CONTROL OF FINITE-STATE, FINITE-MEMORY STOCHASTIC SYSTEMS

Nils R. Sandell, Jr.

Electronic Systems Laboratory

Decision and Control Sciences Group

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139

Department of Electrical Engineering

CONTROL OF FINITE-STATE, FINITE-MEMORY

STOCHASTIC SYSTEMS

by

Nils R. Sandell, Jr.

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

i

CONTROL OF FINITE-STATE, FINITE-MEMORY

STOCHASTIC SYSTEMS

by

Nils R. Sandell, Jr.

B.E.E. University of Minnesota

1970

M.S.   Massachusetts Institute of Technology

1971

E.E.   Massachusetts Institute of Technology

1973

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 1974

Signature of Author... ....
                        Department of Electrical Engineering, May 17, 1974

Certified by.... ....
                                                Thesis Supervisor

Accepted by...
            Chairman, Departmental Committee on Graduate Students

CONTROL OF FINITE-STATE, FINITE-MEMORY
STOCHASTIC SYSTEMS

by

Nils R. Sandell, Jr.

Submitted to the Department of Electrical Engineering on May 17, 1974
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

ABSTRACT

This thesis discusses a generalized problem of stochastic control,
in which multiple controllers with different data bases are present.
The vehicle for the investigation is the finite-state, finite-memory
(FSFM) stochastic control problem. For this problem, the usual
technique of stochastic dynamic programming does not apply. Instead,
optimality conditions are obtained by deriving an equivalent
deterministic optimal control problem.

A FSFM minimum principle is obtained via the equivalent deterministic
problem. The minimum principle suggests the development of a
numerical optimization algorithm, the min-H algorithm. The relation-
ship between the sufficiency of the minimum principle (which is in
general only a necessary condition) and the informational properties
of the problem is investigated.

Dynamic programming functional equations for the FSFM problem are
also obtained from the equivalent deterministic problem. Both the
finite and infinite horizon cases are considered. Numerical
solution of the functional equations is discussed.

To illustrate the general theory, a problem of hypothesis testing
with 1-bit memory is investigated. The discussion illustrates the
application of control theoretic techniques to information processing
problems.

THESIS SUPERVISOR: Michael Athans

TITLE: Professor of Electrical Engineering

ACKNOWLEDGEMENT

I would like to express my deep appreciation to Professor Michael Athans for his indispensable support during the thesis research, and, more generally, for his profound influence on my professional development.

I would like to thank Dr. Hans Witsenhausen (Bell Telephone Laboratories) and Professor Peter Elias for serving as my thesis readers. Those familiar with Dr. Witsenhausen's research will recognize the strong influence of his work on this thesis. Professor Elias' comments were extremely helpful for placing the thesis in perspective.

Numerous interesting discussions with my fellow graduate students contributed to this work. I would particularly like to thank C.Y. Chong, Paul Houpt, David Martin, and Loren Platzman in this respect.

The Publications and Drafting Departments of the M.I.T. Electronic Systems Laboratory deserve thanks for their efficient preparation of this manuscript. Special thanks go to Chris Chiotasso, Nancy Galvin, and Arthur Giordani.

Finally, I would like to thank my wife Yvonne for her love and care during the past four years.

.

DEDICATION

To My Parents

Nils R. Sandell, Sr.

and

Carmela V. Sandell

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Stochastic Control and Large Scale Systems

The fundamental problem of control engineering is illustrated

in Figure 1.1.1. A fixed plant is given with certain variables

(inputs) available for manipulation and other variables (outputs)

available for observation. A controller must be designed to

choose the plant inputs based on the observations so that the plant

behaves in a desired fashion.

In deriving a mathematical model for the plant, phenomena

which cannot be adequately explained by simple deterministic models

are commonly treated as stochastic disturbances. Stochastic

optimal control theory has been developed for problems of this type.

While it is true that the theory at present has been unsuccessful

in producing explicit solutions to practical non-linear problems,

nevertheless the theory provides a useful perspective and a con-

venient framework for deriving suboptimal, but practical and

feasible policies.

Consider for example the Safeguard ballistic missile defense

system, which can be considered a large stochastic control problem.

Of course, the problem is too complicated to be solved in this

input variables

output variables

plant

controller
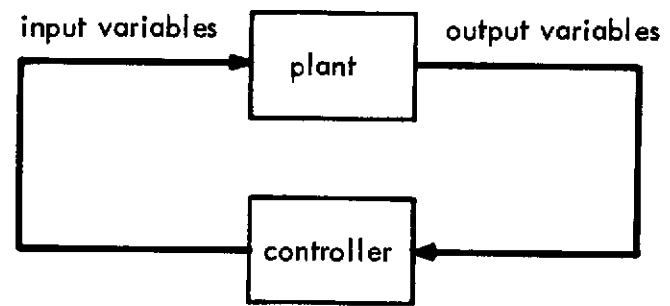
Figure 1.1.1   Control System

formulation, but parts of the problem are tractable. Just to

cite one example, the Kalman filtering theory is used for the

tracking function. But of more fundamental importance is the

perspective available from adopting the stochastic control viewpoint:

the state space formalism, the explicit treatment of uncertainty,

the identification of the computer with the controller and the

radar and missile sites as the sensors and actuators, and the

explicit statement of system goals with their relative importance.

While stochastic control has doubtless been useful for certain

problems, there has recently been an increase in interest in the

more difficult problems of large scale engineering systems.

These systems (Figure 1.1.2) are characterized by the presence of

multiple controllers acting on different data bases and affecting

different aspects of total system performance. Since classical

stochastic control theory is restricted to systems with a single

controller possessing perfect memory of all past sensor outputs

and actuator inputs (the so called classical information pattern),

the need for a generalized theory is apparent. Such a theory must

subsume classical stochastic control, so that explicit optimal

solutions to realistic design problems cannot be expected. But

what can be accomplished is the establishment of a framework in which

the information interface problems that arise in multiple controller

systems can be viewed.
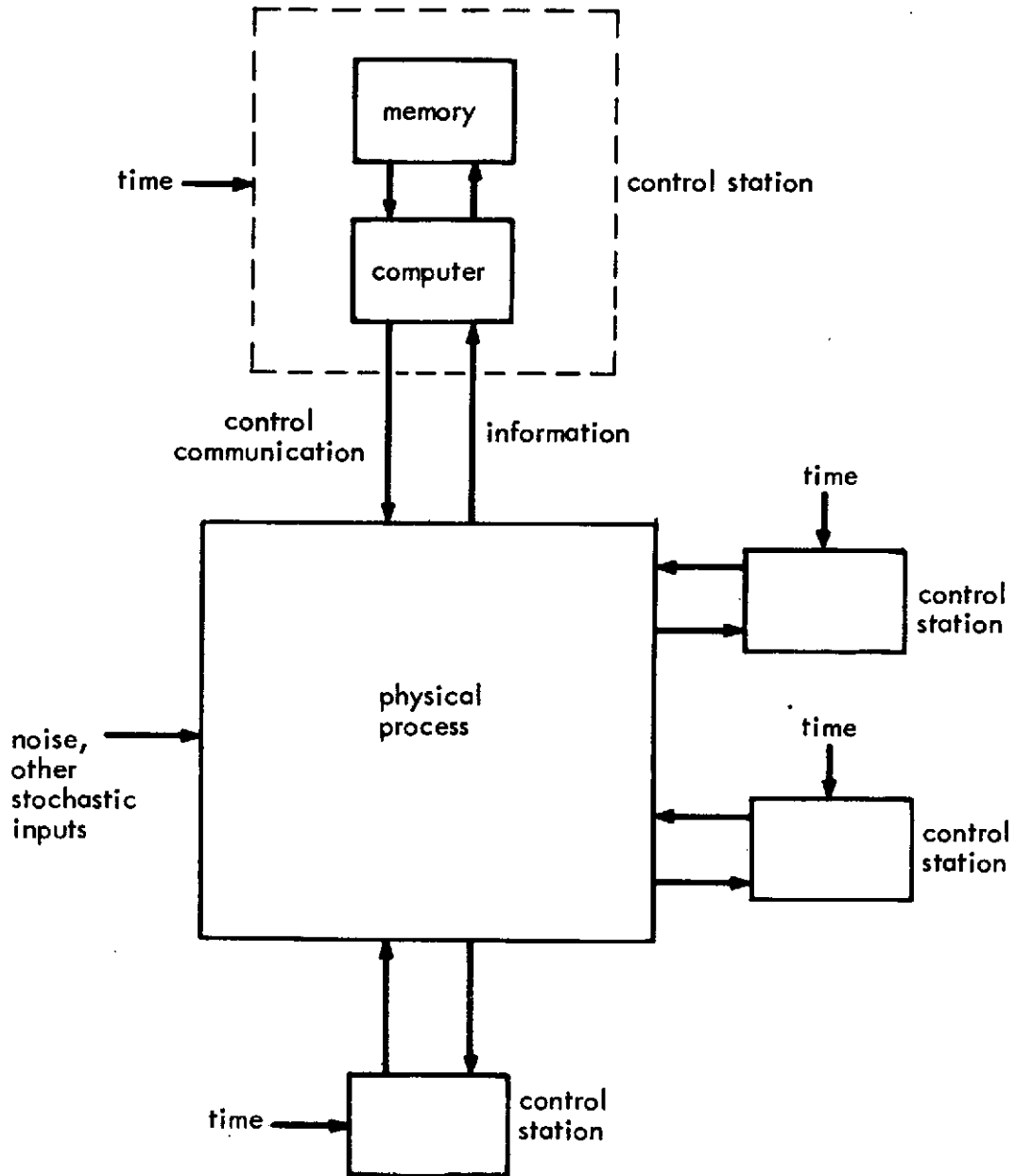
Figure 1.1.2   System with Multiple Controllers

1.2 Background

Tentative steps in the direction of a generalized theory of stochastic control have been taken by a number of workers. Inspiration has come from a number of fields other than classical stochastic control theory. These include the theories of games, statistical decisions, multilevel hierarchical systems, teams, and communications.

The crucial issue in generalized stochastic control is the interaction between information and decision. This issue arises unavoidably in the Von Neuman game theory [V1,L1,O1] due to the presence of more than one player. Unfortunately, attention in game theory has focused on the so called normal form of the game. In this form the dynamical and informational aspects of the game are suppressed by introduction of the notion of strategy. A generalized stochastic control problem can be considered a non zero-sum game, and so it has a normal form. Of course, no insight is gained from this reduction. More useful for non-classical stochastic control is the work that game theorists have performed on the extensive form of the game [K1,K2,D1,T1]. It is here, for example, that the important notion of the information pattern arises.

Another area in which the issue of the interaction between information and decision naturally arises is statistical decision theory. Statistical decision theory is a mathematical discipline that resulted from the infusion of ideas of game theory into the more

traditional statistical theory of Fisher, Neyman and Pearson, and their followers. The synthesis was largely performed by Abraham Wald, and culminated in his book Statistical Decision Functions [Wal]. Wald's formulation is still important, but an alternative formulation by the Bayesian statisticians has grown in popularity [Sal, Ral]. Statistical decision theory contains several ideas important in stochastic control. One example is the notion of a <u>sufficient</u> <u>statistic</u>. Another example is contained in Wald's treatment of the sequential problem. This treatment contains ideas of dual control and of dynamic programming.

The theory of multilevel hierarchical systems is due to Mesarovic [Mel], who drew inspiration from the study of decentralized structures in economics and management [Arl, Soll] and from large scale mathematical programming [Lal, Wisl]. However, Mesarovic's model is deterministic and problems of information flow appear only implicitly. More recently Chong [Cl] has investigated a stochastic version of a two level, hierarchical system in which the interaction between information and control appears explicitly.

Team theory [M1,M2,R1] is closely related to statistical decision theory. According to Radner [Rl], team theory arose from "... attempts by several workers to analyze some of the many-person aspects of organizations that are present even in the absence of many-person game complications...". Team theory is actually a special <u>static</u> case of non-classical stochastic control. It is

important since an explicit solution to the quadratic-Gaussian team problem is known, so that the relative efficacy of different information structures can be compared [R2].

Communications theory is another area that is a special case of non-classical stochastic control. There are two controllers in a communication problem, the encoder and the decoder. By the very nature of the problem, the decoder does not know either the observation (source outputs) or controls (channel inputs) of the encoder. Information theory was invented by Shannon [Sh1,Gal] to deal with problems of this nature.

Although non-classical stochastic control theory has drawn inspiration from a number of cognate disciplines, it is undeniably a direct outgrowth of a critical look at the foundations of classical stochastic control performed by several authors. The fundamental theoretical tool in stochastic control is the dynamic programming algorithm [Bl,Fl,Aol,Hl]. Although the algorithm can only be explicitly carried out in certain special cases, it nevertheless provides a convenient conceptual framework in which theoretical questions of existence, uniqueness, randomization, etc. can be posed and answered. The critical underlying assumption for the validity of dynamic programming is the classical information pattern: one controller with access to all past observations and controls [Chl,St1]. Thus an examination of the foundations of dynamic

programming suggests the non-classical stochastic control problem
as an extension.

Explicit consideration of non-classical stochastic control
theory began with the work of Witsenhausen [W1,W2,W3,W4].
Witsenhausen gave an example of a linear-quadratic-Gaussian
(LQG) stochastic control problem for which the optimal control laws
are nonlinear in [W1]. In [W2], he examined the fundamental issue
of when a general stochastic control problem (or game) is well-
posed. In [W3], the status of the new theory was surveyed, with
the introduction of a useful system of notation and the listing
of a number of "assertions" which might be turned into theorems
by appropriate technical assumptions. In [W4], a maximum
principle (for control laws) was derived.

Non-classical stochastic control has drawn the attention of
other workers. Athans and a number of his students have
investigated suboptimal solutions to certain non-classical
problems [C2,Kw1,Ca1]. Y.C. Ho and his student K.C. Chu have
classified information patterns and identified some for which the
optimal control laws for the LQG case are linear [Ho1,Chu1]. Aoki
has found a suboptimal solution for the control sharing information
pattern [Ao2]. Bismut has given an example in which the
interaction between information and control is clearly exhibited
[Bi1], and Sandell and Athans [S1] have used Bismut's idea to

explicitly characterize the optimal nonlinear solution of the

control-sharing LQG stochastic control problem.

1.3 Summary of Thesis

Research in non-classical stochastic control to date has been handicapped by the absence of a rich class of tractable examples. In classical stochastic control, the LQG (linear-quadratic-Gaussian) problems are a readily solvable class useful for motivation and for practical applications. Unfortunately, the solution of the non-classical LQG problem is difficult and unknown [W1,S1].

The present work is aimed at easing this difficulty. Attention is restricted to the case of finite-state, finite-memory (FSFM) stochastic systems. For these problems, an elegant and elementary theory can be developed. The optimality conditions for these problems have a special structure that can be exploited to develop numerical optimization techniques. Evidence of the importance of the problem is given by the interest in a special case of the problem in the operations research literature [How1,How2].

The FSFM model is introduced in Chapter II. It is demon-strated that a number of apparently more general problems can be reduced to FSFM stochastic control problems An example of a FSFM problem is given that illustrates the important notion of a signaling strategy. The chapter concludes with the derivation of a deterministic optimal control problem equivalent to the FSFM stochastic control problem.

A FSFM minimum principle is derived in Chapter III. The minimum principle is a necessary condition for optimality, but is not sufficient in general as is shown by a simple example. However, in the absence of signaling strategies, the minimum principle can be strengthened to give a sufficient condition. A numerical optimization algorithm, the Min-H algorithm, is developed based on the minimum principle.

The dual dynamic programming functional equations for forward and backward induction are stated in Chapter IV. Several approaches to the numerical solution of these equations are suggested, and their implementation is illustrated by an example.

Chapter V considers the infinite horizon version of the FSFM problem. The Value and Policy Iteration methods are derived for a version of the problem with discounted cost, and their numerical implementation discussed. Policy Iteration is illustrated by an example. This example has the interesting property that the optimal control law sequence is non-stationary.

In Chapter VI, the problem of hypothesis testing of Bernoulli trials with a 1-bit memory is considered. Application of the minimum principle suggests a class of non-obvious, but intuitively desirable strategies. This result provides considerable justification for the use of control-theoretic methods in information theoretic problems.

Chapter VII consists of conclusions and suggestions for future research.

1.4 Contributions of Thesis

The major contributions of this research are:

(1) The formulation of the FSFM problem.

(2) The minimum principle and the person-by-person
min-H algorithm for the FSFM problem.

(3) The relation of the information properties of the
FSFM problem to the optimality conditions.

(4) The extension of the Sondik algorithm to the FSFM
problem.

(5) Formulation of the infinite horizon FSFM problem with
discounting.

(6) Extension of Value and Policy Iteration methods to
the FSFM problem.

(7) Extension of Sondik's implementation of Policy Iteration
to FSFM problems.

(8) Demonstration of the potential value of control-theoretic
methods in information handling systems via the hypothesis
testing problem.

CHAPTER II

THE FSFM STOCHASTIC CONTROL PROBLEM

In this chapter, the finite-state, finite-memory stochastic
control problem is introduced. It is shown that FSFM problems are
a fairly general class of non-classical stochastic control problems.
An example is given illustrating the interesting signaling strategies
that occur in FSFM problems. The chapter concludes with the development
of a determistic optimal control problem equivalent to the FSFM problem.

2.1 Formulation

The systems studied are described by the state equation

$$x(t) = f_t(x(t-1), u(t), q(t)) \qquad (2.1.1)$$

where $x(t) \in X_t$ for $t = 0,1,2,\ldots,T$ and $u(t) \in U_t$, $q(t) \in Q_t$ for
$t = 1,2,\ldots,T$. The finite sets X_t, U_t, and Q_t are referred to as the
state set, the input set, and the uncertainty set, respectively.
Associated with the state equation is a cost function

$$J = \phi_T(x(T)) + \sum_{t=1}^{T} h_t(x(t-1), u(t)) \qquad (2.1.2)$$

where $h_t : X_{t-1} \times U_t \rightarrow R$ and $\phi_T : X_T \rightarrow R$ (R = real numbers).

The interpretation of the equations is as follows. The state
equation (2.1.1) models some controlled, uncertain physical process.
The variables $x(t)$ represent the possible states of the process, the
variables $u(t)$ are the inputs of the controller, and the variables $q(t)$

represent the stochastic effects present. The system's performance is
measured by its cost of operation as expressed by (2.1.2).

The designer's problem is to specify the system controller.
The controller is specified by a sequence of <u>control laws</u>

$$\gamma_t : X_{t-1} \rightarrow U_t \qquad t = 1,2,\ldots,T. \qquad (2.1.3)$$

The interpretation of the control law is that when the process is in
state $x(t-1)$, the controller applies input $u(t) = \gamma_t(x(t-1))$. The fact
that all control laws are not feasible (due to various physical
constraints) is recognized by specifying the set of admissible control
laws

$$\Gamma_t \subset U_t^{X_{t-1}} \qquad (2.1.4)$$

at time t, $t = 1,2,\ldots,T$. The designer is constrained to choosing
$\gamma = (\gamma_1,\ldots,\gamma_T) \in \Gamma$, where

$$\Gamma = \Gamma_1 \times \Gamma_2 \times \ldots \times \Gamma_T. \qquad (2.1.5)$$

An admissible control law sequence $\gamma \in \Gamma$ will be called a <u>design</u>, and the
set Γ will be referred to as the <u>set of admissible designs</u>.

The design γ should be chosen so that the system operates with
minimum cost. Notice, however, that the cost (2.1.2) of operation of
the system is not determined solely by γ, but depends on the (uncertain)
values of $x(0)$, $q(1)$, $\ldots$, $q(T)$. The difficulty is resolved by adopting
a Bayesian viewpoint: all the uncertain variables are assumed to be
random variables with a known joint probability distribution.

In the FSFM model, it is assumed that probability functions

$\pi_0 : X_0 \rightarrow [0,1]$ and $p_t : Q_t \rightarrow [0,1]$ are given. The probability space

(Ω, F, P) is then defined as follows. The sample space Ω and field of

events F are

$$\Omega = X_0 \times Q_1 \times Q_2 \times \ldots \times Q_T \qquad (2.1.6)$$

$$F = P(\Omega) \qquad (2.1.7)$$

where $P(\Omega)$ is the power set of Ω (set of all subsets of Ω). The probab-

ility of a point $\omega = (x_0, q_1, q_2, \ldots, q_T) \in \Omega$ is

$$P(\{\omega\}) = \pi(x_0) \, p_1(q_1) \, p_2(q_2) \, \ldots \, p_T(q_T) \qquad (2.1.8)$$

and the probability of an arbitrary event of F is the sum of the

probabilities of its points.

Given $\gamma \in \Gamma$, the corresponding expected value of J can be computed

in several ways. Define

$$X = X_0 \times X_1 \times \ldots \times X_T, \qquad (2.1.9)$$

$$U = U_1 \times U_2 \times \ldots \times U_T. \qquad (2.1.10)$$

The system of feedback equations

$$x(t) = f_t(x(t-1), \gamma_t(x(t-1)), q(t)), \quad t = 1,2,\ldots,T \quad (2.1.11)$$

has a unique solution $(x(0), x(1), \ldots, x(T)) \in X$ for each $(x(0), q(1),$

$\ldots, q(T)) \in \Omega$. This is a trivial consequence of the casual nature of

the setup.[1] Thus, there exists an unique solution map

$$S_\gamma : \Omega \rightarrow X \qquad (2.1.12)$$

that gives the sequence of states resulting from a given design and a given sequence of stochastic inputs. Defining

$$X = P(X), \qquad (2.1.13)$$

a probability space $(X, \ X, \ PS_\gamma^{-1})$ is defined, where the probability PS_γ^{-1} is defined by

$$PS_\gamma^{-1}(x) = P(\{\omega : x = S_\gamma(\omega)\}). \qquad (2.1.14)$$

Similarly, define the map $R_\gamma : \Omega \rightarrow X \times U$ by

$$R_\gamma = S_\gamma \times (\gamma \circ S_\gamma). \qquad (2.1.15)$$

The corresponding probability space is $(X \times U, \ X \times U, \ PR_\gamma^{-1})$, where PR_γ^{-1} is defined in a similar fashion to PS_γ^{-1}.

Recall that the cost function J is a map $J : X \times U \rightarrow R$. Let $i_X : X \rightarrow X$ be the identity map, then maps $J_\gamma : \Omega \rightarrow R$ and $\tilde{J}_\gamma : X \rightarrow R$ can be defined by

$$J_\gamma = J \circ R_\gamma, \qquad (2.1.16)$$

$$\tilde{J}_\gamma = J \circ (i_X \times \gamma). \qquad (2.1.17)$$

---

[1] The crucial importance of the concept of casuality in assuring the existance of solutions to feedback equations has been demonstrated (for quite different models) by Witsenhausen [W2] and Willems [Wil].

Finally, the expected value of J can be computed in the following three ways[1]:

$$EJ_\gamma = \int_\Omega J_\gamma(\omega) \, dP(\omega),\qquad\qquad (2.1.18)$$

$$E_\gamma J = \int_{X \times U} J(x,u) \, dPR_\gamma^{-1}(x,u),\qquad\qquad (2.1.19)$$

$$E_\gamma J_\gamma = \int_X \tilde{J}_\gamma(x) \, dPS_\gamma^{-1}(x).\qquad\qquad (2.1.20)$$

But by theorem 39.C of Halmos [Hal],

$$EJ_\gamma = E_\gamma J = E_\gamma J_\gamma \triangleq J(\gamma).\qquad\qquad (2.1.21)$$

Thus the FSFM stochastic control problem is to find $\min_{\gamma \in \Gamma} J(\gamma)$, and the minimizing control law sequence $\gamma*$.

Since X and U are finite, it is clear that Γ is finite. Therefore, the cost functional $J(\gamma)$ can in principle be evaluated for each $\gamma \in \Gamma$, and the result tabulated. Since a finite set of real numbers always has a minimum, an optimal control law sequence exists, although it may not be unique. Moreover, since the minimum of a convex combination of a finite set of real numbers cannot be less than the smallest such number, it is clear that randomized designs offer no advantage.[2]

---

[1]For the finite spaces considered here,
$$\int_X f(x) \, dP(x) = \sum_X f(x) \, P(x).$$
Notice the notation EJ_γ, $E_\gamma J$, $E_\gamma J_\gamma$ indicating the dependence of the function and/or probability measure on γ.

[2]A randomized design is a sequence $\lambda_\gamma, \gamma \in \Gamma$, of numbers satisfying $\lambda_\gamma \geq 0$, $\sum_{\gamma \in \Gamma} \lambda_\gamma = 1$. If Ω is the set of such numbers, J is extended to Ω by the definition $J(\lambda) = \sum_{\gamma \in \Gamma} \lambda_\gamma J(\gamma)$. The use of randomized strategies is crucial in game theory [V1].

2.2 Generality of the Model

The FSFM model is motivated by the control system described in section 1.1. Besides the obvious limitation of the finiteness assumptions, there are several features of the general engineering control system of section 1.1 which are apparently not reflected in the FSFM model. The purpose of this section is to establish the generality of the FSFM model. It will be shown that the features of the general engineering control system can be incorporated in the FSFM model. This will be accomplished by reducing a set of apparently more general problems to the finite-state, finite-memory problem.

First consider the case in which the control laws are allowed to depend on the state only through a noisy observation

$$y(t) = g_t(x(t), \theta(t)) \qquad (2.2.1)$$

where $\theta(t) \in \Theta_t$, $y(t) \in Y_t$, and Θ_t, Y_t are finite sets. The random variables $\theta(t)$ are such that $\{x(0), \theta(0), q(1), \ldots, \theta(T-1), q(T)\}$ from a sequence of independent random variables. The problem is reduced to the preceeding by letting $X_t \times Y_t$ be the new state set, $Q_t \times \Theta_t$ be the new uncertainty set and

$$\begin{bmatrix} x(t) \\ \\ y(t) \end{bmatrix} = \begin{bmatrix} f_t(x(t-1), u(t), q(t)) \\ \\ g_t(f_t(x(t-1), u(t), q(t)), \theta(t)) \end{bmatrix} \qquad (2.2.2)$$

be the new state equation. The set Γ_t consists of maps $\gamma_t : X_{t-1} \times Y_{t-1} \to U_t$ satisfying $\gamma_t(x_1, y) = \gamma_t(x_2, y)$ for all $y \in Y_{t-1}$ and

$x_1, x_2 \in X_{t-1}$.

As a second example, suppose there are m observation equations

$$y^i(t) = g_t^i (x(t), \theta^i(t)) \qquad (2.2.3)$$

where for $i = 1, \ldots, m$, $y^i(t) \in Y_t^i$, $\theta^i(t) \in \Theta_t^i$, and Y_t^i, Θ_t^i are finite sets. The random variables $\theta^i(t)$ satisfy independence conditions similar to those of the variables of the first example. Moreover, suppose that

$$U_t = U_t^1 \times U_t^2 \times \ldots \times U_t^m \qquad (2.2.4)$$

and that $u^i(t)$ is to be chosen on the basis of observation of $y^i(t-1)$ alone. This is a case of the <u>dynamic team</u> [M2]. The reduction to the FSFM problem is accomplished by a state augmentation similar to that of (2.2.4). In this case,

$$\Gamma_t = \Gamma_t^1 \times \Gamma_t^2 \times \ldots \times \Gamma_t^m \qquad (2.2.5)$$

where Γ_{t+1}^i consists of maps from $X_t \times Y_t^1 \times \ldots \times Y_t^m$ to U_{t+1}^i that depend only on the variable $y^i(t) \in Y_t^i$.

For a third example, consider the case in which the control laws are restricted to dependence on a finite <u>memory set M_t</u>. The state space for this problem is $X_t \times M_t$, the control space is $U_t \times M_t$, and the state equation is

$$\begin{bmatrix} x(t) \\ m(t) \end{bmatrix} = \begin{bmatrix} f_t(x(t-1), u(t), q(t)) \\ v(t) \end{bmatrix} \qquad (2.2.6)$$

where $v(t) \in M_t$. The control laws $\tilde{\gamma}_t : X_{t-1} \times M_{t-1} \to U_t \times M_t$ are of the

form $\tilde{\gamma}_t = (\gamma_t, \eta_t)$, where $\gamma_t : X_{t-1} \times M_{t-1} \to U_t$, and $\eta_t : X_{t-1} \times M_{t-1} \to$

M_t is the memory update function.

For a fourth example, suppose that there are two control stations.

Control station 1 can communicate with control station 2 through a

channel described by the equation

$$r^2(t) = w_t^2 (s^1(t), \varepsilon^{12}(t)) \qquad (2.2.7)$$

where $s^1(t) \in S_t^1$ is the signal sent by control station 1, $r^2(t) \in R_t^2$

is the signal received by control station 2, and $\varepsilon^{12}(t) \in E_t^{12}$ is a

noise process. It is assumed that R_t^2, S_t^1, and E_t^{12} are finite sets,

and that the random variables of the sequence $\{\varepsilon^{12}(t)\}_{t=1}^T$ are in-

dependent of each other and all other random variables of the system.

This situation is handled by adding (2.2.7) to the state equations,

letting the control space of control station 1 be $U_t^1 \times S_t^1$, and by

letting the observation space of control station 2 be $Y_t^2 \times R_t^2$. The

control laws of control station 1 are the form $\tilde{\gamma}_t^1 : X_{t-1} \times M_{t-1}^1 \to$

$U_t^1 \times M_t^1 \times S_t^1$ where $\tilde{\gamma}_t^1 = (\gamma_t^1, \eta_t^1, \sigma_t^1)$. Here $\sigma_t^1 : X_{t-1} \times M_{t-1}^1 \to$

S_t^1 is the encoder of control station 1.

As a final example, suppose that the cost function is of the form

$$J = \phi_T(x(0), x(T)). \qquad (2.2.8)$$

(This formulation is important when communication or statistical decision

problems are considered as FSFM problems.) This situation is handled

by redefining the state space to be $X_t \times X_t$, and adding an equation of

the form

$$z(t) = z(t-1), \quad t = 1,2,\ldots,T, \tag{2.2.9}$$

to the state equations, where $z(0) = x(0)$.

It should be clear at this point that most of the important features of the general engineering control system of section 1.1 have been captured by the FSFM model. It is worth emphasizing that the memory management and communication handling tasks of the control stations can be incorporated into the FSFM problem. Thus the crucial data processing problems of systems with multiple controllers can be examined on an equal footing with the choice of actuator inputs.

2.3 Other Formulations

The case $\Gamma_t = U_t^{X_{t-1}}$ of the FSFM stochastic control problem is

the case of <u>complete state information</u> and has been extensively studied,

principally in the operations research literature [B1, How1, How2, H1,

Ku1]. The problem is usually referred to as a <u>Markovian decision</u>

<u>process</u>, and the formulation is slightly different. The state is not

defined by a state equation of the form (2.1.1), but is instead defined

as a controlled Mankov chain with transition probability $p_{ij}^{u}(t)$.

This is the probability of a transition from state i to state j at

time t when input u is applied to the system. Of course, (2.2.1) defines

a Markov chain with transition probabilities

$$p_{ij}^{u}(t) = p_t(\{q : j = f_t(i,u,q)\}). \qquad (2.3.1)$$

Since it is not difficult (in the finite state case) to realize a

given controlled Markov chain by a state equation, the two formulations

are in fact equivalent.[1]

An extension of the preceeding problem is the case of <u>incomplete</u>

<u>state information</u> treated extensively in both the control and operations

research literature [As1, H1, Dy1, St1, Ao1, So1, Sm1].[2] This problem

is also (for the finite-state, finite-horizon case) equivalent to a FSFM

stochastic control problem[3]. The incomplete state information is

---

[1]Establishing the equivalence of the two formulations for the case of
continuous state space is more difficult and (to the author's knowledge)
an unresolved problem.

[2]Control theorists have concentrated on the continuous state space case.
The treatment is usually quite formal; certain conditional probability
densities which may or may not be well defined are used extensively.

[3]The infinite-horizon version of the problem cannot be conveniently
handled by the FSFM techniques.

described by observations of the form (2.2.1) that can be adjoined to
the state equation as in (2.2.2). Moreover, all previous observations
and controls are remembered. Therefore, the memory set is
$M_t = Y_0 \times U_1 \times Y_1 \times \ldots \times U_{t-1} \times Y_{t-1}$, and the memory update functions
are constrained to sequentially storing the observations and controls
as they occur. Although the incomplete state information problem is
a special case of the general FSFM problem, the powerful perfect
memory assumption allows special techniques to be used that do not
apply to more general FSFM problems. These special techniques will be
discussed in more detail later.

The case $T=1$ of the FSFM problem includes both the non-sequential
Bayesian statistical decision problem [Sal, Ra1] and the team decision
problem [M1, M2, R1, R2] (for finite sets). The sequential Bayesian
problem (with perfect memory) is actually a special case of the
Markovian decision problem with incomplete state information and is
therefore a FSFM model. A sequential problem (hypothesis testing)
with a 1-bit (hence imperfect) memory is treated in Chapter 6.

Witsenhausen has given several stochastic control models that are
slightly less general than the FSFM model when restricted to finite sets
[W3, W4]. Witsenhausen shows that any sequential stochastic control
problem can be reduced to a certain <u>standard form</u>. The FSFM model is
a sequential stochastic control problem if the sets Γ_t satisfy the
condition $\Gamma_t = \{\gamma_t \in U_t^{X_{t-1}} : \gamma_t^{-1}(U_t) \subset \mathcal{D}_{t-1}\}$ for $t=1,2,\ldots,T$, where
$U_t = P(U_t)$ and $\mathcal{D}_t$ is a subfield of $X_{t-1} = P(X_{t-1})$. In this case, the FSFM
model is said to have a <u>simple information constraint</u>. Thus the FSFM

problem is more general than the sequential stochastic control problem since the most general constraint on the control laws is assumed.

Even if a stochastic control problem has a simple information constraint, it may be preferable to reduce the problem to a FSFM model rather than to Witsenhausen standard form. As Witenhausen says, "... alternative reductions leading to standard models with simpler state spaces may be possible in specific cases" [W4]. For problems with stochastic inputs which are independent from one time to the next, reduction to the FSFM model rather than to the standard model results in a simpler state space, but a more complicated state equation. It may be possible to formulate a FSFM problem with a fixed finite state set while the corresponding standard model requires a growing state set. This is an important computational advantage in general, and a crucial advantage when the infinite horizon problem is considered. In fact, the motivation for the development of the FSFM model was the development of a special class of Witsenhausen-type models for which an infinite horizon problem could be formulated.

Games in extensive form are a class of problems more general than FSFM problems. The original formulation due to Von Neuman and Morgenstern [V1] was improved upon by Kuhn [K1, K2] and subsequently by Aumann [Au1] and Witsenhausen [W2]. The theory of extensive games is more general than stochastic control theory in two significant ways. First, there are in general N players, each with a different cost function. Second, the theory of extensive games (in the Kuhn and Witsenhausen formulations) does not require that the time order in which the various decision

variables are selected is fixed in advance. The fact that there is
more than one cost function is the essential complication of game
theory as opposed to control theory. However, as Witsenhausen [W2]
has pointed out, the non-sequential ordering of decision variables in
extensive game theory is also perfectly appropriate in the context of
control theory. However, aside from Witsenhausen's causality
condition for well-posedness [W2], esentially nothing is known about
non-sequential stochastic control problems.

The FSFM model is related particularly closely to the Kuhn model
of an extensive game. According to Kuhn[1], an extensive game is <u>game</u>
<u>tree</u> with

(i)   a partition of the vertices with alternatives into the
<u>chance moves</u> P_0 and <u>player moves</u> P_1, ..., P_n

(ii)   a partition of the moves of P_i into <u>information sets</u>

(iii)   a probability distribution on the alternatives of the
information sets of P_0

(iv)   an n-tuple of real numbers for each terminal vertex.

An example of Kuhn-type extensive game is shown in Figure 2.3.1.
There is one chance move in P_0 with four alternatives. Each alternative
consists of the choice of an outcome of tossing two pennies. Thus
each outcome occurs with probability $\frac{1}{4}$. There are four moves in P_1,
and player one's information set is equal to P_1. Thus player one does
not know the outcome of the first chance move. He has to guess if the
pennies match or don't match. If he guesses correctly, he gets to keep

---

[1]See [K2] for a complete exposition.
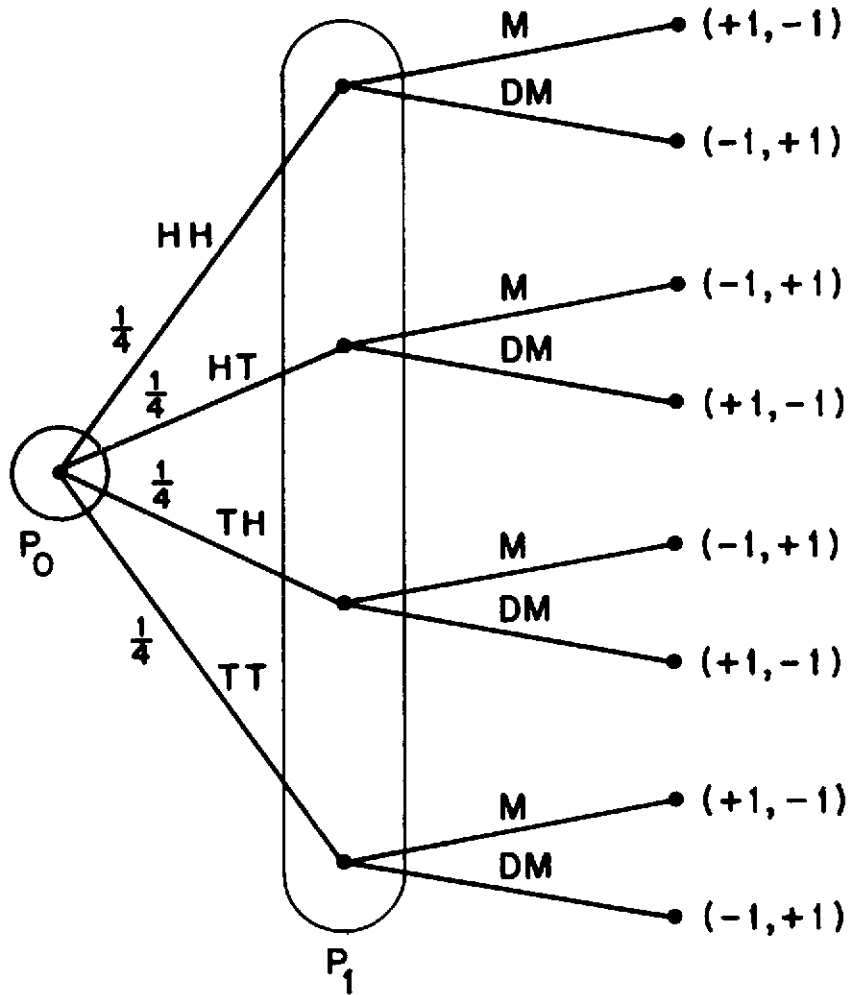
Figure 2.3.1 Matching Pennies

his own penny and player two's penny (the payoff is $(+1, -1)$). If

he guesses incorrectly, he loses his penny to player two (the payoff is

$(-1, +1)$).

Every FSFM problem can be reduced to a Kuhn extensive game. It might

be thought that the reduction is accomplished by identifying the player's

alternatives with the controller's inputs, but this is not always

possible. Suppose, for example, that $X_0 = \{1,2\}$, $U_1 = \{0,1\}$, and

$\Gamma_1 = \{\gamma_1, \tilde{\gamma}_1\}$, where $\gamma_1(1) = 1$, $\gamma_1(2) = 0$ and $\tilde{\gamma}_1 = 1 - \gamma_1$.[1] Clearly,

the game tree for this problem must have its first seven nodes as

in Figure 2.3.2, with vertices 1 and 2 in the set of moves of

player one (the only player). However, it is not possible to partition

P_1 into information sets so that the restriction that the same alternative

must be chosen for each vertex in a given information set is equivalent

to the restriction that the control law must lie in Γ_1. The point is

that restricting the control laws to lie in an arbitrary subset of

$U_t^{X_{t-1}}$ is a more general restriction than one based on information.

Thus, it is in general necessary to identify the player's alternatives

with the set of control laws. This is undesirable since the game does

not exhibit the information properties of the FSFM problem. However,

it will be shown in Chapter 3 that the first reduction (identifying

alternatives with controller inputs) is possible for FSFM problems with

simple information constraint.

---

[1] The choice of Γ_1 seems unnatural, but has appeared in the literature
[Stal]. The control laws in Γ_1 are the closed-loop control laws; those
in $U_1^{X_0} - \Gamma_1$ are the open-loop control laws.
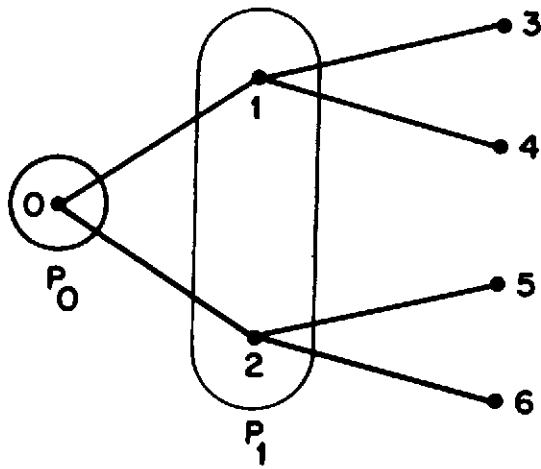
Figure 2.3.2   Game Tree for FSFM Problem

2.4 Example

In this section, an example of FSFM stochastic control problem is given. The problem is sufficiently simple that a solution can be written down by inspection. However it does illustrate the signaling strategy, a key phenomenon that occurs only in non-classical (as opposed to classical) stochastic control problems.

Figure 2.4.1 illustrates the problem considered. The initial state $x(0)$ is random, with $P(x(0)=1) = P(x(0)=2) = \frac{1}{2}$. The objective is to choose the controls $u(1)$, $u(2)$ so that $x(0) = x(2)$. If $x(0) \neq x(2)$, there is a penalty of 1 unit, and there is an additional penalty of $k \geq 0$ units if $x(1) = 3$. The control $u(1)$ at time 1 is allowed to depend on $x(0)$.

If the problem is to be a classical stochastic control problem, the control at t=2 must be allowed to depend on $x(0)$ and $u(1)$. In this case the solution is trivial. For t=1, always choose u=1. For t=2, choose u=1 if x=1 and u=0 if x=2. The resulting expected cost is EJ=0.

Suppose on the other hand that the control $u(2)$ is allowed to depend on the event $x(2)=3$ only. Then, if $k < 1$, an optimal strategy at t=1 is to choose $u(1)=1$ if $x(0)=1$ and $u(1)=0$ if $x(0)=2$. The corresponding optimal strategy for t=2 is to choose $u(2)=0$ if $x(2)=3$ and $u(2)=1$ if $x(2)=1$ or $x(2)=2$. The expected cost is $EJ=\frac{1}{2}k$.

The strategy employed in the choice of the first control for the second case is referred to as a signaling strategy. The interpretation of this statement is the following. If $x(0)=2$, the first controller moves the state to $x(1)=3$, which is undesirable for control purposes (there is a
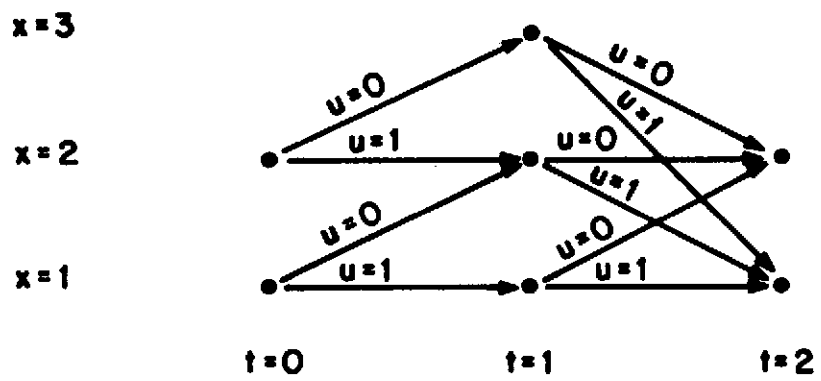
Figure 2.4.1   Example Illustrating Signaling Strategy

penalty k < 1). However, the second controller is able to see that x(1)=3, and so he unfailingly knows that the first state was x(0)=2. He can then avoid the penalty for being in the wrong terminal state.

The terminology signaling strategy arises in the theory of extensive games [T1]. If the present example is viewed as a (1-player) extensive game, it has a Kuhn game tree [K1,K2] as shown in Figure 2.4.2. Notice that the states have been eliminated, and only the sequence of decisions exhibited (the choice x_0=1 or x_0=2 is a decision due to nature). Note that the vertices of the game tree are partitioned into information sets. Thus in Figure 2.4.2b, the second decision must be made on the basis only of the knowledge that the event x(0) = 2 and u(1)=0 did or did not occur. Thus the control law must pick out the same alternative for each vertex within a given information set.

In terms of the game tree, the notion of a signaling strategy can be given a precise definition. Consider the information set U_2 in Figure 2.4.2b. The set of all vertices following the choice u=0 does not contain the set V_2, so according to Thompson's definition [T1], U_2 is a signaling information set, and any strategy (control law) defined on U_2 is a signaling strategy. In constrast, the set of all vertices following the choice u=0 for U_2 in Figure 2.4.2a contains V_3, and a similar statement holds for V_4. Thus U_2 in Figure 2.4.2a is not a signaling information set. This situation may be summed up succinctly as follows. In V_4 and V_3, the player (controller) remembers everything he knew in U_2 (Figure 2.4.2a). In V_2, the player has forgotten nature's choice and his own previous decision.
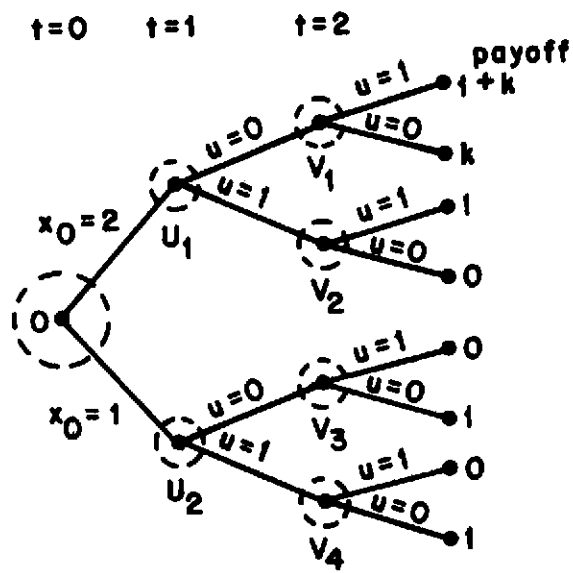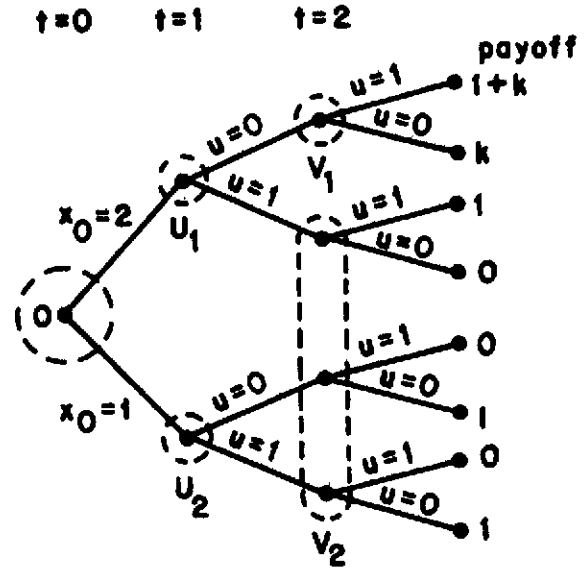
Figure 2.4.2   (a) Perfect State Observation (b) Imperfect State Observation

In the next chapter, a minimum principle is established for the FSFM stochastic control problem, and it is verified that the optimal strategies satisfy the minimum principle. The importance of the concept of the signaling strategy is that when there are no signaling strategies present, the minimum principle can be strengthened to give a sufficient condition.

2.5 The Equivalent Deterministic Problem

In this section, a deterministic optimal control problem equivalent to the FSFM problem is derived. The equivalent problem can be used to obtain necessary and sufficient conditions for the optimality of sequence γ* of control laws for the FSFM problem.

Since the FSFM stochastic control problem with simple information constraint is a special case of the general sequential stochastic control problem, it could be transformed to Witsenhausen's standard form and the general optimality conditions applied [N4]. However, the FSFM problem has a special structure that can be usefully exploited in the development of optimality conditions. These conditions are expressed in terms of the equivalent deterministic problem derived in this section. The deterministic problem for certain important special cases of the FSFM problem has a state space of fixed, finite dimension in contrast to the growing state space required in general. Moreover, the assumption of a simple information constraint is unnecessary.

It should not be inferred from the preceding remarks that the equivalent deterministic problem derived in this section has the most efficient state space for all stochastic control problems that can be cast in the FSFM format. In fact, for perfect memory problems, and for certain sequential hypothesis testing problems, more efficient equivalent deterministic problems can be derived utilizing the special structure of these problems.

The state space of the deterministic problem equivalent to the FSFM is the set of probability vectors on the original state set X_t. Since

X_t is finite, there is no loss in generality in assuming that $X_t = \{1,2,\ldots,n_t\}$.

Let Π_t be the set of probability (row) vectors in R^{n_t}, ie., $\sum_{i=1}^{n_t} \pi_i = 1$

and $\pi_i \geq 0$, $i = 1,2,\ldots,n_t$.

For $\gamma_t \in \Gamma_t$, let $h^{\gamma_t}(t)$ be the (column) vector in R^{n_t} with components

$h_t(i, \gamma_t(i))$, $i = 1,2,\ldots,n_t$. Similarly, let ϕ_T be the column vector

in R^{n_t} with components $\phi_T(i)$, $i=1,2,\ldots,n_T$.

Finally, for each $\gamma_t \in \Gamma_t$, define matrices $P^{\gamma_t}(t)$ with components

$$P_{ij}^{\gamma_t}(t) = P_t(\{q : j = f_t(i, \gamma_t(i), q)\}) \qquad (2.5.1)$$

where $i \in X_{t-1}$ and $j \in X_t$. Clearly, $P^{\gamma_t}(t)$ is a stochastic matrix

(its rows sum to one, and its elements are non-negative). Notice that the

matrices $P^{\gamma_t}(t)$, $\gamma_t \in \Gamma_t$, can be determined by the matrices $P^{u_t}(t)$,

$u_t \in U_t$, where $P^{u_t}(t)$ is the stochastic matrix with components

$$P_{ij}^{u_t} = P_t(\{q : j = f_t(i, u_t, q)\}) \qquad (2.5.2)$$

for $i \in X_{t-1}$, $j \in X_t$. If $\gamma_t(i) = u_t$, then row i of $P^{\gamma_t}(t)$ is equal to

row i of $P^{u_t}(t)$.

Let $\pi(0) = \pi_0$, and define $\pi(t)$ by the equations

$$\pi(t) = \pi(t-1) P^{\gamma_t}(t) \qquad (2.5.3)$$

for $t = 1,2,\ldots,T$. Clearly, $\pi(t)$ corresponds to the marginal probability

measure of PS_γ^{-1} on X_t. That is, $\pi_i(t)$ is the unconditional probability

that $x(t) = i$ when the control law sequence $\gamma = (\gamma_1, \ldots, \gamma_T)$ is used.

It follows immediately that

$$J(\gamma) = E_\gamma J_\gamma = \pi(T) \ \phi(T) + \sum_{t=1}^{T} \pi(t-1) \ h^{\gamma_t}(t). \qquad (2.5.4)$$

Therefore, the FSFM stochastic control problem is equivalent to the deterministic problem of minimizing (2.5.4) subject to (2.5.3).

Further insight into the nature of the equivalent deterministic problem (2.5.3), (2.5.4) can be obtained by considering randomized strategies. Attention is restricted to the class of behavioral strategies [K2]. This is a subclass of the general class of randomized strategies defined in Section 2.1.

A behavioral randomization is a set of non-negative numbers $\{\lambda_{\gamma_t}(t)\}$ satisfying

$$\sum_{\gamma_t \epsilon \Gamma_t} \lambda_{\gamma_t}(t) = 1 \qquad (2.5.5)$$

for $t = 1,2,\ldots,T$. In this case, the control law $\gamma_t \ \epsilon \ \Gamma_t$ is chosen with probability $\lambda_{\gamma_t}(t)$ independently of the choice of γ_τ, $t \neq \tau$. Notice that it is not possible to coordinate the choice of strategies over time (unless the strategy at every stage is pure[1]) so that behavioral randomization is not the most general randomization.

In terms of the behavioral strategy, the state equation (2.5.3) becomes

$$\pi(t) = \pi(t-1) \ (\sum_{\gamma_t \epsilon \Gamma_t} \lambda_{\gamma_t}(t) P^{\gamma_t}(t)), \quad t = 1,2,\ldots,T, \qquad (2.5.6)$$

where $\pi(0) = \pi_0$. The cost function is

---

[1]The behavioral strategy is pure if $\lambda_{\gamma_t}(t) = 1$ for some $\gamma_t \ \epsilon \ \Gamma_t$, $t = 1, 2, \ldots, T$.

$$J(\gamma) = \pi(T) \ \phi_T + \sum_{t=1}^{T} \pi(t-1) \left(\sum_{\gamma_t \epsilon \Gamma_t} \lambda_{\gamma_t}(t) \ h^{\gamma_t}(t) \right). \quad (2.5.7)$$

Equations (2.5.6) - (2.5.7) show that the FSFM problem is equivalent to a deterministic optimal control problem with bilinear state dynamics and bilinear cost functional. Moreover, since the optimal strategy is known to be pure (as pointed out in Section 2.1), the problem is known a priori to be "bang-bang". The fact that the FSFM problem is equivalent to a bilinear problem is intriguing since there has been a considerable amount of research devoted to these systems recently [Br1, Mo1, Wil]. However, this equivalence will not be exploited in the sequel.

In general, the FSFM model is an efficient representation of a given stochastic control problem when the state set of the FSFM problem is a fixed, finite set not too much larger than the original state set. This will generally be the case when the controller has a fixed, finite memory, the noise is independent from stage-to-stage, and the cost has a stage-wise additive structure. For problems of this type, the equivalent deterministic problem has a state space of fixed, finite dimension, in contrast to the growing state space required by the Witsenhausen standard form. This simplification is achieved by admitting slight complications into the structure of the deterministic problem. Thus the matrices corresponding to the $P^{\gamma_t}(t)$ are stochastic matrices with all elements either zero or one and only a terminal cost is required for the deterministic problem equivalent to the Witsenhausen standard form.

When the controller has perfect memory, its memory set expands and so must its state set. Thus the deterministic version of the corresponding

FSFM problem requires a growing state space. A more efficient equivalent deterministic problem is obtained by taking the conditional probability vector of the original state set given past observations as the deterministic state. This approach has been followed, implicitly or explicitly, by a number of authors [Aol, Asl, Sol, Sml, Sawl].

CHAPTER III

THE FSFM MINIMUM PRINCIPLE

In this chapter, a minimum principle is stated and derived. The minimum principle is a necessary condition for optimality, but is not sufficient in general. However, in the absence of signaling control laws, the minimum principle can be strengthened to obtain a sufficient condition.

A numerical optimization algorithm based on the minimum principle is developed. It is shown that the algorithm always converges to a person-by-person extremal.

3.1 Derivation of the FSFM Minimum Principle

In the previous section, it was shown that the FSFM stochastic control problem is equivalent to a deterministic optimal control problem with cost functional

$$J(\gamma) = \pi(T) \; \phi_T + \sum_{t=1}^{T} \pi(t-1) \; h^{\gamma_t}(t) \qquad (3.1.1)$$

and state equations

$$\pi(t) = \pi(t-1) \; P^{\gamma_t}(t) \quad , \quad t = 1, 2, \ldots, T \qquad (3.1.2)$$

where $\pi(0) = \pi_0$ is given. Notice that each Γ_t is a discrete set, so that the convexity assumption required for application of the discrete minimum principle [Hall, Holl] is not satisfied. Therefore, the proof presented in this section proceeds from first principles.

Since the state dynamics of the equivalent deterministic problem are linear in the state, it is useful to consider the adjoint system of (3.1.2). Let Φ be the set of (column) vectors $\phi \in R^n$. Then the product

$$\pi\phi = \sum_{x \in X} \pi_x \phi_x \qquad (3.1.3)$$

is defined in accord with the usual matrix-vector notation. Holding ϕ fixed, a linear functional on Π is defined, and conversely.

Define the forced adjoint or costate equation

$$\phi(t-1) = P^{\gamma_t}(t) \ \phi(t) + h^{\gamma_t}(t) \qquad (3.1.4)$$

for $t = 1, 2, \ldots, T$, where $\phi(T) = \phi_T$ is the terminal cost vector.

Lemma 3.1.1

Let $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_T)$ be fixed control law sequence. Let the corresponding state and costate sequences be defined by (3.1.2) and (3.1.4) where $\pi(0) = \pi_0$ and $\phi(T) = \phi_T$. Then,

$$\pi(t) \ \phi(t) = \sum_{\tau=t+1}^{T} \pi(\tau-1) \ h_\tau^{\gamma_\tau}(\tau) + \pi(T) \ \phi(T). \qquad (3.1.5)$$

Proof

The proof is by backward induction. Equation (3.1.5) is clearly valid for $t = T$. If it is valid for general t, then

$$\pi(t-1) \ \phi(t-1) = \pi(t-1) \ (P^{\gamma_t}(t) \ \phi(t) + h^{\gamma_t}(t))$$

$$= \pi(t) \ \phi(t) + \pi(t-1) \ h^{\gamma_t}(t) \qquad (3.1.6)$$

$$= \sum_{\tau=t}^{T} \pi(\tau-1) \ h^{\gamma_\tau}(\tau) + \pi(T) \ \phi(T)$$

so that (3.1.5) is valid for t-1. Therefore, the equation is valid for
t = T, T-1, ..., 1.

Theorem 3.1.2   (FSFM Minimum Principle)

If the sequence $\gamma^0 = (\gamma_1^0, \gamma_2^0, ..., \gamma_T^0)$ is optimal for the FSFM
stochastic control problem and $\pi^0(t)$, $\phi^0(t)$ are the associated state and
costate sequences satisfying

$$\pi^0(t) = \pi^0(t-1) \, P^{\gamma_t^0}(t), \quad \pi^0(0) = \pi_0 \qquad (3.1.7)$$

$$\phi^0(t-1) = P^{\gamma_t^0}(t) \, \phi^0(t) + h^{\gamma_t^0}(t), \quad \phi^0(T) = \phi_T \qquad (3.1.8)$$

then

$$\pi^0(t-1) \, P^{\gamma_t^0}(t) \, \phi^0(t) + \pi^0(t-1) \, h^{\gamma_t^0}(t)$$

$$\qquad (3.1.9)$$

$$\leqslant \pi^0(t-1) \, P^{\gamma_t}(t) \, \phi^0(t) + \pi^0(t-1) \, h^{\gamma_t}(t)$$

for all $\gamma_t \in \Gamma_t$, for all t = 1, 2, ..., T.

Proof

From Lemma 3.1.1 and equation (3.1.1),

$$J(\gamma_1^0, ..., \gamma_{t-1}^0, \gamma_t^0, \gamma_{t+1}^0, ..., \gamma_T^0)$$

$$= \sum_{\tau=1}^{t-1} \pi^0(\tau-1) \, h^{\gamma_\tau^0}(\tau) + \pi^0(t-1) \, \phi^0(t-1) \qquad (3.1.10)$$

$$= \sum_{\tau=1}^{t-1} \pi^0(\tau-1) \, h^{\gamma_\tau^0}(\tau) + \pi^0(t-1) \, P^{\gamma_t^0}(t) \, \phi^0(t) + \pi^0(t-1) \, h^{\gamma_t^0}(t).$$

Similarly,

$$J(\gamma_1^{\ 0}, \ \ldots, \ \gamma_{t-1}^{\ \ 0}, \ \gamma_t, \ \gamma_{t+1}^{\ \ 0}, \ \ldots, \ \gamma_T^{\ 0}) \qquad (3.1.11)$$

$$= \sum_{\tau=1}^{t-1} \pi^0(\tau-1) \ h^{\gamma_\tau^{\ 0}}(\tau) \ + \ \pi^0(t-1) \ p^{\gamma_t}(t) \ \phi^0(t) \ + \ \pi^0(t-1) \ h^{\gamma_t}(t).$$

Notice that $\pi(\tau)$ is independent of γ_t, $\tau < t$, and $\phi(\tau)$ is independent of γ_t, $\tau \geqslant t$.

Since $\gamma_1^{\ 0}$, $\gamma_2^{\ 0}$, $\ldots$, $\gamma_T^{\ 0}$ is optimal,

$$J(\gamma_1^{\ 0}, \ \ldots, \ \gamma_{t-1}^{\ \ 0}, \ \gamma_t^{\ 0}, \ \gamma_{t+1}^{\ \ 0}, \ \ldots, \ \gamma_T^{\ 0})$$

$$\qquad (3.1.12)$$

$$\leqslant J(\gamma_1^{\ 0}, \ \ldots, \ \gamma_{t-1}^{\ \ 0}, \ \gamma_t, \ \gamma_{t+1}^{\ \ 0}, \ \ldots, \ \gamma_T^{\ 0})$$

and (3.1.9) follows immediately.

Although the minimum principle is a necessary condition for optimality, it is not a general sufficient. This hardly is surprising, since only the condition (3.1.12) of the optimal control law sequence has been utilized. Other control law sequences than the optimal can satisfy (3.1.12). Such sequences are called extremal. Thus Theorem 3.1.2 has the key ingredients of a minimum principle. The Hamiltonian minimization is global since every $\gamma_t \in \Gamma_t$ must be tested. However the overall minimization of the cost functional is local, since the test is performed for a single, isolated time instant. This is completely analogous to the continuous time situation in which large variations in the control for infinitesimal time intervals (the "strong variations" of the calculus of variations) are used to derive the minimum principle [Pol, Al].

3.2 Examples

In this section, two examples illustrating the application of the minimum principle are given.

Example 1

This example shows that the minimum principle is not in general a sufficient condition for optimality. The example has two stages (T=2) and is defined as follows:

$$X_0 = X_1 = X_2 = \{1,2\}$$

$$U_1 = U_2 = \{0,1\}$$

$$Q_1 = \{1\}, \quad Q_2 = \{1,2,3\}$$

$$\pi_0 = [1 \ 0]$$

$$\phi_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$h^{\gamma_t}(t) \equiv 0$$

The sets of admissible control laws $\Gamma_1 = \Gamma_2$ have just two elements - the control law whose value is always 1 and the control law whose value is always 2. The probabilities of the elements of Q_2 are

$$p_2(1) = \frac{1}{2}, \quad p_2(2) = \frac{1}{4}, \quad p_2(3) = \frac{1}{4}.$$

The state transition functions $f_1 : X_0 \times U_1 \times Q_1 \rightarrow X_1$ and $f_2 : X_1 \times U_2 \times Q_2 \rightarrow X_2$ are defined by

$$f_1(1,0,1) = 1, \quad f_1(2,0,1) = 2, \quad f_1(1,1,1) = 2, \quad f_1(2,1,1) = 1$$

$$f_2(1,0,1) = 1, \quad f_2(1,0,2) = 2, \quad f_2(1,0,3) = 2$$

$$f_2(2,0,1) = 1, \quad f_2(2,0,2) = 1, \quad f_2(2,0,3) = 2$$

$$f_2(1,1,1) = 1, \quad f_2(1,1,2) = 1, \quad f_2(1,1,3) = 2$$

$$f_2(2,1,1) = 2, \quad f_2(2,1,2) = 2, \quad f_2(2,1,3) = 2$$

It is not hard to verify that the corresponding transition matrices

at t=1 are

$$P^0(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad P^1(1) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and at t=2 are

$$P^0(2) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix} \quad P^1(2) = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ 0 & 1 \end{bmatrix}.$$

Suppose that $\gamma_1^* \equiv 0$, and that $\gamma_2^* \equiv 0$. It is necessary to compute

$\pi^*(1)$ and $\phi^*(1)$ in order to apply the minimum principle. These are easily

found:

$$\pi^*(1) = [1 \quad 0]$$

$$\phi^*(1) = \begin{bmatrix} \frac{1}{2} \\ \frac{3}{4} \end{bmatrix}$$

The cost is $J(\gamma_1^*, \gamma_2^*) = \frac{1}{2}$. Note that

$$\pi^*(0) \ P^0(1) \ \phi^*(1) = \frac{1}{2} < \pi^*(0) \ P^1(1) \ \phi^*(1) = \frac{3}{4}$$

$$\pi^*(1) \ P^0(2) \ \phi^*(2) = \frac{1}{2} < \pi^*(1) \ P^1(2) \ \phi^*(2) = \frac{3}{4}$$

so that the necessary conditions of the minimum principle are satisfied.

However, if $\hat{\gamma}_1 \equiv 1$ and $\hat{\gamma}_2 \equiv 1$, then $J(\hat{\gamma}_1, \hat{\gamma}_2) = 0$ so that (γ_1^*, γ_2^*)

is not optimal.

Example 2

This example shows that the optimal control laws determined in

section 2.4 for the example considered there satisfy the minimum principle.

The problem as formulated in section 2.4 has state sets $X_0 = \{1,2\}$,

$X_1 = \{1,2,3\}$, and $X_2 = \{1,2\}$. The control sets are $U_1 = U_2 = \{0,1\}$, and

the uncertainty sets are $Q_1 = Q_2 = \{1\}$. The transition functions are

illustrated in Figure 2.4.1. The cost function is

$$J = h_2(x(1)) + g(x(0), x(2)) \qquad (3.2.1)$$

where

$$h_2(x(1)) = \begin{cases} k & x(1) = 3 \\ 0 & x(1) \neq 3 \end{cases}$$

$$g(x(0), x(2)) = \begin{cases} 0 & x(0) = x(2) \\ 1 & x(0) \neq x(2) \end{cases} .$$

Since the cost function does not have the stagewise additive form

(2.1.2), it is necessary to augment the state to put the problem into

the FSFM formulation. The idea is to carry along $x(0)$ in the state

equations so that the term $g(x(0), x(2))$ can be written in terms of the

terminal value of the augmented state.

When new state sets $X_0 = \{1,2\}$, $X_1 = \{1,2,3,4\}$, $X_2 = \{1,2,3,4\}$ are

defined, the state transition diagram of Figure 3.2.1 results. Clearly,

u = 0

x(1) = 3

u = 1

x(0) = 2
x(2) = 2

new state 4

u = 0

u = 0

u = 0

new state 2

x(0) = 2

u = 1

u = 1

x(0) = 2
x(1) = 2

x(0) = 2
x(2) = 1

new state 3

new state 1

u = 0

x(0) = 1

u = 1

u = 0

x(0) = 1
x(1) = 2

u = 1

x(0) = 1
x(2) = 2

new state 2

u = 0

u = 1

x(1) = 1

x(0) = 1
x(2) = 1

new state 1
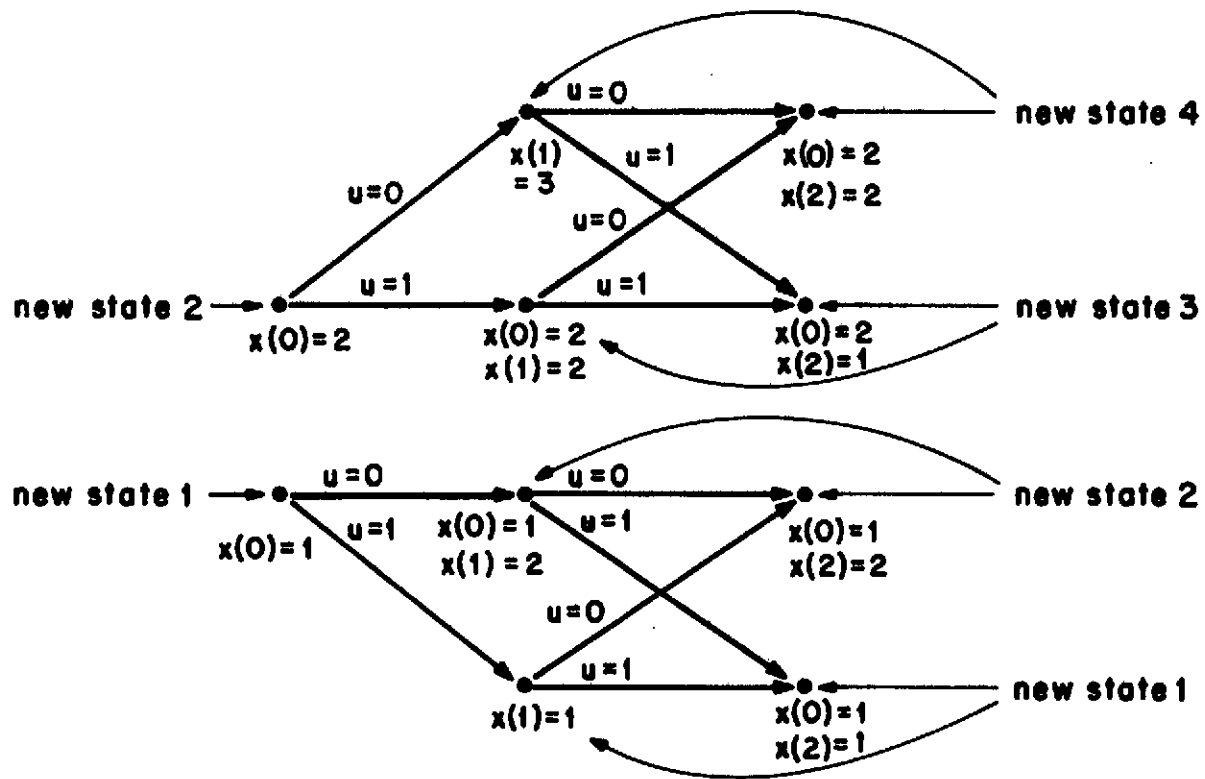
Figure 3.2.1   State Transition Diagram

the cost is

$$J = h_2(x(1)) + \phi_2(x(2)) \tag{3.2.2}$$

where

$$h_2(x(1)) = \begin{cases} k & x(1) = 4 \\ 0 & x(1) \neq 4 \end{cases}$$

$$\phi_2(x(2)) = \begin{cases} 1 & x(2) = 2,3 \\ 0 & x(2) = 1,4 \end{cases}.$$

The equivalent deterministic problem can be written down by inspection of Figure 3.2.1:

$$h(1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad h(2) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ k \end{bmatrix}, \qquad \phi(2) = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$P^1(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P^0(0) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P^1(1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P^0(1) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(As pointed out in section 2.5, the matrices $P^{\gamma_t}(t)$ can be found if the matrices $P^{u_t}(t)$ are known.)

For the case $\Gamma_t = U_t^{x_{t-1}}$, the optimal control laws found in section 2.4 are $\gamma_1^*(1) = 1$, $\gamma_1^*(2) = 1$, and $\gamma_2^*(1) = 1$, $\gamma_2^*(2) = 0$, $\gamma_2^*(3) = 0$, $\gamma_2^*(4) = 0$. The corresponding $\pi^*(1)$ and $\phi^*(1)$ are

$$\pi^*(1) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \ ,$$

$$\phi^*(1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ k \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ k \end{bmatrix} \ .$$

Therefore,

$$\pi^*(0) \ P^{\gamma_1^*}(1) \ \phi^*(1) = 0 \ ,$$

$$\pi^*(1) \ P^{\gamma_2^*}(2) \ \phi^*(2) = 0 \ .$$

Since all numbers in the problem are non-negative, $\gamma^* = (\gamma_1^*, \gamma_2^*)$ clearly satisfies the conditions of the minimum principle.

For the case $\Gamma_1 = U_1^{x_0}$, $\Gamma_2 = \{\gamma_2 \ \epsilon \ U_2^{x_1} : \gamma_2(1) = \gamma_2(2) = \gamma_2(3)\}$, the optimal control laws are $\gamma_1^*(1) = 1$, $\gamma_1^*(2) = 0$, and $\gamma_2^*(1) = \gamma_2^*(2) = \gamma_2^*(3) = 1, \gamma_2^*(4) = 0$. The corresponding $\pi^*(1)$ and $\phi^*(1)$ are

$$\pi^*(1) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

$$\phi^*(1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ k \end{bmatrix}.$$

Therefore,

$$\pi^*(0) \; P^{\gamma_1^*}(1) \; \phi^*(1) = \frac{1}{2} k$$

$$\pi^*(1) \; P^{\gamma_2^*}(2) \; \phi^*(2) = 0$$

There are three other possible control laws at t=1, and at t=2. These give $\pi^*(0) \; P^{\gamma_1}(1) \; \phi^*(1) = \frac{1}{2} k, \frac{1}{2}, \frac{1}{2}$ and $\pi^*(1) \; P^{\gamma_2}(2) \; \phi^*(2) = 0, \frac{1}{2}, \frac{1}{2}.$ Therefore, the minimum principle is satisfied for $k \leqslant 1$.

Note, however, that the control law sequence $\bar{\gamma} = (\bar{\gamma}_1, \bar{\gamma}_2)$, where $\bar{\gamma}_1 \equiv 1$, $\bar{\gamma}_2 \equiv 1$ also satisfies the minimum principle. Since the control law sequence $\bar{\gamma}$ has $J(\bar{\gamma}_1, \bar{\gamma}_2) = \frac{1}{2} > J(\gamma_1^*, \gamma_2^*) = \frac{1}{2} k$ (for $k < 1$), $\bar{\gamma}$ is not optimal. This is a good illustration of the fact that satisfaction of the minimum principle assures only that the control law sequence can not be improved by changing the control law at a single stage. The optimal strategy γ^* is a signaling strategy so that coordination is required: it is no use to employ the signaling control law γ_1^* unless the second stage control law utilizes the information. Conversely, a second stage control law that attempts to utilize signaling information that is not forthcoming is worthless. The need to consider signaling strategies is the fundamental reason why the study of non-classical stochastic control is much more difficult than the study of classical stochastic control.

3.3 Signaling and Sufficiency

The novelty of non-classical stochastic control is the presence of signaling strategies. To explore the implications of this fact, it is necessary to restrict attention to a certain subclass of FSFM problems.

Definition 3.3.1

The FSFM problem (2.1.1)-(2.1.2) is said to have a simple information constraint if

$$\Gamma_t = \{\gamma_t \in U_t^{X_{t-1}} : \gamma_t^{-1}(U_t) \subset F_{t-1}\} \qquad (3.3.1)$$

for $t = 1,2,...,T$, where $U_t = P(U_t)$ and F_{t-1} is a subfield of $X_{t-1} = P(X_{t-1})$.

The reason for restricting attention to FSFM problems with simple information constraints is that these problems can be readily identified with a corresponding Kuhn model of an extensive game (see section 2.3 and reference [K2]).

Suppose that a FSFM problem with simple information constraint is given. Let the sets X_0, Q_1, U_1, Q_2, ..., U_T have n_0, n_1, m_1, n_2, ..., m_T elements, respectively. The rank 0 move[1] of the corresponding game tree has n_0 alternatives. For $1 \leq t \leq T$, the rank 2t-1 move has n_t alternatives and the rank 2t move has m_t alternatives. Thus every play has rank 2T + 1 (Figure 3.3.1).

---

[1] A move is a vertex of the game tree with alternatives; a play is a (terminal) vertex without alternatives. The rank of a move or play is the number of moves that preceed it. See Kuhn [K2] for details.

choice
of x(0)

choice
of q(1)

choice
of u(1)

choice
of q(2)

choice
of u(T)

move

n_0
alternatives

n_1
alternatives

m_1
alternatives

play

m_T
alternatives
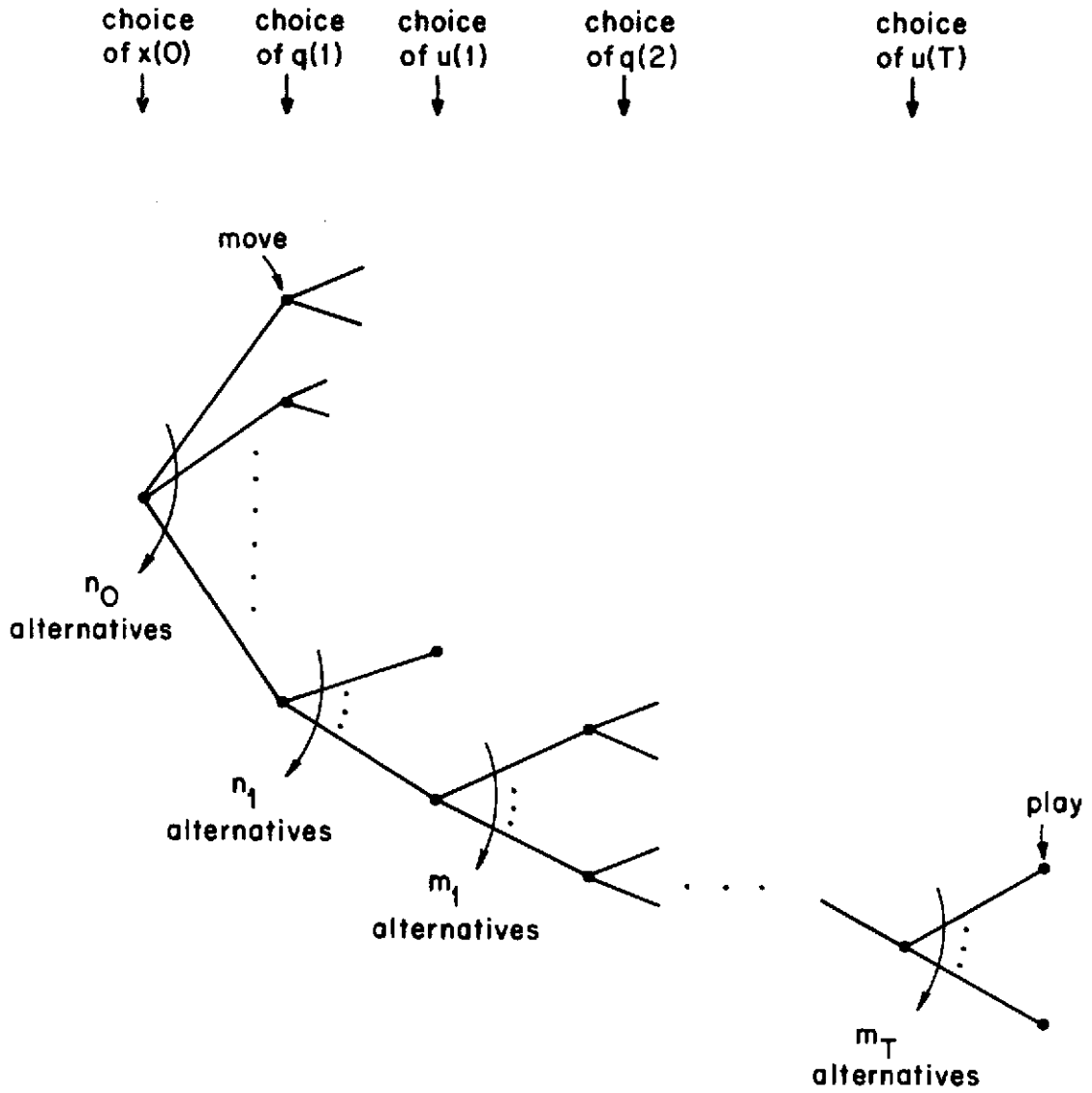
Figure 3.3.1   Game Tree for FSFM Problem With
Simple Information Constraint

The chance moves P_0 are the moves with rank 0, 1, 3, ...,

2T-1, and the moves P_1 of player 1 (the only player) are the moves

with rank 2, 4, ..., 2T. Each alternative of the initial (rank 0)

move of the game tree corresponds to an element of X_0. Similarly, the

alternatives of moves with rank 2t-1 correspond to elements of Q_t, and

moves with rank 2t correspond to elements of U_t.

Each information subset of P_0 contains a single point of P_0. The

information sets of P_1 are defined by the atoms[1] of F_t as follows. Notice

that the system equations (2.1.1) define a map

$$S_t : X_0 \times Q_1 \times U_1 \times \ldots \times Q_t \times U_t \rightarrow X_t \qquad (3.3.2)$$

which takes an initial state and a sequence of inputs and gives

corresponding state. Each atom F of F_t defines a set

$$\{(x(0), q(1), u(1), \ldots, q(t), u(t)) : S_t (x(0), q(1), u(1), \ldots,$$
$$q(t), u(t)) \epsilon F\} \subset X_0 \times Q_1 \times U_1 \times \ldots \times Q_t \times U_t. \qquad (3.3.3)$$

Since there is a one-to-one correspondence between the set $X_0 \times Q_1 \times U_1 \times$

$\ldots \times Q_t \times U_t$ and the moves of order 2t + 1 of the game, the partition

induced on $X_0 \times Q_1 \times U_1 \times \ldots \times Q_t \times U_t$ by the atoms of F_t induces a

partition on the corresponding set of moves. Thus each atom F ϵ F_t gives

rise to a single information set for player one containing moves of player

1. As a consequence, all the moves of given information set are of the

---

[1]An <u>atom</u> of a field F is a set F ϵ F such that if E ϵ F and E $\subset$ F, then
either E = ϕ or E = F. The atoms of a finite field always exist and form
a partition [Hal].

same rank. This is not surprising, since the problem is sequential [W2].

To finish the specification of the game, the probabilities of the

chance moves must be defined and the terminal cost specified. If an

information set of P_0 contains a move of rank 2t-1, its alternative

corresponding to q ϵ Q_t is chosen with probability $p_t(q)$. The terminal

cost is determined by the fact that the plays are in one-to-one

correspondence with X_0 x Q_1 x U_1 x ... x Q_T x U_T. Thus each play determines

a complete state-control trajectory for which J can be evaluated. This

value of J is the cost associated with the play.

In game theory, a strategy for player 1 is the assignment of a single

alternative to each information set. For FSFM problems with simple

information constraint, a control law is the assignment of a point in

U_t to each atom of F_{t-1} (since γ_t is constrained to be F_{t-1} measurable).

Because of the manner in which the information sets have been constructed

above, there is clearly a one-to-one correspondence between the control

laws of a FSFM problem with simple information constraint and its

corresponding extensive game form. Thus the same notation γ will be

used to describe either a control law sequence or a strategy for the

equivalent extensive game.

The equivalence between the extensive game and FSFM forms of a problem

is best understood by example. Figure 3.3.2 illustrates the extensive

game form of the FSFM problem considered in the previous section when

$$F_1 = \{\phi, \{1\}, \{2\}, X.\}$$
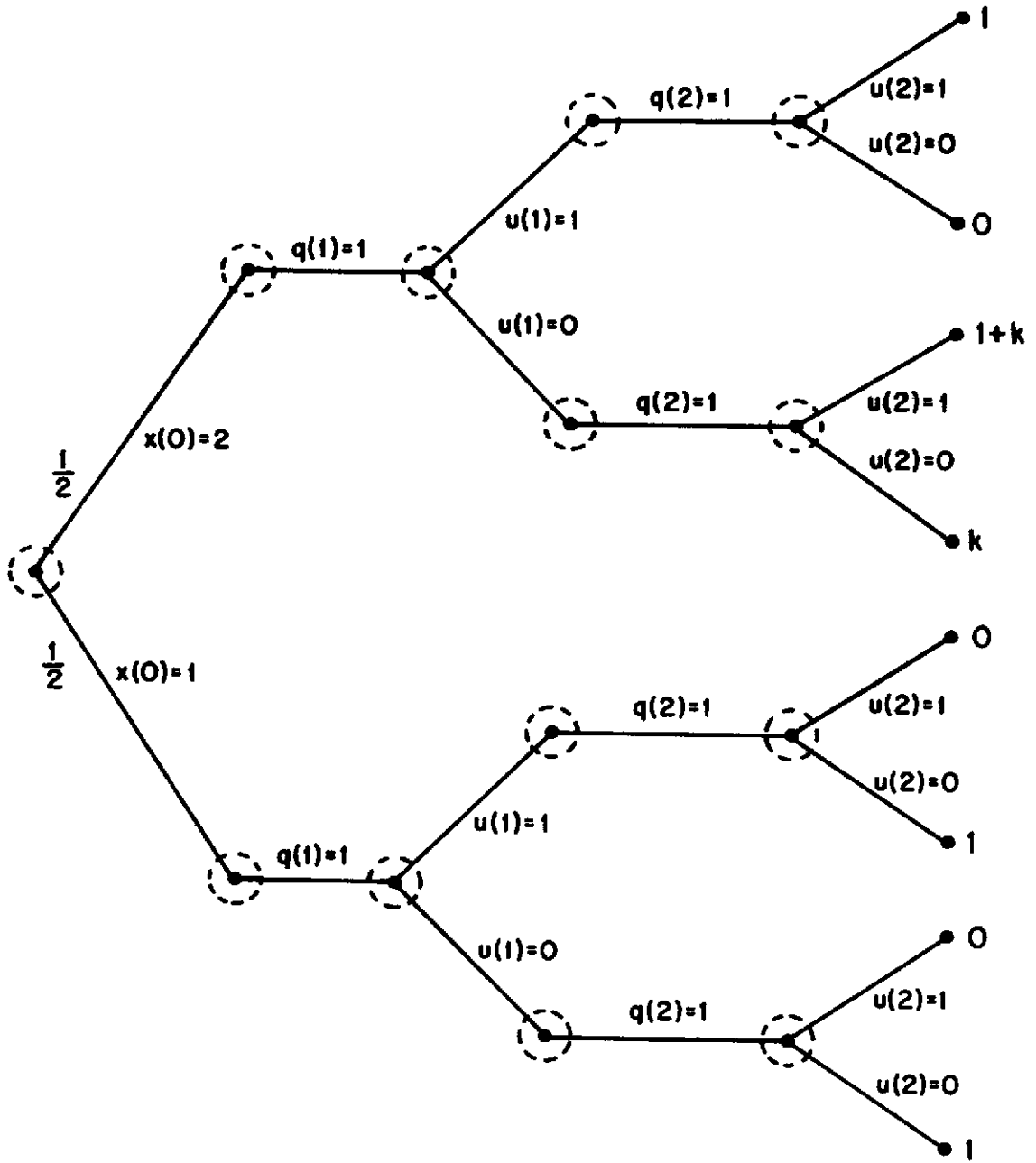
Figure 3.3.2   Extensive Game Form of FSFM Problem
With Full State Information

and $F_2 = \{\phi, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}$,

$\{2,4\}, \{3,4\}, \{1,2,3\}, \{2,3,4\}, \{1,3,4\}, \{1,2,4\}$, $X_1\}$

(full state information). Figure 3.3.3 illustrates the extensive game form when

$$F_1 = \{\phi, \{1\}, \{2\}, X_0\}, \quad F_2 = \{\phi, \{4\}, \{1,2,3\}, X_1\}.$$

The equivalence between FSFM problems and extensive games can be extended to FSFM models with information constraint.

Definition 3.3.2

The FSFM problem (2.1.1)-(2.1.2) is said to have an <u>information constraint</u> if

$$U_t = U_t^{\ 1} \times U_t^{\ 2} \times \ldots \times U_t^{\ m} \tag{3.3.4}$$

$$\Gamma_t = \Gamma_t^{\ 1} \times \Gamma_t^{\ 2} \times \ldots \times \Gamma_t^{\ m} \tag{3.3.5}$$

for $t = 1, 2, \ldots, T$, where

$$\Gamma_t^{\ i} = \{\gamma_t^{\ i} \in U_t^{\ X_{t-1}} : (\gamma_t^{\ i})^{-1}(U_t^{\ i}) \subset F_{t-1}^{\ i}\} \tag{3.3.6}$$

where $U_t^{\ i} = P(U_t^{\ i})$ and $F_{t-1}^{\ i}$ is a subfield of $X_{t-1} = P(X_{t-1})$.

Since the equivalence will not be used in the sequel, the construction of a Kuhn extensive game model equivalent to the FSFM model with information constraint will be omitted.

Since an equivalence has been established between FSFM models with simple information constraint and Kuhn extensive game models, the notions
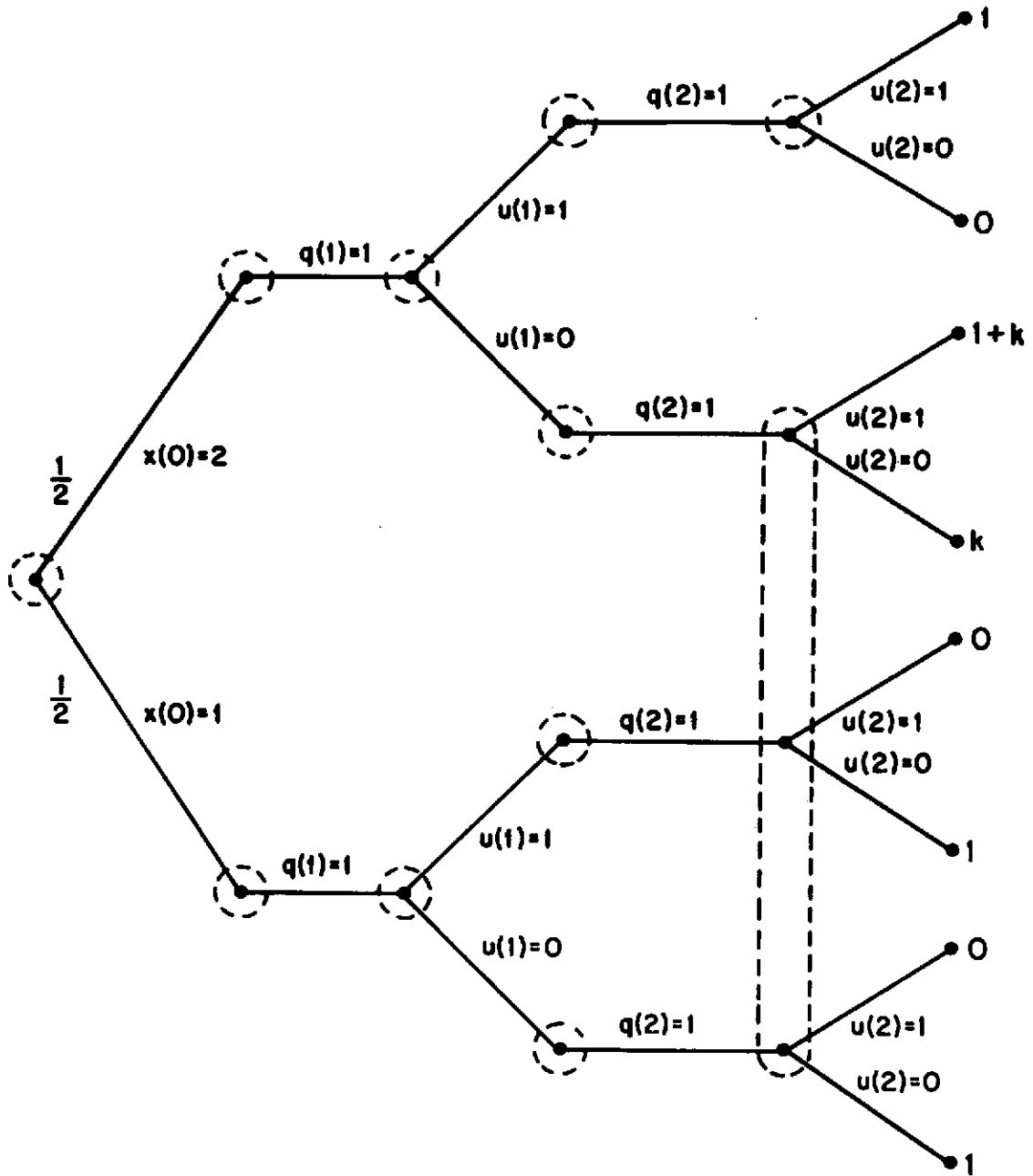
Figure 3.3.3   Extensive Game Form of FSFM Problem
With Partial State Information

of signaling strategy and perfect recall can now be precisely defined.

The following definitions and propositions are stated for 1-player games,

but can be easily extended to n-person games.

Definition 3.3.3 [K2]

A move Z of player 1 (n=1) is called <u>possible</u> when playing γ if it has

non-zero probability of occurring when the strategy γ is used. An

information set I for player 1 is called <u>relevant</u> when playing γ if

some $Z \in I$ is possible when playing γ.

Proposition 3.3.1

A move Z for player 1 is possible when playing γ if and only if γ

chooses all alternatives on the path W_Z from the origin to Z which are

incident at moves of player 1.[1]

Proof

See reference [K2], page 201.

Definition 3.3.4 [K2]

A game G is said to have <u>perfect recall</u> if I is relevant when playing

γ and $Z \in I$ implies that Z is possible when playing γ for all I, Z

and γ.

Definition 3.3.5 [T1]

Let I be an information set for player 1, and let I_u = {moves following

some move in I by alternative u}. Then I is a <u>signaling information set</u>

---

[1] All chance moves are assumed to occur with non-zero probability.

for player 1 if, for some u and some information set J of player 1,

$I_u \cap J \neq \phi$ and $J \not\subset I_u$.

Proposition 3.3.2 [T1]

A game G has perfect recall if and only if player 1 has no signaling information sets.

Proof

See reference [T1], page 268.

The following proposition is <u>not</u> valid for general games, but is a special property of 1-person (stochastic control) problems.

Proposition 3.3.3

Let G be a 1-person game with perfect recall, and let I be an arbitrary information set of the player. If I is not relevant when playing γ, then the probability of any move in I is zero under γ. If I is relevant when playing γ, then the probability of any move in I is positive under γ. Moreover, if I is relevant under any other strategy $\tilde{\gamma}$, then the probabilities of any move of I under γ and $\tilde{\gamma}$ are the same.

Proof

If I is not relevant when playing γ, then by definition no move of I is possible when playing γ. Thus the probability of any such move is zero when γ is used.

If I is relevant when playing γ, then every move of I is possible when playing γ since G has perfect recall. Thus the probability of any such move is positive when γ is used.

If $Z \in I$ is possible when playing γ, by Proposition 3.3.1 γ must choose all alternatives on the path W_Z from the origin to Z which are incident at moves of player 1. All other alternatives on W_Z are incident at chance moves, and the probability of Z under γ is simply the product of the probabilities of these alternatives. But this probability is the same for $\tilde{\gamma}$, since $\tilde{\gamma}$ likewise chooses all alternatives on the path W_Z incident at moves of player 1.

At this point, the preceeding definitions and propositions are applied to the FSFM problem.

Definition 3.3.6

A FSFM stochastic control problem is said to have <u>perfect recall</u> if it has a simple information constraint and the corresponding extensive game has perfect recall.

Definition 3.3.7

A control law γ_t for a FSFM problem with simple information constraint is said to be a <u>signaling control law</u> if an atom of F_{t-1} gives rise to a signaling information set in the corresponding extensive game.

Corollary 3.3.4

A FSFM stochastic control problem with simple information constraint has perfect recall if and only if it has no signaling control laws.

Proof

This is a direct consequence of the definitions, the construction of the equivalent extensive game, and Proposition 3.3.2.

Theorem 3.3.5

Suppose that a FSFM stochastic control problem with perfect recall is given. Let A be an atom of F_{t-1}. Then, for any control sequence, either the probability of all states in A is zero, or the probability of each state is a positive constant independent of γ.

Proof

By construction, the probability of a state $x(t-1) \in A$ under γ is equal to the probability of the corresponding set of moves in the information set I generated by A. Therefore, the theorem follows immediately from Proposition 3.3.3.

The property of FSFM problems with perfect recall expressed by Theorem 3.3.5 makes it possible to strengthen the minimum principle to achieve a sufficient condition for optimality.

Definition 3.3.8

Let the set of state probability vectors reachable at time t, $1 \leqslant t \leqslant T$, when the initial state probability vector is π_0 be denoted

$$r_t(\pi_0) = \{\pi_0 \; P^{\gamma_1}(1) \; P^{\gamma_2}(2) \; \ldots \; P^{\gamma_t}(t) \; : \; \gamma_1 \in \Gamma_1, \; \gamma_2 \in \Gamma_2,$$
$$\ldots, \; \gamma_t \in \Gamma_t\}. \tag{3.3.7}$$

$r_t(\pi_0)$ is called the <u>reachable set</u> $(r_0(\pi_0) = \{\pi_0\})$.

Definition 3.3.9

Suppose that the control law sequence $\gamma^* = (\gamma_1^*, \; \gamma_2^*, \ldots, \; \gamma_T^*)$ satisfies the condition

$$\pi(t-1) \ P^{\gamma_t^*}(t) \ \phi^*(t) + \pi(t-1) \ h^{\gamma_t^*}(t)$$

$$\leqslant \pi(t-1) \ P^{\gamma_t}(t) \ \phi^*(t) + \pi(t-1) \ h^{\gamma_t}(t) \tag{3.3.8}$$

for all $\gamma_t \ \epsilon \ \Gamma_t$, for all $\pi(t-1) \ \epsilon \ r_{t-1} \ (\pi_0)$ where

$$\phi^*(t-1) = P^{\gamma_t^*}(t) \ \phi^*(t) + h^{\gamma_t^*}(t) \tag{3.3.9}$$

for $t = 1,2,\ldots,T$ $(\phi^*(T) = \phi_T)$. Then γ^* is said to be <u>universally</u> <u>extremal</u>.

Lemma 3.3.6

Any universally extremal control law sequence is optimal.

Proof

The proof proceeds by induction on the number of stages T.

Suppose $T = 1$. Then

$$J(\gamma_1) = \pi(0) \ h^{\gamma_1}(1) + \pi(1) \ \phi(1)$$
$$= \pi(0) \ h^{\gamma_1}(1) + \pi(0) \ P^{\gamma_1}(1) \ \phi(1) \tag{3.3.10}$$

so that any extremal is optimal.

Suppose the lemma is valid for problems with $T-1$ stages. It must be established that the lemma is valid for problems with T stages.

Assume that $(\gamma_1^*, \ \gamma_2^*, \ \ldots, \ \gamma_T^*)$ is universally extremal. It follows immediately that $(\gamma_2^*, \ \gamma_3^*, \ \ldots, \ \gamma_T^*)$ is universally extremal for the problem with cost

$$J(\gamma_2, \ \ldots, \ \gamma_T; \ \pi(1)) = \sum_{t=2}^{T} \pi(t-1) \ h^{\gamma_t}(t) + \pi(T) \ \phi(T) \tag{3.3.11}$$

for any $\pi(1) \in r_1(\pi_0)$. Therefore, by the induction hypothesis,

$$J(\gamma_2^*, \ldots, \gamma_T^*; \pi(1)) \leqslant J(\gamma_2, \ldots, \gamma_T; \pi(1)) \qquad (3.3.12)$$

for all $\pi(1) \in r_1(\pi_0)$ and for all $\gamma_2 \in \Gamma_2, \ldots, \gamma_T \in \Gamma_T$. Moreover, since

$$J(\gamma_1, \gamma_2, \ldots, \gamma_T) = \pi(0) \, h^{\gamma_1}(1) + J(\gamma_2, \ldots, \gamma_T; \pi(0) \, P^{\gamma_1}(1)) \qquad (3.3.13)$$

it follows that

$$J(\gamma_1, \gamma_2^*, \ldots, \gamma_T^*) \leqslant J(\gamma_1, \gamma_2, \ldots, \gamma_T) \qquad (3.3.14)$$

for all $\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \ldots, \gamma_T \in \Gamma_T$.

But the assumption that $(\gamma_1^*, \gamma_2^*, \ldots, \gamma_T^*)$ is universally extremal implies that

$$J(\gamma_1^*, \gamma_2^*, \ldots, \gamma_T^*) = \pi(0) \, h^{\gamma_1^*}(1) + \pi(0) \, P^{\gamma_1^*}(1) \, \phi^*(1)$$

$$\leqslant \pi(0) \, h^{\gamma_1}(1) + \pi(0) \, P^{\gamma_1}(1) \, \phi^*(1) = J(\gamma_1, \gamma_2^*, \ldots, \gamma_T^*)$$

$$(3.3.15)$$

for all $\gamma_1 \in \Gamma_1$. The lemma follows from (3.3.15) and (3.3.14).

Notice from the proof of Lemma 3.3.6 that the existence of a universally extremal control law sequence γ^* implies the unusual fact that the problems

$$\min_{\gamma_t \in \Gamma_t, \ldots, \gamma_T \in \Gamma_T} J(\gamma_1, \ldots, \gamma_{t-1}, \gamma_t, \ldots, \gamma_T) \qquad (3.3.16)$$

for $\gamma_1 \in \Gamma_1, \ldots, \gamma_{t-1} \in \Gamma_{t-1}$ have a common solution $(\gamma_t^*, \ldots, \gamma_T^*)$.

Thus the existence of a universal extremal would seem to be rather unlikely.

From this viewpoint, the following property of FSFM problems with perfect recall seems rather remarkable.

Theorem 3.3.7

Every FSFM problem with perfect recall has a universally extremal control law sequence.

Proof

The proof is constructive. The control laws γ_t are defined by choosing their values on the atoms of F_{t-1}.

Consider the case for t=T. Let A_{T-1}^i be an atom of F_{T-1}, i = 1,2,..., P. For simplicity of notation, suppose that A_{T-1}^1 contains the first ℓ_1 states of X_{T-1}, A_{T-1}^2 contains states $\ell_1 + 1$ through ℓ_2 of X_{T-1}, etc. Notice that

$$\pi(T-1) \; P^{\gamma_T}(T) \; \phi(T) \; + \; \pi(T-1) \; h^{\gamma_T}(T) \tag{3.3.17}$$

$$= \sum_{i=1}^{p} \; \sum_{j=\ell_{i-1}+1}^{\ell_i} \; \pi_j(T-1) \left[\sum_{k=1}^{n} P_{jk}^{u_p(T)}(T) \; \phi_k(T) + h_j^{u_p(T)}(T) \right]$$

where n is the number of states in X_{T-1}, $\ell_0 = 0$, and $u_p(T)$ is the value of γ_T on the pth atom of F_{T-1}.

The decomposition (3.3.17) makes the construction of γ_T^* clear. By Proposition 3.3.5, every vector $\pi(T-1) \in r_{T-1}(\pi_0)$ either has $\pi_i(T-1)=0$, i = $\ell_i + 1$, ..., ℓ_{i+1}, or has $\pi_i(T-1) = \hat{\pi}_i(T-1)$, i = $\ell_i + 1$, ..., ℓ_{i+1}, where each $\hat{\pi}_i(T-1)$ is a fixed number. Therefore, γ_T^* takes the value $u_p^*(T)$ on the pth atom of F_{T-1}, where

$$\min_{u \in U_T} \sum_{j=\ell_{i-1}+1}^{\ell_i} \hat{\pi}_j (T-1) \left[\sum_{k=1}^{n} P_{jk}^{u}(T) \, \phi_K(T) + h_j^{u}(T) \right]$$

$$= \sum_{j=\ell_{i-1}+1}^{\ell_i} \hat{\pi}_j (T-1) \left[\sum_{k=1}^{n} P_{jk}^{u_p^*(T)}(T) \, \phi_K(T) + h_j^{u_p^*(T)}(T) \right] .$$

$$(3.3.18)$$

The construction of the remaining γ_t^* is completed by applying an analogous procedure to

$$\pi(t-1) \, P^{\gamma_t}(t) \, \phi^*(t) + \pi(t-1) \, h^{\gamma_t}(t). \qquad (3.3.19)$$

Theorem 3.3.7 is primarily of theoretical and conceptual importance. Problems with perfect recall are more efficiently handled by deriving an equivalent deterministic problem that has a conditional probability vector for the deterministic state. (The conditioning is with respect to the field F_{t-1}. Special cases of this procedure are implicit in the usual stochastic dynamic programming algorithm [Ao1, St1, As1] and the algorithm of Sandell and Athans for the 1-step delay problem [S1].

3.4 Finite-Set Team Problem

A minimum principle was derived in section 3.1, and its properties considered in the following sections. It is interesting to see if feasible numerical algorithms based on the minimum principle can be developed. A start in this direction in contained in the present section, where the numerical solution of the finite set team problem is considered.

In the sequel, the convenient notation

$$\int_X f(x)\ d\pi(x)\ =\ \sum_{x \in X} f(x)\ \pi(x)$$

will be used.

Let X, U_1, U_2, $\ldots$, U_p be finite sets with n, m_1, m_2, $\ldots$, m_p elements, respectively.

Let π be a probability measure on X, and $h : X \times U_1 \times U_2 \times \ldots \times U_p \to R$ a given real-valued function.

The finite set team problem[1] is:

$$\min_{\gamma_1\ \in\ \Gamma_1}\ \min_{\gamma_2\ \in\ \Gamma_2}\ \ldots\ \min_{\gamma_p\ \in\ \Gamma_p}\ \int h(x,\ \gamma_1(x),\ \gamma_2(x),\ \ldots,\ \gamma_n(x))\ d\pi(x)$$

$$(3.4.1)$$

where $\Gamma_1 \subset U_1{}^X$, $\Gamma_2 \subset U_2{}^X$, $\ldots$, $\Gamma_p \subset U_p{}^X$.

Note that for given γ_1, γ_2, $\ldots$, γ_p, the integral above can be computed

---

[1] When $\Gamma_i = \{\gamma_i \in U_i{}^X : \gamma_i{}^{-1}(U) \subset F_i\}$, where $U = P(U)$, and F_i is a subfield of $X = P(X)$, for $i = 1, \ldots, n$, then the finite set team problem is a special case of the more general formulation of Marshak and Radner [M2]. Of course, in this case the finite set team problem is a FSFM problem with information constraint.

with n-1 multiplications and n-1 additions. Therefore the finite set team problem can be solved with at most

$$(n-1) \cdot m_1^n \cdot m_2^n \ldots m_p^n \text{ multiplications,}$$

$$(n-1) \cdot m_1^n \cdot m_2^n \ldots m_p^n \text{ additions,}$$

$$\text{and } m_1^n \cdot m_2^n \ldots m_p^n \text{ comparisons,}$$

since Γ_i has at most m_i^n elements. Of course, there is additional overhead required to compute $h(x, \gamma_1(x), \gamma_2(x), \ldots, \gamma_p(x))$ for given $x, \gamma_1, \gamma_2, \ldots, \gamma_p$. However, this is ignored in the following discussion.

For certain special cases, one can do better.

Case 1   Perfect state observation.

In this case,

$$\Gamma_1 = U_1^X, \quad \Gamma_2 = U_2^X, \quad \ldots, \quad \Gamma_p = U_p^X.$$

It is easy to see that the problem is solved by computing

$$\min_{u_1 \in U_1} \min_{u_2 \in U_2} \ldots \min_{u_p \in U_p} h(x, u_1, u_2, \ldots, u_p) \quad (3.4.2)$$

for each $x \in X$. Therefore, the problem is solved with

No multiplications,

No additions,

m sets of $m_1 \cdot m_2 \ldots m_p$ comparisons.

This is a considerable saving. However, the problem is of limited interest since the stochastic aspect of the problem is trivial. Notice

that the probability measure π does not affect the solution at all.

Case 2 Common measurement.

In this case, the admissible control laws are measurable with respect to the field F determined by a given finite partition $\{A_1, A_2, \ldots A_k\}$ of X. This means that each control law is constant on each atom $A_i \subset X$ of the field F.

Notice that

$$\int_X h(x, \gamma_1(x), \gamma_2(x), \ldots, \gamma_p(x)) \ d\pi(x) \tag{3.4.3}$$

$$= \sum_{i=1}^{k} \int_{A_i} h(x, \gamma_1(x), \gamma_2(x), \ldots, \gamma_p(x) \ d\pi(x).$$

Since each γ_i is constant on P_i the problem can be solved by k minimizations of the form

$$\min_{u_1} \min_{u_2} \ldots \min_{u_p} \int_{A_i} h(x, u_1, u_2, \ldots, u_p) \ d\pi(x) \tag{3.4.4}$$

Each such problem requires

$$(\ell_i - 1) \cdot m_1 \cdot m_2 \ldots m_p \quad \text{multiplications,}$$

$$(\ell_i - 1) \cdot m_1 \cdot m_2 \ldots m_p \quad \text{additions,}$$

$$m_1 \cdot m_2 \ldots m_p \quad \text{comparisons,}$$

where ℓ_i = number of elements of X in atom A_i of F; note that $\sum_{i=1}^{k} \ell_i = n$. Therefore a total of

$(n-k) \cdot m_1 \cdot m_2 \ldots m_p$ multiplications

$(n-k) \cdot m_1 \cdot m_2 \ldots m_p$ additions

k sets of $m_1 \cdot m_2 \ldots m_p$ comparisons

are required.

This problem corresponds to the usual Bayesian statistical decision problem. Such problems are usually treated by a-posteriori analysis [Ra1]. That is, the quantity

$$E \{h(x, \gamma_1(x), \ldots, \gamma_p(x)) \mid A_i\}$$

$$= E \{h(x, u_1, \ldots, u_p) \mid A_i\}$$

(3.4.5)

where $\gamma_1(x) = u_1$, $\gamma_2(x) = u_2$, $\ldots$, $\gamma_p(x) = u_p$, for all $x \in A_i$, is minimized for each $A_i \in F$. Note that

$$E \{h(x, u_1, \ldots, u_p) \mid A_i\}$$

$$= \frac{\int_{A_i} h(x, u_1, \ldots, u_p) \, d\pi(x)}{\int_{A_i} d\pi(x)} .$$

(3.4.6)

The probability π is normalized to give the conditional probability on A_i in the Bayesian formulation, but this is unnecessary. Therefore, a posteriori analysis is equivalent to the preceeding analysis.

Case 3 Team decision problem.

In this case, Γ_i consists of control laws measurable with respect to the field F^i generated by the partition

$$\{ A_1{}^i, A_2{}^i, \ldots, A_{k_i}{}^i \}.$$

Clearly, there are

$$m_1{}^{k_1} \cdot m_2{}^{k_2} \ldots m_p{}^{k_p}$$

possible control laws. Evidently, then,

$$(n-1) \cdot m_1{}^{k_1} \cdot m_2{}^{k_2} \ldots m_p{}^{k_p} \quad \text{multiplications}$$

$$(n-1) \cdot m_1{}^{k_1} \cdot m_2{}^{k_2} \ldots m_p{}^{k_p} \quad \text{additions}$$

$$m_1{}^{k_1} \cdot m_2{}^{k_2} \ldots m_p{}^{k_p} \quad \text{comparisons}$$

are required to solve the problem.

This figure can be improved upon, but only slightly. For simplicity, assume $\max_{1 \leqslant i \leqslant p} m_i{}^{k_i} = m_p{}^{k_p}$. If $\gamma_1, \gamma_2, \ldots, \gamma_{p-1}$ are given, then

$$\min_{\gamma_p} \int h(x, \gamma_1(x), \ldots, \gamma_{p-1}(x), \gamma_p(x)) \, d\pi(x) \qquad (3.4.7)$$

can be computed as in case 2. This gives $\gamma_p{}^*(\cdot; \gamma_1, \gamma_2, \ldots, \gamma_{p-1})$.

Then

$$\min_{\gamma_1} \min_{\gamma_1} \ldots \min_{\gamma_{p-1}} \int h(x, \gamma_1(x), \ldots, \gamma_{p-1}(x),$$

$$\gamma_p{}^*(x; \gamma_1, \ldots \gamma_{p-1})) \, d\pi(x) \qquad (3.4.8)$$

can be computed. However this procedure cannot be iterated further, since $\gamma_p{}^*$ depends on the <u>entire</u> functions $\gamma_1, \gamma_2, \ldots, \gamma_{p-1}$. In other words, attempting to solve

$$\min_{u_{p-1}} \int_{A_i^{p-1}} h(x, \gamma_1(x), \ldots, u_{p-1}, \gamma_p^*(x; \gamma_1, \ldots, \gamma_{p-1}) \, d\pi(x)$$

(3.4.9)

will not work, since changing the value of γ_{p-1} on the atom A_i^{p-1}

changes γ_p^*. But γ_p^* affects the value of the preceeding integral

over $X - A_i^{p-1}$. Therefore γ_{p-1}^* cannot be obtained by independent

optimizations of integrals over atoms of F^{p-1} - these optimizations are

coupled through γ_p^*.

Therefore, the best that can be done is

$$(n-k_p) \cdot m_p \cdot m_1^{k_1} \cdot m_2^{k_2} \ldots m_{p-1}^{k_{p-1}} \quad \text{multiplications}$$

$$(n-k_p) \cdot m_p \cdot m_1^{k_1} \cdot m_2^{k_2} \ldots m_{p-1}^{k_{p-1}} \quad \text{additions}$$

$$k_{p-1} \cdot m_1^{k_1} \cdot m_2^{k_2} \ldots m_{p-1}^{k_{p-1}} \quad \text{sets of } m_p \text{ comparisons.}$$

This gets formidable very fast. Suppose:

p = 3   (3 controllers)

$m_1 = m_2 = m_3 = 10$ (10 controls)

n = 100   (100 states)

$k_1 = k_2 = k_1 = 2$ (2 observations)

Assuming that a floating point multiplication requires 10^{-5} seconds to

perform, and that a floating point addition requires 10^{-6} seconds[1], about

110 seconds of central processing unit time of a modern high-speed

computer are required just to perform the additions and multiplications.

If there are three observations, this increases to 1.1×10^5 seconds $\approx$

300 hrs! Thus even problems of a rather modest size tax the capabilities

---

[1] These numbers are approximately correct for the IBM 370/165.

of modern computers.

Clearly, some approach other than exhaustive elimination must be employed. One such approach is the following algorithm.

Algorithm

1. Guess control laws $\gamma_1{}^0$, $\gamma_2{}^0$, ..., $\gamma_p{}^0$. Compute

$$J^0 = \int_X h(x, \gamma_1{}^0(x), \gamma_2{}^0(x), \dots, \gamma_p{}^0(x)) \quad d\pi(x)$$

Set $i = 1$, $j = 0$.

2. Solve the problem

$$\tilde{J} = \min_{\tilde{\gamma}_i \in \Gamma_i} \int h(x, \tilde{\gamma}_1(x), \tilde{\gamma}_2(x), \dots, \tilde{\gamma}_p(x)) \quad d\pi(x)$$

where $\tilde{\gamma}_k = \gamma_k{}^j \quad k > i$

and $\tilde{\gamma}_k = \gamma_k{}^{j+1} \quad k < i$.

Let some minimizing γ_i above be denoted $\gamma_i{}^{j+1}$. (Keep $\gamma_i{}^j$ if it is a minimizing control law).

   If $i < p$, set $i = i+1$ and return to 2.

   If $i = p$, check $\tilde{J} < J^j$.

   If $\tilde{J} < J^j$, set $j = j+1$, $i = 1$ and return to 2.

   If $\tilde{J} = J^j$, stop.

Definition [M2]

A set of control laws $\tilde{\gamma}_1$, $\tilde{\gamma}_2$, ..., $\tilde{\gamma}_p$ is <u>person-by-person optimal</u> for the team decision problem if

$$\int h(x_1 \tilde{\gamma}_1(x), \ldots, \tilde{\gamma}_{i-1}(x), \tilde{\gamma}_i(x), \tilde{\gamma}_{i+1}(x), \ldots, \tilde{\gamma}_p(x)) \, d\pi(x)$$

$$= \min_{\gamma_i \in \Gamma_i} \int h(x, \tilde{\gamma}_1(x), \ldots, \tilde{\gamma}_{i-1}(x), \gamma_i(x), \tilde{\gamma}_{i+1}(x), \ldots,$$

$$\tilde{\gamma}_p(x)) \, d\pi(x) \tag{3.4.10}$$

The interpretation of person-by-person optimality is that no team member can <u>unilaterally</u> decrease the cost. Thus person-by-person optimality is a necessary, but not sufficient, condition for optimality.

Theorem 3.4.1

After a finite number of steps, the preceeding algorithm converges to a person-by-person optimal solution.

Proof

Let the set of numbers J such that

$$J = \int h(x, \gamma_1(x), \ldots, \gamma_p(x)) \, d\pi(x) \tag{3.4.11}$$

for some $\gamma_1 \in \Gamma_1, \ldots, \gamma_p \in \Gamma_p$ be denoted S. Since S is finite (it has $\leqslant m_1^{k_1} m_2^{k_2} \ldots m_p^{k_p}$ elements) its elements can be arranged in descending order,

$$S = (J_1, J_2, \ldots, J_\ell), \quad J_i > J_{i+1}.$$

Consider the set of positive numbers

$$T = (J_1 - J_2, \ldots, J_{\ell-1} - J_\ell)$$

and let $\varepsilon = \inf T$. Note that $\varepsilon > 0$.

Consider the difference $J^j - \tilde{J}$ defined in the previous algorithm.

Clearly, either $J^j - \tilde{J} = 0$, or $J^j - \tilde{J} \geqslant \varepsilon$. By induction,

$$J^j \leqslant J^0 - j\varepsilon. \tag{3.4.12}$$

Therefore eventually $J^j = \tilde{J}$, since inf S is finite. But $J^j = \tilde{J} \Rightarrow$ $(\gamma_1^{j+1}, \ldots, \gamma_p^{j+1})$ is person-by person optimal.

The algorithm requires

$$\sum_{i=1}^{p} (n-k_i)\, m_i \quad \text{multiplications}$$

$$\sum_{i=1}^{p} (n-k_i)\, m_i \quad \text{additions}$$

k_1 sets of m_1 comparisons, ..., k_p sets of m_p comparisons per iteration.

Thus the previous example requires $\approx .033$ seconds per iteration for 2 observations, and slightly less time for 3 observations.

The algorithm will always improve a suboptimal strategy, unless that strategy is already person-by-person optimal. It will not produce a globally optimal strategy in general, however. Thus the algorithm is a reasonable approach to the problem. The approach is similar in philosophy to using a gradient algorithm to solve a nonlinear optimization problem.

3.5 The Min-H Algorithm

A substantial number of numerical algorithms have been suggested for the solution of deterministic optimal control problems. The most natural of these for the FSFM problem is the min-H algorithm, which is intimately related to the minimum principle. The min-H algorithm was initially suggested by Kelley [Kel]. Platzman [Pl1] has shown that the algorithm is equivalent to Howard's policy iteration method for Markovian decision processes, and has suggested its application to the imperfect state information case of that problem.

To simplify the notation, the sets X_t and U_t are assumed to have a constant cardinality for $0 \leqslant t \leqslant T$.

Algorithm 3.5.1 (Min-H)

1. Guess γ_1^0, γ_2^0, ..., γ_T^0. Set $j = 0$.

2. Compute $\phi^j(T)$, $\phi^j(T-1)$, ..., $\phi^j(1)$ using γ_T^j, ..., γ_1^j in the adjoint equation $(\phi^j(T) = \phi_T)$. Set $t = 1$.

3. Choose γ_t^{j+1} to minimize $\pi^{j+1}(t-1) \ P^{\gamma_t}(t) \ \phi^j(t) + \pi^{j+1}(t-1) \ h^{\gamma_t}(t)$.[1]

   $(\pi^{j+1}(0) = \pi_0)$

4. If $t < T$, compute $\pi^{j+1}(t) = \pi^{j+1}(t-1) \ P^{\gamma_t^{j+1}}(t)$.

   Set $t = t+1$, and go to 3.

5. If $t = T$, test $J^{j+1} < J^j$, where

$$J^j = \sum_{t=1}^{T} \pi^j(t-1) \ h^{\gamma_t^j}(t) + \pi^j(T) \ \phi_T.$$

---

[1] If γ_t^{j+1} is not unique, choose arbitrarily but with preference for γ_t^j if it is in the minimizing set.

If $J^{j+1} < J^j$, set $j = j+1$, $t = 0$, and go to 2.

If $J^{j+1} = J^j$, stop.

Theorem 3.5.1

The preceeding algorithm converges in a finite number of steps to an extremal solution.

Proof

The proof is completely analogous to the proof of Theorem 3.4.1.

The reason for the strong analogy between the algorithms of section 3.4 and this section is that both are embodiments of the **method of orthogonal search**. The method of orthogonal search applies to the problem

$$\min_{x_1} \ldots \min_{x_n} h(x_1, \ldots, x_n). \tag{3.5.1}$$

The procedure is to fix all of the variables but one and to minimize over that variable. This is done repeatedly, so that the cost decreases monotonically. Convergence is (essentially) assured, but the convergence will not in general be to the optimal solution without further assumptions (e.g., convexity of h).

It is important to note that, as applied to this problem, the min H algorithm is __exactly__ equivalent to orthogonal search. This is a consequence of the fact that the dynamics and cost are linear in π_t. The more general case in optimal control theory is that the Hamiltonian gives only a linear approximation to the optimal cost-to-go. To quote Kelly [Kel]

"In adopting the control y = y*(t) generated by min H* as our next approximation, we must risk the violation of our linearizing assumptions, for this may represent a large step process."

This difficulty will not occur in the preceeding algorithm.

Notice that at each iteration, the quantity

$$\pi(t-1)(P^{\gamma_t}(t) + h^{\gamma_t}(t)) \qquad (3.5.2)$$

must be computed for all $\gamma \in \Gamma_t$. Evidently each such computation requires

$n^2 + n-1$ multiplications

$n^2 + n-1$ additions

for each $\gamma \in \Gamma_t$ Since the number of γ is on the order of m^n (m = # controls, n = # states), this computation appears to be hopeless for even moderately sized problems.

However, a deeper look at the structure of the problem shows that the situation can be improved considerably. Define

$$P_{ij}^{u}(t) = p_t(\{q : j = f_t(i, u, q)\}) \qquad 1 \leqslant i,j \leqslant n \qquad (3.5.3)$$

It follows that

$$P_{ij}^{\gamma}(t) = P_{ij}^{u}(t) \qquad j = 1, \ldots, n \qquad (3.5.4)$$

when $u = \gamma(i)$. Therefore, for all $\gamma \in \Gamma_t$ each row of $P^{\gamma}(t)$ is a row of $P^u(t)$ for some $u \in U_t$. There are precisely nm such rows. Therefore the column vectors

$$P^u(t) \ \phi(t) + h^u(t)$$

can be computed with mn^2 multiplications and then the column vectors

$$p^\gamma(t) \; \phi(t) \; + \; h^\gamma(t)$$

can be formed by selecting the appropriate elements from the set
$\{p^u(t) \; \phi(t) \; + \; h^u(t)\}$.

Thus the quantities

$$\pi(t-1) \; (p^\gamma(t) \; \phi(t) \; + \; h^\gamma(t))$$

can be computed with

$$2 \; mn^2 \quad \text{multiplications}$$

$$\approx mn \; (n-1) \; + \; n \; + \; n^2 \; \frac{m^{n-1}-1}{m-1} \quad \text{additions}$$

(assuming Γ has $\approx m^n$ elements). This is a considerable improvement, especially since multiplications take approximately 10 times longer to perform than additions. However, the number of additions is still too large. Further improvement can only be made in the light of assumptions on the nature of Γ.

<u>Case 1</u> $\Gamma_t = U_t^{X_t}$ (perfect state measurement)

This is the simplest case. Simply choose $\gamma_t^*(i) = u^*$, where

$$\sum_j P_{ij}^{u^*}(t) \; \phi_j(t) \; + \; h_i^{u^*}(t) \; = \; \min_u \; \sum_j P_{ij}^u(t) \phi_j(t) \; + \; h_i^u(t) \qquad (3.5.5)$$

This requires mn multiplications, mn additions, and n sets of m comparisons.

Case 2 (imperfect state observations)

Γ_t consists of control laws that are measurable with respect to the field generated by a finite partition $\{A_{t-1}^1, A_{t-1}^2 \ldots, A_{t-1}^k\}$ of X_{t-1}.

Choose $\gamma_t^*(i) = u^*$ for all $i \in A_{t-1}^\ell$, where

$$\sum_{i \in A_{t-1}^\ell} \pi_i(t-1) \left\{ \sum_j P_{ij}^{u^*}(t) \ \phi_j(t) + h_i^{u^*}(t) \right\}$$

(3.5.6)

$$= \min_{u \in U_t} \sum_{i \in A_{t-1}^\ell} \pi_i(t-1) \left\{ \sum_j P_{ij}^{u}(t) \ \phi_j(t) + h_i^{u}(t) \right\} .$$

This requires about $m(n^2+n)$ multiplications, $mn^2 + m(n-k)$ additions, and k sets of m comparisons.

By now the close connection between cases 1 and 2 of section 4 and the above should be apparent. This is a consequence of the fact that the problem

$$\min_{\gamma_t \in \Gamma_t} \pi(t-1) \ (P^{\gamma_t}(t) \ \phi(t) + h^{\gamma_t}(t))$$

is precisely a team problem as defined in section 4. Clearly, the analysis of Case 3 of the first section can also be extended.

Case 3 (dynamic team problem)

$$\Gamma_t = \Gamma_t^1 \times \Gamma_t^2 \times \ldots \Gamma_t^k$$

Γ_t^i consists of control laws measurable with respect the field generated by the partition

$$\left\{ {}^i A_{t-1}^1, \ {}^i A_{t-1}^2, \ \ldots, \ {}^i A_{t-1}^{k_i} \right\} \qquad i = 1, \ldots, k$$

of X_{t-1} (k controllers).

As in section 4, the combinatorics of the problem are overwhelming, so that resort must be made to the notion of person-by-person optimality.

Make the following notational convention:

$$\gamma_t : X_{t-1} \to U_t^1 \times U_t^2 \times \ldots \times U_t^n$$

$$\gamma_t = (\gamma_t^1, \gamma_t^2, \ldots, \gamma_t^k).$$

Then

$$J(\gamma_1, \gamma_2, \ldots, \gamma_T)$$

$$= J(\gamma_1^1, \ldots, \gamma_1^n, \gamma_2^1, \ldots, \gamma_2^n, \ldots, \gamma_T^1, \ldots, \gamma_T^n).$$

Definition

A sequence

$$\gamma^* = (\gamma_1^*, \ldots, \gamma_T^*) = (\gamma_1^{1*}, \ldots, \gamma_1^{n*}, \ldots, \gamma_T^{1*}, \ldots, \gamma_T^{n*})$$

is said to be a person-by-person extremal if

$$J(\gamma_1^{1*}, \ldots, \gamma_t^{i*}, \ldots, \gamma_T^{k*})$$

$$\leqslant J(\gamma_1^{1*}, \ldots, \gamma_t^{i}, \ldots, \gamma_T^{k*}) \quad \text{for all} \quad \gamma_t^i \in \Gamma_t^i,$$

$$i = 1, \ldots, k, \quad t = 1, \ldots, T. \tag{3.5.7}$$

Every optimal control law sequence is a person-by-person extremal, but the converse need not be true. Algorithms 3.4.1 and 3.5.1 can be combined to given an algorithm that always converges to a person-by-person extremal. The order of minimization is

$$\gamma_1{}^1, \ \gamma_2{}^1, \ \ldots, \ \gamma_T{}^1, \ \gamma_1{}^2, \ \gamma_2{}^2, \ \ldots, \ \gamma_T{}^2, \ \ldots, \ \gamma_1{}^k, \ \gamma_2{}^k, \ \ldots, \ \gamma_T{}^k.$$

Thus K forward and backward sweeps of the state and costate equations

are required per iteration. The number of multiplications required is

(exclusive of the state and costate computation) is

$$T\left(\sum_{i=1}^{k} m_i(n^2 + n) \right)$$

and

$$T\left(\sum_{i=1}^{k} m_i(n^2 + n - k_i) \right)$$

additions are required with

$$T \sum_{i=1}^{k} k_i$$

sets of m_i comparisons.

Notice that person-by-person approach is consistent with the minimum

principle approach:

1. both approaches given necessary conditions for optimality

2. both approaches are sufficient only under convexity assumptions
   that do not hold in general

3. An initial guess is improved, but the improvement may stop
   short of optimal.

These facts are consequences of the fact that the person-by-person and

min H algorithms are actually both concrete realizations of orthogonal

search.

CHAPTER IV

DYNAMIC PROGRAMMING FOR THE FSFM PROBLEM

In this chapter, the dual dynamic programming equations for forward

and backward induction are presented. These equations follow

immediately from classical dynamic programming theory [B1] as applied

to the equivalent deterministic problem of Chapter II.

Numerical solution of the dynamic programming equations is a dif-

ficult task. Three approaches are suggested. The first is the usual

technique of replacing the continuous state space with a discrete grid.

The second exploits the fact that the reachable and coreachable sets

of the problem are finite. The third approach applies an algorithm of

Sondik [Sol,Sml] developed for the Markovian decision problem with

incomplete state information to the FSFM problem.

The chapter closes with an example.

4.1 The Equations for Forward and Backward Induction

Recall from Chapter 2 that the deterministic optimal control problem

equivalent to the FSFM problem is to minimize

$$J(\gamma) = \pi(T)\phi(T) + \sum_{t=1}^{T} \pi(t-1) \, h^{\gamma_t}(t) \tag{4.1.1}$$

for $\gamma \varepsilon \Gamma$ subject to

$$\pi(t) = \pi(t-1) \, P^{\gamma_t}(t), \quad t = 1,2,\ldots,T \tag{4.1.2}$$

$$\pi(0) = \pi_0 \qquad\qquad (4.1.3)$$

where $\Gamma = \Gamma_1 \times \Gamma_2 \times \ldots \times \Gamma_T$ is a finite set.

Define the functions

$$V_T(\pi) = \pi\phi_T , \qquad\qquad (4.1.4)$$

$$V_t(\pi) = \min_{\gamma_{t+1}\varepsilon\Gamma_{t+1}} \left[\pi h^{\gamma_{t+1}}(t+1) + V_{t+1}(\pi P^{\gamma_{t+1}}(t+1)) \right] , \qquad (4.1.5)$$

$$W_0(\phi) = \pi_0\phi , \qquad\qquad (4.1.6)$$

$$W_{t+1}(\phi) = \min_{\gamma_{t+1}\varepsilon\Gamma_{t+1}} \left[W_t(P^{\gamma_{t+1}}(t+1) \; \phi + h^{\gamma_{t+1}}(t+1)) \right] . \qquad (4.1.7)$$

The following theorems describe the dynamic programming algorithms for backward and forward induction. The proofs are a simple application of dynamic programming theory [B1].

Theorem 4.1.1 (Backward Induction)

Let the map $\delta_{t+1}: \Pi_t \rightarrow \Gamma_{t+1}$ be defined for $t = 0,1,\ldots,T-1$ by by $\delta_{t+1}(\pi(t)) = \gamma_{t+1}$ if γ_{t+1} is a minimizing control law in (4.1.5) for $\pi = \pi(t)$. Define the quantities $\pi^*(t)$, γ_t^* by

$$\pi^*(0) = \pi_0 , \qquad\qquad (4.1.8)$$

$$\gamma_t^* = \delta_t \; (\pi^*(t-1)), \quad t = 1,2,\ldots,T, \qquad (4.1.9)$$

$$\pi^*(t) = \pi^*(t-1)P^{\gamma_t^*}(t), \quad t = 1,2,\ldots,T. \qquad (4.1.10)$$

Then $\gamma^* = (\gamma_1^*,\gamma_2^*,\ldots,\gamma_T^*)$ is an optimal control law sequence with corresponding sequence $\pi^*(0),\pi^*(1),\ldots,\pi^*(T)$ of optimal states.

Theorem 4.1.2 (Forward Induction)

Let the map $\delta_{t+1}: \Phi_{t+1} \to \Gamma_{t+1}$ be defined for $t = 0,1,\ldots,T-1$ by

$\delta_{t+1}(\phi(t+1)) = \gamma_{t+1}$ if γ_{t+1} is the minimizing control law in (4.1.7)

for $\phi = \phi(t+1)$. Define the quantities $\phi^*(t), \gamma_t^*$ by

$$\phi^*(T) = \phi_T \quad , \tag{4.1.11}$$

$$\gamma_t^* = \delta_t(\phi^*(t)), \quad t = 1,2,\ldots,T, \tag{4.1.12}$$

$$\phi^*(t-1) = P^{\gamma_t^*}(t)\phi^*(t) + h^{\gamma_t}(t), \quad t = 1,2,\ldots,T. \tag{4.1.13}$$

Then $\gamma^* = (\gamma_1^*, \gamma_2^*, \ldots, \gamma_T^*)$ is an optimal control law sequence with

corresponding sequence $\phi^*(0), \phi^*(1), \ldots, \phi^*(T)$ of optimal costates.

There are a number of comments to be made about Theorems 4.1.1

and 4.1.2. First notice that although the original state space

X_t is finite, the dynamic programming algorithm must be carried out

in the uncountable state space Π_t of the equivalent deterministic

problem. This is due to the following requirement of dynamic program-

ming as expressed by Bellman [B2].

"After any number of decisions, say k, we wish the effect of

the remaining N-k stages of the decision process on the total return

to depend only upon the state of the system at the end of the k-th

decision and the subsequent decisions."

This statement may be regarded as a definition of what constitutes

a _state_ for purposes of dynamic programming. It is well-known in

classical stochastic control that the state for dynamic programming

purposes is not the physical state, but is instead the probability

distribution of the physical state conditioned on past measurements

and inputs. The approach of this chapter, in which the <u>unconditional</u>

distribution of the physical state is the dynamic programming state

is due to Witsenhausen [W4].

Second, notice that both the forward and backward algorithms

require two passes. If backward induction is applied, the first

(backward) pass determines δ_t from equations (4.1.4) and (4.1.5), and

then the second (forward) pass eliminates the dependence of the

control laws on the state $\pi(t)$. Similarly, for forward induction,

the first (forward) pass determines δ_t from equations (4.1.6) and

(4.1.7), and then the second (backward) pass eliminates the dependence

of the control laws on the costate $\phi(t)$. Thus although only the

control laws corresponding to the sequence $\pi^*(0), \pi^*(1), \ldots, \pi^*(T)$ or

$\phi^*(0), \phi^*(1), \ldots, \phi^*(T)$ are of interest, the optimal control laws

for all such sequences must be computed. The original problem

has effectively been <u>embedded</u> into an entire class of similar

problems.

Third, Theorems 4.1.1 and 4.1.2 exhibit a particularly striking

duality. The function $V_t(\cdot)$ gives the optimal cost-to-go at time t

for a given state probability vector π. The function $W_t(\cdot)$ gives

the optimal cost-to-go at time t for a given cost-to-go vector ϕ.

The vector $\pi(t)$ summarizes the effect of past control laws, the

vector $\phi(t)$ summarizes the effect of future control laws, and their

scalar product $\pi(t)\phi(t)$ is the expected cost-to-go.

Finally, although Theorems 4.1.1 and 4.1.2 provide an elegant

dual set of sufficient conditions for the FSFM problem, development

of feasible numerical algorithms based on these theorems is difficult.

Some possible approaches are discussed in the next section.

4.2 Numerical Solution

Since the dynamic programming algorithms for forward and backward induction formulated in the previous section have a continuous state space, some type of discretization is necessary for digital computer solution. The straightforward approach is to define a partition over the space Π_t or Φ_t. Note, however, that if Π_t and Φ_t are n_t-dimensional and each dimension is partitioned into 100 elements (for a 1% partition), the grid has 100^{n_t} elements! This is the well-known **curse of dimensionality**. Although numerous heuristic schemes have been suggested to minimize this difficulty (see [Lar1] for example), none has been widely accepted.

An alternative procedure utilizes the special structure of the state and costate equations. Recall that the reachable sets $r_t(\pi_0)$ are defined by equation (3.3.7). Similarly, the **coreachable sets** $\rho_t(\phi_T)$ are defined by

$$\rho_T(\phi_T) = \{ \phi_T \} , \qquad (4.2.1)$$

$$\rho_{t-1}(\phi_T) = \left\{ P^{\gamma_t}(t)\phi(t) + h^{\gamma_t}(t) : \gamma_t \varepsilon \Gamma_t , \phi(t) \varepsilon \rho_t(\phi_T) \right\} , \qquad (4.2.2)$$

for $t = 1,2,\ldots, T$. The interesting fact is that although Π_t and Φ_t are continuous, the sets $r_t(\pi_0)$ and $\rho_t(\phi_T)$ are discrete. Since the minimizations in (4.1.5) and (4.1.7) need only be carried out for $\pi \varepsilon r_t(\pi_0)$ and $\phi \varepsilon \rho_{t+1}(\phi_T)$, respectively, it is really unnecessary to consider Π_t and Φ_t. The difficulty with this approach is that the sets r_t and ρ_t, although finite, will in general be quite large.

Upper bounds to the number of elements of $r_t(\pi_0)$ and $\rho_t(\phi_T)$ are

$$card(r_t(\pi_0)) \leq \sum_{\tau=1}^{t} card(\Gamma_\tau) \qquad (4.2.3)$$

$$card(\rho_t(\phi_T)) \leq \sum_{\tau=t+1}^{T} card(\Gamma_\tau) \qquad (4.2.4)$$

where card (A), the cardinality of a finite set A, is the number of

elements of A. Although these bounds will in general not be achieved

(since two distinct control law sequences can lead to the same state

or costate), the sets $r_t(\pi_0)$ and $\rho_t(\phi_T)$ will still be too large to

be conveniently computed except in particularly simple cases.

A third approach to the numerical solution of the dynamic

programming equations uses the special structure of the problem in

a different way. The key observation is that the functions

W_t, V_t are piecewise linear and concave.

Proposition 4.2.1

Consider the functions $V_t(\cdot)$ defined by (4.1.4), (4.1.5). For

$t = 0,1,\ldots,T$, there is a finite set of (column) vectors A_t such that

$$V_t(\pi) = \min_{a(t)\in A_t} \pi a(t) \qquad (4.2.5)$$

Proof

The proof proceeds by a backward induction argument. Note that

at $t = T$,

$$V_T(\pi) = \pi\phi_T \qquad (4.2.6)$$

so that (4.2.5) holds with

$$A_T = \left\{ \phi_T \right\} . \tag{4.2.7}$$

If it is assumed that

$$V_{t+1}(\pi) = \min_{a(t+1)\varepsilon A_{t+1}} \pi a(t+1) , \tag{4.2.8}$$

then

$$V_t(\pi) = \min_{\gamma_{t+1}\varepsilon \Gamma_{t+1}} \left[\pi h^{\gamma_{t+1}}(t+1) + V_{t+1}(\pi P^{\gamma_{t+1}}(t+1)) \right]$$

$$= \min_{\gamma_{t+1}\varepsilon \Gamma_{t+1}} \left[\pi h^{\gamma_{t+1}}(t+1) + \min_{a(t+1)\varepsilon A_{t+1}} \pi P^{\gamma_{t+1}}(t+1) a(t+1) \right]$$

$$= \min_{\gamma_{t+1}\varepsilon \Gamma_{t+1}} \min_{a(t+1)\varepsilon A_{t+1}} \left[\pi h^{\gamma_{t+1}}(t+1) + \pi P^{\gamma_{t+1}}(t+1) a(t+1) \right]$$

$$= \min_{a(t)\varepsilon A_t} \pi a(t) \tag{4.2.9}$$

where

$$A_t = \left[h^{\gamma_{t+1}}(t+1) + P^{\gamma_{t+1}}(t+1) a(t+1) : \gamma_{t+1}\varepsilon \Gamma_{t+1}, a(t+1)\varepsilon A_{t+1} \right]. \tag{4.2.10}$$

Therefore, the proposition is true for all t, $0 \leq t \leq T$.

A similar result holds for the function $W_t(\cdot)$.

Proposition 4.2.2

Consider the functions $W_t(\cdot)$ defined by (4.1.6), (4.1.7). For

$t = 0,1,\ldots,T$, there is a finite set of (row) vectors B_t such that

$$W_t(\phi) = \min_{b(t) \in B_t} b(t)\phi . \qquad (4.2.11)$$

Proof

The proof proceeds by a forward induction argument completely

analogous to the proof of Proposition 4.2.1, and is therefore omitted.

A representation similar to (4.2.5) has appeared in the

literature on Markovian decision processes with incomplete state

information. Evidently, Astrom [As1] was the first to use the

representation for a specific example. However, Sondik [So1,Sm1]

has systematically exploited the representation to derive an

algorithm for the backward equation. It is shown below that the

Sondik algorithm can be directly applied to the backward equations

arising in the FSFM problem. Moreover, the algorithm can be dualized

to apply to the forward equations.

Note that the set A_t defined in the proof of Proposition 4.2.1

is, in general, larger than necessary. For a given $a^*(t) \in A_t$, there

may not exist a $\pi \in \Pi_t$ such that

$$\min_{a(t) \in A_t} \pi a(t) = \pi a^*(t) . \qquad (4.2.12)$$

To find a smaller set $\tilde{A}_t \subset A_t$ satisfying the above condition, it is

necessary to look at the problem more closely.

Suppose that a set $\tilde{A}_t$ is given that contains a minimizing element

of A_t in (4.2.12) for every $\pi \epsilon \Pi_t$. Suppose that $\tilde{A}_t$ has j_t elements,

$$\tilde{A}_t = \left\{ a^1(t), \ldots, a^{j_t}(t) \right\} .$$

(4.2.13)

Since $\tilde{A}_t$ is a subset of the set A_t defined by (4.2.10), it follows that

$$\text{card}(\tilde{A}_t) \leq \sum_{\tau=t+1}^{T} (\text{card } \Gamma_\tau)$$

for $t = 0,1,\ldots,T-1$, and $\text{card}(\tilde{A}_T) = 1$. Let the sets $R_j(t)$ be defined

for $j = 1,2,\ldots,j_t$ by

$$R_j(t) = \left\{ \pi \epsilon \Pi_t : \min_{a(t) \epsilon \tilde{A}_t} \pi a(t) = \pi a^j(t) \right\} .$$

(4.2.14)

Assume that $\tilde{A}_t$ is chosen so that $R_j(t) \neq \phi$. Clearly,

$$\Pi_t = \bigcup_{j=1}^{j_t} R_j(t) .$$

(4.2.15)

Lemma 4.2.3

The sets $R_j(t)$, $j = 1,2,\ldots,j_t$ and $t = 0,1,\ldots,T$, are convex

sets with linear boundaries.

Proof

Note that

$$R_j(t) \cap R_k(t) = \left\{ \pi \epsilon R_j(t) \cup R_k(t) : \pi(a^j(t) - a^k(t)) = 0 \right\},$$

(4.2.16)

so that if there is a boundary between $R_j(t)$ and $R_k(t)$, it is linear.

To prove convexity, suppose $\pi^1, \pi^2 \epsilon R_j(t)$. Then

$$\pi^1 a^j(t) \leq \pi^1 a(t)$$

(4.2.17)

$$\pi^2 a^j(t) \leq \pi^2 a(t) \tag{4.2.18}$$

for all $a(t) \varepsilon \tilde{A}_t$. Therefore, all $\lambda \varepsilon [0,1]$,

$$\lambda \pi^1 a^j(t) + (1-\lambda)\pi^2 a^j(t) \leq \lambda \pi^1 a(t) + (1-\lambda)\pi^2 a(t) \tag{4.2.19}$$

for all $a(t) \varepsilon \tilde{A}_t$. Equation (4.2.19) implies that $\lambda \pi^1 + (1-\lambda)\pi^2 \varepsilon R_j(t)$.

By assumption, for each $a(t) \varepsilon \tilde{A}_t \subset A_t$, there exist $\lambda_t \varepsilon \Gamma_t$ and $a(t+1) \varepsilon \tilde{A}_{t+1}$ such that

$$a(t) = h^{\gamma_{t+1}}(t+1) + P^{\gamma_{t+1}}(t+1)a(t+1). \tag{4.2.20}$$

Lemma 4.2.4

For all $\pi \varepsilon R_j(t), \delta_{t+1}(\pi) = \gamma_{t+1}^j$. I.e., δ_t is constant over the sets $R_j(t)$.

Proof

From (4.2.14),

$$\min_{a(t) \varepsilon A_t} \pi a(t) = \pi a^j(t) \tag{4.2.21}$$

for all $\pi \varepsilon R_j(t)$. Let

$$a^j(t) = h^{\gamma_{t+1}^j}(t+1) + P^{\gamma_{t+1}^j}(t+1)a(t+1) \tag{4.2.22}$$

for some $\gamma_{t+1}^j \varepsilon \Gamma_{t+1}$, $a(t+1) \varepsilon \tilde{A}_{t+1}$. Then the lemma follows from (4.2.9), (4.2.21), and (4.2.22).

At this point, it is clear that applying the backward algorithm is equivalent to computing

$$\{(R_j(t-1), \gamma_t^j, a^j(t-1), j=1,2,\ldots,j_t\} \tag{4.2.23}$$

for $t = 1,2,\ldots,T$. These quantities are computed as follows.

Suppose that $\{R_j(t)\}$ and $\{a^j(t)\}$ have been computed. (At $t=T$, $j_T = 1$, $R_1(T) = \Pi_T$, and $\tilde{A}_T = \{\phi_T\}$). Consider some fixed, arbitrary $\pi \varepsilon \Pi_{t-1}$. Compute

$$\min_{\gamma_t \in \Gamma_t} \quad \min_{a(t) \in \tilde{A}_t} \quad \pi(h^{\gamma_t}(t) + P^{\gamma_t}(t)a(t))$$

$$= \pi(h^{\gamma_t^1}(t) + P^{\gamma_t^1}(t)a^j(t)), \qquad (4.2.24)$$

where $\pi P^{\gamma_t^1}(t) \in R_j(t)$, and let

$$a^1(t-1) = h^{\gamma_t^1}(t) + P^{\gamma_t^1}(t)a^j(t). \qquad (4.2.25)$$

At this stage, one point $\pi \in R_1(t-1)$ has been found, the first

point $a^1(t-1)$ of $\tilde{A}_{t-1}$ has been obtained, and γ_t^1 determined. Next. the

boundary of $R_1(t-1)$ is determined. Notice that $\pi \in R_1(t-1)$ if and only if

$$\pi a^1(t-1) \leq \pi(h^{\gamma_t}(t) + P^{\gamma_t}(t)a(t)) \qquad (4.2.26)$$

for all $\gamma_t \in \Gamma_t$, $a(t) \in \tilde{A}_t$. A point $\pi \in R_1(t-1)$ is on the boundary of

$R_1(t-1)$ if and only if there exist $\hat{\gamma}_t$, $\hat{a}(t)$ such that

$$\pi a^1(t-1) = \pi(h^{\hat{\gamma}_t}(t) + P^{\hat{\gamma}_t}(t)\hat{a}(t)). \qquad (4.2.27)$$

This condition can be tested by solving the following linear

program. The problem is to minimize

$$\pi(h^{\hat{\gamma}_t}(t) + P^{\hat{\gamma}_t}(t)\hat{a}(t) - a^1(t-1)) \qquad (4.2.28)$$

over $\pi \in \Pi_{t-1}$, subject to

$$\pi(h^{\gamma_t}(t) + P^{\gamma_t}(t)a(t) - a^1(t-1)) \geq 0 \qquad (4.2.29)$$

for all $\gamma_t \in \Gamma_t$, $a(t) \in \tilde{A}_t$. The first $\hat{\gamma}_t \in \Gamma_t$ and $\hat{a}(t) \in \tilde{A}_t$ for which

the minimum of (4.2.28) is zero define $\gamma_t^2 = \hat{\gamma}_t$, and

$$a^2(t-1) = h^{\hat{\gamma}_t}(t) + P^{\hat{\gamma}_t}(t)\hat{a}(t). \qquad (4.2.30)$$

By repeatedly solving the linear program, in a similar manner all

the vectors $a^j(t-1)$ and control laws γ_t^j corresponding to regions

bordering $R_1(t-1)$ are determined. The procedure is then repeated

until all the regions $R_j(t-1)$ with the corresponding vectors

$a^j(t-1)$ and control laws $\gamma_t^{\ j}$ are determined. Repetition of the

algorithm permits the computation of all the quantities in (4.2.23).

The resulting algorithm is summarized in Figure 2.4.1. The

corresponding dual algorithm is illustrated by Figure 2.4.2.

Of course, after one of the algorithms is carried out, a sweep

of the state or costate equations is required to eliminate the

dependence of the control laws.

Sondik's algorithm is an attempt to circumvent the curse of

dimensionality that arises when the state space Π_t is partitioned

by a grid. The algorithm has the desirable properties that

(i) it is exact

(ii) the partition[1] may be considerably coarser than a

naive grid partition.

However, the number of elements of the partition is not known a priori,

and can be expected to increase rapidly with increasing T. Moreover,

the irregular nature of the partition sets makes computer storage

awkward.

Three alternative approaches to the solution of the dynamic

programming functional equations have been given in this section. All

of the approaches are extremely limited with respect to the size of

---

[1] The term partition is used in an informal sense here, since distinct
elements of the partition can share a common boundary and therefore
have a nonempty intersection.
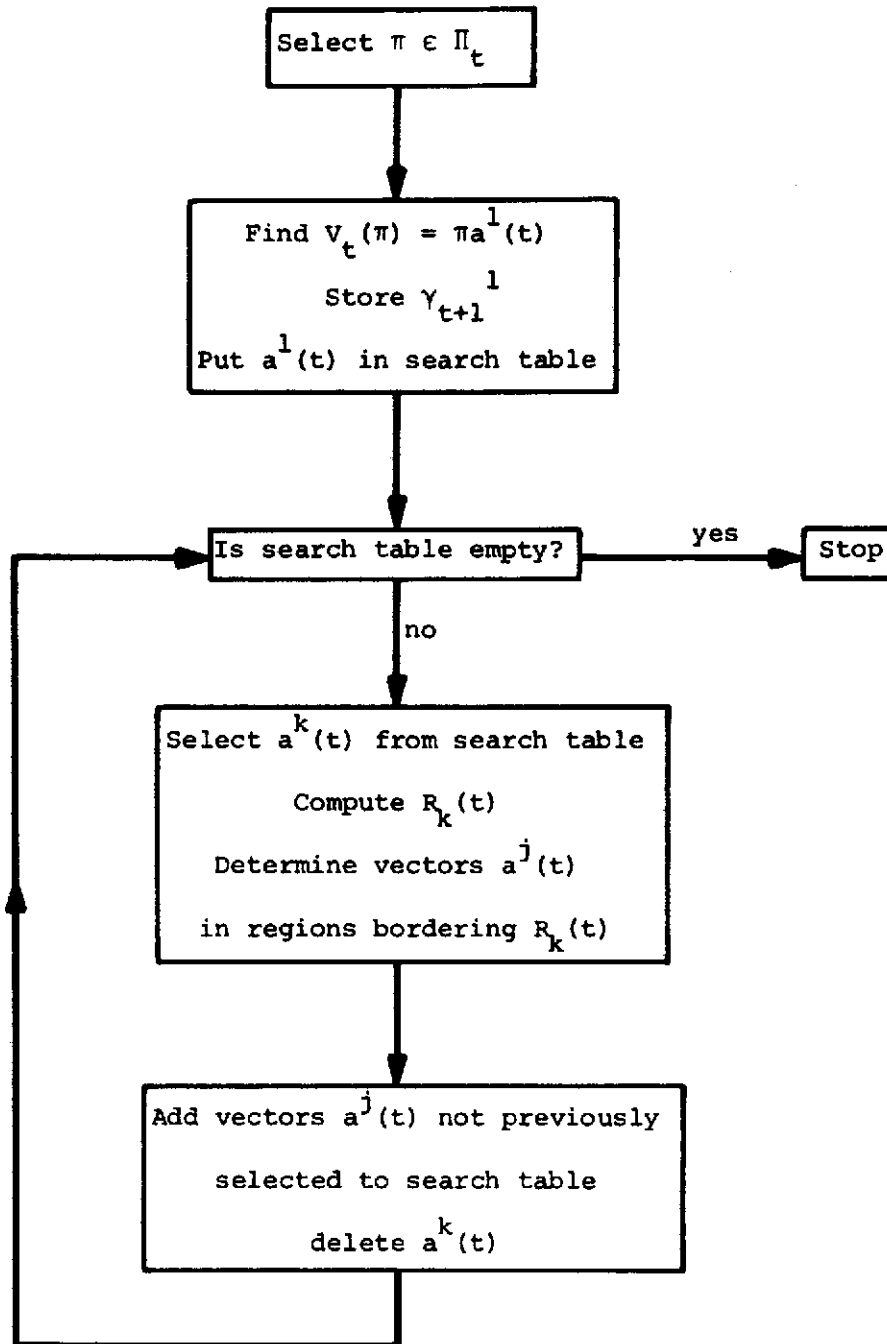
Figure 4.2.1   Sondik's Algorithm Applied to FSFM Problem
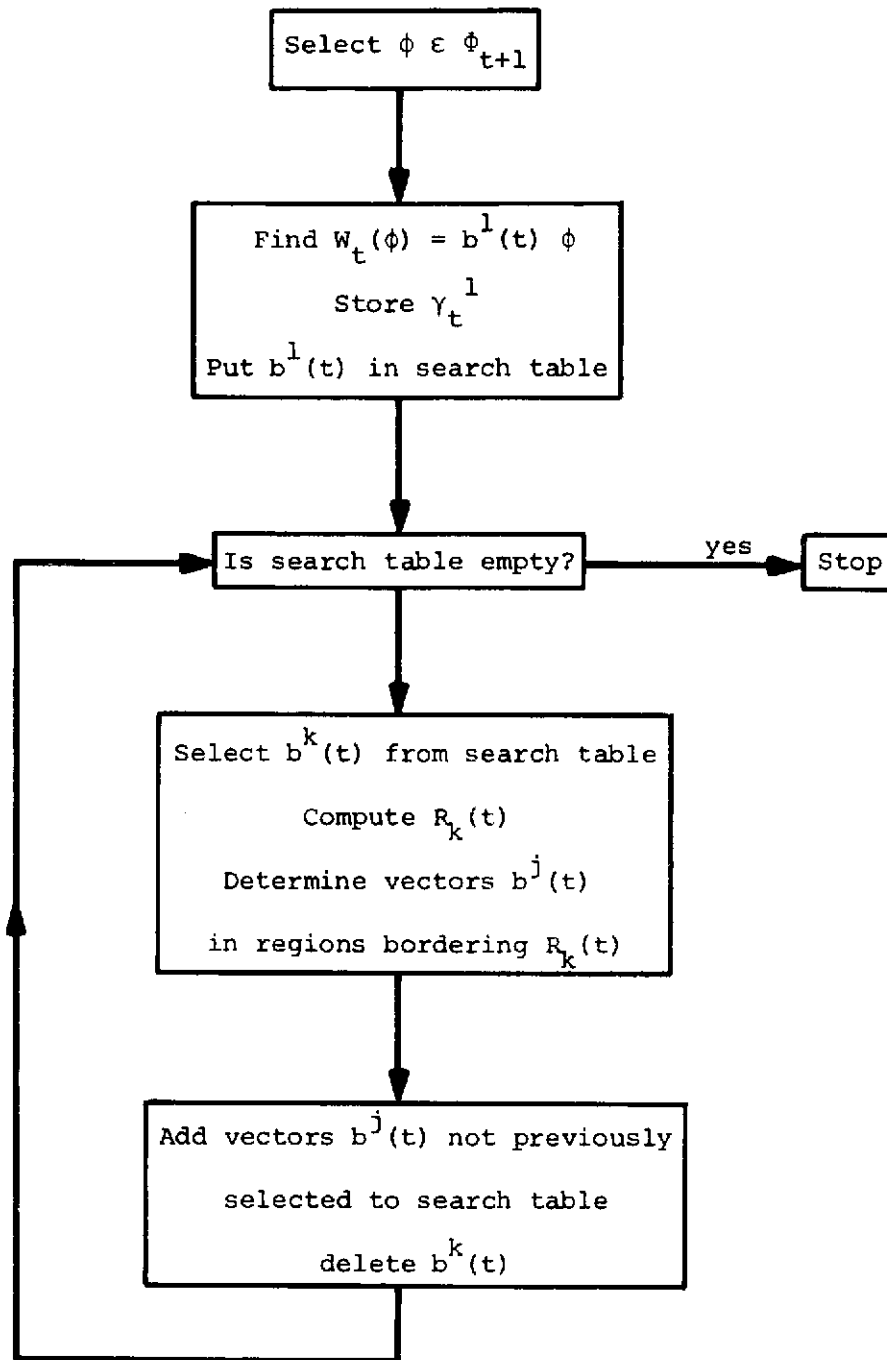
Figure 4.2.2   Dual Version of Sondik's Algorithm
Applied to the FSFM Problem.

problems they can handle for reasons that have been discussed. This
situation is not surprising since it is well known that dynamic program-
ming for problems with continuous state space is an exceedingly
difficult computation problem.

Thus, the situation for the FSFM problem is quite similar to
that encountered in deterministic optimal control theory. Due to the
computational difficulties associated with its use, dynamic programming
is seldom applied to numerical solution of optimal control problems.
Instead algorithms based on the minimum principle are used, even
though these algorithms may converge to extremal, rather than optimal,
solutions. Nevertheless, these algorithms, when combined with appro-
priate engineering judgement, have been found to produce solutions
that are often highly superior to those developed on the basis of
intuition alone. An indication that the min-H algorithm developed in
Chapter 3 can play a similar role for the FSFM problem is provided by
the analysis in Chapter 6.

4.3 Example

In this section, the two approaches to the numerical application of the dynamic programming algorithm developed in the previous section are applied to the second example of section 3.2. Only the backward equations are illustrated.

Recall that the example has state sets

$$X_0 = \{1,2\} \ , \ X_1 = \{1,2,3,4,\} \ , \ X_2 = \{1,2,3,4\} \quad \text{and}$$

control sets $U_1 = \{0,1\} \ , \ U_2 = \{0,1\}$. The parameters of the equivalent deterministic problem are $(0 \leq k < 1)$:

$$h(1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad , \quad h(2) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ k \end{bmatrix} \quad , \quad \phi(2) = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$P^1(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P^0(0) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P^1(1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P^0(1) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Attention is restricted to the case in which

$$\Gamma_1 = U_1^{X_0}, \text{ but } \Gamma_2 = \{\gamma_2 \in U_2^{X_0} : \gamma_1(1) = \gamma_2(2) = \gamma_2(3)\}.$$

The reachable set $r_1(\pi_0)$ has four elements:

$$r_1(\pi_0) = \left\{ \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}, \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}, \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \right\}.$$

The function

$$V_1(\pi) = \min_{\gamma_2 \in \Gamma_2} \left[\pi h^{\gamma_2}(2) + \pi P^{\gamma_2}(2)\phi_2 \right]$$

must be evaluated for each $\pi \in r_1(\pi_0)$, and a corresponding minimizing control law tabulated.

The result is

$$\delta_2(\pi) \equiv \gamma_2^* ,$$

where $\gamma_2^*(x) = 1$ for $x = 1,2,3$, $\gamma_2^*(4) = 0$, and

$$V_1\left(\begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix} \right) = \frac{1}{2} k ,$$

$$V_2\left(\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \right) = \frac{1}{2} ,$$

$$V_3 \left(\left[0 \ \tfrac{1}{2} \ 0 \ \tfrac{1}{2} \right] \right) = \tfrac{1}{2} k \quad ,$$

$$V_4 \left(\left[0 \ \tfrac{1}{2} \ \tfrac{1}{2} \ 0 \right] \right) = \tfrac{1}{2} \quad .$$

Next the function

$$V_0(\pi) = \min_{\gamma_1 \epsilon \Gamma_1} V_1(\pi P^{\gamma_1}(1))$$

must be computed for $\pi \epsilon r_0(\pi_0) = \{\pi_0\}$ and the corresponding control

law tabulated. The result is

$$V_1 \left(\left[\tfrac{1}{2} \ \tfrac{1}{2} \right] \right) = \tfrac{1}{2} k$$

with $\delta_1(\pi) = \gamma_1^*$

where $\gamma_1^*(1) = 1$, $\gamma_1^*(2) = 0$. This is in agreement with the results

obtained earlier.

The solution is now recomputed using the Sondik algorithm. The

algorithm starts with

$$R_1(2) = \Pi_2$$

and

$$A_2 = \tilde{A}_2 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\} \quad .$$

For $t = 1$,

$$A_1 = \left\{ h^{\gamma_2}(2) + P^{\gamma_2}(2)\phi_2 : \gamma_2 \epsilon \Gamma_2 \right\} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1+k \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ k \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1+k \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ k \end{bmatrix} \right\} \quad .$$

It is obvious that the first and third vectors of A_1 are not required

for $\tilde{A}_1$. Let the arbitrary point $\pi = [1 \ 0 \ 0 \ 0] \ \epsilon \ \Pi_1$ be selected.

Performing the minimization

$$\min_{\gamma_2 \epsilon \Gamma_2} \pi(h^{\gamma_2}(2) + P^{\gamma_2}(2)\phi_2)$$

gives $\gamma_2^1(4) = 0$, $\gamma_2^1(1) = \gamma_2^1(2) = \gamma_2^1(3) = 1$, and

$$a^1(1) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ k \end{bmatrix}.$$

Checking the inequalities that define $R_1(1)$,

$$\pi a^1(1) \leq (h^{\gamma_2}(2) + P^{\gamma_2}(2)\phi_2)$$

for all $\gamma_2 \ \epsilon \ \Gamma_2$, it is verified that

$$R_1(1) = \{\pi : \ \pi_3 \leq \pi_1 + \pi_2\}$$

and that

$$a^2(1) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ k \end{bmatrix}$$

where $\gamma^2 \equiv 0$. The region $R_2(1)$ is defined by the inequalities

$$\pi a^2(1) \leq \pi(h^{\gamma_2}(2) + P^{\gamma_2}(2)\phi_2)$$

for all $\gamma_2 \in \Gamma_2$. Since checking the inequalities shows that

$$R_2(1) = \{\pi : \pi_3 \geq \pi_1 + \pi_2\}$$

so that $\Pi_1 = R_1(1) \bigcup R_2(1)$, the algorithm gives

$$\tilde{A}_1 = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ k \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ k \end{bmatrix} \right\}.$$

For $t = 0$, choose $\pi = [1 \quad 0]$. A control law minimizing

$$\min_{a(1) \in \tilde{A}_1} \pi P^{\gamma_1}(1) \, a(1)$$

is $\gamma_1^1(1) = 1$, $\gamma_1^1(2) = 0$ with

$$a^1(0) = P^{\gamma_1^1}(1) \, a^1(1) = \begin{bmatrix} 0 \\ k \end{bmatrix}.$$

The region $R_1(0)$ is defined by the inequalities

$$\pi a^1(0) \leq P^{\gamma_1^1}(1) a(1)$$

for all $\gamma_1 \in \Gamma_1$, $a(1) \in \tilde{A}_1$. A computation gives

$$R_1(0) = \{\pi : k\pi_2 \leq \pi_1\}$$

with

$$a^2(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and $\gamma_1^2 \equiv 0$. Since it is easily verified that

$$R_2(0) = \{\pi : k\pi_2 \geq \pi_1\}$$

so that $\Pi_0 = R_1(0) \cup R_2(0)$, the algorithm terminates with

$$\tilde{A}_0 = \left\{ \begin{bmatrix} 0 \\ k \end{bmatrix} , \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} .$$

The situation at $t = 0$ is illustrated by Figure 4.3.1.

Since $\pi_0 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \in R_1(0)$ for $k \leq 1$, γ_0^1 is an optimal

control. Therefore,

$$\pi(1) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix} .$$

Since $\pi(1) \in R_1(1)$, γ_2^1 is the optimal control for $t = 2$. This is, of course, in agreement with the earlier computations.
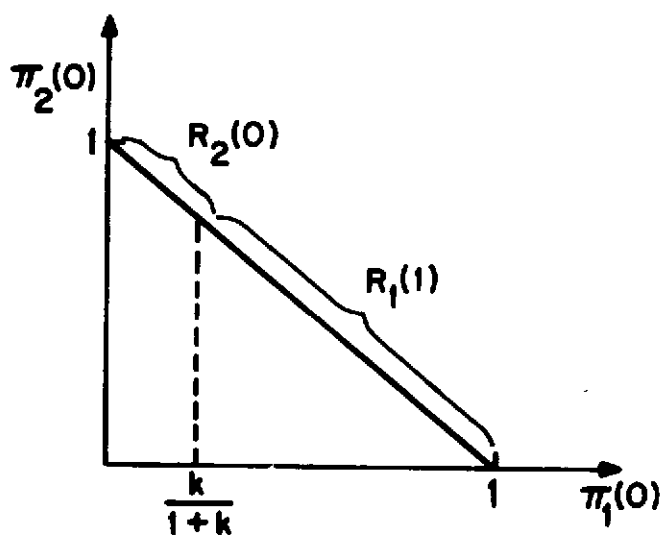
Figure 4.3.1   The Subsets $R^1(0), R^2(0)$ of Π_0.

CHAPTER V

THE INFINITE HORIZON FSFM PROBLEM

In this chapter, time-invariant FSFM models operating over

the infinite time horizon are studied. The infinite horizon model

provides a useful approximation to problems with a finite, but

distant and possibly unknown planning horizon.

The cost criterion studied is the expected discounted cost.

For this criterion, the Value and Policy Iteration methods of

Howard [How1] and Blackwell [B11] can be extended to the FSFM

problem. Moreover, the algorithms of Sondik [So1] implementing

these methods can also be extended to the FSFM problem.

The chapter concludes with an example illustrating the solution

of a simple FSFM problem by the Policy Iteration method. An

important conclusion that can be drawn from the example is

that the optimal control law sequence for an infinite horizon

FSFM problem will be non-stationary in general.

5.1 Formulation

In this chapter, attention is restricted to time invariant FSFM models of the form

$$x(t) = f(x(t-1), u(t), q(t)) \qquad (5.1.1)$$

defined for $t = 1, 2, \ldots$ where the state sets $X_t \equiv X_\infty$, the control sets $U_t \equiv U_\infty$, and the uncertainty sets $Q_t \equiv Q_\infty$ are all independent of time. Moreover, the probability function $p_t \equiv p_\infty$ on $Q_t \equiv Q_\infty$ is assumed time invariant, and the sets $\Gamma_t \equiv \Gamma_\infty$ of admissible control functions are assumed constant.

Let the sets X and Γ be defined as follows

$$X = X_0 \times X_1 \times X_2 \times \ldots \qquad (5.1.2)$$

$$\Gamma = \Gamma_1 \times \Gamma_2 \times \Gamma_3 \times \ldots \qquad (5.1.3)$$

For each $\gamma = (\gamma_1, \gamma_2, \gamma_3, \ldots) \in \Gamma$, a sequence of matrices P^{γ_t} is defined by

$$P_{ij}^{\gamma_t} = p_\infty(\{q : j = f(i, \gamma_t(i), q)\}). \qquad (5.1.4)$$

The matrix P^{γ_t} can be interpreted as a transition probability. That is, for each $i \in X_{t-1}$, a probability measure on X_t is defined with $P_{ij}^{\gamma_t}$ equal to the probability of $j \in X_t$. Since π_0 defines a probability measure on X_0, the theorem of Tulcea [Lo1] can be invoked to establish the existence of a unique probability measure ν^γ on $(X, \mathcal{X})$ $(\mathcal{X} = P(X))$ satisfying

$$\nu^\gamma(A_0 \times A_1 \times \ldots \times A_t \times X_{t+1} \times X_{t+2} \times \ldots)$$

$$= \sum_{i_0 \in A_0} \pi_{i_0}(0) \sum_{i_1 \in A_1} P_{i_0 i_1}^{\gamma_1} \sum_{i_2 \in A_2} P_{i_1 i_2}^{\gamma_2} \ldots \sum_{i_t \in A_t} P_{i_{t-1} i_t}^{\gamma_t} .$$

$$(5.1.5)$$

In particular, from (5.1.5) it follows that

$$\nu(X_0 \times X_1 \times \ldots \times X_{t-1} \times A_1 \times X_{t+1} \times X_{t+2} \times \ldots)$$

$$= \sum_{i_t \in A_t} \pi_{i_t}^\gamma(t) , \qquad (5.1.6)$$

where the sequence $\pi^\gamma(t)$ is defined inductively by

$$\pi^\gamma(0) = 0 , \qquad (5.1.7)$$

$$\pi^\gamma(t) = \pi^\gamma(t-1) P^{\gamma_t} , \quad t = 1,2,\ldots, \qquad (5.1.8)$$

where $\gamma = (\gamma_1, \gamma_2, \ldots)$.

Defining the cost J of operating the system is a delicate problem. One approach might be to define

$$J = \sum_{t=1}^{\infty} h(x(t-1), u(t)) . \qquad (5.1.9)$$

This approach suffers from several defects. If $h(x(t-1), u(t)) > 0$ for all $x(t-1) \in X_{t-1}$, $u(t) \in U_t$, then $J = +\infty$ for every control, and

the cost criterion is useless. If $h(x(t-1),u(t)) \geq 0$ and

$h(x(t-1),u(t)) = 0$ for some $x(t-1)$ ε X_{t-1}, $u(t)$ ε U_t, then the

cost criterion is still infinite useless the non-zero cost states

occur only finitely often. This case might be of interest in some

situations, but is clearly rather special. Similar comments apply

if the direction of the preceeding inequalities is reversed. If

the function h is allowed to assume both positive and negative values,

then there is no assurance that the summation in (5.1.9) is well

defined $\left(\displaystyle\sum_{t=1}^{\infty} (-1)^t = ? \right)$.

A second approach is the definition

$$J = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} h(x(t-1),u(t)) \qquad (5.1.10)$$

of the average cost per unit time. This cost is never

infinite; if

$$k = \sup_{x \varepsilon X_\infty} \sup_{u \varepsilon U_\infty} |h(x,u)| \qquad (5.1.11)$$

then

$$|J| \leq k. \qquad (5.1.12)$$

However, J need not be defined for all sequences $x(t-1)$, $u(t)$. Suppose

that $X_\infty = \{0,1\}$, $h(0) = 0$ and $h(1) = 1$ (independent of u).

Then J is not defined for the sequence

$$x(0) = 0 \qquad (5.1.13)$$

$$x(t) = \begin{cases} 1 & s_{2n-1} \leq t < s_{2n} \\ 0 & s_{2n} \leq t < s_{2n+1} \end{cases} \qquad (5.1.14)$$

where $n = 1,2,3,\ldots$ and the sequence s_n is defined by

$$s_1 = 1 \qquad (5.1.15)$$

$$s_{n+1} = (n+1) \sum_{i=1}^{n} s_i \;, \; n = 1,2,\ldots \qquad (5.1.16)$$

For this sequence,

$$\lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=1}^{T} h(x(t-1),u(t)) = 1 \qquad (5.1.17)$$

but

$$\lim_{T \to \infty} \inf \frac{1}{T} \sum_{t=1}^{T} h(x(t-1),u(t)) = 0. \qquad (5.1.18)$$

A third approach is the definition

$$J = \sum_{t=1}^{\infty} \beta^{t-1} h(x(t-1),u(t)) \qquad (5.1.19)$$

where $0 < \beta < 1$ is the <u>discount rate</u>. Discount factors occur naturally in an economic context when the present value of a stream of future earnings must be determined [An1]. In other contexts, β can be regarded simply as a convergence factor used to achieve an approximation of the average cost (5.1.10).[1] Existence of J is assured. Let k be defined as in (5.1.11), then

---

[1] Ross [Ro1,Ro2] and Mine and Osaki [Mi1] consider the limit as $\beta \to 1$.

$$\sum_{t=1}^{\infty} \beta^t \, |h(x(t-1),u(t)| \leq k \sum_{t=1}^{\infty} \beta^{t-1} = k \, \frac{1}{1-\beta} \qquad (5.1.20)$$

so that the series defining J converges for any sequence of states and controls.

The definition (5.1.19) will be adopted in this chapter, although some further remarks concerning (5.1.10) will be made.

It is necessary to define the expected value of J to state the infinite horizon version of the FSFM problem. By (5.1.20), the summation

$$J_\gamma = \sum_{t=1}^{\infty} \beta^{t-1} h(x(t-1),\gamma_t(x(t-1))) \qquad (5.1.21)$$

exists for all

$$\gamma = (\gamma_1,\gamma_2,\dots) \in \Gamma$$

and sequences

$$(x(0),x(1),x(2),\dots) \in X.$$

Therefore (5.1.21) defines a map

$$J_\gamma : X \to R \qquad (5.1.22)$$

that is automatically measurable with respect to $X = P(X)$. Moreover,

$$E_\gamma |J_\gamma| = \int_X |J_\gamma(x)| \, dv^\gamma(x)$$

$$\leq \int_X \frac{k}{1-\beta} \, dv^\gamma(x) = \frac{k}{1-\beta} \qquad (5.1.23)$$

so that J_γ is integrable. Therefore, the functional

$$J(\gamma) = E_{\gamma}J_{\gamma} = \int_X J_{\gamma}(x)\, d\nu^{\gamma}(x) \qquad\qquad (5.1.24)$$

can be defined. The <u>infinite horizon FSFM problem with discounted</u>

<u>cost criterion</u> is then to minimize $J(\gamma)$, for all $\gamma \in \Gamma$.

Since the sum (5.1.21) is defined for all

$$(x(0), x(1), x(2), \ldots) \in X,$$

and since the bound (5.1.22) holds, application of Lebesque's dominated

convergence theorem (see, e.g., [Rul] or [Sel])

shows that

$$E_{\gamma}\left\{ \sum_{t=1}^{\infty} \beta^{t-1} E_{\gamma}\ h(x(t-1), \gamma_t(x(t-1))) \right\}$$

$$= \sum_{t=1}^{\infty} \beta^{t-1} E_{\gamma}\left\{ h(x(t-1), \gamma_t(x(t-1))) \right\}$$

$$= \sum_{t=1}^{\infty} \beta^{t-1} \pi(t-1) h^{\gamma_t} \qquad\qquad (5.1.25)$$

for all control law sequences $\gamma = (\gamma_1, \gamma_2, \ldots)$ (possibly non-stationary)

where

$$\pi(0) = \pi_0\ , \qquad\qquad (5.1.26)$$

$$\pi(t) = \pi(t-1)\ P^{\gamma_t}\ ,\ t = 1,2,\ldots \qquad\qquad (5.1.27)$$

Therefore the infinite horizon FSFM problem with discounting is

equivalent to the deterministic problem of minimizing (5.1.25) subject

to (5.1.26), (5.1.27).

To apply the method of dynamic programming, the problem defined
by equations (5.1.25) - (5.1.27) is imbedded in a series of similar
problems. Define functions

$$V_t(\pi) = \inf_{\gamma \epsilon \Gamma} \sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi^\gamma(\tau-1) h^{\gamma_{\tau+1}} \tag{5.1.28}$$

for $t \geq 1$, where $\gamma = (\gamma_1, \gamma_2, \ldots)$ and

$$\pi^\gamma(t-1) = \pi, \tag{5.1.29}$$

$$\pi^\gamma(\tau) = \pi^\gamma(\tau-1) P^{\gamma_\tau} \quad , \quad \tau \geq t. \tag{5.1.30}$$

(Notice that the summation in (5.1.28) is independent of the first
$t-1$ components of γ.)

Lemma 5.1.1

For all $s, t \geq 0$, for all $\pi \epsilon \Pi_t \equiv \Pi_\infty$,

$$V(\pi) \overset{\Delta}{=} V_s(\pi) = V_t(\pi) \tag{5.1.31}$$

Proof

Suppose that $t \geq s$. Then the control law sequences

$$\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_{t-1}, \gamma_s, \gamma_{s+1}, \ldots)$$

and

$$\gamma = (\gamma_1, \ldots, \gamma_s, \gamma_{s+1}, \ldots)$$

satisfy

$$\sum_{\tau=t}^{\infty} \beta^{\tau-t} \; \pi^{\hat{\gamma}}(\tau-1)h^{\hat{\gamma}\tau} = \sum_{\tau=s}^{\infty} \beta^{\tau-s} \; \pi^{\gamma}(\tau-1)h^{\gamma\tau} . \qquad (5.1.32)$$

Therefore, the infimum of the above sums must be equal.

Lemma 5.1.2

The functions $V_t(\cdot)$ satisfy the sequence of optimality equations

$$V_t(\pi) = \min_{\gamma_{t+1}\epsilon\Gamma_{t+1}} \left\{ \pi h^{\gamma_{t+1}} + \beta V_{t+1}(\pi P^{\gamma_{t+1}}) \right\} . \qquad (5.1.33)$$

Proof

$$V_t(\pi) = \inf_{\gamma\epsilon\Gamma} \; \sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi^{\gamma}(\tau-1)h^{\gamma\tau}$$

$$= \inf_{\gamma\epsilon\Gamma} \left\{ \pi h^{\gamma_{t+1}} + \beta^{\tau-t-1} \pi^{\gamma}(t-1)h^{\gamma\tau} \right\}$$

$$= \min_{\gamma_{t+1}\epsilon\Gamma_{t+1}} \left\{ \pi h^{\gamma_{t+1}} + \beta \inf_{\gamma_{t+2}\epsilon\Gamma_{t+2}, \gamma_{t+3}\epsilon\Gamma_{t+3}...} \sum_{\tau=t+1}^{\infty} \beta^{\tau-t-1} \pi^{\gamma}(\tau-1)h^{\gamma\tau} \right\}$$

$$= \min_{\gamma_{t+1}\epsilon\Gamma_{t+1}} \left\{ \pi h^{\gamma_{t+1}} + \beta V_{t+1}(\pi P^{\gamma_{t+1}}) \right\} \qquad (5.1.34)$$

Lemma 5.1.3

The function $V(\cdot)$ is the unique bounded solution of the equation

$$V(\pi) = \min_{\gamma_\infty \varepsilon \Gamma_\infty} \left\{ \pi h^{\gamma_\infty} + \beta V(\pi P^{\gamma_\infty}) \right\} \tag{5.1.35}$$

for t arbitrary.

Proof

By Lemmas 5.1.1 and 5.1.2, $V(\cdot)$ is a solution of (5.1.35). Since

$$\left| \sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi^\gamma (\tau-1) h^{\gamma_{\tau+1}} \right|$$

$$\leq \sum_{\tau=t}^{\infty} \beta^{\tau-t} \left| \pi^\gamma (\tau-1) h^{\gamma_{\tau+1}} \right|$$

$$\leq \sum_{\tau=t}^{\infty} \beta^{\tau-t} k = \frac{k}{1-\beta} , \tag{5.1.36}$$

it follows that $V(\cdot)$ is bounded. Therefore, it is enough to show that (5.1.35) has a unique bounded solution.

Let $B(\Pi_\infty)$ be the set of bounded real-valued functions on Π_∞. With the norm

$$||V_1 - V_2|| = \sup_{\pi \varepsilon \Pi_\infty} |V_1(\pi) - V_2(\pi)| , \tag{5.1.37}$$

$B(\Pi_\infty)$ is a Banach space [Ru1]. Define the operation

$$\Lambda : \quad B(\Pi_\infty) \to B(\Pi_\infty)$$
$$V \quad \to \Lambda V \tag{5.1.38}$$

by

$$\Lambda V(\pi) = \min_{\gamma_\infty \epsilon \Gamma_\infty} \left\{ \pi h^{\gamma_\infty} + \beta V(\pi P^{\gamma_\infty}) \right\} .$$ (5.1.39)

Then the problem is to show that the equation

$$\Lambda V = V$$ (5.1.40)

has a unique solution.

To show that (5.1.40) has a unique bounded solution it is sufficient by the Contraction Mapping Theorem for Banach spaces [Si1] to show that Λ is a contraction. But Λ is a contraction by the following argument due to Blackwell [B11]. Notice that

$$\Lambda(V + c) = \Lambda V + \beta c$$ (5.1.41)

for any constant c, and that $V_1 \le V_2$ [1] implies

$$\Lambda V_1 \le \Lambda V_2$$ (5.1.42)

Let $c = ||V_1 - V_2||$, then

$$V_1 \le V_2 + || V_1 - V_2||$$ (5.1.43)

implies that

$$\Lambda V_1 \le \Lambda(V_2 + || V_1 - V_2||) = \Lambda V_2 + \beta ||V_1 - V_2||.$$

By a symmetrical argument,

$$\Lambda V_2 \le \Lambda V_1 + \beta ||V_1 - V_2||$$ (5.1.44)

---

[1] i.e., $V_1(\pi) \le V_2(\pi)$ for all $\pi \epsilon \Pi_\infty$.

so that

$$||\Lambda v_1 - \Lambda v_2|| \leq \beta \, ||v_1 - v_2||. \qquad (5.1.45)$$

Theorem 5.1.4

Let the map $\delta^* : \Pi_\infty \rightarrow \Gamma_\infty$ be defined for $t = 0,1,2,\ldots$ by $\delta^*(\pi) = \gamma_\infty$ if γ_∞ is a minimizing control law in (5.1.35). Define the quantities $\pi^*(t)$, γ_t^* by

$$\pi^*(0) = \pi_0, \qquad (5.1.46)$$

$$\gamma_t^* = \delta^*(\pi^*(t-1)) \ , \quad t \geq 1, \qquad (5.1.47)$$

$$\pi^*(t) = \pi^*(t-1)P^{\gamma_t^*} \ , \quad t \geq 1. \qquad (5.1.48)$$

Then $\gamma^* = (\gamma_1^*, \gamma_2^*, \ldots)$ is an optimal control law sequence with corresponding sequence of optimal states.

Proof

By (5.1.33) and the definition of γ_{t+1}^* ,

$$V_t(\pi^*(t)) = \min_{\gamma_{t+1} \in \Gamma_{t+1}} \left\{ \pi^*(t)h^{\gamma_{t+1}} + \beta V_{t+1}(\pi^*(t)P^{\gamma_{t+1}}) \right\}$$

$$= \pi^*(t)h^{\gamma_{t+1}^*} + \beta V_{t+1}(\pi^*(t+1)) \qquad (5.1.49)$$

for $t = 1,2,\ldots$ Therefore, a simple induction argument establishes that

$$V_0(\pi_0) = \sum_{\tau=1}^{t+1} \beta^{\tau-1} \pi^*(\tau-1)h^{\gamma_\tau^*} + \beta^{t+1} V_{t+1}(\pi^*(t+1)) \qquad (5.1.50)$$

for all t. Recall from (5.1.36) that $V_{t+1}(\pi*(t+1))$ is bounded,

$$V_{t+1}(\pi*(t+1) \leq \frac{k}{1-\beta} \,. \qquad (5.1.51)$$

Therefore,

$$V_0(\pi_0) = \lim_{\tau \to \infty} \left[\sum_{\tau=1}^{t+1} \beta^{\tau-1} \pi*(\tau-1) h^{\gamma_\tau^*} + \beta^{t+1} V_{t+1}(\pi*(t+1)) \right]$$

$$= \sum_{\tau=1}^{\infty} \beta^{\tau-1} \pi*(\tau-1) h^{\gamma_\tau^*} \qquad (5.1.52)$$

which was to be shown.

To employ Theorem 5.1.4, it is necessary to compute the function V. Since Λ is a contraction, V can be computed iteratively by a successive approximation approach. Define the sequence $V_n \in B(\pi)$, $n = 0,1,2,\ldots$, by

$$V_{n+1} = \Lambda V_n \qquad (5.1.53)$$

where V_0 is an arbitrary element of $B(\pi)$. Notice that the notation is ambiguous since it does not distinguish between the forward sequence V_t and the backward sequence V_n. V_n is often referred to as the "cost with n periods remaining" in the literature on Markovian decision procedures [Sol], but this is not meaningful strictly, because there is never a finite number of periods remaining in an infinite horizon problem. Similarly, the statement that the sequence of control laws with n periods remaining is to be determined is not logical. What control law will be used at the first stage?

Lemma 5.1.5

The sequence V_n defined above converges to V. Moreover,

(i) $||\Lambda^n V_0 - V|| \leq \beta^n ||V_0 - V||$, (5.1.54)

(ii) $||\Lambda^n V_0 - V|| \leq \dfrac{\beta^n}{1-\beta} ||\Lambda V_0 - V_0||$. (5.1.55)

Proof

Since Λ is a contraction (Lemma 5.1.3), Lemma 5.1.5 is an immediate consequence of Theorem 1.XVI of Kantorovich and Akilov [Kal].

Notice that (5.1.55) gives a bound on the error at iteration number n that is independent of V and can therefore be precomputed. The method of determining the optimal control law by the iteration (5.1.53) is referred to as Value Iteration [Howl] in the literature on Markovian decision processes. This is in distinction to the method of Policy Iteration introduced by Howard [Howl] and extended by Blackwell [Bll]. The policy iteration method can also be extended to the FSFM problem, as will be demonstrated next.

For any policy $\delta : \Pi_\infty \to \Gamma_\infty$, define the sequence of functions $V_n^\delta : \Pi_\infty \to R$ by

$$V_0^\delta (\pi) \equiv 0 \tag{5.1.56}$$

$$V_{n+1}^{\delta(\pi)} = \pi h^{\delta(\pi)} + V_n^\delta (\pi P^{\delta(\pi)}) \tag{5.1.57}$$

for n = 0,1,2,... Let the operator

$$\Lambda^\delta : B(\Pi_\infty) \rightarrow B(\Pi_\infty)$$

be defined by

$$(\Lambda^\delta V)(\pi) = \pi h^{\delta(\pi)} + V(\pi P^{\delta(\pi)}) . \qquad (5.1.58)$$

Lemma 5.1.6.

For any policy δ, the operator Λ^δ is monotone; i.e., if

$V_1 \leq V_2$, then $\Lambda^\delta V_1 \leq \Lambda^\delta V_2$

Proof

The property follows immediately from (5.1.58).

Theorem 5.1.7 (Howard Policy Iteration)

Let the policy $\hat\delta$ be defined by $\hat\delta(\pi) = \hat\gamma$, where $\hat\gamma$ satisfies

$$\pi h^{\hat\gamma} + \beta V^\delta(\pi P^{\hat\gamma}) = \min_{\gamma_t \in \Gamma_t} \pi h^{\gamma_t} + \beta V^\delta(\pi P^\gamma) \qquad (5.1.59)$$

for some arbitrary policy δ. Then

$$V^{\hat\delta}(\pi) \leq V^\delta(\pi) \qquad (5.1.60)$$

for all $\pi \in \Pi_\infty$. $(V^\delta \overset{\Delta}{=} \lim_{n \to \infty} V_n^\delta$ exists since Λ^δ is a contraction.)

Proof

The method of proof is due to Blackwell [B11].

By (5.1.57)

$$\Lambda^{\hat{\delta}} \, v^{\delta} \leq \Lambda^{\delta} v^{\delta} = v^{\delta} \, . \tag{5.1.61}$$

Since $\Lambda^{\hat{\delta}}$ is monotone,

$$\Lambda^{\hat{\delta}} \Lambda^{\hat{\delta}} \, v^{\delta} \leq \Lambda^{\hat{\delta}} \, v^{\delta} \leq v^{\delta} \tag{5.1.62}$$

and so by induction

$$(\Lambda^{\hat{\delta}})^{n} \, v^{\delta} \leq v^{\delta} \, . \tag{5.1.63}$$

Since $\Lambda^{\hat{\delta}}$ is a contraction (with modulus β),

$$\lim_{n \to \infty} (\Lambda^{\hat{\delta}})^{n} \, v^{\delta} = v^{\hat{\delta}} \leq v^{\delta} \, . \tag{5.1.64}$$

Before turning to the important question of using Value or Policy Iteration as a numerical technique, it is appropriate to comment on the alternative definition (5.1.10) of J.

As mentioned earlier, J may not be well defined by (5.1.10), so that the formal computation

$$E_{\gamma} \left\{ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} h(x(t-1), \gamma_{t}(x(t-1))) \right\}$$

$$= \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E_{\gamma} \left\{ h(x(t-1), \gamma_{t}(x(t-1))) \right\} \tag{5.1.65}$$

is not valid in general. Moreover, the limit on the right hand side

of (5.1.65) need not exist. This has led many authors [Kul,Del ,

Mil] to adapt the definition

$$J(\gamma) = \lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=1}^{T} E_\gamma \{ h(x(t-1),\gamma_t(x(t-1))) \} .$$

(5.1.66)

The limit in (5.1.66) always exists since

$$\left| \frac{1}{T} \sum_{t=1}^{T} E_\gamma \left\{ h(x(t-1),\gamma_t(x(t-))) \right\} \right| \le k$$

(5.1.67)

for all T.

The disadvantage of the definition (5.1.66) is that it is difficult to give a meaningful interpretation to the functional $J(\gamma)$. However, if attention is restricted to the class of stationary control laws $(\gamma_t = \gamma_\infty$ for all t), then a natural interpretation of $J(\gamma)$ is available.

For stationary control laws,

$$P^{\gamma_t} = P^{\gamma_\tau} \triangleq P^\gamma$$

(5.1.68)

$$h^{\gamma_t} = h^{\gamma_\tau} \triangleq h^\gamma$$

(5.1.69)

for all $1 \le t, \tau < \infty$. Moreover, by (5.1.6)

$$E_\gamma \{h(x(t-1),\gamma_t(x(t-1)))\} = \pi^\gamma(t-1) h^\gamma$$

$$= \pi_0 (P^\gamma)^{t-1} h^\gamma.$$

(5.1.70)

But Theorem 2.1 of Doob [Do1] (Chapter 5) states that there exists a

stochastic matrix $P_\infty{}^\gamma$ such that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (P^\gamma)^t = P_\infty{}^\gamma \qquad (5.1.71)$$

Therefore,

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E_\gamma \{h(x(t-1),\gamma_t(t-1))\} = \pi^\gamma h^\gamma \qquad (5.1.72)$$

where

$$\pi^\gamma = \pi_0 P_\infty{}^\gamma \qquad (5.1.73)$$

defines the long-run distribution of the Markov chain. Thus,

$\pi^\gamma h^\gamma$ is the expected cost per transition under the long-run distribution

induced by the stationary control law sequence γ.

Although the average cost criteria has been studied extensively for

problems with finite state space [Howl, Kul, Del, Mil], there are few

results available that apply to continuous state spaces such as Π_∞.

Therefore attention is restricted in the sequel to the expected dis-

counted cost.

5.2 Numerical Solution of the Functional Equation

In the previous section, it was demonstrated that solving the infinite horizon FSFM problem with discounted cost criterion was equivalent to solving the functional equation

$$\Lambda V = V \; . \tag{5.2.1}$$

Two theoretical methods, Value Iteration and Policy Iteration, were described for the solution of (5.2.1).

The solution of functional equations similar to (5.2.1) is a classical problem of dynamic programming. Although numerous suggestions have been advanced (See [B1], [B2], [B3], for example), no satisfactory general algorithm has ever been found. However, Sondik [So1] has recently developed an algorithm utilizing policy improvement for solution of Markovian decision processes with incomplete state information. Since the functional equation that arises in the solution of these processes is similar to (5.2.1), Sondik's algorithm can be applied to the FSFM problem, as will be demonstrated in this section.

The most straightforward approach to the solution of (5.2.1) is to use Value Iteration. Starting the algorithm with $V_0 \equiv 0$, then

$$V_{n+1} = \Lambda V_n \tag{5.2.2}$$

can be obtained using the backward Sondik algorithm as in Figure 4.2.1. The difficulty with this approach is that the algorithm must be

iterated until the factor $\frac{\beta^n}{1-\beta}$ is small enough to insure that V_n approximates V to the desired accuracy. For discount rates β close to one, this requires a large number of iterations. But the Sondik algorithm is practical for only a small number of iterations, in general.

A second approach to the solution of (5.2.1) is to use Policy Iteration. Policy Iteration consists of two steps: value determination and policy improvement [Howl]. These steps are illustrated in Figure 5.2.1. One might conjecture that the Sondik algorithm could be utilized to carry out both steps. However, neither V_n^{δ} nor V^{δ} is in general piecewise linear and concave, as is required for the Sondik algorithm. These same difficulties arise in the partially observable Markovian decision problem, and Sondik has developed an alternative approach [Sol]. As will be demonstrated next, this approach can be applied to the FSFM problem also.

The value determination step of the policy iteration algorithm requires that the functional equation

$$\Lambda^{\delta} V = V \tag{5.2.3}$$

be solved to determine V . For a special class of policies , this equation can be readily solved.

Lemma 5.2.1

Suppose that X_{∞} has n elements so that the (row) vectors in Π_{∞} are n-dimensional. Then for every policy $\delta : \Pi_{\infty} \rightarrow \Gamma_{\infty}$, there exists a map
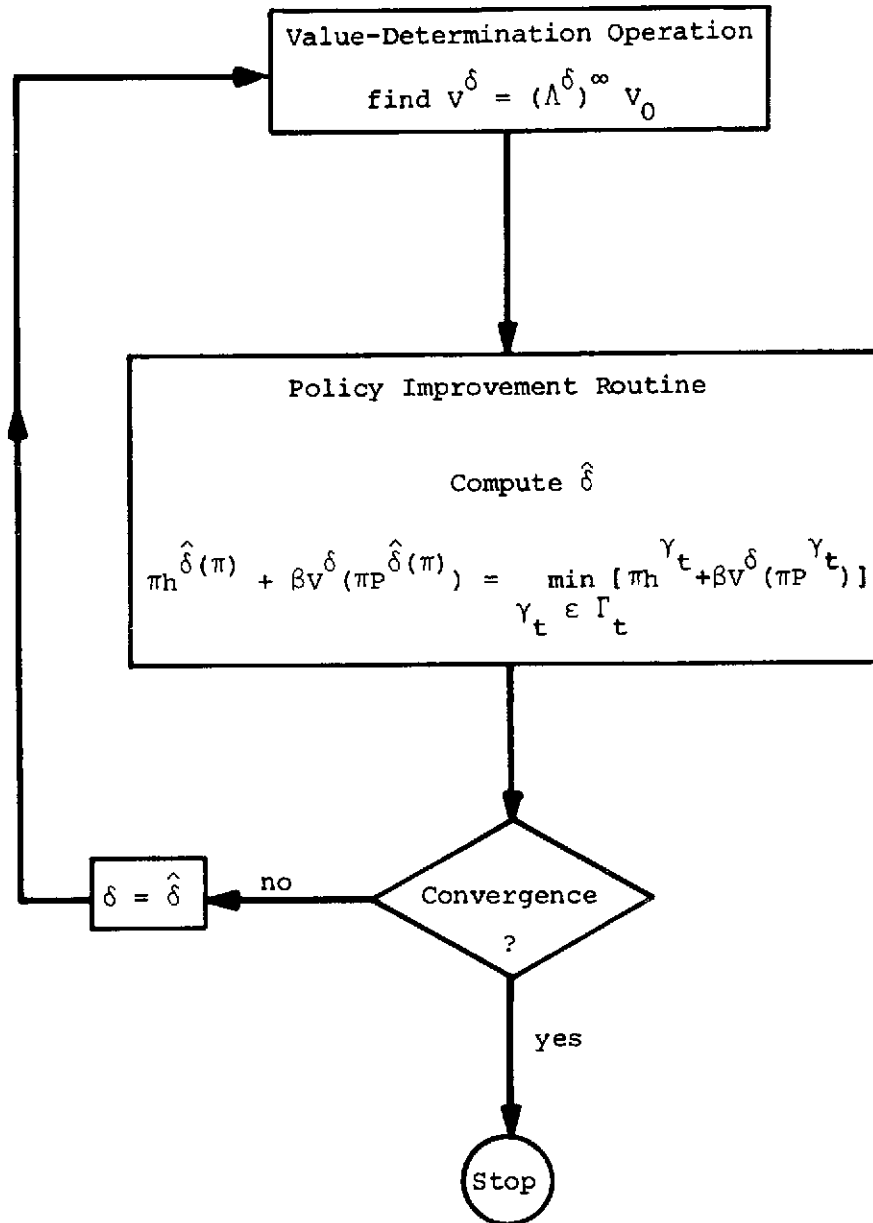
Figure 5.2.1   Policy Iteration (After [Howl], Figure 7.1)

$$a^\delta \quad : \amalg_\infty \to A \ , \tag{5.2.4}$$

where A is the set of n-dimensional column vectors, such that

$$v^\delta(\pi) = \pi a^\delta(\pi) \tag{5.2.5}$$

for all $\pi \ \epsilon \ \amalg_\infty$.

Proof

Recall that v^δ is the unique solution of the functional equation.

$$v^\delta = \Lambda^\delta v$$

$$= \pi h^{\delta(\pi)} + \beta v^\delta(\pi P^{\delta(\pi)}) \tag{5.2.6}$$

Therefore, the representation (5.2.5) is valid if the functional equation

$$a^\delta(\pi) = h^{\delta(\pi)} + \beta P^{\delta(\pi)} a^\delta(\pi) \tag{5.2.7}$$

has a solution. Recall that the space of bounded maps from $\amalg_\infty$ to A, $B(\amalg_\infty;A)$ is a Banach space with the norm

$$||a(\pi)|| = \sup_{\pi \epsilon \Pi} ||a(\pi)||_\infty$$

where

$$||a||_\infty = \max_{1 \leq i \leq n} |a_i| \ . \tag{5.2.9}$$

Moreover,

$$||h^{\delta(\cdot)} + \beta P^{\delta(\cdot)} a_1(\cdot) - h^{\delta(\cdot)} - \beta P^{\delta(\cdot)} a_2(\cdot)||$$

$$= \beta \ ||P^{\delta(\cdot)}(a_1(\cdot) - a_2(\cdot))||$$

$$\leq \beta \ ||a_1(\cdot) - a_2(\cdot)|| \qquad\qquad (5.2.10)$$

so that a unique solution to 5.2.7 exists by the contraction mapping theorem.

The solution to (5.2.7) will not necessarily be piecewise-linear and concave. Suppose, however, that there are sets $R_1, R_2, \ldots, R_j$ that partition Π_∞, and satisfy the following two conditions:

(a) for each $\pi \in R_1$, $\delta(\pi) = \gamma^i$

(b) each $\pi \in R_i$ satisfies $\pi P^{\gamma^i} \in R_{\nu(i)}$

Then (5.2.7) reduces to the system of equation

$$a_i = h^{\gamma^i} + \beta P^{\gamma^i} a_{\nu(i)} \qquad\qquad (5.2.11)$$

for $i = 1, 2, \ldots, j$. The existence and uniqueness of a solution to (5.2.11) can be established by a contraction mapping argument similar to those used previously.

Of course, the reduction of the infinite dimensional equation (5.2.7) to the finite dimensional system of equations (5.2.11) is predicated on the assumption of an appropriate partition of Π_∞. Sondik has found a class of policies for which this assumption is valid [Sol]. To define this class, it is necessary to define some notion. Let

$$T_\delta(A) = \left\{ \pi P^{\delta(\pi)} : \pi \in A \right\}, \qquad\qquad (5.2.12)$$

$$S_\delta{}^n = T_\delta(S^{n-1}) \quad , \quad n \geq 1 \quad , \tag{5.2.13}$$

$$S_\delta{}^0 = \Pi_0 \quad , \tag{5.2.14}$$

$$D_\delta = \text{closure } \{\pi : \delta \text{ is discontinuous at } \pi \}. \tag{5.2.15}$$

A policy δ is <u>finitely transient</u> if and only if there is an integer m such that

$$D_\delta \cap S_\delta{}^m = \phi \tag{5.2.16}$$

(ϕ = null set).

In his thesis [Sol], Sondik establishes two important properties of finitely transient policies:

(a)   A partition $R_1, \ldots, R_j$ with the desired properties

exists if δ is <u>finitely transient</u>, and

(b)   The function V^δ may be approximated arbitrarily closely

by $V^{\hat{\delta}}$, for some finitely transient policy $\hat{\delta}$.

Since Sondik's arguments are readily applied to the FSFM problem, they are not repeated here. However, an example (Figure 5.2.2) is given to illustrate the basic idea.

For the example,
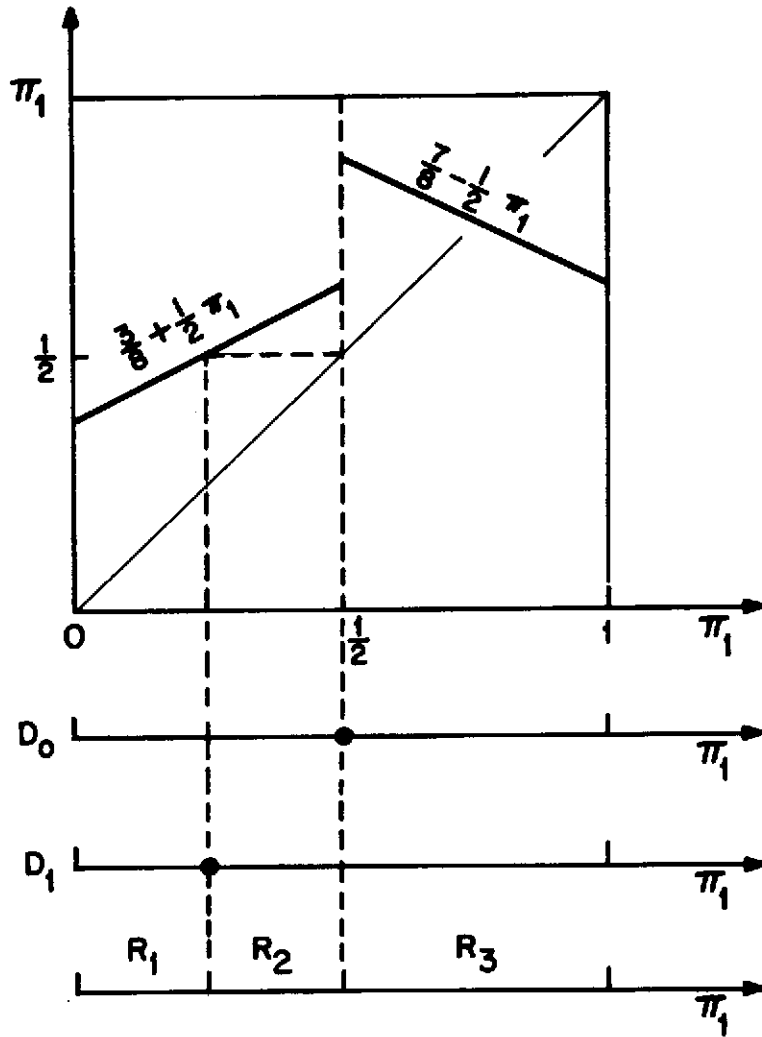
Figure 5.2.2   Construction of the Partition R_1, R_2, R_3.

$$P^{\delta(\pi)} = \begin{cases} \begin{bmatrix} \dfrac{7}{8} & \dfrac{1}{8} \\[2mm] \dfrac{3}{8} & \dfrac{5}{8} \end{bmatrix} = P^{\gamma} & \pi_1 < \dfrac{1}{2} \\[8mm] \begin{bmatrix} \dfrac{3}{8} & \dfrac{5}{8} \\[2mm] \dfrac{7}{8} & \dfrac{1}{8} \end{bmatrix} = P^{\tilde{\gamma}} & \pi_1 \geq \dfrac{1}{2} \end{cases}$$

Notice that $P^{\delta(\pi)}$ is completely characterized by its effect on π_1, since $\pi_2 = 1 - \pi_1$. The sets D^n are defined by

$$D^0 = D_\delta \tag{5.2.17}$$

$$D^{n+1} = \{\pi : \pi P^{\delta(\pi)} \in D^n\}, \quad n \geq 0. \tag{5.2.18}$$

Since $\pi P^{\delta(\pi)} \in R_2$ for $\pi \in R_1$, $\pi P^{\delta(\pi)} \in R_3$ for $\pi \in R_2$, and $\pi P^{\delta(\pi)} \in R^3$ for $\pi \in R^3$, it follows that $\nu(1) = 2$, $\nu(2) = 3$, and $\nu(3) = 3$.

The procedure of constructing the regions R_i from the sets D^n applies to any finitely transient policy. Moreover, an approximate partition can be constructed for an arbitrary policy. The reader is referred to Sondik [Sol] for details.

Although the value determination operation of Policy Iteration can be (approximately) carried out with the use of finitely transient policies, there remains the problem of implementing the policy improvement routine. The difficulty here is that the function V^δ is not

piecewise linear or concave as is required for the Sondik algorithm. But V^δ will be piecewise linear if δ is finitely transient, and Sondik has shown that the concave hull of V^δ can be used in the policy improvement routine [Sol]. Again, Sondik's thesis [Sol] should be consulted for details.

In this section, the numerical solution of the functional equation (5.2.1) has been considered. The emphasis has been on showing that an algorithm recently developed by Sondik for partially observable Markovian decision processes can be adapted for the FSFM problem. It should be pointed out that a FSFM problem will be considerably more difficult to solve than a corresponding partially observed Markovian process. This is due to the fact that the FSFM problem requires that the observation and memory sets be included in the state set, and since the policies in the FSFM problem assign a control law to each state. Thus, a given FSFM problem will be much larger than an analagous partially observed Markovian decision process. Although this advantage is offset somewhat by the simple form of (5.2.1) (relative to the partially observed problem), it is nevertheless true that the technique outlined in this section is feasible only for simple special cases of the FSFM problem.

5.3 Example

In this section, a simple infinite horizon FSFM problem with discounted cost criterion is solved using the policy iteration algorithm outlined in the previous section. The solution illustrates that in contradistinction to the usual discounted infinite horizon Markovian decision process, the optimal control law sequence can be non-stationary.

The problem has $X_\infty = U_\infty = \{1,2\}$ and Γ_∞ contains only the control law whose value is always 1, and the control law whose value is always 2. The parameters of the problem are

$$
P^1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}
\qquad
P^2 = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}
$$

$$
\beta = \frac{1}{2} \qquad h^1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad h^2 = \begin{bmatrix} \frac{3}{32} \\ \frac{35}{32} \end{bmatrix}
$$

Take the starting guess $\delta(\pi) \equiv 1$. Since δ has no discontinuities, it is certainly finitely transient. The algorithm is carried out below.

Policy Evaluation

Since δ has no discontinuities,

$$
v^\delta = \pi a^1
$$

where

$$a^1 = h^1 + \beta P^1 a^1$$

$$\begin{bmatrix} a_1^{\ 1} \\ \\ a_2^{\ 1} \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} a_1^{\ 1} \\ \\ a_2^{\ 1} \end{bmatrix}$$

$$4a_1 = a_1 + a_2$$

$$4a_2 = 4 + a_1 + a_2$$

Therefore,
$$a^1 = \begin{bmatrix} \frac{1}{2} \\ \\ \frac{3}{2} \end{bmatrix} .$$

Policy Improvement

Let $\hat{\delta}(\pi) = \hat{\gamma}$, if

$$\pi h^{\hat{\gamma}} + \beta v^{\delta}(\pi P^{\hat{\gamma}}) = \min_{\gamma} \pi h^{\gamma} + \beta v^{\delta}(\pi P^{\gamma}).$$

Since $v^{\delta}(\pi) = \pi a_1$, this is equivalent to

$$\pi h^{\hat{\gamma}} + \beta \pi P^{\hat{\gamma}} a^1 = \min_{\gamma} \pi h^{\gamma} + \beta \pi P^{\gamma} a^1$$

Since Γ_{∞} has only two elements, it is only necessary to check

$$\pi h^1 + \beta \pi P^1 a^1 \leq \pi h^2 + \beta \pi P^2 a^2$$

$$h^2 + \beta P^2 a^2 = \begin{bmatrix} \dfrac{3}{32} \\[2mm] \dfrac{35}{32} \end{bmatrix} + \dfrac{1}{2} \begin{bmatrix} \dfrac{3}{4} & \dfrac{1}{4} \\[2mm] \dfrac{1}{2} & \dfrac{1}{2} \end{bmatrix} \begin{bmatrix} \dfrac{1}{2} \\[2mm] \dfrac{3}{2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{15}{32} \\[2mm] \dfrac{51}{32} \end{bmatrix} \triangleq a^2$$

$$\pi a^1 \leq \pi a^2$$

$$[\pi_1 \ (1-\pi_1)] \begin{bmatrix} \dfrac{1}{2} \\[2mm] \dfrac{3}{2} \end{bmatrix} \leq [\pi_1 \ (1-\pi_1)] \begin{bmatrix} \dfrac{15}{32} \\[2mm] \dfrac{51}{32} \end{bmatrix}$$

$$\dfrac{4}{32} \pi_1 \leq \dfrac{3}{32}$$

$$\pi_1 \leq \dfrac{3}{4}$$

Therefore,

$$\hat{\delta}(\pi) = \begin{cases} 1 & \pi_1 \leq \dfrac{3}{4} \quad \pi \ \epsilon \ R_1 \\[3mm] 2 & \pi_1 > \dfrac{3}{4} \quad \pi \ \epsilon \ R_2 \end{cases}$$

Test for Convergence

To test for convergence, it is necessary to check

$$v^{\hat{\delta}}(\pi) = \min_{\gamma\epsilon\Gamma} \pi h^{\gamma} + \beta v^{\delta}(\pi P^{\gamma})$$

for all $\pi\epsilon\Pi$.[1]

__Case 1__ $\pi \epsilon R_1$

$$v^{\hat{\delta}}(\pi) = \pi a^1$$

But $\pi a^1 \leq \pi h^2 + \beta\pi P^2 a^1$

for $\pi \epsilon R_1$ by the computation above.

__Case 2__ $\pi \epsilon R_2$

$$v^{\hat{\delta}}(\pi) = \pi a^2$$

But $\pi a^2 \leq \pi h^1 + \beta\pi P^1 a^1$ as above.

Therefore the policy iteration algorithm has converged in a single

step.

Notice if $\pi(0) = (1\ 0)$, then $\delta(\pi(0)) = 2$. However, $\pi(1) = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}$,

so that $\delta(\pi(1)) = 1$. Moreover, $\pi(2) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \pi(t)$ for all $t \geq 2$, so

that $\delta(\pi(t)) = 1$ for all $t \geq 2$. __Thus, the optimal control law__

__sequence is non-stationary!__

---

[1]In general, a criterion such as $\sup_{\pi\epsilon\Pi_{\infty}} \mid v^{\hat{\delta}}(\pi) - \min_{\gamma\epsilon\Gamma} \pi h^{\gamma} + \beta v^{\hat{\delta}}(\pi P^{\gamma}) \mid \leq \epsilon$

would be checked.

CHAPTER VI

EXAMPLE: HYPOTHESIS TESTING WITH 1-BIT MEMORY

In this chapter, a problem of sequential hypothesis testing with a 1-bit memory is considered. The problem is not of enormous intrinsic interest, although substantial work has appeared in literature [Col,Hel, F11, Cha1, Co2, Hil]. However, the problem illustrates the use and limitations of control-theoretic methods in the design of information-handling systems.

6.1 Introduction

In the first five chapters of this thesis, some of the most important theoretical and algorithmic results of modern optimal control theory have been applied to the FSFM problem. It has been pointed out that some of the crucial memory management and communication tasks of information handling systems can be examined within this format. It is felt that the establishment of this framework is a contribution of this research. In the previous chapters, simple examples have been given to illustrate the use and properties of the theorems and algorithms derived. In this chapter, a more substantial example is studied.

Of course, the FSFM framework is not the only way that information handling problems can be treated. In particular, both information theory [Sh1] and the theory of formal languages [Hop1] deal with this important issue.

The relationship between information theory and non-classical

stochastic control theory (which includes the FSFM problem) is clarified

by the following statement of Witsenhausen [W3]:

"The latter [information theory] deals with an essentially simpler

problem, because the <u>transmission</u> of information is considered independently

of its <u>use</u>, long periods of use of a transmission channel are assumed,

and delays are ignored".[1]

Thus in information theory, one does not usually pose the question: "What is

the best code of block length n for a given source and channel?". Instead,

one asks "For a given source and information channel, what is the best

that a code of block length n can do?". Of course, the bounds obtained

as the answer to the latter question throw considerable light on the

former, and thus information theory has had considerable practical

impact. Moreover, obtaining an answer to the first question seems

computationally impossible. But the first question is still of

importance, and it is of interest to examine a framework in which the

question can be raised, even if it can't be answered. The hypothesis-

testing problem considered in this section is of the same nature as the

first question, but is considerably simpler. Thus the problem of this

chapter is studied for paradigmatic rather than pragmatic purposes.

The theory of formal languages deals with more qualitative questions

then the quantitative optimization considered in the FSFM problem. Thus

a typical question in formal language theory is "What class of languages

is accepted by a particular class of finite state machines?". This

---

[1]An exception to the last point is the recent paper of Krich and Berger
[Kr1].

question turns out to be intimately associated with the problem considered

in this chapter.

6.2 Formulation

Suppose that $\{x_i\}$ is a sequence of independent, identically distributed random variables with

$$p(x_i=1) = p, \quad p(x_i=0) = q. \tag{6.2.1}$$

The hypotheses

$$H_0 : p = p_0 \tag{6.2.2}$$

$$H_1 : p = p_1 \tag{6.2.3}$$

are to be tested against one another. A Bayesian viewpoint is adopted, so that there are a priori probabilities λ_0 for H_0 and λ_1 for H_1 ($\lambda_0 + \lambda_1 = 1$), and the cost criterion is the probability of error.

If x_i, $i = 0, 1, \ldots, T-1$, is observed, and the decision is based on these observations, it is well known that the optimal decision is a likelihood ratio test [Val]. Moreover, a sufficient statistic is the number of ones (on zero's) observed [Lil]. Storing this number requires a memory with no more than $\log_2 T$ bits.

An alternative formulation is to assume that only a given memory with less than $\log_2 T$ bits is available. For example, suppose that only one bit is available. Define the sets

$$x_t = \{0,1\} \tag{6.2.4}$$

with corresponding fields

$$X_t = \{\phi, \{0\}, \{1\}, x_t\} \tag{6.2.5}$$

for t = 0, 1, 2, ..., T-1. The memory sets are

$$M_t = \{0,1\} \tag{6.2.6}$$

with corresponding fields

$$\mathcal{M}_t = \{\phi,\ \{0\},\ \{1\},\ M_t\} \tag{6.2.7}$$

for t = 1, 2, ..., T. A set

$$U = \{0,1\} \tag{6.2.8}$$

of terminal decisions is also given.

Let η_t, t = 1, 2, ..., T, and γ_T denote functions

$$\eta_0 : X_0 \rightarrow M_1 \tag{6.2.9}$$

$$\eta_t : M_{t-1} \times X_{t-1} \rightarrow M_t,\ t = 2, 3, ..., T \tag{6.2.10}$$

$$\gamma_T : M_T \rightarrow U. \tag{6.2.11}$$

The functions η_t are the <u>memory update</u> functions, and the function γ_T is the <u>terminal decision function.</u>

Given $\eta_1, \eta_2, ..., \eta_t$, let

$$\tilde{\eta}_t : X_0 \times X_1 \times ... \times X_{t-1} \rightarrow M_t$$

be defined as follows,

$$\tilde{\eta}_1 : X_0 \rightarrow M_1$$
$$x_0 \rightarrow \eta_1(x_0) \tag{6.2.12}$$

and

$$\tilde{\eta}_t : X_0 \times X_1 \times \ldots \times X_{t-1} \to M_t \qquad (6.2.13)$$

$$(x_0, x_1, \ldots, x_{t-1}) \to \eta_t (\tilde{\eta}_{t-1}(x_0, x_1, \ldots, x_{t-2}), x_{t-1})$$

for $t = 2, 3, \ldots, T$. $\tilde{\eta}_T$ is the _memory structure_ induced by the memory update functions $\eta_1, \eta_2, \ldots, \eta_T$.

Define the product space X and product field X by

$$X = X_0 \times X_1 \times \ldots \times X_{T-1}, \qquad (6.2.14)$$

$$X = X_0 \times X_1 \times \ldots \times X_{T-1}. \qquad (6.2.15)$$

Then $\tilde{\eta}_T$ is a map from X to M_T. Define

$$\tilde{\gamma}_T : X_0 \times X_1 \times \ldots \times X_{T-1} \to U \qquad (6.2.16)$$

as the composition of γ_T and $\tilde{\eta}_T$,

$$\tilde{\gamma}_T = \gamma_T \circ \tilde{\eta}_T. \qquad (6.2.17)$$

Each hypothesis induces a probability denoted by $p(\cdot \mid H_0)$ or $p(\cdot \mid H_1)$ on X. (This probability is completely specified by the condition that the probability of a point with m ones and n zeros is $p^m(1-p)^n$ where $p = p_0$ or $p = p_1$ according to $H = H_0$ or $H = H_1$). Define the subsets S_0 and S_1 of X by the equations

$$S_0 = \tilde{\gamma}_T^{-1}(0), \qquad (6.2.18)$$

$$S_1 = \tilde{\gamma}_T^{-1}(1) \qquad (6.2.19)$$

The probability of error corresponding to $\tilde{\gamma}_T$ is

$$P_e(\tilde{\gamma}_T) = p(S_0 \mid H_1) \, \lambda_1 + p(S_1 \mid H_0) \, \lambda_0 \qquad (6.2.20)$$

where $\lambda_0 + \lambda_1 = 1$. ($\lambda_0$ is the a priori probability that H_0 is true.)

The problem of hypothesis testing with 1-bit memory considered in this chapter is to find

$$\min_{\tilde{\gamma}_T} \, P_e(\tilde{\gamma}_T) \qquad (6.2.21)$$

and the functions $\lambda_1, \lambda_2, \ldots, \lambda_T, \gamma_T$ defining the minimizing $\tilde{\gamma}_T$.

Several problems closely related to (6.2.21) have been considered in the literature. Cover [Co1] has considered the preceding problem for the limiting case $T \rightarrow \infty$. Hellman and Cover [He1] have considered the infinite time problem when attention is limited to time invariant, but possibly randomized, memory updates. Flower [Fl1] has considered the finite time problem but with attention again restricted to time invariant, although possibly randomized, memory updates.

6.3 Preliminary Analysis

For fixed $\tilde{n}_T$, the problem is easily solved. Let

$$F_T(\tilde{n}_T) = \tilde{n}_T^{-1}(M_T)$$

(6.3.1)

$$= \{\phi, E, \bar{E}, x\}$$

where

$$E = \tilde{n}_T^{-1}(1)$$

(6.3.2)

$$\bar{E} = \tilde{n}_T^{-1}(0)$$

(6.3.3)

is the __information field__ induced by the memory structure. Then the

Bayes optimal decision is determined by the condition that the a

posteriori probability be maximized [Val]. Therefore, if $m_T = 1$, choose

H_0 $(\gamma_T(1) = 0)$ if

$$p(E \mid H_0) \, \lambda_0 \geqslant p(E \mid H_1) \, \lambda_1$$

(6.3.4)

and choose H_1 $(\gamma_T(1) = 0)$ if

$$p(E \mid H_1) \, \lambda_1 > p(E \mid H_0) \, \lambda_0.$$

(6.3.5)

Similarly, if $m_T = 0$, choose H_0 $(\gamma_T(0) = 0)$ if

$$p(\bar{E} \mid H_0) \, \lambda_0 \geqslant p(\bar{E} \mid H_1) \, \lambda_1$$

(6.3.6)

and choose H_1 $(\gamma_T(0) = 1)$ if

$$p(\bar{E} \mid H_1 \, \lambda_1 > p(\bar{E} \mid H_0) \, \lambda_0.$$

(6.3.7)

Thus the problem reduces to finding the optimal memory structure, or equivalently, the optimal information field that can be obtained from some memory structure.

It is usefull at this point to define the notion of a rectangle. A rectangle $E \subset X_0 \times X_1 \times \ldots \times X_{T-1}$ is a set of the form

$$E = E_0 \times E_1 \times \ldots \times E_{T-1} \tag{6.3.8}$$

where $E_0 \in X_0, E_1 \in X_1, \ldots, E_{T-1} \in X_{T-1}$. The sets E_t are the sides of E.

Lemma 6.3.1

If $E \in X$ is a rectangle, then there is a memory structure $\tilde{n}_T$ such that

$$E = \tilde{n}_T^{-1}(1). \tag{6.3.9}$$

(The memory structure is said to realize E.)

Proof

Let $E = E_0 \times E_1 \times \ldots \times E_{T-1}$. Define

$$n_1 = \begin{cases} 1 & x_0 \in E_0 \\ 0 & x_0 \notin E_0 \end{cases} \tag{6.3.10}$$

$$n_t = \begin{cases} 1 & x_{t-1} \in E_{t-1} \text{ and } m_{t-1} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.3.11}$$

for $t = 2, 3, \ldots, T$. Obviously,

$$(x_0, x_1, \ldots, x_{T-1}) \; \epsilon \; \tilde{\eta}_T^{-1}(1)$$

if and only if $x_0 \; \epsilon \; E_0$, $x_1 \; \epsilon \; E_1$, $\ldots$, and $x_{T-1} \; \epsilon \; E_{T-1}$. But the latter

condition is equivalent to $(x_0, x_1, \ldots, x_{T-1}) \; \epsilon \; E_0 \; x \; E_1 x \; \ldots \; x \; E_{T-1} = E$.

There are eight possible relationships between p_1, p_0, q_1, q_0

(Figure 6.3.1). Because of the obvious symmetries involved, only

$$\underline{\text{Case 1}} \quad p_1 > q_0 \geqslant p_0 > q_1$$

and

$$\underline{\text{Case 2}} \quad p_1 > p_0 \geqslant q_0 > q_1$$

need be considered. For case 1, the following result is available.

Proposition 6.3.2

For case 1, there is an event for which the probability of

error is minimum within the class of rectangles either of the form

$$\{0\}^m \; x \; \{0,1\}^{T-m} \tag{6.3.12}$$

or

$$\{1\}^n \; x \; \{0,1\}^{T-n} \tag{6.3.13}$$

for some integer m, $0 \leqslant m \leqslant T$, or some integer n, $0 \leqslant n \leqslant T$.

Proof

Let F be an arbitrary rectangle. If F has n sides that are $\{1\}$, m

sides that are $\{0\}$, and $T-n-m \geqslant 0$ sides that are $\{0,1\}$, then

$$p_1 > p_0 \geq q_0 > q_1 \longleftrightarrow p_0 > p_1 \geq q_1 > q_0$$

$$p_1 > q_0 \geq p_0 > q_1 \longleftrightarrow p_0 > q_1 \geq p_1 > q_0$$

$$q_1 > q_0 \geq p_0 > p_1 \longleftrightarrow q_0 > q_1 \geq p_1 > p_0$$

$$q_1 > p_0 \geq q_0 > p_1 \longleftrightarrow q_0 > p_1 \geq q_1 > p_0$$
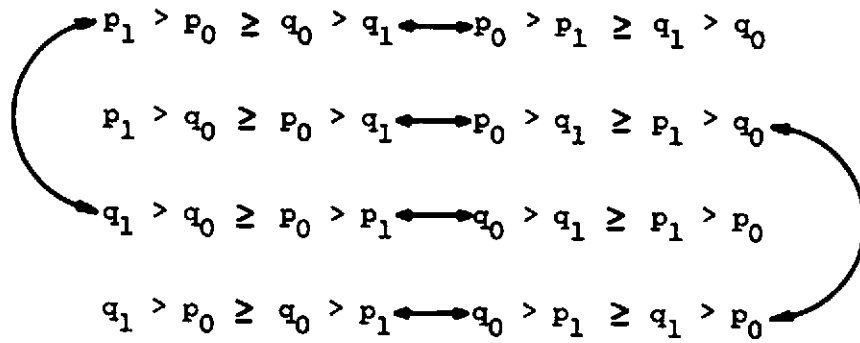
Figure 6.3.1   The Possible Relationships Between
p_0, p_1, q_0, q_1. (Excluding $p_1 = p_0$)

$$p(F \mid H) = p^n q^m \qquad\qquad (6.3.14)$$

where $p = p_0$ or $p = p_1$ according to $H = H_0$ or $H = H_1$. There are three subcases to be examined.

Case 1a

Suppose that the optimal terminal decision function is $\gamma_T \equiv 1$ or $\gamma_T \equiv 0$. In this case, the event F is useless, and can be replaced by an arbitrary event E without increasing the probability of error.

Case 1b

$$\lambda_1 p_1^n q_1^m \geqslant \lambda_0 p_0^n q_0^m \qquad\qquad (6.3.15)$$

$$\lambda_1 (1 - p_1^n q_1^m) \leqslant \lambda_0 (1 - p_0^n q_0^m) \qquad\qquad (6.3.16)$$

In this case, the optimum decision for observation of E is $E \to H_1$, $\bar{E} \to H_0$. The corresponding probability of error is

$$P_e = \lambda_1 (1 - p_1^n q_1^m) + \lambda_0 p_0^n q_0^m. \qquad\qquad (6.3.17)$$

Since

$$\frac{p_1}{q_1} > 1, \qquad\qquad (6.3.18)$$

$$\frac{p_0}{q_0} \leqslant 1, \qquad\qquad (6.3.19)$$

it follows that

$$\lambda_1 p_1^{n+1} q_1^{m-1} > \lambda_0 p_0^{n+1} q_0^{m-1}. \qquad\qquad (6.3.20)$$

Moreover, since

$$\lambda_1 (1 - p_1^{n+1} q_1^{m-1}) = \lambda_1 (1 - p_1^n q_1^m) - \lambda_1 p_1^n q_1^m (\frac{p_1}{q_1} - 1)$$

$$(6.3.21)$$

$$\lambda_0 (1 - p_0^{n+1} q_0^{m-1}) = \lambda_0 (1 - p_0^n q_0^m) - \lambda_0 p_0^n q_0^m (\frac{p_0}{q_0} - 1)$$

$$(6.3.22)$$

and by (6.3.15),(6.3.16), (6.3.18), (6.3.19) it follows that

$$\lambda_1 (1 - p_1^{n+1} q_1^{m-1}) < \lambda_0 (1 - p_0^{n+1} q_0^{m-1}). \qquad (6.3.23)$$

The probability of error for an event E with one side {0} of F changed to {1} is therefore

$$P_e = \lambda_1 (1 - p_1^{n+1} q_1^{m-1}) + \lambda_0 p_0^{n+1} q_0^m \qquad (6.3.24)$$

which is less than the corresponding probability of error for the event F.

Case 1c

$$\lambda_0 p_0^n q_0^m \geqslant \lambda_1 p_1^n q_0^m \qquad (6.3.25)$$

$$\lambda_0 (1 - p_0^n q_0^m) \leqslant \lambda_1 (1 - p_1^n q_1^m) \qquad (6.3.26)$$

By an argument completely analogous to that for case 1b, it can be shown that an event $\bar{E}$ with a side {1} of F changed to a {0} has a lower probability of error.

The proposition is established as follows. If F satisfies case 1a, F may be replaced by an arbitrary event that satisfies case 1b or

case 1c. If no such event exists, than any event, and in particular an

event of the form (6.3.12) or (6.3.13) is optimal. If F satisfies case 1b,

then (by an induction argument) all the sides {0} of F can be changed

to sides {1} without increasing the probability of error. If F

satisfies case 1c, then all the sides {1} of F can be changed to {0}.

Since a rectangular event corresponds to occurance of a particular

substring of $(x_0, x_1, \ldots, x_{T-1})$, it might seem that only rectangular

events could be realized by a 1-bit memory. If this were true, then the

problem would be solved (for case 1) since a complete class[1] of memory

update functions would those determining whether an event of the form

(6.3.12) or (6.3.13) did or did not occur. However, certain non-

rectangular events can be realized by a 1-bit memory.

There are $2^4 = 16$ possible functions $\eta_t : M_{t-1} \times X_{t-1} \to M_t$ for any

$t > 1$, since there are four elements of $M_{t-1} \times X_{t-1}$ and two elements of

M_t. However, symmetry considerations reduce the number of η_t that need

to be considered to the eight listed in Table 6.3.2.

Proposition 6.3.3

Given $\tilde{\eta}_T$, either $\tilde{\eta}_T$ or $1 - \tilde{\eta}_T$ can be constructed from the eight

memory update functions in Table 6.3.2, $T \geqslant 2$.

Proof

The proof proceeds by induction. Notice that any map $\eta : X_{t-1} \times$

$M_{t-1} \to M_t$ is either in Table 6.3.2 or $1 - \eta$ is in Table 6.3.2.

---

[1] A complete class of memory update functions satisfies the condition: for
any choice of λ_0, p_0, p_1, T, the optimum memory update function is in the
class.

η_t	(m_{t-1}, x_{t-1})				interpretation
	0 0	0 1	1 0	1 1	
η_t^1	1	1	1	1	no information
η_t^2	1	1	1	0	
η_t^3	1	1	0	1	picks out one
η_t^4	1	0	1	1	point of
η_t^5	0	1	1	1	$M_{t-1} \times X_{t-1}$
η_t^6	1	1	0	0	gives m_t
η_t^7	1	0	1	0	gives x_t
η_t^8	0	1	1	0	gives $(m_t + x_t)$ mod 2

Table 6.3.2   Eight Possible Memory Update Functions

There is a one-to-one correspondence between functions $\eta_2 : M_1 \times X_1 \rightarrow$
M_2 and functions $\tilde{\eta}_2 : X_0 \times X_1 \rightarrow M_2$, so that the proposition is true for

T=2 by the remark above.

Suppose Proposition 6.3.3 is true for arbitrary T. Note that $\tilde{\eta}_{T+1} =$
$\eta_T \circ (\tilde{\eta}_T \times i_T)$, where $i_T : X_T \rightarrow X_T$ is the identity map on X_T. By

assumption, either $\tilde{\eta}_T$ or $1 - \tilde{\eta}_T$ can be constructed from the table.

If $\tilde{\eta}_T$ can be so constructed, then $\tilde{\eta}_{T+1}$ or $1 - \tilde{\eta}_{T+1}$ can, since

$\tilde{\eta}_{T+1} = \eta_{T+1} \circ (\tilde{\eta}_T \times i_T)$ and since $1 - \tilde{\eta}_{T+1} = (1 - \eta_{T+1}) \circ (\tilde{\eta}_T \times i_T)$.

If $1 - \tilde{\eta}_T$ can be so constructed, modify η_{T+1} so that $\tilde{\eta}_{T+1} = \eta_{T+1} \circ$

$[(1 - \tilde{\eta}_T) \times i_T]$. Then $\tilde{\eta}_{T+1}$ or $1 - \tilde{\eta}_{T+1}$ can be so constructed, since

$1 - \tilde{\eta}_{T+1} = (1 - \eta_{T+1}) \circ [(1 - \tilde{\eta}_T) \times i_T]$ and either η_{T+1} or $1 - \eta_{T+1}$

is in the table. The proposition is therefore valid by the principle

of mathematical induction.

Suppose $\eta_1 = i_0$ (the identity on X_0) and $\eta_2 = \eta_2^8$. Then

$$\tilde{\eta}_2^{-1}(1) = \{ (0,1), (1,0) \} \tag{6.3.27}$$

which is a <u>non-rectangular event</u>. The interpretation is that <u>it is</u>

<u>possible to determine the parity of the string</u> $(x_0, x_1, \ldots, x_{T-1})$

<u>with a 1-bit memory</u>. This does not seem to be a very interesting thing

to know, but complicates the analysis greatly.

An efficient specification of

$$I_T = \bigcup_{\tilde{\eta}_T} \tilde{\eta}_T^{-1}(M_T) \tag{6.3.28}$$

would be useful. I_T is not in general a field, but does contain ϕ, X, and

is closed under complementation. The problem of specifying I_T is precisely the problem of determining the languages that can be accepted by a two state time-varying automaton in T steps. Unfortunately, there appears to be little work on this problem available in the literature [Hop1, Arb1, Bol].

The analysis of this section, while not conclusive, suggests the following conjecture.

Conjecture

The event for which the probability of error is minimum is a rectangular event either of the form

$$\{0\}^m \times \{0,1\}^{T-m} \tag{6.3.29}$$

or

$$\{1\}^n \times \{0,1\}^{T-n}. \tag{6.3.30}$$

In the next section, the problem will be reformulated as a FSFM stochastic control problem. The minimum principle will be used to find events superior to those of the form (6.3.29), (6.3.30). Thus, the above conjecture is false.

6.4 Application of the Minimum Principle

The development in the previous section proceeded independently from the remainder of the thesis. In this section, the problem of hypothesis testing with 1-bit memory is recast to fit within the FSFM format. The utility of the FSFM minimum principle will be illustrated by the derivation of a memory update scheme to serve as a counterexample to the conjecture of the previous section.

Let the variables $x_1(t)$, $x_2(t)$, $u(t)$, $w_0(t)$, $w_1(t)$, $v(t)$ take their values in the set $\{0,1\}$. Suppose that

$$x_1(0), \ w_0(1), \ w_1(1), \ \ldots, \ w_0(T+1), \ w_1(T+1)$$

is a sequence of independent random variables. Assume that $x_1(0)$, $w_0(t)$, $w_1(t)$ take the value 1 with respective probabilities λ_1, p_0, p_1 and let $\lambda_0 = 1 - \lambda_1$, $q_0 = 1 - p_0$, and $q_1 = 1 - p_1$. Moreover, $x_2(0) = 0$ and $m(0) = 0$.

Let state equations

$$x_1(t) = x_1(t-1) \tag{6.4.1}$$

$$x_2(t) = (1 - x_1(t-1)) \ w_0(t-1) + x_1(t-1) \ w_1(t-1) \tag{6.4.2}$$

$$m(t) = v(t) \tag{6.4.3}$$

be defined. Then $x_2(t)$, $t = 1, 2, \ldots, T+1$ is a sequence of independent zero-one random variables with probability p of one and probability q of zero. With probability λ_0, $p = p_0$, and with probability λ_1, $p = p_1$.

The memory is updated by

$$v(t) = \eta_t(m(t-1), x_2(t-1)) \qquad (6.4.4)$$

for $T = 1, 2, \ldots, T+1$ and $u(T+2)$ is specified by

$$u(T+2) = \gamma_{T+2}(m(T+1)). \qquad (6.4.5)$$

The cost is

$$J = (x_1(T+1) - u(T+2))^2. \qquad (6.4.6)$$

Since $x_1(T+1), u_1(T+2) \in \{0,1\}$, the expectation of J under the distribution defined by $\eta_1, \eta_2, \ldots, \eta_{T+1}$ and γ_{T+2} is simply the probability of error. Figure 6.4.1 illustrates the sequence of events.

When the appropriate identifications are made, the preceeding problem can be shown to be equivalent to a FSFM problem. However, it is straightforward to write down the equivalent deterministic problem from the equations above.

Notice that the state set is $\{0,1\} \times \{0,1\} \times \{0,1\}$. This is equivalent to the state set $X_\infty = \{1, 2, 3, 4, 5, 6, 7, 8\}$ when the identifications of Figure 6.4.2 are made. Let $V_\infty = \{0,1\}$. Then the restriction on $\eta_t : X_\infty \to V_\infty$ is that η_t be constant on the sets of the partition

$$\{\{1,5\}, \{2,6\}, \{3,7\}, \{4,8\}\} \qquad (6.4.7)$$

of X_∞. (This corresponds to the fact that η in (6.4.4) cannot depend on $x_1(t)$.) Similarly, $\gamma_{T+2} : X_\infty \to U_\infty = \{0,1\}$ must be constant on the sets of the partition

$$\{\{1,2,5,6\}, \{3,4,7,8\}\} \qquad (6.4.8)$$

t=0:   •transition to $x_1(0)$, $x_2(0)$

      •observation of $x_2(0)$ (no information)

      •memory update $v(1)$ (arbitrary)

t=1:   •transition to $x_1(1)$, $x_2(1)$

      •observation of $x_2(1)$ (observation 1)

      •memory update $v(2)$

            •
            •
            •

t=T-1: •transition to $x_1(T-1)$, $x_2(T-1)$

      •observation of $x_2(T-1)$ (observation T-1)

      •memory update $v(T)$

t=T: •transition to $x_1(T)$, $x_2(T)$

      •observation of $x_2(T)$ (observation T)

      •memory update $v(T+1)$

t=T+1: •transition to $x_1(T+1)$, $x_2(T+1)$

      •choice of control $u(T+2)$

      •observation of $x_2(T+1)$

      •memory update $v(T+2)$ (arbitrary)

Figure 6.4.1   Sequence of Events
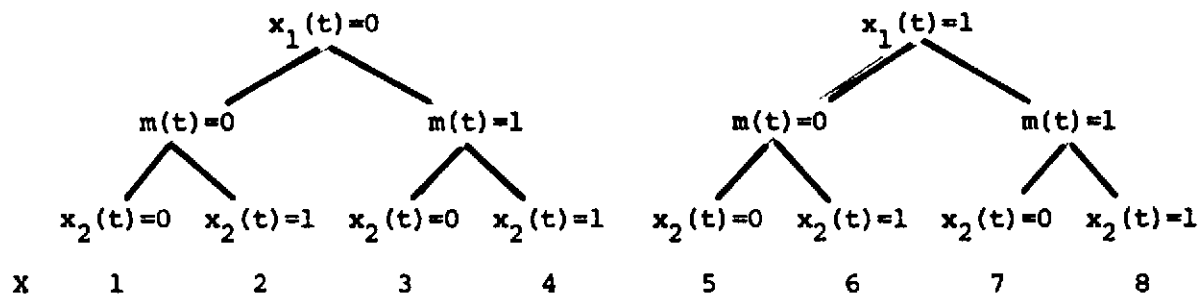
Figure 6.4.2   Definition of the State Set X for the FSFM Problem

of X_∞.

From Figure 6.4.2 and the problem specification, it is easy to compute the parameters of the equivalent deterministic problem. Let P_{ij}^{v} be the probability of a transition from state i to state j when the memory update function is identically equal to v.

$$P^0 = \begin{bmatrix} q_0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_0 & p_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_1 & p_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_1 & p_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_1 & p_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_1 & p_1 & 0 & 0 \end{bmatrix} \qquad P^1 = \begin{bmatrix} 0 & 0 & q_0 & p_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_0 & p_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_0 & p_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_0 & p_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_1 & p_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_1 & p_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_1 & p_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_1 & p_1 \end{bmatrix}$$

Similarly,

$$h_{T+2}^{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad h_{T+2}^{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\pi_0 = [\lambda_0 \ 0 \ 0 \ 0 \ \lambda_1 \ 0 \ 0 \ 0]$$

(All other terms in the cost are zero.)

At this point, attention is restricted to the special case $T = 3$, $p_1 = \frac{3}{4}$, $p_0 = \frac{1}{4}$, and $\lambda_0 = \lambda_1 = \frac{1}{2}$. The optimality of the trial solution

$$\gamma_5^* : \{1,2,5,6\} \to 0, \quad \{3,4,7,8\} \to 1$$

$$\eta_4^* : \{1,2,3,5,6,7\} \to 0, \quad \{4,8\} \to 1$$

$$\eta_3^* : \{1,2,3,5,6,7\} \to 0, \quad \{4,8\} \to 1$$

$$\eta_2^* : \{1,3,5,7\} \to 0, \quad \{2,4,6,8\} \to 1$$

will be tested by the FSFM minimum principle.

The condition

$$\pi^*(2) \; P^{\eta_3^*} \; \phi^*(3) \leqslant \pi^*(2) \; P^{\eta_3} \; \phi^*(3)$$

will be tested first. (Arbitrarily assume $\eta_1^* \equiv 0$.)

$$\pi^*(0) = [\lambda_0 \; 0 \; 0 \; 0 \; \lambda_1 \; 0 \; 0 \; 0]$$

$$\pi^*(1) = \pi^*(0) \; P^{\eta_1^*} = [\lambda_0 q_0 \; \lambda_0 p_0 \; 0 \; 0 \; \lambda_1 q_1 \; \lambda_1 p_1 \; 0 \; 0]$$

$$\pi^*(2) = \pi^*(1) \; P^{\eta_2^*} = [\lambda_0 q_0^2 \; \lambda_0 q_0 p_0 \; \lambda_0 q_0 p_0 \; \lambda_0 p_0^2 \; \lambda_1 q_1^2 \; \lambda_1 q_1 p_1$$
$$\lambda_1 q_1 p_1 \; \lambda_1 p_1^2]$$

$$h^{\gamma_5} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \phi^*(5) = 0$$

$$\phi^*(4) = P^{\eta_5^*} \; \phi^*(5) + h^{\gamma_5^*} = h^{\gamma_5^*}$$

$$\phi*(3) = P^{\eta_4*} \phi*(4) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Substituting $p_0 = \frac{1}{4}$, $p_1 = \frac{3}{4}$, $\lambda_0 = \lambda_1 = \frac{1}{2}$, the following expression for $\pi*(2)$ is obtained.

$$\pi*(2) = \begin{bmatrix} \frac{9}{32} & \frac{3}{32} & \frac{3}{32} & \frac{1}{32} & \frac{1}{32} & \frac{3}{32} & \frac{3}{32} & \frac{9}{32} \end{bmatrix}$$

Notice that

$$P^0 \phi*(3) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad P^1 \phi*(3) = \begin{bmatrix} p_0 \\ p_0 \\ p_0 \\ p_0 \\ q_1 \\ q_1 \\ q_1 \\ q_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

The minimizing $\hat{\eta}_3$ for $\pi*(2) P^{\eta_3} \phi*(3)$ is obtained as follows. Since $\frac{1}{32} < \frac{5}{64}$, $\hat{\eta}_3(1) = \hat{\eta}_3(5) = 0$. Since $\frac{3}{64} < \frac{3}{32}$, $\frac{3}{64} < \frac{3}{32}$, $\frac{5}{64} < \frac{9}{32}$, $\hat{\eta}_3(2) = \hat{\eta}_3(3) = \hat{\eta}_3(4) = \hat{\eta}_3(6) = \hat{\eta}_3(7) = \hat{\eta}_3(8) = 1$. Since $\hat{\eta}_3 \neq \eta_3*$ the sequence

η_1^*, η_2^*, η_3^*, η_4^*, γ_5^* cannot be optimal. However, it can be verified

that η_1^*, η_2^*, $\hat{\eta}_3$, η_4^*, γ_5^* does satisfy the necessary conditions.

The interpretation of this result is obtained with the aid of

Figure 6.4.2. The map $\eta_1^* \equiv 0$ is arbitrarily chosen since the first

observation contains no information. η_2^* simply transmits the first

observation. η_3^*, η_4^* put a 1 in memory if the observation is 1 and the

previous memory state is 1. The net result is a memory structure $\tilde{\eta}_4^*$

that tests for the observation of three 1's. Thus

$$\{(1,1,1)\} = \tilde{\eta}_3^{*-1}(1)$$

is the event realized.

In contrast, $\hat{\eta}_3$ puts a zero in the memory if the previous memory

state was zero and the observation was zero. The net effect is a memory

structure $\tilde{\hat{\eta}}_4$ formed from $(\eta_1^*, \eta_2^*, \hat{\eta}_3, \eta_4^*)$. The memory structure places

a zero in memory if the first two observations were zeros. Otherwise, a

zero or one is placed in memory according to whether the last observation

is zero or one. Thus the non-rectangular event realized is

$$\tilde{\hat{\eta}}_3^{-1}(1) = \{(1,1,1), (0,1,1), (1,0,1)\}.$$

$\tilde{\hat{\eta}}_3^{-1}(1)$ is a considerably closer to the (unrealizable) optimum event

$\{(1,1,1), (0,1,1), (1,0,1), (1,1,0)\}$ than $\tilde{\eta}_3^{*-1}(1)$: only the event

$(1,1,0)$ is misclassified.

There are two noteworthy features of the preceeding analysis. First

application of the minimum principle has resulted in a counterexample to

the conjecture of the previous section. This is impressive since the

problem certainly revolves around the determination of signaling control laws in both the formal and informal meanings of the term. As was pointed out in Chapter 3, the presence of signaling control laws indicates the absence of a universal extremal so that the minimum principle is not necessarily very helpful for this class of problems.

Second, the preceeding example for T=3 suggests more general memory structures for T > 3. For example, consider the event $E \in X_0 \times X_1 \times \ldots \times X_6$,

$$E = \{\{1\} \times \{1\} \times \{1\} \times \{0,1\} \times \{0,1\} \times \{1\}\} \cup \{\overline{\{1\} \times \{1\} \times \{1\}} \times \overline{\{0\} \times \{0\}}\} \times \{1\}\}$$

where $\bar{A}$ is the complement of the set A. For $p_0 = \frac{1}{4}$, $p_1 = \frac{3}{4}$, $\lambda_0 = \lambda_1 = \frac{1}{2}$ as before detection of this event results in a probability of error

$$p_e = \frac{791}{2048} < \frac{832}{2048}$$

where $\frac{832}{2048}$ is the probability of error the event constructed for the case T=3.

Thus a sequence of events of decreasing probability of error can be constructed for $T \to \infty$. However, verification of the optimality of these events requires the application of a sufficient condition of optimality such as the dynamic programming algorithm discussed in the next section.

6.5 Application of Dynamic Programming

In section 2.5, it was stated that the equivalent deterministic problem to the FSFM problem was not always the most efficient such problem. In this section, this statement is justified by demonstrating that the hypothesis testing problem of this chapter is a fact equivalent to a deterministic problem with a two dimensional state space.[1] Although the dynamic programming equations for the two dimensional problem are considerably simplified, the analysis is still too difficult to be completed by hand.

Define the quantities

$$
\begin{bmatrix} \alpha(t) = \text{Prob}(\text{memory} = 1 \mid H_0) \\ \\ \beta(t) = \text{Prob}(\text{memory} = 1 \mid H_1) \end{bmatrix}
$$
(6.5.1)

(6.5.2)

for $t = 0, 1, \ldots, T$. Let $\theta_{ij}(t) = k$ mean that if $m(t-1) = i$, $x(t-1) = j$, then $m(t) = k$, where $i, j, k \in \{0,1\}$. Then the state equations

$$
\begin{bmatrix} \alpha(t) \\ \\ \beta(t) \end{bmatrix} = f_t(\alpha(t-1), \beta(t-1); \theta(t))
$$
(6.5.3)

can be written where

$$f_t(\alpha(t-1), \beta(t-1); \theta(t)) =$$

$$\begin{bmatrix} \theta_{11}(t)p_0+\theta_{10}(t)q_0-\theta_{01}(t)p_0-\theta_{00}(t)q_0 & 0 \\ 0 & \theta_{11}(t)p_1+\theta_{10}(t)q_1-\theta_{01}(t)p_1-\theta_{00}(t)q_0 \end{bmatrix}$$

$$x \begin{bmatrix} \alpha(t-1) \\ \beta(t-1) \end{bmatrix} + \begin{bmatrix} \theta_{01}(t)p_0 + \theta_{00}(t)q_0 \\ \theta_{01}(t)p_1 + \theta_{00}(t)q_1 \end{bmatrix} \qquad (6.5.4)$$

Minimizing the probability of error is equivalent to minimizing $V_T(\alpha,\beta)$, where

$$V_T(\alpha,\beta) = \begin{cases} \frac{1}{2}(1 - \alpha + \beta) & \alpha \geqslant \beta \\ \frac{1}{2}(1 - \alpha + \beta) & \alpha < \beta \end{cases} \qquad (6.5.5)$$

The state equations (6.5.4) with the cost function (6.5.5) define a deterministic optimal control problem equivalent to the problem of hypothesis testing with 1-bit memory.

The dynamic programming backward algorithm for the optimal control problem is

$$V_{t-1}(\alpha,\beta) = \min_{\theta} V_t(f_t(\alpha, \beta; \theta)) \qquad (6.5.6)$$

where V_T is given by (6.5.5). The optimal memory update functions are determined as follows. Let $\theta^\ell(t)$ be the minimizing control for (α,β) ε $R_\ell(t)$, where $R_\ell(t)$ is a convex region with piecewise linear boundary.[1] Then the minimizing control law for (α,β) ε $R_\ell(t)$ is defined by

---

[1]Such a region $R_\ell(t)$ for which the minimizing θ is constant exists by analysis similar to that of Chapter 4.

$$\eta_t^{\ell}(i,j) = k \qquad\qquad\qquad (6.5.7)$$

where $\theta^{\ell}(i,j) = k$.

For the case $p_0 = \frac{1}{4}$, $p_1 = \frac{3}{4}$, $\lambda_0 = \lambda_1 = \frac{1}{2}$, the dynamic programming algorithm has been carried for $t = T$ and $t = T-1$. The results are illustrated in Figures 6.5.1 and 6.5.2. Notice the functions V_T, V_{T-1} are piecewise linear and concave, as might be expected from the analysis in Chapter 4.

Although obtaining a 2-dimensional equivalent deterministic optimal control problem simplifies application of the dynamic programming algorithm, the computation of V_{T-2} is still too complicated to be computed by hand. Solution by digital computer is required.

$\beta(T)$

$R_1(T)$

$$V_T^1(\alpha,\beta) = \frac{1}{2}(1 + \alpha - \beta)$$

$$\gamma_T^1 : \{1\} \to 1$$
$$\{0\} \to 0$$

$R_2(T)$

$$V_T^2(\alpha,\beta) = \frac{1}{2}(1 + \beta - \alpha)$$

$$\gamma_T^2 : \{1\} \to 0$$
$$\{0\} \to 1$$

$\alpha(T)$
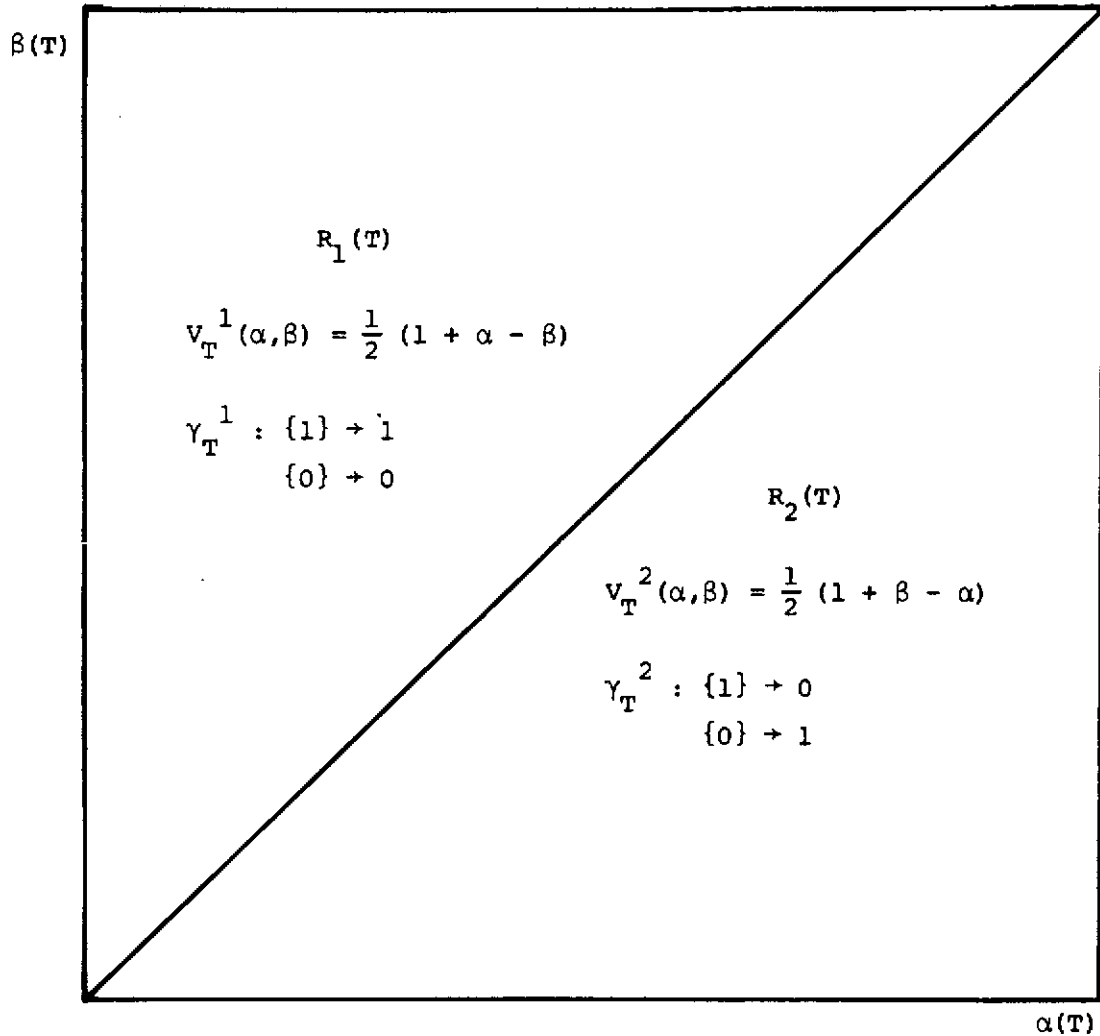
Figure 6.5.1   The Function V_T (Note $\gamma_T^1 : \{1\} \to 1$ means $\gamma_T^1(1) = 1$)

$\beta(T-1)$

$R_1(T-1)$

$V_{T-1}^{1}(\alpha,\beta)=\frac{1}{2}[1+\alpha-\beta]$

η_T^{1} : $\{11,10\} \to 1$
$\{00,01\} \to 0$

$R_2(T-1)$

$V_{T-1}^{2}(\alpha,\beta)=\frac{1}{2}[1+\frac{1}{4}\alpha-\frac{3}{4}\beta]$

η_T^{2}: $\frac{\{11\}}{\{11\}} \to 1 \atop \to 0$

$R_3(T-1)$

$V_{T-1}^{3}(\alpha,\beta) = \frac{1}{2}[\frac{1}{2}+\frac{3}{4}\alpha-\frac{1}{4}\beta]$

η_T^{3}: $\frac{\{00\}}{\{00\}} \to 1 \atop \to 0$

$R_4(T-1)$

$V_{T-1}^{4}(\alpha,\beta) = \frac{1}{4}$

η_T^{4} : $\{01,11\} \to 1$
$\{10,00\} \to 0$

$R_5(T-1)$

$V_{T-1}^{5}(\alpha,\beta) = \frac{1}{2}[1+\frac{1}{4}\beta-\frac{3}{4}\alpha]$

η_T^{5} : $\frac{\{10\}}{\{10\}} \to 1 \atop \to 0$

$R_6(T-1)$

$V_{T-1}^{6}(\alpha,\beta)=\frac{1}{2}[\frac{1}{2}+\frac{3}{4}\beta-\frac{1}{4}\alpha]$

η_T^{6} : $\frac{\{01\}}{\{01\}} \to 0 \atop \to 1$

$R_7(T-1)$

$V_{T-1}^{7}(\alpha,\beta)=\frac{1}{2}[1+\beta-\alpha]$

η_T^{7} : $\{11,10\} \to 1$
$\{00,01\} \to 0$

$\frac{2}{3}$ $\frac{1}{3}$ 1 0

$\alpha(T-1)$
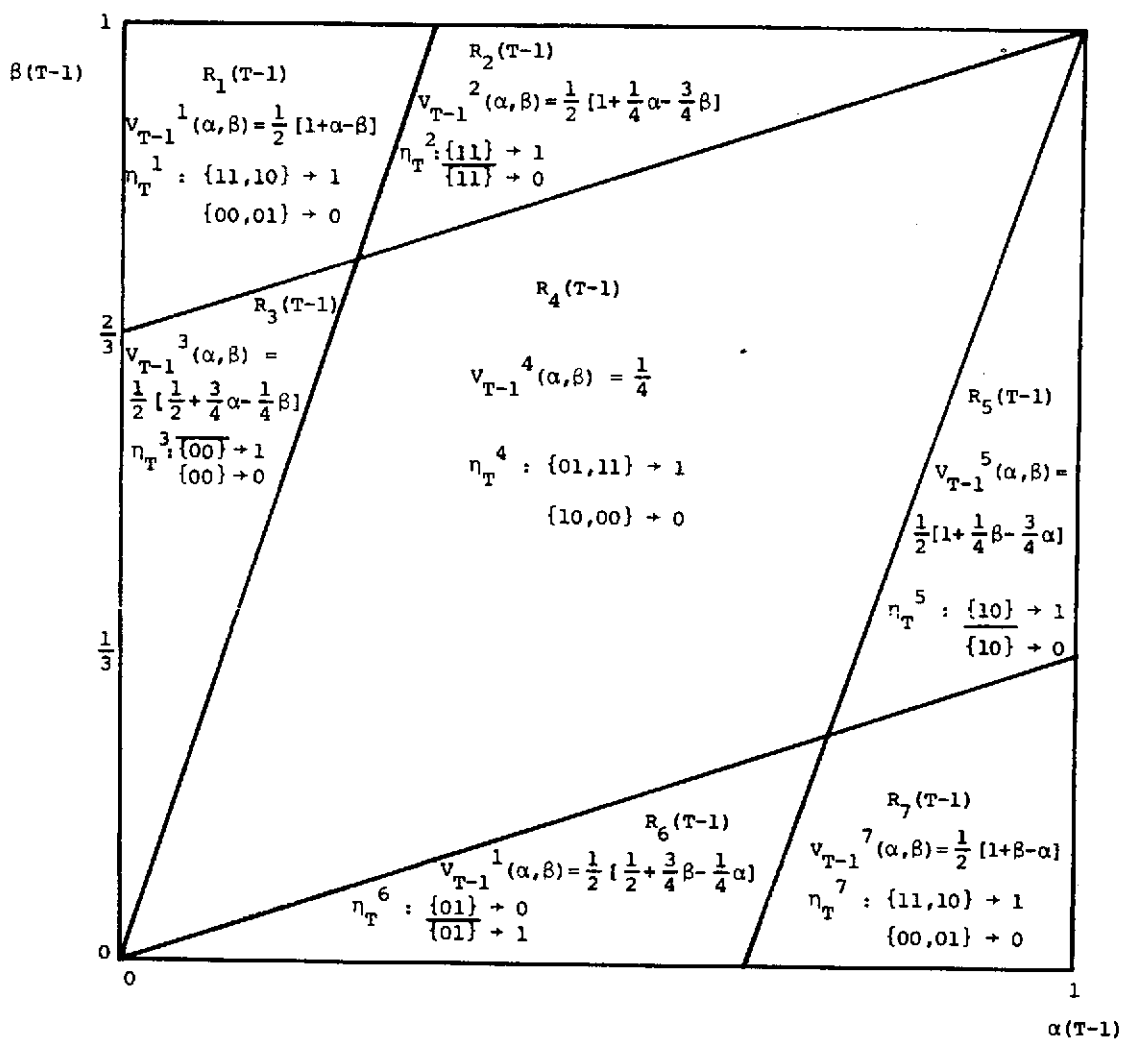
Figure 6.5.2   The Function V_{T-1} (Note η : $\{11\} \to 0$ means $\eta(1,1) = 0$

$\overline{\{11\}}$ = $\{01,10,00\}$ is the complement of $\{11\}$)

CHAPTER VII

SUMMARY, CONCLUSIONS, AND SUGGESTIONS

FOR FURTHER INVESTIGATION

This chapter summarizes the results of the thesis, with a

brief discussion of the conclusions that can be drawn from the

research. A list of possible topics for future investigation

is included.

7.1 Summary

The thesis began in Chapter I with the formulation of a

rather general problem in the design of engineering control systems.

The formulation was intended to motivate the FSFM model studied in

the remainder of the thesis. The FSFM problem is a non-classical

stochastic control problem, and so the existing literature on this

and other closely related topics was briefly surveyed. The chapter

closed with a brief summary of the remaining chapters.

The FSFM model was introduced in Chapter II. It was demon-

strated that a number of apparently more general problems

can be reduced to FSFM models, so that most of the features of the

general engineering control system of Chapter I can be incorporated

in the FSFM formulation. Then an example was given to illustrate

the important signaling strategies that must be considered in

non-classical stochastic control problems. Finally, a deterministic

optimal control problem equivalent to the FSFM problem was derived.

In Chapter III, the FSFM minimum principle was stated and

proved. A Kuhn extensive game model equivalent to the FSFM problem

was obtained so that the notion of a signaling strategy could be

precisely defined. The importance of this concept was established

by proving the existence of a <u>universal extremal</u> for problems <u>without</u>

signaling strategies (i.e., with <u>perfect recall</u>). A numerical

optimization algorithm, the <u>person-by-person min-H algorithm</u>, was

derived based on the minimum principle.

In Chapter IV dynamic programming was considered. As might be

expected, dynamic programming is not a practical procedure for

numerical optimization except for simple special cases.

In Chapter V, the infinite horizon version of the FSFM model

was formulated. The discounted cost criterion was considered since

this criterion led to a well-defined equivalent deterministic

problem. The Value and Policy Iteration methods were extended to

the FSFM problem, as were algorithms of Sondik implementing these

methods.

A problem of hypothesis testing with 1-bit memory was considered

in Chapter VI. Although an optimal solution was not obtained, use

of the minimum principle suggested an interesting class of memory

updates. This result provides some indication that control

theoretic methods can be useful for design of information-handling

systems.

7.2 Conclusions

The fundamental difficulty in non-classical stochastic control
problems in general, and the FSFM problem in particular, is the
occurrence of signaling strategies. This phenomenon, which does
not occur in classical stochastic control, complicates the analysis
in an essential way, since the choice of control laws at different
time instants is tightly coupled. As a consequence, the min-H
algorithm proposed for the numerical solution of FSFM models is
not guaranteed to converge to the globally optimal solution. However,
as illustrated in Chapter 6, the min-H algorithm in conjunction
with some engineering judgement in the choice of the initial
guess can be an effective tool.

The applicability of the algorithms implementing dynamic
programming is more limited. The basic difficulty here is classical;
it is necessary to solve a high dimensional functional equation to
implement dynamic programming. In spite of the large amount of
work devoted to this problem, no generally applicable satisfactory
procedure is available. Thus, the dynamic programming approach is
appropriate only for problems with a rather small state set (around
10 states at most).

Even the min-H algorithm is not adequate to handle large scale
engineering systems directly. The problem is basically combinatorial:
all the observation, memory, and communication sets are lumped with

the state set so that the state set becomes very large. For

example, a system with 100 physical states, and two controllers

each with a 10 state memory set and each observing an output

that takes 10 values requires a FSFM model with 1 million states!

A number of techniques must be employed to handle such a

problem. One generally applicable approach is to remove some of

the redundancy associated with the FSFM representation of the

problem by taking advantage of the factorization of the state set

into the physical and memory sets. For example, notice that with

the memory updates at a particular instant fixed in the above

problem, transition to 990,000 of the states (those corresponding

to memory states not chosen) is impossible. Thus, for computational

work, it is better to retain the factorization of the state set

into the physical state set and the memory set. Other factorizations

may be possible in specific instances.

An important technique in large scale systems theory is

aggregation. As applied to the FSFM model, this technique con-

sists of grouping states together into aggregate states and only

considering transitions between the aggregate states. The resulting

model may closely approximate the original model if the aggregate

states are well chosen, and will be more tractable computationally.

Another possibility involves utilizing any special structure

that occurs in a particular large scale problem. Generally speaking,

the special structure of the problem will be reflected in the fact

that many state transitions will not be allowed. Thus, the

associated state transition matrices will be sparse. A particular

example of this situation has already been mentioned above in

connection with the memory sets. Exploitation of the structural

properties of the transition matrices requires a flexible representa-

tion. One possibility might be to store the transition matrices

as a PL/1 data structure.

To summarize, study of the FSFM model was motivated by the

problems of control and information in large scale systems. The

FSFM model does provide a vehicle for the study of phenomena that

occur in such systems. However, direct solution of large scale

system problems by the algorithms of this thesis will not be possible,

in general, due to limitations on the size of the state set for which

the algorithms are computationally feasible. Techniques such as

aggregation can be used to reduce a large scale system problem to

a computationally feasible size, and any special structure of the

problem should be exploited to mitigate the computational burden.

7.3 Suggestions for Future Research

The study of non-classical stochastic control problems is still at an early stage. Therefore, there are many possibilities for further investigation. Some of these are listed below.

(1) Further study and refinement of the FSFM model.

    (a)   Study of the interaction of communication and control in the FSFM context.

    (b)   Study of the tradeoff between employing signaling strategies and providing additional communication channels.

    (c)   Extension of the analysis of Chapter VI to problems with larger memories.

    (d)   Specialization of the FSFM problem to the case in which the sets involved have an algebraic structure.

    (e)   Determination of upper and lower bounds for the optimal cost without computing the optimal control laws.

(2) Studies aimed at reducing the computational burden.

    (a)   Exploitation of the structure of the FSFM state space as the product of the physical state set with other sets.

    (b)   Replacement of the matrix representation of the FSFM model with one more suited for computational purposes (e.g., a PL/1 data structure).

    (c)   Examination of the possibility of parallel computation

          in the min-H algorithm.

(3) Application of the theory to specific problems.

    (a)   Traffic networks [Houl].

    (b)   Computer communication networks [Kal].

(4) Extensions of the theory

    (a)   To non-sequential stochastic control problems.

    (b)   To FSFM games.

    (c)   Generalization of the signaling strategy notion to

          continuous state spaces.

    (d)   Study of linear designs for linear, quadratic,

          Gaussian problems by techniques similar to those

          developed for the FSFM problem.

REFERENCES

[An1]   R.N. Anthony, Management Accounting Principles, Irwin,
         Homewood, Ill., 1970.

[Ao1]   M. Aoki, Optimization of Stochastic Systems, Academic
         Press, N.Y., 1967.

[Ao2]   M. Aoki, "On Decentralized Linear Stochastic Problems with
         Quadratic Cost", IEEE Trans A.C., Vol. AC-18, No. 3, 1973.

[Arb1]   M.A. Arbib, Theories of Abstract Automata, Prentice-Hall,
         Englewood Cliffs, N.J., 1969.

[Ar1]   K.J. Arrow and F.H. Hahn, General Competitive Analysis,
         Holden-Day, N.Y., 1971.

[As1]   K.J. Astrom, "Optimal Control of Markov Processes with
         Incomplete State Information", J. Math. Anal. Appl.,
         Vol. 10, 1965.

[A1]   M. Athans and P.L. Falb, Optimal Control, McGraw-Hill,
         N.Y., 1966.

[Au1]   R.J. Aumann, "Mixed and Behavior Strategies in Infinite
         Extensive Games", Advances in Game Theory, Princeton
         University Press, Princeton, N.J., 1964.

[B1]   R. Bellman, Dynamic Programming, Princeton University Press,
         Princeton, N.J., 1957.

[B2]   R. Bellman, Adaptive Control Processes: A Guided Tour,
         Princeton University Press, Princeton, N.J., 1961.

[B3]   R. Bellman, Introduction to the Mathematical Theory of
         Control Processes, Academic Press, N.Y., 1967.

[Bi1]   J.M. Bismut, "An Example of Interaction Between Information
         and Control: The Transparency of a Game", IEEE Trans. A.C.,
         Vol. AC-18, No. 5, 1973.

[Bl1]   D. Blackwell, "Discounted Dynamic Programming," Annals
         Math. Statistics, Vol. 36, 1965.

- 182 -

[Bo1]   T.L. Booth, <u>Sequential Machines and Automata Theory</u>,
         Wiley, N.Y., 1967.

[Br1]   R.W. Brockett, "System Theory on Group Manifolds and
         Coset Spaces," <u>SIAM J. Control</u>, Vol. 10, No. 2, 1972.

[Ca1]   C.H. Carpenter, Jr., <u>Some Control and Communication Aspects
         of Stochastic Systems</u>, M.S. Thesis, M.I.T., 1972

[Cha1]   B. Chandrasekaran, "Finite Memory Hypothesis Testing:
         A Critique", <u>IEEE Trans. I.T.</u>, Vol. IT-16, No. 4, 1970.

[Cha2]   B. Chandrasekaran, "Reply to 'Finite Memory Hypothesis
         Testing - Comments on a Critique'", <u>IEEE Trans. I.T.</u>,
         Vol. IT-17, No. 1, 1971

[C1]   C.Y. Chong, <u>On the Decentralized Control of Large-Scale
         Systems</u>, Ph.D. Thesis, M.I.T., 1973.

[C2]   C.Y. Chong, <u>On the Stochastic Control of Coupled Linear
         Systems</u>, M.S. Thesis, M.I.T., 1970.

[Ch1]   H. Chernoff, "Backward Induction in Dynamic Programming",
         unpublished memo, 1963.

[Chu1]   K.C. Chu, "Team Decision Theory and Information Structures
         in Optimal Control Problems - Part II", <u>IEEE Trans. A.C.</u>,
         Vol. AC-17, No. 1, 1972.

[Co1]   T.M. Cover, "Hypothesis Testing with Finite Statistics",
         <u>Annals Math. Statistics</u>, Vol. 40, No. 3, 1969.

[Co2]   T.M. Cover and M.E. Hellman, "Finite Memory Hypothesis
         Testing - Comments on a Critique", <u>IEEE Trans. I.T.</u>,
         Vol. IT-16, No. 4, 1970.

[D1]   N. Dalkey, "Equivalence of Information Patterns and
         Essentially Determinate Games," <u>Contributions to the
         Theory of Games II</u>, Princeton University Press,
         Princeton, N.J., 1953.

[De1]   C. Derman, <u>Finite State Markovian Decision Processes</u>,
         Academic Press, N.Y., 1970.

[Do1]   J.L. Doob, <u>Stochastic Processes</u>, Wiley, N.Y., 1953.

[Dr1]   A.W. Drake, <u>Observation of a Markov Process Through a Noisy Channel</u>, Sc.D. Thesis, M.I.T., 1962.

[Dy1]   E.B. Dynkin, "Controlled Random Sequences," <u>Th. Prob. Appl.</u>, Vol. 10, 1965.

[F1]   A.A. Fel'dbaum, <u>Optimal Control Systems</u>, Academic Press, N.Y., 1965.

[Fl1]   R.A. Flower, <u>Hypothesis Testing with Finite Memory in Finite Time</u>, M.S. Thesis, M.I.T., 1970.

[Gal]   R.G. Gallager, <u>Information Theory and Reliable Communication</u>, Wiley, N.Y., 1968.

[Hall]   H. Halkin, "A Maximum Principle of the Pontryagin Type for Systems Described by Nonlinear Difference Equations", <u>SIAM J. Control</u>. Vol. 4, No. 1, 1966.

[Hal]   P.R. Halmos, <u>Measure Theory</u>, Van Nostrand, N.Y., 1950.

[Hel]   M.E. Hellman and T.M. Cover, "Learning with Finite Memory," <u>Annals Math. Statistics</u>, Vol. 41, No. 3, 1970.

[H1]   K. Hinderer, <u>Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter</u>, Springer-Verlag, Berlin, 1970.

[Hi1]   P.R. Hirschler, "Finite Memory Algorithms for Testing Bernoulli Random Variables", <u>Information and Control</u>, Vol. 24, 1974.

[Ho1]   Y.C. Ho and K.C. Chu, "Team Decision Theory and Information Structures in Optimal Control Problems - Part I", <u>IEEE Trans. A.C.</u>, Vol. AC-17, No. 1, 1972.

[Hol1]   J.M. Holtzman, "On the Maximum Principle for Nonlinear Discrete-Time Systems", IEEE Trans. A.C., Vol. AC-11, No. 1, 1966.

[Hop1]   J.E. Hopcroft and J.D. Ullman, <u>Formal Languages and Their Relation to Automata</u>, Addison-Wesley, Reading, Mass., 1969.

[Hou1]   P.K. Houpt, <u>Decentralized Stochastic Control of Discrete-State Systems with Application to Vehicular Traffic Control</u>, Ph.D. Thesis, M.I.T., 1974.

[Howl]   R.A. Howard, <u>Dynamic Programming and Markov Processes</u>,
         M.I.T. Press, Cambridge, Mass., 1960.

[How2]   R.A. Howard, <u>Dynamic Probabilistic Systems</u>, Vols. I
         and II, Wiley, N.Y., 1971.

[Kal]   R.E. Kahn, "Resource-Sharing Computer Communications
         Networks", <u>Proc. IEEE</u>, Vol. 60, No. 11, 1972.

[Kakl]   J.S. Kakalik, <u>Optimum Policies for Partially Observable
         Markov Systems</u>, M.S. Thesis, M.I.T., 1965.

[Kal]   L.V. Kantorovich and G.P. Akilov, <u>Functional Analysis
         in Normed Spaces</u>, Macmillan, N.Y., 1964.

[Kel]   H.J. Kelley, "Method of Gradients," <u>Optimization
         Techniques</u>, Academic Press, N.Y., 1962.

[Krl]   S.I. Krich and T. Berger, "Coding for a Delay-Dependent
         Fidelity Criterion", <u>IEEE Trans. I.T.</u>, Vol. IT-20,
         No. 1, 1974.

[Kl]   H.W. Kuhn, "Extensive Games", <u>Proceedings of the National
         Academy of Sciences</u>, U.S.A., Vol. 36, 1950.

[K2]   H.W. Kuhn, "Extensive Games and the Problem of Information"
         <u>Contributions to the Theory of Games II</u>, Princeton
         University Press, Princeton, N.J., 1953.

[Kul]   H.J. Kushner, <u>Introduction to Stochastic Control</u>, Holt,
         Rinehart, and Winston, N.Y., 1971.

[Kwl]   R.H. Kwong, <u>On a Stochastic Control Method for Weakly
         Coupled Linear Systems</u>, M.S. Thesis, M.I.T., 1972.

[Larl]   R.E. Larson, <u>State Increment Dynamic Programming</u>,
         Elsevier, N.Y., 1969.

[Lal]   L.S. Lasdon, <u>Optimization Theory for Large Systems</u>,
         Macmillan, N.Y., 1970.

[Lil]   B.W. Lindgren, <u>Statistical Theory</u>, Macmillan, N.Y., 1968.

[Lol]   M. Loeve, <u>Probability Theory</u>, Van Nostrand, N.Y., 1963.

[L1]   R.D. Luce and H. Raiffa, Games and Decisions, Wiley,
        N.Y., 1957.

[M1]   J. Marshak, "Elements for a Theory of Teams",
        Management Science, Vol. 1, 1955.

[M2]   J. Marshak and R. Radner, The Economic Theory of Teams,
        Yale University Press, New Haven, Conn., 1971

[Me1]   M.D. Mesorovic, D. Macko, and Y. Takahara, Theory of
        Hierarchical Multilevel Systems, Academic Press,
        N.Y., 1970.

[Mi1]   H. Mine and S. Osaki, Markovian Decision Processes,
        Elsevier, N.Y., 1970.

[Mo1]   R.R. Mohler and A. Ruberti (Eds.), Theory and Applications
        of Variable Structure Systems, Academic Press,
        N.Y., 1972.

[N1]   John F. Nash, "Equilibrium Points in N-Person Games",
        Proceedings of the National Academy of Science, U.S.A.
        Vol. 36, 1950.

[N2]   John F. Nash, "Non-Cooperative Games", Annals of
        Mathematics, Vol. 54, No. 2, 1951.

[O1]   G. Owen, Game Theory, Saunders, N.Y., 1968.

[P11]   L.K. Platzman, On the Iterative Design of Optimal and
        Suboptimal Controllers, M.S. Thesis, M.I.T., 1973.

[Po1]   L.S. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and
        E. Mishchenko, The Mathematical Theory of Optimal
        Processes, Interscience Publishers, N.Y., 1962.

[R1]   R. Radner, "Team Decision Problems", Annals of Math.
        Statistics, Vol. 33, 1962.

[R2]   R. Radner, "The Evaluation of Information in Organizations",
        Proc. Fourth Berkeley Symp. Math. Statistics and Prob.,
        University of California Press, 1961.

[Ra1]   H. Raiffa and R. Schlaifer, Applied Statistical Decision
        Theory, Harvard University Press, Cambridge, Mass., 1961.

[Ro1]   S.M. Ross, "Non-Discounted Denumerable Markovian Decision
        Models", Annals Math. Statistics, Vol. 39, 1968.

[Ro2]   S.M. Ross, "Arbitrary State Markovian Decision Models",
          Annals Math. Statistics, Vol. 39, 1968.

[Ru1]   W. Rudin, Real and Complex Analysis, McGraw-Hill,
          N.Y., 1966.

[S1]   N.R. Sandell, Jr. and M. Athans, "Solution of Some Non-
          Classical LQG Stochastic Decision Problems", IEEE Trans.
          A.C., Vol. AC-19, No. 2, 1974.

[Sa1]   L.J. Savage, The Foundations of Statistics, Wiley,
          N.Y., 1954.

[Saw1]   Y. Sawaragi and T. Yoshikawa,"Discrete-Time Markovian
          Decision Processes with Incomplete State Observations",
          Annals Math. Statistics, Vol. 41, 1970.

[Se1]   I.E. Segal and R.A. Kunze, Integrals and Operators,
          McGraw-Hill, N.Y., 1968.

[Sh1]   C.E. Shannon, "A Mathematical Theory of Communication",
          Bell System Tech. J., Vol. 27, 1948.

[Si1]   G.F. Simmons, Introduction to Topology and Modern
          Analysis, McGraw-Hill, N.Y., 1963.

[Sm1]   R.D. Smallwood and E.J. Sondik, "The Optimal Control of
          Partially Observable Markov Processes over a Finite
          Horizon", Oper. Res.,Vol. 21, No. 5, 1973.

[Sol1]   D. Solomons, Divisional Performance Measurement and
          Control, Financial Executives Research Foundation, 1965.

[Sol]   E.J. Sondik, The Optimal Control of Partially Observable
          Markov Processes, Ph.D. Thesis, Stanford, University, 1971.

[Sta]   A.N. Starr and Y.C. Ho, "Further Properties of Nonzero-
          Sum Differential Games", Journal of Optimization Theory and
          Applications, Vol. 3, No. 4, 1971.

[St1]   C.T. Striebel, "Sufficient Statistics in the Optimum Control
          of Stochastic Systems", J. Math. Anal. Appl., Vol. 12, 1965.

[T1]   G.L. Thompson, "Signaling Strategies in n-Person Games",
          Contributions to the Theory of Games II, Princeton
          University Press, Princeton, N.J., 1953.

[Va1]   H.L. Van Trees, <u>Detection, Estimation and Modulation Theory</u>, Wiley, N.Y., 1968.

[V1]   J. Von Neuman and O. Morgenstern, <u>Theory of Games and Economic Behavior</u>, Princeton University Press, Princeton, N.J., 1947.

[Wa1]   A. Wald, <u>Statistical Decision Functions</u>, Wiley, N.Y., 1950.

[Wi1]   J.C. Willems, <u>The Analysis of Feedback Systems</u>, M.I.T. Press, Cambridge, Mass., 1971.

[Will] A.S. Willsky, <u>Dynamic Systems Defined on Groups: Structural Properties and Estimation</u>, Ph.D. Thesis, M.I.T., 1973.

[Wis1] D.A. Wismer, Editor, <u>Optimization Theory for Large-Scale Systems</u>, McGraw Hill, N.Y., 1971.

[W1]   H.S. Witsenhausen, "A Counterexample in Stochastic Optimal Control", <u>SIAM J. Control</u>, Vol. 6, 1968.

[W2]   H.S. Witsenhausen, "On Information Structures, Feedback and Causality," <u>SIAM J. Control</u>, Vol. 9, 1971.

[W3]   H.S. Witsenhausen, "Separation of Estimation and Control for Discrete Time Systems", <u>Proc. IEEE</u>, Vol. 59, No. 11, 1971.

[W4]   H.S. Witsenhausen, "A Standard Form for Sequential Stochastic Control", <u>Math. Sys. Th.</u>, Vol. 7, 1973.

BIOGRAPHICAL NOTE

Nils R. Sandell, Jr. was born ███████████████ on
███████████████ He attended public high school in Coon Rapids,
Minnesota, graduating in 1966. He received the B.E.E. degree from
the University of Minnesota in 1970, and the M.S. and E.E. degrees
from the Massachusetts Institute of Technology in 1971 and 1973.

Mr. Sandell has been supported at M.I.T. by the Bell Telephone
Laboratories Graduate Study Program (1970-1971), and by a NSF
Fellowship (1971-1974). He has held summer positions with Bell
Telephone Laboratories (1969, 1970), the Honeywell Systems and
Research Center (1972), and the M.I.T. Center for Advanced Engineering
Study (1973).

Mr. Sandell has published several journal articles, and his
1970 Honors Paper at Minnesota, "Hunting Deer by Optimal Control",
won third prize in the Region IV IEEE student paper contest. He
is a member of Eta Kappa Nu and Tau Beta Pi, having served as an
officer in the Minnesota Chapter of each organization, and is an
associate member of Sigma Xi.

Mr. Sandell is married to the former Yvonne Amelia ████████.