalent to an estimate reachability problem with perfect information, where the estimate is generated by an estimator

derived earlier in Chapter 4. This equivalence should be expected to hold in some form for more general reachability problems not involving an energy constraint for the uncertain quantities. However in such cases the necessary estimator, which will produce some estimate set, will in general require a very complicated implementation.

The second problem considered is a target tube reachability problem with instantaneous ellipsoidal constraints for the uncertain quantities. This problem was first examined in (B1). Our objective in this problem was to obtain suboptimal control schemes that can be practically implemented. We achieved this by reducing the state reachability problem with imperfect state information to a state estimate reachability problem of perfect information. The state estimate is produced by the (suboptimal) bounding ellipsoid filtering algorithm developed earlier in Chapter 4. The resulting control law, though obtained through substantial approximations, involves no more difficulty in its implementation than the corresponding perfect information control law considered in Chapter 3. This control law can be easily modified for the case where the constraint sets for the uncertain quantities are defined in a somewhat different form than those considered in Section 3. Such for example is the case where the ellipsoids specified by the constraints (6.17) are not centered at the origin, or the case where there are no measurement disturbances. In these cases appropriate modifications must be made in the bounding ellipsoid estimator. Another case sometimes appearing in practice is when the measurements are received by the controller with some delay. Thus in time i the controller may know the measurements only up to time k, $k < i$. Under these circumstances the controller must use a

predictor to generate the estimate $\hat{x}_{i|k}$ instead of the filter of Proposition 4.3, and the corresponding estimate reachability problem must be appropriately redefined.

CONCLUSIONS

In this thesis the subject of the feedback control of uncertain systems has been examined for the case where the uncertainty has nonstochastic description. In addition to theoretical investigations an effort has been made to provide design algorithms for the feedback control of uncertain systems which have potential for practical implementation. The set-membership description of the uncertain quantities appears to be attractive from the point of view of the designer who is often faced with a situation where he has an incomplete statistical description of the uncertain quantities. In such cases the designer often subjectively and arbitrarily assigns a probabilistic description to the uncertain quantities with the possible result of a poor mathematical model for the physical problem in hand. It is the author's belief that some of the algorithms in this thesis, particularly the ellipsoidal algorithms of Chapters 3 and 4, can provide a serious alternative to such a procedure when the set-membership description of the uncertainty is available. This is particularly so since, where applicable, these ellipsoidal algorithms lead to designs which have desirable features from the engineering viewpoint. The estimators of Chapter 4 have the same structure as linear minimum variance stochastic estimators, a structure which is desirable from the implementation point of view, and the ellipsoidal algorithm of Chapter 3 leads to a linear control law which again can be implemented with relative ease.

When considering the minimax approach towards a problem of decision under uncertainty one should constantly be aware of the fact that this approach is conservative in nature. In some problems where specified performance tolerances must be met with certainty the minimax approach

is the natural one. However in many other problems the minimax approach may lead to unduly conservative designs, and the algorithms proposed in this thesis should be viewed in the light of this consideration. If the design obtained through the minimax approach is deemed too conservative, other approaches such as a stochastic formulation of the problem can be considered.

Chapters 2 and 5 of the thesis are primarily of theoretical nature, and contain results which require, in general, a substantial computational effort for their use in a practical situation. They are important, however, for providing a general framework for considering minimax problems, for obtaining existence results, for providing insight into the structure of the optimal controller, and for yielding results in special cases such as some of those considered in Chapters 3 and 6. On the other hand, the emphasis in Chapters 3 and 4 and in part of Chapter 6 is in the development of algorithms which have potential for practical applications. These algorithms are applicable to the case of a linear discrete-time system where the constraint sets for the uncertain quantities are, or can be approximated by, ellipsoids. Although these algorithms have some attractive features, their performance has not as yet been sufficiently evaluated either analytically or by simulation. Furthermore the question of the quality of the approximation involved in these algorithms remains as yet unresolved. Thus some research and simulations are required to provide more insight into the merits and the drawbacks of these algorithms.

Other areas where further research is required are the situations where the system is nonlinear and/or continuous-time. The feedback control

problem which involves a nonlinear system and a set-membership description of the uncertainty requires, in general, excessive computational effort for its solution as discussed in Chapter 5. In this area optimal or nearly-optimal algorithms that are computationally feasible can be expected only for problems with special structure. The same appears to be true for the state estimation problem involving a nonlinear system. The state estimation problem involving a continuous-time linear system and either an energy constraint or instantaneous ellipsoidal constraints on the uncertain quantities presents no more difficulty than its discrete-time counterpart and has been considered in (B2). The feedback control problem involving a continuous-time system appears to present considerably greater technical difficulties than its discrete-time counterpart. This problem is essentially a differential game for which a saddle point in pure strategies is not necessarily assumed to exist, and is worthy of careful consideration.

Finally is should be mentioned that while in general the complete characterization of the optimal controller in a minimax control problem is a very difficult task, the same is true of stochastic optimal control problems with the exception of the case of the separation theorem for linear systems and quadratic criteria. Unfortunately no result comparable in elegance to the stochastic separation theorem appears to exist in connection with any particular minimax control problem.

APPENDIX I

ON THE THEORY OF CONVEX FUNCTIONS

In this appendix some definitions and results are summarized concerning convex functions defined on a finite dimensional Euclidean space. Only the results that are necessary for the developments in Chapter 2 are presented here. A complete exposition of the theory can be found in (R1).

The range of the function that we will be concerned with is the extended real line $R_e = [-\infty, \infty]$. The conventions adopted concerning arithmetic operations on R_e which involve $-\infty$ or ∞ are as follows:

Concerning addition we have:

$$a + \infty = \infty + a = \infty, \text{ for } a \in (-\infty, \infty]$$

$$a - \infty = -\infty + a = -\infty, \text{ for } a \in [-\infty, \infty)$$

Concerning multiplication we have:

$$a\infty = \infty a = \infty, \quad a(-\infty) = (-\infty)a = -\infty, \quad \text{for } a \in (0, \infty]$$

$$a\infty = \infty a = -\infty, \quad a(-\infty) = (-\infty)a = \infty, \quad \text{for } a \in [-\infty, 0)$$

$$0\infty = \infty 0 = 0 = 0(-\infty) = (-\infty)0$$

The sums $\infty - \infty$, $-\infty + \infty$ are undefined and are avoided. Under these rules addition and multiplication are commutative and associative, and the distributive law

$$a(a_1 + a_2) = aa_1 + aa_2$$

holds provided the sum $(a_1 + a_2)$ is neither of the forbidden sums $\infty - \infty$ and $-\infty + \infty$.

The cancellation laws hold as follows

$$a + a_1 = a + a_2 \implies a_1 = a_2, \quad \text{for } a \neq -\infty, \infty$$

$$aa_1 = aa_2 \implies a_1 = a_2, \quad \text{for } a \neq 0, -\infty, \infty$$

Order on the extended real line is defined in the natural way. Concerning inequalities we have the cancellation laws:

$$a + a_1 \leq a + a_2 \implies a_1 \leq a_2, \quad \text{for } a \neq -\infty, \infty$$

$$aa_1 \leq aa_2 \implies a_1 \leq a_2, \quad \text{for } a \in (0, \infty)$$

$$aa_1 \leq aa_2 \implies a_1 \geq a_2, \quad \text{for } a \in (-\infty, 0)$$

One of the advantages of the extended real line is that it is closed under taking the supremum or the infimum of any of its subsets with the additional convention that for the empty set

$$\inf \phi = \infty, \quad \sup \phi = -\infty$$

The familiar minimax inequality

$$\sup_{y \in Y} \inf_{x \in X} \mathcal{G}(x, y) \leq \inf_{x \in X} \sup_{y \in Y} \mathcal{G}(x, y)$$

holds for any function $\mathcal{G} : R^n \times R^m \to [-\infty, \infty]$ and any sets $X \subseteq R^n$, $Y \subseteq R^m$.

In calculations which involve the supremum or infimum operation care sometimes must be exercised so that the forbidden sums $\infty - \infty$, $-\infty + \infty$ do not appear. For example if $f_1 : R^n \to (-\infty, \infty]$, $f_2 : R^m \to (-\infty, \infty]$ are functions and X, Y are subsets of R^n and R^m respectively we have

$$\inf_{\substack{x, y \\ x \in X, y \in Y}} [f_1(x) + f_2(y)] = \inf_{x \in X} \inf_{y \in Y} [f_1(x) + f_2(y)]$$

$$= \inf_{x \in X} [f_1(x) + \inf_{y \in Y} f_2(y)]$$

only if either $-\infty < \inf_{y \in Y} f_2(y)$ or $f_1(x) < \infty$, $\forall x \in X$.

Such calculations are common in dynamic programming algorithms and will be used frequently in the text of the thesis.

We now introduce some of the notions related to convex functions. Let $f: R^n \rightarrow [-\infty, \infty]$ be a function.

Definition A.1: The epigraph of f is the subset of R^{n+1}

$$\text{epi} f = \{(x, \mu) \mid x \in R^n, \mu \in R, \mu \geq f(x)\}$$

Definition A.2: The function f is called convex if the set epi $f \subset R^{n+1}$ is convex. If $-\infty < f(x)$, $\forall x \in R^n$ this is equivalent to

$$f[(1-\lambda)x + \lambda y] \leq (1-\lambda)f(x) + \lambda f(y), \quad \forall \lambda \in (0,1), \forall x, y \in R^n$$

Definition A.3: The convex hull of a function f, denoted by conv f, is the convex function which has as epigraph the set conv (epif) (convex hull of epif).

Definition A.4: A convex function f is said to be proper if $-\infty < f(x)$, $\forall x \in R^n$ and $f(x) < \infty$ for at least one $x \in R^n$. It is said to be closed if epif is a closed set.

Definition A.5: The closure of a proper convex function f, denoted by clf, is the closed proper convex function which has as epigraph the set cl(epif) (closure of the set epif).

Concerning closed proper convex functions we have the following proposition:

Proposition A.1: Let f be a convex proper function. The following

conditions are equivalent

    (a) f is closed.

    (b) The level sets $\{x \mid f(x) \leq a\}$ are closed, $\forall\, a \in R$.

    (c) f is lower semicontinuous.

<u>Proof</u>: See Theorem 7.1 in Reference (R1).

<u>Definition A.6</u>: The <u>effective domain</u> of a convex function f is the convex set

$$\mathrm{dom}\, f = \{x \mid f(x) < \infty\}$$

A subset L of R^n is called <u>affine</u> (linear manifold) if $(1 - \lambda) x + \lambda y \in L$, $\forall\, x, y \in L$, $\forall\, \lambda \in R$. Given now a convex set C in R^n the <u>affine hull</u> of C is the smallest affine set that contains C. With these definitions we have:

<u>Definition A.7</u>: The <u>relative interior of the effective domain</u> of a convex function f, denoted ri (domf), is the interior of the set domf relative to its affine hull.

<u>Definition A.8</u>: A convex function f is said to be <u>positively homogeneous</u> if

$$f(\lambda x) = \lambda f(x), \quad \forall\, x \in R^n, \quad \forall\, \lambda \in (0, \infty)$$

An example of a positively homogeneous convex function is the <u>support function</u> of a convex set C

$$\sigma(x \mid C) = \sup_{x^* \in C} \langle x, x^* \rangle$$

Concerning continuity of convex functions we have:

<u>Proposition A.2</u>: Let f be a proper convex function on R^n. Then the restriction of f to any subset C of domf which is open relative to the affine

hull of domf is continuous. In particular the restriction of f to ri (domf) is continuous. This implies that a convex function which is finite on all of R^n is continuous.

Proof: See Theorem 10.1 of Reference (R1).

Some important operations involving convex functions will now be introduced:

Proposition A.3: If f_1, f_2 are proper convex functions in R^n the function $f_1 + f_2$ is convex. It is proper if $\text{domf}_1 \cap \text{dom} f_2 \neq \phi$.

Proof: See Theorem 5.2 in Reference (R1).

Proposition A.4: If f_1, f_2 are proper convex functions in R^n the function f defined by

$$f(x) = \inf_{y} \{f_1(x - y) + f_2(y)\}$$

is convex. The function f is denoted as $f = f_1 \square f_2$ and the operation $\square$ is called infimal convolution.

Proof: See Theorem 5.4 in Reference (R1).

Proposition A.5: Let f_i, $i \in I$, be convex functions, where I is an arbitrary index set. Then the function f defined by

$$f(x) = \sup_{i \in I} f_i(x)$$

is convex.

Proof: See Theorem 5.5 in Reference (R1).

Proposition A.6: Let A be a linear transformation from R^n to R^m. Then

for each convex function g on R^m, the function gA defined by

$$(gA)(x) = g(Ax)$$

is convex in R^n. For each convex function h on R^n, the function Ah defined by

$$(Ah)(y) = \inf_{Ax=y} h(x)$$

is convex on R^m (Notice that in accordance with the convention $\inf \phi = \infty$ we have $(Ah)(y) = \infty$ for all y which are not in the range of A). The function Ah is called the <u>image of h under A</u> and the function gA is called the <u>inverse image of g under A</u>.

<u>Proof</u>: See Theorem 5.7 in Reference (R1).

We now introduce the notion of the recession function of a convex function. This notion is extremely helpful in proving closure and properness of functions resulting from functional operations introduced earlier as well as in proving existence of solutions in convex optimization problems.

Let C be a nonempty convex set in R^n. We say that C recedes in the direction of the vector $z \neq 0$ if and only if $x + \lambda z \in C$ for every $\lambda \geq 0$ and $x \in C$. The set of all vectors $z \in R^n$ that satisfy this condition together with $z = 0$ is called the <u>recession cone</u> of C, denoted by 0^+C.

<u>Definition A.9</u>: Let f be a proper convex function. The <u>recession function</u> $f0^+$ of f is the convex function which has as epigraph the set $0^+(\text{epi } f)$.

<u>Proposition A.7</u>: The recession function $f0^+$ of a proper convex function f is a positively homogeneous proper convex function given by

$$(f0^+)(z) = \sup \{f(x+z) - f(x) | x \in \text{dom } f\}$$

If f is closed then $f0^+$ is closed also.

Proof: See Theorem 8.5 in Reference (R1).

Definition A.10: A direction defined by a vector $z \neq 0$ is called a <u>direction</u> <u>of recession</u> of the proper convex function f if $(f0^+)(z) \leq 0$. It is called a direction in which f is constant if $(f0^+)(z) = (f0^+)(-z) = 0$.

Thus a proper convex function f every direction of recession of which is a direction in which it is constant is characterized by the fact that $(f0^+)(z) \geq 0$, and $(f0^+)(z) = 0$ implies $(f0^+)(-z) = 0$.

Some criteria for properness and closure of functions resulting from functional operations of convex functions will now be given.

Proposition A.8: If f_1, f_2 are closed proper convex functions and $f_1 + f_2$ is not identically ∞ then $f_1 + f_2$ is a closed proper convex function and

$$(f_1 + f_2)0^+ = f_1 0^+ + f_2 0^+$$

Proof: See Theorem 9.3 in Reference (R1).

Proposition A.9: Let f_1, f_2 be closed proper convex functions in R^n such that there exists no $z \in R^n$ such that

$$(f_1 0^+)(z) + (f_2 0^+)(-z) > 0$$

$$(f_1 0^+)(-z) + (f_2 0^+)(z) \leq 0$$

Then $f_1 \square f_2$ is a closed proper convex function and the infimum in the equation

$$(f_1 \square f_2)(x) = \inf_y \{f_1(x - y) + f_2(y)\}$$

is attained by some y for each x. Moreover

$$(f_1 \square f_2)0^+ = f_1 0^+ \square f_2 0^+$$

Proof: See Corollary 9.2.1 in Reference (R1).

Proposition A.10: Let f_i, $i \in I$, be closed proper convex functions where I is an arbitrary index set. Then the function f defined by

$$f(x) = \sup_{i \in I} f_i(x)$$

either is the constant ∞ function or it is a closed proper convex function and $f0^+$ is given by

$$(f0^+)(z) = \sup_{i \in I} (f_i 0^+)(z)$$

Proof: See Theorem 9.4 in Reference (R1).

Proposition A.11: Let h be a closed proper convex function on R^n, and let A be a linear transformation from R^n to R^m. Assume that there exists no $z \in R^n$ such that $Az = 0$, $(h0^+)(z) \le 0$ and $(h0^+)(-z) > 0$. Then the function Ah, where

$$(Ah)(y) = \inf_{Ax=y} h(x)$$

is a closed proper convex function and $(Ah)0^+ = A(h0^+)$. Moreover for every y such that $(Ah)(y) < \infty$ the infimum in the definition of Ah is attained for some x.

Proof: See Theorem 9.2 in Reference (R1).

Proposition A.12: Let g be closed proper convex function on R^m, and let A be a linear transformation from R^n to R^m. Assume that the function gA defined by

$$(gA)(x) = g(Ax)$$

is not identically ∞. Then gA is a closed proper convex function and
$$(gA)0^+ = (g0^+)A.$$

Proof: See Theorem 9.5 in Reference (R1).

As an application of the above propositions consider the function H_k in R^n defined by

$$H_k(x_k) = \inf_{u_k} \{E_{k+1}(A_k x_k + B_k u_k) + g_k(u_k)\} \tag{A.1}$$

where $E_{k+1} : R^n \to (-\infty, \infty]$, $g_k : R^m \to (-\infty, \infty]$ are closed proper convex functions and $A_k : R^n \to R^n$, $B_k : R^m \to R^n$ are given linear transformations. The function H_k above is of great interest in Chapter 2. Assume that H_k is not identically ∞ and consider the following assumptions:

Assumption R: Every direction of recession of each of the functions E_{k+1} and g_k is a direction in which this function is constant.

Assumption C: The recession function of g_k is of the form

$$(g_k 0^+)(z) = \infty \quad \text{for } z \neq 0, \quad (g_k 0^+)(0) = 0$$

The following proposition holds:

Proposition A.13: Under either Assumption R or Assumption C the function H_k of equation (A.1) is given by

$$H_k = [E_{k+1} \,\square\, (-B_k)g_k] A_k \tag{A.2}$$

and it is a closed proper convex function. Furthermore for each $x_k \in R^n$ the infimum in equation (A.1) is attained by some $u_k \in R^m$, and in the case of Assumption R every direction of recession of the function H_k is a direction

in which H_k is constant.

Proof: Under either Assumption R or Assumption C the conditions of Proposition A.11 are satisfied so that the function $(-B_k)g_k$ given by:

$$[(-B_k)g_k](y) = \inf_{y=-B_k u_k} g_k(u_k)$$

is a closed proper convex function and $[(-B_k)g_k]0^+ = (-B_k)g_k 0^+$.

We have now

$$H_k(x_k) = \inf_{u_k} \{E_{k+1}(A_k x_k + B_k u_k) + g_k(u_k)\}$$

$$= \inf_{\substack{y \quad u_k \\ y=-B_k u_k}} \inf \{E_{k+1}(A_k x_k - y) + g_k(u_k)\}$$

$$= \inf_{y} \{E_{k+1}(A_k x_k - y) + \inf_{\substack{u_k \\ y=-B_k u_k}} g_k(u_k)\}$$

$$= \inf_{y} \{E_{k+1}(A_k x_k - y) + (-B_k)g_k(y)\}$$

$$= [E_{k+1} \,\square\, (-B_k)g_k](A_k x_k)$$

$$= [[E_{k+1} \,\square\, (-B_k)g_k]A_k](x_k)$$

where by our assumptions none of the forbidden sums $\infty - \infty$ or $-\infty + \infty$ appears in the above algebra. Thus equation (A.2) is proved.

The relation $[(-B_k)g_k]0^+ = (-B_k)g_k 0^+$ implies in the case of Assumption R that every direction of recession of the function $(-B_k)g_k$ is a direction in which this function is constant as can be easily seen by applying the appropriate definitions and Proposition A.11. In the case of Assumption C we have

$$[\,(-B_k)g_k]\,0^+\,(z)\ =\ \infty\ \text{for}\ z \neq 0, \qquad [\,(-B_k)g_k]\,0^+\,(0) = 0.$$

In both cases the conditions of Proposition A.9 are satisfied so that the function $E_{k+1}\square(-B_k)g_k$ is a closed proper convex function and by Proposition A.12 the function H_k of equation (A.2) is also a closed proper convex function.

To see that the infimum in equation (A.1) is attained observe that if x_k is such that $H_k(x_k) = \infty$ then the infimum is attained for every $u_k \in R^n$. If x_k is such that $H_k(x_k) < \infty$ then attainement of the infimum follows by making use of the conclusions of Propositions A.9 and A.11 in the equation

$$H_k(x_k)\ =\ \inf_{\substack{y \\ y = -B_k u_k}} \{E_{k+1}(A_k x_k - y) + \inf_{u_k} g_k(u_k)\}$$

In the case of Assumption R the conclusion that every direction of recession of the function H_k is a direction in which it is constant follows from the equation

$$H_k 0^+\ =\ [\,E_{k+1}0^+\square(-B_k)g_k 0^+]\,A_k$$

which holds by Propositions A.9, A.11 and A.12, Q.E.D.

The important notion of the conjugate function of a convex function is now introduced.

<u>Definition A.11</u>: Let f be a convex function in R^n. The <u>conjugate function</u> f^* is defined as

$$f^*(x^*)\ =\ \sup_x\,\{<x,x^*> -f(x)\}\ =\ -\inf_x\,\{f(x) - <x,x^*>\}$$

and is a closed convex function in R^n, proper if and only if f is proper. Moreover $(\text{clf})^* = f^*$ and $(f^*)^* = \text{clf}$.

An important example of a pair of conjugate functions is the <u>indicator</u> <u>function</u> and the <u>support function</u> of a closed convex set C. We have

$$\delta(x|C) = 0 \quad \text{if } x \epsilon C, \quad \delta(x|C) = \infty \quad \text{if } x \not\epsilon X$$

$$\delta^*(x^*|C) = \sigma(x^*|C) = \sup_{x \epsilon C} <x^*, x>$$

Concerning conjugate functions of convex functions resulting from the operations introduced earlier we have a duality between addition and infimal convolution, and between the image and the inverse image of a convex function under a linear transformation, in accordance with the following propositions:

<u>Proposition A.14</u>: Let f_1, f_2 be proper convex functions in R^n. Then

$$(f_1 \square f_2)^* = f_1^* + f_2^*$$

$$(cl f_1 + cl f_2)^* = cl(f_1^* \square f_2^*)$$

and if $\text{ri}(\text{dom} f_1) \cap \text{ri}(\text{dom} f_2) \neq \phi$ the closure operation can be ommited from the second equation.

<u>Proof</u>: See Theorem 16.4 in Reference (R1).

<u>Proposition A.15</u>: Let A be a linear transformation from R^n to R^m. For any convex function h on R^n we have

$$(Ah)^* = h^*A'$$

For any convex function g on R^m we have

$$[(clg)A]^* = cl(A'g^*)$$

and if there exists an x such that $Ax \epsilon \text{ri}(\text{dom} g)$, the closure operation can be ommited from the second equation.

Proof: See Theorem 16.3 in Reference (R1).

The notion of the directional derivative and the related notion of the subdifferential is fundamental in the problem of finding the minimum of a convex function.

Let f be any function from R^n to $[-\infty, \infty]$ and let x be a point where f is finite. The <u>one-sided directional derivative</u> of f at x with respect to a vector $y \in R^n$ is defined as the limit

$$f'(x;y) = \lim_{\lambda \to 0} \frac{f(x + \lambda y) - f(x)}{\lambda} \tag{A.3}$$

if it exists (∞ and $-\infty$ being allowed as limits).

<u>Proposition A.16</u>: Let f be a convex function, and let x be a point where f is finite. Then for every y the limit in equation (A.3) exists and the function $f'(x;y)$ is a positively homogeneous convex function of y.

Proof: See Theorem 23.1 in Reference (R1).

<u>Definition A.12</u>: A vector $x^* \in R^n$ is said to be a <u>subgradient</u> of a convex function f in R^n at a point x if

$$f(z) \geq f(x) + \langle x^*, z - x \rangle, \qquad \forall z \in R^n \tag{A.4}$$

The set of all subgradients of f at x is called the <u>subdifferential of f at x</u>, it is denoted by $\partial f(x)$, and it is a closed convex set. If $\partial f(x) \neq \phi$ then f is said to be <u>subdifferentiable at x</u>.

<u>Proposition A.17</u>: Let f be a proper convex function. For $x \notin \text{dom } f, \partial f(x)$ is empty. For $x \in \text{ri}(\text{dom } f)$, $\partial f(x)$ is nonempty and $f'(x;y)$ is the support function of $\partial f(x)$. Finally $\partial f(x)$ is a compact set if and only if $x \in \text{int}(\text{dom } f)$.

Proof: See Theorem 23.4 in Reference (R1).

It should be noted that if f is differentiable at a point x then $\partial f(x)$ consists of a single point the gradient $\nabla f(x)$. Another important case to note is when f is the indicator function of a closed convex set C. Then for $x \in C$, $\partial f(x) = \partial \delta(x|C) = \{x^* | 0 \geq <x^*, z - x>, \quad \forall z \in C\}$, i.e., $\partial \delta(x|C)$ is the set of all vectors normal to C at x.

Duality is prevalent in the theory of subgradients due to the following fact:

Proposition A.18: For any proper convex function f and any x the following conditions are equivalent:

    (a) $x^* \in \partial f(x)$

    (b) $< z, x^* > - f(z)$ achieves its supremum in z at $z = x$.

    (c) $f(x) + f^*(x^*) = < x, x^* >$

Proof: See Theorem 23.5 in Reference (R1).

We also have:

Proposition A.19: Let f_1, f_2 be proper convex functions on R^n and let $f = f_1 + f_2$. Then

$$\partial f(x) \supset \partial f_1(x) + \partial f_2(x) , \quad \forall x \in R^n$$

If $ri(dom f_1) \cap ri(dom f_2) \neq \phi$ then actually

$$\partial f(x) = \partial f_1(x) + \partial f_2(x) , \quad \forall x \in R^n$$

Proof: See Theorem 23.8 in Reference (R1).

Proposition A.20: Let $f(x) = h(Ax)$, where h is a proper convex function on R^m and A a linear transformation from R^n to R^m. Then

$$\partial f(x) \supset A'\partial h(Ax), \qquad \forall x \in R^n$$

If the range of A contains a point of ri (dom h) then

$$\partial f(x) = A'\partial h(Ax), \qquad \forall x \in R^n$$

Proof: See Theorem 23.9 in Reference (R1).

If $\{x_n\}$ is a sequence converging to a point $x \in R^n$ it is not generally true that for any $y \in R^n$ the sequence $\{f'(x_n; y)\}$ converges to $f'(x; y)$. The following proposition however is useful in some cases.

Proposition A.21: Let f be a convex function on R^n and let C be an open convex set on which f is finite. Let $\{f_n\}$ be a sequence of convex functions finite on C and converging pointwise to f on C. Let $x \in C$ and let $\{x_n\}$ be a sequence of points in C converging to x. Then for any $y \in R^n$ and any sequence $\{y_n\}$ converging to y we have

$$\lim_{n \to \infty} \sup f'_n(x_n; y_n) \le f'(x; y)$$

Proof: See Theorem 24.5 in Reference (R1).

As an application of the above we prove the following proposition which will be useful in Chapter 2. This proposition is a generalization of results in References (D1) and (Da1).

Proposition A.22: Let $\mathcal{G} : R^n \times R^m \to (-\infty, \infty]$ be a function and let Y be a compact subset of R^m. Assume further that for every vector $y \in Y$ the function $\mathcal{G}(\cdot, y) : R^n \to (-\infty, \infty]$ is a closed proper convex function. Consider the function f defined as

$$f(x) = \sup_{y \in Y} \mathcal{G}(x, y)$$

Then if f is finite somewhere it is a closed proper convex function. Furthermore if $\text{int}(\text{dom}\, f) \neq \phi$ and $\mathcal{G}$ is continuous on the set $\text{int}(\text{dom}\, f) \times Y$ then for every $x \in \text{int}(\text{dom}\, f)$ we have

$$\partial f(x) = \text{conv}\, \{\partial \mathcal{G}(x, \bar{y}) \mid \bar{y} \in \overline{Y}(x)\}$$

where $\overline{Y}(x)$ is the set

$$\overline{Y}(x) = \{\bar{y} \in Y \mid \mathcal{G}(x, \bar{y}) = \max_{y \in Y} \mathcal{G}(x, y)\}$$

Proof: The fact that f is a closed proper convex function follows immediately from Proposition A.10. Let $x \in \text{int}(\text{dom}\, f)$ and let C be an open convex neighborhood of x such that f is finite on C. Then $\mathcal{G}$ is finite and continuous on $C \times Y$ since $C \subset \text{int}(\text{dom}\, f)$. Now for any $z \in R^n$ let $\{x_n\} = \{x + \lambda_n z\}$ be a sequence of vectors in C with $\lambda_n > 0$, $\{\lambda_n\} \to 0$. Let also $\{\bar{y}_n\}$ be a sequence of vectors such that $f(x_n) = \mathcal{G}(x_n, \bar{y}_n)$, i.e., $\bar{y}_n \in \overline{Y}(x_n) \subset Y$. Such a sequence exists by the compactness of Y and the continuity of the function $\mathcal{G}(x_n, \cdot)$ on Y. Furthermore by compactness of Y the sequence $\{\bar{y}_n\}$ has a subsequence which converges to a vector $\bar{y} \in Y$. Thus without loss of generality we can assume that the sequence $\{x_n\}$ is selected so that the corresponding sequence $\{\bar{y}_n\}$ converges to the vector $\bar{y} \in Y$. We now prove that in fact $\bar{y} \in \overline{Y}(x)$, i.e., that $\mathcal{G}(x, \bar{y}) = \max_{y \in Y} \mathcal{G}(x, y)$. We have by the continuity of f on C for any $\epsilon > 0$

$$f(x) - \epsilon \leq f(x_n) \qquad \text{for } n \geq N_1$$

and by the continuity of $\mathcal{G}$ on $C \times Y$

$$f(x_n) = \mathcal{G}(x_n, \bar{y}_n) \leq \mathcal{G}(x, \bar{y}) + \epsilon \qquad \text{for } n \geq N_2$$

and therefore

$$f(x) = \max_{y \in Y} \mathcal{G}(x, y) \leq \mathcal{G}(x, \bar{y}) + 2\epsilon,$$

and since the above inequality holds for any $\epsilon > 0$ we conclude $\max_{y \in Y} \mathcal{G}(x, y) = \mathcal{G}(x, \bar{y})$ and $\bar{y} \in \bar{Y}(x)$,

Now we have:

$$\frac{f(x + \lambda_n z) - f(x)}{\lambda_n} = \frac{\mathcal{G}(x + \lambda_n z, \bar{y}_n) - \mathcal{G}(x, \bar{y})}{\lambda_n}$$

$$\leq \frac{\mathcal{G}(x + \lambda_n z, \bar{y}_n) - \mathcal{G}(x, \bar{y}_n)}{\lambda_n} = \mathcal{G}'(x, \bar{y}_n; z) + \frac{0(\lambda_n)}{\lambda_n}$$

with $\lim_{n \to \infty} \frac{0(\lambda_n)}{\lambda_n} = 0$. Taking limits in the above inequality we obtain

$$f'(x; z) \leq \lim_{n \to \infty} \sup \mathcal{G}'(x, \bar{y}_n; z) \tag{A.5}$$

Now the sequence of functions of x $\{\mathcal{G}(\cdot, \bar{y}_n)\}$ converges pointwise to the function $\mathcal{G}(\cdot, \bar{y})$ on the open set C by the continuity of $\mathcal{G}$ on CxY, and by applying Proposition A.21

$$\lim_{n \to \infty} \sup \mathcal{G}'(x, \bar{y}_n; z) \leq \mathcal{G}'(x, \bar{y}; z)$$

Using the above in relation (A.5) we obtain

$$f'(x; z) \leq \mathcal{G}'(x, \bar{y}; z) \tag{A.6}$$

On the other hand we have for every vector $\hat{y} \in \bar{Y}(x)$

$$\frac{f(x + \lambda_n z) - f(x)}{\lambda_n} = \frac{\mathcal{G}(x + \lambda_n z, \bar{y}_n) - \mathcal{G}(x, \hat{y})}{\lambda_n}$$

$$\geq \frac{\mathcal{G}(x + \lambda_n z, \hat{y}) - \mathcal{G}(x, \hat{y})}{\lambda_n}$$

Taking limits in the above inequality we obtain

$$f'(x;z) \geq \varphi'(x, y;z), \qquad \forall y \in \overline{Y}(x) \qquad \qquad (A.7)$$

From relations (A.6) and (A.7) it follows that

$$f'(x;z) = \max_{\overline{y} \in \overline{Y}(x)} \varphi'(x, \overline{y};z)$$

and since by Proposition A.17 $\varphi'(x, \overline{y};\cdot)$ is the support function of the convex compact set $\partial\varphi(x, \overline{y})$ and $f'(x;\cdot)$ is the support function of the convex compact set $\partial f(x)$ it follows:

$$\partial f(x) = \text{conv}\{\partial\varphi(x, \overline{y})|\overline{y} \in \overline{Y}(x)\} \qquad Q.E.D.$$

Consider now a closed proper convex function of f and the problem of finding its minimum in R^n. The set of points $\overline{x} \in R^n$ such that

$$f(\overline{x}) = \inf_x f(x)$$

will be called the __minimum set__ of f. We have the following proposition:

__Proposition A.23__: The following statements are valid for any closed proper convex function f and its conjugate f^*.

(a) $\inf_x f(x) = -f^*(0)$. Thus f is bounded below if and only if $0 \in \text{dom} f^*$.

(b) A vector $\overline{x}$ belongs to the minimum set of f if and only if $0 \in \partial f(\overline{x})$.

(c) The minimum set of f is $\partial f^*(0)$. Thus the infimum of f is attained if and only if f^* is subdifferentiable at 0. This condition is satisfied in particular when $0 \in \text{ri}(\text{dom} f^*)$; moreover one has $0 \in \text{ri}(\text{dom} f^*)$ if and only if every direction of recession of f is a direction in which f is constant.

(d) The minimum set of f is a nonempty compact set if and only if $0 \in \text{int}(\text{dom } f^*)$. This holds if and only if f has no directions of recession.

(e) The minimum set of f consists of a unique vector x if and only if f^* is differentiable at 0 and $x = \nabla f^*(0)$.

Proof: See Theorem 27.1 in Reference (R1).

The above Proposition illustrates the fundamental role of the sub-differential in convex minimization problems and shows the importance of the recession function in such problems.

APPENDIX II

In this Appendix we present the proofs of Propositions 3.1 and 3.2. We begin with the proof of Proposition 3.1. For the purpose of clearer presentation a few lemmas, some of which are well known, will be given first:

<u>Lemma A.1</u>: Let Σ_1, Σ_2 be positive definite symmetric $n \times n$ matrices. Then

   (a) $\Sigma_1 \leq \Sigma_2$ if and only if $\Sigma_1^{-1} \geq \Sigma_2^{-1}$

   (b) There exist positive scalars μ, ν such that

$$\nu \Sigma_2 < \Sigma_1 < \mu \Sigma_2$$

<u>Proof</u>: (a) For all $y \in R^n$, $(y'\Sigma_1^{-1}y)^{1/2} = \sup_{x'\Sigma_1 x \leq 1} \langle x,y \rangle \geq \sup_{x'\Sigma_2 x \leq 1} \langle x,y \rangle = (y'\Sigma_2^{-1}y)^{1/2}$ Q.E.D.

   (b) For any two norms in R^n, $||\cdot||_1$, $||\cdot||_2$, there exist positive scalars μ, ν such that

$$\nu^{1/2}||x||_2 < ||x||_1 < \mu^{1/2}||x||_2 \qquad \text{for all } x \in R^n$$

Taking $||x||_1 = (x'\Sigma_1 x)^{1/2}$, $||x||_2 = (x'\Sigma_2 x)^{1/2}$ the result follows Q.E.D.

<u>Lemma A.2</u>: Let F be an $n \times n$ matrix such that for every eigenvalue $\lambda = a + bi$ of F we have $a^2 + b^2 < 1$. Then there exists a positive definite symmetric matrix M such that

$$F'MF < M$$

<u>Proof</u>: This lemma is a direct consequence of the fact that if $\rho(F) = \max \{\sqrt{a^2 + b^2} \,|\, \lambda = a + bi, \ \lambda : \text{eigenvalue of } F\}$ then for every $\epsilon > 0$ there exists

a Euclidean norm $||x|| = (x'Mx)^{1/2}$ such that

$$\rho(F) < ||F|| = \sup_{x} \frac{||Fx||}{||x||} = \sup_{x} \frac{(x'F'MFx)^{1/2}}{(x'Mx)^{1/2}} < \rho(F) + \epsilon$$

(see for example, Reference (I1)

Since $\rho(F) < 1$ there exists a positive definite symmetric matrix M such that

$$\sup_{x} \frac{(x'F'MFx)^{1/2}}{(x'Mx)^{1/2}} < 1$$

implying $F'MF < M$ Q.E.D.

Lemma A.3: Let $\{\Sigma_k\}$ be a sequence of positive definite symmetric $n \times n$ matrices such that $\Sigma_k \leq \Sigma_{k+1} \leq M$ for all k, where M is a positive definite symmetric matrix. Then the sequence $\{\Sigma_k\}$ converges (in any norm in R^{n^2}) to a positive definite symmetric matrix Σ_∞.

Proof: This lemma is a special case of a result for positive operators in Hilbert space [Kal] and has appeared in this form in Reference (Wo2).

Lemma A.4: Consider the sequences of matrices $\{K_k\}$, $\{\Sigma_k\}$ generated by the equations

$$K_{k-1} = A'(K_k^{-1} - GQ^{-1}G' + BR^{-1}B')^{-1}A + \Psi \tag{A.8}$$

$$K_n = \Psi$$

$$\Sigma_{k-1} = (A - BL)'(\Sigma_k^{-1} - GQ^{-1}G')^{-1}(A - BL) + \Psi + L'RL \tag{A.9}$$

$$\Sigma_n = \Psi$$

where A, B, G, L are given matrices of dimension $n \times n$, $n \times m$, $n \times r$, $m \times n$ respectively and Ψ, Q and R are given positive definite symmetric matrices. Assume that for all k the matrix Σ_k is positive definite and symmetric and

that

$$GQ^{-1}G' < \Sigma_k^{-1} \qquad \text{for all } k$$

Then we have for all k

$$0 < K_k \leq \Sigma_k \text{ and } GQ^{-1}G' < K_k^{-1}$$

<u>Proof</u>: We prove the lemma by induction. For $k = N$ it holds. Assume $0 < K_k \leq \Sigma_k$. It will be proved that $0 < K_{k-1} \leq \Sigma_{k-1}$. For convenience write

$$M = (\Sigma_k^{-1} - GQ^{-1}G')^{-1} \geq (K_k^{-1} - GQ^{-1}G')^{-1}$$

Then from equations (A.8) and (A.9)

$$\Sigma_{k-1} - K_{k-1} = (A-BL)'M(A-BL) + \Psi + L'RL$$
$$- A'(K_k^{-1} - GQ^{-1}G' + BR^{-1}B')^{-1}A - \Psi$$

$$\geq (A-BL)'M(A-BL) + L'RL - A'(M^{-1} + BR^{-1}B')^{-1}A$$

By using the well known matrix identity

$$(M^{-1} + BR^{-1}B')^{-1} = M - MB(B'MB + R)^{-1}B'M$$

in the above inequality and by expanding we obtain

$$\Sigma_{k-1} - K_{k-1} \geq A'MA + L'(B'MB+R)L - L'B'MA - A'MBL$$
$$- A'MA + A'MB(B'MB+R)^{-1}B'MA$$

$$= L'(B'MB+R)L - A'MBL - L'B'MA$$
$$+ A'MB(B'MB+R)^{-1}B'MA$$

$$= [L - (R+B'MB)^{-1}B'MA]'(B'MB+R)[L - (R+B'MB)^{-1}B'MA]$$

$$\geq 0$$

Hence $K_{k-1} \leq \Sigma_{k-1}$. Since $GQ^{-1}G' < K_k^{-1}$ it follows from equation (A.8) that $0 < K_{k-1}$ and since $GQ^{-1}G' < \Sigma_{k-1}^{-1} \leq K_{k-1}^{-1}$ we also obtain $GQ^{-1}G' < K_{k-1}^{-1}$ Q.E.D.

Lemma A.5: Consider the sequence of matrices $\{\Sigma_k\}$ generated by the equation

$$\Sigma_{k-1} = F'(\Sigma_k^{-1} - GQ^{-1}G')^{-1}F + \Psi + L'RL \tag{A.10}$$

$$\Sigma_n = \Psi \tag{A.11}$$

where Ψ, R, Q, G, F, L are given matrices of appropriate dimension, Ψ, R and Q are positive definite and symmetric and for every eigenvalue $\lambda = a + bi$ of the $n \times n$ matrix F we have $a^2 + b^2 < 1$.

Let M be a positive definite symmetric matrix such that $F'MF < M$, let q be a positive scalar such that

$$GQ^{-1}G' < qM^{-1} \tag{A.12}$$

and let μ be a positive scalar such that

$$F'MF < (1 - q\mu)M, \quad q\mu < 1 \tag{A.13}$$

The existence of such a scalar μ is guaranteed since by Lemma A.1 there exists a positive scalar ν such that $\nu M < M - F'MF$ and any scalar μ with $0 < \mu \leq \frac{\nu}{q}$ satisfies the inequality (A.13). Assume further that the matrices Ψ, R and L are such that

$$\Psi + L'RL < \frac{\mu}{1-q\mu}[(1-q\mu)M - F'MF] \tag{A.14}$$

Then the sequence $\{\Sigma_k\}$ converges to a positive definite symmetric matrix $\Sigma_{-\infty}$. Furthermore the matrices Σ_k are positive definite and for all k

$$GQ^{-1}G' < \Sigma_k^{-1} \ , \quad \Sigma_k < \mu M \tag{A.15}$$

<u>Proof</u>: It will be proved under our assumptions that

$$0 < \Sigma_k \leq \Sigma_{k-1} < \mu M \tag{A.16}$$

then by Lemma A.3 convergence will follow and furthermore the inequality (A.15) will be satisfied since from (A.12) and (A.13) we have

$$GQ^{-1}G' < qM^{-1} < \frac{1}{\mu}M^{-1} < \Sigma_k^{-1}$$

The relation (A.16) will be proved by induction. We have from (A.14) $0 < \Psi = \Sigma_N < \mu M$ which also implies $(\Psi^{-1} - GQ^{-1}G')^{-1}$ is positive definite and from equation (A.10) we have $\Psi = \Sigma_N \leq \Sigma_{N-1}$. Assume that $\Sigma_{k+1} \leq \Sigma_k < \mu M$. Then

$$\begin{aligned}
\Sigma_{k-1} &= F'(\Sigma_k^{-1} - GQ^{-1}G')^{-1}F + \Psi + L'RL \\
&\geq F'(\Sigma_{k+1}^{-1} - GQ^{-1}G')^{-1}F + \Psi + L'RL = \Sigma_k
\end{aligned}$$

and also

$$\begin{aligned}
\Sigma_{k-1} &= F'(\Sigma_k^{-1} - GQ^{-1}G')^{-1}F + \Psi + L'RL \\
&\leq F'(\frac{1}{\mu}M^{-1} - GQ^{-1}G')^{-1}F + \Psi + L'RL \\
&\leq F'(\frac{1}{\mu}M^{-1} - qM^{-1})^{-1}F + \Psi + L'RL \\
&= \frac{\mu}{1-q\mu}F'MF + \Psi + L'RL < \mu M
\end{aligned}$$

where the last inequality follows from relation (A.14). Thus we obtain $\Sigma_k \leq \Sigma_{k-1} < \mu M$ and the induction proof is complete Q.E.D.

We are now ready to state the proof of Proposition 3.1:

<u>Proof of Proposition 3.1</u>: It is required to prove that there exists a scalar

β_1 such that for every β with $0 < \beta \leq \beta_1$ there exist positive scalars a_1, b_1 such that for every a, b with $0 < a \leq a_1$, $0 < b \leq b_1$ the sequence $\{K_k\}$ generated by the equation

$$K_{k-1} = A'[(1-\beta)K_k^{-1} - \frac{1-\beta}{\beta}GQ^{-1}G' + \frac{1}{b}BR_1^{-1}B']^{-1}A + a\Psi_1 \qquad (A.17)$$

$$K_N = a\Psi_1 \qquad (A.18)$$

converges to a positive definite symmetric matrix $K_{-\infty}$ and furthermore we have

$$\frac{1}{\beta}GQ^{-1}G' < K_k^{-1} \qquad \text{for all k} \qquad (A.19)$$

The Lemma A.3 will be used to reduce the proof of the proposition to proving a different statement. We first make the following observation:

If for some β, a, b the inequality (A.19) holds for all k then

$$K_k \leq K_{k-1} \qquad \text{for all k} \qquad (A.20)$$

We prove this fact by induction. For k = N we have from equation (A.17) $a\Psi_1 = K_N \leq K_{N-1}$. Assume $K_{k+1} \leq K_k$. Then from (A.17)

$$\begin{aligned} K_{k-1} &= A'[(1-\beta)K_k^{-1} - \frac{1-\beta}{\beta}GQ^{-1}G' + \frac{1}{b}BR_1^{-1}B']^{-1}A + a\Psi_1 \\ &\geq A'[(1-\beta)K_{k+1}^{-1} - \frac{1-\beta}{\beta}GQ^{-1}G' + \frac{1}{b}BR_1^{-1}B']^{-1}A + a\Psi_1 = K_k \end{aligned}$$

or $K_k \leq K_{k-1}$.

Now if we could find a positive definite symmetric matrix S such that

$$K_k \leq S \qquad \text{for all k} \qquad (A.21)$$

and furthermore

$$\frac{1}{\beta}GQ^{-1}G' < S^{-1} \qquad (A.22)$$

then the inequality (A.19) would be satisfied for all k and from the relations (A.20) and (A.21) we would have $K_k \leq K_{k-1} \leq S$ for all k. This in turn would imply by Lemma A.3 that the sequence $\{K_k\}$ converges to a positive definite symmetric matrix $K_{-\infty}$.

Thus in order to prove the proposition it is sufficient to demonstrate a positive scalar $\beta_1 < 1$ and for every β, $0 < \beta \leq \beta_1$ positive scalars a_1, b_1 such that for all a, b, $0 < a \leq a_1$, $0 < b \leq b_1$ there exists a matrix S satisfying relations (A.21) and (A.22).

Since the pair (A, B) is stabilizable there exists an m x n matrix L such that the matrix (A-BL) is stable. Let β_1 be a scalar, $0 < \beta_1 < 1$ such that the matrix

$$F_1 = (1 - \beta_1)^{-1/2}(A - BL)$$

is also stable. Clearly such a scalar exists and for every β, $0 < \beta \leq \beta_1$, the matrix

$$F = (1 - \beta)^{-1/2}(A - BL) \tag{A.23}$$

is also stable. It will be shown below that β_1 satisfies the requirements of the proposition.

Let now for any β, $0 < \beta \leq \beta_1$, $\overline{A} = (1 - \beta)^{-1/2}A$, $\overline{B} = (1 - \beta)^{-1/2}B$. The equation (A.17) can be rewritten as

$$K_{k-1} = \overline{A}'[K_k^{-1} - \frac{1}{\beta}GQ^{-1}G' + \frac{1}{b}\overline{B}R_1^{-1}\overline{B}']^{-1}\overline{A} + a\Psi_1 \tag{A.24}$$

and since the matrix F of equation (A.23) can be written as $F = \overline{A} - \overline{B}L$, by using Lemma A.4 we obtain that

$$K_k \leq \Sigma_k \qquad \text{for all k} \tag{A.25}$$

where Σ_k is the solution of the equation

$$\Sigma_{k-1} = F'(\Sigma_k^{-1} - \frac{1}{\beta}GQ^{-1}G')^{-1}F + a\Psi_1 + bL'R_1L \qquad (A.26)$$

$$\Sigma_N = a\Psi_1 \qquad (A.27)$$

provided that $0 < \Sigma_k$ for all k and

$$\frac{1}{\beta}GQ^{-1}G' < \Sigma_k^{-1} \qquad \text{for all k.} \qquad (A.28)$$

Now by Lemma A.5 if

$$a\Psi_1 + bL'R_1L < \frac{\mu}{1-q\mu}[(1-q\mu)M - F'MF] \qquad (A.29)$$

where M is a positive definite symmetric matrix and q, μ are positive scalars such that

$$F'MF < M \qquad (A.30)$$

$$\frac{1}{\beta}GQ^{-1}G' < qM^{-1} \qquad (A.31)$$

$$F'MF < (1 - q\mu)M \qquad (A.32)$$

the sequence $\{\Sigma_k\}$ generated by equation (A.26) converges to a positive definite symmetric matrix $\Sigma_{-\infty}$ and we have $\frac{1}{\beta}GQ^{-1}G' < \Sigma_k^{-1}$ for all k and $\Sigma_k < \mu M$ for all k. Thus for a, b satisfying the relation (A.29) we have from (A.25) that for $S = \mu M$

$$K_k < S \qquad \text{for all k}$$

Furthermore since $0 < q\mu < 1$ from equation (A.31)

$$\frac{1}{\beta}GQ^{-1}G' < qM^{-1} < \frac{1}{\mu}M^{-1} = S^{-1}$$

Therefore for $S = \mu M$ the relations (A.21) and (A.22) are satisfied and consequently the sequence $\{K_k\}$ converges to a positive definite symmetric matrix $K_{-\infty}$ and the inequality (A.19) is satisfied for every a, b satisfying the inequality (A.59). It is clear from Lemma A.1 that there exist positive scalars a_1, b_1 such that for every a, b, $0 < a \leq a_1$, $0 < b \leq b_1$ the inequality (A.19) is satisfied. Any such scalars a_1, b_1 satisfy the requirements of the proposition. Q.E.D.

We next present the proof of Proposition 3.2.

<u>Proof of Proposition 3.2</u>: We will prove that for the state of the closed-loop system

$$x_{k+1} = (A - BL)x_k \qquad (A.33)$$

with

$$L = (R + B'F_{-\infty}B)^{-1}B'F_{-\infty}A.$$

and for any positive integer N we have

$$x_N'K_{-\infty}x_N + \sum_{k=0}^{N-1} x_k'(\Psi + L'RL)x_k < x_0'K_{-\infty}x_0 \qquad (A.34)$$

Then from the positive definiteness of $(\Psi + L'RL)$ asymptotic stability of the system (A.33) follows.

To prove the relation (A.34) we will use the following identity which is familiar from Riccati equation theory. This identity can be verified in a straight forward manner

$$K_{-\infty} = (A - BL)'F_{-\infty}(A - BL) + \Psi + L'RL \qquad (A.35)$$

We also have

$$F_{-\infty}^{-1} = (1-\beta)(K_{-\infty}^{-1} - \frac{1}{\beta}GQ^{-1}G') < K_{-\infty}^{-1} - \frac{1}{\beta}GQ^{-1}G' \leqq K_{-\infty}^{-1}$$

implying

$$K_{-\infty} < F_{-\infty} \qquad\qquad (A.36)$$

By using relations (A.35) and (A.36) we now have:

$$x_N' K_{-\infty} x_N + \sum_{k=0}^{N-1} x_k'(\Psi + L'RL)x_k$$

$$< x_N' F_{-\infty} x_N + \sum_{k=0}^{N-1} x_k'(\Psi + L'RL)x_k$$

$$= x_{N-1}'[(A-BL)'F_{-\infty}(A-BL) + \Psi + L'RL]x_{N-1} + \sum_{k=0}^{N-2} x_k'(\Psi + L'RL)x_k$$

$$= x_{N-1}' K_{-\infty} x_{N-1} + \sum_{k=0}^{N-2} x_k'(\Psi + L'RL)x_k$$

$$- -$$

$$< x_1' K_{-\infty} x_{N-1} + x_0'(\Psi + L'RL)x_0$$

$$< x_0'[(A-BL)'F_{-\infty}(A-BL) + \Psi + L'RL]x_0 = x_0' K_{-\infty} x_0$$

or

$$x_N' K_{-\infty} x_N + \sum_{k=0}^{N-1} x_k'(\Psi + L'RL)x_k < x_0' K_{-\infty} x_0$$

Q.E.D.

REFERENCES

(A1)   Aoki, M., "Optimization of Stochastic Systems", Academic Press, New York, N.Y., 1967.

(At1)   Athans, M., and Falb, P.L., "Optimal Control", McGraw Hill, New York, 1966.

(B1)   Bertsekas, D.P., and Rhodes, I.B., "On the Minimax Reachability of Target Sets and Target Tubes", Automatica, March, 1971.

(B2)   Bertsekas, D.P., and Rhodes, I.B., "Recursive State Estimation for a Set-Membership Description of the Uncertainty", IEEE Transactions on Automatic Control, AC-16, April, 1971.

(B3)   Bertsekas, D.P., and Mitter, S.K., "Steepest Descent for Optimization Problems with Nondifferentiable Cost Functionals", presented in the Fifth Annual Princeton Conference of Information Sciences and Systems, Princeton, N.Y., March, 1971.

(Bl1)   Blackwell, D., and Girshick, M.A., "Theory of Games and Statistical Decisions", John Wiley and Sons, Inc., New York, N.Y., 1954.

(Br1)   Bryson, A.E., and Ho, Y.C., "Applied Optimal Control", Blaisdell Publishing C., Waltham, Mass., 1969.

(Bra1)   Bram, J., "The Lagrange Multiplier Theorem for Max-Min with Several Constraints", J. SIAM Appl. Math., Vol. 14, No. 4, 1966.

(D1)   Dem'yanov, V.F., "The Solution of Several Minimax Problems", Kibernetica, Vol. 2, No. 6, 1966.

(D2)   Dem'yanov, V.F., and Rubinov, A.M., "Minimization of Functionals in Normed Spaces", SIAM J. Control, Vol. 6, 1968, pp. 73-88.

(D3)   Dem'yanov, V.F., and Rubinov, A.M., "Approximate Methods in Optimization Problems", Am. Elsevier Publ. Co., New York, 1970.

(Da1)   Danskin, J.M., "The Theory of Max-Min with Applications", J. SIAM Appl. Math., Vol. 14, No. 4, 1966.

(Da2)   Danskin, J.M., "The Theory of Max-Min", Springer Verlag, 1967.

(Del)   Delfour, M.C., and Mitter, S.K., "Reachability of Perturbed Systems and Min-Sup Problems", SIAM J. Control, Vol. 7, No. 4, 1969.

(F1)   Feldbaum, A.A., "Theory of Dual Control", I, II, III, IV, Automation and Remote Control, Vol. 21, No. 9, No. 11, 1960, Vol. 22, No. 1, No. 2, 1961.

REFERENCES (Cont'd)

(F2)   Feldbaum, A.A., "Optimal Control Systems", Academic Press, New York, N.Y., 1965,

(Fr1)   Fraser, D.C., and Potter, J.E., "The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters", IEEE Transactions AC-14, No. 4, 1969.

(G1)   Gunckel, T.L., and Franklin, G.F., "A General Solution for Linear, Sampled-Data Control", Trans. ASME, J. Basic Engrg., Series D, Vol. 85, 1963, pp. 197-201.

(H1)   Hnyilicza, E., "A Set-Theoretic Approach to State Estimation", M.S., Thesis, Department of Electrical Engineering, M.I.T., 1969.

(He1)   Heins, W., and Mitter, S.K., "Conjugate Convex Functions, Duality, and Optimal Control Problems I: Systems Governed by Ordinary Differential Equations", Information Sciences, Vol. 2, 1970, pp. 211-243.

(I1)   Isaacson, E., and Keller, H.B., "Analysis of Numerical Methods", John Wiley and Sons, New York, 1966.

(Is1)   Isaacs, R., "Differential Games", John Wiley and Sons, New York, N.Y., 1965.

(J1)   Joseph, P.D., and Tou, J.T., "On Linear Control Theory", AIEE Transactions, Vol. 80, Part II, 1961, pp. 193-196.

(K1)   Kalman, R.E., and Koepcke, R.W., "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes", Transactions ASME, Vol. 80, 1958.

(Ka1)   Kantorovich, L.V., and Akilov, G.P., "Functional Analysis in Normed Spaces", MacMillan, New York, 1964.

(Ku1)   Kuhn, H.W., "Extensive Games", Proc. Nat. Acad. Sci. Wash. Vol. 36, 1950, pp. 570-576.

(Ku2)   Kuhn, H.W., "Extensive Games and the Problem of Information", Ann. Math. Studies, Vol. 28, 1953, pp. 193-216.

(L1)   Lee, E.B., and Markus, L., "Foundations of Optimal Control Theory", John Wiley and Sons, New York, N.Y., 1967.

(Lu1)   Luenberger, D.G., "Optimization by Vector Space Methods", John Wiley and Sons, New York, N.Y., 1969.

REFERENCES (Cont'd)

(Lu2)   Luenberger, D.G., "Control Problems with Kinks", IEEE Transactions, AC-15, No. 5, 1970.

(P1)   Pontryagin, L.S., Boltyanskii, V., Gankrelidze, R., and Mishchenko, E., "The Mathematical Theory of Optimal Processes", Interscience Publishers, Inc., New York, 1962.

(Ps1)   Pschenichnyi, B.N., "Dual Methods in Extremum Problems," I, II, Kibernetica, Vol. 1, No. 3, pp. 89-95, Vol. 1, No. 4, pp. 64-69, 1965.

(R1)   Rockafellar, R.T., "Convex Analysis", Princeton University Press, Princeton, N.J., 1970.

(R2)   Rockafellar, R.T., "Conjugate Convex Functions in Optimal Control and the Calculus of Variations", Journal of Math. Analysis and Applications, Vol. 32, 1970, pp. 174-222.

(Ra1)   Rauch, H.E., Tung, F., Striebel, C.T., "Maximum Likelihood Estimates of Linear Dynamic Systems", AIAA Journal, Vol. 3, No. 8, 1965, pp. 1445-1450.

(Rh1)   Rhodes, I.B., "Optimal Control of a Dynamic System by two Controllers with Conflicting Objectives", Ph.D. Dissertation, Department of Electrical Engineering, Stanford University, 1967.

(S1)   Schweppe, F.C., "Recursive State Estimation; Unknown but Bounded Errors and System Inputs", IEEE Transactions, AC-13, No. 1, 1968.

(S2)   Schweppe, F.C., "Uncertain Dynamic Systems", Class Notes 6.606, M.I.T., 1969.

(S3)   Schweppe, F.C., "Uncertain Dynamic Systems", Academic Press, New York, N.Y., (to appear).

(Sa1)   Salmon, D.M., "Minimax Controller Design", IEEE Transactions, AC-13, No. 4, 1968.

(Sc1)   Schlaepfer, F.M., "Set Theoretic Estimation of Distributed Parameter Systems", Rep. ESL-R-413, M.I.T., 1970.

(St1)   Striebel, C.T., "Sufficient Statistics in the Optimal Control of Stochastic Systems", J. of Math. Analysis and Applications, Vol. 12, 1965.

REFERENCES (Cont'd)

(Sw1) Sworder, D.D., "Minimax Control of Discrete Time Stochastic Systems", <u>SIAM J. Control,</u> Vol. 2, 1965, pp. 443-449.

(Sw2) Sworder, D.D., "Optimal Adaptive Control Systems", Academic Press, New York, N.Y., 1966.

(Su1) Sussman, R., "Optimal Control of Systems with Stochastic Disturbances", Electronics Research Lab., University of California, Berkeley, Report No. 63-20, November, 1963.

(T1) Tse, E.T., "On the Optimal Control of Linear Systems with Incomplete Information", Report ESL-R-412, M.I.T., January, 1970.

(T2) Tse, E.T., and Athans, M., "Optimal Minimal-Order Observer-Estimators for Discrete Linear Time-Varying Systems", <u>IEEE Transactions,</u> AC-15, No. 4, 1970.

(V1) Von Neumann, J. and Morgenstern, D., "Theory of Games and Economic Behavior", Princeton University Press, Princeton, N.J., 1944.

(W1) Witsenhausen, H.S., "Minimax Control of Uncertain Systems", Rep. ESL-R-269, M.I.T., May, 1966.

(W2) Witsenhausen, H.S., "A Minimax Control Problem for Sampled Linear Systems", <u>IEEE Transactions,</u> AC-13, No. 1, 1968.

(W3) Witsenhausen, H.S., "Sets of Possible States of Linear Systems Given Perturbed Observations", <u>IEEE Transactions,</u> AC-13, No. 5, 1968.

(Wa1) Wald, A., "Statistical Decision Functions", John Wiley and Sons, New York, N.Y., 1950.

(Wo1) Wonham, W.M., "On Pole Assignment in Multi-input Controllable Linear Systems", <u>IEEE Transactions on Automatic Control,</u> Vol. AC-12, No. 6, 1967.

(Wo2) Wonham, W.M., "On a Matrix Riccati Equation of Stochastic Control", <u>SIAM J. of Control,</u> Vol. 6, No. 4, 1968.

(Wo3) Wonham, W.M., "On the Separation Theorem of Stochastic Control", <u>SIAM J. Control,</u> Vol. 6, 1968.