PM 58

ZERO MEMORY CONTROLLERS IN STOCHASTIC CONTROL

Graham Seddon

A thesis submitted for the Degree of Doctor of Philosophy

December 1979

Department of Computing and Control

Imperial College of Science and Technology

University of London

ABSTRACT

To extend the results of classical stochastic control theory to cater for more general information patterns is extraordinarily difficult. This thesis is an exposition of the nature of this difficulty. A simple, scalar, discrete-time LQG problem is examined with both classical and zero-memory controllers. Topics such as second-guessing and the conflict between control and communication are thus introduced.

For the finite-time case, a dynamic programming approach is presented in which a suitable probability density function is selected as a state variable. Again both the zero-memory and classical cases are examined; indeed the classical case is considered in some detail since it has been the source of some confusion in the past. To contrast zero-memory and classical control, a discrete state example is worked in full. A number of sub-optimal techniques, including perturbation analysis, are also discussed, together with a cautionary formulation of the zero-memory problem which is plausible, attractive, and also quite wrong.

Considerable attention is devoted to the zero-memory infinite-horizon problem. Because of the practical significance of stationary control, various approximation techniques are also considered in detail - Edgeworth approximation, moment formulation, and perturbation analysis.

Continued reference is made to the famous Witsenhausen problem, for which it is ultimately shown that the necessary condition for optimality has a unique solution amongst the class of bounded control laws.

The initial motivation for this thesis was provided by the Witsenhausen problem and a latter chapter reviews various contributions since its publication in 1968.

Emphasis throughout is on concepts and discussion; abstract mathematics and intricate technicalities have been avoided where possible.

ACKNOWLEDGEMENTS

I would like to acknowledge the generous support and encouragement which my supervisor, Dr. J.M.C. Clark, has provided over an exceptionally long period. Tribute is also due to my wife, Toni, who has not only tolerated my continuing pre-occupation with this research but has also accepted the unfortunate role of typist.

Financial support during the period 1968-1971 was afforded by the Science Research Council.

CONTENTS

VOLUME 1

		Page
Chapter 1	PROLOGUE	
1.1	Introduction	1.1
1.2	Background Information	1.2
1.3	The Approach of this Thesis	1.8
1.4	Material to be Presented	1.11
1.5	Contributions of this Thesis	1.14
Chapter 2	AN INTRODUCTORY EXAMPLE	
2.1	Introduction	2.1
2.2	Simple Two-Stage Example	2.6
	2.2.1 Classical Information Pattern	2.6
	2.2.2 Nested Observation Case	2.17
	2.2.3 Zero Memory Problem	2.25
2.3	Summary and Conclusions	2.46
Chapter 3	A DYNAMIC PROGRAMMING APPROACH	
3.1	Introduction	3.1
3.2	Statement of Stochastic Optimal Control Problem	3.3
	3.2.1 Description of the Stochastic Control System	3.4
	3.2.2 Evolution of the Stochastic Control System	3.7
	3.2.3 Criterion of Performance	3.14
3.3	Dynamic Programming Formulation	3.18
	3.3.1 Dynamic Programming and the Stochastic Optimal Control Problem	3.18

3.3.2 Zero-Memory Case 3.20

3.3.3 Classical Information Pattern 3.33

3.3.4 General Information Pattern 3.54

3.4 Summary and Conclusions 3.61

Chapter 4 A DISCRETE STATE EXAMPLE

4.1 Introduction 4.1

4.2 Problem Definition 4.3

4.3 Solution for Zero-Memory Case 4.6

4.4 Solution for Classical Information Pattern Case 4.8

4.5 Comparison Between Two Cases 4.12

4.6 Summary and Conclusions 4.28

Chapter 5 SOME PLAUSIBLE HYPOTHESES EXAMINED AND SOME SUB-OPTIMAL TECHNIQUES

5.1 Introduction 5.1

5.2 The Classical Information Pattern Problem 5.2

5.2.1 Dual Control 5.2

5.2.2 Open Loop Optimal Feedback Control 5.4

5.2.3 Optimal Linear Control 5.8

5.3 The Zero Memory Problem 5.12

5.3.1 The Neo-Classical Information Pattern Problem 5.12

5.3.1.1 The Useless Information Case 5.20

5.3.1.2 The Valuable Information Case 5.28

5.3.1.3 The Witsenhausen Problem 5.32

5.3.2 The Indirect Approach 5.34

5.3.3 Open Loop Optimal Feedback Control 5.40

5.3.4 Optimal Linear Control 5.48

5.3.5 Optimal Polynomial Control 5.54

5.4 Summary and Conclusions 5.56

Chapter 6 A PERTURBATION APPROACH TO THE STOCHASTIC OPTIMAL CONTROL PROBLEM

6.1 Introduction 6.1

6.2 A General Examination of the Perturbation Technique 6.5

6.2.1 The Methodology of the Perturbation Technique 6.5

6.2.2 The Accuracy of the Approximation 6.10

6.2.3 Taylor Series Expansion of Some Relevant Quantities 6.15

6.3 Perturbation Solution of Classical Information Pattern Problem 6.21

6.3.1 Optimal Control at Time N-1 6.23

6.3.2 Optimal Control at Time N-2 6.24

6.3.3 Optimal Control at Time k Given Optimal Control at Time k+1 6.25

6.3.4 Optimal Cost and Explicit Control Laws 6.27

6.3.5 Comments on the Analysis 6.29

6.4 Perturbation Approach to Zero Memory Problem 6.37

6.4.1 Statement of Problem 6.38

6.4.2 Dynamic Programming Formulation 6.39

6.4.3 Derivation of Necessary Condition 6.41

6.4.4 Perturbation Solution 6.43

6.4.5 Comments on the Result 6.46

6.5 Summary and Conclusions 6.48

Chapter 7 OPTIMAL CONTROL OF STATIONARY SYSTEMS

7.1	Introduction	7.1
7.2	The Infinite Horizon Zero-Memory Problem	7.3
	7.2.1 Statement of the Problem	7.7
	7.2.2 Discussion of the Problem	7.11
	7.2.3 Summary	7.47
7.3	Zero-Memory Optimal Linear Controllers	7.51
	7.3.1 Problem Statement	7.53
	7.3.2 State Propogation under Linear Control	7.55
	7.3.3 The Discount Problem	7.56
	7.3.4 The Average Cost per Period Problem	7.61
	7.3.5 The Ergodic Problem	7.63
	7.3.6 Summary	7.69
7.4	Further Examination of the Ergodic Problem	7.72
	7.4.1 How Good is the Linear Controller?	7.72
	7.4.2 Design for Linear Multi-Dimensional Systems	7.87
	7.4.3 Optimality of the Static Linear Control Law	7.93
	7.4.4 Approximation of Methods	7.94
	7.4.5 Summary	7.129
7.5	Further Examination of the Discount Problem	7.132
	7.5.1 Successive Approximation Method	7.132
	7.5.2 Asymptotic Behaviour of Solution	7.139
	7.5.3 Summary	7.141
7.6	Conclusions	7.141

Chapter 8 FURTHER EXAMINATION OF WITSENHAUSEN'S PROBLEM

8.1 Introduction 8.1

8.2 Optimal Linear Control for Witsenhausen Problem 8.5

8.3 Necessary Condition for Witsenhausen Problem 8.11

8.4 Attempts to Show that T is a Contraction Mapping 8.15

8.5 Conclusions 8.32

Chapter 9 SOME COMMENTS ON RECENT DEVELOPMENTS

9.1 Introduction 9.1

9.2 Separation and Estimation 9.9

9.3 Information Structures 9.20

9.4 Delayed Sharing Information Patterns 9.29

9.5 Linear Solutions in Non-Classical Problems 9.33

9.6 Communications and Control 9.35

9.7 Large Scale Systems 9.40

9.8 Conclusions 9.43

Chapter 10 EPILOGUE

10.1 Introduction 10.1

10.2 Review of Thesis 10.3

10.3 Conclusions 10.7

REFERENCES R1-8

VOLUME 2

Appendix 1 LIST OF FREQUENTLY OCCURING SYMBOLS A1.1

Appendix 2 ANALYSIS OF THE LINEAR QUADRATIC GAUSSIAN PROBLEM WITH NESTED INFORMATION

Page

A.2.1 Introduction A2.1

A.2.2 Two Stage Example - Nested Observation Case A2.1

A.2.3 Two Stage Example - Classical Information Pattern Problem A2.4

A.2.4 Derivation of Filtering Equations A2.7

A.2.5 Sufficient Statistics A2.12

A.2.6 Comments A2.17

Appendix 3 SOLUTION OF TEST EXAMPLE OF CHAPTERS 4 AND 5

A.3.1 Solution for Zero-Memory Case A3.1

A.3.2 Solution for Classical Information Pattern Case A3.18

A.3.3 Elimination of $p(x_1 \mid I_1)$ from γ_1^0 A3.34

A.3.4 Neo-Classical Solution of Discrete State Problem A3.39

A.3.5 Neo-Classical Solution of Witsenhausen's Problem A3.48

Appendix 4 PERTURBATION ANALYSIS

A.4.1 Perturbation Solution of Classical Information Pattern Problem A4.1

A.4.1.1 Optimal Control at Time N-1 A4.1

A.4.1.2 Optimal Control at Time N-2 A4.4

A.4.1.3 Optimal Control at Time k A4.13

A.4.2 Perturbation Solution of Witsenhausen's Problem A4.18

A.4.2.1 Dynamic Programming Formulation A4.18

A.4.2.2 Determination of Approximation to Optimal Control Law A4.26

Appendix 5 ANALYSIS OF THE STATIONARY PROBLEM

A.5.1	Some Popular Results for Stationary Problems	A5.1
	A.5.1.1 The Discount Problem	A5.1
	A.5.1.2 The Average Return per Period Problem	A5.15
A.5.2	Linear System, Quadratic Cost Stationary Problem	A5.27
	A.5.2.1 The Discount Problem	A5.29
	A.5.2.2 The Average Cost per Period Problem	A5.33
A.5.3	Approximate Optimal Linear Controller for Zero-Memory Discount Problem	A5.37

CHAPTER 1

Prologue

		Page
1.1	Introduction	1.1
1.2	Background Information	1.2
1.3	The Approach of this Thesis	1.8
1.4	Material to be Presented	1.11
1.5	Contributions of this Thesis	1.14

1.1 Introduction

This thesis is about the role which the so-called information pattern of a problem plays in stochastic optimal control. Whilst a fairly general approach will be followed, particular attention will be paid to what, in the parlance of generalised control theory, is called the zero-memory problem and, in the parlance of large scale control systems theory, is called decentralised control.

In this introductory chapter, we shall provide appropriate background information for the material which will be presented in subsequent chapters. In addition we shall explain the expository and self-contained approach to presentation which has been adopted, and we shall then briefly review the contents of this thesis. Finally, we shall summarise the contribution which it is claimed the material and the manner of its presentation represent.

1.2 Background Information

Stochastic control theory deals with dynamical systems, described by difference or differential equations, and subject to disturbances which are characterised as stochastic processes. It was Norbert Wiener who first had the insight to observe that all cybernetic situations - those involving communication and control - are essentially stochastic. If there are no chance phenomena present, then there is no need for feedback and hence, in principle, no problem.

Given a system and a criterion, the classical problem of stochastic optimal control is to find the control law which minimises the objective. The tools to solve this problem are of fairly recent development, and indeed are still under development. Thus we have the famous Kalman-Bucy filtering equations, the Separation Theorem, and the Certainty Equivalence Principle, where the concept of Dynamic Programming has been widely used to establish such results. However, control theoreticians realised that many problems outside the traditional field of control engineering were essentially control problems. Perhaps, with some extensions, control theory could make contributions in areas such as operational research, economics, game theory, and so on. The term Generalised Control Theory was coined in 1970 by Ho (23) to accomodate the extension of conventional control theory to these new areas. Apparently, generalisation was required in three directions.

First, conventional control theory was concerned with the notion of a single controller whereas for Generalised Control Theory multiple con-

trollers were required. Thus the static team problem of Radner (38) could now be interpreted as a one-stage generalised control problem. Our mutiple controllers now represent the players in Radner's team; each controller makes an observation and on the basis of this implements a control action. Since the order of play is unimportant, with the system only changing state when all controllers have acted, we have a one-stage problem. Clearly, multiple controllers were required if the new theory was to accommodate any game-theory situations.

Next, the question of the information available to a controller needed generalisation. This is the so-called information pattern of the problem. To accommodate this aspect of the problem, it was necessary to extend the idea of multiple controllers even further - to 'multiple-multiple' controllers. Thus in the static team problem we have multiple controllers C_i $i = 1,2,\ldots N$ each acting at the same time instant. What was necessary now was the idea of multiple controllers acting autonomously at multiple time instants, that is C_{it} $i = 1,2,\ldots N_t, t = 1,2,\ldots T$ We now have a total of $\sum_{t=1}^{T} N_t$ controllers in the problem, each one free to do as it pleases. Each controller must determine its action on the basis of the information I_{it} available to it. A specification of the inter-relationships between the various I_{it} defines the information pattern of the problem. In this way it is possible to model communications delays, the degree of information sharing, finite-memory controllers and so on. In this framework, the conventional control problem is considered to be composed of T autonomous controllers, each acting at successive time instants. This is an important change in interpretation. Hitherto, the

idea of a single controller in the conventional control problem has obscured the importance of the information pattern. Since the same controller had been involved, implicit assumptions about memory capacity and the sharing of information had been made. These assumptions were now highlighted and their importance began to be appreciated.

Finally, Generalised Control Theory needed the simultaneous consideration of multiple objective functions. In a game situation, in which the players compete, there are always two objective functions. One player will attempt to maximise what the other attempts to minimise. But competition is not the only reason why multiple cost functions will occur. When co-operation is present, we have to model situations such as co-alition for example. Each partner in the co-alition has his own objective function, which will conflict to some extent with the aspirations of the other members of the co-alition. What then are the optimal compromises? Can we do better if we trust each other?

We have presented a discussion of Generalised Control Theory since it provided a convenient unified approach to several issues which were receiving attention from control theoreticians at the time. For example in 1969, Wonham (59) had offered some informal remarks on the current status and future direction of stochastic control. In this article, he suggests that the control problem of stochastic control is an elucidation of what one really understands by information flow around a cybernetic loop, and that stochastic control will come into its own not in the study of small numbers of differential equations, but in the study of large systems, such as human organisations, with some flexibility of structure. The

remarks of Wonham emphasise another feature of problems outside the realm of conventional control theory which must be accommodated by any new theory. Methods of applying control theory to large scale systems had to be developed. If we look at a realistic large scale system, such as a management structure, we see that it has a number of characteristics:-

i) Delegation of responsibility takes place and a hierarchial structure evolves. This is due to the fact that central control of the entire enterprise is beyond the intellectual capacity of one individual. In control theory, this limitation can be expressed as a constraint on the computational ability of a controller.

ii) The flow of information, both horizontally and vertically, is limited. A manager does not need to know, and does not want to know, everything about everybody else in the organisation. He has enough problems of his own. In control theory, the flow of information is expressed by the information pattern of the problem.

Thus large scale systems theory must, in control theory terms answer such questions as "Given computational constraints, what is the best information pattern and degree of de-centralisation - that is, what C_{it}, I_{it} do we need?" The key problems in the study of large systems are thus those of computation and information.

In other areas of optimisation theory, notably linear programming, some success had been achieved in the area of computation. For large

scale systems with a block diagonal structure for the constraint matrix, it is possible to decompose the problem into a number of interlinked sub-problems to which are adjoined a number of co-ordinating constraints. The computation is thus effectively de-centralised and could be performed, if required, simultaneously on a number of small capacity computers. In control theory, attempts have been made to imitate these decomposition procedures by introducing various interaction constraints into the formulation.

The information aspects of the problem had received attention in other fields, for example Radner (43). In control theory however the information pattern of the problem had been relatively neglected.

Notice that large scale system theory is just a special case of Generalised Control Theory, since we can easily formulate such notions as hierarchial control in the context of Generalised Control Theory by specification of the information pattern. Consider the single stage multi-level system depicted below:-

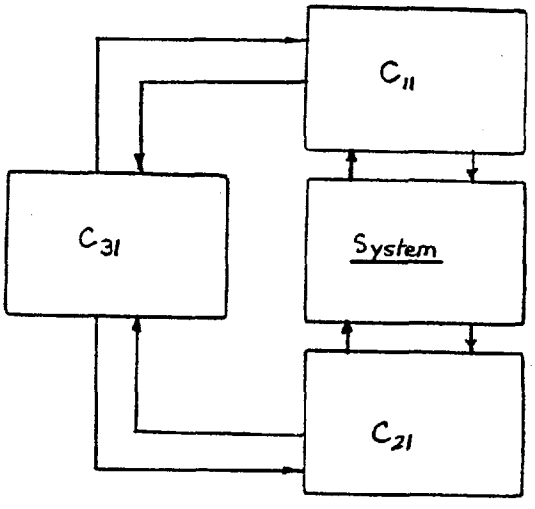


Fig. 1.1 Multi-Level Single Stage Control

Controller C_{31} might not in fact have any direct influence on the system. Thus his control actions can be visualised as information actions - his role is to simply analyse and transmit information from C_{21} to C_{11} and vice versa. Direct communication is ruled out presumably by the computational limitations of C_{11} and C_{21} - they need the assistance of C_{31} to co-ordinate their limited ability.

Above we have reviewed how conventional control theory can be extended into new application areas. These new application areas are of immense practical significance; for example:-

1) Decentralised and Distributed computing networks.
2) Electrical power system generation and Transmission networks.
3) Social and economic systems.
4) Traffic Control Networks.

Clearly, the state-of-the-art implementation of these systems is well in advance of the theory.

1.3 The Approach of this Thesis

We have emphasised above the importance of the information pattern of a control problem. In this thesis we shall examine the topic of information pattern in some detail. The author's interest in the subject was stimulated by the publication in 1968 by Witsenhausen (56) of a now famous counter-example in which he examined the optimal control of a very simple linear system with quadratic cost and gaussian noise. Given complete information, the sclution was well established, but Witsenhausen showed that if the information to a controller was restricted then the optimal control law would be non-linear. Moreover, the example demonstrated very clearly that communication as well as control was a feature of the control law. This thesis is a result of (fruitless) attempts to find the optimal control law for the Witsenhausen problem.

We shall consider a fairly conventional control system throughout, but we shall consider a particular form of information pattern - which we have called zero-memory information here. Thus we have the usual feedback loop but the control law at any instant is a function of the current observation only. The situation is illustrated below:-

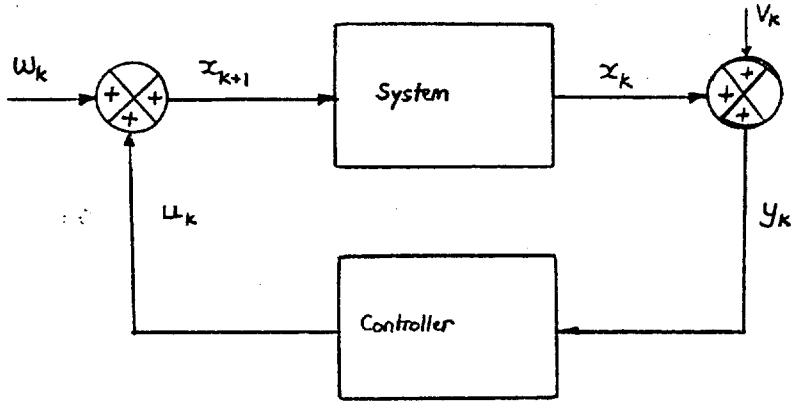


Fig. 1.2 Zero Memory Control System

When a controller has access to all previous observations and controls, the information pattern is usually called classical. Hence what we have here is a non-classical information pattern, although there is certainly nothing non-classical about the problem itself. It is a very traditional servo mechanism problem.

Notice that the zero-memory problem is an example of what in the terminology of large scale systems would be called a de-centralised system. Viewing the controller at each time instant as an autonomous entity, there is no sharing of information between successive controllers. Thus the controllers are decentralised since they enjoy no common information. Nor in this case is there any higher level control to co-ordinate their activities. Zero-memory problems are not restricted in their importance to large scale systems, however. They can occur quite naturally in systems of much lower dimension. For example, consider the air traffic control problem in which communication between the ground and the aircraft is lost. Alternatively, consider the servomechanism problem in which the controller has no memory for some reason (such as cost, or speed of response).

We remarked in the previous section that the information pattern of a problem has received scant attention from control theoreticians and this has tended to obscure its importance. Thus throughout this thesis we shall consider the zero memory problem and its classical information counterpart in parallel. This will not only permit comparisons between the two versions of the problem to be made but it will perhaps remove some of the con-

fusion regarding the classical problem resulting from a lack of explicit consideration of the information pattern.

The approach throughout will be expository. We shall consider only discrete time systems and we shall generally examine scalar examples. Our intention is to examine concepts, and this objective is not served by introducing either the delicate details necessary for continuous time systems nor the complexity of matrix algebra. In addition, the development will be totally self-contained. However a self-contained treatment requires that certain tedious algebra is also presented, which inevitably will distract from the main development of the thesis. Consequently a compromise approach has been adopted; wherever possible digressions, messy analysis, intricate calculations and so on have been relegated to an appropriate Appendix.

1.4 Material to be Presented

In Chapter 2 we shall present a discussion of a simple two-stage linear system with quadratic cost and gaussian noise. This discussion is intended to introduce the topic of information pattern in a concrete fashion. It will also illustrate at an early stage what it is that makes the zero-memory problem an order of magnitude more difficult than the classical problem.

Dynamic programming will be utilised in Chapter 3 to formalise the ideas developed in Chapter 2. We shall also provide a somewhat more extensive definition of the problem and its information pattern. Again, both the classical and zero-memory versions of the problem will be considered. The techniques which we derive will be applied to obtain comparative solutions of a finite state system in Chapter 4. To illustrate techniques, the computations are presented in full. Since the computations themselves are not central to any further development, they are presented in a separate Appendix.

In Chapter 5 some sub-optimal techniques and some common pitfulls will be discussed. We shall introduce such terms as dual control and open loop optimal feedback control. We shall also point out the dangers of testing proposed techniques on the basis of inadequate test examples.

An important sub-optimal technique, perturbation analysis, will be examined in Chapter 6. We shall use the technique for an examination of our familiar LQG problem and the Witsenhausen problem. This chapter, like Chapter 4, is characterised by the complexity of the algebra which is

generated.

Optimal stationary control is discussed in Chapter 7. We shall present a review of the traditional theory of stationary control as applied to Markov chains and classsical LQG systems. We shall then develop a corresponding theory for our zero-memory LQG problem, considering the discount version, the average cost per period version, and the ergodic version. In this chapter we shall also discuss a number of approximation techniques, such as Edgeworth approximation, moment formulation, and perturbation analysis. We shall also consider the problem of finding optimal linear control laws for the zero-memory problem in its infinite-horizon and finite-time forms.

Chapter 8 is devoted totally to further examination of the Witsenhausen problem. In particular we shall attempt, unsuccessfully, to show that under some circumstances a linear control law can be optimal for the Witsenhausen problem. All that we are able to produce is a uniqueness result which requires that the optimal control law is bounded.

The work in all preceding chapters, with the exception of Chapter 8, was completed prior to 1972. Reporting of the work has however been delayed by a variety of factors, not the least being an obsession with achieving the elusive objectives of Chapter 8. Thus Chapter 9 presents some comments on recent developments which are designed to set our remarks in a more contemporary context. It is gratifying that our development has stood the test of time remarkably well and our somewhat careful self-contained approach is amply justified.

A summary of our work is presented in Chapter 10, together with some conclusions and suggestions for further development. The importance of information patterns, particularly in the context of large scale systems is again emphasised. In fact it is argued that with the advent of the microprocessor and the so-called information explosion, the development of a suitable theory is now a matter of some urgency.

1.5 Contributions of this Thesis

The material presented inthis thesis is the sole responsibility of the author and, as explained above with the exception of Chapter 8 and obviously Chapter 9, was completed prior to 1972. Thus the question arises as to whether its merits should be judged in the context of 1972 or in the context of the present day. In view of the remarks of Chapter 9, it is felt that the question does not really matter. The research was prompted by a desire to find a solution to the Witsenhausen problem and in this we were unsuccessful. But then the problem still remains unresolved. So how does the system designer handle problems with non-classical information patterns? We have presented a number of approximation techniques which can be applied to both classical and non-classical problems to help in this respect.

In 1968 when the Witsenhausen problem was published, it was assumed that the classical problem was well understood. It is argued here that this was not the case and that considerable confusion existed in the field at that time. It is hoped that the contents of Chapter 3 and Chapter 5 will help to remove any confusion which still persists. It is interesting to note that several articles have appeared in the literature since 1972 trying to shed light on the classical information pattern problem, particularly the LQG version. These articles are discussed in Chapter 9. More unfortunately, when some authors have been concerned with non-classical information patterns a number of conceptual errors have arisen. These

errors, mentioned in Chapter 10, could perhaps have been avoided by adopting the careful approach taken here in Chapter 3 and 5. More encouraging, when authors have been successful in examining non-classical information problems, such as delay sharing information patterns, they have adopted a dynamic programming approach very much akin to that presented in Chapter 3.

In summary, if we had to produce a fly-sheet sales pitch describing the merits of this thesis it would probably be composed as follows:-

i) Presents a self contained conceptual introduction to the role which the information pattern plays in a stochastic control problem. Readable by anybody who wants to know more about this very important topic.

ii) Presents a dynamic programming approach which can, with appropriate modification, be applied to a number of problems involving non-classical information patterns. Emphasises the role of probability distributions as 'state variables' in the dynamic programming framework. Offers further insight into the classical problem as well as the zero-memory problem.

iii) Presents examples of various approximation procedures which can be employed when the classical results are not applicable. Perturbation analysis is given extensive examination.

Notice that it is not claimed that this thesis is either a survey or a bibliography. For a start, in 1972 there was little to survey of direct relevance anyway. What there was has been cited where appropriate. Secondly, very extensive surveys of the topics considered here have appeared recently in the literature (for example, Sandell et al (48)). Our prime concern has been to present a self contained introduction to the role of the information pattern of a problem and to offer some approximation techniques should the designer require them.

CHAPTER 2

An Introductory Example

		Page
2.1.	Introduction	2.1
2.2.	Simple Two-Stage Example	2.6
	2.2.1. Classical Information Pattern	2.6
	2.2.2. Nested Observation Case	2.17
	2.2.3. Zero Memory Problem	2.25
2.3.	Summary and Conclusions	2.46

2.1 Introduction

As mentioned in the previous introductory chapter we shall not adopt the conventional approach of formulating a fairly general problem, proposing a correspondingly general technique for its solution, and then specialising the problem and its solution to a particular example. Instead a reverse approach will be taken. First, three seemingly very simple problems will be examined. These problems will in fact be identical except for the fact that the controller in each problem will have available a different information pattern. In the first problem the controller will have available all past and present observations, and all past values of control (the classical information pattern). In the second problem, the controller will have available all past and present observations (the nested observation problem). In the third problem the controller will have only the present observation (the zero memory problem). These examples will illustrate very graphically just how significant an effect the information pattern of a problem can have on the complexity of its solution.

What is more we shall attempt to develop solutions to these problems by a very direct approach. The techniques and concepts which we thus introduce will then be applied in developing a more general dynamic programming formulation of the stochastic optimal control problem in Chapter 3.

It will be seen in the general description of the problem which follows, that we are studying nothing more than a Linear system, with Quadratic cost, and Gaussian noise (the so-called LQG problem). Why then should an entire chapter be devoted to the solution of a problem which, at least for the classical information pattern, is reputed to be

well understood ? The reason is simple. The problem is apparently not well understood at all. Further comment will be made on this point in the conclusions to this chapter. For the moment let us merely remark that in the current literature, text books especially, several approaches are presented to the problem of determining the optimal control laws for the LQG system. All too frequently these approaches are based on a variety of misconceptions, which although for the LQG problem with classical information pattern produce the right answers, soon become apparent once a non-classical information pattern is studied. In some cases, the information pattern is not even clearly stated.

The purpose of this chapter then is to introduce various key concepts by means of a concrete example. There is also one other important function. It is to describe an example which will receive considerable attention in later Chapters - this is the so-called Witsenhausen problem mentioned in the introductory chapter.

Let us first define the process we wish to control -

$$x_{n+1} = x_n + u_n + w_n \quad \text{---} \quad \text{dynamics equation - (2.2.1.)}$$

$$y_n = x_n + v_n \quad \text{---} \quad \text{observation equation - (2.2.2.)}$$

Here x_n is the value of the state at time n which for simplicity is scalar, u_n is the value of control applied at time n. The variables $\{w_n \mid n = 0, 1, 2 \cdots\}$ introduce a noise component into the dynamics. They represent a sequence of independent Gaussian random variables with:-

$$P\left\{w_n \leq \Omega_n\right\} = \int_{-\infty}^{\Omega_n} \frac{\Phi}{\sigma_w}\left\{\frac{t}{\sigma_w}\right\} dt$$

where,

$$\Phi(t) = \frac{exp}{\sqrt{2\pi}} \left\{ \frac{-t^2}{2} \right\}$$

The variable y_n is the observation of the state of the process at time n. The variables $(v_n | n = 0, 1, 2 - - -)$ introduce a noise component into the observation equation and again they are a sequence of independent Gaussian random variables with:-

$$P\left\{ v_n \leq \psi_n \right\} = \int_{-\infty}^{\psi_n} \frac{\Phi}{\sigma_v} \left\{ \frac{t}{\sigma_v} \right\} dt$$

The initial state of the system x_0 cannot be observed exactly since the observation is corrupted by noise. It is assumed that x_0 is an independent Gaussian random variable with:-

$$P\left\{ x_0 \leq X_0 \right\} = \int_{-\infty}^{X_0} \frac{\Phi}{\sigma_0} \left\{ \frac{t}{\sigma_0} \right\} dt$$

The process can be described in the traditional block diagram form as below:-

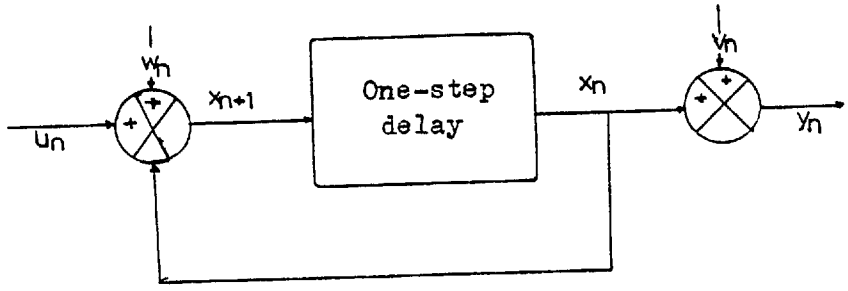


Fig. 2.1. Stochastic Optimal Control Problem

For simplicity we shall consider a 'two stage' process, where a control value u_0 is applied at time 0 and a control value u_1 is applied at time 1. The terminal state is x_2.

It is convenient for discussion purposes to visualise the control situation as being composed of two distinct controllers. Controller 0 acts at time 0 and has available some control data I_0 which he will utilise to determine the appropriate control action u_0 by means of his control policy γ_0 ; that is:-

$$u_0 = \gamma_0(I_0)$$

Similarly Controller 1 acts at time 1 and has available some control data I_1 which will be used to determine control action u_1, by means of control policy γ_1 that is:-

$$u_1 = \gamma_1(I_1) .$$

We are now in a position to state the optimal control problem. We wish to find those control policies γ_0 and γ_1 such that the expected value of the following expression is minimised:-

$$x_2^2 + k^2 u_0^2$$

The variables x_0, w_0, w_1, v_0, and v_1 are the so-called primitive random variables. We shall now examine three versions of the problem as loosely defined above. As mentioned earlier, these versions will be generated by imposing three different structures on the control data I_0 and I_1 thus creating three different information patterns.

In the analysis which follows various expected values will be written explicitly in terms of the appropriate probability density functions. This is generally a rather long-winded approach. However it will be demonstrated later that, in a dynamic programming formulation of the stochastic optimal control problem, these probability density functions assume the role of the 'state' variable. Hence, at the risk of being tedious, these probability density functions are stated quite explicitly.

2.2 Simple Two-Stage Example

2.2.1 Classical Information Pattern

Basically, in the classical information pattern version of the problem, it is assumed that Controller 1 knows what Controller 0 observed and precisely what he did as a result of that observation. Thus:-

$$I_0 = (y_0)$$

$$I_1 = (y_1, y_0, u_0)$$

This information pattern is illustrated in the diagram below.

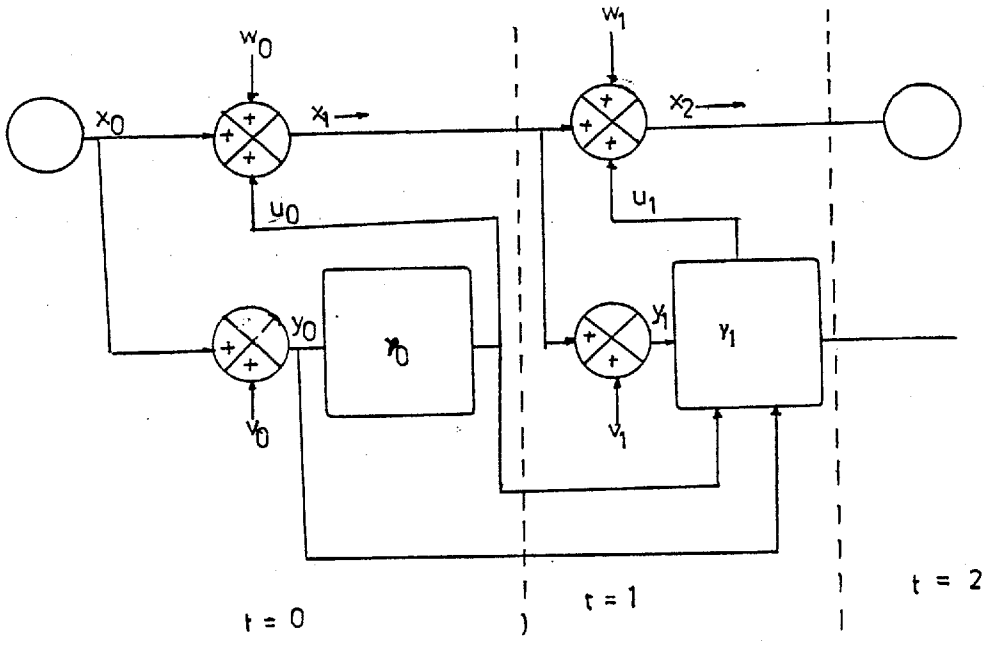


Fig. 2.2 Stochastic Optimal Control Problem - Classical Information Pattern

Compare this diagram with the traditional form of diagram shown earlier. The difference between the two forms of representation is that in this case the evolution of the process as time increases is explicitly shown. The obvious disadvantage of this representation is that the diagram rapidly becomes unwieldy as more stages of the process are considered. It does however, have the considerable advantage that the information pattern and the dynamic and observation process can be demonstrated jointly.

First consider the problem confronting Controller 1. The additive nature of the cost function and the fact that nothing that Controller 1 can do will influence the value of 'previous' random variables means that Controller 1 can restrict his attention to the task of minimising the expected value of x_2^2 . More precisely, Controller 1 wishes to apply a control value u_1 such that $E[x_2^2 \mid y_0, y_1, u_0, u_1]$ is minimised. It is our problem to find him a policy γ_1 for determining u_1 from his available control data (y_0, y_1, u_0); that is, $u_1 = \gamma_1(y_0, y_1, u_0)$

Consider some arbitary control value u_1 applied by Controller 1. Let us examine how a choice of u_1 influences $E[x_2^2 \mid y_0, y_1, u_0]$

First

$$E[x_2^2 \mid y_0, y_1 u_0] = \int_{-\infty}^{+\infty} x_2^2 p(x_2 \mid y_0, y_1, u_0, u_1)\, dx_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} x_2^2 p(x_2 \mid x_1, y_0, y_1, u_0, u_1)\, p(x_1 \mid y_0, y_1, u_0, u_1)\, dx_1\, dx_2$$

Some comment on notation is in order here. The function referred to by the notation $p(x_2 \mid x_1, y_0, y_1, u_0, u_1)$ is the probability density function of the -

random variable x_2 conditioned on the random variables x_1, y_0, y_1, u_1, u_0.

Usually in writing a probability density function as $p(x)$, we will mean the probability density function of the random variable x. When the identity of the corresponding random variable is not clear from the argument of the probability density function, we shall explicitly write the probabiltiy density function as $p_x(\ )$.

Now $\int_{-\infty}^{X_2} R_{X_2}(t \mid x_1, y_0, y_1, u_0, u_1)\, dt$

$= P\left\{ x_2 \leq X_2 \mid x_1 = X_1,\ y_0 = Y_0,\ y_1 = Y_1,\ u_0 = U_0, u_1 = U_1 \right\}$

$= P\left\{ w_1 \leq X_2 - (U_1 + X_1) \mid x_1 = X_1, y_0 = Y_0, y_1 = Y_1, u_0 = U_0, u_1 = U_1 \right\}$, from the dynamics equation

$= P\left\{ w_1 \leq X_2 - (U_1 + X_1) \right\}$, since w_1 is an independent random variable.

$= \int_{-\infty}^{X_2 - (X_1 + U_1)} \frac{\Phi}{\sigma_w}\left\{ \frac{t}{\sigma_w} \right\} dt$

$= \int_{-\infty}^{X_2} \frac{\Phi}{\sigma_w}\left\{ \frac{t' - (X_1 + U_1)}{\sigma_w} \right\} dt'$

This leads to an interpretation of $E[x_2^2 \mid y_0, y_1, u_0, u_1]$ as follows:-

$$\underset{\gamma}{E}[x_2^2 \mid y_0, y_1, u_0, u_1] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_2^2 \frac{\Phi}{\sigma_w}\left\{\frac{x_2-(x_1+u_1)}{\sigma_w}\right\} p(x_1 \mid y_0, y_1, u_0, u_1)\, dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} (\sigma_w^2 + u_1^2 + x_1^2 + 2x_1u_1)\, p(x_1 \mid y_0, y_1, u_0, u_1)\, dx_1$$

The p.d.f. $p(x_1 \mid y_0, y_1, u_0, u_1)$ is independent of u_1 since u_1 cannot influence x_1.

Thus,

$$E[x_2^2 \mid y_0, y_1, u_0, u_1] = u_1^2 + \int_{-\infty}^{\infty} x_1^2\, p(x_1 \mid y_0, y_1, u_0)\, dx_1 + 2u_1\int_{-\infty}^{\infty} x_1\, p(x_1 \mid y_0, y_1, u_0)\, dx_1 + \sigma_w^2$$

This expression is minimised when

$$u_1 = -\int_{-\infty}^{\infty} x_1\, p(x_1 \mid y_1, y_0, u_0)\, dx_1, \text{ or alternatively}$$

$$\gamma_1(y_0, y_1, u_0) = -E[x_1 \mid y_1, y_0, u_0]$$

and the corresponding minimal cost is $E[x_1^2 \mid y_1, y_0, u_0] - \left\{E[x_1 \mid y_1, y_0, u_0]\right\}^2 + \sigma_w^2$. It must be appreciated that the above deductions regarding the optimal policy γ_1 in no way solve our problem. The control value u_1 is still totally unknown because it depends upon some function $p(x_1 \mid y_1, y_0, u_0)$. Can we compute this function ? The answer is affirmative. Consider the physical significance of $p(x_1 \mid y_1, y_0, u_0)$ - it is the conditional probability density function of x_1 given y_0, y_1 and u_0.

Now:-

$$p(x_1 \mid y_1, y_0, u_0) = \frac{p(x_1, y_1 \mid y_0, u_0)}{\int p(x_1, y_1 \mid y_0, u_0)\, dx_1}$$

Thus:-

$$p(x_1 \mid y_1, y_0, u_0) = \frac{p(y_1 \mid x_1, y_0, u_0)\, p(x_1 \mid y_0, u_0)}{\int_{-\infty}^{\infty} p(y_1 \mid x_1, y_0, u_0)\, p(x_1 \mid y_0, u_0)\, dx_1}$$

$$= \frac{p(y_1 \mid x_1, y_0, u_0) \int_{-\infty}^{\infty} p(x_1 \mid x_0, y_0, u_0)\, p(y_0 \mid x_0, u_0)\, p(x_0)\, dx_0}{p(y_1 \mid x_1, y_0, u_0) \int_{-\infty}^{\infty} p(x_1 \mid x_0, y_0, u_0)\, p(y_0 \mid x_0, u_0)\, p(x_0)\, dx_0 dx_1}$$

The probability density functions involved in this expression can all be determined from the dynamic and observation equations describing the process. Thus we have:-

i) $$\int_{-\infty}^{x_1} p_{x_1}(t \mid x_0\ y_0\ u_0)\, dt = \int_{-\infty}^{x_1} \frac{\Phi}{\sigma_w}\left\{\frac{t - (x_0 + u_0)}{\sigma_w}\right\} dt,$$

by using a similar argument as that used to determine $p(x_2 \mid x_1, y_1, y_0, u_1, u_0)$.

This implies:-

$$p(x_1 \mid x_0\ y_0\ u_0) = \frac{\Phi}{\sigma_w}\left\{\frac{x_1 - (x_0 + u_0)}{\sigma_w}\right\}$$

ii) $$\int_{-\infty}^{y_1} p_{y_1}(t \mid x_1, y_0, u_0)\, dt = P(y_1 \le Y_1 \mid x_1 = X_1, y_0 = Y_0, u_0 = U_0)$$

$$= P(x_1 + v_1 \le Y_1 \mid x_1 = X_1, y_0 = Y_0, u_0 = U_0],$$

by the observation equation

$$= P(v_1 \le Y_1 - X_1],$$

since v_1 is an independent random variable.

$$= \int_{-\infty}^{Y_1 - X_1} \frac{\phi}{\sigma_v}\left\{\frac{s}{\sigma_v}\right\} ds$$

$$= \int_{-\infty}^{Y_1} \frac{\phi}{\sigma_v}\left\{\frac{t - X_1}{\sigma_v}\right\} dt$$

Thus:-

$$p(y_1 | x_1, y_0, u_0) = \frac{\phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\}$$

iii) By a similar argument to (ii) it is easy to show that,

$$p(y_0 | x_0, u_0) = \frac{\phi}{\sigma_v}\left\{\frac{y_0 - x_0}{\sigma_v}\right\}$$

In what follows, the above substitutions for the particular problem we are studying here will be introduced without comment.

Thus for the process under consideration,

$$p(x_1 | y_1, y_0, u_0) = \frac{\frac{\phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \int_{-\infty}^{\infty} \frac{\phi}{\sigma_w}\left\{\frac{x_1 - (x_0 + u_0)}{\sigma_w}\right\} \frac{\phi}{\sigma_v}\left\{\frac{y_0 - x_0}{\sigma_v}\right\} \frac{\phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\} dx_0}{\int_{-\infty}^{\infty} \frac{\phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \int_{-\infty}^{\infty} \frac{\phi}{\sigma_w}\left\{\frac{x_1 - (x_0 + u_0)}{\sigma_w}\right\} \frac{\phi}{\sigma_v}\left\{\frac{y_0 - x_0}{\sigma_v}\right\} \frac{\phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\} dx_0 dx_1}$$

Using the simple algebraic identity:-

$$\frac{\phi}{\sigma_a}\left\{\frac{x-a}{\sigma_a}\right\}\frac{\phi}{\sigma_b}\left\{\frac{x-b}{\sigma_b}\right\} = \frac{\phi}{\sigma_m}\left\{\frac{x-\frac{a\sigma_b^2+b\sigma_a^2}{\sigma_b^2+\sigma_a^2}}{\sigma_m}\right\}\frac{\phi}{\sigma_n}\left\{\frac{a-b}{\sigma_n}\right\}$$

where $\sigma_m^2 = \sigma_a^2\sigma_b^2/(\sigma_a^2+\sigma_b^2)$, and $\sigma_n^2 = \sigma_a^2+\sigma_b^2$,

the above expression simplifies to:-

$$p(x_1|y_1,y_0,u_0) = \frac{\Phi}{\sigma_{1/1}}\left\{\frac{x_1 - \hat{x}_{1/1}}{\sigma_{1/1}}\right\},$$

where $\hat{x}_{1/1} = (y_1\sigma_{1/0}^2 + x_{1/0}\sigma_v^2)/(\sigma_{1/0}^2+\sigma_v^2)$, $\sigma_{1/1}^2 = \sigma_{1/0}^2\sigma_v^2/(\sigma_{1/0}^2+\sigma_v^2)$,

$$\hat{x}_{1/0} = \hat{x}_{0/0} + u_0, \quad \sigma_{1/0}^2 = \sigma_{0/0}^2 + \sigma_w^2,$$

$$\hat{x}_{0/0} = y_0\sigma_0^2/(\sigma_0^2+\sigma_v^2), \quad \sigma_{0/0}^2 = \sigma_v^2\sigma_0^2/(\sigma_0^2+\sigma_v^2)$$

Thus $\gamma_1^{\circ}(y_1,y_0,u_0) = -\hat{x}_{1/1}$

$$= \frac{-\left[y_1\left\{\frac{\sigma_v^2\sigma_0^2}{\sigma_0^2+\sigma_v^2}+\sigma_w^2\right\} + u_0\sigma_v^2 + \frac{y_0\sigma_v^2\sigma_0^2}{\sigma_0^2+\sigma_v^2}\right]}{\left[\sigma_v^2 + \frac{\sigma_v^2\sigma_0^2}{\sigma_0^2+\sigma_v^2} + \sigma_w^2\right]}$$

and

$$E\left[x_2^2|y_1,y_0,u_0,\gamma_1^{\circ}(y_1,y_0,u_0)\right] = E\left[x_1^2|y_1,y_0,u_0\right] - E^2\left[x_1|y_1,y_0,u_0\right] + \sigma_w^2$$

$$= \sigma_{1/1}^2 + \sigma_w^2$$

Consider next the problem facing Controller 0. He must select his policy γ_0° (so that his action $u_0^{\circ} = \gamma_0^{\circ}(y_0)$) minimises the expected value of $x_2^2+k^2u_0^2$, given the control data y_0.

Let us evaluate $E(x_2^2+k^2u_0 \mid y_0,u_0)$.

$$E(x_2^2+k^2u_0^2 \mid y_0,u_0)$$

$$= \int x_2^2 p(x_2 \mid y_0) dx_2 + E(k^2u_0^2 \mid y_0,u_0)$$

$$= \int x_2^2 p(x_2 \mid y_0,y_1,u_0,u_1) p(y_1,u_1 \mid u_0,y_0) dx_2 dy_1 du_1 + k^2u_0^2$$

$$= \int x_2^2 p(x_2 \mid y_0,y_1,u_0,u_1) p(u_1 \mid y_1,u_0,y_0) p(y_1 \mid y_0,u_0) dx_2 dy_1 du_1 + k^2u_0^2$$

In its present form this expression is rather meaningless. What do we mean for example by the probability density function $p(u_1 \mid y_1,y_0,u_0)$? To obtain the correct interpretation we must introduce a control policy γ_1 for Controller 1, where $u_1 = \gamma_1(y_0,y_1,u_0)$.

Then we obtain

$$p(u_1 \mid y_1,y_0,u_0) = \delta(u_1 - \gamma_1(y_0,y_1,u_0)), \ \delta \text{ the Dirac delta function}$$

and

$$E(x_2^2 + k^2u_0^2 \mid y_0u_0)$$

$$= \int x_2^2 p(x_2 \mid y_0,y_1u_0, \gamma_1(y_1,y_0,u_0)) p(y_1 \mid y_0,u_0) dy_1 dx_2 + k^2u_0^2$$

Instead of assuming an arbitrary control policy γ_1 for Controller 1, Controller 0 knows the exact policy γ_1^0 which Controller 1 will adopt. This has already been evaluated. What is more:-

$$\int x_2^2 p(x_2 \mid y_0, y_1, u_0, \gamma_1^0(y_1, y_0 u_0))\, dx_2 = \sigma_{1/1}^2 + \sigma_w^2$$

Thus we have:-

$$E(x_2^2 + k^2 u_0^2 \mid y_0, u_0) = \sigma_{1/1}^2 + \sigma_w^2 + k^2 u_0^2$$

This expression is minimised when $u_0 = \gamma_0^0(y_0) = 0$ for all y_0. Let us summarise the results we have obtained:-

(i) Controller 0 does nothing. He leaves all the work for Controller 1 to do.

(ii) Controller 1 adopts a control policy of the form:-

$$\gamma_1^0(y_1, y_0, u_0) = -\hat{x}_{1/1}$$

$$= = -\frac{\left[y_1 \left\{ \frac{\sigma_v^2 \sigma_0^2}{\sigma_0^2 + \sigma_v^2} + \sigma_w^2 \right\} + \frac{y_0 \sigma_v^2 \sigma_0^2}{\sigma_0^2 + \sigma_v^2} \right]}{\left[\sigma_v^2 + \frac{\sigma_v^2 \sigma_0^2}{\sigma_v^2 + \sigma_0^2} + \sigma_w^2 \right]}$$

(iii) The optimal value of the cost for the problem is:-

$$\bar{\sigma}_{1/1}^2 + \sigma_w^2$$

where:- $\bar{\sigma}_{1/1}^2$ is defined on page 2.12

We have derived the above results for an extremely simple system. We have also explicitly stated the problem in terms of the underlying probability density functions. The reason for this is that later in Chapter 3 we shall introduce a general dynamic programming formulation of the problem, in which the 'state' of the system is represented by the probability density function of the process variable given the control data I.

Before we look at a problem with a slightly different information pattern, let us recall the significant steps in the analysis.

1. Consider Controller 1. Evaluate $E(x_2^2 | y_0, y_1, u_0, u_1)$ in terms of u_1 and $p(x_1 | y_0, y_1, u_0)$. Minimisation with respect to u_1 leads to an expression for u_1 in terms of a functional involving $p(x_1 | y_0, y_1, u_0)$. It turns out that the control data y_0, y_1, u_0 is sufficient to enable $p(x_1 | y_0, y_1, u_0)$ to be determined. It is a Gaussian density function with the mean given by a simple linear function of the variables y_0, y_1, u_0 and the variance inde-

pendent of y_0, y_1 and u_0 . This enables u_1 to be determined as a function γ_1° of y_1, y_0, and u_0 .

2. Consider Controller 0. He must evaluate $E(x_2^2+k^2u_0^2 \mid y_0,u_0)$. This expression is however meaningless unless Controller 0 assumes some control policy γ_1 for Controller 1. The sensible assumption is that Controller 1 applies an optimal policy γ_1° , as calculated above. This means that - $E(x_2^2+k^2u_0^2 \mid y_0,u_0)$ can be written as follows:-
γ_1°

$$E_{\gamma_1^\circ}(x_2^2+k^2u_0^2 \mid y_0,u_0) = \int x_2^2 p_{\gamma_1^\circ}(x_2 \mid y_0,u_0)\,dx_2+k^2u_0^2$$

$$= \int x_2^2 p(x_2 \mid y_0,u_0,y_1,\gamma_1^\circ(y_0,y_1,u_0))\; p(y_1 \mid y_0,u_0)\,dy_1\,dx_2+k^2u_0^2$$

But $\int x_2^2 p(x_2 \mid y_0,u_0,y_1,\gamma_1^\circ(y_0,y_1,u_0))\,dx_2$ was evaluated in stage 1 above. Substituting this value into the above expression, integrating with respect to the observation y_1 and then minimising with respect to u_0 yields the optimal value of the control policy $\gamma_0^\circ(y_0)$.

2.2.2 Nested Observation Case

The version of the problem to be discussed here will differ from that of the previous section only in the information pattern. Now it is assumed that:-

$$I_0 = (y_0)$$

$$I_1 = (y_1, y_0)$$

In other words, Controller 0 knows his observation y_0 . Controller 1 knows his own observation y_1 and Controller 0's observation y_0. Unlike the classical information pattern case, however, Controller 1 does not know what Controller 0 did as a result of observation y_0 . This situation is illustrated in the diagram below:-

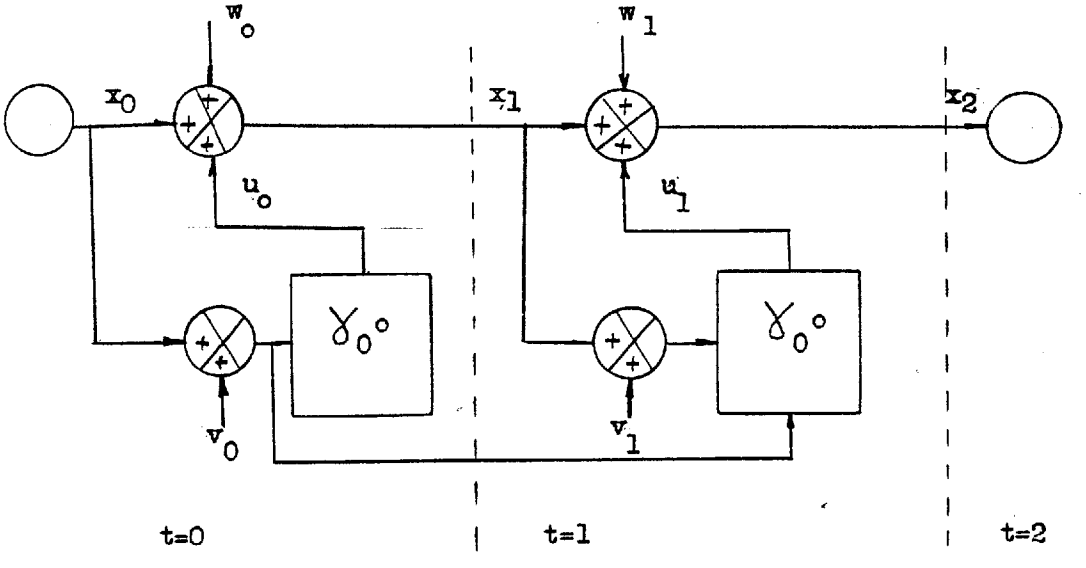


Fig. 2.3 Stochastic Optimal Control Problem - Nested Observation Information Pattern.

Let us examine Controller 1's approach to the problem. He wishes to minimise $E(x_2^2|y_1,y_0,u_1)$ with respect to u_1. As before:-

$$E(x_2^2 \mid y_1,y_0,u_1) = \int x_2^2 p(x_2 \mid y_1,y_0,u_1)\,dx_2$$

$$\text{and } p(x_2 \mid y_1,y_0,u_1) = \int p(x_2 \mid x_1,y_1,y_0,u_1)\; p(x_1 \mid y_1,y_0)\,dx_2dx_1$$

$$\text{Thus, } E(x_2^2 \mid y_1,y_0,u_1) = \int x_2^2 \frac{\Phi}{\sigma_w}\left\{\frac{x_2-(x_1+u_1)}{\sigma_w}\right\} p(x_1|y_1,y_0)\,dx_2dx_1$$

$$= \sigma_w^2 + \int(x_1^2+u_1^2+2x_1u_1)\; p(x_1 \mid y_1,y_0)\,dx_2dx_1$$

Again, this expression is minimised when:-

$$u_1 = -\int x_1 p(x_1 \mid y_1,y_0)\,dx_2dx_1$$

$$= -E(x_1 \mid y_1,y_0)$$

with corresponding

$$E(x_2^2 \mid y_1,y_0) = E(x_1^2 \mid y_1,y_0) - E^2(x_1 \mid y_1,y_0)+\sigma_w^2$$

Just as in the classical information pattern case, Controller 1 still has the problem of determining $p(x_1|y_1,y_0)$ in order to compute u_1. This probability density function cannot however be determined unless the control policy γ_0 adopted by Controller 0 is known to Controller 1. Suppose we assume an arbitrary control policy γ_0^a. Thus the control

value $u_0 = \gamma_0^a(y_0)$ is also assumed and we can parameterise the unknown probability density function with $\gamma_0^a(y_0)$, representing it as $p(x_1 | y_1, y_0, \gamma_0^a(y_0))$. But this is precisely the probability density evaluated in the previous section, and we know that:-

$$p(x_1 | y_1, y_0, \gamma_0^a(y_0)) = \frac{\Phi}{\sigma_{1/1}} \left\{ \frac{x - \hat{x}_{1/1}}{\sigma_{1/1}} \right\} ,$$

with $x_{1/1}$ and $\sigma_{1/1}$ as defined on page 2.12 (with the exception that $\gamma_0^q(y_0)$ replaces u_0).

Thus:- $$u_1^o = \gamma_1^o(y_1, y_0, \gamma_0^a(y_0)) = - \hat{x}_{1/1}$$

$$- \frac{\left[y_1 \left\{ \frac{\sigma_v^2 \sigma_o^2}{\sigma_v^2 + \sigma_o^2} + \sigma_w^2 \right\} + \gamma_o^a(y_o) \sigma_v^2 + \frac{y_o \sigma_v^2 \sigma_o^2}{\sigma_o^2 + \sigma_v^2} \right]}{\left[\sigma_v^2 + \frac{\sigma_v^2 \sigma_o^2}{\sigma_v^2 + \sigma_o^2} + \sigma_w^2 \right]}$$

Notice that this does not solve Controller 1's problem. All we have done is to parameterise his policy with an arbitrary control value $\gamma_0^a(y_0)$. Next we examine Controller 0's situation. His control data is simply y_0 and precisely the same arguments used in the classical information pattern version can be used to determine γ_0^o.

$$E(x_2^2 \mid y_0,u_0) = \int x_2^2\, p(x_2 \mid y_0,u_0)\,dx_2$$
$$\gamma_1^o,\gamma_o^a(y_0) \qquad\qquad \gamma_1^o,\gamma_0^a(y_0)$$

$$= \iint x_2^2 p(x_2 \mid y_0,y_1,\gamma_0^a(y_0))\,p(y_1 \mid y_0,u_0)\,dx_2\,dy_1$$
$$\gamma_1^o$$

This expression was evaluated by Controller 1. Its value is $\sigma_{1/1}^2 + \sigma_w^2$

Thus the optimal value u_0^o can be determined as a function of y_0 and $\gamma_0^a(y_0)$.

that is:-

$$u_0^o = \bar{\gamma}_0^o(y_0,\gamma_0^a(y_0))$$

In deriving this policy $\bar{\gamma}_0^o$ we have assumed that Controller 1 assumes an arbitrary control policy γ_0^a for Controller 0. Rather than assume an arbitrary policy, Controller 1 can in fact assume the optimal policy which we have just computed. If we put $u_0^a = \gamma_0^a(y_0)$, then we obtain $u_0^o = \bar{\gamma}_0^o(y_0,u_0^o)$, which leads to the equations

$$u_0^o = \gamma_0^o(y_0)$$

and $$u_1^o = \bar{\gamma}_1^o(y_1,y_0,\gamma_0^o(y_0)) = \gamma_1^o(y_1,y_0)$$

Thus we see that in the nested observation case, an optimal solution can be derived almost as easily as in the classical case. In fact the easiest way to solve the nested observation problem is

to solve the classical problem first, obtaining control laws - $\overline{\gamma}_0^\circ(y_0)$ and $\overline{\gamma}_1^\circ(y_1, y_0, u_0)$. We then simply insert $u_0 = \overline{\gamma}_0^\circ(y_0)$ in the equation for $\overline{\gamma}_1^\circ$ to obtain the optimal control laws γ_0° and γ_1° for the nested observation case:-

that is $$\gamma_0^\circ(y_0) = \overline{\gamma}_0^\circ(y_0),$$

and $$\gamma_1^\circ(y_0, y_1) = \overline{\gamma}_1^\circ(y_0, y_1, \overline{\gamma}_0(y_0))$$

Thus we have demonstrated a very close mathematical equivalence between the classical information pattern problem and the nested observation problem. The optimal control laws for the nested observation problem can be trivially obtained from the classical information pattern problem by simply following the substitution procedure described above. In our analysis of the problem we have deliberately attempted to determine the control laws directly without consideration of the classical problem. This approach was adopted in order to highlight a fundamental difference between the classical information pattern problem and the nested observation problem. In the classical case Controller 1 need make no assumptions concerning the behaviour of Controller 0. All that Controller 1 wants to know is what control value Controller 0 applied. He is not interested in whether or not this control value is the optimal one given Controller 0's observation. Even if Controller 0 does not behave optimally, it makes no difference to Controller 1's judgement of what constitutes an optimal control law for Controller 1.

This is because $p(x_1|y_1,y_0,u_0)$ is independent of γ_0. In the nested observation case Controller 1 can only determine his policy if he knows the policy adopted by Controller 0. In our analysis, Controller 1's optimal policy had to be parameterised by an arbitrary policy adopted by Controller 0. In fact, in the nested observation case, the control policy of Controller 1 must contain an unknown variable representing what Controller 1 thinks Controller 0 has done. This unknown variable had to be introduced into the derivation of the control policy for Controller 0. However, Controller 0 removed the mystery about the variable by making the assumption that Controller 1's opinion of what Controller 0 has done will be correct, i.e. $u_0^o = u_0^a$. This is not an unreasonable assumption. It is simply based on the prior assumption that Controller 1 believes Controller 0 will act optimally. Thus, if γ_0^o is the optimal control policy applied by Controller 0, i.e. - $u_0^o = \gamma_0^o(y_0)$, then Controller 1 will put $\gamma_1^a = \gamma_0^o$, thus obtaining $\gamma_0^a(y_0) = \gamma_0^o(y_0) = u_0^o$. For Controller 0 this yields the equation:-

$$u_0^o = \bar{\gamma}_0^o(y_0, u_0^o)$$

which provides a method of determining γ_0^o and consequently γ_1^o. Suppose we consider the case when Controller 0 does not act optimally, for example suppose the control computer breaks down and a control value $\bar{u}_0^a$ is applied. In this case Controller 1 can still calculate $u_0^o = \gamma_0^o(y_0)$ and can thus evaluate the probability density function $p(x_1 \mid y_1,y_0,u_0^o)$. However this will lead to the wrong calculation of the optimal control value, since this calculation should be based on $p(x_1 \mid y_1,y_0,\bar{u}_0^a)$, not $p(x_1 \mid y_1,y_0,u_0^o)$. This possibility does not arise the case of the classical information pattern problem because Controller 1

always knows what Controller 0 has done, irrespective of whether it is right or wrong, and can thus compute his optimal control based on $p(x_1 \mid y_1, y_0, \bar{u}_0^a)$.

This difference between the classical information pattern problem and the nested observation problem is analogous to the difference between open loop and closed loop optimal control in the deterministic case. Mathematically, the two problems are equivalent and there is no apparent advantage in including in the control data the previous control values. From a control engineering point of view, however, there is considerable advantage in knowing the previous control values, since then the optimal control policy at some time n becomes immune to deviations in the application of the appropriate optimal control value at times prior to n. Another way of viewing this situation is to regard the probability density function $p(x_1 \mid y_1, y_0, u_0)$ as the 'state' of the system in the classical case. On the other hand, in the nested observation case $p(x_1 \mid y_1, y_0, \gamma_0^o(y_0))$ is an anticipated value of the state assuming the actual control applied at time 0 equals $\gamma_0^o(y_0)$.

The above discussion raises the question of modelling situations in which the control computer is not reliable. Suppose:-

$$u_0^o = \gamma_0^o(y_0) + w_0$$

where w_0 is some random noise. Then w_0 can be included in the dynamic equations of the system and the control computer can be assumed to be perfectly reliable.

Similarly suppose:-

$$u_0^o = \gamma_0^o(y_0 + v_0),$$

where v_0 is some random noise again. In this case v_0 can be included in the observation equations of the system and again the control computer can be assumed to be perfectly reliable.

It is, however, possible to visualise instances in which the unreliability of the computer control cannot be modelled in this way. In this case it might be more appropriate for Controller 1 to alter his assumption that Controller 0 acts optimally. For example, Controller 1 might wish to base his calculation of his optimal control on the assumption that Controller 0 attempts to maximise the cost function. Controller 1 is thus engaged in a game-like situation, (although Controller 0 is not playing the same game). In other words, Controller 1 is trying to minimise the maximum damage that Controller 0 can do.

This completes our discussion of the nested observation case. It is hoped that the comments made above have illustrated that the nested observations case and the classical information pattern case are by no means trivially equivalent as is often implied in the literature. What is more, we have introduced rather simply the element of second-guessing which characterises the zero-memory problem to be discussed in the next section. Second-guessing in the context of the nested observation case means Controller 1 must somehow determine an estimate of what Controller 0 has done, and then Controller 0 must determine his own estimate of what Controller 1 thinks Controller 0 has done. These ideas will be made more explicit in the next section.

2.2.3. Zero-Memory Problem

Again, the version of the problem to be discussed here will differ from the two cases considered in the previous section only in the information pattern. Now it is assumed that:-

$$I_0 = y_0$$
$$I_1 = y_1$$

In other words, each Controller only knows his own observation of the process. There is no transmission of information from Controller 0 to Controller 1. If we visualise the Controller as a single Controller acting at successive time intervals, then the observation at time 0 is not known at time 1. Thus, this type of information pattern is sometimes referred to as zero memory. The situation is illustrated below:-

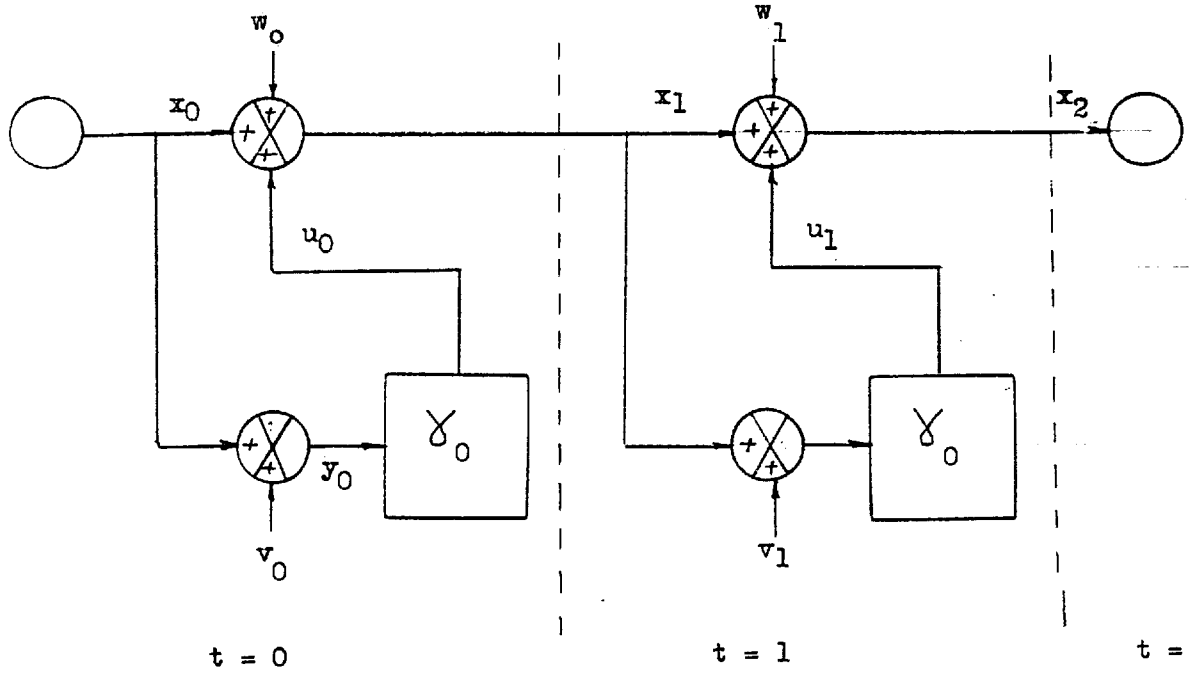


Fig. 2.4 Stochastic Optimal Control Problem - Zero Memory Information Pattern.

We shall adopt precisely the same approach to the problem as in the sections dealing with the classical information pattern problem and the nested observation problem. In this case however, we shall consider two possible approaches to the problem. We shall show that, whereas in the previous problems these approaches are equivalent, in the zero memory case they are not equivalent.

2.2.3.1. Direct Approach

We use the term 'direct' here simply to describe the problem which Controller 0 poses himself. We shall assume that Controller 0 sets himself the problem of minimising $E[x_2^2 \mid y_0]+k^2 u_0^2$, and he assumes that Controller 1 acts optimally.

Consider Controller 1's approach to the problem. He wishes to minimise $E[x_2^2 \mid y_1,u_1]$ with respect to u_1. As before:-

$$E[x_2^2 \mid y_1,u_1] = \int x_2^2\, p(x_2 \mid y_1)\,dx_2$$

$$= \int x_2^2\, p(x_2 \mid x_1,y_1,u_1)\; p(x_1 \mid y_1)\,dx_1\,dx_2$$

$$= \int x_2^2\, \frac{\Phi}{\sigma_w}\left\{\frac{x_2-(x_1+u_1)}{\sigma_w}\right\} p(x_1 \mid y_1)\,dx_1\,dx_2$$

$$= \sigma_w^2 + \int (x_1^2+u_1^2+2x_1u_1)\; p(x_1 \mid y_1)\,dx_1$$

This expression is minimised when,

$$u_1 = u_1^{\circ} = -\int x_1 p(x_1 \mid y_1)\,dx_1,$$

with corresponding,

$$E[x_2^2 \mid y_1,u_1^{\circ}] = E[x_1^2 \mid y_1] - \left(E[x_1 \mid y_1]\right)^2 + \sigma_w^2$$

Just as in the previous cases Controller 1 still has the problem of determining $p(x_1 \mid y_1)$ in order to compute u_1°. Just as in the case of nested observations this probability density cannot be determined unless the control policy γ_0 adopted by Controller 0 is known to Controller 1. Let us assume an arbitrary policy γ_0^a. Then we can compute $p(x_1 \mid y_1, \gamma_0^a)$ as follows:-

$$p(x_1 \mid y_1, \gamma_0^a) = \frac{p(y_1 \mid x_1)p(x_1, \gamma_0^a)}{\int p(y_1 \mid x_1)p(x_1, \gamma_0^a)\,dx_1}$$

$$= \frac{p(y_1 \mid x_1) \int p(x_1 \mid x_0, y_0, \gamma_0^a)\, p(y_0 \mid x_0)p(x_0)\, dx_0\, dy_0}{\int p(y_1 \mid x_1) \int p(x_1 \mid x_0, y_0, \gamma_0^a)p(y_0 \mid x_0)p(x_0)\, dx_0\, dy_0\, dx_1}$$

$$= \frac{\frac{\Phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\}\int\frac{\Phi}{\sigma_w}\left\{\frac{x_1-(x_0+\gamma_0^a(y_0))}{\sigma_w}\right\}\frac{\Phi}{\sigma_v}\left\{\frac{y_0-x_0}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}\,dx_0\,dy_0}{\int\frac{\Phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\}\int\frac{\Phi}{\sigma_w}\left\{\frac{x_1-(x_0+\gamma_0^a(y_0))}{\sigma_w}\right\}\frac{\Phi}{\sigma_v}\left\{\frac{y_0-x_0}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}\,dx_0\,dy_0}$$

A fundamental difference should be noted between this expression and the corresponding one obtained for the nested observation case. In the zero memory case above, we have integrated over y_0. This means that in this case, the probability density function of x_1 given the control data must be parameterised by the entire function γ_0^a. In the nested observation case it was sufficient to parameterise this probability density function merely by the function value $\gamma_0^a(y_0)$.

Let us now consider the problem facing Controller 0. Suppose he determined his optimal policy as γ_0^o. Now Controller 1's optimal policy is determined once γ_0^o is defined. Thus Controller 0 can compute the expected value of the cost. $\underset{\gamma_1^o}{E}\left[x^2 \mid y_0, u_0\right] + k^2 u_0^2$

If we make the following definitions:-

$$N(\gamma_0^\circ, y_1, x_1) = \iint \frac{\phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x_1-(x_0+\gamma_0^\circ(y_0))}{\sigma_\omega}\right\}\frac{\phi}{\sigma_v}\left\{\frac{y_0-x_0}{\sigma_v}\right\}\frac{\phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}dx_0\,dy_0$$

$$M(\gamma_0^\circ(y_0), y_1, x_1, y_0) = \frac{\int \frac{\phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x_1-(x_0+\gamma_0^\circ(y_0))}{\sigma_\omega}\right\}\frac{\phi}{\sigma_v}\left\{\frac{y_0-x_0}{\sigma_v}\right\}\frac{\phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}dx_0}{\int \frac{\phi}{\sigma_v}\left\{\frac{y_0-x_0}{\sigma_v}\right\}\frac{\phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}dx_0}$$

Then we obtain

$$E_{\gamma_1^\circ}\left[x_2^2 \mid y_0, u_0\right] + k^2 u_0^2$$

$$= \left[\left\{\frac{\int x_1 N(\gamma_0^\circ, y_1, x_1)\,dx_1}{\int N(\gamma_0^\circ, y_1, x_1)\,dx_1}\right\}^2 + \sigma_\omega^2 - 2x_1\left\{\frac{\int x_1 N(\gamma_0^\circ, y_1, x_1)\,dx_1}{\int N(\gamma_0^\circ, y_1, x_1)\,dx_1}\right\} + x_1^2\right] M(\gamma_0^\circ(y_0), y_1, x_1, y_0)\,dx_1\,dy_1 + k^2\gamma_0^\circ(y_0)$$

The problem facing Controller 0 is to find that function γ_0° which minimises this expression, it is clearly a daunting prospect.

An example of a two stage stochastic control problem with a zero memory controller was first presented in the literature by Witsenhausen[1]. The problem studied by Witsenhausen is described by the following equations:-

$$\left.\begin{aligned} x_1 &= x_0 + u_1 \\ x_2 &= x_1 - u_2 \end{aligned}\right\} \quad \text{State Equations}$$

$$\left.\begin{aligned} y_0 &= x_0 \\ y_1 &= x_1 + v \end{aligned}\right\} \quad \text{Observation Equations}$$

$$k^2 u_0^{\,2} + x_2^{\,2} \quad \text{Cost Function}$$

$$\left.\begin{aligned} u_1 &= \bar{\gamma}_1(y_0) \\ u_2 &= \bar{\gamma}_2(y_1) \end{aligned}\right\} \quad \text{Admissable Controllers}$$

The problem is to determine $\bar{\gamma}_1$ and $\bar{\gamma}_2$ such that the expected value of the cost function is minimised. The observation noise v is assumed to be Gaussian with mean zero and variance σ_v^2. Witsenhausen studied two versions of the problem. One version corresponded to x_0 described by a two point symmetric probability distribution function. The second version corresponded to x_0 described by a Gaussian probability density function; this is the version of the problem we are primarily interested in here. We shall assume x_0 has a zero mean and variance σ_0^2.

1. H.S. Witsenhausen "A Counter-Example in Stochastic Optimal Control". (56)

We can reduce our formulation of the zero memory problem to Witsenhausen's problem as follows:-

i) Let $\sigma_w \to 0$. Then $\frac{\Phi}{\sigma_w}\left\{\frac{x_1-(x_0+\gamma_0(y_0))}{\sigma_w}\right\} \to \delta\left\{x_1-(x_0+\gamma_0(y_0))\right\}$

ii) Let $\sigma_v \to 0$ in observation equation at initial stage. That is:-

$$\frac{\Phi}{\sigma_v}\left\{\frac{y_0 - x_0}{\sigma_v}\right\} \to \delta\left\{y_0 - x_0\right\}$$

iii) Let $\gamma_1 = \bar{\gamma}_2$, and $\gamma_0 = \bar{\gamma}_1$. It is also convenient to introduce the function f where $f(x_0) = x_0 + \bar{\gamma}_1(x_0) = x_0 + \bar{\gamma}_1(y_0)$.

Introducing the above simplifioations, we obtain the following expression for $E(x_2^2 \mid y_0,u_0) + k^2u_0^2$

$$E(x_2^2 \mid y_0,u_0)+k^2u_0^2 = \int\left\{\left[\frac{\int f(x_0)\ \frac{\Phi}{\sigma_v}\left\{\frac{y_1-f(x_0)}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}\,dx_0}{\int \frac{\Phi}{\sigma_v}\left\{\frac{y_1-f(x_0)}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}\,dx_0}\right]^2 - 2f(y_0)\left[\frac{\int f(x_0)\ \frac{\Phi}{\sigma_v}\left\{\frac{y_1-f(x_0)}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}\,dx_0}{\int \frac{\Phi}{\sigma_v}\left\{\frac{y_1-f(x_0)}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}\,dx_0}\right]\right.$$

$$+ \; f^2(y_0) \Bigg\} \frac{\Phi}{\sigma_v} \left\{ \frac{y_1 - f(y_0)}{\sigma_v} \right\} \; dy_1 + k^2 \left\{ y_0 - f(y_0) \right\}^2$$

In order to introduce some brevity into the algebra, it is convenient to introduce, as Witsenhausen does, the following definitions:-

$$D_f(y) = \int \frac{\Phi}{\sigma_v} \left\{ \frac{y - f(x)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{x}{\sigma_0} \right\} dx$$

and $$N_f(y) = \int f(x) \frac{\Phi}{\sigma_v} \left\{ \frac{y - f(x)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{x}{\sigma_0} \right\} dx$$

Note that:-

$$D_f'(y) = \frac{d(D_f(y))}{dy} = -\int \frac{(y-f(x))}{\sigma_v^2} \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{x}{\sigma_0} \right\} dx$$

$$= \frac{-yD_f(y)}{\sigma_v^2} + \int \frac{f(x)}{\sigma_v^2} \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{x}{\sigma_0} \right\} dx$$

That is, $$\frac{D_f'(y)}{D_f(y)} = \frac{-y}{\sigma_v^2} + \frac{\int \frac{f(x)}{\sigma_v^2} \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{x}{\sigma_0} \right\} dx}{\int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{x}{\sigma_0} \right\} dx}$$

Thus we obtain:-

$$E(x_2^2 \mid y_0,u_0) + k^2u_0^2 = \int_{\gamma_1^0} \left[\left\{ \frac{N_f(y_1)}{D_f(y_1)} \right\}^2 - 2f(y_0) \left\{ \frac{N_f(y_1)}{D_f(y_1)} \right\} + f^2(y_0) \right] \frac{\Phi}{\sigma_v} \left\{ \frac{y_1 - f(y_0)}{\sigma_v} \right\} dy_1 + (k^2(y_0 - f(y_0))^2$$

Notice that Controller 0 is faced with the problem of minimising this expression at y_0. However, this does not represent an easy scalar minimisation problem. The functions $D_f(y_1)$ and $N_f(y_1)$ are in fact functionals of the unknown function $f(y_0)$. The value of the above expression depends not merely on the value of f at y_0, but on the value of f at every point y_0 in the range $(-\infty,+\infty)$. Thus we might as well present Controller 0 with the problem of minimising -

$E(\underset{\gamma_1^0}{E}(x_2^2 \mid y_0,u_0)) + k^2 E(y_0 - f(y_0))^2$. We then obtain

$$E(x_2^2) + E(k^2(y_0 - f(y_0))^2)$$
$$\gamma_1^0$$

$$= \iint_{\gamma_1^0} \left[\left\{ \frac{N_f(y_1)}{D_f(y_1)} \right\}^2 - 2f(y_0) \left\{ \frac{N_f(y_1)}{D_f(y_1)} \right\} + f^2(y_0) \right] \frac{\Phi}{\sigma_v} \left\{ \frac{y_1 - f(y_0)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{y_0}{\sigma_0} \right\} dy_0 dy_1$$

$$+ k^2 E\left[(y_0 - f(y_0))^2 \right]$$

$$= \int \left\{ \frac{N_f^2(y_1)}{D_f(y_1)} - \frac{2N_f^2(y_1)}{D_f(y_1)} \right\} dy_1 + E(f^2(y_0))$$

$$+ k^2 E([y_0 - f(y_0)]^2)$$

$$= \int \frac{-N_f^2(y_1)}{D_f(y_1)} dy_1 + E(f^2(y_0)) + k^2 E[(y_0 - f(y_0))^2]$$

After some algebra, we obtain for the expected value of the cost function using function $\chi_0(x) = f(x) - x$ the expression:-

$$E(x_2^2) + E(k^2(y_0 - f(y_0))^2) = \sigma_v^2 - \int \left\{ \frac{d}{dy_1} \{ D_f(y_1) \} \right\}^2 \frac{\sigma_v^4 dy_1}{D_f(y_1)}$$

$$+ k^2 E\left[(y_0 - f(y_0))^2 \right]$$

The problem facing the Controller is that of minimising the value of the expression with respect to f. As was pointed out earlier, this is not a trivial problem because the functions $D_f(y)$ are in fact functionals involving the entire range of f and not merely its value at y_0.

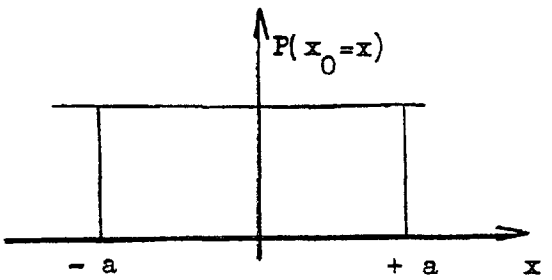
In his paper, Witsenhausen proves:-

i) a solution to this problem exists

ii) this solution is not linear, that is the optimum f(x) is not of the form a + bx.

Attempts to find the solution to this seemingly simple stochastic control have so far been fruitless. Witsenhausen demonstrates that in this zero memory control problem Controller 0 is faced with playing a dual role, with conflicting objectives:-

i) He can use his control value to pass information to Controller 1 about the state of the system at time 0. For example, suppose x_0 is described by a two point symmetric distribution as shown below:-

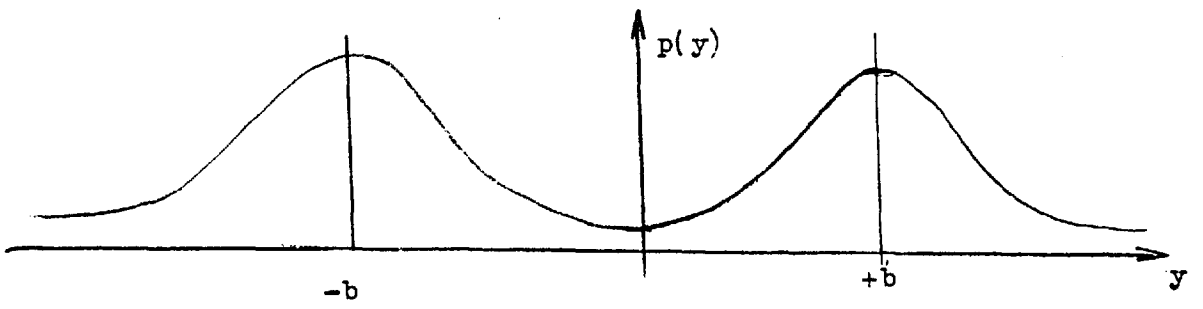


Since x_0 is perfectly observed, Controller 0 can select for example the function f as follows:-

$$f(y_0) = -b \quad \text{if } y_0 = -a$$

$$f(y_0) = +b \quad \text{if } y_0 = +a$$

where $|b| >> |a|$. Controller 1's observation of the state is y_1, where $y_1 = f(y_0) + v$. Thus y_1 will be distributed according to the diagram below:-



It is apparent that Controller 1 can with a great degree of certainty deduce the state of the system at time 1 and hence apply that control which will reduce x_2 to zero. Acting in this role, Controller 0 is utilising his control opportunity not for control purposes but for communication purposes: he is passing signals indirectly to Controller 1 about the state of the system and letting him do the controlling.

ii) Alternatively Controller 0 can play the genuine Controller. He can drive x_1 to zero and leave Controller 1 to do nothing.

Which role Controller 0 adopts will depend on the parameters of the problem (k^2, σ_v^2, and a). For a particular combination of these parameters Witsenhausen shows that the optimal value b is not unique and Controller 0 can adopt either role with the same minimum expectation of cost.

What this simple example shows is a fundamental characteristic of all stochastic control problems which do not exhibit perfect memories. That is, the optimum control functions must reflect a compromise between the communication of information and the control of the system. It is this interaction of communication and control which leads to some of the difficulties in determining the optimal control policies.

We shall return to Witsenhausens problem later. However, first let us demonstrate another feature of the non-perfect memory situation. This is the 'second-guessing' which arises when selecting the optimum policy f. Earlier we derived the following identity:-

$$E(x_2^2 \mid y_1,u_0) + k^2u_0^2 = \int \left\{ \left\{ \frac{N_f(y_1)}{D_f(y_1)} \right\}^2 - 2f(y_0) \left\{ \frac{N_f(y_1)}{D_f(y_1)} \right\} + f^2(y_0) \right\} \frac{\Phi}{\sigma_v} \left\{ \frac{y_1-f(y_0)}{\sigma_v} \right\} dy_1 + k^2u_0^2$$

Let us interpret the terms in this expression. Note that $y_1=f(y_0)+v$.

Thus,

$$\frac{N_f(y_1)}{D_f(y_1)} = \frac{\int f(y_0) \frac{\Phi}{\sigma_v} \left\{ \frac{y_1-f(y_0)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{y_0}{\sigma_0} \right\} dy_0}{\int \frac{\Phi}{\sigma_v} \left\{ \frac{y_1-f(y_0)}{\sigma_v} \right\} \frac{\Phi}{\sigma_0} \left\{ \frac{y_0}{\sigma_0} \right\} dy_0}$$

$$= E(f(y_0) \mid y_1)$$

This leads to the following expression:-

$$E(x_2^2 \mid y_0,u_0) + k^2u_0^2 = E((E(f(y_0) \mid y_1))^2 \mid y_0) - 2f(y_0)\, E(E(f(y_0) \mid y_1) \mid y_0) + f^2(y_0) + k^2(f^2(y_0) - 2y_0 f(y_0) + y_0^2)$$

Controller 0's problem is to select f such that this expression is minimised for every y_0. This would be the usual trivial problem which occurred in both the classical information pattern problem and the nested observation problem were it not for the appearance of such terms as $E(E(f(y_0) \mid y_1) \mid y_0)$. This is Controller 0's estimate of Controller 1's estimate of Controller 0's action at y_0 - in other words, Controller 0 is 'second guessing'. To evaluate these 'second guesses' requires knowledge of the entire f - not just its value at y_0. Similar terms appear of course in the classical information pattern problem and the nested observation problem. The crucial point is however that the nesting of the control data for successive controllers makes the evaluation of these second guesses trivial. Thus, for the nested observation case, we have,

$$E(E(f(y_0) \mid y_1, y_0)^2 \mid y_0) = f^2(y_0),$$

and $$E(E(f(y_0) \mid y_1, y_0) \mid y_0) = f(y_0)$$

This leads to the following simplification of Controller 0's problem:-

$$\underset{\gamma_0}{\min} E(x_2^2 \mid y_0, u_1) = k^2(f^2(y_0) - 2y_0 f(y_0) + y_0^2)$$

i.e. $$f^0(y_0) = y_0$$

i.e. $$\gamma_0^0(y_0) = 0$$

Thus, if the control data is nested, then the optimum policy is for Controller 0 to do nothing and then leave Controller 1 to drive x_2 to zero. This corresponds with our solutions obtained earlier for the nested case and with what is obvious by inspection of the problem. What we have just shown by consideration of the second guessing aspects of the problem is simply another demonstation of the interaction of communication and control in the zero memory problem.

This concludes the discussion of our so-called direct approach to the problem. In the next section we consider an indirect approach.

2.2.3.2. An Indirect Approach

Here we have introduced the word 'Indirect' to decribe the problem which Controller 0 poses for himself. We shall see that although the arguments which lead up to Controller 0's interpretation of the problem are quite plausible, this interpretation is not equivalent to the one which follows from the direct approach discussed above.

We showed in Section 2.2.3.1. above that Controller 1's optimal policy is $\gamma_1^{\circ}(y_1) = -\int x_1 p(x_1 \mid y_1)\,dx_1$, and that the resulting cost function as far as Controller 1 is concerned is $J_1^{\circ}(y_1, \gamma_0^{\circ})$ where:-

$$J_1^{\circ}(y_1, \gamma_0^{\circ}) = E(x_1^2 \mid y_1, \gamma_0^{\circ}) - E^2(x_1 \mid y_1, \gamma_0^{\circ}) + \sigma_w^2$$

$$= \int x_1^2\, p(x_1 \mid y_1, \gamma_0^{\circ})\,dx_1 - \left(\int x_1 p(x_1 \mid y_1, \gamma_0^{\circ})\,dx_1\right)^2 + \sigma_w^2$$

But we have evaluated $p(x_1 \mid y_1 \gamma_0^{\circ})$ for arbitrary γ_0^{a} earlier. Let us put $\gamma_0^{a} = \gamma_0^{\circ}$. Then we have :-

$$p(x_1 \mid y_1, \gamma_0^{\circ}) = \frac{N(\gamma_0^{\circ}, y_1, x_1)}{\int N(\gamma_0^{\circ}, y_1, x_1)\,dx_1}$$

, where N is as defined on page 2.30

and

$$J_1^{o}(y_1, \gamma_0^{o}) = \frac{\int_{x_1^2} N(\gamma_0^{o}, y_1, x_1)\, dx_1}{\int N(\gamma_0^{o}, y_1, x_1)\, dx_1}$$

$$- \left\{ \frac{\int x_1\ N(\gamma_0^{o}, y_1, x_1)\ dx_1}{\int N(\gamma_0^{o}, y_1, x_1)\ dx_1} \right\}^2$$

$$+ \sigma_w^2$$

But what does this expression represent? It is Controller 1's estimate of that component of the ultimate cost which Controller 1 can influence (i.e. x_2^2). Controller 1, if he knows γ_0^{o}, will act optimally and the contribution to the ultimate cost which results from this action will be given by the expression above. Controller 0 would like to select γ_0^{o} to minimise this expression on Controller 1's behalf. But Controller 0's position is complicated by two facts:-

i) He dosen't know y_1. That is, it might be useful to know y_1, but our formulation of the problem dosen't allow it.

ii) He must consider the contribution $k^2 \gamma_0^{o2}(y_0)$ to the final value of the cost.

Suppose then we state Controller 0's problem as follows - find that function γ_0^0 such that $J_0^0(y_0, \gamma_0^0) = E(J_1^0(y_1, \gamma_0^0) \mid y_0) + k^2 \gamma_0^2(y_0)$ is minimised. Let us state explicitly the expression for - $J_0^0(y_0, \gamma_0^0)$:-

$$J_0^0(y_0, \gamma_0^0)$$

$$= \left\{ \frac{\int x_1^2 N(\gamma_0^0, y_1, x_1)\, dx_1}{\int N(\gamma_0^0, y_1, x_1)\, dx_1} \right.$$

$$\left. - \left\{ \frac{\int x_1 N(\gamma_0^0, y_1, x_1)\, dx_1}{\int N(\gamma_0^0, y_1, x_1)\, dx_1} \right\}^2 \right\}$$

$$\cdot M(\gamma_0^0(y_0), y_1, x_1, y_0)\, dx_1\, dy_1$$

$$+ k^2 \gamma_0^{0^2}(y_0)$$

In our current interpretation of the problem, it is Controller 0's task to select that function γ_0^0 which minimises this expression.

Comparing this with the task we set Controller 0 in our previous interpretation when we had to select γ_0° to minimise $E(x_2^2 | y_0, u_0) + k^2 u_0^2$ γ_1° (see page 2.39), we see there is considerable difference in the two approaches to the problem. In order to demonstrate the difference more markedly let us reduce the above expression to the corresponding one for Witsenhausen's problem. This leads to:-

$$J_0^\circ(y_0, \gamma_0^\circ) = E(J_1^\circ(y_1, \gamma_0^\circ) \mid y_0) + k^2 \gamma_0^{\circ 2}(y_0)$$

$$= E(E(f^2(y_0) \mid y_1) \mid y_0) - E\left\{\left[E(f(y_0) \mid y_1)\right]^2 \mid y_0\right\}$$

$$+ k^2 \left\{ y_0 - f(y_0) \right\}^2$$

Compare this with the problem confronting Controller 0 as determined by our earlier direct approach. (Page 2.37)

$$E(x_2^2 \mid y_0, u_0) + k^2 u_0^2 = E\left((E(f(y_0) \mid y_1))^2 \mid y_0 \right)$$

$$- 2 f(y_0) E(E(f(y_0) \mid y_1) \mid y_0)$$

$$+ f^2(y_0) + k^2 \left\{ y_0 - f(y_0) \right\}^2$$

γ_1°

In deriving the above we have omitted the simple but tedious algebraic manipulations involved. What we have shown is that our two interpretations of the problem are quite dissimilar. This is by no means surprising. Let us summarise these two interpretations:-

(i) In the first instance we selected a γ_1° to minimise $E(x_2^2 | y_1)$. This γ_1° was given by $\gamma_1^\circ(y_1) = -\int x_1 p(x_1 | y_1)\, dx_1$. Next we turned to the problem of finding $\gamma_0^\circ(y_0)$, assuming γ_1° as defined above. We had to select a γ_0° which would minimise the expression - $E(x_2^2 | y_0; \gamma_0^\circ \text{ and } \gamma_1^\circ) + k^2 u_0^2$. Explicitly in the Witsenhausen problem this meant the minimisation of:-

$$E(x_2^2 \mid y_0, \gamma_1^\circ, \gamma_0^\circ) + k^2 u_0^2 = E\left(\left[E(f(y_0) \mid y_1)\right]^2 \mid y_0\right)$$

$$- 2f(y_0) E(E(f(y_0) \mid y_1) \mid y_0)$$

$$+ f^2(y_0) + k^2 (y_0 - f(y_0))^2$$

(ii) For the second interpretation we selected a γ_1° to minimise $E(x_2^2 \mid y_1)$. As in the first case above γ_1° was given by -

$\gamma_1^\circ(y_1) = -\int x_1 p(x_1 \mid y_1)\, dx_1$. This leads to an expression for -

$E(x_2^2 \mid y_1, \gamma_1^\circ) = E(x_1^2 \mid y_1) - (E(x_1 \mid y_1))^2$.

Finally we sought that γ_0° would minimise:-

$$E(E(x_1^2 \mid y_1) \mid y_0) - E(\left[E(x_1 \mid y_1)\right]^2 \mid y_0) + k^2 u_0^2$$

$$= E(E(f^2(y_0) \mid y_1) \mid y_0) - E(\left[E(f(y_0) \mid y_1)\right]^2 \mid y_0) + k^2 (y_0 - f(y_0))^2.$$

Thus in our first interpretation, we seek that γ_0° which minimises $E(x_2^2 \mid y_0) + k^2 u_0^2$.

γ_1°

In the second case, we seek that γ_0° which minimises $E(E(x_2^2 \mid y_1) \mid y_0) + k^2 u_0^2$

That these two approaches are not equivalent is not surprising, since, for any γ_1, γ_2, in general.

$$E(E(x_2^2 \mid y_1) \mid y_0) \neq E(x_2^2 \mid y_0)$$
$$\gamma_1 \gamma_2 \qquad\qquad \gamma_1 \gamma_2$$

So, why then have we been at such pains to show that the two approaches are not equivalent. The answer is quite simple. The two approaches are equivalent in the case of both the classical information pattern problem and the nested observation problem. In fact we exploited the equivalence in our derivation of the optimal control laws for these two cases. This equivalence is due to the nesting of the control data at each instant. In the nested observation case we have:-

$$E(E(x_2^2 \mid y_1, y_0) \mid y_0) = E(x_2^2 \mid y_0)$$

The temptation is thus to attempt to analyse non-nested control data problems using the same form of identity. As we have shown, in this case this approach leads to a version of the problem confronting Controller 0 that is different from his problem when he directly attempts to minimise $E(x_2^2 \mid y_0, \gamma_1^\circ) + k^2 \gamma_0^{\sigma^2}(y_0)$.

More will be said of this matter in Chapter 5, for the moment let us simply remark that neither approach has so far been shown to be appropriate.

2.3. Summary and Conclusions

In this chapter we have restricted our attention to the study of a very simple two stage process. It is regretted if the discussion has seemed either verbose, slow, or long-winded. The justification for such a close examination of this simple concrete example is that it has enabled some fundamental concepts of stochastic optimal control to be introduced. Were it not for the fact that considerable confusion appears to exist in the field of stochastic optimal control, this chapter would have been unnecessary. Thus the author has found it impossible to find a single text book which contains a valid derivation of the optimal control policies for a problem with even a classical information pattern. Invariably, the analysis which is presented produces the right answers, but by very dubious methods. It is in fact quite remarkable how insensitive the problems discussed in this chapter are to a variety of conceptual errors. It is perhaps only when the significance of the information pattern is considered that these errors become apparent. Regarding the research literature, Striebel[1] - remarks that "the reluctance of authors in this field to state clearly the nature of a control function has led to considerable confusion". In other words, the information pattern has been largely ignored and interested readers have been forced to deduce for themselves upon what control data the control policy acts. In the next chapter we shall apply the techniques we have demonstrated in this chapter to more general classes of problem. It will be possible then to present a more detailed and critical survey of published

1. C. Striebel "Sufficient Statistics in the Optimum Control of Stochastic Systems". (49)

material on the subject.

Let us summarise then the major conclusions we can draw from this introductory chapter:-

1) The information pattern of a particular stochastic control problem is of prime importance. In the classical information pattern problem, solutions can be fairly readily obtained. For the nested observation information pattern problem, subject to appropriate assumptions, the solution was only marginally more complex. On the other hand, for the zero-memory information pattern problem, even for the simple system studied here, solutions are difficult to find.

2) The system we studied had certain convenient properties. It was linear, it was subject to Gaussian process and observation noises, and it had an additive quadratic cost function. Also the initial state was distributed according to a Gaussian probability density function. For the classical information pattern problem, these properties imply that, whatever control laws are applied , the distribution of the state at each time instant n will be given by probability density functions of the form $p(x_n|y_0,y_1 -- y_n;u_0,u_1 -- u_{n-1})$ which are themselves Gaussian, with the mean of the distribution computable as a simple linear function of $y_0,y_1 -- y_n$ and $u_0,u_1 --u_{n-1}$. Similarly for the nested observation case, $p(x_n|y_0,y_1,-y_n)$ is also Gaussian in structure. For the zero-memory problem, however, we know nothing of the structure of $p(x_n|y_n)$ even when all the above

properties are present. The additivity of the cost function implies that the optimisation problem can be reduced to a series of optimisation sub-problems. For the classical information pattern problem the quadratic nature of the cost function implies that the optimal control value at time n is a simple linear function of $E(x_n|y_0,y_1, -- y_n,u_0,u_1 -- u_{n-1})$. Since $E(x_n|y_0,y_1 -- y_n;u_0,u_1 -- u_{n-1})$ is a linear function of $y_0,y_1, -- y_n$ and $u_0,u_1 -- u_{n-1}$, this implies that the optimal control value at time n is a linear function of $y_0,y_1 -- y_n$ and $u_0,u_1 -- u_{n-1}$.

If the density function $p(x_n|y_0,y_1, -- y_n,u_0,u_1 -- u_{n-1})$ is Gaussian in structure, this enables the distribution to be summarised by a pair of so-called sufficient statistics, namely its mean $\hat{x}_n$ and the variance σ_n^2. For the classical information pattern problem, it is thus possible to derive a recursive algorithm for the generation of $\hat{x}_n$ and σ_n and to express the appropriate control laws in terms of $\hat{x}_n$ and σ_n. These recursive relations take the general form:-

$$\hat{x}_n = L_1(\hat{x}_{n-1}, y_n, u_{n-1}),$$

$$\sigma_n = L_2(\sigma_{n-1})$$

and $$u_n^{\circ} = L_3(\hat{x}_n),$$

where L_1, L_2, and L_3 are linear functions of their arguments. The advantage of introducing these relationships are two fold:-

i) it greatly condenses the algebraic complexity involved in deriving optimal control policies.

ii) the control computer can simply store from one instant to the next the current value of $\hat{x}_n$ (σ_n is not data dependant). Thus, there is no need to store $y_0, y_1, -- y_n$, and u_0, u_1, u_{n-1}. It is sufficient to store $\hat{x}_n$, although it must be borne in mind that $\hat{x}_n$ is always ultimately a function of $y_0, y_1, - y_n$, and $u_0, u_1 -- u_{n-1}$. We simply condense the information contained in this control data to a single parameter $\hat{x}_n$.

It is the elegance of the solution to the classical information pattern problem when sufficient statistics are introduced which has led to the general acceptance of the technique. Unfortunately it is also likely that one reason why general confusion has existed in the field of stochastic optimal control, is that authors have directly attempted to introduce sufficient statistics without first justifying their validity.

If any of the properties mentioned above are relaxed, then the approach we have adopted for the classical information pattern problem is still applicable, and in theory a solution can be derived. We will still have a scalar minimisation problem to solve at each time instant, although this could, however, be computationally quite difficult and it would be unlikely that we could exploit the notion of a sufficient statistic.

A possible approach to such computational difficulties using a perturbation technique is presented in Chapter 6.

3) It was illustrated that the zero memory information pattern problem is characterised by a functional minimisation problem - that is, we are required to find a function γ which minimises a functional $F\{\gamma\}$, where $F\{\gamma\}$ depends on the entire function γ and not simply on its value at a particular point. It will be seen later that this is a general property of all stochastic optimal control problems. The classical information pattern problem is a degenerate example which fortunately permits a reduction to a scalar minimisation problem.

4) Another general property of all stochastic optimal control problems is that each controller must adopt a dual role. Not only must he control the process, he must also seek to impart information regarding his knowledge of the process to subsequent controller via his control actions. This interaction of communication and control is another characteristic of the general stochastic optimal control problem and leads to the notion of "second guessing". The interaction of communication and control was illustrated most graphically in Witsenhausen's problem when the initial state was described by a symmetric two-point distribution function. In this particular problem it was clear that Controller 0 could act either as a Controller or as a Communicator. Either role was optimal.

This concludes our discussion of the introductory example. We have seen that for all but the most degenerate examples of information pattern, the solution to the stochastic optimal control problem can be very difficult. We have mentioned the concept of sufficient statistics and intimated how these sufficient statistics can be exploited to produce a very elegant technique for controlling a linear system with Gaussian noise, additive quadratic cost, and classical information pattern. We have not introduced the closely related concept of separation. Loosely stated, the separation principle states that, for those systems which exhibit the separation property, the stochastic control problem can be separated into two sub-problems as follows:-

i) The estimation problem - this involves the construction of an estimate $E(x_n|I_n)$ of the state of the system x_n at time n, given the control data I_n.

ii) The control problem - this involves the construction of the sequence of optimal control policies $\bar{\gamma}_n^{\,o}$ $(n= 0,1, -- N)$ which optimise a control problem similar to the original one, but with the state vector replaced by $E(x_n|I_n)$ in the process equation and the cost function. This sub-problem has the important property that it is perfectly observable.

The separation theorem then states that if $\gamma_n^{\,o}$ is the optimal control at time n for the original problem, then:-

$$\gamma_n^{\,o}(I_n) = \bar{\gamma}_n^{\,o}(E(x_n|I_n)).$$

Closely related to the separation principle is the so-called certainty-equivalence principle. In the context of the above discussion, when it holds this principle simply states that γ_n^0 is that control policy which would be obtained if all the noise in the system were ignored and we simply studied the resulting deterministic system.

A discussion of both the above topics has been omitted from our introductory example since to have included them would have obscured the basic concepts we have been attempting to introduce.

In the next chapter we shall generalise our stochastic optimal control problem considerably. We shall also attempt to organise the techniques we have presented here into a general dynamic programming formulation of the problem. In so doing a precise notation will be introduced to represent the various quantities which we have used somewhat loosely in this chapter. Also topics such as sufficient statistics, the separation theorem, and the certainty equivalence principle will be discussed in more detail than has been appropriate in the present chapter.

CHAPTER 3

A Dynamic Programming Approach

		Page
3.1	Introduction	3.1
3.2	Statement of Stochastic Optimal Control Problem	3.3
	3.2.1 Description of the Stochastic Control System	3.4
	3.2.2 Evolution of the Stochastic Control System	3.7
	3.2.3 Criterion of Performance	3.14
3.3	Dynamic Programming Formulation	3.18
	3.3.1 Dynamic Programming and the Stochastic Optimal Control Problem	3.18
	3.3.2 Zero-Memory Case	3.20
	3.3.3 Classical Information Pattern	3.33
	3.3.4 General Information Pattern	3.54
3.4	Summary and Conclusions	3.61

3.1. Introduction

The purpose of this chapter is to discuss a dynamic programming formulation of the general discrete time stochastic optimal control problem. In our statement of the problem we shall impose only one restriction - that the cost function is additive with respect to time and its components are positive.

It will be shown that a stochastic optimal control problem involves the determination of a sequence of control functions γ_n, $n = 0, 1, 2, \cdots$. Our dynamic programming approach will attempt to decompose this problem into a sequence of interlinked sub-problems which provide necessary conditions for the funtions $\{\gamma_n\}$. It will be demonstrated, however, that a general dynamic programming formulation which is valid for all information patterns cannot be derived. Even when a dynamic programming formulation can be adopted, it will be shown that generally we must find solutions to sub-problems involving minimisation with respect to an entire function and not merely the value of the functions at a particular point. Only in very special cases can the sub-problems be further reduced to problems involving minimisation with respect to a function value.

We shall consider two specific types of information pattern; the classical information pattern problem and the zero memory problem. For the classical information pattern problem, further comments regarding the separation theorem and the certainty equivalence principle will be made.

In addition to its prime purpose of discussing a dynamic programming framework for the general stochastic optimal control problem, this chapter will also serve to introduce a precise notation which will be employed in subsequent chapters. If the previous chapter demonstrated nothing else, then it certainly demonstrated that some abbreviated notation is required:-

i) to condense the complex expressions which arise in stochastic optimal control problems, and

ii) to avoid the possibility of any ambiguity in our interpretation of these expressions.

In the next section we shall discuss the formulation of a general stochastic optimal control problem, and we shall begin to develop an appropriate notation.

3.2. Statement of the Stochastic Optimal Control Problem

This section will be devoted to the task of developing a precise interpretation of what we mean when we use the term "stochastic optimal control problem". This is not quite as trivial as it may seem at first sight. We shall begin by discussing the system to be controlled; next we shall consider the evolution of this system under an arbitrary sequence of control functions; and finally we shall state the stochastic optimal control problem.

3.2.1. Description of the Stochastic Control System

We shall consider a system which is mathematically described by two equations. The first equation will be referred to as the process equation and it will describe the evolution of the state variable X from one discrete time instant to the next. This process will not necessarily be time invariant. The process equation to be considered will be as follows:-

$$X_{n+1} = F_n (X_n, U_n, W_n),$$

where X_{n+1} represents the state variable at time (n+1),

X_n represents the state variable at time n,

U_n represents the control applied at time n, and

W_n represents some noise process which corrupts the evolution of the state variable.

It is assumed that $X_n \in S_n$ where S_n is an arbitrary state space. For example, $S_n = R^{p_n}$ indicates that the state variable X_n is an element in a p_n-dimensional vector with components $\left\{ X_n^{\,i} \,\middle|\, i = 1,2,- p_n \right\}$, where each $X_n^{\,i}$ is some point of the extended real line $\left[-\infty, +\infty \right]$ Alternatively, each X_n can be an element of a p_n-dimensional 'discrete' state space. In this case X_n is again a p_n-dimensional vector with components $\left\{ X_n^{\,i} \,\middle|\, i = 1,2 -- p_n \right\}$, but now each $X_n^{\,i}$ is an element in the range of some invertible function which has as its domain some subset of the set of integers I.

Similarly it is assumed that $U_n \in C_n$ where C_n is an arbitrary control space. Again, as for S_n, C_n can be continuous or discrete and of arbitrary dimension, q_n, say.

The noise process $\{W_n\}$ is in fact a sequence of independent random variables. Probabilistically, all that we require of W_n is that it is described by an appropriate distribution function P_n^W. We can assume that W_n is an element in, say, an r_n-dimensional space Ω_n. Note that we do not need to assume that P_n^W is time-invariant

Finally we should make some remarks about the state transition function F_n. This function maps a point from the product space - $S_n \times C_n \times \Omega_n$ into the space S_{n+1}. In order to derive the appropriate 'probabilistic' state transition we will require that F_n is measurable with respect to the σ-field generated by $S_n \times C_n \times \Omega_n$. Also, the initial state of the system X_0 is assumed to be a random variable. It is described by a distribution function P_0^X, and takes values in some state space S_0 of dimension p_0 say.

Next, let us turn our attention to the second component in our system - the observation equation. It is assumed that the state of the system is not perfectly observable at any instant in time. The general form of the observation equation will be as follows:-

$$Y_n = G_n(X_n, V_n),$$

where Y_n represents the observation at time n,

X_n represents the state variable at time n, and

V_n represents some noise process which corrupts our observation of the state.

In this case it is assumed that $Y_n \in O_n$, where O_n is a space of dimension h_n, say. The argument X_n has been defined above. The noise process $\{V_n\}$ is like $\{W_n\}$, a sequence of independent random variables. Each random variable V_n is associated with a distribution function P_n^V and takes values in a t_n-dimensional space $\mathcal{V}_n$.

The observation function G_n maps a point from the product space $S_n \times V_n$ into the observation space O_n. It will be assumed that G_n is measurable with respect to the σ-field generated by $S_n \times \mathcal{V}_n$.

Finally let us consider the cost function. It is required that in operating the system, the function J is minimised, where J is defined as follows:-

$$J(X^N, U^{N-1}) = \sum_{n=0}^{N-1} L_n(X_n, U_n) + L_N(X_N)$$

The only restriction on J for the moment is that it is formed as a sum of non-negative terms, each one of which depends only upon the 'current' state and, where appropriate, the 'current' control action. In the definition of J we have introduced the notation X^N and U^N, defined as follows:-

$$X^N \triangleq (X_0, X_1, X_2, \text{---} X_N)$$

$$U^{N-1} \triangleq (U_0, U_1, U_2, \text{---} U_{N-1})$$

We have also introduced the notion of a stopping time N. We are not interested in what the system does after time N. Our only ambition is to control the system such that $J(X^N, U^{N-1})$ is minimised and this function depends only upon variables occurring up to and including time N. Let us consider this in more detail by studying the evolution of the system.

3.2.2. Evolution of the Stochastic Control System

In this section it is intended merely to make some general observations regarding the evolution of the stochastic control system.

The system starts off in some (unknown) state X_0, an observation Y_0 is taken, a control U_0 is applied, and then, at time $t = 1$, the system moves to state X_1. The same sequence of events takes place at time $t = 1$; namely, an observation Y_1 is taken, a control U_1 is applied, and then, at time $t = 2$, the system moves to state X_2. This sequence is repeated at every time instant until the system has moved into state X_N at time $t = N$. By successive substitutions of the state transition functions F_n, we can derive an expression for $J(X^N, U^{N-1})$ in terms of the so called primitive random variables X_0, $\{W_i \mid i = 1 - N - 1\}$, together with values for U^{N-1}. In other words,

$$J(X^N, U^{N-1}) = J^1(X_0, W^{N-1}, U^{N-1}),$$

where J^1 is a known function of its arguments. This will be the value of the cost function for a particular realisation of the appropriate primitive random variables and a particular sequence of control values U^{N-1}. We would like the value of the cost function to be minimised for every realisation of the primitive random variables X_0 and W^{N-1} by appropriate selection of the control values U^{N-1}. But what do we mean by the control values U^{N-1}? Our present formulation would seem to imply that U^{N-1} is a sequence of a priori constants.

If this was the case then it would be impossible to achieve our objective; all that we could do would be to select the U^{N-1} such that the average value of J^1 is minimised. Alternatively, the expression involving J^1 above seems to suggest that U^{N-1} could be deduced as some function of X_0 and W^{N-1}, so that J^1 is minimised for every X_0 and W^{N-1}. But this is not physically possible. For a start, since they are not observable, how can we compute the X_0 and W^{N-1}? Secondly, the control values at some time n say, might depend on future values of $\{W_i,\ i = n+1, \ldots N-1\}$.

What is wrong with our formulation is that the control values must be expressed as functions of the available information at each time instant. Let us examine what we mean by the available information at some time instant n, say. Trivially, the available information first must include a priori information relating to the system, such as knowledge of the functions $\{F_n\}$, $\{G_n\}$ together with the distribution functions $\{P_n^w\}$, $\{P_n^v\}$, and P_0^x. It is assumed that this information is always available and will no longer be considered. But, in addition to the a priori information, the available information will include some records of what has already happened in the system. The most complete set of information available at time n, will thus consist of all past records of observation and control. In other words - complete information is knowledge of $\{Y_0, Y_1, Y_2 \ldots Y_n\}$ and $\{U_0, U_1, U_2, \ldots U_{n-1}\}$.

But generally we will not have this complete information set and we must derive our control values from a knowledge only of some subset of this complete information set. Let us denote this subset by I_n,

that is:-

$$I_n \subseteq \{Y_0, Y_1, Y_2 - Y_n;\ U_0, U_1, U_2 - U_{n-1}\}.$$

When I_n is defined for every n we have defined the information pattern of the problem. We shall call I_n the control data at time n.

Note that we can generalise the formulation even further by considering several controllers acting at the same time instant. Each of these controllers will generally have available a different control data set, composed typically of a collection of the components of each observation vector Y_k and control vector U_k. In the literature this class of problem has been called the multi-person dynamic team problem and sometimes the de-centralised control problem. Examples of such systems do not normally occur in the traditional field of control engineering; they more commonly occur in the control of large scale organisations where information sharing between controllers might become impossible or, at least, expensive. This type of problem will be considered in more detail in Chapter 9 - it will not be examined further here.

Returning to our 'single person' control problem, we now permit the control action U_k at some time k to be a function of the control data I_k, that is,

$$U_k = \gamma_k (I_k)$$

The function γ_k is known interchangeably as the control function at time k, the control law at time k, and the control policy at time k. We shall for the moment consider no restrictions on γ_k other than

that it is a "measurable" function - from the Cartesian product space implied by I_k to the space of control values C_k. This requirement of measurability is simply inserted to ensure that various operations defined later have a sound technical foundation. In practice it represents no significant restriction on the class of admissable control policies. The symbol $\textcircled{H}_n$ will be used to denote the class of admissable control laws at time n.

In our definition of the information pattern of the problem we have been guilty of some abuse of terminology. We have defined the control data as some subset of the complete information set.

$$\text{i.e. } I_k \subseteq \{Y_0, Y_1, \text{---} Y_k; U_0, U_1 \text{---} U_{k-1}\}$$

What we really mean by this is that :-

i) the subscripts defining the observations available for the determination of control at time k are a subset of the set $\{0,1,2, \text{---} k\}$. We could denote this subset by M_k.

ii) the subscripts defining the control values available for the determination of control at time k are a subset of $\{0,1, \text{---} k-1\}$. In this case, we could denote this subset by N_k.

The information pattern of the problem is then defined when (M_k, N_k) is defined for every time k. Rather than discuss information patterns in terms of subsets of subscripts, it is perhaps more convenient to introduce a further function into our formulation.

We shall call this function the information function B_k at time k. It is introduced primarily for notational convenience. Recall that complete information at time k was defined as the complete record of all observations up to time k and all control values up to time k-1. Let us denote this information set by $\hat{I}_k$, in other words,

$$\hat{I}_k \triangleq \left\{Y_0, Y_1, Y_2, \text{---} Y_k; U_0, U_1, \text{---} U_{k-1}\right\}^T$$

Here we are treating $\hat{I}_k$ as a vector of dimension $\sum_{n=1}^{k} h_n + \sum_{n=1}^{k-1} q_n$

Recall also that the control data I_k was loosely defined as a subset I_k of $\hat{I}_k$. Then the information function B_k is defined as follows:-

$$I_k = B_k(\hat{I}_k)$$

Thus B_k is simply a function which maps the complete information set $\hat{I}_k$ into the actual control data I_k at time k. The specification of B_k for every k defines the information pattern of the problem. Generally B_k will simply be a permutation matrix, composed of 1's and 0's, arranged to select the appropriate elements from $\hat{I}_k$.

For example, for the classical information pattern problem - $(I_k = \hat{I}_k)$, we have $B_k = U^k$ where U^k is the identity matrix of appropriate dimension, that is,

$$I_k = U^k \hat{I}_k$$

For the zero memory problem in which the only information available for control is the current observation, we have -

$$I_0 = \begin{bmatrix} 1 \end{bmatrix} \quad \begin{bmatrix} Y_0 \end{bmatrix} \; ;$$

$$I_1 = \begin{bmatrix} 0 \vdots 1 \vdots 0 \end{bmatrix} \quad \begin{bmatrix} Y_0 \\ Y_1 \\ U_0 \end{bmatrix} \; ;$$

$$I_2 = \begin{bmatrix} 0 \vdots 0 \vdots 1 \vdots 0 \vdots 0 \end{bmatrix} \quad \begin{bmatrix} Y_0 \\ Y_1 \\ Y_2 \\ U_0 \\ U_1 \end{bmatrix} \; ;$$

and so on.

Let us now consider the evolution of the stochastic control system. We shall assume that some known sequence of control laws $\gamma_0, \gamma_1, \gamma_2, \text{ -- } \gamma_N$ is specified. The system starts off in some (unknown to the controller) state X_0, an observation Y_0 is taken, an information set I_0 becomes available, a control $\gamma_0(I_0)$ is applied, and at time t=1, the system progresses to state X_1, where,

$$\begin{aligned} X_1 &= F_0 (X_0, \gamma_0(I_0), W_0) \\ &= F_0 (X_0, \gamma_0(B_0(\hat{I}_0)) W_0) \\ &= F_0 (X_0, \gamma_0(B_0(Y_0)) W_0) \\ &= F_0 (X_0, \gamma_0(B_0(G_0(X_0, V_0))), W_0) \end{aligned}$$

that is $X_1 = F_0^{\gamma_0}(X_0, V_0, W_0)$.

A similar sequence of events now takes place at time t = 1. An observation Y_1 is taken, an information set I_1 becomes available, a control $\gamma_1(I_1)$ is applied, and at time t = 2, the system progresses to state X_2, where:-

$$
\begin{aligned}
X_2 &= F_1(X_1, \gamma_1(I_1), W_1) \\
&= F_1(F_0^{\gamma_0}(X_0, V_0, W_0), \gamma_1(B_1(\hat{I}_1)), W_1) \\
&= F_1(F_0^{\gamma_0}(X_0, V_0, W_0), \gamma_1(B_1(G_0(X_0, V_0), G_1(X_1, V_1), \\
&\quad \gamma_0(B_0(G_0(X_0, V_0))), W_1) \\
&= F_1(F_0^{\gamma_0}(X_0, V_0, W_0), \gamma_1(B_1(G_0(X_0, V_0), \\
&\quad G_1(F_0^{\gamma_0}(X_0, V_0, W_0), V_1), \gamma_0 B_0(G_0(X_0, V_0))), W_1) \\
&= F_1^{\gamma_0 \gamma_1}(X_0, V_0, V_1, W_0, W_1)
\end{aligned}
$$

Again a similar sequence of events will take place at time t = 3, and so on until the terminal time t = n is reached. It is apparent from the above discussion that the state variable X_k at any time k can be determined from a known function as follows:-

$$X_k = F_k^{\gamma_0 \gamma_1 \cdots \gamma_k}(X_0; V_0, V_1, \cdots V_{k-1}; W_0, W_1, \cdots W_{k-1})$$

Notice that X_k is a known function of the primitive random variables X_0, $\{V_i : i = 0 - k-1\}$ and $\{W_i : i = 0 - k-1\}$. Thus X_k is a well defined random variable.

3.2.3. The Criterion of Performance

A similar iterative use of the process and observation equations enables the control value U_k at some time k to be expressed as a function of the primitive random variables. In other words,

$$U_k = H_k^{\gamma_0, \gamma_1, \cdots \gamma_k}(X_0; V_1, V_2, \cdots V_k; W_0, W_1 \cdots W_{k-1})$$

Thus we can now express the cost function J as follows:-

$$J(X^N, U^{N-1}) = \sum_{n=0}^{N-1} L_n(X_n, U_n) + L_N(X_N)$$

$$= \sum_{n=0}^{N-1} L_n^{\gamma_0, \gamma_1 \cdots \gamma_n}(X_0; V_0, V_1, \cdots V_n; W_0, W_1 \cdots W_{n-1})$$

$$+ L_n^{\gamma_0, \gamma_1 \cdots \gamma_{N-1}}(X_0; V_0, V_1, \cdots V_n; W_0, W_1 \cdots W_{n-1})$$

It is apparent that J is thus a random variable, being a function of the primitive random variables. Its value for a particular realisation of the primitive random variables will depend on the choice of control policies $\gamma_0, \gamma_1 \cdots \gamma_{N-1}$. Our objective is to select an appropriate sequence of control policies such that the value of J is minimised. But this is not a well defined problem because, as we have just shown, J is a random variable. Several alternatives are available to us for resolving this difficulty:-

1) We can attempt to find a sequence of control policies - $\gamma_0, \gamma_1, \gamma_2, \cdots \gamma_N$ such that the most likely value of J is minimised.

2) We can attempt to find a sequence of control policies - $\gamma_0, \gamma_1, \text{---} \gamma_N$ such that the worst possible value of J is minimised, irrespective of the liklihood of this value of J actually occurring.

3) We can attempt to find a sequence of control policies - $\gamma_0, \gamma_1, \gamma_2, \text{---} \gamma_N$ such that the average or expected value of J is minimised. This is the most common approach, but it should be noted the selection of this criterion of performance is based on the assumption that the "experiment" will be repeated a sufficient number of times to justify taking the average value. In the case of industrial control problems this might be a reasonable approach. In other applications it might not be so suitable.

4) In a sense, the three cases discussed above represent extremes. It might be quite reasonable to adopt a mixture of these cases in order to measure the performance of our stochastic control system. For example, suppose we have some sequence of control laws - $\gamma_0, \gamma_1, \gamma_2 \text{---} \gamma_N$. For each value of X_0, we could compute the average value of J with respect to the primitive random variables $\{W_i : i = 0,1,N\}$ and $\{V_i : i = 0,1, \text{--} N\}$. We could then consider the most likely value of this average value with respect to the probability distribution of X_0. The optimal control problem would then be to select that sequence of control laws $\gamma_0, \gamma_1, \text{---} \gamma_N$ which minimise this quantity.

As mentioned above, the criterion which is most commonly selected is the average or expected value of J. This is the criterion which we shall generally use here and so we shall introduce some special notation for dealing with it.

First, let us denote the sequence $\gamma_0, \gamma_1, \gamma_2, \text{---} \gamma_{N-1}$ by Γ_{N-1}. Secondly, the expected value of J will be denoted by V. Thus,

$$\Gamma_n \triangleq \{\gamma_0, \gamma_1, \gamma_2, \text{---} \gamma_n\}, \text{ and}$$

$$V(\Gamma_n) \triangleq E(J(X^N, U^{N-1}))$$

$$= E\left\{\sum_{n=0}^{N} L_n^{\Gamma_{n-1}}(X_0; V_0, V_1 \text{ -- } V_n; W_0, W_1 \text{ -- } W_n)\right.$$

$$\left. + L_N^{\Gamma_{N-1}}(X_0; V_0, V_1 \text{ -- } V_N; W_0, W_1, \text{ -- } W_N)\right\}$$

Thus we can at last state the stochastic optimal control problem. It is to select that sequence of control laws Γ_{N-1} such that $V(\Gamma_{N-1})$ is minimised. Each control law γ_n must be selected from the class of admissable control laws $(H)_n$. To cater for the situation in which there exists no Γ_{N-1}^{o} such that $V(\Gamma_{N-1}^{o})$ is minimal, we introduce the infimum notation; that is

$$V^{o} = \inf_{\Gamma_{N-1}} (V(\Gamma_{N-1}))$$

$$\gamma_n \in (H)_n \text{ for every } n = 0,1,\text{---- } N-1.$$

In the event that this infimum is not attained within the class of admissable controls, then subject to certain regularity conditions we can find an admissable control sequence Γ_{N-1} which has the property that

$$V^{\circ} + \epsilon \geqslant V(\Gamma_{N-1}) \text{ for any } \epsilon > 0.$$

This type of control law is usually referred to as ϵ - optimal control; the topicis only mentioned here and it will not be pursued in detail.

Our discussion of the statement of the stochastic optimal control problem has been somewhat lengthy. It is considered that this is unavoidable for the same reasons which justified the equally lengthy discussion of the simple examples presented in Chapter 2. It is important to establish the fact that a stochastic optimal control problem involves the selection of a sequence of control <u>functions</u>. Any approach which begins with an attempt to find a sequence of control <u>values</u> is conceptually incorrect (although ultimately it might provide the right answers). Control values are random variables which cannot even be defined until the control function is specified. And the control function cannot be specified until the information pattern has been defined. Failure to appreciate these points has led to considerable confusion in the field.

Finally the selection of an appropriate criterion of performance should be carefully considered. The one selected here is the one which is commonly adopted throughout the literature and therefore the one about which most if known. However, the lack of any results for those classes of problem which have non-classical information patterns ($I_k \neq \hat{I}_k$) indicates that a re-consideration of the criterion of performance might be fruitful.

3.3 Dynamic Programming Formulation

3.3.1 Dynamic Programming and the Stochastic Optimal Control Problem

Below we repeat the general problem of stochastic optimal control. We have chosen to parameterise the cost functional V with the probability density function p_0, which describes the initial state of the system X_0. Expectation is with respect to the primitive random variables. We have:-

$$V(\Gamma_{N-1}, p_0) = E(J(X^N, U^{N-1}))$$

$$= E\Big(\sum_{n=0}^{N-1} L_N \gamma_0 -- \gamma_n(X_0; \{V_k : k = 0--n\}; \{W_k : k = 0,1,--n-1\})\Big)$$

$$+ L_N \gamma_0 -- \gamma_{N-1}(X_0; \{V_k : k = 0--n\}; \{W_k\ k = 0--N-1\})$$

The problem is then to find $V^o(p_0)$ such that,

$$V^o(p_0) = \inf\ V(\Gamma^{N-1}, p_0)$$

$$\Gamma^{N-1} \in \textcircled{H}_1 \times \textcircled{H}_2 \times \textcircled{H}_3 -- \textcircled{H}_{N-1}$$

Following the usual approach of dynamic programming[1], we must introduce a partial cost function which provides the cost from some time k in terms of the 'state variable' at time k. At some time k, the dynamic system will be in a state X_k and controls $\gamma_0, \gamma_1 -- \gamma_{k-1}$ will have been applied. The extra cost which will be incurred in evolving from X_k at time k to X_N at time N will be denoted by V_k where

$$V_k = E\Big(\sum_{n=k}^{N-1} L_n(X_n, U_n) + L_N(X_N)\Big)$$

$$= E\Big(\sum_{n=k}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N)\Big)$$

1. See for example O.L.R. Jacobs 'Introduction to Dynamic Programming

Before proceeding to the general case, let us first consider two specific cases. These will be the zero-memory case and the classical information pattern case.

3.3.2 Zero-Memory Case

For this type of information pattern we have $I_n = Y_n$ for all n. In other words the control data at each time instant is composed only of the current observation. The Controller cannot recall what has been observed or done at any time instant prior to the current time instant. Although this case is one about which very little is known, it does reduce to a very simple dynamic programming formulation. The 'state variable' to be selected is the probability distribution function of the dynamic state variable. That is, our 'state variable' will be the infinite dimensional vector $P_n(X)$ where $P_n(X)$ obeys an appropriate state transition equation

State Transition Equation

With the probability distribution function $P_n(x)$ which describes the state vector X_n at time n selected as our 'state variable', the next problem is to derive a suitable transition equation which expresses the function $P_{n+1}(X)$ in terms $P_n(X)$. For the very general spaces proposed in our statement of the problem in Section 3.2., a rigorous derivation of a state transition equation would be dependent upon a considerable degree of measure theoretic technicalities. It is felt that such a derivation would obscure the main development of a dynamic programming formulation of the problem. What is more, for the particular examples which will be studied later, such a level of rigour is not required.

For our present purposes we shall adopt a formal approach. Thus we shall assume that the various distribution functions are differentiable and that suitable probability density functions can be defined.

Suppose that $p_{n+1}(X_{n+1})$ denotes the probability density function of the state X_{n+1} at time n+1. Also we shall assume that $p(X_{n+1}, Y_n, X_n)$ denotes the joint probability density function of X_{n+1}, the observation Y_n at time n, and the state X_n at time n. Then,

$$P(X_{n+1}) = \iint_{S_n\ O_n} p(X_{n+1}, Y_n, X_n)\ dY_n\ dX_n$$

where S_n denotes the range space of X_n, and O_n denotes the range space of Y_n. Thus, by the chain rule for conditional probability density functions:-

$$p_{n+1}(X_{n+1}) = \iint_{S_n O_n} p(X_{n+1} \mid X_n, Y_n)\ p(Y_n \mid X_n)\ p_n(X_n)\ dY_n\ dX_n$$

Thus we have our required state transition equation, providing the two indicated density functions $p(X_{n+1} \mid X_n, Y_n)$ and $p(Y_n \mid X_n)$ can be provided. In applications these two density functions can usually be derived without difficulty from the process and observation equations respectively.

Thus,

$$P\left\{X_{n+1} \leqslant x_{n+1} \mid X_n = x_n;\ Y_n = y_n\right\}$$

$$= P\left\{F_n(X_n, \gamma_n(Y_n), W_n) \leqslant x_{n+1} \mid X_n = x_n;\ Y_n = y_n\right\}$$

$$= P\left\{W_n \leqslant F_n'(x_{n+1}, x_n, \gamma_n(y_n))\right\}$$

$$= \int_{-\infty}^{F_n'(x_{n+1},x_n,\gamma_n(y_n))} dP_n^{w}(s)$$

For convenience we have assumed the function F_n' above can be defined. Finally, by a suitable change of variable, $s=F_n^{*}(t,x_n,\gamma_n(y_n))$ we obtain the expression:-

$$P\left\{X_{n+1} \leqslant x_{n+1} \mid X_n=x_n, Y_n=y_n\right\} = \int_{-\infty}^{x_{n+1}} dP_n^{w}(F_n'(t,x_n,\gamma_n(y_n)))$$

Thus we have obtained an expression for the probability density function $p(X_{n+1} \mid X_n, Y_n)$. An example of this construction was presented in Chapter 2. Note that the probability density function depends on the control law γ_n through its value $\gamma_n(Y_n)$ at Y_n

In a similar way we can derive an expression for $P(Y_n \mid X_n)$ from the observation equation of the system. Thus:-

$$P\left\{Y_n \leqslant y_n \mid X_n = x_n\right\} = P\left\{G_n(X_n,V_n) \leqslant y_n \mid X_n = x_n\right\}$$

$$= P\left\{V_n \leqslant G_n'(x_n,y_n)\right\}$$

$$= \int_{-\infty}^{G_n'(x_n,y_n)} dP_n^{v}(s)$$

For convenience we have assumed that the function $G_n'(x_n,t)$ above can be defined. Finally, by a suitable change of variables, $s = G_n'(x_n,t)$, we obtain the expression:-

$$P\left\{Y_n \leqslant y_n \mid X_n = x_n\right\} = \int_{-\infty}^{y_n} dP_n^{v}(G_n'(x_n,t))$$

Thus we have demonstrated the manner in which $p(X_{n+1} \mid X_n, Y_n)$ and $p(Y_n \mid X_n)$ can be derived from the process and observation equations. It follows from our equation,

$$p(X_{n+1}) = \int\int p(X_{n+1} \mid X_n, Y_n)\, p(Y_n \mid X_n)\, p(X_n)\, dX_n\, dY_n$$

that if we know the probability density function of $p_n(X_n)$ of the state X_n at time n, then we can derive the probabiltiy density function $p_{n+1}(X_{n+1})$ of the state X_{n+1} at time $n+1$. It is obvious that since $p(X_{n+1} \mid X_n, Y_n)$ depends on $\gamma_n(Y_n)$ and since we are integrating over Y_n, the function $p_{n+1}(X_{n+1})$ must depend on the entire function γ_n and not merely its value $\gamma_n(Y_n)$ at Y_n.

As we have already warned, the construction above has been very formal. A more rigorous discussion of the problem of deriving a state transition equation is however not appropriate here.

For the moment let us assume that a suitable state transition equation involving the probability density function $p_n(X_n)$ can always be defined in the form of the equation above. Notice that, if $\mathcal{L}_n$ indicates the transformation from $p_n(X_n)$ to $p_{n+1}(X_{n+1})$, we have symbolically:-

$$p_1 = \mathcal{L}_0(p_0, \gamma_0)$$

$$p_2 = \mathcal{L}_1(p_1, \gamma_1) = \mathcal{L}_1(\mathcal{L}_0(p_0, \gamma_0), \gamma_1) = \mathcal{L}_1'(p_0, \gamma_0, \gamma_1)$$

$$p_{n+1} = \mathcal{L}_n(p_n, \gamma_n) = \mathcal{L}_n(\mathcal{L}_{n-1}'(p_0, \gamma_0, \gamma_1 \cdots \gamma_{n-1}), \gamma_n) = \mathcal{L}_n'(p_0, \gamma_0, \gamma_1 \cdots \gamma_n),$$

where the functions $p_n(X_n)$ are denoted as p_n to indicate that we mean functions and not function values.

Thus we see that the probability density function p_n, our state, is dependent upon:-

i) its initial condition

ii) the control laws applied up to including time (n-1).

What we must next investigate is that knowledge of p_n provides sufficient information to enable it to be used in a suitable dynamic programming formulation of the problem.

Iterative Relationship Involving Partial Cost

Suppose we define the partial cost from time k onwards as follows:-

$$V_k(p_k, \Gamma_k^{N-1}) \triangleq E\left(\sum_{n=k}^{N-1} L_n(X_n, \gamma_n(Y_n)) + L_N(X_N)\right)$$

We must first verify that this partial cost is indeed a function of the (postulated) arguments p_k and $\Gamma_k^{N-1} = \gamma_k, \gamma_{k+1} -- \gamma_{N-1}$

Taking a typical term in this expression, we have,

$$L_n(X_n, \gamma_n(Y_n)) = L_n^{\gamma_n, \gamma_{n-1}}(X_{n-1}, V_{n-1}, W_{n-1}, V_n),$$

by substituting ,

$$X_n = F_{n-1}(X_{n-1}, \gamma_{n-1}(Y_{n-1}), W_{n-1}),$$

and $Y_n = G_n(X_n, V_n)$

Carrying on with this substitution, we eventually have:-

$$L_n(X_n, \gamma_n(Y_n)) = L_n^{\gamma_n, \gamma_{n-1} -- \gamma_k}(X_k, W_k, -- W_{n-1}, V_k, -- V_n)$$

Clearly then,

$$E\Big(\sum_{n=k}^{N-1} L_n(X_n, \gamma_n(Y_n)) + L_N(X_N)\Big)$$

$$= E\Big(\sum_{n=k}^{N-1} L_n \gamma_n, \gamma_{n-1} -- \gamma_k(X_k, W_k, -- W_{n-1}, V_k, -- V_n)$$

$$+ E(L_N \gamma_{N-1}, -- \gamma_k(X_k, W_k -- W_{N-1}, V_k -- V_N)) \triangleq E(L'_k(X_k, W_k -- W_{N-1}, V_k -- V_N))$$

where expectation can now be defined with respect to the random variables X_k and the primitive random variables W_k -- W_{N-1}, and V_k -- V_N.

Thus we have shown that knowledge of:-

i) the probability distribution p_k of X_k at time k, and

ii) the control functions γ_k, γ_{k+1}, -- γ_{N-1}

is sufficient to enable the costs from time k onwards to be calculated. In other words, the probability distribution p_k satisfies our requirement for a state variable - it summarises all past history of the system.

Notice that we still must derive a suitable iterative relationship involving V_k and V_{k+1}. This is easily achieved as follows:-

$$V_{k+1}(p_{k+1}, \Gamma_{k+1}^{N-1}) = E(L'_{k+1}(X_{k+1}; \Gamma_{k+1}^{N-1}; W_{k+1} -- W_{N-1}; V_{k+1} -- V_{N-1}))$$

and $$V_k(p_k, \Gamma_k^{N-1}) = E(L_k(X_k, \gamma_k(Y_k)))$$

$$+ E(L'_{k+1}(F_k(X_k, \gamma_k(G_k(X_k, V_k) W_k), \Gamma_{k+1}^{N-1}, W_{k+1} -- W_{N-1}, V_{k+1} -- V_{N-1})$$

To be of use, we must demonstrate that the second term in this expression is equivalent to:-

$V_{k+1}(p_{k+1}\vdash^{N-1}_{k+1})$, as defined above and where,

$$p_{k+1}(X_{k+1}) = \iint_{S_k O_k} p(X_{k+1}|X_k, \gamma_k(Y_k))\, p(Y_k|X_k)\, p_k(X_k)\; dX_k dY_k$$

Again a rigorous discussion of this equivalence has been omitted for clarity. Essentially it is merely a complex application of a well-known result from probability theory. Loosely stated, this result says that if $Y = G(X)$ where X is random variable with distribution function F_X and G is a measurable function defined on the range space of X, then Y is a random variable whose distribution function F_Y can be determined from F_X as follows:-

$$\begin{aligned} F_Y(y) &= P\left\{Y \leq y\right\} \\ &= P\left\{G(X) \leq y\right\} \\ &= P\left\{X \in G^{-1}(-\infty, y]\right\} \\ &= \int_{G^{-1}(-\infty, y]} dF_X(x) \end{aligned}$$

What is more, for any measurable function L defined on the range space of Y, we have:-

$$E(L(G)) = \int L(G(X))\, dF_X = \int L(Y)\, dF_Y$$

In our case, dF_X is specified by $p_n(X)$ and dF_Y is specified by - $p_{n+1}(X)$. The function G is the composition of the process equation and the observation equation. The rigorous analysis which we are omitting would merely justify what is intuitively obvious anyway.

What we have shown is that:-

$$V_k(p_k, \Gamma_k^{N-1}) = E(L_k(X_k, \gamma_k(Y_k)) + V_{k+1}(p_{k+1} = \mathcal{L}_k(p_k, \gamma_k), \Gamma_{k+1}^{N-1})$$

At last we can state the dynamic programming formulation of the zero-memory stochastic optimal control problem.

It is as follows:-

Theorem 3.1

$$\text{If } V_{k+1}^{o}(p_{k+1}) \triangleq \inf_{\Gamma_{k+1}^{N-1}} E\left(\sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(Y_n)) + L_N(X_N)\right)$$

$$\text{and } p_{k+1}(X_{k+1}) \triangleq \iint_{X_k O_k} p(X_{k+1} \mid X_k, Y_k; \gamma_k)\, p(Y_k \mid X_k)\, p_k(X_k)\, dX_k dY_k$$

then,

$$V_k^{o}(p_k) = \inf_{\gamma_k} \left[E(L_k(X_k, \gamma_k(Y_k))) + V_{k+1}^{o}(p_{k+1} = \mathcal{L}_k(p_k, \gamma_k))\right]$$

where the notation $p_{k+1} = \mathcal{L}_k(\gamma_k, p_k)$ is introduced to indicate that p_{k+1} is to be expressed as a functional of γ_k and p_k given by the 'state' transition equation.

Proof

This can easily be verified as follows:-

$$V_k^{\circ}(p_k) \triangleq \inf_{\Gamma_k^{N-1}} \left\{ E(L_k(X_k, \gamma_k(Y_k))) + V_{k+1}(p_{k+1} = \mathcal{L}_k(P_k, \gamma_k'), \Gamma_{k+1}^{N-1}) \right\}$$

$$= \inf_{\gamma_k} \left\{ E(L_k(X_k, \gamma_k(Y_k)) + \inf_{\Gamma_{k+1}^{N-1}} V_{k+1}(p_{k+1} = \mathcal{L}_k(p_k, \gamma_k), \Gamma_{k+1}^{N-1}) \right\}$$

$$= \inf_{\gamma_k} \left\{ E(L_k(X_k, \gamma_k(Y_k)) + V_{k+1}^{\circ}(p_{k+1} = \mathcal{L}_k(p_k, \gamma_k)) \right\}$$

In order to use this formulation, we also make the trivial definition:-

$$V_N^{\circ}(p_N) \triangleq E(L_N(X_N))$$

Thus we have decomposed our problem which originally involved finding a minimum with respect to an entire collection of control laws into a sequence of interlinked control problems, each one involving minimisation with respect to a single control law.

Symbolically, the algorithm for finding the optimum control laws $\gamma_0^{\circ}, \gamma_1^{\circ}, \gamma_2^{\circ}, \ldots \gamma_{N-1}^{\circ}$ is as follows:-

i) Compute $V_N^{\circ}(p_N) = E(L_N(X_N))$

ii) Find γ_{N-1}° such that:-

$$V_{N-1}^{\circ}(p_{N-1}) \triangleq E(L_{N-1}((X_{N-1}), \gamma_{N-1}^{\circ}(Y_{N-1}))) + V_N^{\circ}(p_N = \mathcal{L}_{N-1}(\gamma_{N-1}^{\circ}, p_{N-1}))$$

$$= \inf_{\gamma_{N-1}} \left\{ E(L_{N-1}(X_{N-1}, \gamma_{N-1}(Y_{N-1}))) + V_N^{\circ}(p_N = \mathcal{L}_{N-1}(\gamma_{N-1}^{\circ}, p_{N-1})) \right\}$$

This yields γ_{N-1}° as a functional of p_{N-1}. Let us denote this as,

$$\gamma_{N-1}^{\circ} = \mathcal{J}_1(p_{N-1})$$

iii) Find γ_{N-2}° such that:-

$$V_{N-2}^{\circ}(p_{N-2}) = E(L_{N-2}(X_{N-2}, \gamma_{N-2}^{\circ}(Y_{N-2}))) + V_{N-1}^{\circ}(p_{N-1} = \mathcal{L}_{N-2}(\gamma_{N-2}^{\circ}, p_{N-2}))$$

$$= \inf_{\gamma_{N-2}} \left\{ E(L_{N-2}(X_{N-2}, \gamma_{N-2}(Y_{N-2}))) + V_{N-1}^{\circ}(p_{N-1} = \mathcal{L}_{N-2}(\gamma_{N-2}, p_{N-2})) \right\}$$

Again this yields γ_{N-2}° as a functional of p_{N-2}; let us denote this as:-

$$\gamma_{N-2}^{\circ} = \mathcal{J}_2(p_{N-2})$$

iv) The above process is repeated for each time step n, until the starting time is reached and we have:-

$$V_0^{\circ}(p_0) = E(L_0(X_0, \gamma_0^{\circ}(Y_0))) + V_1^{\circ}(p_1 = \mathcal{L}_0(\gamma_0^{\circ}, p_0))$$

$$= \inf_{\gamma_0} \left\{ E(L_0(X_0, \gamma_1(Y_0))) + V_1^{\circ}(p_1 = \mathcal{L}_0(\gamma_0, p_0)) \right\}$$

and $\gamma_0^{\circ} = \mathcal{J}_N(p_0)$

At this point let us consider what has been achieved. We have have determined $V_0^{\circ}(p_0)$, where,

$$V_0^{\circ}(p_0) = \inf_{\Gamma_0^{N-1}} V_0(p_0, \Gamma_0^{N-1})$$

$$= \inf_{\Gamma_0^{N-1}} E\left\{ \sum_{n=0}^{N-1} L_n(X_n, \gamma_n(Y_n)) + L_N(X_N) \right\}$$

Thus in determining $V_0^{\circ}(p_0)$ we have found a sequence of control laws $\gamma_0^{\circ}, \gamma_1^{\circ} -- \gamma_{N-1}^{\circ}$, each parameterised by the 'simultaneous' probability density $p_0, p_1, -- p_{N-1}$, which when applied will result in expected cost for the system which is minimum. (If the infima are obtained outside the admissable class of control laws, then we can find control laws which will result in an expected cost which is within ϵ of this minimum cost for any $\epsilon > 0$).

Having determined the optimal sequence $\gamma_0^{\circ}, \gamma_1^{\circ}, \gamma_2, -- \gamma_{N-1}^{\circ}$ the dependence upon the 'simultaneous' probability density functions $p_0, p_1, p_2 -- p_{N-1}$ can be eliminated by use of the state transition equation as follows:-

$$\gamma_0^{\circ} = \mathcal{F}_N(p_0)$$

$$\gamma_1^{\circ} = \mathcal{F}_{N-1}(p_1) = \mathcal{F}_{N-1}(\mathcal{L}_0(\gamma_0^{\circ}, p_0)) = g_1(p_0)$$

$$\gamma_2^{\circ} = \mathcal{F}_{N-2}(p_2) = \mathcal{F}_{N-2}(\mathcal{L}_1(\gamma_1^{\circ}, p_1)) = \mathcal{F}_{N-2}(\mathcal{L}_1'(\gamma_0^{\circ}, \gamma_1^{\circ}, p_0))$$

$$= g_2(p_0)$$

$$\gamma_{N-1}^{\circ} = \mathcal{J}_1(p_{N-1}) = \mathcal{J}_1(\mathcal{L}_{N-2}(\gamma_{N-2}, p_{N-2})) = \mathcal{J}_1(\mathcal{L}'_{N-2}(\gamma_0^{\circ}, \gamma_1^{\circ}, \cdots \gamma_{N-1}^{\circ}, p_0))$$

$$= G_{N-1}(p_0)$$

Thus we see that the optimal control policies have at last be determined as a functional of the initial state probability density function p_0 (distribution function P_0^x). In eliminating the dependence on p_0 we have generated an 'open-loop' solution in the sense discussed in Chapter 2. The control law $\gamma_k^{\circ} = G_k(p_0)$ is optimal only if all previous controllers act optimally. If they do not then $\gamma_k^{\circ} = G_k(p_0)$ will not be optimal. This is, however, a somewhat trivial point since $\gamma_k^{\circ} = \mathcal{J}_{N-k}(L_{k-1}(\gamma_0, \gamma_1, \cdots \gamma_{k-1} p_0))$ will always yield the optimal control law, irrespective of whether $\gamma_0, \gamma_1, \cdots \gamma_{k-1}$ are optimal control laws.

The major difficulty in applying the algorithm we have developed is finding functions γ_k which satisfy the fundamental equation of dynamic programming, namely,

$$V_k^{\circ}(p_k) = \inf_{\gamma_k} \left\{ E(L_k, (X_k, \gamma_k(Y_k))) + V_{k+1}^{\circ}(p_{k+1} = \mathcal{L}_k(p_k, \gamma_k)) \right\}$$

Finding a minimising γ_k for this expression involves minimisation with respect to an entire function and not simply its value at a particular point. This was demonstrated by the zero-memory example presented in Chapter 2. Even for that very simple case finding γ_k° proved to be an impossible task. If however we restrict our attention to systems which have a discrete state space, a discrete space of observations, and a discrete set of control values then the optimal

γ_k° can always be found by direct enumeration all the possible functions γ_k. If we have an observation space composed of k_0 points and k_c control values are allowed, then there will be $(k_c)^{k_0}$ different functions to consider. This could represent a considerable algebraic task for all but the most trivial systems. However, the use of an algebraic manipulation computing tool (such as "FORMAC") coupled with a rational tree searching algorithm such as "branch and bound" might enable larger systems to be solved.

Discrete space systems will not be pursued further in this thesis since emphasis will be placed on more general systems. The use of FORMAC as an aid to analysis will however be considered in connection with perturbation techniques in Chapter 6.

One final point should be made regarding the zero-memory problem. Given the fundamental equation of dynamic programming,

$$V_k^{\circ}(p_k) = \inf_{\gamma_k} \left\{ E\left(L_k(X_k, \gamma_k(Y_k)) + V_{k+1}^{\circ}(p_{k+1} = \mathcal{L}_k(p_k, \gamma_k) \right) \right\},$$

there is a temptation to take conditional expectations and then interchange the outer expectation with the infimum, obtaining a sub-problem which involves minimising with respect to $\gamma_k(Y_k)$ - thus resolving the intrinsic difficulty of finding a minimum with respect to an entire function. This will be examined more closely in Chapter 5, where it will be shown to be inappropriate. Some comment has already been made on the point in the 'Indirect approach to the Witsenhausen problem' discussed in Chapter 2.

3.3.3 Classical Information Pattern

For this type of information pattern, we have $I_n = (Y^n, U^{n-1})$ for all n=0,---N-1, and where,

$$Y^n = Y_0, Y_1, Y_2 \text{ ---} Y_n$$

$$U^{n-1} = U_0, U_1, U_2 \text{ ---} U_{n-1}$$

In other words the controller at each time instant always knows his current observation Y_n and also all information available to the previous controller together with the actual control value that the previous controller applied.

We shall proceed exactly as in the previous section. First we shall identify a suitable 'state variable', next we shall attempt to use this 'state variable' to derive an iterative relationship involving the cost, and finally dynamic programming arguments will be used to develop an algorithm for determining the optimal control laws. This class of problem has been dealt with extensively in the literature and no attempt will be made to rigorously justify our approach. This can be found elsewhere.

State Transition Equation

We saw in the previous section that an adequate 'state variable' was provided by the probability density function p_n of the state x_n at time n, where,

$$p_{n+1}(X_{n+1}) = \int_{S_n}\int_{\delta_n} p(X_{n+1} | X_n, \gamma_n(Y_n)) p(Y_n | X_n) p_n(x_n) dX_n dY_n$$

The question which arises is whether or not this 'state variable' is adequate in the classical information pattern problem. In other words, is knowledge of p_n sufficient to enable:-

$$E\left[\sum_{k=n}^{N-1} L_k(X_k, \gamma_k(Y^k, U^{k-1})) + L_N(X_N)\right]$$

to be determined? Does it summarise sufficient information about previous contol laws to define the above expression, given $\gamma_n, \gamma_{n+1} \cdots \gamma_{N-1}$. Consider a typical term in this expression such as:-

$$J_k = L_k(X_k, \gamma_k(Y^k, U^{k-1})$$

Then by successive substitution we have:-

$$J_k = L_k(F_{k-1}(X_{k-1}, \gamma_{k-1}(Y^{k-1}, U^{k-2}), W_{k-1}), \gamma_k(Y^k, \gamma_{k-1}(Y^{k-1}, U_k^{k-2}))$$

$$J_k = L_k'(\Gamma_n^k, X_n, W_n \cdots W_{k-1}, V_{n+1} \cdots V_k, Y^n, U^{n-1})$$

If we imitate the procedures of section 3.3.2 for the zero-memory problem, then we see that $E(J_k)$

$$= \int L_k'(\Gamma_n^k; X_n; W_n \cdots W_{k-1}, V_{n+1} \cdots V_k; Y^n, U^{n-1})$$

$$p(Y^n, U^{n-1} \mid X_n) p_n(X_n) d(Y^n, U^{n-1})\ dX_n d(W_n \cdots W_{k-1}; V_{n+1} \cdots V_k)$$

Although this quantity involves $p_n(X_n)$ as we were hoping to prove, it also involves the probability density function $p(Y^n, U^{n-1} \mid X_n)$.

But what is $p(Y^n, U^{n-1} \mid X_n)$? For the classical information pattern problem we know (Y^n, U^{n-1}), so $p(Y^n, U^{n-1} \mid X_n)$ is rather meaningless. Rather obviously, we should be looking at conditional expectations $E\left[J_k \mid Y^n, U^{n-1}\right]$, rather than unconditional expectations $E\left[J_k\right]$. This enables us to propose (Y^n, U^{n-1}) as a state variable, since knowledge of it enables $E\left[J_k \mid Y^n, U^{n-1}\right]$ to be computed. Unfortunately (Y^n, U^{n-1}) has the strange property that its dimension increases with time; moreover it is clearly not extendable to the nested observation case.

The problem is resolved by exploiting the fact that for the classical information pattern problem, the control data is nested. This enables some remarkable simplification to be accomplished. Since the dynamic programming formulation depends mostly upon this nesting property, the arguments will be developed using an arbitrary nested information pattern. Only at the end when it is necessary to make some more specific remarks about the classical information pattern problem will the peculiar features of this information pattern be introduced.

Thus we assume that the control data at time n is denoted by I_n and the control data at time n+1 is denoted by I_{n+1}. Since the information patterns are nested, we are assuming that,

$$I_0 \subset I_1 \subset I_2 \cdots I_n \subset I_{n+1} \subset \cdots \subset I_{N+1}$$

At some time k+1, a particular control data set denoted by I_{k+1} will be available and we can express the conditional cost function to go based on I_{k+1} as follows:-

$$E\left[\sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \,\middle|\, I_{k+1}\right]$$

in this expression I_n is a random variable of which the dimension increases as n increases. Since the information pattern is nested, let us write I_n as follows:-

$$I_n = \left\{I_{n-1}, \beta_n'(Y_n), \beta_n''(U_{n-1})\right\}$$

where β_n' and β_n'' will depend upon the particular information pattern. By using the observation equation and the control law γ_{n-1}, we have:-

$$I_n = \left\{I_{n-1}, \beta_n'(G_n(X_n, V_n)); \beta_n''(\gamma_{n-1}(I_{n-1}))\right\}$$

$$I_n = \left\{I_{n-1}, H_n'(X_n, V_n, \gamma_{n-1}(I_{n-1}))\right\}$$

By a similar process as used for the zero-memory problem, we can recursively substitute for the random variable X_n using the process equation,

$$X_n = F_{n-1}(X_{n-1}, \gamma_{n-1}(I_{n-1}), W_{n-1})$$

Taking a typical term in the conditional cost function to go, we have

$$L_p(X_p, \gamma_p(I_p))$$

$$= L_p(X_p, \gamma_p(I_{p-1}, H_p(X_p, V_p, \gamma_{p-1}(I_{p-1}))))$$

$$= \bar{L}_p(F_{p-1}(X_{p-1}, \gamma_{p-1}(I_{p-1}), W_{p-1}), \gamma_p, V_p, W_{p-1})$$

By continuing this successive substitution, we will eventually obtain,

$$L_p(X_p, \gamma_p(I_p))$$

$$= L_p'(X_{k+1}, \gamma_{k+1}(I_{k+1}), \Gamma^p_{k+2}, V^p_{k+2}, W^{p-1}_{k+1})$$

For notational convenience we have shown this expression depending on the entire functions Γ^p_{k+2}. For a particular realisation of - V^p_{k+2}, W^{p-1}_{k+1}, and I_{k+1}, only the values of Γ^p_{k+2} for this realisation are required. However, since we are proposing to integrate out V^p_{k+2} and W^{p-1}_{k+1}, the entire functions Γ^p_{k+2} will eventually be required so they have been introduced now somewhat prematurely to simplify the notation.

Also notice that although we have used the notation Γ^p_{k+2} to denote what we have called the entire functions $\{\gamma_{k+2}, \gamma_{k+2}, \text{---} \gamma_p\}$ this does not mean that we require at this stage that each γ_i is defined over its entire domain. Each γ_i is a function defined on a product space I_i. Let us denote the nesting of the control data as follows:-

$I_{k+1} = I_k \overset{\vee}{U} I_{k+1}$, where U denotes union. Thus:-

$$I_{k+2} = I_{k+1} \overset{\vee}{U} I_{k+2} = I_k \overset{\vee}{U} I_{k+1} \overset{\vee}{U} I_{k+2}$$

$$I_p = I_{p-1} \overset{\vee}{U} I_p = I_k \overset{\vee}{U} I_{k+1} \text{---} \overset{\vee}{U} I_p$$

Thus we see that if we wish to integrate out V^p_{k+2} and W^{p-1}_{k+1}, we only need various sections of $\gamma_i; i = k+2, \text{---}, p.$ defined at I_k. In other words, what we really mean by Γ^p_{k+2} is a sequence of functions $\gamma^{I_{k+1}}_{k+2}$, $\gamma^{I_{k+1}}_{k+3}, \text{---} \gamma^{I_{k+1}}_p$ where -

$$\gamma^{I_{k+1}}_{k+2} = \gamma_{k+2}(I_{k+1}, \overset{v}{I}_{k+2}),$$

$$\gamma^{I_{k+1}}_{k+2} = \gamma_{k+3}(I_{k+1}, \overset{v}{I}_{k+2}, \overset{v}{I}_{k+3})$$

and so on, where I_{k+1} is for the moment fixed.

Thus:-

$$E\left[\sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \,\middle|\, I_{k+1}\right]$$

$$= \int \sum_{n=k+1}^{N-1} L'_n(X_{k+1}, \gamma_{k+1}(I_{k+1}), \Gamma^n_{k+2}, V^n_{k+2}, W^{n-1}_{k+1})\, dP(X_{k+1} \mid I_{k+1})\, dP(W^{n-1}_{k+1}, V^n_{k+2}),$$

where $L_N(X_N)$ has been absorbed into $L'_{N-1}(X_{N-1}, \gamma_{N-1}(I_{N-1}))$

Next consider the conditional cost to go from time k, given the control data I_k. This is derived by precisely the same development as above and we obtain:-

$$E\left[\sum_{n=k}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \,\middle|\, I_k\right]$$

$$E\left[L_k(X_k, \gamma_k(I_k)) + \sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \,\middle|\, I_k\right]$$

$$= E\left[L_k(X_k, \gamma_k(I_k)) + \sum_{n=k+1}^{N-1} L'_n(X_{k+1}, \gamma_{k+1}(I_{k+1}), \Gamma^n_{k+2}, V^n_{k+2}, W^{n-1}_{k+1}) \,\middle|\, I_k\right]$$

where $I_{k+1} = \left\{I_k, H'_{k+1}(X_{k+1}, V_{k+1}, \gamma_k(I_k))\right\}$

and $X_{k+1} = F_k(X_k, \gamma_k(I_k), W_k)$

Thus,

$$E\left[\sum_{n=k}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \,\middle|\, I_k\right]$$

$$= \int L_k(X_k, \gamma_k(I_k)) p(X_k | I_k)\, dX_k$$

$$+ \int \sum_{n=k+1}^{N-1} L'_n(X_{k+1}, \gamma_{k+1}(I_k, \overset{v}{I}_{k+1}), \Gamma^n_{k+2}, V^n_{k+2}, W^{n-1}_{k+1})\, dP(W^{n-1}_{k+1}, V^n_{k+2})$$

$$dP(X_{k+1}, \overset{v}{I}_{k+1} | I_k),$$

where $I_{k+1} = I_k \overset{v}{U} I_{k+1}$ by nesting property and where $dP(X_{k+1}, \overset{v}{I}_{k+1} | I_k)$ is the Lebesgue-Stieltjes measure generated from $dP(X_k | I_k)$, P_k^w, and $P_{k+1}^{\overset{v}{}}$ by the transformations:-

$$\overset{v}{I}_{k+1} = H_{k+1}(X_{k+1}, V_{k+1}, \gamma_k(I_k))$$

$$X_{k+1} = F_k(X_k, \gamma_k(I_k), W_k))$$

In stating this equality we are employing a conditional version of the theorem exploited when discussing the zero-memory problem. We can derive a formal expression for $dP(X_{k+1}, \overset{v}{I}_{k+1} | I_k)$ by assuming that various probability density functions exist. Thus,

$$E\left[\sum_{n=k}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \,\middle|\, I_k\right]$$

$$= \int L_k(X_k, \gamma_k(I_k)) p(X_k | I_k)\, dX_k$$

$$+ \int \sum_{n=k+1}^{N-1} L'_n(X_{k+1}, \gamma_{k+1}(I_k, \overset{v}{I}_{k+1}) \Gamma^n_{k+2}, v^n_{k+2}, W^{n-1}_{k+1})\, dP(W^{n-1}_{k+1}, V^n_{k+2})$$

$$p(X_{k+1} | \overset{v}{I}_{k+1}, I_k)\, dX_{k+1}\; p(\overset{v}{I}_{k+1} | I_k)\, d\overset{v}{I}_{k+1}$$

$$= \int L_k(X_k, \gamma_k(I_k))\, p(X_k \mid I_k)\, dX_k$$

$$+ \int E\left[\sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \mid I_{k+1}\right] p(\overset{v}{I}_{k+1} \mid I_k)\, d\overset{v}{I}_{k+1}$$

where $p(X_{k+1} \mid I_{k+1})$ for the inner expectation is generated from $p(X_k \mid I_k)$ and P_k^W and P_{k+1}^V via the transformations,

$$\overset{v}{I}_{k+1} = H_{k+1}(X_{k+1}, V_{k+1}, \gamma_k(I_k))$$

and $$X_{k+1} = F_k(X_k, \gamma_k(I_k), W_k)$$

Thus,

$$p(X_{k+1} \mid \overset{v}{I}_{k+1}, I_k) = \frac{p(X_{k+1}, \overset{v}{I}_{k+1} \mid I_k)}{p(\overset{v}{I}_{k+1} \mid I_k)}$$

$$= \frac{p(\overset{v}{I}_{k+1} \mid X_{k+1}, I_k)\, p(X_{k+1} \mid I_k)}{p(\overset{v}{I}_{k+1} \mid X_{k+1}, I_k)\, p(X_{k+1} \mid I_k)\, dX_{k+1}}$$

$$= \frac{p(\overset{v}{I}_{k+1} \mid X_{k+1}, I_k) \int p(X_{k+1} \mid X_k, \gamma_k(I_k))\, p(X_k \mid I_k)\, dX_k}{\int p(\overset{v}{I}_{k+1} \mid X_{k+1}, I_k) \int p(X_{k+1} \mid X_k, I_k)\, p(X_k \mid I_k)\, dX_k\, dX_{k+1}}$$

and,

$$p(\overset{v}{I}_{k+1} \mid I_k) = \int p(\overset{v}{I}_{k+1} \mid X_{k+1}, I_k) \int p(X_{k+1} \mid X_k, I_k)\, p(X_k \mid I_k)\, dX_k\, dX_{k+1}$$

Suppose we define V_{k+1} as follows:-

$$V_{k+1}(p(X_{k+1} \mid I_{k+1}), \Gamma_{k+2}^{N-1}, \gamma_{k+1}(I_{k+1})) \triangleq E\left[\sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \mid I_{k+1}\right]$$

Then we see that,

$$V_k(p(X_k \mid I_k), \Gamma_{k+1}^{N-1}, \gamma_k(I_k)) = \int L_k(X_k, \gamma_k(I_k)) p(X_k \mid I_k) dX_k$$

$$+ \int V_{k+1}(p(X_{k+1} \mid I_{k+1}), \Gamma_{k+2}^{N-1}, \gamma_{k+1}(I_k, \overset{v}{I}_{k+1})) p(\overset{v}{I}_{k+1} \mid I_k) d\overset{v}{I}_{k+1}$$

where $p(X_{k+1} \mid I_{k+1}) = \dfrac{p(I_{k+1} \mid X_{k+1}, I_k) \int p(X_{k+1} \mid X_k, I_k) p(X_k \mid I_k) dX_k}{\int p(I_{k+1} \mid X_{k+1}, I_k) \int p(X_{k+1} \mid X_k, I_k) p(X_k \mid I_k)\, dX_k\, dX_{k+1}}$,

Thus it will be seen that knowledge of $p(X_k \mid I_k)$ at any time k is sufficient to enable the conditional cost to go from time k to be computed. This conditional cost to go we have denoted as - $V_k(p(X_k \mid I_k), \Gamma_k^{N-1}, \gamma_k(I_k))$ and a useful looking iterative relationship involving V_k and V_{k+1}, linked by our state transition equation, has been derived. Regarding the state transition equation notice that all the unknown probability density functions can easily be determined:-

$$P(\overset{v}{I}_{k+1} \in A \mid X_{k+1}, I_k) = P\left\{H_{k+1}(X_{k+1}, V_{k+1}, \gamma_k(I_k) \in A \mid X_{k+1}, I_k\right\}$$

$$= P\left\{V_k \in \bar{H}_{k+1}^{-1}(A)\right\}$$

which can be determined from P_{k+1}^{V}.

$$P(X_{k+1} \in A \mid X_k, I_k) = P\left\{ F_k(X_k, \gamma_k(I_k), W_k) \in A \mid X_k, I_k \right\}$$

$$= P\left\{ W_k \in \bar{F}_k^{-1}(A) \right\}$$

which can be determined from P_k^W. Notice that in general both $p(\overset{\vee}{I}_{k+1} \mid X_k, I_k)$ and $p(X_{k+1} \mid X_k, I_k)$ will depend upon $\gamma_k(I_k)$. Let us denote that state transition equation as,

$$p(X_{k+1} \mid I_{k+1}) = \ell_k(\overset{\vee}{I}_{k+1}, \gamma_k(I_k), p(X_k \mid I_k))$$

Equipped with a state transition equation and a corresponding iterative relationship involving the cost we can now proceed with a dynamic programming formulation of the nested information pattern problem.

Theorem 3.2

Suppose,

$$V_{k+1}^{\circ}(p(X_{k+1} \mid I_{k+1})) \triangleq \inf_{\substack{\gamma_{k+1}(I_{k+1}), \\ \Gamma_{k+2}^{N-1}(I_{k+1})}} \left\{ V_{k+1}(p(X_{k+1} \mid I_{k+1}) \Gamma_{k+1}^{N-1}, \gamma_{k+1}(I_{k+1})) \right\}$$

where we have explicitly denoted the restriction of Γ_{k+2}^{N-1} to those sections defined at I_{k+1}.

Then V_k° satisfies the following iterative relationship,

$$V_k^{\circ}(p(X_k \mid I_k)) = \inf_{\gamma_k(I_k)} \left\{ \int L_k(X_k, \gamma_k(I_k)) p(X_k \mid I_k) \, dX_k \right.$$

$$+\int V_{k+1}^{\circ}(p(X_{k+1} \mid I_{k+1}))\,p(\overset{\vee}{I}_{k+1} \mid I_k)\,d\overset{\vee}{I}_{k+1}\Big\}$$

where $$p(X_{k+1} \mid I_{k+1}) = \frac{p(\overset{\vee}{I}_{k+1} \mid X_{k+1}, I_k)\int p(X_{k+1} \mid X_k, I_k)\,p(X_k \mid I_k)\,dX_k}{\int p(I_{k+1} \mid X_{k+1}, I_k)\int p(X_{k+1} \mid X_k, I_k)\,p(X_k \mid I_k)\,dX_k}$$

Proof

This can be demonstrated easily as follows:-

By definition,

$$V_k^{\circ}(p(X_k \mid I_k) \triangleq \inf_{\substack{\gamma_k^{I_k} \\ \Gamma_{k+1}^{N-1}(I_k)}} \Big\{\int L_k(X_k, \gamma_k(I_k))\,p(X_k \mid I_k)\,dX_k$$

$$+ \int V_{k+1}^{\circ}(\mathcal{L}_k(\overset{\vee}{I}_{k+1}, \gamma_k(I_k), p(X_k \mid I_k))), \Gamma_{k+2}^{N-2}, \gamma_{k+1}(I_k, \overset{\vee}{I}_{k+1})\,p(\overset{\vee}{I}_{k+1} \mid I_k)\,dI_{k+1}$$

$$= \inf_{\gamma_k(I_k)} \Big\{\int L_k(X_k, \gamma_k(I_k))\,p(X_k \mid I_k)\,dX_k$$

$$+ \inf_{\substack{\gamma_{k+1}^{I_k} \\ \Gamma_{k+2}^{N-1}(I_k)}} \int V_{k+1}(\mathcal{L}_k(\overset{\vee}{I}_{k+1}, \gamma_k(I_k), p(X_k \mid I_k)), \Gamma_{k+2}^{N-1}, \gamma_{k+1}(I_k, \overset{\vee}{I}_{k+1})\,p(\overset{\vee}{I}_{k+1} \mid I_k)\,d\overset{\vee}{I}_{k+1}\Big\}$$

$$= \inf_{\gamma_k(I_k)} \Big\{\int L_k(X_k, \gamma_k(I_k))\,p(X_k \mid I_k)\,dX_k$$

$$+\int \inf_{\gamma_{k+1}(I_{k+1})} V_{k+1}(\mathcal{L}_k(\overset{\vee}{I}_{k+1}, \gamma_k(I_k), p(X_k|I_k)), \Gamma_{k+2}^{N-1}(I_{k+1}), \Gamma_{k+2}^{N-1}, \gamma_{k+1}(I_k, \overset{\vee}{I}_{k+1}))\, p(\overset{\vee}{I}_{k+1}|I_k)\, d\overset{\vee}{I}_{k+1}\Big\}$$

since V_{k+1} depends only on the functions γ_{k+1} and Γ_{k+2}^{N-1} through their values at $(I_k, \overset{\vee}{I}_{k+1})$. Also $p(I_{k+1}|I_k)$ does not depend on $\Gamma_{k+1}^{I_{k+1}}$, although it should be noted that $p(\overset{\vee}{I}_{k+1} \mid I_k)$ will depend on $p(X_k \mid I_k)$.

Thus,

$$V_{k+1}^{\circ}(p(X_k|I_k)) = \inf_{\gamma_k(I_k)} \Big\{\int L_k(X_k, \gamma_k(I_k))\, p(X_k \mid I_k)\, dX_k$$

$$+ \int V_{k+1}^{\circ}\mathcal{L}_k(\overset{\vee}{I}_{k+1}, \gamma_k(I_k), p(X_k \mid I_k)))\, p(\overset{\vee}{I}_{k+1} \mid I_k)\, dI_{k+1}$$

which completes the proof.

In the above proof we have used the following well known Lemma:-

Lemma

$$\inf_{\gamma \in \Gamma} \int f(x, \gamma(x))\, dx = \int \inf_{\gamma(x) \in U} f(x, \gamma(x))\, dx$$

Proof

Let $I(x) \triangleq \inf_{\gamma(x) \in U} (f(x, \gamma(x))$ and $J(\gamma) \triangleq \int f(x, \gamma(x))dx$

Then (i) $I(x) \leqslant f(x, \gamma(x))$ for all $\gamma(x) \in U$

$\Rightarrow J^* = \int I(x)\, dx \leqslant \int f(x, \gamma(x))\, dx$ for all $\gamma \in \Gamma$

$\Rightarrow J^* \leqslant J(\gamma)$ for all γ

(ii) Consider any $\epsilon > 0$ and, for every x, select $\bar{\gamma}(x)$ such that,

$$I(x) + \epsilon \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \geq f(x, \bar{\gamma}(x))$$

then $J^* + \epsilon \geq J(\bar{\gamma})$

Thus $J^* = \inf_{\gamma} \int f(x, \gamma(x))dx = \int \inf_{\gamma(x)} f(x, \gamma(x))dx$

Algorithm for $\gamma_0^\circ, \gamma_1^\circ, \gamma_2^\circ \cdots \gamma_{N-1}^\circ$

Let us consider how this theorem assists in computing an optimal sequence of controls $\gamma_0^\circ, \gamma_1^\circ, \gamma_2^\circ \cdots \gamma_{N-1}^\circ$. The algorithm is as follows.

(i) Calculate $V_N^\circ(p(X_N|I_N)) \triangleq \int L_N(X_N, \gamma_N(I_N))p(X_N|I_N)dX_N$

(ii) Calculate $V_{N-1}^\circ(p(X_{N-1}|I_{N-1})$

$$= \inf_{\gamma_{N-1}(I_{N-1})} \left\{ \int L_{N-1}(X_{N-1}, \gamma_{N-1}(I_{N-1}))p(X_{N-1}|I_{N-1})dX_{N-1} \right.$$

$$\left. + \int V_N^\circ(\mathcal{L}_{N-1}(\check{I}_N, \gamma_{N-1}(I_{N-1}), p(X_{N-1}|I_{N-1}))p(\check{I}_N|I_{N-1})\check{d}I_N \right\}$$

(iii) Calculate $V_{N-2}^\circ(p(X_{N-2}|I_{N-2}))$

$$= \inf_{\gamma_N(I_{N-2})} \left\{ \int L_{N-2}(X_{N-2}, \gamma_{N-2}(I_{N-2}))p(X_{N-2}|I_{N-2})dX_{N-2} \right.$$

$$\left. + \int V_{N-1}^\circ(\mathcal{L}_{N-2}(\check{I}_{N-1}, \gamma_{N-2}(I_{N-2}), p(X_{N-2}|I_{N-2}))p(I_{N-1}|I_{N-2})dI_{N-1} \right\}$$

(iv) The above procedure is repeated until ultimately we obtain,

$$V_0^{\circ}(p(X_0|I_0)) = \inf_{\Gamma_0^{N-1}(I_0)} E\left[\sum_{n=0}^{N-1} L_N(X_n, \gamma_n(I_n)) + L_N(X_N) \,\middle|\, I_0\right]$$

and from this it then follows trivially that:-

$$J_0^{\circ}(p(X_0)) \triangleq \int V_0^{\circ}(p(X_0|I_0))p(I_0|X_0)p(X_0)\,dX_0\,dI_0$$

$$= \inf_{\Gamma_0^{N-1}} E\left[\sum_{n=0}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N)\right]$$

In other words, we have determined the optimal cost as a function of the initial condition $p(X_0)$ as desired. Let us examine the procedure in greater depth:-

(i) at each stage k we must find the optimal (or ϵ-optimal) control value $\gamma_k^{\circ}(I_k)$. That is, we only require the value of γ_k° at I_k. This is a considerable advantage over the situation for the zero memory problem where we must find the entire optimal (or ϵ-optimal) control function, which as we have learned from the examples in Chapter 2, can be a difficult problem.

(ii) at each stage k, the optimal control value $\gamma_k^{\circ}(I_k)$ is parameterised by the state variable $p(X_k|I_k)$. That is

$$\gamma_k^{\circ}(I_k) = \mathcal{F}_k(p(X_k|I_k)),$$

and in the general case, for each I_k, $\gamma_k^{\circ}(I_k)$ is parameterised by an infinite dimensional quantity In order to eliminate $p(X_k|I_k)$ from our control laws, we must compute the full sequence $\gamma_{N-1}^{\circ}, \gamma_{N-2}^{\circ}, \cdots \gamma_0^{\circ}$ as defined above, and then use some estimation equation $p(X_{n+1}|I_{n+1}) = \mathcal{L}_n(\overset{v}{I}_{n+1}, p(X_n|I_n), \bar{\gamma}_n(I_n))$ to compute each $p(X_n|I_n)$ $n=0,1,\cdots N-1$ and thereby eliminate it from the expression for $\gamma_n^{\circ}(I_n)$. The procedure is as follows:-

Compute $\gamma_{N-1}^{\circ}(I_{N-1}) = \mathcal{Y}_{N-1}(p(X_{N-1}|I_{N-1}))$

$$\gamma_{N-2}^{\circ}(I_{N-2}) = \mathcal{J}_{N-2}(p(X_{N-2}|I_{N-2}))$$
$$\vdots$$
$$\gamma_0^{\circ}(I_0) = \mathcal{J}_0(p(X_0|I_0))$$

But $p(X_0|I_0) = \dfrac{p(I_0|X_0)p(X_0)}{\int p(I_0|X_0)p(X_0)\,dX_0}$, which is known

Thus $\bar{\gamma}_0^{\circ}(I_0) = \mathcal{J}_0'(p(X_0), I_0)$,

$$\bar{\gamma}_1(I_1) = \mathcal{J}_1(p(X_1|I_1))$$
$$= \mathcal{J}_1(\mathcal{L}_0(p(X_0|I_0), \bar{\gamma}_0(I_0), \overset{v}{I}_1)$$
$$= \mathcal{J}_1'(p(X_0), I_1)$$
$$\bar{\gamma}_2(I_2) = \mathcal{J}_2(p(X_2|I_2))$$
$$= \mathcal{J}_2(\mathcal{L}_1(p(X_1|I_1), \overset{v}{I}_2, \bar{\gamma}_1(I_1))$$
$$= \mathcal{J}_2'(p(X_0), I_2)$$ by successive substitution,

and so on.

In eliminating $p(x_k \mid I_k)$ from the expression for γ_k by the above recursive procedure, we are substituting 'open-loop' for 'closed-loop' control in the sense described in the previous Chapter. Our control is still closed-loop however in that u_k is a function of the control data I_k and this is a random variable which is not known with certainty. It is open-loop however in that at all instants up to time k the optimal control value has been applied according to $\gamma_i(I_i)$ $i = 0,1,\cdots k$. If $u_i \neq \gamma_i(I_i)$ for some i then we will not be in the state we think we are and $\gamma_k(I_k) = \mathcal{F}_k(\mathcal{L}_k(p(x_{k-1} \mid I_{k-1}), \check{I}_k, \gamma_{k-1}(I_{k-1}))$ may not be the optimal control to apply.

In appendix 2, we present details of the application of Theorem 3.2 to the traditional linear system with quadratic cost and gaussian noise (the LQG problem). Two versions are studied - the classical information pattern problem and the nested observation problem. It will be seen that Theorem 3.2 alone is not sufficient to enable a solution to be found. In addition it is necessary to exploit the particular structure of the problem to show that $p(x_n \mid I_n)$ is guassian (if x_0, $\{v_i\}$ and $\{w_i\}$ are gaussian). This property is used to derive the famous Kalman filtering equations. The gaussianity of $p(x_n \mid I_n)$ means that it can be summarised in terms of its equivalent statistics; that is, its mean $x_{n/n}$ and its variance $\sigma^2_{n/n}$. Moreover, it turns out that the variance $\sigma^2_{n/n}$ is independent of the control laws which are applied. This property, which is sometimes called neutrality, is crucial since it permits a tremendous simplification in the complexity of the problem, enabling us to replace the infinite dimensional quantity $p(x_n \mid I_n)$

in Theorem 3.2 with its so-called sufficient statistic $x_{n/n}$, which is finite dimensional. Here the term sufficient is used with respect to the optimisation problem. We thus obtain Theorem 3.3 (see Appendix page A2.12) as a particular version of Theorem 3.2 in which $x_{n/n}$ is substituted for $p(x_n \mid I_n)$.

Comments on Classical Information Pattern Problem

We have now concluded our dynamic programming formulation of the classical information pattern problem. We have had to work surprisingly hard. The problem has been dealt with in detail because it is the author's experience that, despite a number of correct results on the subject, all manner of dubious arguments are used in their derivation. In fact Chapter 5 will be devoted to describing a number of common pitfalls for this and other classes of problem.

It is not claimed that the analysis leading up to Theorem 3.2 is in any sense the best that can be presented. But it does have the advantage of being explicit and direct. It must also be admitted however that some amplification, at least of the notation, is desirable:-

We have demonstrated that the probability density function $p(X_n|I_n)$ is suitable as a 'state' variable for our problem. We showed that knowledge of $p(X_n|I_n)$ summarised sufficient information about past controls in order to enable the conditional cost to go to be evaluated and we defined,

$$V_k(p(X_k I_k), \Gamma_{k+1}^{N-1}, \gamma_k(I_k)) \triangleq E\left[\sum_{k=k}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \mid I_k\right]$$

This is somewhat ambiguous and is worthy of further consideration. The notations $\gamma_k(I_k)$, Γ_{k+1}^{N-1}, and $p(X_k(I_k)$ denote quantities defined at I_k; knowledge of these values alone is sufficient to determine the conditional cost to go. Thus the variable I_k is not needed as an individual argument in V_k. It is implicit in the other arguments.

Thus we have,

$$V_k(p(X_k|I_k), \Gamma_{k+1}^{N-1}, \gamma_k(I_k))\Big| = E\left[\sum_{L=k}^{N-1} L_n(X_n, \gamma_n(I_n)) + L_N(X_N) \mid I_k = I_k'\right]$$

It is this implicit dependence upon I_k which enables I_k to be more or less ignored until we wish to bring it back. In other words, it permits the separation of the problem into an optimisation problem and an estimation problem as follows:-

Optimisation: Determine $\gamma_k^{\circ}(I_k)$ as a functional of $p(X_k|I_k)$. Clearly I_k will not appear explicitly in this problem. We shall obtain:-

$$\gamma_k^{\circ} = \mathcal{F}_k(p(X_k|I_k))$$

Estimation: Express $p(X_k|I_k)$ now as an explicit function of I_k and obtain $\gamma_k^o(I_k) = G(I_k)$. In the general sense this is an estimation problem.

Some authors have expressed the notion that I_k is implicit in V_k by saying that I_k is fixed. This is a somewhat misleading description however.

As was stated at the outset, our formulation was primarily for the nested information case and we have chosen to treat the classical information pattern as a special case of the former. The disadvantage of this approach is of course that, somewhat trivially, functions such as $\delta_i(U_i - \gamma_i(I_i))$ keep appearing and then cancelling out. We could avoid all this for the classical information pattern problem by taking as our state transition equation,

$$p(X_{n+1}|I_{n+1}) = \frac{p(Y_{n+1}|X_{n+1})\int p(X_{n+1}|X_n,U_n)p(X_n|I_n)dX_n}{\int p(Y_{n+1}|X_{n+1})\int p(X_{n+1}|X_n,U_n)p(X_n|I_n)dX_n}$$

There is no fundamental difference between this approach and the one adopted earlier; which one is adopted is largely a matter of taste. The author has an aversion to mentioning U_n's at all since it is preferred to emphasize that functions γ_i are involved and that these are functions of I_i.

The above discussion has been presented to explain why nowhere have terms such as U_k^o been introduced. We have preferred to continually demonstrate the implicit dependence upon I_k.

The analysis of the simple scalar versions of the classical information pattern problem and the nested observation problem has demonstrated that, as far as a dynamic programming formulation is concerned, there is little difference between the two problems. It can easily be demonstrated in fact that for either problem,

$$\gamma_k^{\circ}(I_k) = \mathcal{F}_k(p(X_k|I_k))$$

where $\mathcal{F}_k$ is identical for both problems. Thus the optimisation parts of the problem are the same. The only difference between the two problems is in the estimation. There is no mandatory reason why the state transition equation used to express $p(X_{k+1}|I_{k+1})$ in terms of $p(X_k|I_k)$ should be used in the reverse order to enable $p(X_k|I_k)$ to be estimated. The right answers will be obtained by using the right value of $p(X_k|I_k)$ - however this is obtained. It so happens that in the classical information pattern case $p(X_k|I_k)$ can be estimated with considerably more certainty than in the nested observation problem. Since, in the classical case we know all past control values, we can determine $p(X_k|I_k)$ using the estimation equations,

$$p(X_k|I_k) = \frac{p(Y_k|X_k)\int p(X_k|X_{k-1},U_{k-1})p(X_{k-1}|I_{k-1})\,dX_{k-1}}{\int p(Y_k|X_k)\int p(X_k|X_{k-1},U_{k-1})p(X_{k-1}|I_{k-1})\,dX_{k-1}dX_k}$$

For the nested observation case we have to use the estimation equations,

$$p(X_k|I_k) = \frac{p(Y_k|X_k)\int p(X_k|X_{k-1},\gamma_{k-1}^{\circ}(I_{k-1})p(X_{k-1}|I_{k-1})\,dX_{k-1}}{\int p(Y_k\,X_k)\int p(X_k\,X_{k-1},\gamma_{k-1}^{\circ}(I_{k-1}))p(X_{k-1}|I_{k-1})\,dX_{k-1}}$$

These just happens to be the same as the state transition equations - although now we are proposing that we iterate from k-1 to k instead of the other way round. The correct estimation of $p(X_k|I_k)$ for the nested observation case depends on the assumption that each control applied is given by $\gamma_{k-1}^{o}(I_{k-1})$ - the optimal control value. If, of course, at some preceding instant the control applied is not equal to $\gamma_{k-1}^{o}(I_{k-1})$ the estimate of $p(X_k|I_k)$ will be wrong and consequently so will the optimal control $\hat{\gamma}_k(p(X_k|I_k))$. The point is that in both problems we must assume that at future instants the optimal control value will be given by $\gamma_n^{o}(I_n)$, $n>k$. If this is not the case then $\gamma_k^{o}(I_k)$ will not be optimal for either problem, however good our estimate of the state $p(X_k|I_k)$ is. The above comments are only of engineering significance. Mathematically, the two problems are equivalent.

As was stated earlier, the dynamic programming formulation of the classical information pattern problem has proved quite hard work. This is surprising in view of the ease with which a solution was obtained for the examples presented in Chapter 2. One is left to wonder if dynamic programming is the best approach for the classical information pattern problem. An alternative to dynamic programming is the so called person-by-person optimisation technique, where the terminology is borrowed from game theory. It was a variant of this technique which was applied in Chapter 2. More will be said of this approach in the next section where we investigate the solution of problems with general information patterns.

3.3.4 General Information Pattern

At the beginning of 3.3.1, we demonstrated that our stochastic optimal control problem can be posed as follows:-

$$V(\Gamma_{N-1}, p_0) \triangleq E\left[J(X^N, U^N)\right]$$

$$= E\left[\sum_{n=0}^{N-1} L_n^{\gamma_0\gamma_1\cdots\gamma_n}(X_0; \{V_k : k = 1-n\}; \{W_k : k = 1-n\}) + L_N^{\gamma_0\gamma_1\cdots\gamma_{N-1}}(X_0; \{V_k : k = 1-N-1\}; \{W_k : k = 1-N-1\})\right]$$

where p_0 is the joint distribution function of the primitive random variables. The problem then is to find $V^\circ(p_0)$ such that,

$$V^\circ(p_0) = \inf \quad V(\Gamma_{N-1}, p_0)$$

$$\Gamma^{N-1} \in \textcircled{H}_1 \times \textcircled{H}_2 \times \textcircled{H}_3 \times \cdots \times \textcircled{H}_{N-1}$$

In looking for a dynamic programming formulation, we are really trying to decompose this problem into a number of simpler but interlinked sub-problems. In other words, we are attempting to determine infima with respect to a single function and not an entire sequence of functions. We have shown how this can be achieved for both zero-memory problem and the classical information pattern problem. Indeed, for the classical information pattern problem we have the added bonus that we are finding infima with repsect to a function value and not an entire function. Let us next investigate the possibility of a dynamic programming formulation for the general information pattern problem.

As usual, we need to do two things:-

i) Identify some quantity which we can interpret as a 'state' and which is governed by a state transition equation.

ii) Derive an iterative relationship which involves the costs to go, V_k and V_{k+1}, for some times k and k+1 respectively, and where V_k has as arguments the 'state' and the control laws sequence $\gamma_k, \gamma_{k+1} \cdots \gamma_{N-1}$. As far as the state at time k is concerned, it must summarise all past behaviour of the system.

Let us look at this problem in the light of what we have discovered for our two previous special cases:-

i) The unconditional probability density function $p(X_k)$ is not suitable as a state variable for the classical information pattern problem since it will involve the term $p(Y^k, U^{k-1} \mid X_k)$ and this is a rather meaningless quantity since (Y^k, U^{k-1}) is is assumed known in the classical information pattern problem. Thus we must conclude that $p(X_k)$ is not suitable as a state variable for the general information pattern problem.

ii) The other alternative for a state variable which our previous examples suggest is $p(X_k \mid I_k)$ which we saw was suitable for the classical information pattern case. The question is whether or

not it is suitable for the general problem. Let us investigate this question for the zero-memory problem.

Consider the situation at some time k+1 assuming that appropriate density functions exist, we have:-

$$V_{k+1} \overset{\Delta}{=} E\left[\sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(Y_n)) + L_N(X_N) \,\middle|\, Y_{k+1}\right]$$

As ususal, taking a typical term in this expression, we have

$$\begin{aligned} L_n(X_n, \gamma_n(Y_n)) &= L_n(X_n, \gamma_n(G_n(X_n, V_n)) \\ &= L_n^{\gamma_n}(V_n, F_{n-1}(X_{n-1}, \gamma_{n-1}(Y_{n-1}), W_{n-1}) \\ &= L_n^{\gamma_n \gamma_{n-1}}(V_n, X_{n-1}, Y_{n-1}, W_{n-1}) \\ &= L_n^{\gamma_n \gamma_{n-1} \cdots \gamma_{n-1}}(V_n \cdots V_{k+1}, X_{k+1}; W_{n-1} \cdots W_{k+1}) \end{aligned}$$

Thus we see that,

$$E\left[L_n(X_n, \gamma_n(Y_n)) \,\middle|\, Y_{k+1}\right] = C_{k+1}(p(X_{k+1} \,|\, Y_{k+1}), \Gamma_{k+2}^{n}, \gamma_{k+1}(Y_{k+1}))$$

Notice that γ_{k+1} is only required at Y_{k+1}, whilst the entire functions Γ_{k+2}^{N} will be required in order to evaluate the expectation with respect to $p(X_{k+1} \,|\, U_{k+1})$. Next we can evaluate $E\left[L_n(X_n, \gamma_n(Y_n)) \,\middle|\, Y_k\right]$ in the same manner, thus obtaining,

$$E\left[L_n^{\gamma_n \gamma_{n-1} \cdots \gamma_{k+1}(Y_{k+1})}(V_n \cdots V_{k+2}; X_{k+1}; Y_{k+1}; W_n \cdots W_{k+1}) \,\middle|\, Y_{k+1}\right]$$

where we propose to substitute,

$$X_{k+1} = F_k(X_k, \gamma_k(Y_k), W_k)$$

$$Y_{k+1} = G_{k+1}(X_{k+1}; V_{k+1})$$

By precisely the same procedures used for the classical information pattern problem in Section 3.3.3, we obtain,

$$E\left[L_n(X_n, \gamma_n(Y_n)) \middle| Y_k\right] = \int L_n \, \gamma_n \gamma_{n-1} \cdots \gamma_{k+1}(Y_{k+1}) \, (V_{k+2}^n; X_{k+1}; Y_{k+1}; W_{k+1}^n)$$

$$p(X_{k+1}, Y_{k+1} \mid Y_k) \, dX_{k+1} \, dY_{k+1}$$

$$= \int L_n \, \gamma_n \gamma_{n-1} \cdots \gamma_{k+1}(Y_{k+1}) \, (V_{k+2}^n; X_{k+1}; Y_{k+1}; W_{k+1}^n)$$

$$p(X_{k+1} \mid Y_{k+1}, Y_k) \, dX_{k+1} \; p(Y_{k+1} \mid Y_k) \, dY_{k+1}$$

$$= \int C_{k+1}(p(X_{k+1} \mid Y_{k+1}, Y_k), \gamma_{k+1}(Y_{k+1}), \Gamma_{k+2}^n)$$

$$p(Y_{k+1} \mid Y_k) \, dY_{k+1}$$

$$\neq \int C_{k+1}(p(X_{k+1} \mid Y_{k+1}), \gamma_{k+1}(Y_{k+1}), \Gamma_{k+2}^n) p(Y_{k+1} \mid Y_k) \, dY_{k+1}$$

Thus we see that $p(X_k \mid Y_k)$ is not suitable as a state variable since we cannot derive a satisfactory iterative relationship involving the conditional costs to go based upon it. This is not surprising since we are no longer dealing with nested information patterns. Of course, in searching for state variables and related conditional costs, we are

striving to find costs which at time $k = 0$ are conditioned on Y_0 and which can easily be related back to the unconditional cost at time $k = 0$. In this case we are then assured of a solution to our original statement of the problem which is to find a minimum of the expected unconditional cost. In other words we find a sequence of control laws which minimise $E\left[J \middle| Y_0\right]$ for all Y_0. This then ensures that $E(J)$ is minimised. If we are prepared to admit different forms of cost function then it may be possible to find some form of dynamic programming approach. For example, suppose we chose to minimise,

$$E(J \mid Y_{n-1} \mid Y_{n-2} \mid \cdots \mid Y_0)$$

(where it must be carefully defined), then it may well be that a dynamic programming approach can be found. This and other related topics will be discussed in the Chapter 5. For the moment, let us merely assert that a dynamic programming formulation for the general information pattern problem cannot easily be found and therefore is probably not worth pursuing further.

Thus we must conclude that the only approach to the problem is to treat it as one vast closely coupled problem in which we have to find an infimum with respect to an entire sequence of control laws. We are stuck with the problem as stated at the beginning of this section:-

Find $V^{\circ}(p_0)$ such that,

$$V^{\circ}(p_0) = \inf \; V(\Gamma_{N-1}, p_0)$$

$$\Gamma_{N-1} \in \textcircled{H}_1 \times \textcircled{H}_2 \times \textcircled{H}_{N-1}$$

We can, rather trivially find a necessary condition for the members of the optimal sequence $\Gamma^{\circ}_{N-1} = \gamma_0^{\circ}, \gamma_1^{\circ}, \gamma_2^{\circ}, \cdots \gamma_{n-1}^{\circ}$. It is as follows:-

$$V(\gamma_0^{\circ}, \gamma_1^{\circ}, \cdots \gamma_i^{\circ}, \cdots \gamma_{N-1}^{\circ}, p_0) = \inf_{\gamma_i \in \Gamma_i} V(\gamma_0^{\circ}, \gamma_1^{\circ}, \cdots \gamma_i, \cdots \gamma_{N-1}^{\circ}, p_0)$$

This condition is sometimes referred to as person-by-person optimality. By using variational claculus, the condition can conceivably be used to generate a set of necessary conditions from which the optimal control law sequence can be constructed.

Notice that we can write the condition in a somewhat different format by taking expectations with respect to θ, the primitive random variables.

$$E(J(\theta, \gamma_0^{\circ}, \gamma_1^{\circ}, \cdots, \gamma_i^{\circ}, \cdots \gamma_{N-1})) = \inf_{\gamma_i \in \Gamma_i} E_{\theta}\left[J(\theta, \gamma_0^{\circ}, \gamma_1^{\circ}, \cdots, \gamma_i, \cdots \gamma_{N-1}^{\circ})\right]$$

$$= \inf_{\gamma_i \in \Gamma_i} E_{I_i}\left[E_{\theta}\left[J(\theta, \gamma_0^{\circ}, \gamma_1^{\circ}, \cdots \gamma_i \cdots \gamma_{N-1}^{\circ}) \mid I_i\right]\right]$$

Formally, since $p(I_i)$ does not depend upon γ_i, we can take the infimum inside the outer expectation, obtaining -

$$E_{\theta}\left(J(\theta,\gamma_0^{\circ},\gamma_1^{\circ}\text{--},\gamma_i^{\circ},\text{--}\gamma_{N-1}^{\circ})\right) = E_{I_i}\left[\inf_{\gamma_i \in \Gamma_i} E_{\theta}\left[J(\theta,\gamma_0^{\circ},\gamma_1^{\circ},\text{--}\gamma_i,\text{--}\gamma_{N-1}^{\circ}) \mid I_i\right]\right]$$

There is a temptation to suspect that this modified necessary condition implies that we have a scalar minimisation problem as opposed to one involving minimisation with respect to an entire function. This unfortunately is not the case since $E_{\theta}\left[J(\theta,\gamma_0^{\circ},\gamma_1^{\circ},\text{--}\gamma_i\text{--},\gamma_{N-1}^{\circ}) \mid I_i\right]$ will generally still depend on γ_i throughout its entire range and not merely its value at $\gamma_i(I_i)$. Thus techniques such as differentiation with respect to $\gamma_i(I_i)$ and setting the resultant expression equal to zero should be treated with care. As we have seen from the Witsenhausen example in Chapter 2, second guessing terms such as $E(F_i(\gamma_i(I_i)) \mid I_k \mid I_i)$ will occur and these can only be evaluated if the entire function - $\gamma_i(I_i)$ is available. Thus variational techniques must be employed to obtain the necessary conditions.

3.4 Summary and Conclusions

One of the major objectives of this chapter has been to introduce a system of notation which will be used throughout the remainder of this thesis. It is not claimed that the notation adopted is the best in the sense that it is either the most economical or the least ambiguous; these are to a certain extent contradictory ideals. What has been selected is more of a compromise between the two. Thus we have used the notation $p(X_n \mid I_n)$ to denote the probability density function which maps a point in the product space of X_n and I_n (which is itself a product space) into a point on the real line. It can be justifiably argued that this notation might cause confusion with the point $p(X_n \mid I_n)$ which is the value of the density function at X_n, I_n. Thus alternatives such as $p_{n/n}$, $p_{n/n}(\quad)$, $p_{Xn/In}$ and so on can be proposed to represent the density function but these suffer from the disadvantage that they are more cumbersome and in some cases the random variables which they describe are not obvious. The difficulty arises of course because in probability theory the symbol p is traditionally used to denote probabiltiy density functions and the argument is left to indicate which random variable the probabiltiy density function describes. Consequently there is always a lack of discrimination between a function and a function value. Despite this, the traditional approach has its advantages; not the least being that it is widely accepted. This is the reason it has been adopted here. The arguments for and against, at least for probabiltiy density functions, do seem a little tedious since for an arbitrary argument a function is always defined by its values anyway. No confusion seems likely regarding the

interpretation of $p(X_n|I_n), p(X_n)$, and so on in the analysis presented earlier.

One inadequacy of the notation $p(X_n|I_n)$ and $p(X_n)$ is that these quantities have no meaning until the previous control laws or at least control values have been specified. Thus it could be argued that these controls should be explicitly shown, as $p(X_n|I_n;\Gamma_n(I_n))$ or $p(X_n;\Gamma_{n-1})$ for example. In the interestsof brevity this dependence upon previous controls has not been included in the notation. And in any case, since we are interpreting these probability density functions as state variables, it is thus implicit that they depend upon past controls in one way or another. In deterministic control theory it is not usual to write the state variable as $X(t,U(t)\ o\leq t\leq t)$ and there is no real reason why an equivalent notation should be adopted here.

The above remarks all apply generally to probability density functions. The situation regarding controls is a little more delicate. Here it is important to distinguish quite clearly between functions γ and function values $\gamma(I)$ and a consistent approach has been adopted. In fact we have stubbornly resisted all temptation to write U for $\gamma(I)$ even when such a substitution seems desirable. This is because we have preferred to consider to U as control data and $\gamma(I)$ as something to be determined. This intransigence does admittedly lead to the inclusion of some Dirac delta functions such as $\delta(U-\gamma(I))$ which do not lend elegance to the analysis. On the other hand, it does enable the nested information case to be treated very generally in its own right without appeal to the results for the classical information pattern. It is

also possible that the approach will permit the inclusion of randomised control strategies into the analysis, although this has not been considered here. All of the notation has been defined as it has been required. A summary of the notation and symbols used is presented for reference purposes in Appendix 1 .

Much of the notation was introduced when a precise statement of the stochastic optimal control problem was presented in Section 3.2. The various features of the stochastic optimal control problem were carefully described in order to emphasize a number of important points.

i) The system to be controlled has two major components, one component is described by the so-called process equation and the other by the so-called observation equation.

ii) **Next** we considered the evolution of the system and concluded that our problem of minimising the cost function has no meaning until the controls to be applied at each time instant have been specified as function of the available data.

iii) **This** led us on to a discussion of what is meant by available data in terms of the information pattern of a problem. A general definition of the concept of an information pattern was given and various particular cases, such as the zero-memory problem and the classical information pattern problem, were identified.

iv) Finally we saw that even when the information pattern has been specified and the evolution of the system described in terms of the various control functions applied at each instant, the function still depends on a number of random variables, termed

the primitive random variables. This led to a consideration of a number of ways of eliminating the dependence on the primitive random variables. One approach, and the one which will generally receive the greatest attention throughout the remainder of this thesis, is to define the objective as finding a sequence of control functions which will minimise the expected value of the cost function. Various alternatives to this approach were mentioned however.

Having at last defined the problem, attention was directed towards the problem of finding solutions to it. The technique of Dynamic Programming was discussed in this respect and attempts were made to apply the technique to three classes of problem - the zero memory problem, the nested information problem, and then the general information pattern problem.

For the first two cases, the zero memory problem and the nested information pattern problem, mathematical rigour was deliberately avoided so that the underlying concepts of the approach should not be obscured. However since a careful evolutionary approach was adopted for the derivations of the theorems for the two cases considered, it is felt that rigour could be introduced fairly straightforwardly. The major difficulties will be in establishing the regularity of the various probability distribution functions and in handling the technicalities which must be introduced to support the definition of those conditional distribution functions which are required. No rigorous

justification of the nested information pattern results has been attempted. This problem has been reported extensively in the literature. Comparitively recently, Striebel[1] has provided a comprehensive and rigorous development of the classical information pattern case. This development is not, however, based on dynamic programming principles.

The analysis of the zero memory problem concluded with the derivation of Theorem 3.1, which provides a technique for decomposing the problem into a number of interlinked sub-problems. In the sense that only the current data at each time instant is available, this Theorem can be viewed as demonstrating a form of Separation property for the zero - memory problem. The difficulty in the application of Theorem 3.1 is that it requires the finding of an infimum with respect to an entire function and not merely a function value. As was seen in the example of the zero-memory problem considered in Chapter 2, finding a infimum with respect to an entire function can be a difficult task.

The next type of information pattern considered was the nested information pattern case and the results for this problem were expressed in Theorem 3.2. This theorem has the distinct advantage that it requires an infimum to be found with respect to a function value which is a much more attractive prospect than that demanded by Theorem 3.1. Nevertheless even for a simple LQG problem it was quickly discovered that Theorem 3.2 alone was not sufficient to solve the problem. It was found necessary to anticipate the structure of the state variable $p(X_n \mid I_n)$. It was discovered that whatever control function γ_k was applied for $k \leq n$, the state variable $p(X_n \mid I_n)$ would always have

1. C. Striebel "Optimal Control of Discrete Time Stochastic Systems". (49)

Gaussian structure, with mean $\bar{X}_{n/n}$ and variance $\bar{\sigma}^2_{n/n}$. Moreover, it was shown that even if $\bar{X}_{n/n}$ depends on the data and previous control values, $\bar{\sigma}^2_{n/n}$ was independent of both these quantities. This remarkable good fortune enabled the 'state' to be replaced by $\bar{X}_{n/n}$ where $\bar{X}_{n/n}$ is a so-called equivalent statistic. The benefits of this simplification were expressed in the form of Theorem 3.3 (which is only applicable to the examples considered). Generalisations of Theorem 3.3. for an arbitrary LQG problem are available from several sources in the literature. The most satisfactory seems to be due to Striebel[1].

Finally, for a general information pattern problem, it was not found possible to derive a suitable dynamic programming formulation. All that we could do was restate the problem in terms of so-called person-by-person optimality. For a general information pattern problem we must consider the complete sequence of control functions at once. Decoupling the problem into a sequence of inter-linked sub-problems does not appear feasible.

The author must accept full responsibility for the entire contents of this chapter. At the time this work was being completed, there was little interest in the zero-memory problem and it is believed that results relating to it, such as Theorem 3.1, had never appeared in the literature. Regarding the nested information pattern problem, the contents of Theorem 3.3 and Theorem 3.2 had been long established. Two notable papers, regarding Theorem 3.2, were published independently by Aoki[2] and Astrom[3]. However neither of these papers consider the impact which the nesting of the information pattern makes upon the comparative ease with which a solution is obtained. Both authors tend to emphasize the im- (please continue over/..)

1. Striebel "Sufficient Statistics in the Optimum Control of Stochastic Systems" 1965 (49) 2. Aoki "Optimal Control of Partially Observable Markov Systems" 1965 (2) 3. Astrom "Optimal Control of Markov Processes with Incomplete State Information" 1965 (4).

portance of establishing a Markov property for various collections of random variables. This enables a state transition equation to be proposed but alone this is not sufficient. It must be demonstrated also that this state transition equation can be incorporated in a useful recursive relationship involving costs. Thus Aoki cites as a prime example of a Markov process the zero-memory problem characterised by the admissable class of control policies of the form $U_n = \gamma_n(Y_n)$. In this case it is true that,

$$p(X_{n+1}, Y_{n+1} \mid X^n, Y^n) = p(X_{n+1} \mid X_n = U_n = \gamma_n(Y_n))\, p(Y_{n+1} \mid X_{n+1})$$

$$= p(X_{n+1}, Y_{n+1} \mid X_n, Y_n),$$

and so $\{(X_{n+1}, Y_{n+1})\}$ represents a Markov chain. Much of Aoki's work is then dedicated to demonstrating that quantities such as (X_n, Y^n) also represent Markov chains. One is therefore strongly tempted to conclude that once the Markov property has been established, the general technique (which is Theorem 3.2) proposed by Aoki can easily be applied. As we have seen, this is not the case! Astrom's approach is somewhat parallel, although the analysis is restricted to discrete Markov chains so that mathematical rigour can be incorporated fairly easily. But again the importance of the nesting of the information in enabling us to write $E\left[J \mid I_k\right] = E\left[E\left[J \mid I_{k+1}\right] \mid I_k\right]$ is not stressed. It must be admitted however that both authors do quite explicitly state the information pattern upon which their control laws are based. Unfortunately they do not mention how significant this information pattern is. What is more, as we have seen in our statement of the problem in Section 3.2, it is important to view the problem as one evolving in time and only

to consider the problem as well defined when control functions have been specified. Both authors however develop their analysis in terms of control values.

As mentioned earlier, Striebel provided a very satisfactory discussion of the use of sufficient statistics in the classical information pattern problem. Unfortunately the introduction of the notion of sufficient statistics tended to obscure the general case.

Thus it was concluded that a satisfactory general dynamic programming analysis of the nested information pattern problem was not available. This provides the motivation for the careful derivation of Theorem 3.2 provided here. It is believed that this derivation highlights a number of points which are worthy of emphasis:-

i) We deal with sections of control function sequences at each time instant.

ii) The control data becomes an implicit variable in the formulation and can be ignored until it is desired to derive control functions which explicitly depend upon it.

iii) The formulation throughout incorporates control functions. It is just lucky that we can reduce the problem to one involving control values.

It could well be argued that the contents of this Chapter are merely a restatement of a number of facts which at the time had already been established to everybody's satisfaction. Evidence that this was not the case, and perhaps still is not the case, is provided by Chapter

5 which is devoted to considering a number of pitfalls which threaten the traveller in stochastic optimal control.[1]

However, in order to consolidate the techniques developed in this Chapter and to provide an example which demonstrates how easily misunderstandings can arise, the next Chapter is devoted to the solution of a very simple discrete space example.

1. Support for this view has been provided only comparatively recently by the publication in 1977 by O.L.R. Jacobs (30) of a paper which points out the deficiencies in the then current literature and develops a result very similar to Theorem 3.2.

CHAPTER 4

A Discrete State Example

		Page
4.1	Introduction	4.1
4.2	Problem Definition	4.3
4.3	Solution for Zero- Memory Case	4.6
4.4	Solution for Classical Information Pattern Case	4.8
4.5	Comparison Between the Two Cases	4.12
4.6	Summary and Conclusions	4.28

4.1 Introduction.

The major purpose of this chapter is to demonstrate an application of the techniques which were developed in the previous chapter. Although in the course of our discussion we have already introduced a number of examples, these have always been of the "linear, quadratic, gaussian" variety. Such examples are ideal for the study of classical information pattern problems but, as we have seen in Chapter 2 with the Witsenhausen problem, they are intractable when a non-classical information pattern is considered.

The problem to be studied here will be extremely simple, involving a discrete state space and controls applied only at two instants in time. Additionally, the cost function will depend only upon the terminal state of the system and no attempt will be made to cost the control effort. The implication of this restriction is that there will be no conflict between control and communication which was a feature of the zero-memory problem studied in Chapter 2. Finally, to reduce the cardinality of the problem, a bang-bang control is all that is permissable where the control values can only assume the values ± 1. Despite its deliberate simplicity, the problem still retains much of the flavour of the traditional stochastic control problem.

Optimal solutions will be derived for both the zero-memory and classical information versions of the problem. It will be seen that although the problem is almost trivial in construction, the determination of these optimal solutions is extremely tedious and involved.

In fact, the task of extending the problem to involve the determination of the control laws at further instants in time appears to be an impossible prospect. When finally, the optimal control laws have been computed for the two versions of the problem, it will be seen that only a very minor difference in these control laws appears. This difference however will be shown to have an interesting physical interpretation.

In addition to our prime objective of demonstrating the application of Theorems 3.1 and 3.2 of Chapter 3 to a particular example, the current chapter also prepares the way somewhat for Chapter 5, which discusses some of the common misunderstandings which can arise when non-classical information patterns are considered. Our example, however, does not refute these misunderstandings; it merely shows how easily they can be motivated and subsequently supported by inadequate test examples.

Please note that the solutions presented in this chapter have been derived via the direct application of Theorems 3.1 and 3.2 in order to demonstrate a general technique. We have not exploited the linearity of the system nor the quadratic cost. For the example considered here, the classical information pattern version could have been solved using the Certainty Equivalence Principle (see Chapter 9 - Section 9.2.4 - Page 9.15).

4.2 Problem Definition

In terms of the notation and terminology introduced in the previous Chapter, we shall consider a system with process equation defined as follows:-

$$x_{n+1} = x_n + u_n + w_n \quad \text{for } n = 0,1,$$

where the process noise w_n is a discrete random variable described by the probabilities:-

$$P\begin{bmatrix} w_n = 1 \\ w_n = 0 \\ w_n = -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \qquad \text{for } n = 0,1$$

In other words, P_n^w is a time invariant jump function. Also our observation equation is defined as:-

$$y_n = x_n + v_n, \quad \text{for } n = 0,1,$$

where the observation noise v_n is a discrete random variable described by the probabilities

$$P\begin{bmatrix} v_n = 1 \\ v_n = 0 \\ v_n = -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \qquad \text{for } n = 0,1$$

In this case P_n^v is also a time invariant jump function.

The initial state of the process at time n = 0 is also assumed to be a discrete random variable described by the probabilities:-

$$P_0^X \begin{bmatrix} x_0 = 1 \\ x_0 = 0 \\ x_0 = -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

that is, P_0^X is a jump function.

The cost associated with the process is simply $J = x_2^2$. Our control values u_n for n = 0,1 can be selected from the space $\mathcal{C}_n$ which is composed of the two elements ± 1. Our task is to determine functions γ_n for n = 0,1 such that if $u_n = \gamma_n(I_n)$, where I_n is the control data available at time n, and which has yet to be defined, then the expected value of J is minimised.

The possible evolution of the system is depicted in the diagram shown on the next page:-

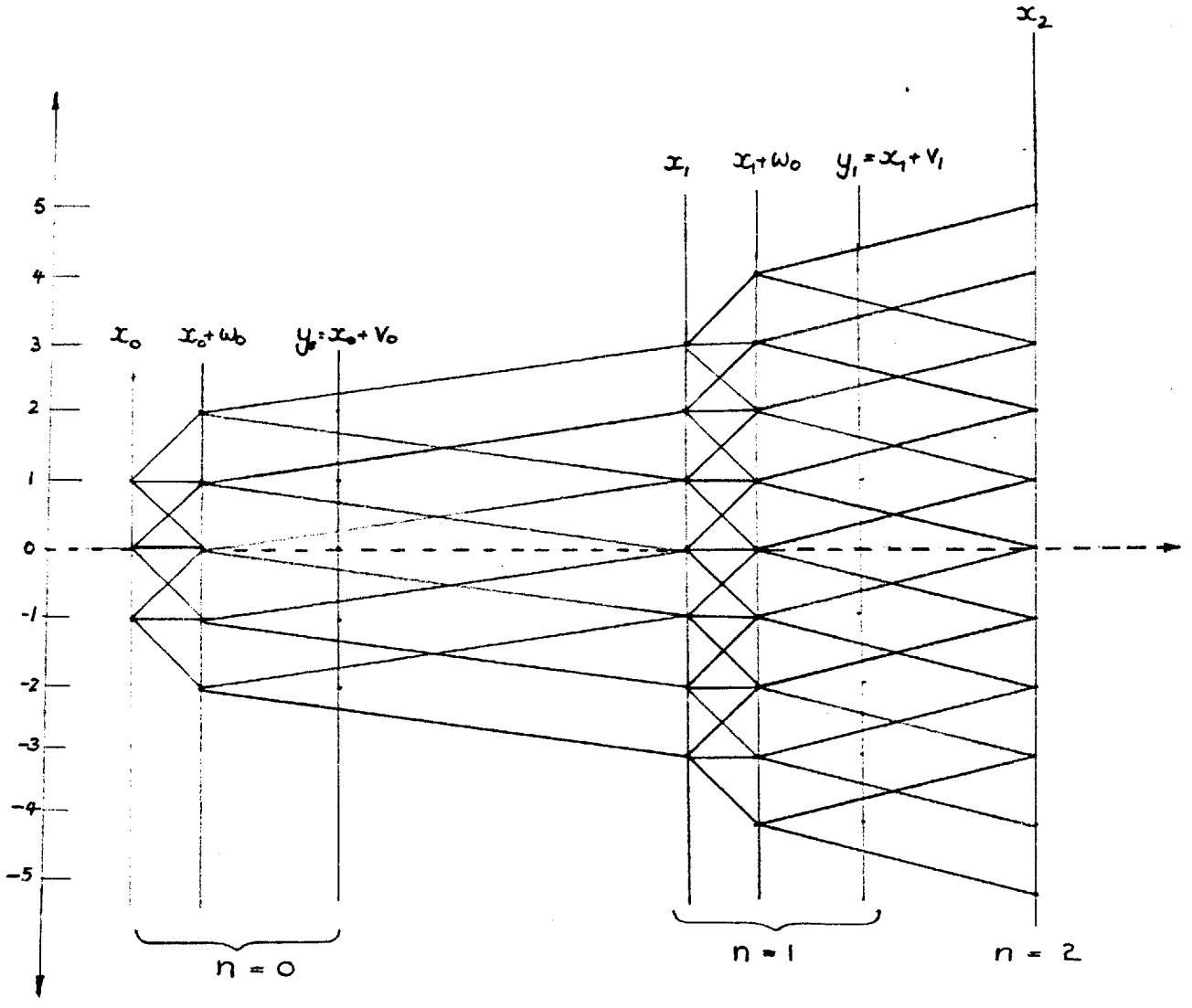


Fig. 4.1 Discrete State Example

4.3 Solution for Zero-Memory Problem

In this problem, the information available to the controller at time n is merely the current observation y_n.

In what follows, only the results obtained at various stages throughout the solution process are presented; the intricate arithmetic details have been relegated to Appendix 3.1.

Adopting the notation of Theorem 1 in Chapter 3 we have:-

$$V_1^{\circ}(p_1) = \inf_{\gamma_1} \int_{x_2}^{2} p(x_2)\,dx_2, \text{ where } p(x_2) = \mathcal{L}(p_1, \gamma_1)$$

$$= \frac{5}{3} + 3\bar{p}_1 - \frac{1}{3}\bar{p}_3 - \frac{2\ \mathrm{sgn}\,(\bar{p}_3 - \bar{p}_5)}{3} - \frac{1}{3}\bar{p}_5 + 3\bar{p}_7,$$

and $\gamma_1^{\circ}(Y_r) = -\ \mathrm{sgn}\,(\sum_i X_i\, p_{ir})$,

where $\bar{p}_i = P\{x_1 = X_i\}$, $i = 1,2, \text{---} 7$

$$p_{ir} = P\{x_1 = X_i \mid y_1 = Y_r\}, \quad i = 1,2, \text{---} 7, \ r = 1,2, \text{---} 9$$

The next step in the analysis is to substitute for $\{p_i\}$ in this expression using the state transition equation:-

$$p(x_1) = \iint p(x_1 \mid x_0, y; \gamma_0)\, p(y_0 \mid x_0)\, p(x_0)\ dx_0\ dy_0,$$

and to find V_0° and γ_0° such that

$$V_0^{\circ}(p_0) = \inf_{\gamma_0^{\circ}} V_1^{\circ}(\mathcal{L}(p_0, \gamma_0)),$$

where $p_1 = \mathcal{L}(p_0, \gamma_0)$ denotes the state transition equation above. The evaluation of $V_1^0(\mathcal{L}(p_0, \gamma_0))$ for any γ_0 is very tedious. Moreover, there are five possible values of y_0, hence there are 2^5 possible policies for γ_0 and for each of these we must in theory calculate $V_1^0(\mathcal{L}(p_0, \gamma_0))$. It turns out that not all 32 policies γ_0 have to be considered, we can restrict ourselves to those with $\gamma_0(Y_1) = -1$ and $\gamma_0(Y_5) = +1$. This means that we need only consider 2^3 possible policies (see Appendix 3.1). Assuming the initial distribution $p_1 = p_2 = p_3 = \frac{1}{3}$, we obtain:-

Observ. / Control v	Y_1	Y_2	Y_3	Y_4	Y_5	$V_1^0(\mathcal{L}(p_0, \gamma_0))$
	2	1	0	-1	-2	
$\gamma_0(Y_v)$	-1	-1	-1	-1	+1	$\frac{130}{81}$
$\gamma_0(Y_v)$	-1	+1	-1	-1	+1	$\frac{146}{81}$
$\gamma_0(Y_v)$	-1	-1	+1	-1	+1	$\frac{138}{81}$
$\gamma_0(Y_v)$	-1	+1	+1	-1	+1	$\frac{146}{81}$
$\gamma_0(Y_v)$	-1	-1	-1	+1	+1	$\frac{126}{81}^*$
$\gamma_0(Y_v)$	-1	+1	-1	+1	+1	$\frac{138}{81}$
$\gamma_0(Y_v)$	-1	-1	+1	+1	+1	$\frac{126}{81}^*$
$\gamma_0(Y_v)$	-1	+1	+1	+1	+1	$\frac{130}{81}$

The optimal costs have been marked with an asterisk. Note that the optimal cost is not unique. It does not appear to matter what we do when $y_1 = 0$ is observed. At this point we have essentially solved the zero-memory problem. All that we need to do is eliminate the dependence of γ_1° on $p(x_1)$ and the problem will be completely solved. This will be demonstrated in Section 4.5.

4.4 Solution for the Classical Information Pattern Case

For this version of the problem, the information I_n available to the controller at each time instant n consists of the current observations y_n and all previous observations and controls $(y_i, u_i \quad i = 0 \text{ --- } n-1)$

Again the tedious arithmetic details of the solution have been relegated to the Appendix (appendix 3.2). Here only results obtained at various stages throughout the solution process have been presented.

Adopting the notation developed in Chapter 3, we have

$$V_2^{\circ}(p(x_2 \mid I_2)) = \int x_2^2 p(x_2 \mid I_2)\, dx_2$$

and

$$V_1^{\circ}(p(x_1 \mid I_1)) = \inf_{\gamma_1(I_1)} \int V_2^{\circ}\, p(x_2 \mid I_2) p(\check{I}_2 \mid I_1)\, d\check{I}_2$$

where,

$$p(x_2 \mid I_2) = \frac{p(\check{I}_2 \mid x_2, I_1) \int p(x_2 \mid x_1, \gamma_1(I_1)) p(x_1 \mid I_1)\, dx_1}{\int p(\check{I}_2 \mid x_2, I_1) \int p(x_2 \mid x_1, \gamma_1(I_1)) p(x_1 \mid I_1)\, dx_1 dx_2}$$

and,

$$p(\check{I}_2|I_1) = \int p(\check{I}_2|x_2) \int p(x_2|x_1, \gamma_1(I_1)) p(x_1|I_1)\, dx_1\, dx_2$$

Proceeding as in Appendix 3.1, we obtain

$$\gamma_1^{\circ}(I_r) = -\,\mathrm{sgn}(\Sigma_j X_j\, p_{jr}),$$

and

$$V_1^{\circ}(p(x_1|I_1)) = \Sigma_j X_j^2\, p_{jr} + \frac{5}{3} - 2\,\mathrm{sgn}(\Sigma X_j\, p_{jr})\, \Sigma X_j\, p_{jr},$$

where

$$p_{jr} = p(x_1 = X_j \;\; I_1 = I_r)$$

and I_r denotes the control data at time 1; that is, an element from the set $O_2 \times O_1 \times C_1$, where $O_2 = (Y_i \mid i = 1,2,\text{---}9)$, $O_1 = (Y_j \mid j = 1,2,\text{---}5)$, and $C_1 = (u_k \mid k = 1,2)$. Notice that we do not have to be concerned at this stage with the specific elements I_r. The control law γ_1° is parameterised by $p(x_1|I_1)$ and this is an implicit function of I_1; for the moment we can forget about the details of I_1.

The next step is to compute:-

$$V_0^{\circ}(p(x_0|I_0)) = \inf_{\gamma_0(I_0)} \int V_1^{\circ}(p(x_1|I_1)) p(I_1|I_0)\, dI_1 \quad ,$$

where

$$p(x_1|I_1) = \frac{p(\check{I}_1|x_1) \int p(x_1|x_0, \gamma_0(I_0)) p(x_0|I_0)\, dx_0}{\int p(\check{I}_1|x_1) \int p(x_1|x_0, \gamma_0(I_0)) p(x_0|I_0)\, dx_0\, dx_1}$$

and

$$p(\check{I}_1|I_0) = \int p(\check{I}_1|x_0, I_0) \int p(x_1|x_0, \gamma_0(I_0)) p(x_0|I_0)\, dx_0 dx,$$

It turns out that the optimal control γ_0^o is given by:-

$$\gamma_0^o(Y_s) = -\operatorname{sgn}(p_{1s} - p_{3s}),$$

where $p_{1s} = P\{x_1 = X_1 | y_1 = Y_s\}$,

and $p_{3s} = P\{x_1 = X_3 | y_1 = Y_s\}$

with $s = 1,2, \text{---}5$.

Assuming the initial distribution $p_1 = p_2 = p_3 = \frac{1}{3}$, we obtain:-

s	1	2	3	4	5
Y_s	2	1	0	-1	-2
$p_{1s}-p_{3s}$	1	$\frac{1}{2}$	0	$-\frac{1}{2}$	-1
$\gamma_0(Y_s)$	-1	-1	± 1	+1	+1
V_0^o	$\frac{13}{9}$	$\frac{25}{18}$	$\frac{46}{27}$	$\frac{58}{18}$	$\frac{13}{9}$

After computing the probabilities of the various Y_s, we finally obtain an expected value of the optimal cost of $\frac{122}{81}$. Compare this with $\frac{126}{81}$ for the zero-memory problem.

At this point we have essentially solved the classical information pattern version of the problem. As for the zero-memory problem, all that remains to be done is to express the γ_1° explicitly in terms of the control data I_1 rather than the 'state' $p(x_1 | I_1)$. This will be done in the next section.

4.5 Comparison Between the Two Cases

We have now computed the optimal solutions for both the zero-memory and classical information pattern versions of our simple discrete state problem. As we might have expected, the extra information which the classical information pattern makes available to the controller is of value - it results in an expected cost of $\frac{122}{81}$ as opposed to $\frac{126}{81}$ for the zero memory case. But note that the optimal control laws at time n = 0 are identical for each problem. And at time n = 1, the optimal control laws are at least structurally identical in the sense that each is defined by,

$$\gamma_1(I_1) = - \operatorname{sgn} E\left[X_1 \mid I_1\right],$$

where I_1 is the information available to the controller at time n = 1. In view of this strong similarity between the optimal control laws, it is not clear where the benefits of the extra information in the classical information pattern version are arising. Obviously we need to examine more deeply the form of $\gamma_1(I_1)$ for each of our information patterns.

Optimal Control $\gamma_1(I_1)$ for Zero-Memory Problem

Let us take as our optimal control law at time n = 0 the following function:-

s	1	2	3	4	5
Y_s	2	1	0	-1	-2
$\gamma_0^o(Y_s)$	-1	-1	-1	1	1

Using this function γ_0° we can compute the state $p_1(x_1)$ at time n = 1 using our state transition equation $p_1(x_1) = \mathcal{L}(p_0(x_0), \gamma_0^{\circ})$, or alternatively,

$$p_1(x_1) = \iint p(x_1 | x_0, \gamma_0(y_0))\; p(y_0 | x_0)\; p_0(x_0)\; dx_0\; dy_0$$

Adopting our earlier notation,

$$p_k^{\gamma} = \sum_r \sum_j p_{kjr}^{\gamma}\; p_{rj}\; p_j$$

where,

$$p_k^{\gamma} = p(x_1 = X_k)$$

$$p_{kjr}^{\gamma} = p(x_1 = X_k | x_0 = X_j, y_0 = Y_r; \gamma_0^{\circ}) = p(w_0 = X_k - (X_j + \gamma_0^{\circ}(Y_r)))$$

$$p_{rj} = p(y_0 = Y_r | x_0 = X_j) = p(v_0 = Y_r - X_j)$$

$$p_j = p(x_0 = X_j)$$

Using this expression for $p_1(x_1)$, we can next compute the optimal control law γ_1° at time n = 1. This is given by:-

$$\gamma_1^{\circ}(Y_r) = -\operatorname{sgn}\left\{\sum_j X_j\; p_{jr}\right\},$$

which as we saw earlier, is of the form:-

r	1	2	3	4	5	6	7	8	9
Y_r	+4	+3	+2	+1	0	-1	-2	-3	-4
$\gamma_1^{\circ}(Y_r)$	-1	-1	-1	-1	$-\operatorname{sgn}$ $p_3 - p_5$	+1	+1	+1	+1

Thus, to determine $\gamma_1^{\circ}(Y_r)$ completely, we only need to compute - sgn $(p_3 - p_5)$.

Now $p_3^{\gamma_0^{\circ}} = \Sigma \Sigma \cdot p_{3jr}^{\gamma_0^{\circ}} \; p_{rj} \; p_j$, where,

$p_{3jr} = p(\omega_0 = 1 - (X_j + \gamma_0^{\circ}(Y_r)))$ =

		r	1	2	3	4	5
		Y_r	2	1	0	-1	-2
		$\gamma_0^{\circ}(Y_r)$	-1	-1	-1	1	1
1	1		1/3	1/3	1/3	1/3	1/3
2	0		0	0	0	1/3	1/3
3	-1		0	0	0	1/3	1/3
j	X_j						

and

$p_{5jr} = p(\omega_0 = -1 - (X_j + \gamma_0^{\circ}(Y_r)))$ =

		r	1	2	3	4	5
		Y_r	2	1	0	-1	-2
		$\gamma_0^{\circ}(Y_r)$	-1	-1	-1	1	1
1	1		1/3	1/3	1/3	0	0
2	0		1/3	1/3	1/3	0	0
3	1		1/3	1/3	1/3	1/3	1/3
j	X_j						

Also, we have $p_{rj} = P\left\{V_0 = Y_r - X_j\right\}$ as below:-

$p_{rj} = p(v_0 = Y_r - X_j)$ =

		1	2	3	j
		1	0	-1	X_j
1	2	1/3	0	0	
2	1	1/3	1/3	0	
3	0	1/3	1/3	1/3	
4	-1	0	1/3	1/3	
5	-2	0	0	1/3	
r	Y_r				

Thus $\sum_r p_{3jr}\, p_{rj} \triangleq \bar{P}_{3j}$ =

$$\begin{bmatrix} \frac{1}{3} \\ \frac{1}{9} \\ \frac{2}{9} \end{bmatrix}$$

and $\sum_r p_{5jr}\, p_{rj} \triangleq \bar{P}_{5j}$ =

$$\begin{bmatrix} \frac{1}{3} \\ \frac{2}{9} \\ \frac{1}{3} \end{bmatrix}$$

Next, we can compute p_3^{γ} and p_5^{γ} as follows:-

$$p_3^{\gamma} = \sum_j \bar{P}_{3j}\, p_j = 1/9 + 1/27 + 2/27 = 6/27$$

and

$$p_5^{\gamma} = \sum_j \bar{P}_{5j}\, p_j = 1/9 + 2/27 + 1/9 = 8/27$$

Thus, $\operatorname{sgn}(p_3^{\gamma} - p_5^{\gamma}) = -$, and $\gamma_1(Y_1 = 0) = +1$. For the complete control law at time n = 1 we have:-

r	1	2	3	4	5	6	7	8	9
Y_r	4	3	2	1	0	-1	-2	-3	-4
$\gamma_1^o(Y_r)$	-1	-1	-1	-1	+1	+1	+1	+1	+1

A useful physical verification of the probabilities $\bar{P}_{3j}$ and $\bar{P}_{5j}$ can be achieved by considering the evolution of the system under the control $\gamma_0^{\,o}$. This is depicted in the diagram on the next page. This diagram will be referred to later.

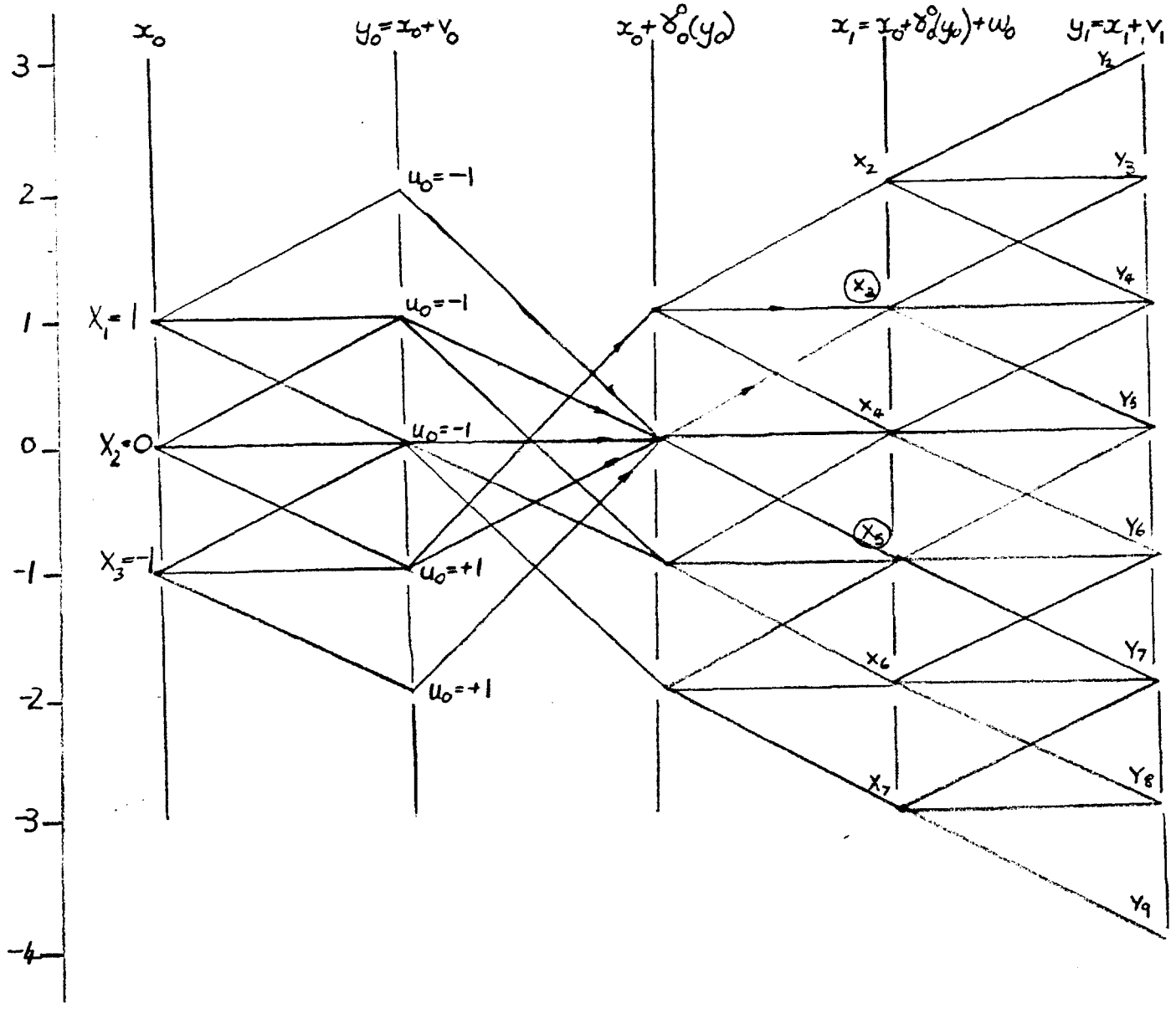


Fig. 4.2 Discrete State System - Evolution Under Control Law γ_0^0

From the above diagram it can be seen that to reach $X_3 = +1$ there are six possible routes, each with a probability of $\frac{1}{27}$, as follows:-

	w_0	v_0	x_0
i)	0	-1	0
ii)	1	1	1
iii)	1	0	1
iv)	1	-1	1
v)	1	0	-1
vi)	1	-1	-1

Thus $$p(x_1 = X_3 \mid x_0 = X_1) = \sum_k \sum_j \frac{p\{x_1 = X_3, x_0 = X_1, w_0 = W_j, v_0 = V_k\}}{p(x_0 = X_1)}$$

that is $$p(x_1 = X_3 \mid x_0 = X_1) = \frac{3}{27} \cdot \frac{3}{1} = \frac{1}{3}$$

Similarly $$p(x_1 = X_3 \mid x_0 = X_2) = \frac{1}{27} \cdot \frac{3}{1} = \frac{1}{9}$$

and $$p(X_1 = X_3 \mid x_0 = X_3) = \frac{2}{27} \cdot \frac{3}{1} = \frac{2}{9}$$

A similar procedure can be used to obtain $\bar{P}_{5j}$. In this case, to reach $X_5 = -1$, there are eight possible routes, again each with a probability of $\frac{1}{27}$, as follows:-

	w_0	v_0	x_0
i)	0	0	0
ii)	0	1	0
iii)	-1	1	1
iv)	-1	0	1
v)	-1	-1	1
vi)	-1	0	-1
vii)	-1	+1	-1
viii)	+1	+1	-1

In this case $p(x_1 = X_5 \mid x_0 = X_1) = \frac{3}{27} \cdot \frac{3}{1} = \frac{1}{3}$,

$p(x_1 = X_5 \mid x_0 = X_2) = \frac{2}{27} \cdot \frac{3}{1} = \frac{2}{9}$,

and $p(x_1 = X_5 \mid x_0 = X_3) = \frac{3}{27} \cdot \frac{3}{1} = \frac{1}{3}$

Let us next examine in detail the control $\gamma_1^{\circ}(I_1)$ for the classical information pattern problem.

Optimal Control Law $\gamma_1(I_1)$ for Classical Information Pattern Problem

As before, we have for the optimal control law at time n = 0 the following function:-

s	1	2	3	4	5
Y_s	2	1	0	-1	-2
$\gamma_0^o(Y_s)$	-1	-1	-1	+1	+1

The optimal control law at time n = 1 is given by the expression:-

$$\gamma_1^o(I_r) = -\operatorname{sgn}\left\{\sum_j X_j\, p_{jr}\right\},$$

where p_{jr} is a component of our 'state' vector at time n = 1. The state vector at time n = 0 has already been computed; it is described by,

$$p_{ps} = P\left\{X_0 = X_p \,\middle|\, y_0 = Y_s\right\} =$$

		1	2	3	4	5	s
		2	1	0	-1	-2	Y_s
1	1	1	$\frac{1}{2}$	$\frac{1}{3}$	0	0	
2	0	0	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	0	
3	-1	0	0	$\frac{1}{3}$	$\frac{1}{2}$	1	
p	X_p						

Also, our state transition equation is described by,

$$p_{jr} = \frac{p_{nmjs} \sum_{p} p_{jps}^{\gamma_0^o} p_{ps}}{\sum_{j} p_{nmjs} \sum_{p} p_{jps}^{\gamma_0^o} p_{ps}}$$

where $p_{nmjs} = P\left\{y_1 = Y_n, u_0 = U_m \mid X_1 = X_j, y_0 = Y_s\right\}$

$$p_{jps}^{\gamma_0^o} = P\left\{x_1 = X_j \mid x_0 = X_p, y_0 = Y_s, \gamma_0^o\right\}$$

$$p_{ps} = P\left\{x_0 = X_p \mid y_0 = Y_s\right\}.$$

Now $p_{nmjs} = P\left\{y_1 = Y_n, u_0 = U_m \mid x_1 = X_j, y_0 = Y_s\right\}$

$$= P\left\{y_1 = Y_n \mid X_1 = X_j\right\} \cdot P\left\{u_0 = U_m \mid y_0 = Y_s\right\}$$

$$= p_{nj}\, p_{ms}$$

Thus

$$p_{jr} = \frac{p_{nj} \sum_{p} p_{jps}^{\gamma_0^\circ} p_{ps}}{\sum_{j} p_{nj} \sum_{p} p_{jps}^{\gamma_0^\circ} p_{ps}}, \qquad \text{where } S = 1,2,3,4,5 \text{ and } n = 1,2,3,4,5,6,7,8,9$$

In theory, we must compute p_{jr} for each value of $j = 1,2,3,4,5,6,7$, and each value of s and n - this means that we must compute 225 values of p_{jr}. However all that we need is - sgn $\Sigma X_j p_{jr}^{\gamma}$; thus we only need $-\text{sgn}(\Sigma X_j p_{jr}^{\gamma})$ for $s = 1,2,3,4,5$, and $n = 1,2,3,4,5,6,7,8,9$. In other words, the control law at time n = 1 has 45 arguments in theory.

$$\gamma_1(I_r) = -\text{sgn} \sum_{j} X_j p_{rj}$$

$$= -\text{sgn}\left\{ \sum_{j} X_j p_{nj} \sum_{p} p_{jps}^{\gamma_0^\circ} p_{ps} \right\},$$

since the denominator is always positive (or zero). But it is apparent that not all the 45 arguments are realisable. For example, $y_1 = Y_1$ can never occur; $y_1 = Y_2$ and $y_0 = Y_1$ can never occur simultaneously. What is more, we can see that $x_1 = X_1$ cannot occur and neither can $x_1 = X_7$. Can these observations be exploited to simplify the algebra? Unfortunately it does not appear that there is any method that is much simpler than directly evaluating the terms involved in the above expression for

$\gamma_1(I_r)$. The details of this are presented in Appendix 3.3. Notice that above we stated that the control law at time n=1 has 45 arguments. However, if there are 5 possible observations at time n = 0, 2 possible control values at time n = 0, and 9 possible observations, then there should be 90 arguments for γ_1. What we have done is omit the dependence on the control values; hence we have a nested information pattern rather than a classical information pattern problem. We have chosen this slight re-formulation of the problem in order to simplify the presentation of the results. We obtain:-

$\gamma_1^{\circ}(I_1)$ =

		1	2	3	4	5	S
		2	1	0	-1	-2	Y_s
1	4						
2	3				-1		
3	2	-1	-1	-1	-1	-1	
4	1	-1	-1	-1	-1	-1	
5	0	$\pm$1	+1	+1	-1	$\pm$1	
6	-1	+1	+1	+1	+1	+1	
7	-2	+1	+1	+1	+1	+1	
8	-3		+1	+1			
9	-4			+1			
n	Y_n						

In this arrary, only those controls have been entered which correspond to control data which can actually occur. Let us compare this control law with that derived for the zero-memory problem. For convenience it is displayed below:-

$$\gamma_1^{\circ}(I_1) =$$

n	Y_n	
1	4	-1
2	3	-1
3	2	-1
4	1	-1
5	0	+1
6	-1	+1
7	-2	+1
8	-3	+1
9	-4	+1

It is immediate by inspection of these two control laws that the extra information provided by knowledge of the observation y_0 at time $n = 0$ is only of value when $y_1 = 0$ is observed at time $n = 1$. For all other observations y_1 at time $n = 1$, knowledge of y_0 is of no use whatsoever since the same control is applied whatever the value of y_0. This is not the case however when $y_1 = 0$ is observed. For this observation, the control value s for the classical information pattern and zero memory problems are identical for all but one case - this has been ringed in the array for the classical information pattern problem. It corresponds to control data for the classical information pattern problem of $(y_1 = 0;\ y_0 = -1)$. In this case the optimal control value is $u_1^{\circ} = -1$.

For the zero memory problem (with control data $y_1 = 0$ only) the optimal control value is $u_1^{\circ} = +1$.

The extra information which knowledge of y_0 imparts when $y_1 = 0$ is observed has a very appealing physical interpretation. Referring to Fig. 4.2, it can be seen that an observation $y_0 = -1$ can lead only to two possible 'quasi-states' $x_0 + \gamma_0^{\circ}(y_0)$ of +1 and 0 at time n = 1. The 'quasi-state' $x_0 + \gamma_0^{\circ}(y_0)$ value of 0 leads to an observation of $y_1 = 0$ implying $x_1 = +1, 0$, or -1 with equal liklihood. The 'quasi-state' $x_0 + \gamma_0^{\circ}(y_0)$ value of 1, however, leads to an observation of $y_1 = 0$ implying $x_1 = +1$ or 0 with equal liklihood. In the first instance ($x_0 + \gamma_0^{\circ}(y_0) = 0$) it dosen't matter what we choose for our control value, whereas in the second instance ($x_0 + \gamma_0^{\circ}(y_0) = 1$) there is a definite advantage in applying a control of -1. Thus for ($y_1 = 0$, $y_0 = -1$), there is a gain if $u_1 = -1$ is selected. On the other hand for $y_1 = 0$, and $y_0 \neq -1$ there is apparently a greater chance that $x_1 < 0$ than $x_1 > 0$ and the corresponding optimal control $u_1 = +1$ should be selected. Some physical interpretation should perhaps be offered regarding the indifference of the controller to observations of ± 2 at time n = 0. These observations imply that the state is perfectly observable in that the only possible values are $x_0 = \pm 1$ respectively. Applying the optimal controls as calculated for γ_0° leads to $x_0 + \gamma_0^{\circ}(y_0) = 0$ for either case. Applying the same arguments as above, it dosen't matter what we choose for our control value.

Finally let us make a remark concerning what is perhaps the most interesting feature of our problem. Our development has been such that the optimal control laws for the zero-memory and classical information pattern-problems have been derived independently of each other. However on comparing these control laws, we see that there is very little difference in the control laws. In fact, the extra information provided by a knowledge of y_0 is only of value if control data ($y_1 = 0$, $y_0 = -1$) is encountered at time $n = 1$. The possibility thus arises of only communicating the observation y_0 if $y_0 = -1$ is observed. The probability of this observation y_0 is $\frac{2}{9}$. If communication costs are expensive this gives some scope for economy. In other words, we only transmit y_0 if it is essential (that is, if it equals -1). For all other cases we can be content to let the controller carry on oblivious of what the previous observation was. This leads us to an apparently even more intractable version of the stochastic optimal control data in which we let the control data at time n be a function of the control data at times preceding n. In terms of our formulation in Chapter 3, this means that the information function β_k now become genuine functions of the arguments $\hat{I}_k$ and not merely permutation matrixes. What is more, the information functions β_k are not specified for us and we must find optimal sequences not only for $\{\gamma_k\}$ but for $\{\beta_k\}$ also. The temptation is to make the assumption that γ_k and β_k will only occur as functions of functions, that is only terms involving $\gamma_k(\beta_k(\hat{I}_k))$ will occur and we can restrict our attention to functions $\delta_k(\hat{I}_k) = \gamma_k(\beta_k(\hat{I}_k))$ which optimise the classical information pattern problem. However this assumption is only valid

for problems which do not attempt to cost the transmission of information. What is needed is a useful measure of the cost of information transmission. If such a measure can be defined, perhaps then it will be possible to introduce sufficient flexibility into the model to enable some **decoupling** between control and communication to be introduced. As we saw in Chapter 2 for the Witsenhausen problem, the conflict between control and communication is a recurring feature in stochastic optimal control when differing information patterns are considered. However, the question of costing the transmission of information will not be considered further.

4.6 Summary and Conclusions

The primary purpose of this chapter has been to provide an example of the application of the dynamic programming formulation discussed in Chapter 3. The example itself was extremely simple and yet even so the analysis was tedious and complicated. What is more, it would be very difficult to extend the analysis to determine the optimal control laws γ_0^o, γ_1^o, and γ_2^o for a three-stage process. However, our example does have the virtue that we did at least find solutions for both the zero-memory and classical information pattern versions of our problem. Moreover the example was not so trivial that these solutions were totally identical.

It was seen that the control laws at time n = 0 for both versions of the problem were identical. At time n = 1, there was however some difference between the control laws. The optimal control law for the classical information pattern problem was found to have an appealing intuitive interpretation. The extra information provided by the knowledge of the earlier observation enables the controller to identify a particular situation which is worthy of individual attention (that is, for $y_1 = 0$ and $y_0 = -1$, $u_1^o = -1$; for all other y_0, $u_1^o = +1$).

This leads to the notion that the control data itself should be some (unknown) function of the preceding control data and not merely some given function of the preceding control data. It was argued that this notion however is only of relevance if the cost of information transmission can be included in the formulation. Otherwise the problem simply reduces to the standard classical information pattern.

Regarding the simplicity of our example, two points can be made:-

(i) Our cost function did not include a component related to the control effort. Since the control can only be $\pm$ 1 and nothing else, measures such as u^2 or $|u|$ would be constant in value anyway whatever the control policy was. Thus no control term was included in the cost function.

(ii) The noise in the observation process was just sufficient to complicate the issue beyond the trivial. In other words, a particular observation y could arise from a number of different states x. But consider a n observation noise defined as follows:-

$$p\left\{\begin{array}{l} V_n = 1-\epsilon \\ V_n = 0 \\ V_n = -(1-\epsilon) \end{array}\right\} = \left[\begin{array}{c} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{array}\right] \quad \text{for any } 0 \leqslant \epsilon < 1$$

In this case a particular observation y can only arise from a particular state x and the noise is such as to guarantee perfect observation of the state x_n at each time n. The solution to the problem would then be relatively straightforward and is provided by numerous references in the literature. At $\epsilon = 0$ however, there is a monumental jump in the complexity of the problem. This observation is of significance when we attempt to apply perturbation techniques in Chapter 6.

The next chapter is a discussion of the many pitfalls which abound when the effect of the information pattern in a stochastic optimal control problem is considered. The example discussed above will provide a convenient vehicle for demonstrating how easily some of these hypotheses are motivated and in some cases it will provide a suitable counter example to refute their validity.

CHAPTER 5

Some Plausible Hypotheses Examined and Some Sub-Optimal Techniques

		Page
5.1	Introduction	5.1
5.2	The Classical Information Pattern Problem	5.2
	5.2.1 Dual Control	5.2
	5.2.2 Open Loop Optimal Feedback Control	5.4
	5.2.3 Optimal Linear Control	5.8
5.3	The Zero Memory Problem	5.12
	5.3.1 The Neo-Classical Information Pattern Problem	5.12
	5.3.1.1 The Useless Information Case	5.20
	5.3.1.2 The Valuable Information Case	5.28
	5.3.1.3 The Witsenhausen Problem	5.32
	5.3.2 The Indirect Approach	5.34
	5.3.3 Open Loop Optimal Feedback Control	5.40
	5.3.4 Optimal Linear Control	5.43
	5.3.5 Optimal Polynomial Control	5.54
5.4	Summary and Conclusions	5.56

5.1 Introduction

On a number of occasions in the previous Chapters we have suggested various approaches to the solution of the stochastic optimal control problem which we have then shown to be incorrect. The purpose of the present chapter is to discuss these misconceptions collectively, along with a number of other plausible approaches to our problem. The opportunity will also be taken to present a discussion of some sub-optimal solutions to the classical information pattern problem. Some of these sub-optimal techniques have been reported fairly recently in the literature (although their origins can be traced back to authors such as Dreyfus,[1] Feldbaum,[2] and others). The intention will be to discuss these techniques within the context of the theory which has been developed here, particularly in Chapter 3.

The major portion of the chapter will however, be devoted to reporting misconceptions which arise for a non-classical information pattern version of the problem; in particular the zero-memory problem. It must be admitted that the author has unfortunately his own contribution to make here. Much of the discussion of the zero-memory problem will be concerned with a dynamic programming formulation which looks perfectly plausible. A more critical examination of the formulation will however reveal that it is nothing more than a modified formulation of the classical information pattern problem. Luckily this formulation will find an application in Chapter 6 where a perturbation analysis will be considered.

1. Dreyfus S.E. "Some Types of Optimal Control of Stochastic Systems" 1964 (19).

2. Feldbaum A.A. "Optimal Control Systems" 1965 (20)

5.2 The Classical Information Pattern Problem

Regarding the classical information pattern problem, there is fortunately little to report concerning errors in formulation. It has been remarked elsewhere in this thesis that the stochastic optimal control problem involves the determination of a sequence of control functions and not merely control values. That authors persist in not clearly making this distinction is perhaps the most significant criticism of their work, together with the consequence that the information pattern of the problem is never defined. This observation, as we have remarked before, was first made by Striebel. In what follows a number of terms are introduced from the literature describing the classical information pattern problem. These are introduced at this stage merely as a convenience for classification and the present discussion is in no way intended to be a survey of the literature on the topic.

5.2.1 Dual Control

The Controller which we are attempting to find in our formulation of the classical information pattern problem is usually called a dual-controller. This terminology was originally introduced by Feldbaum and results from his observation that the optimal control law must not only ensure good control; it must also ensure good estimation of the unknown parameters (state) of the system. These two tasks are in general contradictory since good estimation might require large control signals while good control might require that the control signals are small.

A dual controller thus must compromise between these two tasks. We have already seen this same phenomenon in our examination of the Witsenhausen zero memory problem in Chapter 2. It is interesting to note that it also occurs in the classical information pattern problem.

The dual controller is the most general class of controller and corresponds to true stochastic optimal control. Unfortunately, for any problems apart from the linear quadratic case, the determination of the dual controller is an extremely difficult task. As we have seen in Chapter 3 it involves the solution of the following recursive equation:-

$$V_k^{o}(p(X_k \mid I_k)) = \inf_{\gamma_k(I_k)} E\left[L_k(X_k, \gamma_k(I_k)) + V_{k+1}^{o}(p(X_{k+1} \mid I_{k+1}) \mid I_k \right],$$

where $p(X_k \mid I_k)$ satisfies a suitable state transition equation. Notice we are involved at each stage k in finding an infimum with respect to a function value (parameter optimisation), and this is at least a better prospect than finding an infimum with respect to an entire function. However $p(X_k \mid I_k)$ is an unknown quantity of infinite dimension and only in special cases can we derive an analytic solution.

This difficulty has motivated the search for sub-optimal solutions in a number of directions. For example, the designer can search for linear controls. For the moment we wish to make some comments regarding one particular approach which is somewhat confusingly called optimal feedback control. Whilst the authors themselves in this area have been quite clear on the sub-optimality of the designs which this approach yields, there is however a temptation among their audience to misinterpret the results.

5.2.2 Open Loop Optimal Feedback Control

First let us discuss how the term optimal feedback control arises. Dual control is sometimes referred to as closed-loop control. Whatever its name however, it is characterised by the fact that at time k the controller is aware that at subsequent time instants k+1, k+2,--- N-1 new information will be available from observation upon which the controller can base his control action at those times. In other words, at time k the controller determines a control action U_k which depends upon his current observations (e.g. Y^k, U^{k-1}) together with his knowledge that future observations will be made and that further opportunities for control will be based on these observations. This dependence upon future events was modelled statistically in our formulation by averaging out the relevant primitive random variables, v_{k+1}, v_{k+2} and so on. This was made possible by the not unrealistic assumptions that the observation process equation G_n for n = k+1,--- N-1 is known to the Controller at time k. It is argued that this dependence upon G_n gives rise to the dual nature of the controller, making it possible to utilise the

control to probe the system and learn more about its unknowns. This is particularly relevant for so-called adaptive control systems.

Suppose, however that we postulate that the control at time k cannot be based upon knowledge of the functions G_n for n = k+1,--- N-1. Certainly in this case the control is non-dual. Because the control does depend on the current observation at time k, this type of control is sometimes called feedback control. An interesting example of this is reported by Curry.[1] In this paper the author discusses what he calls M-measurement optimal feedback control. Let us first consider what is termed open loop optimal feedback control (Sic).

Given the usual state equations, measurement equations, cost criterion functions and so on, the open loop optimal feedback control at time k is given by that u_k which minimises $J_{OL,k}$, where

$$J_{OL,k} = E\left[\sum_{j=k}^{N} L_j(X_j, u_j) + L_N(X_N) \,\middle|\, Y^k, U^{k-1}\right]$$

Notice that the function $J_{OL,k}$ depends upon the sequence of controls $(u_k, u_{k+1}, \cdots u_{N-1})$. To evaluate $J_{OL,k}$ also requires knowledge of $p(X_j \mid Y^k, U^{k-1})$ which can be fairly easily determined as an explicit function of Y^k and U^{k-1}. Consequently we shall find a $u_k(Y^k, U^{k-1})$ which minimises $J_{OL,k}$ provided that $u_{k+1}, u_{k+2}, \cdots u_{N-1}$ are also functions of (Y^k, U^{k-1}). This is the inherent weakness of the approach. The $u_{k+1}, u_{k+2}, \cdots u_{N-1}$ we have just computed are immediately discarded once u_k has been applied. At the next time instant k+1 the entire process is repeatedwith the updated information state $(Y^{k+1}\ U^k)$, controls

1. Curry " A New Algorithm for Sub-Optimal Stochastic Control" 1969 (18)

u_{k+1} u_{k+2} ---u_{N-1} are computed, control u_{k+1} is applied, and controls u_{k+2}, u_{k+3}, ---u_{N-1} are discarded. Despite the obvious comments that each u_k is sub-optimal, the approach does have the considerable advantage that it is computationally and analytically feasible, involving simple parameter optimisation at each stage. The technique does however suggest that it is a solution looking for a problem rather than the other way round. The implications of the technique would also become more salient if the author had refrained from using u_p to denote controls values and had explicitly shown their dependence upon the available information by denoting them as $\gamma_p(Y^k, U^{k-1})$.

Thus we see that open loop optimal feedback control is characterised by the fact that no future measurements are involved in the control computation at time k. Suppose that we assume in our computation that one future measurement will be available at time k+n. Then the one measurement cost function is:-

$$J_{OM,k} = E\left\{ \sum_{j=k}^{k+n-1} L_j(x_j, u_j) + E\left\{ J^o_{OL,k+n} \,\middle|\, Y^k, U^{k+n-1} \right\} \middle|\, Y^k, U^{k-1} \right\}$$

In this equation, the optimum open-loop cost function $J^o_{OL,k+n}$ is similar to the earlier equation for $J_{OL,k}$ after minimisation, but it depends on the measurements (Y^k, y_{k+n}) and the controls U^{k+n-1}. The first member of the control sequence $(u_k, u_{k+1}$---) that minimises the above expression for $J_{OM,k}$ is called the one-measurement optimal feedback control at time k.

The problem of finding the optimal u_k above is not easy and involves the solution of an n-stage two point boundary value problem. If, however, the future measurement occurs at the next time interval, then the problem reduces to parameter optimisation over u_k. The technique can, in theory be extended to include M future measurements (hence the term M-measurement optimal feedback control). If M at time k approaches N-k-1, then the optimal feed back controls we derive will approach the truly optimal closed loop control. Thus the technique does offer a compromise between the simplicity of the purely open loop optimal feedback control and the opposite extreme of the intractable closed loop control. This discussion has been presented here in order to introduce the terms feedback control and closed loop control. The reader is reminded that feedback control is sub optimal and based on the somewhat strange assumption that no future measurements will be available. This assumption appears to be motivated by a desire to reduce the problem to a form which is computationally feasible and not by any practical considerations. This same desire to reduce the problem to a form which is easy to handle is probably the source of a number of misconceptions which have appeared in the study of non-classical information patterns and which will be dealt with in the next section. In the next chapter we shall present our approach to the problem of finding useful sub optimal controls.

5.2.3 Optimal Linear Controllers

Another type of sub-optimal controller design can be obtained by restricting the class of functions from which the optimal control laws can be selected. Our formulation as developed in Chapter 3 admits any Borel function as a candidate for an optimal control law. However, suppose we restrict our search to the class of affine functions (that is, functions with a constant and proportional term such as $\gamma(x) = a+bx$). Does this assist in the determination of sub-optimal controls?

Consider an N-stage process, and let a typical control law be denoted by $\gamma_n(I_n) = a_n + \sum_{i=0}^{n} b_{ni}\, y_i + \sum_{i=0}^{n-1} c_{ni} u_i$. Then we can express our cost function in terms of $\left\{(a_n, b_{ni}, c_{nj}) \;\middle|\; n = 0,1---N-1;\ i = 0,1,---n;\ j = 0,1,---n-1\right\}$ and our problem reduces to a parameter optimisation problem, although the details of a particular problem could make it quite complicated. But at least it is a problem in a space of finite dimension. This technique can be symbolically expressed as follows:-

$$J^{o}(p_0) = \min_{\{a_n, b_{ni}, c_{nj}\}} J\left\{\{a_n, b_{ni}, c_{nj}\}, p_0\right\}$$

Notice that we have not attempted to decompose the above minimisation problem into a sequence of sub problems each involving minimisation with respect to a single triple (a_k, b_{ki}, c_{kj}). We still have a single minimisation problem involving the simultaneous minimisation of J with respect to an entire sequence of parameters $\{a_n, b_{ni}, c_{nj}\}$. This problem can be solved numerically by a variety of methods.

The idea of restricting our attention to linear control laws can be quite trivially incorporated into our dynamic programming formulation of the classical information pattern problem. In chapter 3 we derived the following iterative relationship for determining the optimal control sequence:-

$$V_k^{\circ}(p(X_k \mid I_k)) = \inf_{\gamma_k(I_k)} E\left[L_k(X_k, \gamma_k(I_k)) + V_{k+1}^{\circ}(p(X_{k+1} \mid I_{k+1})) \mid I_k \right],$$

where $p(X_k \mid I_k)$ satisfies a suitable state transition equation. But if we know $p(X_k \mid I_k)$ this is a simple parameter optimisation problem anyway, involving merely minimisation with respect to $\gamma_k(I_k)$. So what can we gain by restricting ourselves to linear control laws? The answer is that we might be able to remove the infinite dimensionality of the problem embodied in $p(X_k \mid I_k)$ - expressing $p(X_k \mid I_k)$ in terms of p_0, I_k and $\{a_n, b_{ni}, c_{nj}\}$ for n=0,1,--- k-1. In other words p_0, I_k and $\{a_n, b_{ni}, c_{nj}\}$ assume the role of the 'state' of the system. To demonstrate this procedure properly would require a careful detailed analysis similar to that presented in deriving Theorem 3.2 in Chapter 3. And it would not seem that a dynamic programming approach then offers us any advantage over the straight forward parameter optimisation approach proposed earlier.

The idea of restricting the class of control laws to linear functions has been introduced because it is sometimes used somewhat confusingly wtih dynamic programming to derive results for the LQG problem. For example, in the book "Applied Optimal Control" Ho[1] presents very brief proof of the Certainty Equivalence Principle. The argument used to justify the approach runs as follows (private discussion with Ho):-

1. Ho and Bryson "Applied Optimal Control" 1969 (26)

Suppose $p(X_{N-1} \mid I_{N-1})$ is Gaussian, then the optimal control γ_{N-1} is linear and $p(X_N \mid I_N)$ is also Gaussian. By induction, therefore if $p(X_0 \mid I_0)$ is Gaussian, γ_0 is linear and $p(X_1 \mid I_1)$ is Gaussian.

Let us look at this argument closely. First, if $p(X_n \mid I_n)$ is Gaussian, then $p(X_{n+1} \mid I_{n+1})$ is Gaussian, whatever the structure of the control law γ_n. We utilised this fact in Chapter 3. The linearity of γ_n does not matter. Next the induction argument needs to be examined. We must be careful. The analysis for γ_{N-1} is essentially that for a single stage process. The next step in the induction argument is really to prove that if the result holds for an N-n+1 stage process it also can be applied to an extra time interval at the <u>end</u> of the process. The extra time interval must be added on at the <u>beginning</u> to satisfy the dynamic programming concept. This is usually where the difficulty lies, as witnessed by the considerable jump in complexity from a 1 stage process to a 2 stage process. The difficulty however is far less significant if we first assume that $p(X_n \mid I_n)$ is Gaussian. Finding γ_n is then easy (and it is linear) and we can then append our results for the N+n-1 stage process quite conveniently. Such induction arguments should be treated with caution for the LQG case. They can be used to great advantage to simplify the algebraic complexity of the problem. But it must be remembered that they rely on the fact that $p(X_n \mid I_n)$ is Gaussian whatever the form of γ_{n-1}.

What Ho has proved is a result which is valid among the class of linear controllers. It does not show that his control is optimal over the class of all control laws. Ho has in fact published a technique -

which shows that for the LQG the optimal control is linear. He does this by reducing the N stage LQG problem to a single stage problem. Linearity then follows easily. This approach, which is discussed in Chapter 9, provides a very elegant analysis of the LQG problem.

One final comment should be made concerning linearity. We have shown that for the LQG problem $p(X_n \mid I_n)$ is Gaussian whatever the form of the control law. It is not true however that $p(X_n)$ is also Gaussian whatever the form of the control law. For $p(X_n)$ to be Gaussian the control law must be linear. Thus linearity arguments can also be associated with an ambiguous definition of the state. For the LQG problem it turns out of course that $p(X_n)$ is Gaussian but this is coincidental.

In the next section we shall consider some misconceptions associated with the zero-memory problem.

5.3 The Zero-Memory Problem

On several occasions in the preceding chapters we have suggested approaches to the solution of the zero memory problem which we have then abandoned because we have shown they were inappropriate. The purpose of this section is to catalogue these approaches collectively, together with some further examples of what appear to be fairly popular misconceptions.

The first case we shall consider involves an attempt to formulate the zero-memory problem in the framework of dynamic programming. This particular formulation, which is the author's own contribution, was motivated by a desire to obtain a sequence of sub problems, each one involving minimisation with respect to a function value and not an entire function. It will be seen that the formulation which is obtained is in fact an alternative formulation of the nested information (or equivalently the classical information) pattern problem. As such it is believed to be of value in its own right and has been awarded the somewhat pompous title of the Neo-Classical approach.

5.3.1 The Neo-Classical Information Pattern Problem

To simplify the algebra in what follows, we shall assume that the cost function is simply a function of the final state - $\phi(X_N)$. As far as the algebra is concerned, there is no loss of generality. This does not imply of course that results obtained for cost functions of the form $\phi(X_N)$ can be applied in more general situations - they certainly cannot. All that we are doing is simplifying the formulation.

In Chapter 3 various candidates were examined carefully to see which would qualify as state variables in a dynamic programming formulation. Our tests were tedious but at least reliable. Let us see what happens if we are a little more careless in selecting a state variable.

Suppose we propose (y_n, p_n) as a state variable for the zero memory problem. Here y_n is the observation and p_n is the probabiltiy density function of the state x_n, both at time n.

Let us first derive an iterative relationship involving the expected cost at time n+1. We have:-

$$E[\phi(x_N) | y_{n+1}, p_{n+1}]$$

$$= \int \phi(x_N) \text{ ------ } \int p(x_{n+2} | x_{n+1}, \gamma_{n+1}(y_{n+1})) p(x_{n+1} | y_{n+1}) \, dx_{n+1} \cdots dx_N$$

$$= \frac{\int \phi(x_N) \text{ ----- } \int p(x_{n+2} | x_{n+1}, \gamma_{n+1}(y_{n+1})) p(y_{n+1} | x_{n+1}) p_{n+1}(x_{n+1}) \, dx_{n+1} \cdots dx_N}{\int p(y_{n+1} | x_{n+1}) p_{n+1}(x_{n+1}) \, dx_{n+1}}$$

$$\triangleq V_{n+1}(y_{n+1}, \Gamma_{n+1}, p_{n+1}), \text{ where } \Gamma_{n+1} \text{ denotes functions } \gamma_{n+1}, \gamma_{n+2}, \cdots \gamma_{N-1}$$

Similarly consider,

$$E[\phi(x_N)|y_n, p_n]$$

$$= \frac{\int \phi(x_N) \cdots \int p(x_{n+2}|x_{n+1}, \gamma_{n+1}(y_{n+1})) p(y_{n+1}|x_{n+1}) \int p(x_{n+1}|x_n, \gamma_n(y_n))}{\int p(y_n|x_n) p_n(x_n)\, dx_n} \cdot p(y_n|x_n) p_n(x_n)\, dx_n \cdots dx_N$$

$\triangleq V_n(y_n, \Gamma_n, p_n)$, where Γ_n denotes functions $\gamma_n, \gamma_{n+1}, \cdots \gamma_{N-1}$.

We now propose to establish an iterative relationship between V_n and V_{n+1}, with y_n, p_n as the state variable. Consider:-

$$\int V_{n+1}(y_{n+1}, \Gamma_{n+1}, p_{n+1})\, p(y_{n+1}|y_n)\, dy_{n+1}$$

$$= \int \left[\frac{\int \phi(x_N) \cdots \int p(x_{n+2}|x_{n+1}, \gamma_{n+1}(y_{n+1})) p(y_{n+1}|x_{n+1}) p_{n+1}(x_{n+1})\, dx_{n+1}}{\int p(y_{n+1}|x_{n+1}) p_{n+1}(x_{n+1})\, dx_{n+1}} \right] \cdot p(y_{n+1}|y_n)\, dy_{n+1}$$

Now

$$p(y_{n+1}|y_n) = \frac{\int p(y_{n+1}|x_{n+1}) \int p(x_{n+1}|x_n, \gamma_n(y_n))\, p(y_n|x_n) p_n(x_n)\, dx_n\, dx_{n+1}}{\int p(y_n|x_n) p_n(x_n)\, dx_n}$$

and we shall make the substitution:-

$$p(x_{n+1}) = \frac{\int p(x_{n+1} \mid x_n, \gamma_n(y_n))\, p(y_n \mid x_n)\, p_n(x_n)\, dx_n}{\int p(y_n \mid x_n)\, p_n(x_n)\, dx_n}$$

Thus, after some fortuitous cancelling out, we obtain:-

$$\int V_{n+1}(y_{n+1}, \Gamma_{n+1}, p_{n+1})\, p(y_{n+1} \mid y_n)\, dy_{n+1}$$

$$= \int \left[\frac{\int \phi(x_N) \cdots \int p(x_{n+2} \mid x_{n+1}, \gamma_{n+1}(y_{n+1}))\, p(y_{n+1} \mid x_{n+1}) \int p(x_{n+1} \mid x_n, \gamma_n(y_n))}{\int p(y_n \mid x_n)\, p_n(x_n)\, dx_n} \right.$$

$$\left. \cdot\, p(y_n \mid x_n)\, p_n(x_n)\, dx_n\, dx_{n+1} \cdots dx_N \right] dy_{n+1}$$

$$= V_n(y_n, \Gamma_n, p_n)$$

Thus, for a general cost function of the form $\sum_{n=0}^{N-1} L_n(X_n, U_n) + L_N(x_N)$, we have derived the following iterativerelationships involving the conditional costs to go:-

$$V_n(y_n, p_n, \Gamma_n) = E\left[L_n(x_n, \gamma_n(y_n)) + V_{n+1}(y_{n+1}, p_{n+1}, \Gamma_{n+1}) \mid y_n \right]$$

where, $p_{n+1}(x_{n+1}) = \dfrac{\int p(x_{n+1} \mid x_n, \gamma_n(y_n))\, p(y_n \mid x_n)\, p_n(x_n)\, dx_n}{\int p(y_n \mid x_n)\, p_n(x_n)\, dx_n}$

Equipped with this, we can now utilise dynamic programming arguments to determine a useful necessary condition for optimality. This is stated below as Theorem 5.1:-

Theorem 5.1

Suppose $V^{o}_{n+1}(y_{n+1}, p_{n+1}) \triangleq \inf_{\Gamma_{n+1}} V_{n+1}(y_{n+1}, p_{n+1}, \Gamma_{n+1})$.

Then $V^{o}_{n}(y_n, p_n) = \inf_{u_n} E\left[L_n(x_n, u_n) + V^{o}_{n+1}(y_{n+1}, p_{n+1}) \mid y_n\right]$,

where

$$p(x_{n+1}) = \frac{\int p(x_{n+1} \mid x_n, u_n)\, p(y_n \mid x_n)\, p_n(x_n)\, dx_n}{\int p(y_n \mid x_n)\, p_n(x_n)\, dx_n}$$

Proof

By definition

$$V^{o}_{n}(y_n, p_n) = \inf_{\gamma_n(y_n), \Gamma_{n+1}} E\left[L_n(x_n, \gamma_n(y_n)) + V_{n+1}(y_{n+1}, p_{n+1}, \Gamma_{n+1}) \mid y_n\right]$$

$$= \inf_{u_n, \Gamma_{n+1}} E\left[L_n(x_n, u_n) + V_{n+1}(y_{n+1}, \mathcal{L}_n(p_n, y_n, u_n), \Gamma_{n+1}) \mid y_n\right]$$

$$= \inf_{u_n} \Big\{ E\big[L_n(x_n, u_n) \,|\, y_n\big) + \int \Big\{ \inf_{\Gamma_{n+1}} V_{n+1}\big(y_{n+1}, \mathcal{I}_n(p_n, y_n, u_n), \Gamma_{n+1}\big) \Big\} p(y_{n+1}|y_n)\, dy_{n+1} \Big\}$$

$$= \inf_{u_n} \Big\{ E\big[L_n(x_n, u_n) + V^{\circ}_{n+1}\big(y_{n+1}, \mathcal{I}_{n+1}(p_n, y_n, u_n), \Gamma_{n+1}\big) \,|\, y_n\big] \Big\}$$

In the above proof we have omitted the rather obvious justification for commuting the infimum and expectation operations. All this looks very promising for the zero-memory problem. We have decomposed our original problem into a sequence of sub problems, as usual. But each sub-problem involves minimisation with respect to a function value $\gamma_n(y_n)$ and not the entire function γ_n. Compare this with Theorem 3.1. We appear to have derived something of value. Imitating the procedures we have adopted before, we can determine each optimal control value u_n in terms of y_n and p_n. Hopefully, we can work back from time N-1 to time 0, and obtain a sequence of control laws:-

$$\gamma^{\circ}_{N-1}(y_{N-1}, p_{N-1}(\ \ ))$$

$$\gamma^{\circ}_{N-2}(y_{N-2}, p_{N-2}(\ \ ))$$

$$\gamma_1^{o}(y_1, p_1(\ ))$$

$$\gamma_0^{o}(y_0, p_0(\ ))$$

Having obtained γ_0^{o} in this way, we then have the problem of eliminating the dependence γ_n^{o} upon p_n. This can be achieved by means of a 'sweep forward' using the state transition equation.

The whole technique sounds fine. Unfortunately, it is wrong and, as we said earlier, it is merely an alternative formulation of the nested information pattern problem. This can be seen by looking a little more closely at our state transition equation:-

$$p_{n+1}(x_{n+1}) = \frac{\int p(x_{n+1} \mid x_n, \gamma_n(y_n)) p(y_n \mid x_n) p_n(x_n)\, dx_n}{\int p(y_n \mid x_n) p_n(x_n)\, dx_n}$$

$$\triangleq \mathcal{L}_n(p_n, y_n, \gamma_n(y_n))$$

When we determine our optimal control value u_n, we obtain u_n as a function y_n and p_n, that is:-

$$u_n^{o} = \gamma_n^{o}(y_n, p_n)$$

But p_n is determined by our state transition equation,

$$p_n = \mathcal{L}_{n-1}(p_{n-1}, y_{n-1}, \gamma_n(y_{n-1})),$$

thus we have,

$$u_n^{o} = \bar{\gamma}_n^{o}(y_n, y_{n-1}, \gamma_{n-1}(y_{n-1}), p_{n-1})$$

Clearly, by recursive application of the state transition equation, we ultimately obtain:-

$$u_n^{\circ} = \hat{\gamma}_n^{\circ}(y_n, y_{n-1} \cdots y_0; \gamma_n(y_{n-1}); \gamma_{n-2}(y_{n-2}), -\gamma_0(y_0); p_0)$$

In other words, u_n° is the nested information (or classical information) pattern optimal control value. What we failed to appreciate in our earlier analysis was that p_n is an implicit function of y_{n-1} and p_{n-1}; p_{n-1} is an implicit function of y_{n-2} and p_{n-2}; and so on. Thus u_n is a function of y_n and, implicitly, $y_{n-1}, y_{n-2}, \cdots y_0$.

It might be suspected that even if we had failed to spot our error initially, it would have soon become apparent when we attempt to eliminate p_n by what we called a 'sweep forward using the state transition equation'. We then would have obtained u_n as a function of $y_n, y_{n-1}, \cdots y_0$, and at least suspicion should have been aroused. Not so.

Perhaps anticipating the difficulty subconciously, the trick is to invent a new state transition equation:-

$$p_{n+1}(x_{n+1}) = \iint p(x_{n+1} \mid x_n, \gamma_n(y_n)) p(y_n \mid x_n) p_n(x_n) \, dx_n \, dy_n$$

This is the state transition equation we introduced in Chapter 3 when the zero-memory problem was discussed. Quite naturally, this works ideally. We now have:-

$$p_{n+1} = M_n(p_n, \gamma_n),$$

and we can use this equation to eliminate p_n from our control laws. We thus obtain functions which depend upon the current observation alone, as required:-

$$u_0 = \gamma_0^{\circ}(y_0, p_0) \qquad = \overset{V}{\gamma}_0^{\circ}(y_0)$$

$$u_1 = \gamma_1^{\circ}(y_1, M(p_0, \overset{V}{\gamma}_0^{\circ})) \qquad = \overset{V}{\gamma}_1^{\circ}(y_1)$$

$$u_{N-1} = \gamma_{N-1}^{\circ}(y_{N-1}, M_{n-2}(p_{n-2}, \overset{V}{\gamma}_{n-2})) \qquad = \overset{V}{\gamma}_{N-1}(y_{N-1})$$

But let us reflect what in fact we have done. First we have found the optimal nested information pattern solution, in terms of the current observations and the (implicit) conditional probability density function of the state. Next, we have simply averaged out those components of the information pattern which are unknown to the controller at time n.

Viewed in this light, we see that the technique is nothing more than a sub-optimal approach to the zero memory problem. It is interesting to note however that the technique does yield truly optimal solutions in a surprising number of examples. It is this fact which has lead some workers to suggest that the technique itself is correct since it produces the right answers, sometimes. This assertion will be demonstrated briefly by three examples:-

5.3.1.1 The Useless Information Case

The following example was suggested by Y.C. Ho. It is a particularly degenerate example and demonstrates several pathological properties:-

(i) the information contained in the observations is of no value

(ii) the classical information pattern solution can be used to obtain the zero-memory solution

The evolution of the states of the system is depicted in the diagram below:-

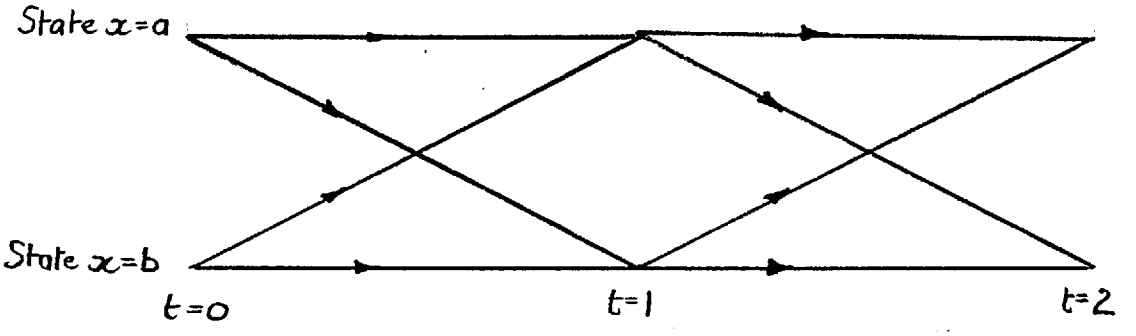


Fig. 5.1 Useless Information Case

Here the state can reside in either state 'a' or state 'b'. Two controls designated U (up) and D (down) are available to influence the probability that the system occupies state i at time n, given that it occupied state j at time n-1. These transition probabilities are defined below for t = 1:-

$p(x_2 \mid x_1;\ u = U)$ =

U	a	b
a	$\frac{1}{2}$	$\frac{1}{2}$
b	$\frac{3}{4}$	$\frac{1}{4}$

X_2 (columns) / X_1 (rows)

$p(x_2 \mid x_1;\ u = D)$ =

D	a	b
a	$\frac{1}{2}$	$\frac{1}{2}$
b	$\frac{1}{4}$	$\frac{3}{4}$

x_2 (columns) / X_1 (rows)

From these transition probabilities we see that the control is only effective in state 'b' at time t=1. Similarly, the transition probabilities at t=0 are defined as follows:-

$p(x_1 \mid x_0;\ u = U)$ =

U	a	b
a	$\frac{3}{4}$	$\frac{1}{4}$
b	$\frac{1}{2}$	$\frac{1}{2}$

x_1 (columns)

$p(x_1 \mid x_0;\ u = D)$ =

D	a	b	x_1
a	$\frac{1}{4}$	$\frac{3}{4}$	
b	$\frac{1}{2}$	$\frac{1}{2}$	
x_0			

At time t=0, control is only effective if the state is 'a'. The observation η of the state x at any time instant i is defined below,

$p(\eta_i \mid x_i)$ =

	A	B	η_i
a	$\frac{3}{4}$	$\frac{1}{4}$	
b	$\frac{1}{2}$	$\frac{1}{2}$	
x_i			

The initial distribution x_0 is defined as:-

$$p_0 = P \begin{bmatrix} x_0 = a \\ x_0 = b \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

Finally, a cost of 0 is incurred when the final state x_2 is 'a', and a cost of 1 is incurred when x_2 is b. The problem is to determine the zero memory control laws γ_0° and γ_1° such that the expected cost is minimised.

The problem can be approached by either the Neo Classical Method suggested by Theorem 5.1 or by the direct numeration of control laws (function minimisation) suggested by Theorem 3.1. Both these approaches will be considered below, although the intricate algebra has been omitted:-

The Neo-Classical Method

We define the unknown probability density function of the state x, at time t=1 as follows:-

$$p_1 = P \begin{bmatrix} x_1 = a \\ x_1 = b \end{bmatrix} = \begin{bmatrix} p \\ 1-p \end{bmatrix}$$

Applying Theorem 5.1 we obtain the following optimal control law at time t=1:-

$$\gamma_1^{\circ}(A) = U$$

$$\gamma_1^{\circ}(B) = U,$$

with associated optimal cost to go given by:-

$$V_1^{\circ}(\eta_1, p) = \frac{1 + 2p}{2(2+2p)} \qquad \text{for } \eta_1 = A$$

$$= \frac{1}{2(2-p)} \qquad \text{for } \eta_1 = B$$

Again, applying Theorem 5.1 we can next determine the optimal control law at time t=0:-

$$\gamma_0^{\circ}(A) = D$$

$$\gamma_0^{\circ}(B) = D,$$

with associated optimal cost to go given by:-

$$V_0^{\circ}(\eta_0, p_0) = .343 \quad \text{for } \eta_p = A$$

$$= .346 \quad \text{for } \eta_p = B$$

Thus we have solved the problem using our so-called Neo Classical approach. Next let us examine direct enumeration of the possible control laws.

Direct Enumeration Method

At time t=1, there are four possible control laws γ_1. By evaluating the cost associated with each control law (in terms of p) we obtain the optimal control. As with the example in Chapter 4 and in fact for all single stage decisions, the determination of γ_1° can be reduced conveniently to evaluation of $\gamma_1^{\circ}(\eta_1)$ instead of the entire γ_1°.

Again we obtain,

$$\gamma_1^{\circ}(A) = U$$

$$\gamma_1^{\circ}(B) = U$$

with optimal costs to go given by:-

$$V_1^{\circ}(\eta_1, p) = \frac{1 + 2p}{2(2+p)} \quad \text{for } \eta_1 = A$$

$$= \frac{1}{2(2-p)} \quad \text{for } \eta_1 = B$$

The next step is to determine $E\left[V_1^{\circ}(\eta_1, p)\right] = J_1^{\circ}(p)$, and then express p in terms of each of the four candidate control functions at time t = 0. We find that the optimal cost is given by the following control law at t = 0:-

$$\gamma_0^{\circ}(A) = D$$

$$\gamma_0^{\circ}(B) = D,$$

together with an associated cost J_0° for this control policy of:-

$$J_0^{\circ}(p_0) = 0.3445$$

This concludes our discussion of the direct enumeration method for these examples.

Comparison of the two-methods

The expected cost J_0^o obtained using the pseudo-classical method is as follows:-

$$\bar{J}_0^o(p_0) \triangleq V_0^o(A,p_0)p(\eta_0 = A) + V_0^o(B,p_0)p(\eta_0 = B)$$

$$= 0.3445 = J_0^o(p_0)$$

Thus we see that the two methods produce the same optimal cost and the temptation is to conclude that the two methods are in some sense equivalent. Unfortunately, they are only equivalent for this particular example as is demonstrated by a brief appraisal of the solution. The optimal control law at time t=1 for the neo-classical problem is independent of p. The implication of this is that the control law is independent of the past observations and controls. Thus the zero memory solution is as 'good' as the classical information pattern solution. (In fact, since the control law is independent of the current observation as well, the conclusion can be drawn that the open-loop solution is also as 'good' as the classical information pattern solution.)

The fact that the zero memory solution is as good as the classical information pattern solution is expressed in the equality of the two optimal costs $\bar{J}_0^o$ and J_0^o. Clearly the extra information provided by the classical information pattern formulation is of no value - hence the name the Useless Information Problem. In fact even the current observations are useless, let alone the previous observations since the control is always independent of the observations.

It may happen, however, that even when the extra information provided by the classical information pattern is of value, the neo-classical approach will still yield optimal control policies for the zero memory problem. This is demonstrated by the next example, which was first introduced in Chapter 4. It was seen there that the extra information of the classical information pattern formulation was of value, resulting in a lower cost for this formulation as opposed to the zero-memory formulation - hence the name the Valuable Information Case.

5.3.1.2 The Valuable Information Case

To demonstrate the use of the neo-classical approach, this example will be worked in full. For ease of reference, the evolution of the system to be studied is illustrated in the diagram below (this is identical to Fig. 4.1):-

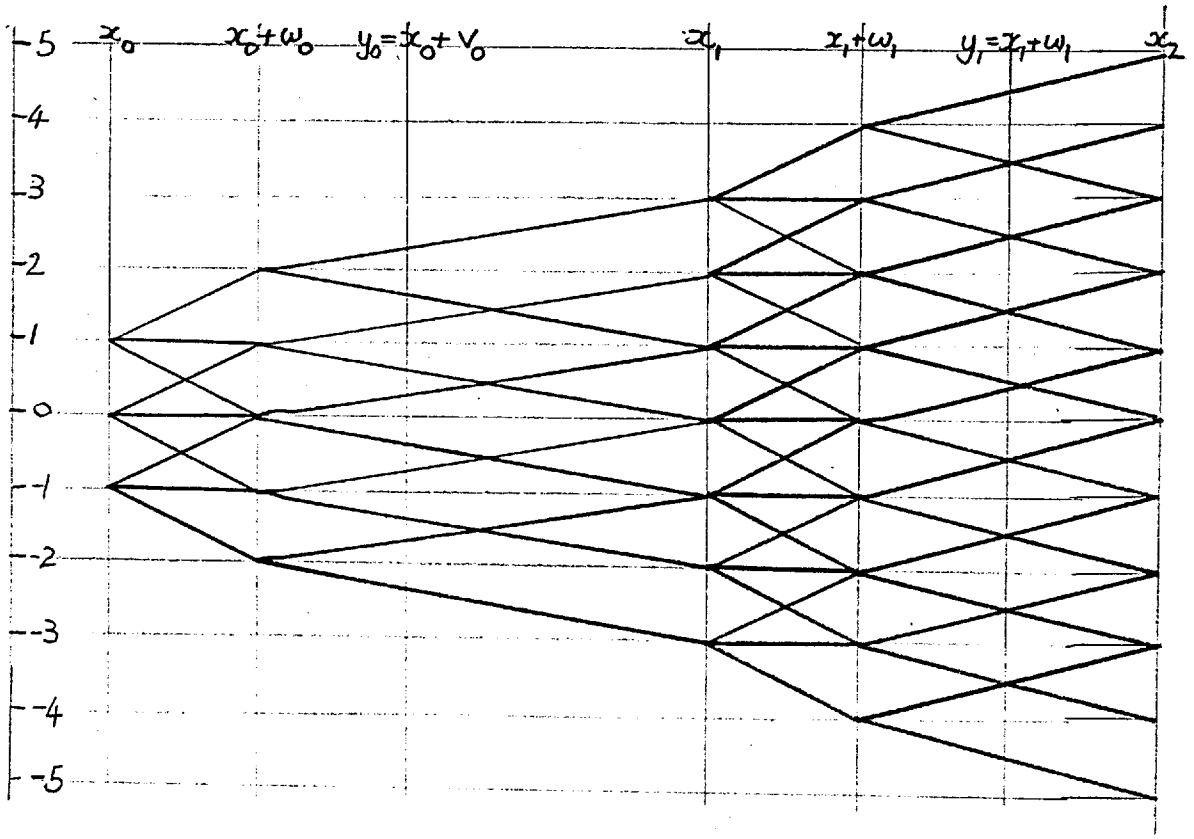


Fig. 5.2 Valuable Information Case

The details of the application of Theorem 5.1 are presented in Appendix 3.4, where it is shown that the optimal cost for the problem is $\frac{122}{81}$ (recall that for the zero memory problem, the expected optimal cost $\frac{126}{81}$, obtained by direct enumeration in Chapter 4).

Thus our example refutes the assertion that the neo-classical approach yields an optimal solution for the zero memory problem. Nevertheless, an optimist might still persist in proposing the technique, since it can be used (in certain cases) to obtain the optimal zero memory control laws.

This is demonstrated as follows:-

Optimal Control at t=0

Y_n	2	1	0	-1	-2
$\gamma_0^\circ(Y_n)$	-1	-1	± 1	+1	+1

Optimal Control at t=0

Y_r	4	3	2	1	0	-1	-2	-3	-4
$\gamma_1^\circ(Y_r)$	-1	-1	-1	-1	$-\text{sgn}(p_3-p_5)$	+1	+1	+1	+1

Thus to compute our optimal control laws, we need to find p_3 and p_5. These are calculated via the 'zero memory state transition equation' as follows:-

$$p(x_1) = \iint p(x_1 \mid x_0, \gamma_0^{\circ}(y_0)) p(y_0 \mid x_0) p_0(x_0)\, dx_0 dy_0,$$

that is:-

$$p_k = \sum_{nj}\sum p_{kjn}^{\gamma_0^{\circ}} p_{nj} p_j,$$

where γ_0° is as specified above with $\gamma_0^{\circ}(0) = -1$. Under this control law, the appropriate values for p_3 and p_5 were evaluated in Chapter 4. They are as follows:-

$$p_3 = \frac{6}{27}$$

$$p_5 = \frac{8}{27}$$

Thus for the 'optimal' control law at time t=1 for the zero memory problem, using the neo-classical approach, we obtain:-

Y_r	4	3	2	1	0	-1	-2	-3	-4
$\gamma_1^{\circ}(Y_r)$	-1	-1	-1	+1	+1	+1	+1	+1	+1

By comparison with Chapter 4, we see that this is indeed the optimal control law for the zero-memory problem. This result therefore tends to support the view of our optimist. Unfortunately, we have simply been

lucky in the choice of our example. The application of the neo-classical technique, followed by a sweep forward using the zero memory state transition equation is a sub-optimal procedure. It so happens that the sub-optimal solution for this example is identical to the truly optimal solution. Below, by means of the Witsenhausen example, we demonstrate that the neo-classical technique does not necessarily give optimal control laws for the zero memory problem.

However, first let us compare our neo-classical solution technique with that employed for the same problem in Chapter 4 using Theorem 3.2. In total, to find the optimal control laws by either technique seems to require the same amount of work - which is surprisingly tedious in both cases. There is however a significant shift of where in the procedure this work is performed. In the approach of Theorem 3.2, the backward induction for time $t = 2$ to time $t = 0$ is executed fairly easily. Unfortunately when we have completed this backward induction, we know very little about the structure of the control laws since they are completely parameterised by a conditional probability density function of the state, about which at this stage we know very little. And to eliminate the conditional probability density function, sweeping forward using our 'state' transition equation involves considerable effort.

Conversely, the backward induction process using the neo-classical approach is more involved than the corresponding stage when Theorem 3.2 is used. It does however have the significant advantage, perhaps, that the control laws are functions of the current observation and a conditional probability density function of the state (this time conditioned

on all observations except the current one). It is this extra argument in the control law which makes the backward induction more tedious. However, since the conditional probability density function of the state is conditioned not on the current observation but on all previous ones, subsequent elimination of this density function is more straightforward. The fact that backward induction in the neo-classical approach gives a control law which is an explicit function of the current observation is utilised in the next Chapter to derive a sub-optimal procedure for the classical information pattern problem.

5.3.1.3 The Witsenhausen Problem

In order to totally refute the neo-classical approach as an optimal technique for the solution of the zero memory problem, we return to the celebrated Witsenhausen example which was first introduced in Chapter 1.

For reference, a brief statement of the problem is presented here:-

State Equations: $x_1 = x_0 + u_0$

$x_2 = x_1 - u_1$

Output Equations: $y_0 = x_0$

$y_1 = x_1 + v$

Cost Function: $k^2(u_1^2) + (x_2)^2, k^2 > 0$

Admissable Controllers: $u_0 = \gamma_0(y_0)$

$u_1 = \gamma_1(y_1)$

The random variable v is Gaussian with zero mean and variance σ_v^2. The random variable x_0 is Gaussian with zero mean and variance σ_0^2.

The details of the application of Theorem 5.1 to this problem are presented in Appendix 3.5, where it is shown that the optimal control laws are:-

$$\gamma_0^0(y_0) = 0$$

and $$\gamma_1^0(y_1) = -y_1 \sigma_0^2/(\sigma_0^2 + \sigma_v^2)$$

The optimal control laws we have derived at time t = 0 and t = 1 are both linear control laws. But Witsenhausen deomstrates in his paper that for this problem, the truly optimal control law at time t = 0 is non-linear. Thus our neo-classical approach does not generally yield optimal control laws for the zero memory problem. If it does, it is a fortuitous accident. As we have said before, the neo-classical approach is just a slight reformulation of the classical information pattern problem. This is confirmed here by the observation that the control laws we have derived here are identical to those obtained in Chapter 2 when the classical information pattern version of the problem was studied.

This completes for the moment our discussion of the neo-classical approach. It will be re-introduced in the next Chapter as the basis for a perturbation approach to the problem of finding sub-optimal controls for the classical information pattern problem.

5.3.2 The Indirect Approach

In Chapter 2, two possible approaches to the solution of a simple zero memory stochastic optimal control problem were discussed. These two approaches were termed the direct approach and the indirect approach. In this section we shall examine the indirect approach within the context of our dynamic programming formulation of the problem.

Briefly, in Chapter 2, we considered a two stage system with cost function $x_2^2 + k^2 u_0^2$. The direct approach posed the following problem for the determination of the optimal control law at time t = 0:-

Find that γ_0° such that $E\left[x_2^2 \mid y_0\right] + k^2 u_0^2$ is minimised, given that at time t = 1, the controller acts optimally. On the other hand, with the indirect approach, the problem posed for the determination of the optimal control law at time t = 0 was as follows:-

Find that γ_0° such that $E\left[E\left[x_2^2 \mid y_2\right] \mid y_1\right] + k^2 u_0^2$ is minimised, again given that at time t = 1, the controller acts optimally.

In the previous section our so-called neo-classical approach to the solution of the zero memory problem was considered in detail. It was seen that the technique floundered owing to the wrong selection of

the state transition equation - the state being selected was one which implicitly depended on past observations. Because of this the technique was nothing more than a reformulation of the classical information pattern problem.

In this section we shall take as our state variable the same two quantities y_k and p_k as were adopted for the neo-classical approach. But now we shall select the correct state transition equation for the zero memory problem. Even so, we shall show that the technique is wrong because, although it provides optimal zero memory control laws, they are optimal for the wrong problem. Moreover, solution still demands minimisation with respect to an entire function.

Consider the cost-to-go from time t = k conditioned on observation y_k, defined as $C_k(p_k, y_k, \Gamma_k^{N-1})$, where:-

$$C_k(p_k, y_k, \Gamma_k^{N-1}) = E\left[L_k(x_k, \gamma_k(y_k)) + L_N(x_N) \,\middle|\, y_k \right],$$

and as usual p_k is our state variable - the probability density function of the state x_k at time t = k - together with the observation y_k. Our state transition equation is defined by:-

$$p_{k+1}(x_{k+1}) = \iint p(x_{k+1} \mid x_k, \gamma_k(y_k)) p(y_k \mid x_k) p_k(x_k)\, dy_k\, dx_k,$$

that is,

$$p_{k+1}(x_{k+1}) = \mathcal{L}_k(p_k, \gamma_k)$$

First let us demonstrate that :-

$$C_k(p_k, y_k, \Gamma_k^{N-1}) \neq E\left[L_k(x_k, \gamma_k(y_k)) + C_{k+1}(\mathcal{L}_k(p_k, \gamma_k), y_{k+1}, \Gamma_k^{N-1}) \,\middle|\, y_k \right]$$

This is most easily done by considering, as in the previous section, a

cost function consisting of the single component $\Phi(x_N)$. Then

$$c_{k+1}(p_{k+1}, y_{k+1}, \Gamma_{k+2}^{N-1}, \gamma_{k+1}(y_{k+1})) = \int\int \phi(x_N)\, p(x_N|x_{N-1}, \gamma_{N-1}(y_{N-1}))\, p(y_{N-1}|x_{N-1})\, dy_{N-1} \cdots$$

$$\cdots \int\int p(x_{N-1}|x_{N-2}, \gamma_{N-2}(y_{N-2}))\, p(y_{N-2}|x_{N-2})\, dy_{N-2} \cdots$$

$$\cdots \frac{\int\int p(x_{k+2}|x_{k+1}, \gamma_{k+1}(y_{k+1}))\, p(y_{k+1}|x_{k+1})\, p_{k+1}(x_{k+1})\, dx_{k+1}\, dx_{k+2} \cdots dx}{\int p(y_{k+1}|x_{k+1})\, p_{k+1}(x_{k+1})\, dx_{k+1}}$$

Now $$p_{k+1}(x_{k+1}) = \int\int p(x_{k+1}|x_k, \gamma_k(y_k))\, p(y_k|x_k)\, p_k(x_k)\, dx_k\, dy_k;$$

and $$p(y_{k+1}|y_k) = \frac{\int\int p(y_{k+1}|x_{k+1})\, p(x_{k+1}|x_k, \gamma_k(y_k))\, p(y_k|x_k)\, p_k(x_k)\, dx_k\, dx_{k+1}}{\int p(y_k|x_k)\, p_k(x_k)\, dx_k}$$

Thus $$E[c_{k+1}(L_k(p_k, \gamma_k), y_{k+1}, \Gamma_{k+1}^{N-1}, \gamma_{k+1}(y_{k+1}))|\, y_k]$$

$$= \int \left\{ \int\int \phi(x_N)\, p(x_N|x_{N-1}, \gamma_{N-1}(y_{N-1}))\, p(y_{N-1}|x_{N-1})\, dy_{N-1} \cdots \right.$$

$$\left. \cdots \int\int p(x_{N-1}|x_{N-2}, \gamma_{N-2}(y_{N-2}))\, p(y_{N-2}|x_{N-2})\, dy_{N-2} \cdots \right.$$

$$\cdots \frac{\int\int p(x_{k+2}|x_{k+1}, \gamma_{k+1}(y_{k+1}))\, p(y_{k+1}|x_{k+1}) \int\int p(x_{k+1}|x_k, \gamma_k(y_k))\, p(y_k|x_k)\, p_k(x_k)\, dy_k\, dx_k \cdots}{\int p(y_{k+1}|x_{k+1}) \int\int p(x_{k+1}|x_k, \gamma_k(y_k))\, p(y_k|x_k)\, p_k(x_k)\, dx_k\, dy_k\, dx_{k+1}}$$

$$\cdots \frac{\int p(y_{k+1}|x_{k+1}) \int p(x_{k+1}|x_k, \gamma_k(y_k))\, p(y_k|x_k)\, p_k(x_k)\, dx_k\, dx_{k+1}\, dy_{k+1}}{\int p(y_k|x_k)\, p_k(x_k)\, dx_k}$$

$$\triangleq \int \Big\{ \iint \phi(x_N)\, p(x_N | x_{N-1}, \gamma_{N-1}(y_{N-1}))\, p(y_{N-1} | x_{N-1})\, dy_{N-1} \cdots$$

$$\cdots \iint p(x_{N-1} | x_{N-2}, \gamma_{N-2}(y_{N-2}))\, p(y_{N-2} | x_{N-2})\, dy_{N-2} \cdots$$

$$\cdots \frac{\iint p(x_{K+2} | x_{K+1}, \gamma_{K+1}(y_{K+1}))\, p(y_{K+1} | x_{K+1})\, \phi_1(x_{K+1}, \gamma_K, p_K)\, dx_{K+1}\, dx_N}{\phi_2(y_{K+1}, \gamma_K, p_K)} \Big\} \cdot$$

$$\cdot \frac{\phi_3(y_{K+1}, \gamma_K(y_K), p_K)\, dy_{K+1}}{\phi_4(y_K, p_K)}$$

$$\neq C_K(p_K, \Gamma_{K+1}^{N-1}, \gamma_K(y_K))$$

since, amongst other things, $\Phi_2 \neq \Phi_3$ and the usual cancelling out does not happen.

As an alternative to our definition of C_k, let us consider the following conditional cost to go given our state variables p_k and y_k.

$$B_K(y_K, p_K, \Gamma_{K+1}^{N-1}, \gamma_K(y_K)) \triangleq E\big[E\big[E \cdots - E\big[\phi(x_N) \big| y_{N-1}\big] \big| y_{N-2}\big] \big| \cdots - \big] \big| y_K\big]$$

Thus $B_K(y_K, p_K, \Gamma_{K+1}^{N-1}, \gamma_K(y_K)) = \dfrac{\int \phi(x_N)\, p(x_N | x_{N-1}, \gamma_{N-1}(y_{N-1}))\, p(y_{N-1} | x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}\, dx_N}{\int p(y_{N-1} | x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \cdot$

$$\cdot \frac{\int p(y_{N-1} | x_{N-1}) \int p(x_{N-1} | x_{N-2}, \gamma_{N-2}(y_{N-2}))\, p(y_{N-2} | x_{N-2})\, p_{N-2}(x_{N-2})\, dx_{N-1}\, dx_{N-2}\, dy_{N-1} \cdots}{\int p(y_{N-2} | x_{N-2})\, p_{N-2}(x_{N-2})\, dx_{N-2}}$$

$$\cdots \frac{\int p(y_{K+1} | x_{K+1}) \int p(x_{K+1} | x_K, \gamma_K(y_K))\, p(y_K | x_K)\, p_K(x_K)\, dx_K\, dx_{K+1}\, dy_{K+1}}{\int p(y_K | x_K)\, p_K(x_K)\, dx_K}$$

The conditional cost B_k is still not fully defined by the above expansion since $p(X_{N-1}), p(X_{N-2})$ --- $p(x_{k+1})$ have not been defined yet. The state transition equation we propose to use is:-

$$p(x_k) = \int\int p(x_k \mid x_{k-1}, y_{k-1}) p(y_{k-1} \mid x_{k-1}) p(x_{k-1}) dy_{k-1} dX_{k-1}$$

that is, $p_k = \mathcal{L}_k(p_{k-1}, \gamma_k)$

By successive substitution for p_n in terms of p_{n-1} and γ_{n-1}, we can express each probability density function $\{p(x_i) \mid i = N-1, \text{---} k+1\}$ in terms of p_k and $\gamma_k(y_k)$ and Γ_{k+1}^{N-1}.

It now follows, almost by definition, that B_k and B_{k+1} are related by the following iterative relationship:-

$$B_k(p_k, y_k, \gamma_k(y_k), \Gamma_{k+1}^{N-1}) = E\left[B_{k+1}(\mathcal{L}_k(p_k, \gamma_k), y_{k+1}, \gamma_{k+1}(y_{k+1}), \Gamma_{k+2}^{N-1}) \,\middle|\, y_k\right]$$

In the above expression we are guilty of some abuse of notation. The cost B_k is, as we have denoted, a functional of $p_k, y_k, \gamma_k(y_k)$, and Γ_{k+1}^{N-1}. But notice that the iterative relationship, by virtue of the dependence of p_{k+1} on γ_k and p_k, requires that the entire function γ_{k+1} is known.

Thus to be more exact, we should write:-

$$B_k'(p_k, y_k, \Gamma_k^{N-1}) = E\left[B_{k+1}'(\mathcal{L}_k(p_k, \gamma_k), y_{k+1}, \Gamma_{k+1}^{N-1}) \,\middle|\, y_k\right],$$

where it is now made clear that to evaluate B_k', the control law γ_k is required and not just $\gamma_k(y_k)$.

The above argument can clearly be extended to include our more general additive cost function, and we obtain:-

$$B_k'(p_k,y_k,\Gamma_k^{N-1}) = E\left[L_k(x_k,\gamma_k(y_k))+B_{k+1}'(\mathcal{L}_k(p_k,\gamma_k),y_{k+1},\Gamma_{k+1}^{N-1})\,\middle|\, y_k\right]$$

At first this looks very promising as far as deriving a useful dynamic programming formulation is concerned. Suppose:-

$$B_{k+1}^{o}(p_{k+1},y_{k+1}) \triangleq \inf_{\Gamma_{k+1}^{N-1}} \left\{B_{k+1}'(p_{k+1},y_{k+1},\Gamma_{k+1}^{N-1})\right\}$$

Then, by our usual dynamic programming arguments, we can show that:-

$$B_k^{o}(p_k,y_k) = \inf_{\gamma_k} \left\{E(L_k(x_k,\gamma_k(y_k))+B_{k+1}^{o}(L_k(p_k,\gamma_k),y_{k+1})\,\middle|\, y_k\right\}$$

But this decomposition is not as attractive as it might appear, since -

1) We still have a problem which involves minimisation with respect to an entire function. This is because, as we have mentioned above, our state transition equation still demands knowledge of γ_k, and not merely $\gamma_k(y_k)$. Despite the fact that y_k is 'fixed', we are stuck with functional minimisation.

2) Even if the minimising sequence $\gamma_{N-1}^{o}, \gamma_{N-2}^{o}, \cdots \gamma_0^{o}$ could be determined, we would have determined control laws which minimise the rather peculiar cost function:-

$$J'(y_0,p_0) = E\left[L_1(X_1,\gamma_1(y_1) + E\left[L_2(x_2,\gamma_2(y_2)+\cdots\right.\right.$$

$$\left.\left.\cdots E\left[L_N(x_N)\,\middle|\, y_{N-1}\right]\,\middle|\, y_{N-2}\,\middle|\,\cdots\, y_0\right]\right.$$

This is far removed from our original statement of the problem which involved the determination of functions which would minimise,

$$J(y_0,p_0) = E\left[\sum_{n=0}^{N-1} L_n(x_n,\gamma_n(y_n))+L_N(X_N)\,\middle|\, y_0\right]$$

Recalling our application of this technique to the Witsenhausen problem in Chapter 2, it is also worth remarking that the method is just as difficult to apply as the direct functional minimisation approach of Theorem 3.1(Chapter 3).

The so-called indirect approach has been discussed here partly for completeness and partly to warn against its careless application in non-classical information pattern problems. In a classical information pattern problem, the cost functions defined above by J' and J are identical, since

$$E\left[E\left[\,\cdot\,\middle|\, y_n,y_{n-1}\right]\middle|\, y_{n-1}\right] = E\left[\,\cdot\,\middle|\, y_{n-1}\right]$$

What is more, for a classical information pattern problem we can use a state transition equation which involves $\gamma_k(y_k)$ and not all γ_k, that is

$$p(x_{k+1}) = \frac{\int p(x_{k=1}\,|\, x_k,\gamma_k(y_k))\,p(y_k\,|\, x_k)\,p(x_k)\,dx_k}{\int p(y_k\,|\, x_k)\,p(x_k)\,dx_k}$$

$$\text{i.e. } p(x_{k+1}) = \mathcal{L}_k(p_k,y_k,\gamma_k(y_k))$$

This means that we get a dynamic programming formulation involving minimisation with respect to a function value. For the zero memory

problem there is a temptation to **imitate** the technique but this is wrong because:-

1) If we use the above state transition equation we shall derive control laws which are implicitly functions of previous observations.

2) The solutions would minimise the wrong cost function, since for the zero-memory problem $J' \neq J$.

The indirect approach for the zero-memory problem as we have formulated it here can perhaps be regarded as a sub-optimal technique, since $J' \geq J$.[1] But we have already suggested that solutions are no more easy to derive by the method. Thus the technique must be classified as possibly interesting, but probably useless. No further comment will be made on the subject.

5.3.3 Open Loop Optimal Feedback Control

In the same way that Open Loop Optimal Feedback control can be specified as a sub-optimal (non-dual) technique for the classical information pattern problem, it is fairly straightforward to propose something similar for the zero-memory problem.

For the zero-memory problem version, we are interested in a cost function at time k of the form

$$J_{OLk}(y_k) = E\left[\sum_{n=k}^{N-1} L_n(x_n,u_n) + L_N(x_N) \,\middle|\, y_k\right]$$

1. Here J' denoted the cost for the zero-memory problem with traditional cost function obtained using controls derived via the indirect approach.

Note that this function depends upon $(u_k, u_{k+1} \cdots u_{N-1})$ and the probability density function $p(x_k \mid y_k)$, which can be expressed as a functional of $p(x_0)$ and Γ_0^{k-1} since,

$$p(x_k \mid y_k) = \frac{p(y_k \mid x_k)p(x_k)}{\int p(y_k \mid x_k)p(x_k)dx_k}$$

To evaluate $p(x_k)$ in terms of Γ_0^{k-1} we would use the familiar state transition equation,

$$p(x_k) = \iint p(x_k \mid x_{k-1}, \gamma_{k-1}(y_{k-1}))p(y_{k-1} \mid x_{k-1})p(x_{k-1})dy_{k-1}\,dx_{k-1}$$

in a recursive manner.

The technique then starts at time $t = 0$ and a u_0° is determined at which the minimum of $J_{OL,0}$ is achieved. This defines a function γ_0°. Notice that the value of $J_{OL,0}$ is minimised under the assumption that $u_1, u_2, \cdots u_{N-1}$ are functions of y_0, solely.

Next the situation at time $t = 1$ is examined. The probability density function $p(x_1)$ can be determined from a knowledge of $p(x_0)$ and γ_0°. This enables $p(x_1 \mid y_1)$ to be evaluated. Next we can determine that u_1° which minimises $J_{OL,1}$. This defines a function γ_1°. This time $J_{OL,1}$ is minimised under the assumption that $u_2, u_3, \cdots u_{N-1}$ are functions of y_1, solely.

This procedure is repeated at every instant in time until finally control u_{N-1} is determined. Let us compare the differences between

the solutions which the technique provides for the classical information pattern problem and the zero-memory problem.

For the classical information pattern problem, the technique will always be pessimistic in the sense that each control value is computed on the assumption that no more information will be available for future control actions than is available for the current control action.

For the zero-memory problem two conflicting assumptions are made. First, it is assumed that the current observation will be available for future control actions. Second, it is assumed that the future control actions, at some time t say, will not have available the current observation at time t. The first assumption is optimistic in the sense that it assumes some information will be available in the future, although in practice it will not. The second assumption is pessimistic in the sense that it ignores the value of the current observations at future time instants. Since it can perhaps be argued that the current observation is of more value than the previous observations, the sentiment arises that overall the technique will be pessimistic.

Let us briefly consider the application of the technique to the examples considered above in Section 5.3.1.:-

i) For the Useless Information Problem, the technique will probably produce the right answers. This statement is made because the optimal controls are 'open loop' at both time

instants anyway, for both the zero memory and classical information pattern versions of the problem. Hence it would be surprising if the open-loop optimal feedback control technique produced a different set of control laws.

ii) For the Valuable Information Case, it is a little more difficult to predict the form of the open loop optimal feedback control laws. They can, however, be easily computed:-

$$J_{OL,O}(y_0) = E\left[x_2^2 \mid y_0\right]$$

$$= E\left[(x_1 + u_1 + w_1)^2 \mid y_0\right]$$

$$= E\left[(x_0 + u_0 + w_0 + u_1 + w_1)^2 \mid y_0\right]$$

$$= E\left[x_0^2 \mid y_0\right] + 2 + \frac{4}{3} + 2u_0\, E\left[x_0 \mid y_0\right] + 2u_1\, E\left[x_0 \mid y_0\right]$$

This expression is minimised when $u_0^o = -\operatorname{sgn} E\left[x_0 \mid y_0\right]$

$$= -\operatorname{sgn} \frac{\int x_0 p(y_0 \mid x_0) p(x_0)\, dx_0}{\int p(y_0 \mid x_0) p(x_0)\, dx_0}$$

$$= -\operatorname{sgn} \frac{\sum_j X_j p_{r_j} p_j}{\sum_j p_{r_j} p_j}$$

Evaluating this expression for u_0^o gives:-

$$\bar{\gamma}_0^o \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} - \\ - \\ - \\ + \\ + \\ + \end{bmatrix},$$

a similar function for $\bar{\gamma}_1^o$ being obtained. This gives a value for $J_{OL,0}(y_0)$ as follows:-

$$J_{OL,0} \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{11}{6} \\ \frac{10}{3} \\ \frac{11}{6} \\ \frac{1}{3} \end{bmatrix}$$

Assuming values for $p(y_0)$ as computed earlier, we obtain an expected value for $J_{OL,0}$ of $\frac{162}{81}$. Comparing this with the truly optimal cost of $\frac{126}{81}$, we see that the technique is indeed pessimistic.

Let us next compute the open loop optimal feedback control laws at the next time instant t = 2.

$$J_{OL,1} = \left[E \; x_2^2 \,\middle|\, y_1 \right]$$

$$= E\left[(x_1 + u_1 + w_1)^2 \,\middle|\, y_1 \right]$$

$$= E\left[x_1^2 \,\middle|\, y_1 \right] + 1 + \frac{2}{3} + 2u_1 E\left[x_1 \,\middle|\, y_1 \right]$$

Again, this expression is minimised when $u_1^{\circ} = -\operatorname{sgn} E\left[x_1 \,\middle|\, y_1 \right]$.

$$\text{i.e. } \bar{\bar{\gamma}}_1^{\circ}(y_1) = -\operatorname{sgn} \left\{ \frac{\int x_1 p(y_1 \,|\, x_1) p(x_1) \, dx_1}{\int p(y_1 \,|\, x_1) p(x_1) dx_1} \right\},$$

whence $p(x_1) = \iint p(x_1 \,|\, x_0, \bar{\bar{\gamma}}_0^{\circ}(y_0)) p(y_0 \,|\, x_0) p(x_0) \, dx_0 \, dy_0$.

Notice that the above expression for $\bar{\bar{\gamma}}_1^{\circ}$ is identical with the expression for the truly optimal control law γ_1° obtained by using direct enumeration in Chapter 4. What is more, the control law $\bar{\gamma}_0^{\circ}$ is also identical with γ_0° obtained in Chapter 4. Thus the state variable $p(X_1)$ is the same, whether $\bar{\gamma}_0^{\circ}$ or γ_0° is used to compute it. And this implies that $\bar{\bar{\gamma}}_1^{\circ} = \gamma_1^{\circ}$.

Thus we see that the open loop optimal feedback control technique is yet another approach, like the 'neo-classical' technique, which can yield truly optimal control policies. It is a sub-optimal technique which for some examples will give genuine optimal control laws.

(iii) For the Witsenhausen problem as we might expect, we enjoy no such luck. The open loop optimal feedback control at time t = 0 is given by:-

$$\bar{\gamma}_0^{\circ}(y_0) = -E\left[x_0 \,\middle|\, y_0\right] = -y_0$$

Notice that this is different from the solution to the classical problem at time t = 0 which is $\underline{\gamma}_0^{\circ} = 0$. As we have remarked, Witsenhausen has shown that the optimal γ_0° for the zero-memory problem is non-linear, whereas $\bar{\gamma}_0^{\circ}$ above is linear. Hence the Witsenhausen example as usual provides the acid test for the technique.

This concludes our discussion of the open loop optimal feedback control technique. Our discussion of the technique has been rather sparse and extremely loose. This is because for our purposes here the technique is of little interest. This does not by any means imply that the technique itself is of no value. On the contrary, its simplicity makes it a very attractive method of providing some form of sensible feedback control in zero memory problems. To the author's knowledge however very little has been published regarding the technique. It would be interesting to know for example how close the open loop optimal control laws come to truly optimal control laws. We have seen two examples where they are identical. In cases where they are not identical, it would be useful to know what is the effect on the expected value of the cost function, in other words precisely how much sub optimal the control laws are. This would invlove the determination of bounds on

the cost function, which could possible be expressed in terms of the optimal costs of problems which in some sense are simpler, for example classical information pattern problems.

5.3.4 Optimal Linear Control

In this section we propose to discuss briefly yet another sub-optimal technique. Instead of searching for control laws among the class of all (Borel measurable) functions, we shall restrict our search to the class of all linear functions. To be more exact we shall restrict our attention to the class of affine functions (constant + proportional term). This approach has immediate and highly significant benefits:-

(i) Recall Theorem 3.1 of Chapter 3 which proved the validity of the following recursive relation for the zero-memory problem:-

$$V_k^{\circ}(p_k) = \inf_{\gamma_k \in B} \left[E\left[L_k(x_k, \gamma_k(y_k)) + V_{k+1}^{\circ}(p_{k+1}) \right] \right],$$

where $p_{k+1}(x_{k+1}) = \iint p(x_{k+1} \mid x_k, y_k; \gamma) p(y_k \mid x_k) p_k(x_k) dx_k dy_k$.

Here B denotes the class of admissable control functions, in this case the Borel Measurable functions. Let us denote the class of affine functions by $\mathcal{A}$, in other words:-

$$\gamma_k \in \mathcal{A} \Rightarrow \gamma_k(y) = a_k + b_k y$$

Then the following is a direct consequence of Theorem 3.1. It is called Theorem 5.2 here merely for reference purposes, although it is nothing more than a restatement of Theorem 3.1 in slightly different form:-

Theorem 5.2

Let $V_{k+1}^{a}(p_{k+1}) = \inf_{\Gamma_{k+1}^{N-1} \in a} E\left[\sum_{n=k+1}^{N-1} L_n(x_n, \gamma_n(y_n)) + L_N(x_N)\right]$

and $p_{k+1}(x_{k+1}) = \iint p(x_{k+1} \mid x_k, y_k; a_k, b_k) p(y_k(x_k) p_k(x_k) dx_k dy_k$

Then $V_k^{a}(p_k) = \inf_{a_k, b_k} E\left[L_k(x_k, a_k + b_k y_k) + V_{k+1}^{a}(\mathcal{L}_k(p_k; a_k, b_k))\right]$,

where the notation $\mathcal{L}_k(p_k, a_k, b_k)$ denotes the probability density function $p(x_{k+1})$ as determined by our state transition equation above when $\gamma_k \in a$.

The significant advantage of Theorem 5.2 over Theorem 3.1 is that we have a parameter optimisation problem. Our original formulation as embodied in Theorem 3.1 invloved optimisation with respect to an entire function. As we have seen in our various examples, this usually implies that for discrete state examples optimal solutions can only be determined by means of look-up-tables. For continuous state examples we have a problem in infinite dimensions. However, if we restrict our attention to the class of affine control laws then we have a finite dimensional optimisation problem. We only need to find an infimum with respect to the quantities a_k and b_k. This is of obvious advantage in computing the control laws γ_k.

(ii) For linear systems with gaussian process and observation noises, together with a gaussian distribution of the initial state x_0, con-

fining ourselves to the affine class has another immense benefit. This is the fact that p_k for any k will be gaussian also. This property has already been mentioned on several occasions elsewhere in this thesis. It can easily be demonstrated for a simple scalar example as follows:-

Process Equation $x_{k+1} = Ax_k + Bu_k + w_k \quad p(w_k) = \frac{\Phi}{\sigma_w}\left\{\frac{w_k}{\sigma_w}\right\}$

Observation Equation $y_k = Cx_k + V_k \quad p(V_k) = \frac{\Phi}{\sigma_v}\left\{\frac{v_k}{\sigma_w}\right\}$

Thus under a linear control law $u_k = \gamma_k(y_k) = a_k + b_k y_k$, we have:-

$$p(x_{k+1}) = \iint p(x_{k+1} \mid x_k, y_k; a_k, b_k)\, p(y_k \mid x_k)\, p(x_k)$$

Assuming that $p(x_k)$ is gaussian of mean $\bar{x}_k$ and variance σ_k^2,

$$p(x_{k+1}) = \iint \frac{\Phi}{\sigma_w}\left\{\frac{x_{k+1} - (Ax_k + Ba_k + Bb_k y_k)}{\sigma_w}\right\} \frac{\Phi}{\sigma_v}\left\{\frac{y_k - Cx_k}{\sigma_v}\right\} \frac{\Phi}{\sigma_k}\left\{\frac{x_k - \bar{x}_k}{\sigma_k}\right\} dx_k\, dy$$

By a variety of techniques, we can show that x_k has a gaussian probability density function of mean $\bar{x}_{k+1}$ and variance σ_{k+1}^2 where:-

$$\bar{x}_{k+1} = Ba_k + (A + Bb_k C)\,\bar{x}_k, \text{ and}$$

$$\sigma_{k+1}^2 = (A + B b_k C)^2 \sigma_k^2 + B^2 b_k^2 \sigma_v^2 + \sigma_w^2$$

Thus if $p(x_0)$ is gaussian and γ_0 is linear, then $p(x_1)$ is gaussian. By induction therefore $p(x_k)$ is also gaussian if $\gamma_1, \gamma_2 \text{ --- } \gamma_{k-1}$ are linear control laws.

The fact that $p(x_k)$ is gaussian has the considerable advantage that it can be described by the sufficient statistics $\bar{x}_k$ and σ_k^2, where these sufficient statistics are described by the linear 'state transition equation':-

$$\begin{bmatrix} \bar{x}_{k+1} \\ \sigma^2_{k+1} \end{bmatrix} = \mathcal{L}_k \begin{bmatrix} \bar{x}_k & a_k \\ \sigma_k^2, & b_k \end{bmatrix}$$

This now enables theorem 5.3 below to be proposed as a special case of Theorem 5.2:-

Theorem 5.3

For a linear system with gaussian noises and gaussian initial state, let

$$V^a_{k+1}\begin{pmatrix} x_{k+1} \\ \sigma^2_{k+1} \end{pmatrix} = \inf_{\Gamma_{k+1}^{N-1} \in a} E\left[\sum_{n=k+1}^{N-1} L_n(X_n, \gamma_n(y_n)) + L_N(x_N)\right]$$

and $$\begin{bmatrix} \bar{x}_{k+1} \\ \sigma^2_{k+1} \end{bmatrix} = \mathcal{L}_k \begin{bmatrix} \bar{x}_k & a_k \\ \sigma_k^2 & b_k \end{bmatrix},$$

$$\text{Then } V_k^a\left(\begin{matrix}\bar{x}_k \\ \sigma_k^2\end{matrix}\right) = \inf_{a_k, b_k} E\left[L_k(x_k, a_k + b_k y_k) + V_{k+1}^a\left(\mathcal{L}_k\left[\begin{matrix}\bar{x}_k \\ \sigma_k^2\end{matrix}, \begin{matrix}a_k \\ b_k\end{matrix}\right]\right)\right]$$

It might be thought that the above is a useful result. It is generally suggested that optimal linear control laws are attractive because they are easy to determine and to implement. And even though such control laws may not be truly optimal, the resulting performance is sufficiently close to the overall optimal performance criterion to render the search for more sophisticated control laws of little importance to the system designer. Indeed, a major proportion of what follows in this thesis is devoted to the problem of showing how close optimal linear control laws can come in matching the performance of truly optimal control laws.

Whilst the arguments are often valid, they fail when applied to our linear zero memory system. This is because the optimal linear control is not easy to determine as suggested above. Consider the usual algorithm implied by the above optimality equation. First we obtain the optimal a_{N-1} and b_{N-1} in terms of $\bar{x}_{N-1}$ and σ_{N-1}^2. This is usually straightforward, leading to an expression for V_{N-1}^o in terms of $\bar{x}_{N-1}$ and σ_{N-1}^2. Next we must express $\bar{x}_{N-1}$ and σ_{N-1}^2 in terms of a_{N-2}, b_{N-2}, $\bar{x}_{N-2}$, and σ_{N-2}^2, to obtain an expression $V_{N-2}(a_{N-2}, b_{N-2}, \bar{x}_{N-2}, \sigma_{N-2}^2)$ which must be minimised with respect to a_{N-2}, b_{N-2}. Admittedly, this is a scalar minimisation problem but even so it is impossible to derive convenient explicit expressions for the optimal a_{N-2} and b_{N-2}. Typically, we obtain a quartic equation to solve for the optimal a_{N-2} and b_{N-2}. Effectively, this prohibits further application of the backward induction procedure for the determination of the optimal $a_{N-3}, b_{N-3}, a_{N-4}, b_{N-4}$, and so on. This problem is dis-

cussed more fully in Chapter 7 in the more general context of the infinite horizon discount problem, where an approximation method is proposed.

5.3.5 Optimal Polynomial Control

In the previous section we considered the problem of determining the best affine laws for the zero memory problem. In this section we shall briefly consider control laws of the form $\gamma_k(y) = \sum_{i=0}^{n} b_{ki}\, y^i$ Clearly, Theorem 5.2 can be trivially extended to cater for such polynomial control laws. At each stage k we now have to find an infimum with respect to the sequence $\left\{ b_{ki} \mid i = 0 \text{ --}n \right\}$ This is still a finite dimensional parameter optimisation problem and at least is simpler than the formulation of Theorem 3.1.

The technique however suffers from the disadvantage that each b_{ki} will be a function of the infinite dimensional quantity p_k. However, for the important practical case of a linear system with gaussian noises, a non linear control law of the type suggested will result in a probability density function $p(x_k)$ which is generally not gaussian. In other words, we cannot summarise $p(x_k)$ by a pair of sufficient statistics such as $\bar{x}_k$ and σ_k^2, as we did for the determination of the optimal linear control law in the previous section. The problem of representing $p(x_k)$ adequately represents a severe computational difficulty; a number of techniques are available for resolving this difficulty.

One such technique is based on a Monte Carlo simulation method, in which the optimal parameterised control laws are determined by a first order gradient method. This is reported by C.W. Chia.[1]

Another technique would be to represent each p_k by its moments, although this would involve numerical integration and again some form

1. C.W. Chia "Determination of Optimal Controls for Partially Observable Stochastic Systems with Non-Classical Information Pattern" 1970.(11)

of simulation would be necessary. Alternatively p_k can be represented by some polynomial approximation. In Chapter 7 p_k is represented by means of Hermite polynomials, where in this case a stationary version of the zero memory problem is considered. This case has the advantage that simulation can be avoided.

5.4 Summary and Conclusions

As was stated at the beginning of the Chapter, our intention has been two fold. Firstly, to present a discussion of some of the popular pitfalls in the dynamic programming formulation of stochastic optimisation problems, and secondly to introduce a number of sub-optimal techniques.

For the classical information pattern problem we introduced the term dual control to convey the notion that a genuinely optimal control law must fulfil a dual role. It must of course control the system in some desired manner. But in addition it can be used to probe the system and thereby transmit information to the controller at subsequent time instants. Following the discussion of dual control laws, we then introduced a type of non dual control law which is confusingly referred to as "Open Loop Optimal Feedback Control". We saw that this sub-optimal technique is non-dual in the sense that the control laws at any time instant ignore the fact that future controls will be based on future observations. The technique has the one significant merit however that it is easy to use. Our review of the classical information pattern problem was concluded with a mention of optimal linear control laws and one type of misconception based on linearity arguments was examined.

The discussion of the zero memory problem was concerned with what we called the 'neo-classical' approach. It was seen that the neo classical approach was nothing more than a reformulation of the classical or

nested information pattern problem, with the unfortunate property that it looks as though it will solve the zero-memory problem. It must be admitted that the author has been guilty of enthusiastically proposing the approach as appropriate for the solution of the zero memory problem. However, since the author was not alone in this, the conceptual errors involved in applying the approach have been reported in full. In fact it is surprisingly difficult to find a zero memory problem which will refute the technique conclusively. One example was considered for which the technique gives the correct optimal cost and the correct optimal control laws. A second example was considered for which the technique gives the correct contrd laws but the wrong optimal cost. The acid test of the technique was provided by the celebrated Witsenhausen example, for which both wrong controls and wrong costs were obtained.

Next a technique which we called here 'The Indirect Approach' was discussed. It was shown that the technique is both incorrect and difficult to apply in the zero-memory problem, involving minimisation with respect to an entire function. It was pointed out however that the approach is quite appropriate for the classical information pattern problem, and it is this which probably accounts for the temptation to imitate the approach for the zero memory problem.

Having discussed two wrong approaches to the zero memory problem (which it could be argued are in themselves attractive sub optimal techniques), a number of more traditional sub-optimal techniques were introduced. We considered a zero memory version of "Open Loop Optimal Feedback Control" which it was shown is at least easy to apply. Next

the important practical class of optimal linear control laws was discussed. Finally we mentioned optimal polynomial control laws very briefly. The computation of these optimal control laws is certainly possible in theory, although in practice their determination will involve complex numerical techniques based on Monte Carlo simulation methods.

In the next chapter we shall discuss another sub optimal approach for the classical or nested information pattern problem. Essentially this is a perturbation approach.

CHAPTER 6

A Perturbation Approach to the Stochastic Optimal Control Problem

		Page
6.1	Introduction	6.1
6.2	A General Examination of the Perturbation Technique	6.5
	6.2.1 The Methodology of the Perturbation Technique	6.5
	6.2.2 The Accuracy of the Approximation	6.10
	6.2.3 Taylor Series Expansion of Some Relevant Quantities	6.15
6.3	Perturbation Solution of Classical Information Pattern Problem	6.21
	6.3.1 Optimal Control at Time N-1	6.23
	6.3.2 Optimal Control at Time N-2	6.24
	6.3.3 Optimal Control at Time k Given Optimal Control at Time k+1	6.25
	6.3.4 Optimal Cost and Explicit Control Laws	6.27
	6.3.5 Comments on the Analysis	6.29
6.4	Perturbation Approach to Zero Memory Problem	6.37
	6.4.1 Statement of Problem	6.38
	6.4.2 Dynamic Programming Formulation	6.39
	6.4.3 Derivation of Necessary Condition	6.41
	6.4.4 Perturbation Solution	6.43
	6.4.5 Comments on the Result	6.46
6.5	Summary and Conclusions	6.48

6.1. Introduction

A method of obtaining approximate analytical solutions to awkward non linear problems is the so-called Perturbation Method (see for example, "Non-Linear Mathematics" by Saaty (45)). This technique has its widest application in the deterministic theory of differential equations. However the conceptual approach of the perturbation method can be adopted in an attempt to find approximate solutions to the stochastic optimal control problem which is our concern here.

Essentially the method introduces a parameter ϵ into the formulation of the problem, where it is assumed that ϵ is a small number. The parameter ϵ is introduced in such a way that it multiplies the 'non-linearity' which is giving us trouble - put another way, when $\epsilon = 0$ the 'non-linearity' disappears and we have a problem which is relatively easy to solve. By expanding our formulation about $\epsilon = 0$ using a Taylor series, we obtain a sequence of connected sub-problems. One sub-problem will correspond to the case $\epsilon = 0$ and we assume that the sub-problem can be solved. Assuming this solution to the first sub-problem (the so-called generating solution), the next sub-problem will be linear and hence we can expect to find a solution to that sub-problem. Generally we shall derive an infinite sequence of sub-problems, each sub-problem being linear if the solutions to preceding sub-problems are known. Usually, of course, the process is terminated after a finite number of sub-problems have been solved and we end up with an approximate solution to our original problem. The next question which naturally arises is: how good is our approximate solution, or alternatively, can we estimate the degradation in performance which will occur if we use our approximate - continued over/....

solution? Finding explicit answers to this question appears to be quite difficult. Our discussion above of the perturbation method has merely been of an introductory character. In the next section we shall discuss the philosophy of the perturbation method in more detail.

Perturbation techniques are not new to the stochastic optimal control problem. One notable contribution to the subject has been made by Fleming[1], who presented a very deep and rigorous study of a continuous time, perfectly observable system with noise in the process equation. A variation of this problem had previously been reported by Kushner[2] who considered the same problem but where the noise in the process equqtion occurs as random impulses at known instants in time rather than as white noise.

The problem which will be examined here is fundamentally different from those cited above in the sense that we shall not assume perfect observation. Our small parameter ϵ will be introduced into the observation equation, for example:-

$$y = x + \epsilon v$$

For the case $\epsilon = 0$, we have a perfectly observable system, although it is not deterministic since there is still noise in the process equation. For the approach to be useful, the problem corresponding to $\epsilon = 0$ must be solvable. To ensure that this is so, in the examples to be considered later we shall restrict our attention to linear systems with quadratic costs and gaussian noises. As usual, two versions of the problem will be considered, the classical information pattern problem and the zero-

1. Fleming "Stochastic Control for Small Noise Indensities" 1970 (21)
2. Kushner "Near Optimal Control in the Presence of Small Stochastic Perturbations" 1965 (35)

memory problem. In studying the classical information pattern problem, we shall utilise Theorem 5.1 derived in Chapter 5 - our so called neo-classical approach. This approach is selected because it immediately yields control laws in terms of the current observation y_n, and - $p(x_n \mid y_0, y_1 \cdots y_{n-1}; u_0, u_1 \cdots u_{n-1})$. One might wonder why anyone should bother with a perturbation analysis of the classical information pattern LQG problem, when solutions to this problem are well known. Recall however that the crucial feature in the classical information pattern LQG problem dynamic programming formulation is that $p(x_n \mid y_0, y_1 \cdots y_{n-1}; u_0, u_1 \cdots u_{n-1})$ is gaussian at each time instant only if the initial state x_0 has a gaussian probability density function $p(x_0)$. Moreover the variance of $p(x_n \mid y_0, y_1 \cdots y_{n-1}; u_0, u_1 \cdots u_{n-1})$ is independent of the observations and this results in very fortuitous simplification. But if $p(x_0)$ is not gaussian, what then? The perturbation analysis below is directed towards providing at least a partial answer to this problem.

Notice that since the problem is linear and quadratic we could invoke the Certainty Equivalence Principle (see Chapter 9 - Section 2.4) to provide some assistance in finding a solution. Simply stated, this tells us that the optimal control law γ_k^o will be of the form

$$u_k^o = \gamma_k^o(E(x_k \mid y_0, y_1, \cdots y_k; u_0, u_1 \cdots u_{k-1})),$$

where γ_k^o is that control law which is optimal for a corresponding deterministic system obtained by setting all random disturbances permanently at their mean value. Since our intention is to demonstrate the general application of the perturbation technique we shall not pursue this approach here.

Having examined the classical information pattern problem, we shall then consider the zero memory problem via Theorem 3.1 (Chapter 3) and encounter some disappointing results. First, however, let us introduce the basic principles of the perturbation method.

6.2 A General Examination of the Perturbation Technique

In what follows we shall implicitly assume that all the various functions which we encounter are well behaved (strongly regular). This means that derivatives of appropriate order must exist and, where a change in the order of differentiation is implied, they must be continuous.

6.2.1 The Methodology of the Perturbation Technique

Consider some function V of the variables u and ϵ. We are interested in finding that function u^o of ϵ such that $V(u^o(\epsilon), \epsilon)$ is minimised for every value of ϵ.

Generally u will be a control (value), V will be the cost function and ϵ will be a parameter such that when $\epsilon = 0$ any non-linearity in the problem vanishes.

A necessary condition for minimality is $\frac{\partial V(u, \epsilon)}{\partial u} = 0$. Generally, we would attempt to solve this equation to determine the required function u^o. For problems of the type we are interested here, however, finding a solution of the equation $\frac{\partial V(u, \epsilon)}{\partial u} = 0$ is assumed to be a complex task and too difficult to contemplate. Instead we seek a method which is computationally feasible and which will provide approximate solutions to our necessary condition $\frac{\partial V(u, \epsilon)}{\partial u} = 0$.

Let us assume that this equation is satisfied by the solution - $u = u^o(\epsilon)$. In other words:-

$$\frac{\partial V(u^o(\epsilon), \epsilon)}{\partial u} = 0 \text{ for every } \epsilon$$

Next we apply the finite version of Taylor's theorem to expand the LHS of this equation about $\epsilon = 0$. We obtain:-

$$\left[\frac{\partial V}{\partial u}(u^0(\epsilon),\epsilon)\right]_{\epsilon=0} + \epsilon\left[\frac{d}{d\epsilon}\left(\frac{\partial V}{\partial u}(u^0(\epsilon),\epsilon)\right)\right]_{\epsilon=0}$$

$$+\frac{\epsilon^2}{2!}\left[\frac{d^2}{d\epsilon^2}\left(\frac{\partial V}{\partial u}(u^0(\epsilon),\epsilon)\right)\right]_{\epsilon=0}$$

$$+\frac{\epsilon^3}{3!}E_v(u^0(\epsilon),\epsilon)$$

$$= 0,$$

where E_v represents the error function.

By comparing the coefficients of ϵ, we obtain the following sequence of subsidiary necessary conditions:-

i) $\left[\frac{\partial V}{\partial u}(u^0(\epsilon),\epsilon)\right]_{\epsilon=0} = 0$ — the constant term

ii) $\left[\frac{d}{d\epsilon}\left[\frac{\partial V}{\partial u}(u^0(\epsilon),\epsilon)\right]\right]_{\epsilon=0} = 0$ — coefficient of ϵ

iii) $\left[\frac{d^2}{d\epsilon^2}\left(\frac{\partial V}{\partial u}(u^*(\epsilon),\epsilon)\right)\right]_{\epsilon=0} = 0$ — coefficient of ϵ^2

The coefficient of ϵ^3 in our original equation can be viewed as a function E_V of ϵ. We shall return to this function later. For the moment, let us write our subsidiary necessary conditions above more explicitly. We shall assume that the solution u^0 can be expressed as a finite Taylor series expansion of the form:-

$$u^*(\epsilon) = u_0 + \epsilon u_1 + \frac{\epsilon^2}{2!}u_2 + \frac{\epsilon^3}{3!}E_u(\epsilon)$$

Then for the constant term, we have:-

$$\left[\frac{\partial V}{\partial u}(u^*(\epsilon),\epsilon)\right]_{\epsilon=0} = \frac{\partial V}{\partial u}(u_0,0) = 0 \qquad \text{— (i)}$$

For the coefficient of ϵ, omitting the arguments of V for convenience, we have:-

$$\left[\frac{d}{d\epsilon}\left(\frac{\partial V}{\partial u}(u^*(\epsilon),\epsilon)\right)\right]_{\epsilon=0} = \left[\frac{\partial^2 V}{\partial u \partial \epsilon} + \frac{\partial^2 V}{\partial u^2}\cdot\frac{\partial u}{\partial \epsilon}\right]_{\epsilon=0}$$

that is, $$\frac{\partial^2 V}{\partial u \partial \epsilon}(u_0,0) + u_1 \frac{\partial^2 V}{\partial u^2}(u_0,0) = 0 \qquad \text{— (ii)}$$

For the coefficient of ϵ^2, we have:-

$$\left[\frac{d^2}{d\epsilon^2}\left(\frac{\partial V}{\partial u}(u^0(\epsilon),\epsilon)\right)\right]_{\epsilon=0} = \left[\frac{d}{d\epsilon}\left(\frac{\partial^2 V}{\partial u \partial \epsilon} + \frac{\partial^2 V}{\partial u^2}\frac{du^0}{d\epsilon}\right)\right]_{\epsilon=0}$$

$$= \left[\frac{\partial^3 V}{\partial u \partial \epsilon^2} + \frac{\partial^3 V}{\partial u^2 \partial \epsilon}\frac{du}{d\epsilon} + \frac{\partial^3 V}{\partial u^3}\left(\frac{du}{d\epsilon}\right)^2 + \frac{\partial^3 V}{\partial u^2 \partial \epsilon}\frac{du}{d\epsilon} + \frac{\partial^2 V}{\partial u^2}\frac{d^2 u}{d\epsilon^2}\right]_{\epsilon=0}$$

that is,

$$\frac{\partial^3 V}{\partial u \partial \epsilon^2}(u_0,0) + 2\frac{\partial^3 V}{\partial u^2 \partial \epsilon}(u_0,0)\,u_1 + \frac{\partial^3 V}{\partial u^3}(u_0,0)\,u_1^2 + \frac{\partial^2 V}{\partial u^2}(u_0,0)\,u_2 = 0 \qquad \text{— (iii)}$$

Let us examine the above equations (i), (ii), and (iii) more closely. Equation (i) provides us with a very convenient method of finding u_0. We simply put $\epsilon = 0$ in the original function $V(u, \epsilon)$, differentiate with respect to u, equate to zero and hopefully solve for u_0. Notice that we anticipate that solving for u_0 will be a fairly easy task since it corresponds to solving a version of our original problem in which the

non-linearity is absent. Thus, let us assume that u_0 can easily be found.

Next, consider equation (ii):-

$$\frac{\partial^2 V}{\partial u \partial \epsilon}(u_0,0) + u_1 \frac{\partial^2 V}{\partial u^2}(u_0,0) = 0$$

The term $\frac{\partial^2 V(u_0,0)}{\partial u^2}$ can be easily obtained by setting $\epsilon = 0$ in $V(u,\epsilon)$ and differentiating twice with respect to u. The resulting expression is then evaluated at u_0, which is now known. Similarly, the term $-\frac{\partial^2 V(u_0,0)}{\partial u \partial \epsilon}$ is most conveniently evaluated by differentiating $V(u,\epsilon)$ with respect to ϵ and then setting $\epsilon = 0$. (This gives the first term in a Taylor series expansion of $V(u,\epsilon)$). The resulting expression is then differentiated **once** with respect to u and evaluated at u_0. What we are then left with is a linear equation to solve for u_1, fairly easily yielding u_1 in terms of two known functions of u_0.

Finally, consider equation (iii):-

$$\frac{\partial^3 V}{\partial u \partial \epsilon^2}(u_0,0) + 2\frac{\partial^3 V}{\partial u^2 \partial \epsilon}(u_0,0) + \frac{\partial^3 V}{\partial u^3}(u_0,0)u_1^2 + \frac{\partial^2 V}{\partial u^2}(u_0,0)u_2 = 0$$

As before the various terms in this equation can be obtained by differentiating with respect to u the appropriate coefficients in Taylor series expansion of $V(u,\epsilon)$. This gives us again a linear equation to solve for u_2. At this point we have found u_0, u_1, u_2 - that is we have a second order approximation for $u^0(\epsilon)$.

6.2.2 The Accuracy of the Approximation

In the analysis above, we have restricted our attention to the determination of the first three terms in the Taylor series expansion of the optimal control. Clearly, at the expense of a considerable leap in the degree of algebraic complexity, we can extend the technique to provide the first n terms in the Taylor series expansion of the optimal control; we could seek an approximation of the form $u_n^\circ(\epsilon)$, where

$$u_n^\circ(\epsilon) \triangleq u_0 + \epsilon u_1 + \frac{\epsilon^2}{2!} u_2 + \cdots + \frac{\epsilon^n}{n!} u_n,$$

and where

$$u^\circ(\epsilon) = u_n^\circ(\epsilon) + \frac{\epsilon^{n+1}}{(n+1)!} E_u(n, \epsilon)$$

Notice also that in our analysis above, no approximations have been made. The analysis was exact since remainder terms were retained throughout for both $\frac{\delta V}{\delta u}(u(\epsilon), \epsilon)$ and $u(\epsilon)$ itself. It could of course happen that $u_0 = u_1 = u_2 = 0$ and then we have achieved precisely nothing. We would have to find the remainder term $E_u(n, \epsilon)$ and this could be just as difficult as our original problem. In a particular example, we are proposing to use $u_n(\epsilon)$ as an approximation to $u^\circ(\epsilon)$. For this to be useful, we wish to ensure therefore that, in neglecting $E_u(n, \epsilon)$, the resultant deviation of the approximate control from its optimal value produces a corresponding deterioration in the performance criterion V which is acceptable.

To be precise, the deterioration in V can be written as follows:-

$$\left| V(u^{\circ}(\epsilon), \epsilon) - V(u_n(\epsilon), \epsilon) \right| \triangleq K_n(\epsilon)$$

To apply our method with confidence we would like to know $K_n(\epsilon)$ exactly. This appears to be difficult, since $K_n(\epsilon)$ is defined in terms of $u^{\circ}(\epsilon)$ which by the definition of our problem is not easy to find. Nevertheless, failing complete knowledge of $K_n(\epsilon)$, it would be useful to know some of its properties. In particular, we would like to find a bound $M_n(\epsilon)$ such that,

$$M_n(\epsilon) \geq K_n(\epsilon)$$

Given any $\delta > 0$ it would then perhaps be possible to determine some $N(\delta, \epsilon)$ such that $n \geq N(\delta, \epsilon)$ implies $K_n(\epsilon) \leq \delta$. We could then guarantee our approximation u_n to provide a deterioration in performance no greater than δ, for a particular value of ϵ. In other words, we require,

$$\lim_{n \to \infty} M_n(\epsilon) = 0$$

The above remarks apply for any value of ϵ. However, we are not generally interested in any value of ϵ; we are primarily interested in $\epsilon = 1$. Our Taylor series expansions on the other hand were executed about the point $\epsilon = 0$. Thus we can refine our requirement on $M_n(\epsilon)$ as follows:-

$$\lim_{n \to \infty} \lim_{\epsilon \to 1^-} M_n(\epsilon) \longrightarrow 0.$$

Even if we fail to find a bound $M_n(\epsilon)$ with this ideal property, it would still be useful to find some $M_n(\epsilon)$ such that for any n, say $n = 2$, we could at least predict the accuracy of our approximation for a given ϵ.

Unfortunately the determination of suitable bounds $M_n(\epsilon)$ is in general an impossible task, of the same order of complexity as solving our original non-linear problem. It is not possible to propose an algorithm for the determination of $M_n(\epsilon)$ in the same way that $u_n(\epsilon)$ is determined. In particular problems of course we might be able to find an appropriate $M_n(\epsilon)$ - whether we can or cannot will depend on the details of our problem.

This difficulty in finding bounds $M_n(\epsilon)$ represents a fundamental inadequacy of the perturbation method. Unfortunately it is also a feature of the method which is rarely commented upon in the literature. Instead emphasis is usually laid on the behaviour of various remainder terms as ϵ tends to zero. This approach to the analysis can be used to rigorously show that our approximate solution approaches the true optimum solution (at $\epsilon = 0$) for sufficiently small ϵ. But as we have mentioned above, we are not particularly interested in small ϵ; we are more concerned with $\epsilon = 1$. For the stochastic optimal control problem to be considered later we shall see that a small ϵ is equivalent to a small noise variance. Thus the usefulness of the technique will be determined by the magnitude of the noise variance. For large noise variance the technique would appear to offer very little - it is only of value for small noise variance. It is also worthwhile recalling the example which was studied in

detail in Chapter 4. The observation noise was defined as follows:-

$$p\left\{\begin{array}{l} V_n = 1-\epsilon \\ V_n = 0 \\ V_n = -(1-\epsilon) \end{array}\right\} = \left[\begin{array}{c} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{array}\right] \quad \text{for } 0 \leqslant \epsilon < 1,$$

or alternatively,

$$p\left\{\begin{array}{l} V_n = \delta \\ V_n = 0 \\ V_n = -\delta \end{array}\right\} \left[\begin{array}{c} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{array}\right] \quad \text{for } 0 < \delta \leqslant 1$$

If we were to pursue the perturbation approach and find expansions about $\delta = 0$, we would encounter some difficulties. For δ in the range $0 < \delta < 1$ the solution is trivial, however at $\delta = 1$ the problem changes immediately from one which essentially involves perfect state observation to one involving noisy state observation. In other words there is a 'discontinuity' at $\delta = 1$.

In the light of all the pessimism about the usefulness of the perturbation method, one might wonder what point there is in considering it all. The answer if three fold:-

i) For a particular problem, we might be able to quantify the errors in such a way that we can guarantee our approximations, although this possibility is based more on faith than realism.

ii) We can dream of typical text-book problems in which the first few terms of the approximation give us a clue to the identity of the closed form solution to the problem. For the Witsenhausen problem, for example, we know the solution exists and is unique. Maybe the first few terms in the approximate solution will indicate what the likely exact solution is.

iii) We might genuinely be interested in small ϵ and be perfectly prepared to accept results only valid for a small noise variance.

In the stochastic optimal control problems which will be considered later in this Chapter, the analysis will be concerned with small ϵ. In such analysis, it is customary to introduce the following notation to describe the properties of various remainder terms which are encountered:-

i) $O(\epsilon^2)$ is to be interpreted as a function of ϵ^2 with the property that

$$\lim_{\epsilon \to 0} \frac{O(\epsilon^2)}{\epsilon^2} \leqslant \infty$$

In other words, $O(\epsilon^2)$ is of the order of ϵ^2.

ii) $o(\epsilon^2)$ is to be interpreted as a function of ϵ^2 with the property that

$$\lim_{\epsilon \to 0} \frac{o(\epsilon^2)}{\epsilon^2} = 0$$

In other words, $o(\epsilon^2)$ is of order less than ϵ^2.

6.2.3 Taylor Series Expansion of Some Relevant Quantities

In the examples which will be studied later, we will be concerned with stochastic optimal control problems in which the observation equation is of the form:-

$$y = x + \epsilon v,$$

where x is a random variable described by an unknown probability density function p_x, v is random variable representing the observation noise and described by a Gaussian probabiltiy density function p with mean 0 and variance σ_v^2.

As we have seen from our previous examinations of the stochastic optimal control problem, a dynamic programming formulation will lead to minimisations of expressions which include such terms as $E\left[x|y\right]$. This term, according to our observation equation, is now a function of ϵ and will be denoted as such. Let us consider the problem of expanding - $E\left[x|y\right]\ (\epsilon)$ as a Taylor series about $\epsilon = 0$. Two approaches to this problem will be examined. The first one, which is perhaps the most direct, will be shown to be unsuccessful; the second one will provide the required approximation:-

i) The Direct Approach

$$E\left[x|y\right]\ (\epsilon) = \int x\ p(x|y)\ dx$$

$$= \frac{\int x\ p(y|x)\ p_x(x)\ dx}{\int p(y|x)\ p_x(x)\ dx}$$

Now $p(y|x)$ is the only function which depends on ϵ;

$$P\{y \leq Y | x = X, \epsilon\} = P\left\{v \leq \frac{Y-X}{\epsilon}\right\}$$

$$= \int_{-\infty}^{\frac{Y-X}{\epsilon}} \frac{\phi}{\sigma_v}\left\{\frac{t}{\sigma_v}\right\} dt$$

$$= \int_{-\infty}^{Y} \frac{\phi}{\epsilon\sigma_v}\left\{\frac{s-X}{\epsilon\sigma_v}\right\} ds$$

Thus:-

$$p(y|x,\epsilon) = \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-x}{\epsilon\sigma_v}\right\}$$

$$= \frac{1}{\epsilon\sigma_v\sqrt{2\pi}} \exp\left\{\frac{-(y-x)^2}{2\epsilon^2\sigma_v^2}\right\}$$

and:-

$$E[x|y](\epsilon) = \frac{\int x\left(\frac{1}{\epsilon\sigma_v\sqrt{2\pi}} \exp\left\{\frac{-(y-x)^2}{2\epsilon^2\sigma_v^2}\right\}\right) p_x(x)\,dx}{\int \left(\frac{1}{\epsilon\sigma_v\sqrt{2\pi}} \exp\left\{\frac{-(y-x)^2}{2\epsilon^2\sigma_v^2}\right\}\right) p_x(x)\,dx}$$

To derive a Taylor series expansion of $E\left[x|y\right]\ (\epsilon)$ about $\epsilon = 0$, we must obtain derivatives of the above expression up to the appropriate order. It is apparent that this represents a complex and messy task. What is more, the Taylor series expansion we should obtain will not be very useful. It will be of the form:-

$$E\left[x|y\right]\ (\epsilon) = y + 0 + 0 + 0 + 0 + R(\epsilon, y),$$

Where R represents a remainder term. Apparently $E\left[x|y\right]\ (\epsilon)$ is demonstrating the same pathological behaviour when expanded about $\epsilon = 0$ as does the notorious function $\exp\left[\frac{1}{x}\right]^2$ when it is expanded about $x = 0$. This is not surprising when we note the appearance of components such as $\exp\left[-\frac{(y-x)^2}{2\epsilon^2\sigma_v^2}\right]$ in the above expression for $E\left[x|y\right]\ (\epsilon)$. To resolve this difficulty in expanding $E\left[x|y\right]\ (\epsilon)$ a change of variable is recommended.

ii) Change of Variable

Again we consider $E\left[x|y\right]\ (\epsilon)$ but now it turns out that it is the function p and not the function ϕ which must be expanded.

Without loss of generality, we can consider an observation equation of the form

$$y = x - \epsilon v,$$

where the minus sign is introduced merely as a notational convenience. In this case, we can show that:-

$$E[x/y](\epsilon) = \frac{\int x \frac{\phi}{\epsilon\sigma_v}\left\{\frac{x-y}{\epsilon\sigma_v}\right\} p_x(x)\,dx}{\int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{x-y}{\epsilon\sigma_v}\right\} p_x(x)\,dx}$$

Next, we make the substitution $s = (x - y)/\epsilon\sigma_v$, yielding

$$E[x/y](\epsilon) = \frac{\int (y+\epsilon\sigma_v s)\,\phi(s)\,p_x(y+\epsilon\sigma_v s)\,ds}{\int \phi(s)\,p_x(y+\epsilon\sigma_v s)\,ds}$$

This expression can now be fairly easily expanded about $\epsilon = 0$. We obtain

i) $$E[x/y](\epsilon)\Big|_{\epsilon=0} = y$$

ii) $$\left[\frac{d}{d\epsilon}\left[E[x/y](\epsilon)\right]\right]_{\epsilon=0} = 0$$

iii) $$\left[\frac{d^2}{d\epsilon^2}\left[E[x/y](\epsilon)\right]\right]_{\epsilon=0} = \frac{2p_x'(y)\sigma_v^2}{p_x(y)}$$

where $$p_x'(y) = \frac{dp_x(x)}{dx}\Big|_{x=y}$$

Thus we have:-

$$E[x|y](\epsilon) = y + \epsilon^2\sigma_v^2 \frac{p_x'(y)}{p_x(y)} + o(\epsilon^2)$$

As a further useful example, consider the expansion of $E\left[x^2 \mid y\right](\epsilon)$ about $\epsilon = 0$. This is most easily accomplished as follows:-

$$E[x^2|y](\epsilon) = E\left[y^2 + \epsilon^2 v^2 + 2\epsilon v y \mid y\right](\epsilon)$$

$$= y^2 + 2\epsilon y\, E[v|y](\epsilon) + \epsilon^2 E[v^2|y](\epsilon)$$

Now

$$E[v^2|y](\epsilon) = \frac{\int v^2 p(y|v) p(v)\, dv}{\int p(y|v) p(v)\, dv}$$

$$= \frac{\int v^2 p_x(y+\epsilon v) \frac{\phi}{\sigma_v}\left\{\frac{v}{\sigma_v}\right\} dv}{\int p_x(y+\epsilon v) \frac{\phi}{\sigma_v}\left\{\frac{v}{\sigma_v}\right\} dv}$$

Thus $E[v^2|y](\epsilon)\Big|_{\epsilon=0} = \sigma_v^2$

Similarly

$$E[v|y](\epsilon) = \frac{\int v\, p_x(y+\epsilon v)\frac{\phi}{\sigma_v}\left\{\frac{v}{\sigma_v}\right\}dv}{\int p_x(y+\epsilon v)\frac{\phi}{\sigma_v}\left\{\frac{v}{\sigma_v}\right\}dv},$$

with $$E[v|y](\epsilon)\Big|_{\epsilon=0} = 0,$$

and $$\frac{d}{d\epsilon}\left(E[v|y](\epsilon)\right)\Big|_{\epsilon=0} = \sigma_v^2\frac{p_x'(y)}{p_x(y)}$$

Thus:-

$$E[x^2|y](\epsilon) = y^2 + \epsilon^2\sigma_v^2\left\{1 + 2y\frac{p_x'(y)}{p_x(y)}\right\} + o(\epsilon^2)$$

The above expansions of $E\left[x \mid y\right]$ and $E\left[x^2 \mid y\right]$ have been presented primarily as examples to demonstrate the technique of expanding about $\epsilon = 0$. The technique in principle is simple but, for higher orders of ϵ, it is also extremely laborious. In what follows, no attempt will be made to detail the arduous algebra underlying the results which are derived.

6.3 Perturbation Solution of Classical Information Pattern Problem

In the previous chapter, a somewhat obscure dynamic programming formulation of the classical information pattern problem was introduced and termed the 'neo-classical' approach. This formulation was presented as Theorem 5.1 which is repeated here for convenience.

Theorem 5.1.

$$\text{Suppose } V_{n+1}^{\circ}(y_{n+1},p_{n+1}) \triangleq \inf_{\Gamma_{n+1}} V_{n+1}(y_{n+1},p_{n+1},\Gamma_{n+1})$$

$$\text{Then } V_n^{\circ}(y_n,p_n) = \inf_{u_n} E\left[L_n(x_n,u_n) + V_{n+1}^{\circ}(y_{n+1},p_{n+1}) \mid y_n\right],$$

$$\text{where } p_{n+1}(x_{n+1}) = \frac{\int p(x_{n+1} \mid x_n,u_n)p(y_n \mid x_n)p_n(x_n)dx_n}{\int p(y_n \mid x_n)p_n(x_n)dx_n}$$

In this section it is proposed to utilise Theorem 5.1 together with a perturbation analysis to derive approximate solutions to a particular classical information pattern problem. The problem to be examined is chosen principally to illustrate the technique, whilst keeping the algebra to manageable proportions. Thus we shall consider a scalar linear system with quadratic costs and Gaussian process and observation noise. The only difference between this problem and the traditional LQG problem for which the solution, as we have seen, is well known is that for our problem here the initial distribution of the state $p(\mathbf{x}_0)$ is not assumed to be Gaussian. This extension to a general $p(\mathbf{x}_0)$ in fact represents a distinct jump in the complexity of the problem. No longer

can we make the crucial simplifications which follow from the fact that $p(x_n)$ at each time n is Gaussian, with a variance which is independent of past observations and controls. As we remarked earlier, we could invoke the Certainty Equivalence Principle for this linear, quadratic problem. Instead we wish to study a general approach which will work when the entire function $p(x_n)$ is required, not simply its mean statistic. It will turn out that the perturbation technique provides us with a method of reducing the problem to an approximation of finite dimension.

The problem we shall consider is as follows:-

$x_{n+1} = x_n + u_n + \delta w_n$ - process equation, $n = 0,1,---,N-1$

$y_n = x_n - \epsilon v_n$ - observation equation,

where w_n is Gaussian with mean zero and variance σ_w^2 and v_n is Gaussian with mean zero and variance σ_v^2. The cost function, the expected value of which is to be minimised, is:-

$$J = \sum_{n=0}^{N-1} \left(x_n^2 + u_n^2 \right) + x_N^2$$

In this system ϵ and δ are small scalar constants. The problem is to determine functions $\Gamma_{N-1}^{\circ} = \gamma_0^{\circ}, \gamma_1^{\circ}, ---\gamma_{N-1}^{\circ}$ such that when $u_n = \gamma_n^{\circ}(y_0, y_1, y_2 \text{ --- } y_n; u_0, u_1, u_2 \text{ --- } u_{n-1})$, then $E\left[J\right]$ is minimised with respect to all other sequences of functions Γ_{N-1}. Here we make no assumptions about the probability distribution of the state x_0 at time 0. It is arbitrary and the corresponding probabiltiy density function will be denoted by $p(x_0)$.

It is proposed to obtain approximations to the optimal control laws as polynomials in ϵ (up to ϵ^2). The parameter δ is introduced to

facilitate the removal of the process noise, should this be necessary. For example, $\delta \to 0$ if we wish to approach the Witsenhausen problem. For the moment δ is not significant, it is merely a scalar multiplier. Let us now derive the approximate optimal solutions; these will be computed at time N-1, N-2, and a general time n. We have previously computed the optimal control at time N-1, but the computation is repeated here for completeness.

In what follows it should be remembered, in the light of our discussion in Chapter 5, that when we write a function such as $p_n(x_n)$ what we really mean is $p_n(x_n \mid y_0, y_1, y_2 \cdots y_{n-1}; \gamma_0, \gamma_1 \cdots \gamma_{n-1})$. The conditioning observations and control functions γ_i have been omitted more for brevity rather than clarity.

6.3.1 Optimal Control at time N-1

As usual,

$$V_N^\circ(y_N, p_N) \triangleq \frac{\int x_N^2 p(y_N/x_N) p_N(x_N)\, dx_N}{\int p(y_N/x_N) p_N(x_N)\, dx_N}$$

and, from Theorem 5.1,

$$V_{N-1}^\circ(y_{N-1}, p_{N-1}) = \min_{u_{N-1}} E\left[x_{N-1}^2 + u_{N-1}^2 + V_N^\circ(y_N, p_N) \mid y_{N-1}\right],$$

where
$$p_N(x_N) = \frac{\int p(x_N/x_{N-1}, u_{N-1}) p(y_{N-1}/x_{N-1}) p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}/x_{N-1}) p_{N-1}(x_{N-1})\, dx_{N-1}}$$

and
$$p(y_N/y_{N-1}) = \frac{\int p(y_N/x_N) \int p(x_N/x_{N-1}, u_{N-1}) p(y_{N-1}/x_{N-1}) p_{N-1}(x_{N-1})\, dx_{N-1}\, dx_N}{\int p(y_{N-1}/x_{N-1}) p_{N-1}(x_{N-1})\, dx_{N-1}}$$

Following the usual procedure (see Appendix 4.1.1), we obtain

$$u_{N-1}^{o} = \frac{1}{2} \; E\left[x_{N-1} \mid y_{N-1}\right] \quad ,$$

and $$V_{N-1}^{o}(y_{N-1},p_{N-1}) = 2E\left[x_{N-1}^{2} \mid y_{N-1}\right] - \frac{1}{2}\left\{E\left[x_{N-1} \mid y_{N-1}\right]\right\}^{2} + \delta^{2}\sigma_{w}^{2}$$

as we would expect the analysis at time N-1 presents no difficulty, there is no need to resort to perturbation analysis. This will not be the case at time N-2.

6.3.2 Optimal Control at time N-2

It is shown in appendix 4.1.2 that after application of Theorem 5.1, we are led to the problem of solving the equation:-

$$V_{N-2}^{o}(y_{N-2},p_{N-2}) = \min_{u_{N-2}} \left\{3u_{N-2}^{2} + 3E\left[x_{N-2}^{2} \mid y_{N-2}\right] + 4u_{N-2}E\left[x_{N-2} \mid y_{N-2}\right]\right.$$

$$\left. + \; 3\delta^{2}\sigma_{w}^{2} - \frac{1}{2}E\left[E\left(x_{N-1} \mid y_{N-1}\right)\right]^{2} \mid y_{N-2}\right]$$

It is the appearance of the term $E\left[\left(E\left[x_{N-1} \mid y_{N-1}\right]\right)^{2} \mid y_{N-2}\right]$ in the above expression which makes difficult the determination of the optimal u_{N-2}. And it is to resolve this difficulty that a perturbation technique is introduced.

The details of the analysis are tedious and consequently are presented in Appendix 4.1.2. It is shown there that:-

$$u_{N-2}^{o} = -\frac{3}{5}\left\{ y_{N-2} + \epsilon^2 \sigma_v^2 \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} \right\} + o(\epsilon^2),$$

and that:-

$$V_{N-2}(y_{N-2}, p_{N-2}) = \frac{8}{5} y_{N-2}^2 + \frac{16}{5} y_{N-2} + \epsilon^2 \sigma_v^2 \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} + \frac{5}{2}\delta^2 \sigma_w^2 + 3\epsilon^2 \sigma_v^2 + o(\epsilon^2).$$

6.3.3 Optimal Control at time k given optimal solution at time k+1

In the traditional LQG stochastic optimal control problem it is possible to develop useful recursive relations describing the optimal costs and control. In the same way, similar recursive relations can be developed for the problem we are examining here using perturbation analysis.

Suppose that at time k+1 the optimal cost to go V_{k+1}^{o} is given by an expression of the form below:-

$$V_{k+1}^{o}(y_{k+1}, p_{k+1}) = a_{k+1} y_{k+1}^2 + b_{k+1} y_{k+1} \epsilon^2 \sigma_v^2 \frac{p'_{k+1}(y_{k+1})}{p_{k+1}(y_{k+1})} + c_{k+1} \epsilon^2 \sigma_v^2 + d_{k+1} \delta^2 \sigma_w^2 + o(\epsilon^2)$$

Then we can derive recursive relations involving $a_{k+1}, b_{k+1}, c_{k+1}$, and d_{k+1} which can be used to determine the optimal cost to go V_k^{o} and the corresponding optimal control law u_k^{o} at time k. The procedure follows that same pattern as the procedure used to determine V_{N-2}^{o} and u_{N-2}^{o}.

The details of the analysis are presented in Appendix 4.1.3. For V_k^o we obtain:-

$$V_k^o(y_k,p_k) = a_k y_k^2 + b_k y_k \epsilon^2 \sigma_v^2 \frac{p_k'(y_k)}{p_k(y_k)} + c_k \epsilon^2 \sigma_v^2 + d_k \sigma_w^2 + o(\epsilon^2),$$

where the coefficients are defined by the following recursive relations, and their indicated starting values:-

$$a_k = 1 + 2a_{k+1}/(1 + a_{k+1}) \; ; \; a_{N-1} = \frac{3}{2}$$

$$b_k = 2(1 + 2a_{k+1})/(1 + a_{k+1}) \; ; \; b_{N-1} = 3$$

$$c_k = 1 + c_{k+1} \; ; \; c_{N-1} = 2$$

$$d_k = a_{k+1} + d_{k+1} \; ; \; d_{N-1} = 1$$

The optimal control law at time k is given by

$$u_k^o = \gamma_k^o(y_k,p_k) = -e_k \left\{ y_k + \epsilon^2 \sigma_v^2 \frac{p_k'(y_k)}{p_k(y_k)} \right\} + o(\epsilon^2),$$

where $$e_k = \frac{a_{k+1}}{1 + a_{k+1}}$$

6.3.4 Optimal Cost and Explicit Control Laws

At this point we have more or less completed the perturbation analysis of the classical information pattern problem. There is very little left to do. By working backwards from time N-1, we can find each control value u_i^o as a function γ_i^o of y_i and $p_i(x_i)$, for $i = 0,1,---N-1$. Also we can very simply determine the optimal cost $E\left[J^o\right]$ for the problem as follows:-

$$E[J^o(p_o)] = \int V_o^o(y_o,p_o)p(y_o)\,dy_o$$

$$= \int \left\{ a_o^2 y_o^2 p_o(y_o) + \frac{\epsilon^2\sigma_v^2 a_o}{2}\left\{ 4y_o p_o'(y_o) + y_o^2 p_o''(y_o)\right\}\right\} dy_o$$

$$+ c_o\epsilon^2\sigma_v^2 + d_o\delta^2\sigma_w^2 + o(\epsilon^2)$$

Finally, the last step to perform before we have totally completed our solution of the problem is to eliminate the dependence of u_i^o upon $p_i(x_i)$ by elimination using the state transition equation. Recall that the general form of the control law at time k is:-

$$u_k^o = \gamma_k^o(y_k,p_k) = -e_k\left\{ y_k + \epsilon^2\sigma_v^2\frac{p_k'(y_k)}{p_k(y_k)}\right\} + o(\epsilon^2)$$

Since $p_0(x_0)$ is assumed known, we already have u_0 in the desired form, that is we can write,

$$u_o = \bar{\gamma}_o^o(y_o) = -e_o\left\{ y_o + \epsilon^2\sigma_v^2\frac{p_o'(y_o)}{p_o(y_o)}\right\} + o(\epsilon^2)$$

Next consider the task of eliminating $p_1(x_1)$ from our expression for u_1^o. We have:-

$$u_1^o = \gamma_1^o(y_1, p_1) = -e_1\left\{ y_1 + \epsilon^2\sigma_v^2 \frac{p_1'(y_1)}{p_1(y_1)} \right\} + o(\epsilon^2)$$

To eliminate the dependence of u_1^o upon $p_1(x_1)$ we must use our state transition equation, which in its general form, has been defined elsewhere as follows:-

$$p_{k+1}(x_{k+1}) = \frac{\int p(x_{k+1} \mid x_k, u_k)\, p(y_k \mid x_k)\, p_k(x_k)\, dx_k}{\int p(y_k \mid x_k)\, p_k(x_k)}$$

Notice however that we are not concerned with the entire function $p_{k+1}(x_{k+1})$. All that we require for our $o(\epsilon^2)$ approximation for u_k^o is an $O(\epsilon)$ approximation for $p_{k+1}'(y_{k+1})/p_{k+1}(y_{k+1})$. As we have seen before this is given by:-

$$\frac{p_{k+1}'(y_{k+1})}{p_{k+1}(y_{k+1})} = -\left\{ \frac{y_{k+1} - (y_k + u_k)}{\delta^2\sigma_w^2} \right\} + O(\epsilon)$$

Thus for u_1^o we obtain:-

$$u_1^o = -e_1\left\{ y_1 - \epsilon^2\sigma_v^2 \left\{ \frac{y_1 - (y_0 + u_0)}{\delta^2\sigma_w^2} \right\} \right\} + o(\epsilon^2)$$

that is,

$$u_1^o = \bar{\gamma}_1^o(y_1, y_0) = -e_1\left\{ y_1 - \epsilon^2\sigma_v^2\left\{\frac{y_1-(y_0-e_0y_0)}{\delta^2\sigma_w^2}\right\}\right\} + O(\epsilon^2)$$

$$= -e_1\left\{ y_1 - \frac{\epsilon^2\sigma_v^2}{\delta^2\sigma_w^2}\left(y_1-(1-e_0)y_0\right)\right\} + O(\epsilon^2)$$

In general, we shall obtain for u_k^o,

$$u_k^o = \bar{\gamma}_k^o(y_k, y_{k-1}) = -e_k\left\{ y_k - \frac{\epsilon^2\sigma_v^2}{\delta^2\sigma_w^2}\left(y_k-(1-e_{k-1})y_{k-1}\right)\right\} + O(\epsilon^2)$$

where e_k and e_{k-1} have been defined previously.

Notice that, for an $O(\epsilon^2)$ approximation, the optimal control only depends on the previous and current observations. What is more, apart from γ_0^o, the optimal control is independent of the initial probability distribution $p(x_0)$ of x_0. Further comments on this and other matters will be made below.

6.3.5 Comments on the Analysis

Above we have presented in some detail an example of the application of perturbation analysis to a stochastic optimal control problem with

a classical information pattern. The only difference between our problem and the traditional LQG problem is that here we made no assumption about the probability distribution of the state x_0 at time 0. As we have seen in earlier Chapters, the traditional LQG theory relies on the fact that $p_0(x_0)$ is Gaussian. This assumption, for the classical information pattern problem, implies that for any time n, $p_n(x_n \mid I_n)$ is Gaussian with a variance which is independent of I_n. When, as we have assumed here, $p_0(x_0)$ is arbitrary we can make no predictions about the structure of $p_n(x_n \mid I_n)$ and an exact analysis becomes impossible. What we have presented above is a demonstration of how perturbation techniques can be applied to derive approximate solutions to the case when $p_0(x_0)$ is arbitrary.

In this presentation, the perturbation technique was demonstrated in detail and in theory the same procedure could be applied to any problem, providing the solutions at $\epsilon = 0$ can be obtained. The full step by step procedure was described with this possibility in mind. The only concession made to brevity was to omit some details of the very tedious algebra which is involved in applying the method. Note that even without appeal to the Certainty Equivalence Principle for the particular problem examined here, the analysis can be reduced to almost trivial proportions by applying a trick similar to that employed in the traditional LQG problem:-

Recall that the optimal cost at time N-1 is given by V^{o}_{N-1}, where:-

$$V^{o}_{N-1}(y_{N-1}, p_{N-1}) = 2E[x_{N-1}|y_{N-1}] - \frac{1}{2}\left(E[x_{N-1}|y_{N-1}]\right)^2 + \delta^2\sigma_w^2$$

$$= \frac{3}{2}E[x^2_{N-1}|y_{N-1}] + \frac{1}{2}\left\{E[x^2_{N-1}|y_{N-1}] - \left(E[x_{N-1}|y_{N-1}\right)^2 + \delta^2\sigma_w^2\right.$$

Now

$$E[x^2_{N-1}|y_{N-1}] = y^2_{N-1} + \epsilon^2\sigma_v^2\left\{1 + 2y_{N-1}\frac{p'_{N-1}(y_{N-1})}{p_{N-1}(y_{N-1})}\right\} + o(\epsilon^2)$$

and

$$\left(E[x_{N-1}|y_{N-1}]\right)^2 = y^2_{N-1} + 2\epsilon^2\sigma_v^2 y_{N-1}\frac{p'_{N-1}(y_{N-1})}{p_{N-1}(y_{N-1})} + o(\epsilon^2)$$

Thus

$$E[x_{N-1}|y_{N-1}] - \left(E[x_{N-1}|y_{N-1}]\right)^2 = \epsilon^2\sigma_v^2 + o(\epsilon^2)$$

which gives us:-

$$V^{o}_{N-1}(y_{N-1}, p_{N-1}) = \frac{3}{2}E[x^2_{N-1}|y_{N-1}] + \frac{1}{2}\epsilon^2\sigma_v^2 + \delta^2\sigma_w^2 + o(\epsilon^2)$$

This simplification in the expression for V^{o}_{N-1} is equivalent to the simplification in the traditional LQG problem when it is assumed that the variance of $p_n(x_n|I_n)$ is independent of I_n.

At each stage in the backward induction process we can exploit the fact that the $o(\epsilon^2)$ approximation to the variance of $p(x_n \mid I_n)$ is independent of I_n to greatly lighten the algebraic load. We ulitmately obtain recursion relations which are equivalent to those presented earlier. This short cut method however, can only be used for linear systems with quadratic costs and gaussian process and observation noises; and even then, only for $o(\epsilon^2)$ approximations. The simplification rests on the fact that the variance (or at least our approximation to it) -

$$E\left[x_k^2 \mid y_k\right] - \left(E\left[x_k \mid y_k\right]\right)^2,$$

is independent of the previous controls and observations.

Let us conclude this section with some general comments regarding our application of the perturbation technique:-

i) Our dynamic programming approach was based on what we have previously called the neo-classical formulation. As we saw in Chapter 5, this approach is nothing more than a slightly different formulation of the traditional dynamic programming approach. Also, it has been previously remarked that this formulation is open to misinterpretation as a formulation of the zero-memory problem. This is because the optimal cost-to-go functions V_n^o depended on y_n and p_n, where it is sometimes forgotten that p_n is the conditional p.d.f. of x_n given - $y_0, y_1, \cdots y_{n-1}$ and $u_0, u_1 \cdots u_{n-1}$. It must be admitted that the notation - continued over/...

adopted in our analysis above does not contribute anything to removing this source of confusion, since we have used terms such as:-

$$E\left[x_k \mid y_k\right],$$

where the motivation for this notation is provided by the working definition:-

$$E\left[x_k \mid y_k\right] \triangleq \frac{\int x_k \; p(y_k \mid x_k) p_k(x_k) dx_k}{\int p(y_k \mid x_k) p_k(x_k) dx_k}$$

However, as we saw in Chapter 5, the state transition equation means that p_k is in fact implicitly conditioned on $y_0, y_1 \cdots y_{k-1}$ and $u_0, u_1, \cdots u_{k-1}$. Thus:-

$$E\left[x_k \mid y_k\right] = \frac{\int x_k \; p(y_k \mid x_k) p(x_k \mid y_0, y_1, \cdots y_{k-1}; u_0, u_1 \cdots u_{k-1}) dx_k}{\int p(y_k \mid x_k) p(x_k \; y_0, y_1, \cdots y_{k-1}; u_0, u_1, \cdots u_{k-1}) dx_k}$$

$$= \int x_k \; p(x_k \mid y_0, y_1, \cdots y_k; u_0, u_1, \cdots u_{k-1}) dx_k,$$

and this is the expected value of x_k conditioned on all past and present observations and controls, and not merely the present observation as the notation would imply.

ii) In the perturbation **analysis,** note that ϵ always occurs as a multiplier of σ_v. Thus when we speak of small ϵ, we mean small σ_v and we can interpret all our analysis as an $o(\sigma_v^2)$ approximation to the optimal solution.

iii) The only difference between the problem we considered in this section and the traditional LQG problem is that here we make no assumptions about the structure of $p_0(x_0)$. In the traditional problem $p_0(x_0)$ is gaussian. For an $o(\epsilon^2)$ approximation however, we have seen that $p_0(x_0)$ only influences the control law at time $t = 0$ - all other control laws at time $t > 0$, are functions only of the current observation and the preceding observation. In no way do these control laws depend on $p_0(x_0)$. We would expect of course that for higher order approximations, the effect of $p_0(x_0)$ would be propagated further into control laws at times $t > 0$.

iv) By accident, we have derived the $o(\epsilon^2)$ approximation to the optimal solution of a non-classical information pattern problem. Because the $o(\epsilon^2)$ control law at each time $t = n$ is a function of only y_n and y_{n-1}, our solutions are also $o(\epsilon^2)$ approximations to the optimal solutions of the non-classical information pattern problem in which the control data I_n is defined as follows:-

$$I_n = (y_n, y_{n-1})$$

It is an interesting speculation that higher order approximations (say $o(\epsilon^4)$) will be functions of a more retrospective portion of the historical data (say y_n, y_{n-1}, y_{n-2}, and y_{n-3}). Supposing such a speculation were true, we would then have derived an $o(\epsilon^4)$ approximation to the optimal solution of a second non-classical information pattern problem; one in which I_n is defined as follows:-

$$I_n = (y_n, y_{n-1}, y_{n-2}, y_{n-3}).$$

This kind of result is of little value in calculating general approximations (say $o(\epsilon^n)$) to a particular non-classical information pattern problem - the zero memory problem for example. However, such a result could be useful in assessing the value of different information patterns without having to find the corresponding optimal control laws. For example, we have shown that the value of information between the classical information pattern problem with I_n defined as:-

$$I_n = (y_n, y_{n-1}, y_{n-2} \text{ --- } y_0),$$

and the non classical information pattern problem with I_n defined as:-

$$I_n = (y_n, y_{n-1}),$$

is $o(\epsilon^2)$.

v) For completeness, we should remark that for $\epsilon \rightarrow 0$, our solutions approach the correct solutions obtained at $\epsilon = 0$. Our solutions for $\epsilon \rightarrow 0$ can be determined by inspection and it is a trivial matter to verify that these are the solutions that are obtained for the problem in which $\epsilon = 0$.

vi) In the preceding discussion we have mentioned on several occasions the possibility of determining higher order approximations. Unfortunately we have no guarantee that a convenient short cut method of the type discussed earlier can be utilised. In this event we are forced to embark upon a course which employs the full laborious technique presented in **Appendix 4. For approxi-**

mations of greater accuracy than $o(\epsilon^2)$, the prospect is depressing.

In an attempt to automate the tedium of the technique, considerable computer experimentation was performed using a symbolic manipulation language. The language used was FORMAC[1] (Formula Manipulation Compiler), which was available on an IBM 7094 computer on an experimental basis. Despite the tremendous assistance which this facility offers, it was discovered that, even for our simple problem here, the length of the expressions involved rapdily exceeded the capacity of the system. It understood that later versions of FORMAC, available as a superset of PL/1, do not suffer from such severe limitations and further experimentation might be fruitful. In the words of IBM, "FORMAC can be used to obtain symbolic solutions in problem areas which heretofore could only be approached numerically".

This concludes our comments on the perturbation approach to the classical information pattern problem. In the next section we shall attempt to apply the technique to a zero memory problem.

1) 7090/94 FORMAC - An Experimental Formula Manipulation Compiler. Order No. 7090-R2 3538 PRP

6.4 Perturbation Approach to Zero Memory Problem

In this section, a perturbation approach to the Witsenhausen problem will be examined. As we have seen before, the Witsenhausen problem is a particularly simple example and yet it contains the essential features of a non-classical information pattern problem. In other words, the determination of the optimal control laws is an extraordinarily difficult task and, as yet, has proved impossible. We shall see that the perturbation approach does provide what could be useful approximations to the optimal control law. However, we shall also see that the control laws is a linear approximation and as such offers us no clues to the identity of the mysterious non linear optimal control law. Witsenhausen, as we have stated in Chapter 2, has shown that this control law exists, that it is unique, and that in some instances outperforms the optimal linear control law. Unfortunatlely, nobody knows what it is.

We have already introduced the Witsenhausen example on several occasions. For ease of reference the problem will be re-stated here. We shall then use Theorem 31 of Chapter 3 to derive a necessary condition for the optimal control law at time $t = 0$. Variational calculus will then be employed to derive a necessary condition suitable for perturbation analysis. This form of necessary condition (a Gateaux differential) is of importance in its own right and we shall return to examine it more closely in Chapter 8. In what follows, every attempt will be made to make the development as brief as possible, without becoming mystical.

6.4.1 Statement of Witsenhausen's Problem

Consider the dynamic system,

$$x_1 = x_0 + u_0$$

$$x_2 = x_1 + u_1$$

with observation equations,

$$y_0 = x_0$$

$$y_1 = x_1 + v,$$

where x_0 represents a Gaussian random variable with mean zero and variance σ_0^2 and v represents a Gaussian random variable with mean zero and variance σ_v^2. In what follows we shall generally use the notation $p_0(x)$ to denote the probability density function of x_0. Only when we wish to denote the specific nature of $p_0(x)$ will the notation $\frac{\Phi}{\sigma_0}\left\{\frac{x}{\sigma_0}\right\}$ be introduced. The admissable controllers are defined as follows:-

$$u_0 = \gamma_0(y_0)$$

$$u_1 = \gamma_1(y_1)$$

It is desired to select functions $\{\gamma_0, \gamma_1\}$ from the class of Borel measurable functions such that,

$$E\left[k^2(u_0)^2 + (x_2)^2\right], \quad k^2 > 0$$

is minimised.

The information pattern is non-classical because the value of y_0 is known at the first stage but not at the second. Let us denote the class of pairs (γ_0, γ_1) as Γ.

6.4.2 Dynamic Programming Formulation

In Chapter 3 we developed a dynamic programming formulation of the zero memory stochastic optimal control problem. It was expressed as Theorem 3.1:-

Theorem 3.1

$$\text{If } V^0_{k+1}(p_{k+1}) = \inf_{\Gamma^{N-1}_{k+1}} \left\{ E\left[\sum_{n=k+1}^{N-1} L_n(x_n, \gamma_n(y_n)) + L_N(x_N) \right] \right\},$$

$$\text{and } p_{k+1}(x_{k+1}) = \int\int p(x_{k+1} | x_k, \gamma_k(y_k))\, p(y_k | x_k)\, p_k(x_k)\, dy_k\, dx_k,$$

$$\text{then } V^0_k(p^0_k) = \inf_{\gamma_k} \left\{ E[L_k(x_k, \gamma_k(y_k)] + V^0_{k+1}(p_{k+1}) \right\}$$

It is shown in Appendix 4.2.1 that the application of Theorem 3.1 to the Witsenhausen problem results in a requirement to find a function f which minimises a cost functional $J(p_0, \quad, f)$, where:-

$$J(p_0, \epsilon, f) = E\left[k^2(f(x)-x)^2\right] + \epsilon^2 \sigma_v^2 - \epsilon^4 \sigma_v^4 \int \frac{d}{dy}\left[D_f(y)\right] \frac{dy}{D_f(y)}$$

in which

$f(x) = x + \gamma_0(x)$, redefines the control law γ_0

$y = x + \epsilon v$, redefines the observation equation,

and $D_f(y) = \int \frac{\Phi}{\sigma_v}\left\{\frac{y - f(x)}{\sigma_v}\right\} p_0(x)\, dx$

It is proposed to derive a necessary condition for that f° which minimises $J(p_0, \epsilon, f)$ and then attempt to solve this condition of f by employing perturbation analysis.

6.4.3 Derivation of Necessary Condition

The necessary condition to be derived below will be referred to later in Chapter 8 when we consider the possibility of finding conditions which ensure that the condition has a unique solution.

The necessary condition is derived by variational techniques. Suppose the optimal f is f°. Consider a perturbation about f° as follows:-

$$f(x,\lambda) = f^{\circ}(x) + \lambda \delta f(x)$$

where λ is some scalar multiplier and $\delta f(x)$ is any arbitrary function. Then the necessary condition for optimality is:-

$$\left.\frac{\partial J}{\partial \lambda}(p_0, \epsilon, F)\right|_{\lambda=0} = 0$$

Now,

$$\frac{\partial J}{\partial \lambda} = \frac{\partial J_1}{\partial \lambda} - \frac{\partial J_2}{\partial \lambda}$$

where,

$$\frac{\partial J_1}{\partial \lambda} \triangleq k^2 \frac{\partial}{\partial \lambda} E[(f(x,\lambda)-x)^2]$$

$$= k^2 \frac{\partial}{\partial \lambda}\left\{\int (f^0(x)+\lambda \delta f(x)-x)^2 p_0(x)\,dx\right\}$$

that is,

$$\left.\frac{\partial J_1}{\partial \lambda}\right|_{\lambda=0} = k^2 \int 2(f^0(x)-x)\,\delta f(x)\,p_0(x)\,dx$$

and,

$$\frac{\partial J_2}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left\{\epsilon^4 \sigma_v^4 \int \left\{\frac{d}{dy}[D_f(y)]\right\}^2 \frac{dy}{D_f(y)}\right\}$$

After some algebra, we obtain

$$\left.\frac{\partial J_2}{\partial \lambda}\right|_{\lambda=0} = \iint \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-f^0(x)}{\epsilon\sigma_v}\right\}\frac{D_f^{0\prime}(y)}{D_f^0(y)}\left\{\frac{2}{\epsilon^2\sigma_v^2} - 2\left\{\frac{y-f^0(x)}{\epsilon^2\sigma_v^2}\right\}^2 - \frac{D_f^{0\prime}(y)}{D_f^0(y)}\left(\frac{y-f^0(x)}{\epsilon^2\sigma_v^2}\right)\right\}\delta f(x)\,p_0(x)\,dx\,dy,$$

where

$$D_f'(y) \triangleq \frac{d}{dy}[D_f(y)]$$

Thus, our necessary condition can be expressed as:-

$$\left.\frac{\partial J}{\partial \lambda}\right|_{\lambda=0} = k^2 \int 2\,(f^{\circ}(x)-x)\,\delta f(x)\,p_0(x)\,dx$$

$$+\epsilon^2\sigma_v^4 \iint \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-f^{\circ}(x)}{\epsilon\sigma_v}\right\}\frac{D_f'(y)}{D_f(y)}\left\{2\left(\frac{y-f^{\circ}(x)}{\epsilon^2\sigma_v^2}\right)^2+\frac{D_f'(y)}{D_f(y)}\,\frac{(y-f^{\circ}(x))}{\epsilon^2\sigma_v^2}-2\right\}.$$

$$\cdot\ \delta f(x)\,p_0(x)\,dx\,dy$$

Since this condition must be satisfied for any $\delta f(x)$, we must have:-

$$2k^2(f^{\circ}(x)-x)+\epsilon^2\sigma_v^2\,\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-f^{\circ}(x)}{\epsilon\sigma_v}\right\}\frac{D_{f0}'(y)}{D_{f0}(y)}\left\{2\left(\frac{y-f^{\circ}(x)}{\epsilon\sigma_v}\right)^2+\frac{D_{f0}'(y)}{D_{f0}(y)}\,(y-f^{\circ}(x))-2\right\}$$

This condition must be satisfied for every x. Notice that it is still a rather complex integral equation; f° must be defined over its entire domain in order to define $D_f'(y)$ and $D_f(y)$. In the next section we shall explore the possibility of finding useful approximations to f°.

6.4.4 Perturbation Solution

Let us write our necessary condition in a form demonstrating explicitly the dependence upon ϵ :-

$$2k^2\left\{f^{\circ}(x,\epsilon)-x\right\}=\epsilon^2\sigma_v^2\int\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-f^{\circ}(x,\epsilon)}{\epsilon\sigma_v}\right\}\frac{D_{f0}'(y,\epsilon)}{D_{f0}(y,\epsilon)}\left\{2\left(\frac{y-f^{\circ}(x,\epsilon)}{\epsilon\sigma_v}\right)^2\right.$$

$$\left.+\frac{D_{f0}'(y,\epsilon)}{D_{f0}(y,\epsilon)}\,(y-f^{\circ}(x,\epsilon))-2\right\}dy$$

It is intended to expand the RHS of this equation about $\epsilon = 0$ using a Taylor series. Unfortunately this is not possible for any general $f(x, \epsilon)$ - the problem is similar to that encountered when the expansion of $E\left[x \mid y\right]$ (ϵ) was considered earlier in this chapter. What is required is a change of variable, and this requires an invertible $f^{\circ}(x, \epsilon)$. So, suppose:-

$$z \triangleq f^{\circ}(x, \epsilon)$$

and

$$x = g(z, \epsilon)$$

Then

$$D_g(y) = \int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-z}{\epsilon\sigma_v}\right\} p_o(g(z,\epsilon)) \frac{dg}{dz}(z,\epsilon)\,dz,$$

and we have the following necessary condition,

$$2k^2(g(z,\epsilon)-z)$$

$$= \epsilon^2\sigma_v^2 \int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-z}{\epsilon\sigma_v}\right\} \frac{D_g'(y,\epsilon)}{D_g(y,\epsilon)}\left\{2\left(\frac{y-z}{\epsilon\sigma_v}\right)^2 + \frac{D_g'(y,\epsilon)}{D_g(y,\epsilon)}(y-z) - 2\right\} dy$$

for every z. Notice that the assumption that f is invertible is introduced to permit subsequent expansion of the necessary condition.

In his paper, however, Witsenhausen proves that f is only monotone non- decreasing. Hence our assumption should be treated with suspicion.

Notice that for $\epsilon = 0$ in the above condition, we obtain

$$2k^2(g(z,0)-z) = 0,$$

which implies $$g(z,0) - z = 0$$

that is $$f^0(x,0) = x$$

or alternatively $$\gamma_0^0(x,0) = 0$$

This is exactly what we would expect for the perfect observation case ($\epsilon = 0$). The optimal strategy is to do nothing at the first stage and save all the corrections until the second stage. Notice from these considerations that

$$g(z,0) = z$$

and

$$\frac{d}{dz}(g(z,0)) = 1$$

These properties of g are used in Appendix 4.2.2 in the expansion of the necessary condition.

From Appendix 4.2.2 we obtain:-

$$g(z,\epsilon) = z - \frac{\epsilon^4 \sigma_v^4 z}{2k^2} + o(\epsilon^4)$$

that is

$$g(z,\epsilon) = z\left(1 - \frac{\epsilon^4 \sigma_v^4}{2k^2}\right) + o(\epsilon^4)$$

Now $g(z,\epsilon) = x$ and $z = f^0(x,\epsilon)$. Thus:-

$$f^0(x,\epsilon) = x\Big/\left(1 - \frac{\epsilon^4 \sigma_v^4}{2k^2}\right) + o(\epsilon^4)$$

and

$$\gamma_0^0(x,\epsilon) = f^0(x,\epsilon) - x$$

that is,

$$\gamma_0^0(x,\epsilon) = \frac{x\epsilon^4 \sigma_v^4}{2k^2} + o(\epsilon^4)$$

So this is our $o(\epsilon^4)$ approximation to the optimal control law for the Witsenhausen problem. For all our patient and tedious development, the result is not very exciting. Assuming that the algebra is accurate, the result is nevertheless worthy of some comment.

6.4.5. Comments on the Result

First for $\epsilon = 0$, the approximation is, as we have already seen, what we would expect. In the absence of noise, do nothing at the first

stage and leave the corrections to the second stage is the optimal strategy. The policy is still valid for $o(\epsilon^2)$ - approximations to the optimal control. However we begin to see something interesting when an $o(\epsilon^4)$ - approximation is considered. The approximation is linear but notice that correction is not being made to reduce the value of the state. Apparently it is being made to increase the value of the state - in other words, the correction is concerned with communication rather than control. The correction seeks to impart information to the second stage controller rather than perform any controlling action. This phenomenon was first discussed in Chapter 2. Suppose σ_v^4 is large and k^3 small, then a massive control will be applied at the first stage in order to enable the second stage to accurately estimate the state at the second stage and exert a minimising influence.

That the $o(\epsilon^4)$ term in our approximation is more concerned with communication than control is not surprising when we inspect the cost functional that we are attempting to minimise with $\gamma_0(x)$ (or $f^0(x)$). Recall from Section 6.4.2 that this is as follows:-

$$J(p_0, \epsilon, f) = E\left[k^2(f(x) - x)^2\right] + \epsilon^2\sigma_v^2 - \epsilon^4\sigma_v^4 \int (D_f'(y))^2 \, dy/D_f(y)$$

The term $\int (D_f'(y))^2 \, dy / D_f(y)$ is the Fisher Information of the random variable y. As Witsenhausen points out, the designer is trying to find a compromise between:-

i) keeping the cost of the first stage correction small, and

ii) making the Fisher Information of the observation available at the second stage large.

Since the Fisher Information appears as an effect of order $\alpha\epsilon^4$), we would expect terms involving ϵ^4 in our approximation to be concerned with making the Fisher Information large.

6.5 Summary and Conclusions

The purpose of this Chapter has been to introduce the technique of perturbation analysis. Thus some time was spent introducing the basis of technique. It was pointed out that a major drawback of the technique is that it is only valid for small values of ϵ (or σ_v). Unfortunately no convenient method can be found which tells us how small ϵ should be, or conversely, for a given ϵ how accurate our approximation will be. Despite this objection, the technique is still of value. Firstly, the technique does provide valuable insight into the likely structure of the optimal controls for problems which are otherwise insoluble. Secondly, there is always the hope that the coeffiecients of the power series in ϵ will suggest the form of the closed form solution of the problem. This hope is fulfilled in the text book application of the technique to the theory of differential equations. But what of the theory of stochastic optimal control?

The perturbation technique was applied to two simple problems of stochastic control - the classical information pattern problem and the zero-memory (Witsenhausen) problem. The purely mechanical application of the technique to both these problems was characterised by the diabolical complexity of the algebra which it generates. It is possible that a more thoughtful application of the technique can significantly reduce the algebraic load. This was demonstrated very dramatically for the classical information pattern problem. Failing all else, a computer

symbolic manipulation language can be utilised, although early experimentation in this field has shown that the algebra can be beyond the capacity of even the computer.

The perturbation analysis of our two sample problems has provided some valuable insight. The classical information pattern problem which we studied was identical to the traditional LQG problem except in one respect. In our version we made no assumption about the probability density function $p_0(x_0)$ of the initial state x_0. In the traditional problem, this must be Gaussian to ensure that the famous Kalman filtering equations can be utilised successfully, which in turn ensures that the analysis is fairly tractable. If no assumption is made about $p_0(x_0)$ the traditional approach is not applicable and we are forced to resort to approximation techniques. For arbitrary $p_0(x_0)$, our perturbation analysis reveals that, for an $o(\epsilon^2)$ - approximation to the optimal control, the effect of $p_0(x_0)$ can be ignored in all control laws after the first one (γ_0^o). It was also revealed that for an $o(\epsilon^2)$ - approximation, the optimal control law is a function of the present and immediately preceding observations only, despite the availability of all preceding observations. Thus, as a by-product, we obtained the $o(\epsilon^2)$ - approximation to a non-classical information pattern problem in which the control data consists only of the present and immediately preceding observation.

As an example of a zero memory problem we selected the Witsenhausen problem. This is a remarkable problem since it presents clearly

the contradictory nature of the first stage controller's task, which involves both communication and control. The $o(\epsilon^4)$ - approximation to the optimal control at the first stage exhibits components of control (0) and communication $\frac{(x\epsilon^4\sigma_v^4)}{k^2}$.

In applying the perturbation technique to the above examples we have perturbed the problem about $\epsilon = 0$, where ϵ is a small parameter which was introduced into the observation equation:-

$$y = x + \epsilon v$$

Perturbing about $\epsilon = 0$ means perturbing about the case of perfect observation. In other words, the generating solution is that for the perfect observation problem. But as we have seen in our discrete state example of Chapter 4, we should anticipate some discontinuity in moving from the perfect observation case to the noisy observation case. A method for avoiding this discontinuity which suggests itself is to perturb the problem about some other parameter so that the generating solution corresponds to some problem more closely related to the problem at hand. For example, can we model a system in which the generating solution corresponds to the classical information pattern problem.

A possibility is illustrated in the diagram shown overleaf.

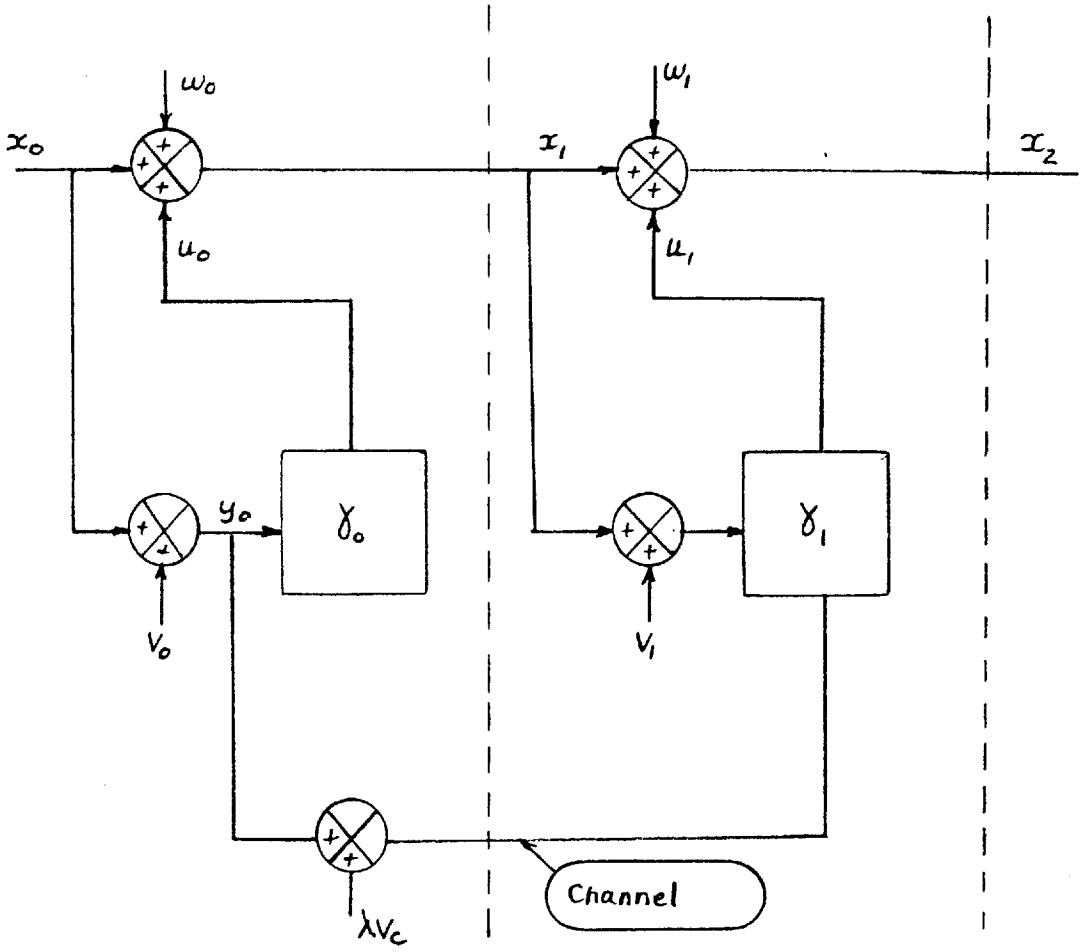


Fig. 6.1 Perturbation about Classical Information Pattern Problem

In the above model, $\lambda = 0$ corresponds to the Classical Information Pattern Problem. For $\lambda \neq 0$, we are modelling the performance of the communication channel, which is of considerable practical importance.

For $\lambda \gg 0$, we intuitively approach the zero-memory problem.

Extending this idea further, why not give the first stage controller some control over the communication channel? A possibility is illustrated on the next page.

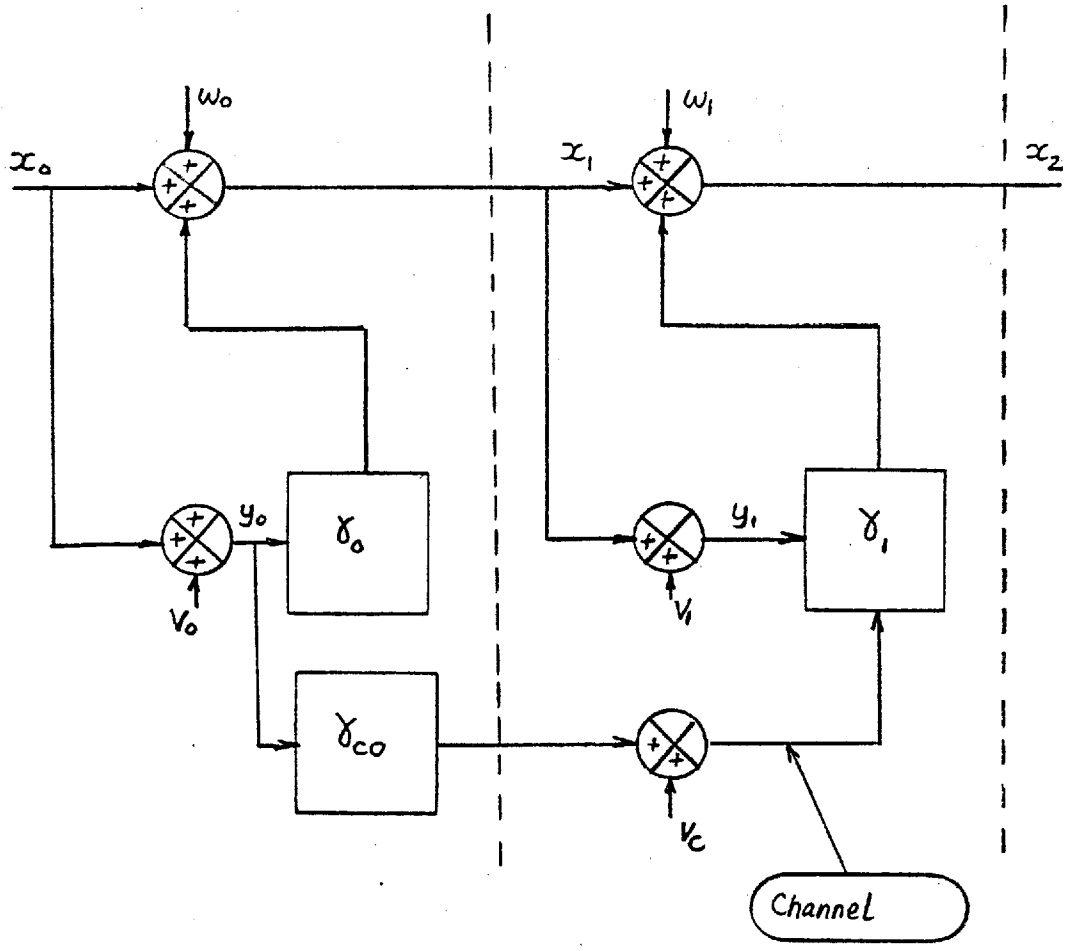


Fig 6:2 Separation of Communication and Control

Here the problem is to select γ_{c0} as well as γ_0 and γ_1, so that the performance of the system is optimised. The communication control γ_{c0} can be regarded as some coding algorithm to facilitate accurate data transmission. In addition a component such as $k' \gamma_{c0}^2(y_0)$ can be added to the cost function to model the cost of the communication link.

The systems which are proposed above are both examples of practical importance, modelling situations in which the control and communication aspects of the problem are salient features. It is hoped that their solution might provide more insight into the solution of non-classical information pattern control problems. It is not suggested however that the solutions of above systems will be easy to obtain - they will be highly non-linear in nature. Nor will they fit into the dynamic programming formulations derived earlier. But perhaps perturbation analysis will pay some dividends and give more clues as to the identity of the elusive solution to Witsenhausen's problem.

CHAPTER 7

Optimal Control of Stationary Systems

		Page
7.1	Introduction	7.1
7.2	The Infinite Horizon Zero-Memory Problem	7.6
	7.2.1 Statement of the Problem	7.7
	7.2.2 Discussion of the Problem	7.11
	7.2.3 Summary	7.47
7.3	Zero-Memory Optimal Linear Controllers	7.51
	7.3.1 Problem Statement	7.53
	7.3.2 State Propogation under Linear Control	7.55
	7.3.3 The Discount Problem	7.56
	7.3.4 The Average Cost per Period Problem	7.61
	7.3.5 The Ergodic Problem	7.63
	7.3.6 Summary	7.69
7.4	Further Examination of the Ergodic Problem	7.72
	7.4.1 How Good is the Linear Controller?	7.72
	7.4.2 Design for Linear Multi-Dimensional Systems	7.87
	7.4.3 Optimality of the Static Linear Control Law	7.93
	7.4.4 Approximation of Methods	7.94
	7.4.5 Summary	7.129

Cont/...

7.5	Further Examination of the Discount Problem	7.132
	7.5.1. Successive Approximation Method	7.132
	7.5.2 Asymptotic Behaviour of Solution	7.139
	7.5.3 Summary	7.141
7.6	Conclusions	7.141

7.1. Introduction

This chapter is concerned with what we have called the stationary problem. Since the word 'stationary' is used in a variety of contexts in the literature, we must be careful to define what the term means here. In the previous chapters we have considered various problems concerning the optimum operation of a stochastic control system over a finite number of stages. The stationary problem we shall examine here is similar to our earlier problems except for three important features:-

i) The number of stages is assumed to be infinite.

ii) The system equations and the random disturbance statistics do not change from one stage to the next. In other words, the system equations and noise processes are stationary.

iii) The function describing the cost per stage does not depend on time (except perhaps for a discount factor). Again, we can say that the cost per stage is stationary.

In practice, we never encounter problems with an infinite number of stages. Nevertheless the mathematical formulation of problems with an infinite number of stages does represent a reasonable approximation for practical problems involving a finite but very large number of stages. Also, the assumptions of stationarity are often satisfied in practice (or represent a good approximation to problems in which the system parameters vary very slowly as time progresses). That is to say, sometimes a reformulation of a non-stationary problem will yield a stationary version of the problem.

Having established that a (hypothetical) stationary problem can closely represent a practical problem, the next question to answer is why we should be interested anyway. Certainly the analysis is no more attractive and requires a level of sophistication beyond that demanded by the finite time counterparts. First, the stationary problem introduces a need for the analysis of the limiting behaviour of certain quantities. Secondly, a rigorous consideration of the probabilistic aspects of the problem require the intricacies of measure-theoretic probability theory. So, why bother?

The answer is that a number of benefits arise which do not appear for the finite time problem:-

i) The analysis produces occasionally very elegant and sometimes surprising results.

ii) The implementation of optimal control laws for infinite time problems is often much simpler. It turns out that in some cases the optimal policy is stationary - in other words the optimal control law does not change from one time instant to the next. Indeed, this property is sometimes included in the definition of a stationary problem.

iii) The optimal stationary control policies can be determined by a variety of very attractive techniques.

The above remarks all apply to the perfect observation and classical information pattern version of the stationary problem. For non-classical information patterns, nothing much is known.

In this chapter we propose to examine a particular example in some detail. This example will be a stationary version of the simple scalar linear, quadratic, gaussian zero-memory problem, the finite time version of which we have already explored in previous chapters. Our earlier work has demonstrated just how difficult to solve is the zero-memory problem. Moreover, we have seen that this difficulty is linked in part to the fact that we cannot make the crucial argument that the 'state' variable $p(x_n)$ at any time $n > 0$ is Gaussian. This leads to our motivation for studying the stationary problem. Given the tendency of sums of identically distributed and independent random variables to approach a gaussian distribution (the laws of large numbers) the thought arises that the state vectors for an infinite time problem might approach a Gaussian distribution. In this case it might then be possible to show that the optimal stationary control policy is linear and consequently relatively easy to determine and implement.

An extensive theory has developed for dealing with the optimal stationary control of both the perfect observation and classical information pattern versions of the stochastic control problem. This theory is reviewed in Appendix 5.1. No attempt is made to prove these results. Such proofs are available from a number of sources.[1] We do not even state the conditions under which these results apply, since we are ultimately interested in deriving a solution to a specific problem and generalities are not our immediate concern.

1. For example - D.P. Bertsekas - Dynamic Programming and Stochastic Control - Academic Press 1976. (8)

Similarly Appendix 5.2 describes how the general theory developed in Appendix 5.1 can be applied to the special case of a linear system with quadratic cost. Again, the treatment is purely descriptive - the intention being to provide background information for the content of this chapter.

In Section 2 we consider the zero-memory stationary problem. A number of theorems are introduced and proved. These are counterparts of corresponding theorems discussed in Appendix 5.1 for the perfect observation version of the problem.

In Section 3 we commence the task of attempting to find solutions to the zero-memory stationary problem. Even when we restrict our attention to the problem of finding the best control laws within the class of linear controllers, the results are disappointing. Only for one version of the problem can we find a convenient algorithm for the best linear control law. For the general stationary problem we encounter severe algebraic complexities and we are forced to relax our problem to that of finding what we will call static (as opposed to stationary) linear control laws.

In Section 4 we consider in detail a particular version of a stationary problem - what we have called here the ergodic problem. We shall first compare the performance of the best linear control law with the optimal control law for the corresponding problem with a classical in-

formation pattern. The experimental results presented indicate that the zero-memory controller performs very well. Next we extend the algorithm for the optimal control law for our simple scalar example to cater for a general multivariable linear system with quadratic cost. Finally we complete Section 4 by examining a number of techniques which are available for finding approximate solutions to the zero-memory problem.

In Section 5 we return to the problem of finding the best linear control laws for the versions of the stationary problem we were unable to solve in Section 3 - notably the discount problem. An approximate solution technique is proposed by using perturbation analysis. Since the solution is derived via the successive approximation method of dynamic programming, the results are interesting not only for the discount problem but also for the finite time zero-memory problem.

7.2 The Infinite Horizon Zero-Memory Problem

In Appendix 5 we discuss at length the traditional theory of infinite horizon stochastic optimal control problems. We consider the two classical problems - the discount problem, and the average cost per period problem. Our discussion is purely descriptive since a rigorous treatment of the topic would have required far more space than is feasible here. Nevertheless, even our descriptive treatment is quite lengthy, which perhaps indicates the inadequacy of a purely formal approach to the subject. When dealing with stationary problems, it appears that we must be extremely careful. It was not however the sole intention of Appendix 5 to merely make this single point. The intention was to provide the context in which the infinite horizon zero-memory problem can be discussed. The classical theory of Appendix 5 will be used as a basis for our treatment of the zero-memory problem. We shall point out those portions of the classical theory which can be adopted for the zero-memory problem and those portions which cannot be adopted, and why they cannot. We shall encounter a familiar difficulty - the requirement to minimise with respect to an entire function and not merely a function value. And from this difficulty stems a host of of associated difficulties.

The principal objective of this Section is to present a number of theorems, and their proofs, which are the zero-memory counterparts of those discussed in the preceding sections for the perfect memory problem. These theorems will provide the basis for our development in subsequent sections.

7.2.1 Statement of the Problem

The system we propose to examine is described by the following process equation:-

$$x_{n+1} = x_n + u_n + w_n \; ; \; n = 0,1,---,$$

where $\{w_n\}$ is a sequence of independent Gaussian random variables with zero mean and variance σ_w^2 for all n. The observation equation is as follows:-

$$y_n = x_n + v_n \; ; \; n = 0,1,---,$$

where v_n is a sequence of independent Gaussian random variables with zero mean and variance σ_v^2 for all n.

The variables x_n and y_n are scalar quantities. The probability distribution of the state variable x_0 at time 0 is Gaussian with mean zero and variance σ_0^2. It will be appreciated that we have selected as an example the simplest system which can be imagined. Notwithstanding this, we shall soon encounter some ugly looking algebra.

As usual for the zero-memory problem, we assume that the control value u_n at some time n is a function γ_n of the current observation y_n only:-

$$u_n = \gamma_n(y_n).$$

It is shown in Appendix 5 that there is great similarity between the solutions of a number of different formulations -

of the stationary problem, particularly for a linear system with quadratic cost, and gaussian noise. Accordingly we shall consider the following three versions of the zero-memory stationary problem.

(i) The Discount Problem

Suppose we define a cost function $V_0(\Gamma_0, p_0)$ as follows:-

$$V_0(\Gamma_0, p_0) \triangleq \lim_{N \to \infty} \; E_{\{w_k, v_k, x_0\}} \left[\sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k)) \right],$$

where Γ_0 is some admissable sequence of control functions $\{\gamma_k \mid k=0,1,2---$ and p_0 is some arbitrary probability density function for the state x_0 at time 0. The discount factor ρ satisfies $0 < \rho < 1$. Next suppose that:-

$$V_0^{\circ}(p_0) \triangleq \inf_{\Gamma_0} V_0(\Gamma_0, p_0); \text{ for all } p_0.$$

The problem is to find the optimal cost $V_0^{\circ}(p_0)$ and the corresponding optimal control law sequence $\Gamma_0^{\circ} = \{\gamma_k^{\circ} \mid k = 0,1,2,---\}$ at which a minimum (if any) is attained. In view of our discussion in Appendix 5.1 above, we should expect that the expectation operator and limit operator in the definition of $V_0(\Gamma_0, p_0)$ can be interchanged without undue difficulty. Moreover we might also expect that V_0° can be determined as the limit of the finite time version of the problem, and that a stationary control law will be optimal. We shall consider these points later.

ii) The Average Cost per Period Problem

In this case we define a cost function $H_0(\Gamma_0, p_0)$ as follows:-

$$H_0(\Gamma_0, p_0) = \lim_{N \to \infty} \; E_{\{w_k, v_k, x_0\}} \left[\frac{1}{N} \sum_{k=0}^{N-1} (x_k^2 + \gamma_k^2(y_k)) \right] .$$

The problem is to find the optimal average cost per period $H_0^{\circ}(p_0)$, as defined below:-

$$H_0^{\circ}(p_0) = \inf_{\Gamma_0} H_0(\Gamma_0, p_0),$$

together with the corresponding optimal control law sequence $\Gamma_0^{\circ} = \{\gamma_k^{\circ} \quad k = 0,1,2---\}$ at which the minimum if any is attained.

As for the discount problem, we should expect that the limit and expectation operators in the definition of $H_0(\Gamma_0, p_0)$ can be interchanged, and that solutions to the problem can be determined as the limit of the finite time version of the problem.

iii) The Ergodic Problem

This problem is so called because of its connection with the ergodic theorem of analysis. Suppose we restrict our attention from the beginning to stationary control law sequences. Then, given the 'state' at any

time n, we can compute the 'state' at time n+1 from the state transition equation, where here the term 'state' is used in the sense of Chapter 3 - the probability density function of the state x_n. Thus,

$$p_{n+1}(x_{n+1}) = \iint p(x_{n+1} \mid x_n, \gamma(y_n))\ p(y_n \mid x_n)\ p_n(x_n)\, dx_n\, dy_n$$

where γ is the stationary control law. Next, we further restrict our attention to those stationary control laws which lead to a stationary distribution $p_\infty(x)$. In other words $p_\infty(x)$ satisfies the asymptotic equation:-

$$p_\infty(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p_\infty(\theta)\ d\theta dy$$

Within the admissable class of control laws for γ, each γ gives rise to a certain p_∞. We can now define the cost functional $V(p_\infty, \gamma)$ as follows:-

$$V(p_\infty, \gamma) = E\left[x^2 + \gamma^2(y)\right]$$

$$= \iint (x^2 + \gamma^2(y))\ p(y \mid x)\ p_\infty(x)\ dydx\ .$$

The problem now is to find the cost $\bar{v}$ where:-

$$\bar{v} = \inf_{\gamma} \iint (x^2 + \gamma^2(y))\ p(y \mid x)\ p_\infty(x)\ dydx,$$

subject to the constraint:-

$$p_\infty(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p_\infty(\theta)\ dyd\theta$$

For this problem we should expect a strong similarity with the average cost per period problem defined previously.

Above we have presented three quite distinct formulations of a stationary zero-memory problem. In the next section we shall examine the connections between these formulations.

In presenting the above formulations we have omitted the demonstration that the cost functionals $V_0^{\circ}(p_0)$, $H_0(p_0)$, $V(p_\infty, \gamma)$ are in fact dependent on the parameters p_0 etc. as specified. To be complete, we should repeat the development of Chapter 3, carefully considering the evolution of the system from one time period to the next. Also, from the outset we have anticipated the role of the probability density function of the state $p_n(x_n)$ as the dynamic programming 'state variable' and embedded the problem in one involving an arbitrary initial state $p_0(x_0)$.

7.2.2 Discussion of the Problem

The results in Appendix 5 are all derived in the same way as limiting form of a finite time dynamic programming formulation. This formulation was used both as means of deriving necessary and sufficient conditions for optimality, and as a basis for the computational procedures. First, consider the discount problem.

7.2.2.1 The Discount Problem

When we turn to the zero memory problem, we can utilise the finite time formulation which was developed in Chapter 3 (Theorem 3.1.) Suppose,

that the cost for an N-stage zero memory version of the problem we have here is $\hat{V}_N(\Gamma_0^N, p_0)$, where:-

$$\hat{V}_N(\Gamma_0^N, p_0) = E\left\{\sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k))\right\},$$

and where $\Gamma_0^N = (\gamma_0, \gamma_1, \gamma_2 \cdots \gamma_{N-1})$

Suppose also that:-

$$\hat{V}_N^0(p_0) = \inf_{\Gamma_0^N} \hat{V}_N(\Gamma_0^N, p_0)$$

We can now trivially extend our finite-time zero memory dynamic programming formulation to include the discount factor, we obtain:-

Theorem 7.1

$$\text{Suppose } \hat{V}_n^0(p'_n) \triangleq \inf_{\Gamma_n^N} E_{p_n'} \left\{\sum_{k=N-n}^{N-1} \rho^{k-N+n} (x_k^2 + \gamma_k^2(y_k))\right\},$$

with $\hat{V}_0^0(p'_0) = 0$, then, for every p'_n, we have:-

$$\hat{V}_n^0(p'_n) = \inf_{\gamma_n} \left\{ E_{p_n'} (x_n^2 + \gamma_n^2(y_n)) + \rho \hat{V}_{n-1}^0(p'_{n-1}) \right\}, \quad n=1,2,\cdots N$$

where $p'_{n-1}(x_{n-1})$ satisfies the state transition equation:-

$$p'_{n-1}(x_{n-1}) = \iint p(x_{n-1} \mid x_n, \gamma_n(y_n)) p(y_n \mid x_n)\, p'_n(x_n)\, dx_n dy_n$$

Notice that we have changed our notation somewhat. The subscript n in the above formulation represents the 'time to go', and not the 'time already spent'. To avoid confusion we have adopted a special notation p'_n for the probability density function of the state with n time periods to go. Note that $p'_n = p_{N-n}$. In addition the notation $E_{p'_n}$ implies expectation with respect to the probability density function p'_n, and the primitive random variables w_{N-n} and v_{N-n}.

As before we can, at least conceptually, use the above formulation to find $\hat{V}_N^o(p_o)$. The question to be answered is: Does $\hat{V}_N^o(p_o)$ tend to $V_o^o(p_o)$ in some sense as $N \to \infty$? In other words, is it true that:-

$$\begin{aligned} V_\infty^o(p_o) &= \lim_{N\to\infty} \hat{V}_N^o(p_o) \\ &= \lim_{N\to\infty} \inf_{\Gamma_o^N} \hat{V}_N(p_o, \Gamma_o^N) \\ &= \inf_{\Gamma_o} \lim_{N\to\infty} \hat{V}_N(p_o, \Gamma_o^N) \\ &= V_o^o(p_o) \ ? \end{aligned}$$

But suppose that we did prove the above equality. It would be of little use because, as we have seen from our attempts to solve the Witsenhausen problem, obtaining a finite solution is impossible for the zero memory problem.

The next question concerns the zero memory analogue of the optimality equation. Intuitively, we would expect the functional $V_0^{\circ}(p)$ to satisfy the equation:-

$$V_0^{\circ}(p) = \inf_{\gamma} \left\{ E_p \left[x^2 + \gamma^2(y) \right] + \rho V_0^{\circ}(p') \right\} \text{ , for every } p$$

where $p'(x) = \iint p(x \mid \theta, \gamma(y)) p(y \mid \theta) p(\theta) \, d\theta dy$

Leaving aside the business of proving that the above functional equation is a necessary and sufficient condition for optimality, we must consider the value to us of the equation. Unlike the development of Appendix 5, the above equation has a state space of infinite dimension - the function space of p, and the minimising γ will be a functional of p. Admittedly, we are only interested in a particular p - the Gaussian density function. But we must ensure that any V_0° we propose as a solution of the optimality equation is a solution for every p, not a particular Gaussian p. It might turn out of course that consideration of p's other than Gaussian is not necessary, but to prove such a proposition would probably involve an extension of a finite time problem, which as we have already suggested is impossible anyway. So the prospect looks bleak.

Nevertheless, let us at least outline the proofs of the above results - namely the convergence of the dynamic programming algorithm, and the optimality equation.

Theorem 7.2 - The Optimality Equation

The optimal value function V_0° satisfies the equation:-

$$V_0^\circ(p) = \inf_{\gamma} \left\{ E_p(x^2 + \gamma^2(y)) + \rho V_0^\circ(p_1) \right\} \text{ , for all } p$$

where

$$p_1(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta) p(\theta)\ d\theta dy$$

(This is the state transition equation)

Proof

Suppose V_0° is the optimal value function. Then

$$V_0^\circ(p_0) = \inf_{\Gamma_0} V_0(p_0, \Gamma_0)$$

$$= \inf_{\Gamma_0} \lim_{N\to\infty} \sum_{k=0}^{N-1} \rho^k E\left[x_k^2 + \gamma_k^2(y_k)\right]$$

$$= \inf_{\Gamma_0} \left\{ \lim_{N\to\infty} \left\{ E_{p_0}(x_0^2 + \gamma_0^2(y_0)) + \sum_{k=1}^{N-1} \rho^k E_{p_1}\left(x_k^2 + \gamma_k^2(y_k)\right) \right\} \right\}$$

$$= \inf_{\gamma_0} \left\{ E_{p_0}(x_0^2 + \gamma_0^2(y_0)) + \rho \inf_{\Gamma_1} \lim_{N\to\infty} \sum_{i=0}^{N-2} \rho^i E_{p_1}(x_i^2 + \gamma_i^2(y_i)) \right\}$$

$$= \inf_{\gamma_0} \left\{ E_{p_0}(x_0^2 + \gamma_0^2(y_0)) + \rho V_0^0(p_1) \right\}$$

Thus V_0^0 satisfies the optimality equation. Above, we have used the notation E_p to denote the expectation with respect to the probability density function p, and the appropriate primitive random variables.

As it stands, the optimality equation is little more than a reformulation of the problem. We can add considerable more meat to our results by introducing the notion of a contraction mapping. A detailed discussion of this topic is postponed until the next Chapter. The following description will suffice for our present purposes.

Contraction Mapping

A mapping $T: B \to B$ is said to be contraction mapping if there exists a scalar $\rho < 1$ such that

$$\| T(V_1) - T(V_2) \| \leq \rho \| V_1 - V_2 \| \quad \text{for all } V_1, V_2 \in B$$

where $\| \quad \|$ denotes some appropriate norm function, and B is the corresponding Banach space,

Fixed Point Theorem

If T is a contraction mapping, then there exists a unique fixed point V_* such that:-

$$T(V_*) = V_*$$

Furthermore, if V_0 is any function and T^k is the composition T----T of T with itself k times, then:-

$$\lim_{k \to \infty} \left|\left| T^k(V_0) - V_* \right|\right| = 0$$

Notice that the functionals V referred to above must be elements of some Banach Space **B**. Thus, to apply the Fixed Point Theorem, we must first define our candidate Banach space together with its associated norm function. We must verify that:-

1) Our space is a linear space.
2) Our norm function satisfies the definition of a norm function.
3) The space is complete in the metric defined by the norm.

Suppose then we define **our** space B as the set of all functions V defined on the set S, where S is the space of all probability density functions p(x) which can be defined. Then it can be verified that B satisfies the axioms of a linear space.

Unfortunately, our definition of B does not permit the selection of a norm function which is consistent with (2) and (3) above and convenient for our purpose here. For example, the definition:-

$$\| V \| = \sup_{p \in S} | V(p) |$$

is of little use since if $V(p) = \int x^2 p(x)\,dx$, then V has infinite norm. Thus we must re-examine our definition of B. The following set-up was suggested by J.M.C. Clark:-

$$\text{Let } S_K = \left\{ p \;\middle|\; \int x^2 p(x)\,dx < K \right\},$$

$$B = \left\{ V \;\middle|\; |V(p)| \leqslant K_1 + K_2 \int x^2 p(x)\,dx \text{ for all } p \in S_K, \text{ some } K_1 K_2 \right\}$$

and $$\| V \| = \sup_{p \in S_K} |V(p)|$$

It can then be easily demonstrated that this norm function satisfies (2) and (3) above, and is clearly finite. Thus we are restricting both the elements of B and the set on which they are defined.

The elements p of S_K are our 'state' variables and they will be transformed according to our state transition equation:-

$$p'(x) = \int\int p(x \mid \theta, y, \gamma) p(y \mid \theta) p(\theta)\,d\theta\,dy,$$

or $$p' = \mathcal{L}_\gamma p.$$

Since we wish to consider only elements in S_K, we must therefore restrict the class of admissable functions so that this is assured. We define U_K as follows:-

$$U_k = \left\{ \gamma \mid p \in S_K \Rightarrow p' = \mathcal{L}_\gamma p \in S_K \right\}.$$

The restrictions we are imposing by the above definitions are not unnatural. The selection of an appropriate K can be determined by an a priori consideration of the parameters of the problem. For example, consider the following:-

Lemma

Suppose p is ϵ -optimal; that is if ,

$$J(\gamma) \overset{\Delta}{=} \lim_{N \to \infty} \frac{L}{N} E \sum_{i=1}^{N} (x_i^2 + \gamma^2(y_i)),$$

and $p^\epsilon = \mathcal{L}_{\gamma^\epsilon} p^\epsilon$

with $J(\gamma^\epsilon) \leqslant \inf_\gamma J(\gamma) + \epsilon$,

Then $\int x^2 p^\epsilon(x)dx < 3\sigma_v^2 + 2\sigma_w^2$ for sufficiently small ϵ .

Proof

We have $\int x^2 p^\epsilon(x)dx < E_{p^\epsilon}(x^2 + \gamma^2(y))$

$$< E_{\tilde{p}}(x^2+(-y)^2)$$ for sufficiently small ϵ ,

where $\tilde{p}$ is generated by $\gamma(y) = -y$ which implies $x_{n+1} = -v_n+w_n$.

Thus,

$$\int x^2\, p(x)\,dx = \sigma_v^2 + \sigma_w^2, \text{ and } \int (-y)^2\, p(x)\,dx = \sigma_v^2 + \sigma_w^2 + \sigma_v^2,$$

and

$$\int x^2\, p(x)\,dx < 3\sigma_v^2 + 2\sigma_v^2 \text{ for sufficiently small } \epsilon .$$

Thus we can take K in our definition to be $3\sigma_v^2 + 2\sigma_w^2$.
In what follows, the notation B_K, S_K, and U_K will be used without further comment.

Suppose we define our contraction mapping T as follows. For $p \in S_K$

$$T(V)(p) = \inf_{\gamma \in U_K} \left\{ E_p(x^2 + \gamma^2(y)) + \rho V(p') \right\},$$

where $p'(x) = \int\int p(x \mid \theta, \gamma(y))\, p(y \mid \theta)\, p(\theta)\, d\theta\, dy$

Given a functional V of p , TV is another functional of p as defined by the above equation. We must now show that T is a contraction mapping. Consider any two functionals V_1 and V_2:-

For any $p \in S_K$ and $\gamma \in U_K$, we have

$$\left\{ E_p(x^2 + \gamma^2(y)) + \rho V_1(p') \right\} = \left\{ E_p(x^2 + \gamma^2(y)) + \rho V_2(p') \right\} + \rho \left\{ V_1(p') - V_2(p') \right\}$$

$$\leq \left\{ E_p(x^2 + \gamma^2(y)) + \rho V_2(p') \right\} + \rho \sup_{p \in S_K} \left| V_1(p) - V_2(p) \right|$$

Taking the infimum of both sides over U_k, we obtain

$$TV_1(p) - TV_2(p) \leq \rho \sup_{p \in S_k} \left| V_1(p) - V_2(p) \right|$$

A similar argument shows that

$$TV_2(p) - TV_1(p) \leq \rho \sup_{p \in S_k} \left| V_1(p) - V_2(p) \right|$$

Thus

$$\left| TV_2(p) - TV_1(p) \right| \leq \rho \sup_{p \in S_k} \left| V_1(p) - V_2(p) \right| \quad \text{for every } p \in S_k,$$

and by taking the supremum over $p \in S_k$ of the left hand side, we obtain

$$\| TV_1 - TV_2 \| \leq \rho \| V_1 - V_2 \|$$

We have thus shown that T is a contraction mapping. This is good news, because it enables us to envoke the elegance of the Fixed Point Theorem to derive a zero memory analogue of the Successive Approximation Theorem (Convergence of the Dynamic Programming Algorithm), and to assert that the Optimality Equation has a unique solution. This is the content of the two theorems presented below.

First consider the Optimality Equation (Theorem 7.2). In terms of our contraction mapping T, it can be expressed as:-

$$V_0^0(p) = TV_0^0(p).$$

Invoking the Fixed Point Theorem, we can assert that there exists a unique solution to this equation, and the optimality equation becomes much more than a necessary condition. Thus we have the following:-

Theorem 7.3

The optimal value functional V_0° for the zero memory problem is the unique solution to the equation:-

$$V_0^\circ(p) = \inf_{\gamma \in U_K} \left\{ E_p(x^2 + \gamma^2(y)) + \rho V_0^\circ(p') \right\} \quad \text{for all } p \in S_K$$

where $p'(x) = \iint p(x \mid \theta, \gamma(y))\, p(y \mid x)\, p(\theta)\, d\theta\, dy$

Furthermore if γ° is a 'functional' of p such that $\gamma^\circ(p)$ attains the infimum in the right hand side of the optimality equation for each $p \in S_K$, then the stationary policy $\{\gamma^\circ\}$ is optimal. Conversely, if $\{\gamma^\circ, \gamma^\circ, \gamma^\circ \text{ ---}\}$ is optimal, then $\gamma^\circ(p)$ attains the infimum in the right hand side for all $p \in S_K$.

Proof

The first part has already been proved. To prove the second part, suppose that there is only one admissable control 'functional' γ^*. Then the stationary cost under this control is still given by the optimality equation, but with the inf missing because there is only one candidate to consider. If $\bar{V}(\gamma^*, p)$ is the stationary cost, then

$$\bar{V}(\gamma^*, p) = E_p(x^2 + \gamma^2(y)) + \rho\, \bar{V}(\gamma^*, p'),$$

which we can write as:-

$$\bar{V}(\gamma^*, p) = T_{\gamma^*} \bar{V}(\gamma^*, p).$$

This is just the zero-memory analogue of the stationarity equation encountered in **Appendix 5.1** The mapping T_{γ^*} can be shown to be a contraction mapping by the same device as used above; T_{γ^*} is simply a special case of T with the class of admissable controls restricted to the single 'functional' γ^*. Thus there is a unique solution $\bar{V}(\gamma^*, p)$ to the above stationarity equation.

Now if $\gamma^*(p)$ minimises the right hand side of the optimality equation, then

$$V_0^{\circ}(p) = \underset{p}{E}\,(x^2 + \gamma^{*2}(y)) + \rho V_0^{\circ}(p'))$$

But this is the stationarity equation and it has a unique solution.

Thus,

$V_0^{\circ}(p) = \bar{V}(\gamma^*, p)$ for every p,

and it follows that $\{\gamma^*\}$ is optimal. Conversely, suppose $\{\gamma^*\}$ is optimal, then

$$V_0^{\circ}(p) = \bar{V}(\gamma^*, p), \text{ and } \bar{V}(\gamma^*, p) = T_{\gamma^*} \bar{V}(\gamma^*, p).$$

Thus

$V_0^{\circ}(p) = T_{\gamma^*} V_0^{\circ}(p)$, which implies that $\gamma^*(p)$ attains the infimum in the optimality equation for each $p \in S_K$.

Next consider the convergence of the dynamic programming algorithm. In terms of our contraction mapping T above, we can state the dynamic programming algorithm of Theorem 7.1 as follows:-

Suppose $\hat{V}_n^{\circ}(p) = \inf_{\Gamma_n^N \in U_K} E_p\left\{\sum_{k=N-n}^{N-1} \rho^{k-N+n}(x_k^2 + \gamma_k^2(y))\right\}$

and $\hat{V}_0^{\circ}(p) = 0$ for all $p \in S_K$. Then

$$\hat{V}_n^{\circ}(p) = T\hat{V}_{n-1}^{\circ}(p), \text{ for } n=1,2,---N.$$

But T is a contraction mapping. Invoking the Fixed Point Theorem, we have:-

$$\lim_{N \to \infty} \left\| T^N V_0 - V_0^{\circ} \right\| = 0,$$

or alternatively,

$$\lim_{N \to \infty} \left\| \hat{V}_N^{\circ} - V_0^{\circ} \right\| = 0.$$

Thus we have:-

Theorem 7.4 - Convergence of the Dynamic Programming Algorithm

The optimal value functionals $\hat{V}_N^{\circ}$ for the finite time zero memory problem converge uniformly in p to the optimal value function V_0° for infinite horizon zero memory problem. That is, for any $\epsilon > 0$ there exists an $N(\epsilon)$ such that for $N > N(\epsilon)$, we have:-

$$\left| \inf_{\Gamma_0^N \in \mathcal{U}_K} E_p \left\{ \sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k)) - \inf_{\Gamma_0^\infty \in \mathcal{U}_K} \lim_{N\to\infty} E_p \left\{ \sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k)) \right\} \right| < \epsilon,$$

for each $p \in S_K$, that is :-

$$\lim_{N\to\infty} \inf_{\Gamma_0^N \in \mathcal{U}_K} E_p \left\{ \sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k)) \right\} = \inf_{\Gamma_0^\infty \in \mathcal{U}_K} \lim_{N\to\infty} E_p \left\{ \sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k)) \right\}$$

We have now derived zero memory analogues for the dynamic programming algorithm, the optimality equation, and the stationarity equation. Before we examine the average cost per period problem, can we find an equivalent of the policy iteration method? This is the context of the following Theorem.

Theorem 7.5 - The Policy Interation Method

Let $\{\gamma'\}$ be an admissable stationary policy and $\bar{V}(\gamma', p)$ be the corresponding value of the cost functional for some $p \in S_K$ Let $\bar{\gamma}(p) \in \mathcal{U}_K$ be a control 'functional' which satisfies:-

$$E_p(x^2 + \bar{\gamma}^2(y)) + \rho \bar{V}(\gamma', p'') = \min_{\gamma \in \mathcal{U}_K} \left\{ E_p(x^2 + \gamma^2(y)) + \rho \bar{V}(\gamma', p') \right\},$$

where $p'(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p(\theta)\ d\theta\ dy,$

and $p''(x) = \iint p(x \mid \theta, \bar{\gamma}(y))\ p(y \mid \theta)\ p(\theta)\ d\theta\ dy$

Then, if $\bar{V}(\bar{\gamma}, p)$ is the value of cost function corresponding to $\bar{\gamma}$ for some $p \in S_K$ we have

$$\bar{V}(\bar{\gamma}, p) \leqslant \bar{V}(\gamma', p) \text{ for all } p \in S_K$$

Furthermore, if the policy $\{\bar{\gamma}\}$ is not optimal, then strict inequality holds above for at least one 'state' $p \in S_K$.

Proof

From the proof of Theorem 7.3, we have:-

$$\bar{V}(\gamma', p) = \underset{p}{E}(x^2 + \gamma'^2(y)) + \rho \bar{V}(\gamma', p')$$

$$\geqslant \underset{p}{E}(x^2 + \bar{\gamma}^2(p)) + \rho \bar{V}(\gamma', p''), \text{ by hypothesis}$$

$$\triangleq T_{\bar{\gamma}}\ \bar{V}(\gamma', p)$$

But $T_{\bar{\gamma}}$ is a monotonic operator; that is:-

if $\bar{V}_1(\gamma, p) \leqslant \bar{V}_2(\gamma, p)$ for all γ, p

then $T_{\bar{\gamma}}\ \bar{V}_1(\gamma, p) \leqslant T_{\bar{\gamma}}\ \bar{V}_2(\gamma, p)$

This follows directly from the definition of $T_{\bar{\gamma}}$:-

$$\bar{V}_1(\gamma, p) \leq \bar{V}_2(\gamma, p) \text{ for all } \gamma, p$$

$$\Rightarrow \bar{V}_1(\gamma, p'') \leq \bar{V}_2(\gamma, p'') \text{ for all } \gamma, p'',$$

where $p''(x) = \int\int p(x \mid \theta, \gamma(y)) p(y \mid \theta)\ p(\theta)\ d\theta\ dy$

$$\Rightarrow \underset{p}{E}(x^2 + \gamma^2(y)) + \rho \bar{V}_1(\gamma, p'') \leq \underset{p}{E}(x^2 + \gamma^2(y)) + \rho \bar{V}_2(\gamma, p'')$$

$$\Rightarrow T_{\bar{\gamma}}\ \bar{V}_1(\gamma, p) \leq T_{\bar{\gamma}}\ \bar{V}_2(\gamma, p) \text{ for all } \gamma, p.$$

Returning to the policy iteration theorem, we have shown above that:-

$$\bar{V}(\gamma', p) \geq T_{\bar{\gamma}}\ \bar{V}(\gamma', p).$$

Thus by repeated application of $T_{\bar{\gamma}}$ to both sides of this inequality, we obtain:-

$$\bar{V}(\gamma', p) \geq T_{\bar{\gamma}}\ \bar{V}(\gamma', p) \geq T_{\bar{\gamma}}^2\ \bar{V}(\gamma', p) \geq \text{---} \geq \lim_{N \to \infty} T_{\bar{\gamma}}^N\ \bar{V}(\gamma', p) \quad \text{(a)}$$

As in Theorem 7.3, $T_{\bar{\gamma}}$ is a contraction mapping and we can invoke the Fixed Point Theorem to assert that :-

$$\lim_{N \to \infty} T_{\bar{\gamma}}^N\ \bar{V}(\gamma', p) = \bar{V}(\bar{\gamma}, p)$$

Thus $\bar{V}(\gamma', p) \geq \bar{V}(\bar{\gamma}, p)$ for all p.

Suppose $\bar{V}(\gamma', p) = \bar{V}(\bar{\gamma}, p)$ for all p. Then from the inequality (a)

above, we have $\bar{V}(\gamma', p) = T_{\gamma} \bar{V}(\gamma', p)$. And, by the hypothesis of the theorem:-

$$\bar{V}(\gamma', p) = T_{\gamma} \bar{V}(\gamma', p) = E(x^2 + \bar{\gamma}^2(y)) + \rho \bar{V}(\gamma', p'')$$

$$\leqslant E(x^2 + \gamma^2(y)) + \rho \bar{V}(\gamma', p') \text{ for all } \gamma$$

But this is simply the optimality equation. Thus $\bar{V}(\gamma', p) = V_0^{\circ}(p)$, and by Theorem 7.3 the stationary policy $\{\gamma'\}$ is optimal. Thus if some control law $\{\bar{\gamma}\}$ is not optimal, then we shall have:-

$$\bar{V}(\bar{\gamma}, p) < \bar{V}(\gamma, p) \text{ for some } p \in S_K.$$

The above theorem provides us with the basis for a (conceptual) policy iteration algorithm:-

(a) Guess an initial stationary policy $\Gamma_0^{\circ} = \{\gamma_0, \gamma_0, \gamma_0 ----\}$, where $\gamma_0 \in U_K$

(b) For k = 0,1,2,----, compute the corresponding cost $\bar{V}(\gamma_k, p)$.

This can be done by using the method of successive approximation; that is:-

$$\bar{V}(\gamma_k, p) = \lim_{N \to \infty} \bar{V}_N(\gamma_k, p),$$

where $\bar{V}_N(\gamma_k, p) \overset{\Delta}{=} E_p(x^2 + \gamma_k^2(y)) + \rho \bar{V}_{N-1}(\gamma_k, p')$

$$= T_{\gamma_k}^N \bar{V}_0(\gamma_k, p), \text{ where } \bar{V}_0(\gamma_k, p) \text{ is arbitrary.}$$

The convergence of $\bar{V}_N(\gamma_k, p)$ to $\bar{V}(\gamma_k, p)$ is assured by application of the Fixed Point Theorem, as in Theorem 7.5 above.

c) If $\bar{V}(\gamma_k, p) = \bar{V}(\gamma_{k-1}, p)$ for all $p \in S_K$ then stop. We have the optimal solution.

d) Otherwise, obtain a new stationary policy, γ_{k+1}, where γ_{k+1} satisfies:-

$$E_p(x^2 + \gamma_{k+1}^2(y)) + \rho \bar{V}(\gamma_k, p'') = \min_{\gamma \in U_K} \left\{ E_p(x^2 + \gamma^2(y)) + \rho \bar{V}(\gamma_k, p') \right\} \text{ for all } p$$

where $p''(x) = \iint p(x \mid \theta, \gamma_{k+1}(y))\, p(y \mid \theta)\, p(\theta)\, d\theta\, dy$

and $p'(x) = \iint p(x \mid \theta, \gamma(y))\, p(y \mid \theta)\, p(\theta)\, d\theta\, dy$

e) Return to step (b), and repeat the procedure.

This at last computes our preliminary discussion of the zero-memory discount problem. We have derived the traditional arrary of elegant theorems and procedures. But again it is important to point out that the results are of only academic interest, if that. This is because they all require the parameter p as a state variable. And p is any probability density contained in S_K. The control laws we are endeavouring to find will be functionals of p.(and functions of y). Finding such functionals of p. is, as we have said earlier, not an enticing prospect. We can further restrict the class of probability density functions to something convenient, such as the set of all gaussian probability density functions. Unfortunately such a restiction would seem to enforce a restriction on the class of admissable controllers

in order to preserve our Banach spaces. Alternatively we could perhaps reduce the dimensionality of the problem by some method of approximation. These topics will be examined later in Sections 3 and 4. Let us next consider the average cost per period problem.

7.2.2.2 The Average Cost per Period Problem

The discussion of Appendix 5.1 concerned with a finite Markov process, in which every state could be reached from every other state under some stationary control law. In the context of our zero memory problem, this would seem to imply that every possible p should be reachable from every other p. As we have seen in the preceding discussion of the discount problem, it may be possible to restrict the class of p which has to be considered to merely those p which are reachable from p_0 under the admissable class of control laws.

Consider the following theorem which provides a condition for the existence of a stationary optimal control which satisfies a certain optimality function.

Theorem 7.6

Assume that there exists a functional $V(p)$ and a constant λ such that :-

1) $$\lambda + V(p) = \inf_{\gamma \in U^p} E\,(x^2 + \gamma^2(y)) + V(p') \quad ,$$

where $$p'(x) = \iint p(x \mid \theta, \gamma(y))\; p(y \mid \theta)\; p(\theta)\; d\theta\; dy = \mathcal{L}_\gamma p(x)$$

2) $V(p_n)$ is uniformly bounded for any sequence $\{p_n\}$ generated by $p_{n+1} = \mathcal{L}_{\gamma_n} p_n$ and an admissable policy $(\gamma_0, \gamma_1, \text{---})$, where $\gamma_n \in U$.

Suppose the policy $\gamma^\circ(p)$ attains the minimum in (1). Then the policy $\{\gamma^\circ(p_0), \gamma^\circ(p_1), \gamma^\circ(p_2)\text{---}\}$, where $p_{n+1} = \mathcal{L}_{\gamma^\circ(p_n)} p_n$ is optimal, and the minimum cost is λ.

Comment

(i) The uniform boundedness condition (2) holds if,

$$U = \left\{ \gamma \mid 0 \leq V(\mathcal{L}_\gamma p) \leq \alpha V(p) + k, \text{ for all } p \right\} \text{ for some } \alpha: 0 \leq \alpha < 1 \text{ and } k > 0,$$

since, $V(p_r) \leq \alpha^r V(p_0) + \frac{k(1-\alpha^r)}{1-\alpha} \leq V(p_0) + \frac{k}{1-\alpha}$

(ii) The uniform boundedness condition (2) is necessary so that $\frac{V(p_N)}{N} \longrightarrow 0$ in the proof below.

(iii) The condition that $p \in S_K$ does not apply in this theorem.

Proof From the proposed optimality equation

$$+ V(p_0) \quad \underset{p_0}{E} (x^2 + \quad ^2(y)) + V(p_1)$$

By repeated application we have:-

$$\lambda + V(p_1) \leq \underset{p_1}{E} (x^2 + \quad ^2(y)) + V(p_2)$$

$$\lambda + V(p_2) \leq \underset{p_2}{E} (x^2 + \quad ^2(y)) + V(p_3)$$

$$\vdots$$

$$\lambda + V(p_{N-1}) \leq \underset{p_{N-1}}{E} (x^2 + \quad ^2(y)) + V(p_N)$$

Adding these inequalities, we obtain:-

$$\lambda \leqslant \frac{1}{N}\left\{\sum_{i=0}^{N-1} E_{p_i}(x^2+\gamma^2(y)) + V(p_N) - V(p_0)\right\}$$

$$= \frac{1}{N}\left\{\sum_{i=0}^{N-1} E_{p_0\{\omega_i\}\{v_i\}}(x^2+\gamma^2(y)) + \frac{V(p_N)-V(p_0)}{N}\right\}$$

with equality if $\gamma(p)=\gamma^{\circ}(p)$ for every p encountered. Taking the limit as $N\to\infty$, we obtain:-

$$\lambda \leqslant \lim_{N\to\infty}\left\{\frac{1}{N}\sum_{i=0}^{N-1} E_{p_0\{\omega_i\}\{v_i\}}(x^2+\gamma^2(y))\right\} = H_0(p_0,\Gamma_0)$$

for every p_0 and admissable policy Γ. Furthermore, equality holds when $\Gamma_0 = \{\gamma^{\circ}(p_0),\gamma^{\circ}(p_1),\gamma^{\circ}(p_2),\text{-----}\}$, and consequently

$$\lambda = H_0^{\circ}(p_0) = \inf_{\Gamma_0} H_0(p_0,\Gamma_0)$$

Notice that this theorem is of little use to us, since it offers us no assistance in either establishing that functions λ and $V(p)$

exist or in finding out what they are. Also notice that the average cost per period λ is independent (by hypothesis) of the initial state p_0.

We can find conditions which guarantee the existence of functions λ and $V(p)$ by establishing a connection between the average cost per period problem and the discount problem. Consider the discounted cost functional $J(\rho, p, \Gamma_0)$, where

$$J(\rho, p, \Gamma_0) = \lim_{N \to \infty} E_p \left\{ \sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y)) \right\} ,$$

and suppose that

$$J_0(\rho, p) = \inf_{\Gamma_0 \in \mathcal{U}} J(\rho, p, \Gamma_0)$$

Now we know from the preceding section (Theorem 7.2) that $J_0(\rho, p)$ is a solution to the optimality equation:-

$$J_0(\rho, p) = \inf_{\gamma \in \mathcal{U}} \left\{ E_p (x^2 + \gamma^2(y)) + \rho J_0(\rho, p') \right\} ,$$

where $p'(x) = \iint p(x \mid \theta, \gamma(y))\, p(y \mid \theta)\, p(\theta)\, d\theta\, dy$

Now let p_a be an arbitrary state, and define:-

$$\bar{V}(\rho, p) = J_0(\rho, p) - J_0(\rho, p_a) \text{ for every } p.$$

Then, by the optimality equation for the discount problem:-

$$\bar{V}(\rho,p) + J_0(\rho,p_a) = \inf_{\gamma \in \mathcal{U}} \left\{ E_p(x^2+\gamma^2(y)) + \rho(\bar{V}(\rho,p') + J_0(\rho,p_a)) \right\}$$

that is:-

$$(1-\rho)\, J_0(\rho,p_a) + \bar{V}(\rho,p) = \inf_{\gamma \in \mathcal{U}} \left\{ E_p(x^2+\gamma^2(y)) + \rho\, \bar{V}(\rho,p') \right\}$$

Suppose that $\lim_{\rho \to 1} (1-\rho)\, J_0(\rho,p_a) = \lambda$,

and $\lim_{\rho \to 1} \bar{V}(\rho,p) = V(p)$

Then $\lambda + V(p) = \inf_{\gamma \in \mathcal{U}} \left\{ E_p(x^2 + \gamma^2(y)) + V(p') \right\}$

In other words, the functions λ and $V(p)$ defined above are solutions of the optimality equation (Theorem 7.6). We can summarise the above analysis in the form of the following theorem:-

Theorem 7.7

Suppose that $J_0(\rho,p)$ is the optimal cost functional for the zero memory discount problem, and suppose that the following limits exist:-

$$\lambda = \lim_{\rho \to 1} (1-\rho)\, J_0(\rho,p_a) \,,$$

$$V(p) = \lim_{\rho \to 1} \left\{ J_0(\rho,p) - J_0(\rho,p_a) \right\} ,$$

where p_a is an arbitrary probability density function. Then the quantities λ and $V(p)$ satisfy the optimality equation of Theorem 7.6:-

$$\lambda + V(p) = \inf_{\gamma \in U_k} \left\{ E_p \left(x^2 + \gamma^2(y) \right) + V(p') \right\}$$

where $p' = \mathcal{L}_\gamma p$

Moreover, suppose that $V(p_n)$ is uniformly bounded for any sequence $\{p_n\}$ generated by $p_{n+1} = \mathcal{L}_{\gamma_n} p_n$ and an admissable policy,

$$(\gamma_0, \gamma_1, \text{---}), \quad \gamma_n \in U$$

and further suppose that the policy $\gamma^\circ(p)$ attains the minimum in the optimality equation.

Then the policy $\{\gamma^\circ(p_0), \gamma^\circ(p_1), \gamma^\circ(p_2)\text{---}\}$, where $p_{n+1} = \mathcal{L}_{\gamma^\circ(p_n)} p_n$, is optimal and the minimum average cost per period is λ.

The above analysis is simply telling us that if we can solve the zero-memory discount problem, then it is worth examining the limits

$$\lim_{\rho \to 1} (1-\rho) J_0(\rho, p_a)$$

and

$$\lim_{\rho \to 1} \left(J_0(\rho, p) - J_0(\rho, p_a) \right)$$

to see if they exist. If they do, and if the minimum is attained by γ° and if the uniform boundedness condition is satisfied, then we have a solution for the average cost per period problem.

We have seen above that the average cost per period problem can be treated as the limit (as $\rho \to 1$) of the discount problem. This observation provides the basis for a dynamic programming approach to the problem; which is described next. The analysis is formal in nature and is intended only to provide motivation for the technique - it is not a rigorous justification of the method.

In Section 2.2.1 above we introduced a contraction mapping T as follows:-

$$T(V(p)) \triangleq \inf_{\gamma \in U_K} \left\{ E_p(x^2 + \gamma^2)) + \rho V(p') \right\} , \quad p \in S_K$$

where $p'(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p(\theta)\ d\theta\ dy$

It is convenient to replace the symbol T by T_ρ to indicate a dependence on the discount factor ρ. In Theorem 7.4, it was shown that the optimal solution of the ρ-discounted problem can be obtained as the limit of the finite time problem. In other words, if $\hat{V}(p) = 0$ and

$$T_\rho^{k+1} \hat{V}(p) \triangleq \inf_{\gamma \in U_K} \left\{ E_p (x^2 + \gamma^2(y)) + \rho T_\rho^k \hat{V}(p') \right\}, \quad p \in S_K$$

with p' as defined above, then

$$\lim_{k\to\infty} T_\rho^k \hat{V}(p) = \inf_{\Gamma_0^{\ell l_k}} \lim_{N\to\infty} E_p \left\{ \sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k)) \right\}$$

$$\triangleq J_0(\rho, p)$$

It may turn out that $T_\rho^k \hat{V}(p) \to \infty$ as $\rho \to 1$ and $k \to \infty$, so let us define quantities that have finite limits as $\rho \to 1$ and $k \to \infty$. As before,

$$\bar{V}^k(\rho, p) \triangleq T_\rho^k \hat{V}(p) - T_\rho^k \hat{V}(p_a), \quad p, p_a \in S_K$$

for every p and some arbitrary p_a. We can now write our iteration expression above as

$$T_\rho^{k+1} \hat{V}(p) = \bar{V}^{k+1}(\rho, p) + T_\rho^{k+1} \hat{V}(p_a)$$

$$= \inf_{\gamma \in \ell l_K} \left\{ E_p (x^2 + \gamma^2(y)) + \rho T_\rho^k \hat{V}(p') \right\}$$

$$= \inf_{\gamma \in \ell l_K} \left\{ E_p (x^2 + \gamma^2(y)) + \rho (\bar{V}^k(p', \rho) + T_\rho^k \hat{V}(p_a)) \right\}$$

Thus:-

$$T_\rho^{k+1}\hat{V}(p_a) - \rho T_\rho^{k}\hat{V}(p_a) + \bar{V}^{k+1}(p,\rho) = \inf_{\gamma \in \Gamma_k}\left\{ E_p(x^2+\gamma^2(y)) + \rho \bar{V}^k(p',\rho)\right\}$$

Let us define the functional $H^k(p,\rho)$ as follows:-

$$H^k(p,\rho) \triangleq \bar{V}^k(p,\rho) + \left\{T_\rho^k\hat{V}(p_a) - \rho T_\rho^{k-1}\hat{V}(p_a)\right\}$$

Then, from our definition of $\bar{V}^k(p,\rho)$, we have:-

$$\bar{V}^k(p,\rho) = H^k(p,\rho) - H^k(p_a,\rho).$$

We can now write our iteration expression as:-

$$H^{k+1}(p,\rho) = \inf_{\gamma \in \Gamma_k}\left\{ E_p(x^2+\gamma^2(y)) + \rho \bar{V}^k(p',\rho)\right\} \quad , p \in S_k$$

with p' as defined earlier and $\bar{V}^k$ as above. It is worth reflecting that the above expression is nothing more than a slightly different form of the familiar dynamic programming formulation of the ρ-discounted problem (Theorem 7.4). All that we have done is change the variables somewhat by introducing an arbitrary state p_a. In the limit, according to Theorem 7.4, we shall have functions $V(p,\rho)$ and $H(p,\rho)$ via the relations:-

$$V(p,\rho) = \lim_{k\to\infty} \bar{V}^k(p,\rho) = \lim_{k\to\infty} \left\{ T_\rho^k \hat{V}(p) - T_\rho^k \hat{V}(p_a) \right\} = J_0(p,\rho) - J_0(p_a,\rho)$$

and:-

$$H(p,\rho) = \lim_{k\to\infty} H^k(p,\rho) = \lim_{k\to\infty} \left\{ \bar{V}^k(p,\rho) + T_\rho^k \hat{V}(p_a) - \rho T_\rho^{k-1} \hat{V}(p_a) \right\}$$

Thus:-

$$V(p,\rho) = J_0(p,\rho) - J_0(p_a,\rho),$$

and:-

$$H(p_a,\rho) = (1-\rho)\, J_0(p_a,\rho),$$

where J_0 is the optimal cost for the discounted cost problem. Let us formally take the limit as $\rho \to 1$ in our iteration expression above. If we define

$$V_*^k(p) \triangleq \lim_{\rho\to 1} \bar{V}^k(p,\rho),$$

and $H_*^k(p) = \lim_{\rho\to 1} H^k(p,\rho)$, then we obtain the algorithm:-

$$H_*^{k+1}(p) = \inf_{\gamma \in \mathcal{L}} \left\{ E_p(x^2 + \gamma^2(y)) + V_*^k(p) \right\}$$

$$V_*^k(p) = H_*^k(p) - H_*^k(p_a)$$

Suppose that, as $k \to \infty$, the above algorithm yields functions

$$H_*(p) = \lim_{k \to \infty} H_*^k(p),$$

and $$V_*(p) = \lim_{k \to \infty} V_*^k(p)$$

Now assume that the limits with respect to ∞ and k can be interchanged. Then we have:-

$$\lim_{\rho \to 1} V(p,\rho) = \lim_{\rho \to 1} \lim_{k \to \infty} \bar{V}^k(p,\rho) = \lim_{k \to \infty} \lim_{\rho \to 1} \bar{V}^k(p,\rho) = \lim_{k \to \infty} V_*^k(p) = V_*(p)$$

$$\lim_{\rho \to 1} H(p,\rho) = \lim_{\rho \to 1} \lim_{k \to \infty} H^k(p,\rho) = \lim_{k \to \infty} \lim_{\rho \to 1} H^k(p,\rho) = \lim_{k \to \infty} H_*^k(p) = H_*(p)$$

Thus we have:-

$$V_*(p) = \lim_{\rho \to 1} V(p,\rho) = \lim_{\rho \to 1} \left\{ J_0(p,\rho) - J_0(p_a,\rho) \right\}$$

and

$$H_*(p_a) = \lim_{\rho \to 1} H(p_a,\rho) = \lim_{\rho \to 1} (1-\rho) J_0(p_a,\rho)$$

But as we have seen earlier (Theorem 7.7), the quantities $H_*(p_a)$ and $V_*(p)$ are solutions of the optimality equation. We can conclude therefore that the iteration procedure derived above is the natural candidate as a successive approximation method for the average cost per period problem and, under suitable assumptions, should yield in the limit the optimal average cost per stage. Since we are interested here in a specific zero-memory problem, we shall not pursue this matter further since generalisations are not our immediate concern. The above discussion is merely background information. Let us next briefly examine what we have called the ergodic problem in Section 7.2.1.

7.2.2.3 The Ergodic Problem

Recall that the problem to be solved is formulated as follows:- Find the functional $V(p_\infty)$, such that:-

$$V(p_\infty) = \inf_{\gamma} \left\{ \underset{p_\infty}{E} (x^2 + \gamma^2(y)) \right\}$$

and where p_∞ satisfies the constraints

$$p_\infty(x) = \iint p(x \mid \theta, \gamma(y))\, p(y \mid \theta)\, p_\infty(\theta)\, d\theta\, dy$$

The connection between this problem and the average cost per period problem is fairly evident, at least formally.

Recall that for the average cost per period problem:-

$$H_0(\Gamma_0, p_0) = \lim_{N\to\infty} E_{\{w_k, v_k\}\, p_0}\left\{\frac{1}{N-1}\sum_{k=0}^{N-1}(x_k^2 + \gamma_k^2(y_k))\right\}$$

$$= \lim_{N\to\infty} \frac{1}{N-1}\left\{\sum_{k=0}^{N-1} E_{\{w_k, v_k\}\, p_0}(x_k^2 + \gamma_k^2(y_k))\right\}$$

where the probability density function p_k is generated via the 'state transition equation'. Suppose we restrict our attention to stationary control policies, $\bar{\Gamma}_0 = (\gamma, \gamma, \gamma)$, which depend only on y_k and are not functionals of p_k.

Then:-

$$H_0(\bar{\Gamma}_0, p_0) = \bar{H}_0(\gamma, p_0) = \lim_{N\to\infty} \frac{1}{N-1}\left\{\sum_{k=0}^{N-1} E_{p_k}(x_k^2 + \gamma_k^2(y_k))\right\}$$

$$= \lim_{k\to\infty} \left\{ \underset{p_k}{E} \left(x_k^2 + \gamma_k^2(y_k) \right) \right\}$$

$$= \underset{\bar{p}}{E} \left(x^2 + \gamma^2(y) \right)$$

where $\bar{p} = \lim_{k\to\infty} p_k$ in some sense.

But $p_{k+1}(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p_k(\theta)\ d\theta\ dy$,

because this is the state transition equation. Let us denote this equation as follows:-

$$p_{k+1} = \mathcal{L}_\gamma(p_k)$$

Under appropriate assumptions of continuity for $\mathcal{L}_\gamma$, we then have:-

$$\mathcal{L}_\gamma(\bar{p}) = \lim_{k\to\infty} \mathcal{L}_\gamma(p_k) = \lim_{k\to\infty} p_{k+1} = \bar{p}$$

Thus we have for the average cost per period problem with stationary control laws:-

$$\inf_\gamma \bar{H}_0(\gamma, p_0) = \inf_\gamma \left\{ \underset{\bar{p}}{E} \left(x^2 + \gamma^2(y) \right) \right\},$$

where $\bar{p}$ satisfies the equation:-

$$\bar{p} = \mathcal{L}_\gamma(\bar{p})$$

But this is precisely the ergodic problem as presented at the beginning of this section. In other words, the ergodic problem is

simply the average cost per period problem restricted to stationary control laws $\{\gamma\}$, which are functions of y_k and do not depend on p_k. In moving from the average cost per period problem to the ergodic problem, we appear to have lost our dependence of the initial probability density function p_0. It may well be that $\inf \bar{H}_0(\gamma, p_0)$ is independent of p_0 (of the irreducible Markov chain) - so that there was no dependence to loose anyway. Alternatively, it may turn out $\inf \bar{H}_0(\gamma, p_0)$ is dependent on p_0. In this case, we should qualify the ergodic problem somewhat. For an arbitrary γ, the state transition equation $p_\infty = L_\gamma(p_\infty)$ might have several solutions. However, we are only interested in that p_∞ which is reachable from p_0 under the control γ; that is, we insist that:-

$$p_\infty = \lim_{k\to\infty} p_k = \lim_{k\to\infty} \mathcal{L}_\gamma(p_{k-1}), \text{ with } p_0 \text{ given.}$$

With this qualification, which in an application is necessary anyway, it is now possible to move from the ergodic problem to the average cost per period problem (with stationary laws independent of p_k). In other words, the two problems are equivalent. To show that the ergodic problem is equivalent to the general non-stationary average cost per period problem for a particular system is therefore simply a matter of showing that the optimal control laws are stationary and independant of p_k for the average cost per period problem.

Some conditions which guarantee the stationarity of the optimal control laws for the average cost per period problem are presented in the form of a theorem below:-

Theorem 7.8

Suppose $\gamma^{\circ}, p^{\circ}$ are the minimising control and probability density for the ergodic problem:-

$$\lambda = \min_{\gamma} \mathrm{E}_{p} (x^2 + {}^2(y)^2), \text{ where } p = \mathcal{L}_{\gamma} p$$

If γ° has the following properties:-

i) γ° grows at most linearly; that is for some k_0, k_1 $|\gamma^{\circ}(y)| \leq k_0 + k_1|y|$

ii) γ° is stabilising; that is for some $\alpha_0 < 1$

$$\left| \mathrm{E}_{L_{\gamma^{\circ}} p} \left[x^2 + \gamma^{\circ}(y)^2 \right] - \lambda \right| \leq \alpha_0 \left| \mathrm{E}_{p} \left[x^2 + \gamma^{\circ}(y)^2 \right] - \lambda \right|$$

then for any initial p_0 with $\mathrm{E}_{p_0} x^2 < \infty$, the stationary policy - $(\gamma^{\circ}, \gamma^{\circ}, \text{---})$ minimises,

$$\lim_{N \to \infty} \frac{1}{N+1} \sum_{0}^{N} \mathrm{E}_{p_0 \Gamma} (x_n^2 + \gamma_n(y_n)^2)$$ in the class of all non-stationary policies $\Gamma = (\gamma_0, \gamma_1, \text{---})$ defined by

$$\Gamma = \left\{ (\gamma_0, \gamma_1, \text{---}) \middle| \exists\ k_0, k_1, k_2, k_3 \text{ such that } \gamma_n \leq k_0 + k_1|y| \right.,$$

$$\left. \text{and } \mathrm{E}_{p_0 \Gamma} (x_n^2 \middle| x_0 = a) \leq k_2 + k_3 a^2 \right\}$$

Proof Let $V(p) = \lim_{N\to\infty} \left\{ \sum_{0}^{N} E_{p,\gamma^\circ} \left[x_n^2 + {}^\circ(y_n)^2 \right] - (N+1)\lambda \right\}$

Then by condition (ii)

$$\left| E\left[x_{n+1}^2 + \gamma^\circ (y_{n+1})^2 \right] - \lambda \right| \leq \alpha_0 \left| E\left[x_n^2 + \gamma^\circ (y_n)^2 \right] - \lambda \right| ,$$

and so,

$$V(p) < (1 + \alpha_0 + \cdots + \alpha_0^n) \left| E_p \left[x_0^2 + \gamma^\circ (y_0)^2 \right] - \lambda \right|$$

$$\leq \frac{1}{1-\alpha_0} \left[E_p\left[x^2 (1 + 3k_1^2) \right] + 3k_1^2 \sigma_v^2 + 3k_0^2 + \lambda \right)$$

$$= k_5 + k_6 \, E_p(x^2)$$

Let $p_1, p_2, \cdots$ be generated by a non-stationary policy Γ with initial density p_0. Then $V(p_n) \leq k_5 + k_6 E\, x_n^2$, which is uniformly bounded in n, because of the restrictions on Γ and p_0. Thus condition (2) of Theorem 7.6 is satisfied.

Also, note that $V(p^\circ) = V(\mathcal{Z}_{\gamma^\circ} p^\circ) = 0$, which means that condition (1) of Theorem 7.6 is satisfied.

Thus the conditions of Theorem 7.6 are all satisfied and the Theorem is proved.

The formulation of the ergodic problem does have the advantage that it is conceptually simpler than the average cost per period problem, and it is generally the formulation which is adopted in the examples which follow. From a practical point of view, stationary control

laws are all that interest us anyway. This completes our discussion of the ergodic problem. Before examining some examples in Section 3, we shall conclude this section by summarising the main points.

7.2.3 Summary

In the preceding sections we have considered in fairly general terms three versions of the zero-memory infinite horizon problem - the discount problem, the average cost per period problem, and the ergodic problem (as we called it).

Our approach was based on the familiar notion of treating the probability density function $p_n(x_n)$ of the state x_n at time n as the state variable. It was possible to proceed with the analysis in terms of 'contraction operator' and invoke the Fixed Point Theorem to establish appropriate convergence and uniqueness properties. Various topics were discussed such as the optimality equation, the convergence of the dynamic programming algorithm, the policy iteration method, and so on. In addition, we demonstrated the traditional connections between the discount problem, the average cost per period problem, and the ergodic problem.

Although our development was concerned with the zero-memory problem, it is interesting to note that it could equally be applied to the classical information pattern case. All that is different is the state vector (it would be $p(x_n \mid I_n)$ and not $p(x_n)$ as here) and the corresponding optimality equation.

In our considerations of the various forms of the zero memory problem, we have derived some apparently elegant theorems. But it must be admitted that their practical value is, at best, dubious. The difficulty is as usual due to the infinite dimension of the state vector p_n. We are required to find various functionals which satisfy some functional equation for every p_n. This means every possible probability density function p_n! Although we can restrict this class to only those probability density functions which are **reachable via the state transition equation** from the initial probability density function p_0, we still have a very broad (and mysterious) class of probability density functions to consider. Only if we restrict the class of admissable control laws does the problem become manageable. For example, if we stipulate that the control law is a linear function of the observation, and if p_0 is Gaussian, then we know that p_n will be Gaussian for all n. In this case we can then reduce our problem to one involving the mean and variance of p_n and the difficulty of infinite dimension is resolved. This will be considered in the next section.

We shall conclude this section on a cautionary note. In formal derivations of optimal control policies (particularly for deterministic systems), it is usual to employ a dynamic programming approach, working backwards from time N. The iteration is repeated until the cost and controls settle down to a 'stationary' (or steady state) value. For a pro-

blem with a classical information pattern, therefore we would iterate backwards from time N to time 0, and obtain a control law at time 0 which is a function of the current observation y_0 only. There is a temptation, particularly for the infinite horizon problem, to believe that the problem facing a zero memory controller at time 0 is identical to that facing a controller with perfect memory at time 0. Hence it is assumed that the optimal stationary control law for the zero memory problem is obtained by finding the optimal stationary control law for the classical information pattern problem. As we have seen for the finite time problem this assumption is false. Some policy γ_0 is optimal, not in isolation, but because the structure of future control laws is assumed. Thus for the classical information pattern problem, $\gamma_0(y_0)$ is optimal only when it is known that $\gamma_1(y_0,y_1,u_0)$, $\gamma_2(y_0,y_1,y_2,u_0,u_1)$, and so on will be applied. But this will not be the case for the zero-memory problem and hence $\gamma_0(y_0)$ will not be optimal for the zero-memory problem. This discussion also raises the question of what a stationary policy is in the context of a classical information pattern problem. Clearly it can never be anything like $\gamma_0, \gamma_1, \gamma_2$ as above because these are always mappings from different product spaces. For the classical information pattern problem, a stationary policy is one which with respect to the state variable $p(x_n \mid I_n)$. Thus we have a stationary sequence of control policies $\{\bar{\gamma}\}$, where

$$u_0 = \bar{\gamma}(p(x_0 \mid I_0))$$

$$u_1 = \bar{\gamma}(p(x_1 \mid I_1)),$$

and so on.

As we suggested earlier in this chapter, the stationary problems which we have considered above do produce some elegant results in terms of optimality equations, successive approximation methods, and so on. Unfortunately, these results are in the same class as the dynamic programming formulation developed in Chapter 3, that is, they are virtually useless.

In the next section, we shall discuss the problem of finding optimal linear controllers for our three versions of the stationary zero memory problem. In preceding sections we have hinted that this restriction to linear controllers will produce a tractable problem. Alas, we shall find that only in special cases is this true.

7.3 Zero-Memory Optimal Linear Controllers

In this section we shall consider a linear, quadratic, gaussian problem in which the control law at each instant is constrained to be a linear function of the current observation. If a solution can be found, the ease with which such a control law can be implemented makes this class of problem of great practical significance in engineering applications. Unfortunately, we shall discover that the determination of mere linear controllers is not necessarily an easy task. In fact, we must be very careful to define what we mean by a linear controller. As mentioned above, we require first of all that the control law is a linear function of the current observation; that is:-

$$u = \gamma(y) = b \cdot y$$

To be exact, we should refer to this class of control law as 'proportional', and restrict the term 'linear' to refer to control laws of the form:-

$$u = \gamma(y) = a + b\,y,$$

where a is a constant. This structure of controller is sometimes called 'affine'. However, it is usually fairly easy to demonstrate that the value of a is zero and hence the problem of finding the best affine controller reduces to that of finding the best proportional controller. In what follows we shall use the term linear somewhat loosely to refer to both proportional controllers and affine controllers. In fact we shall usually omit the analysis which demonstrates that the optimal a is always zero.

So far however we have only mentioned the structure of the control law with respect to the current observation. But in the preceding section we developed a number of results which show that the optimal controller will be some functional of the current state, the probability density function $p_n(x_n)$. What of the structure of the controller with respect to $p_n(x_n)$, or alternatively of the structure of b with respect to $p_n(x_n)$? The description of linear controller is not adequate to cope with this. We shall be interested in two types of stationary linear control law. The first one,

$$u_n = \gamma(y_n, p_n) = b\,(F\,(p_n))\,.\,y_n = B_n\,y_n$$

will be called stationary linear control and is characterised by the fact that the proportional term b is a 'stationary' functional of p_n. The second one,

$$u_n = \gamma(y_n) = b.y$$

will be called static linear control and is characterised by the fact that the proportional term b is independent of the current state p_n. Static linear controllers are much easier to find than stationary linear controllers. The question is: are they as good?

Notice that according to our definition of stationary control, the constants of proportionality are not themselves stationary - B_n actually depends on n. Thus in the next section, in the statement of the problems which we shall consider, the structure of the control laws is assumed to

be,

$$u_n = \gamma(y_n, p_n) = a_n + b_n y_n$$

Only in the ergodic problem has this non-stationarity been lifted, because p_n is by definition a constant anyway.

7.3.1 Problem Statement

For ease of reference we shall repeat the statement of the problem previously presented in Section 4.1. Our system is composed of the process equation:-

$$x_{k+1} = x_k + \gamma_k(y) + w_k, \quad k = 0,1,2,---,$$

together with an observation equation:-

$$y_k = x_k + v_k, \qquad k = 0,1,2,----,$$

where x_k and y_k are scalars, $\{w_k\}$ and $\{v_k\}$ are sequences of independent gaussian random variables with zero mean and variance σ_w^2 and σ_v^2 respectively. The state variable x_0 is a gaussian random variable with zero mean and variance σ_0^2.

As a demonstration of the differences between the discount problem, the average cost per period problem, and the ergodic problem we shall consider all three infinite horizon problems:-

i) The Ergodic Problem

The problem here is to determine constants a and b, such that for a control law:-

$$\gamma(y_k) = a + by_k \quad \text{for all } k,$$

the following cost functional is minimised:-

$$V(a,b,p_0) = \lim_{k \to \infty} E(x_k^2 + \gamma^2(y_k))$$

ii) The Discount Problem

The problem here is to determine sequences of scalars $\{a_n\}$ and $\{b_n\}$ such that for a control law:-

$$\gamma_k(y_k) = a_k + b_k y_k \quad \text{for all } k,$$

the following cost functional is minimised:-

$$V_\rho(\{a_n\},\{b_n\},p_0) = \lim_{N \to \infty} E\left\{\sum_{k=0}^{N-1} \rho^k (x_k^2 + \gamma_k^2(y_k))\right\}.$$

iii) The Average Cost per Period Problem

The problem here is to determine sequences of scalars $\{a_n\}$ and $\{b_n\}$ such that for a control law:-

$$\gamma_k(y) = a_k + b_k y_k \quad \text{for all } k$$

the following cost function is minimised:-

$$H_0(\{a_n\},\{b_n\},p_0) = \lim_{N \to \infty} E\left\{\frac{1}{N-1} \sum_{k=0}^{N-1} (x_k^2 + \gamma_k^2(y_k))\right\}$$

The key to the solution of all the above problems is the behaviour of the state transition equation under linear control. This is considered next.

7.3.2 State Propogation under Linear Control

Consider a control law of the form:-

$$\gamma_k(y_k) = a_k + b_k y_k,$$

and assume that the probability density function $p_k(x_k)$ of the state x_k at time k is gaussian with mean $\bar{x}_k$ and variance σ_k^2. The 'state' at time k+1 is given by the state transition equation:-

$$p_{k+1}(x_{k+1}) = \iint p(x_{k+1} \mid x_k, \gamma_k(y_k))\; p(y_k \mid x_k)\; p_k(x_k)\; dx_k\, dy_k$$

$$= \iint \frac{\phi}{\sigma_w}\left\{\frac{x_{k+1} - (x_k + a_k + b_k y_k)}{\sigma_w}\right\} \frac{\phi}{\sigma_v}\left\{\frac{y_k - x_k}{\sigma_v}\right\} \frac{\phi}{\sigma_k}\left\{\frac{x_k - \bar{x}_k}{\sigma_k}\right\} dx_k\, dy_k$$

By the familiar device of completing the square, we could determine $p_{k+1}(x_{k+1})$ from the above expression. It is easier however to use the well known result that if X has a gaussian distribution with mean $\bar{X}$ and variance σ_X^2, and Y has a gaussian distribution with mean $\bar{Y}$ and variance σ_Y^2, then X + Y has a gaussian distribution with mean $\bar{X} + \bar{Y}$ and variance $\sigma_X^2 + \sigma_Y^2$.

(See for example Kingman and Taylor Page 324)[1]. But

$$x_{k+1} = x_k + a_k + b_k y_k + w_k,$$

1. Kingman and Taylor - Introduction to Meausre and Probability 1966 (34)

that is, $x_{k+1} = x_k(1 + b_k) + b_k v_k + w_k + a_k$

Thus x_{k+1} is the sum of three independent gaussian random variables, and repeated application of the above result yields:-

$$P(x_{k+1}) = \frac{\Phi}{\sigma_{k+1}} \left\{ \frac{x_{k+1} - \bar{x}_{k+1}}{\sigma_{k+1}} \right\} ,$$

where $\bar{x}_{k+1} = (1 + b_k)\, \bar{x}_k + a_k ,$

and $\sigma_{k+1}^2 = (1 + b_k)^2\, \sigma_k^2 + b_k^2\, \sigma_v^2 + \sigma_w^2$

Thus, if p_0 is gaussian then under linear control, p_1 is gaussian. Similarly if p_1 is gaussian, under linear control p_2 is gaussian, and so on. Thus for our system here $p_k(x_k)$ is gaussian for all k, with mean and variance defined by the above recursive relations.

7.3.3 The Discount Problem

The general discount problem was discussed earlier in Section 2.2.1 In particular, we have the optimality equation Theorem 7.2,

$$V_\rho^o(p) = \inf_{a,b} \left\{ E(x^2 + (a + by)^2) + \rho V_\rho^o(p') \right\} ,$$

where $p'(x) = \iint p(x \mid \theta, y)\, p(y \mid \theta)\, p(\theta)\, d\theta\, dy$

In addition, it was shown that the finite horizon dynamic programming formulation will converge to the infinite horizon problem (Theorem 7.4). Unfortunately this is of little use because, even when the finite horizon problem is restricted to linear controls, it is not possible to derive an algorithm which provides the optimal control at each time instant in a convenient form. Thus we cannot derive the optimal controls for the finite horizon discount problem via the corresponding finite time version.

It is worthwhile to note a dangerous conceptual pitfall concerned with the infinite horizon discount problem. Suppose we adopt the optimistic attitude that the zero memory problem is very similar to the classical information pattern problem and that consequently we propose the following structure for a functional $\hat{V}(p_1)$

$$\hat{V}(p_1) = c\,E(x_1^2) + d\sigma_v^2 + e\sigma_w^2,$$

where $c = (1+\beta^2) + \rho c(1+\beta)^2$

$$d = \beta^2 + \rho(c\beta^2 + d)$$

$$e = \rho(c+e)$$

and $$\beta = \frac{-\rho c}{1+\rho c}\,\frac{E(x_1^2)}{E(x_1^2)+\sigma_v^2}$$

jointly determine the multipliers c,d,e and the parameter β.

Let us now examine if $V_\rho^\circ(p)$ satisfies our (linear) optimality equation:-

$$\text{Let} \quad \overline{V}(p_o) = \inf_b \left\{ E(x_o^2 + b^2 y_o^2) + \rho \left(c\,E(x_1^2) + d\sigma_v^2 + e\sigma_w^2 \right) \right\}$$

$$= \inf_b \left\{ E(x_o^2)(1+b^2) + b^2\sigma_v^2 + \rho\left(c\,E(x_o^2(1+b^2) + cb^2\sigma_v^2 \right. \right.$$
$$\left. \left. + c\sigma_w^2 + d\sigma_v^2 + e\sigma_w^2 \right\} \right.$$

$$= \inf_b \left\{ E(x_o^2)\left[(1+b)^2 + \rho c(1+b)^2\right] + \sigma_v^2\left[b^2 + \rho(cb^2+d)\right] \right.$$
$$\left. + \rho(c+e)\sigma_w^2 \right\}$$

Thus expression is minimised with respect to b when:-

$$b = b^\circ = -\frac{\rho c}{1+\rho c} \cdot \frac{E(x_0^2)}{E(x_0^2) + \sigma_v^2} \quad ,$$

and the expression above for $\overline{V}(p)$ becomes:-

$$\overline{V}(p_0) = \left[(1+b^{\circ 2}) + \rho c(1+b^\circ)^2 \right] E(x_0^2) + \left[b^{\circ 2} + \rho(cb^{\circ 2} + d) \right] \sigma_v^2 + \rho(c+e)\sigma_w^2$$

In view of the identical structures of b° and β there is now a strong temptation to assert that $\bar{V}(p_0) = \hat{V}(p_0)$, as desired. This is wrong. First, $b^{\circ} \neq \beta$ since the former is a function of $E(x_0^2)$ and the latter a function of $E(x_1^2)$. Secondly, the multiplier c is not a constant - it is an implicit function of $E(x_1^2)$ via the dependence of β on $E(x_1^2)$. Thus, we should write for $\hat{V}(p_1)$:-

$$\hat{V}(p_1) = c(p_1)\, E(x_1^2) + d(p_1)\, \sigma_v^2 + e(p_1)\, \sigma_w^2$$

and in evaluating $\hat{V}(p_1)$ we should not treat c as a constant - we must show the dependence of c on b and $E(x_0^2)$. The assertion that $\bar{V}(p_0) = \hat{V}(p_0)$ and the conclusion that $\hat{V}(p_0)$ is a solution of the optimality equation is only true if either $E(x_0^2) = E(x_1^2)$ or $\sigma_v^2 = 0$. In the first case we have a stationary distribution; and in the second we have a perfectly observable problem. In either case, $b^{\circ} = \beta$ and c can be regarded as a constant.

For the moment we must abandon the problem of finding the optimal stationary linear controls for the discount problem. We shall return to the problem later and attempt to find an approximation method (Section 5). Let us next consider the problem of finding what we have called static controllers. At last we have a problem which is fairly straightforward. We assume that

$$u = \gamma(y,p) = b\, y$$

Under this form of control law, the infinite horizon discount cost can easily be determined in terms of b, σ_0^2, σ_v^2, and σ_w^2.

It is simply a matter of geometric series. We obtain

$$\bar{V}(\rho,b) = \lim_{N\to\infty} E\left\{\sum_{k=0}^{N} \rho^k (x_k^2 + \gamma_k^2 (y_k))\right\}$$

$$= \frac{b^2\sigma_v^2}{(1-\rho)} + \frac{(1+b^2)\sigma_0^2}{1-\rho(1+b)^2} + \frac{(1+b^2)(b^2\sigma_v^2+\sigma_w^2)\rho}{(1-(1+b)^2)(1-\rho)}$$

$$- \frac{(1+b^2)(b^2\sigma_v^2+\sigma_w^2)\rho(1+b^2)}{(1-(1+b)^2)(1-\rho(1+b)^2)}$$

It is now a simple matter to determine that b which minimises $\bar{V}(\rho, b)$. We merely solve the equation:-

$$\frac{d\bar{V}(\rho,b)}{db} = 0$$

The algebra is tedious but straightforward. Moreover, if necessary the analysis can easily be extended to include the general multi-dimensional linear system

$$x_{n+1} = Ax_n + Bu_n + w_n$$

$$y_n = Cx_n + v_n$$

with cost function

$$J(\mathbf{B}, \sigma_0^2) = \lim_{N\to\infty} E\left\{\sum_{k=0}^{N_1} \rho^k (xQx + yRy)\right\} ,$$

where the matrices above all have the necessary properties to guarantee convergence of the infinite sums. This matter will not be pursued here. Let us next consider the average cost per period problem.

7.3.4 The Average Cost per Period Problem

The general form of this problem was discussed earlier in Section 2.2.2 where an optimality equation was proposed. In this section we are interested in finding solutions λ ,$V^{o}(p)$ to the following equation:-

$$\lambda + V^{o}(p) = \inf_{a,b} \left\{ E(x^2 + (a + by)^2) + V^{o}(p') \right\} ,$$

where

$$p'(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p(\theta)\ d\theta\ dy$$

Finding solutions to the above equation, even for the case of linear control, is beset with the same difficulties as those encountered for the discount problem considered in the previous section. Again we must resort to finding static controllers. For an N-stage process under some static linear control law, we have:-

$$J(b) = E\left\{ \sum_{k=0}^{N} (x_k^2 + \gamma_k^2(y_k)) \right\}$$

$$= (N+1)b^2\sigma_v^2 + \frac{(1+b^2)\sigma_o^2(1-(1+b)^{2N+2}}{1-(1+b)^2} + \frac{(1+b^2)(b^2\sigma_v^2+\sigma_w^2)N}{1-(1+b)^2}$$

$$- \frac{(1+b^2)(b^2\sigma_v^2+\sigma_w^2)(1+b^2)(1-(1+b)^{2N})}{(1-(1+b)^2)^2}$$

If we assume that b must yield a stable control law, then $(1+b)^n \to 0$ as $n \to \infty$. Under this condition:-

$$\lim_{N \to \infty} \frac{J(b)}{N} = b^2\sigma_v^2 + \frac{(1+b^2)(b^2\sigma_v^2+\sigma_w^2)}{1-(1+b)^2}$$

Notice that an identical expression is obtained, as we would expect, by $\lim_{\rho \to 1} (1-\rho)\,\bar{V}(\rho,b)$, where $\bar{V}(\rho,b)$ was defined as the discounted cost of the same process in the previous section.

As for the discount problem we can differentiate $J(b)/N$ with respect to b to obtain a necessary condition for the optimal value of b. Although the algebra will become tedious, we can also extend the method to derive optimal static linear controllers for multi-dimensional systems.

So far our attempts at finding optimal linear controllers for the discount problem and the average cost per period problem have been disappointing. All that we have been able to do is find static controllers. We shall consider approximation methods for optimal linear stationary

controllers in Section 5. Let us next examine the ergodic problem - a problem which we have seen is closely associated with the average cost per period problem.

7.3.5 The Ergodic Problem

For this problem we should expect the distinction between stationary control and static control to become irrelevant. This is simply because by definition, the state of the system $p_n(x_n)$ becomes a constant. Thus:-

$$u_n = \gamma(p_n, y_n) = b(p_n)\, y_n = b(p_\infty) y_n = B\, y_n$$

In other words the proportional term is a constant with respect to time. There are two methods of approach to the problem, we can view the ergodic problem as a special case of the average cost per period problem, or we can view the ergodic problem directly as a constrained optimisation problem. Recall that the connection between the average cost per period problem and the ergodic problem was discussed in Section 2.2.3. This will not be pursued here. Instead we shall concentrate upon the constrained optimisation approach. This method is preferred because in many senses it is conceptually easier than viewing the ergodic problem as a special case of the average cost per period problem. For example, Theorem 7.6 could be extended quite trivially to deal with those p which are stationary and those γ which maintain p stationary. But this would amount to little more than a slight reformulation of the problem. And it would obscure the important point that in the ergodic problem we are optimising not merely with respect to γ but also with respect to p, the two quantities being linked by the constraint

imposed by the state transition equation. The idea will be exploited more fully when we examine non-linear controllers in Section 4.

As before we shall dispense with the problem of finding the constant a in the linear control law:-

$$u = \gamma(y) = a + By$$

It is easily shown that the optimal a is zero. It must be zero to guarantee stability, anyway. Thus we shall restrict ourselves to finding the optimal B. Let us formulate as a constrained optimisation problem. Under, linear control starting with $p_0(x_0)$ Gaussian, we know that all future $p_n(x_n)$ will be Gaussian. We can argue that the mean of $p_n(x_n)$ will be zero and let us denote its variance (second moment) as m_2.

Then we are required to minimise J, where

$$J = E(x^2 + B^2 y^2) = (1+B^2)\, m_2 + B^2\, \sigma_v^2$$

subject to the constraint:-

$$m_2 = (1+B)^2 m_2 + B^2\, \sigma_v^2 + \sigma_w^2$$

Note that for a Gaussian $p_n(x_n)$, a second moment constraint is all that is necessary for stationarity. Using the technique of Lagrange multipliers, we have

$$J^a = (1+B^2)m_2 + B^2\, \sigma_v^2 + \lambda\left\{(1+B)^2\, m_2 + B^2\, \sigma_v^2 + \sigma_w^2 - m_2\right\}$$

We wish to minimise this expression with respect to m_2 and B, and to satisfy the constraint between m_2 and B. We have the necessary conditions:-

$$\frac{\partial J^a}{\partial B} = 2B m_2 + 2B\sigma_v^2 + 2\lambda\left[(1+B)m_2 + B\sigma_v^2\right] = 0$$

$$\frac{\partial J^a}{\partial m_2} = (1+B^2) + \lambda(1+B)^2 - \lambda = 0$$

$$\frac{\partial J^a}{\partial \lambda} = (1+B)^2 m_2 + B^2\sigma_v^2 + \sigma_w^2 - m_2 = 0$$

Rearranging the first equation, we then have to solve the set of equations:-

$$(1+B^2) + \lambda(1+B)^2 = \lambda \qquad \text{—— (i)}$$

$$(1+B)^2 m_2 + B^2\sigma_v^2 + \sigma_w^2 = m_2 \qquad \text{—— (ii)}$$

$$B = \frac{-\lambda}{1+\lambda}\,\frac{m_2}{m_2+\sigma_v^2} \qquad \text{—— (iii)}$$

It is interesting to note that equation (i) is the algebraic Riccati equation for the problem, equation (ii) simply the stationary covariance matrix, and equation (iii) shows that:-

$$u_n^{o} = B\, y_n = \frac{-\lambda}{\lambda+1} \quad E(x_n \mid y_n)$$

since the state variable $p_n(x_n)$ is Gaussian. The gain $\frac{-\lambda}{\lambda+1}$ is the Riccati gain of the system.

We can, of course, determine the optimal value of B by more direct methods, in other words by determination of m_2 from the expression above for J.

Thus

$$J = E(x^2 + B^2 y^2) = (1+B^2)\, m_2 + B^2 \sigma_v^2$$

and $$m_2 = (1+B)^2 m_2 + B^2 \sigma_v^2 + \sigma_w^2$$

From the variance equation,

$$m_2 = \frac{B^2 \sigma_v^2 + \sigma_w^2}{1 - (1+B)^2},$$

and hence

$$J = \frac{(1+B^2)\,(B^2 \sigma_v^2 + \sigma_w^2) + B^2 \sigma_v^2}{1 - (1+B)^2}$$

Differentiating and re-arranging we obtain, the necessary condition :-

$$B^4 \sigma_v^2 + 4B^3 \sigma_v^2 - B^2(\sigma_v^2 + \sigma_w^2) + B\sigma_w^2 + \sigma_w^2 = 0$$

This equation can be obtained also by elimination of λ, and m_2 from equations (i), (ii), and (iii) above. To obtain the optimal gain B^o, we must solve the above equation. For the simple scalar system we have here, this is not much of a problem and a simple Newton-Raphson technique can be used quite easily. This topic, especially for multi-dimensional systems is studied in Section 4. Before we leave the ergodic problem, let us draw a few qualitative conclusions from the above equation:-

i) For $\sigma_v = 0$ (no observation noise), we obtain

$$B^2 - B - 1 = 0,$$

that is $B^o = -0.62$

and $J = 1.62\ \sigma_w^2$

Notice that in this case, the optimal linear control law is independent of the process noise

ii) For $\sigma_v = \sigma_w$, we obtain

$$B^4 + 4B^3 - 2B^2 + B + 1 = 0,$$

that is $B^o = -0.411$

and $J = 3.00\ \sigma_w^2$

Denoting the necessary condition as $f(B) = 0$, we have sketched below the behaviour of $f'(B)$, together with the corresponding costs $J(B)$:-

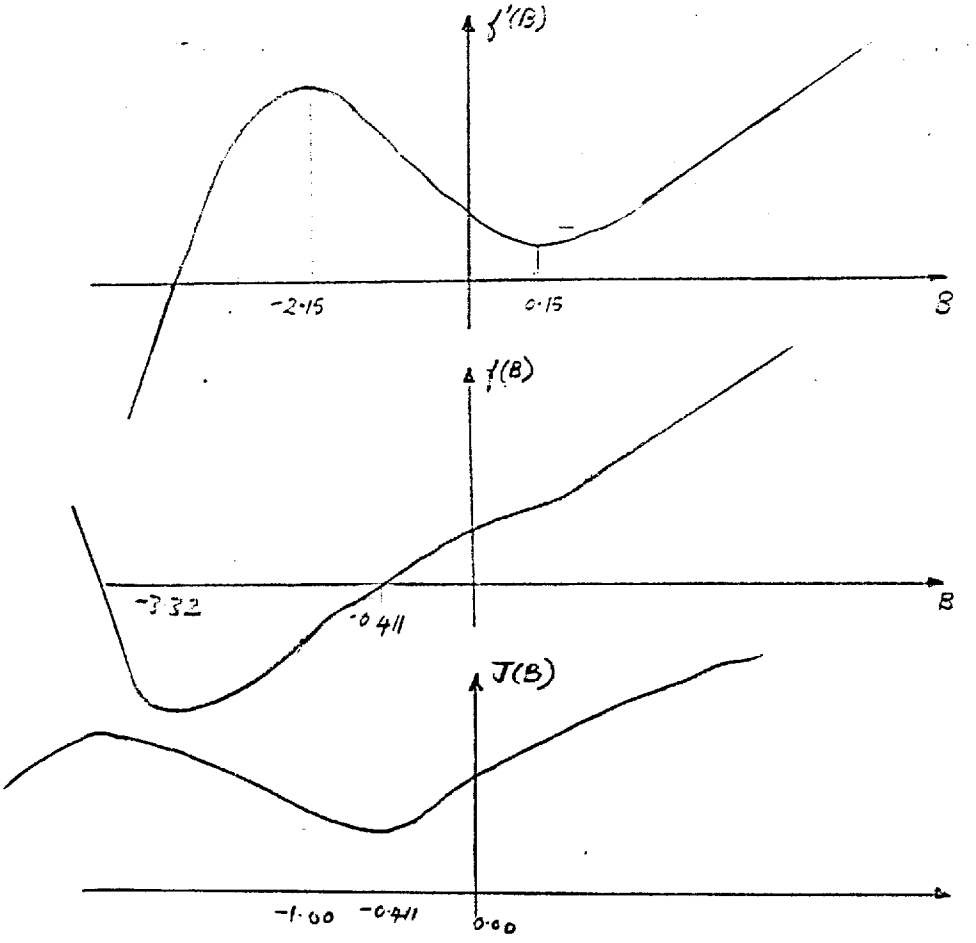


Fig. 7.1 Optimal Zero Memory Static Linear Controller ($\sigma_v = \sigma_w$)

It is clear that for $\sigma_v = \sigma_w$ there is a single local minimum, and that within the region of stability ($-1 < B < 0$), this is a global minimum. For different values of σ_w/σ_v the same qualitative behaviour occurs. Another local minimum will arise but it will always be outside the stability range for B. Thus we can conclude that, for stability, our local minimum will be a global minimum.

In the case $\sigma_v = \sigma_w$ the control is less tight than in (i) - we take less account of the observation. For the moment this concludes our examination of the ergodic problem. We shall examine it more thoroughly in Section 4 below. First however, let us summarise our attempts to find optimal linear controllers for our three versions of an infinite horizon zero memory problem .

7.3.6 Summary

In the preceding sections, we have hinted that by restricting the class of control laws to be linear functions of the current observation, we would generate a problem which we could easily solve. We have shown above that for the discount problem and the average cost per period problem this optimism is misplaced. As we have seen, the difficulty is due to the fact that although we impose a structure on the controller with respect to the observation, we are not imposing a structure with respect to the state $p_n(x_n)$. We could of course propose some structure for γ with respect to $p_n(x_n)$ - say for example:-

$$u_n = \gamma(y_n, p_n) = b(p_n).y_n$$

$$= \beta.m_2.y_n. = B_n.y_n$$

In other words the constant of proportionality is a linear function of the second moment of p_n. Such a controller must at least be as good as the static controllers which were all that we could find in our analysis in Sections 3.2 and 3.3. The problem of finding optimal linear controller s for the discount problem and the average cost per period problem will be examined more thoroughly in Section 5.

The distinction between stationary control and static control disappears for the ergodic problem since p_n becomes by definition a constant. Thus we can assert that static control is just as good as stationary control. That is, if

$$u_n = \gamma(y_n, p_n) = b(p_\infty).y_n$$

$$= B\, y_n.$$

we might as well concentrate on finding constants B rather than functionals $b(.\, p_\infty)$. The technique of using Lagrange multipliers to solve the ergodic problem is new to the author and is certainly conceptually clearer and analytically simpler than the traditional method of considering costs to infinity under linear control. Its major advantage however is that it can easily be extended to examine non-linear control laws for the ergodic problem.

In concluding this section, let us remark that a number of questions are outstanding:-

i) For the ergodic problem, how good is the linear control law?

ii) For the same problem, what is the optimal non-linear control law?

iii) For the same problem, can we develop an algorithm for the efficient numerical determination of B for a general multi-dimensional linear system?

iv) For the discount problem and the average cost per period problem, can we develop a method of approximating the optimal linear stationary control laws?

Items (i), (ii), and (iii) we will consider in the next Section. Item (iv) will be dealt with in Section 5.

7.4 Further Examination of the Ergodic Problem

The summary at the end of the preceding Section outlined a number of unresolved questions. These questions will be considered next.

7.4.1 How Good is the Linear Controller

We want to know how good the linear controller is compared to the general non-linear optimal controller. The trouble is, of course, that we do not know what the general non-linear optimal controller is. We can however get a very useful measure of the efficiency of the linear controller by comparing its performance with the performance of the same system under the optimal control of a controller with a classical information pattern.

Suppose $V_{\not\equiv M}^{L}$ is the performance of a zero memory ergodic problem under linear control, $V_{\not\equiv M}^{o}$ is the performance of the same problem under general optimal non-linear control, and that V_{CI}^{o} is the performance of the corresponding classical information pattern problem. Then

$$V_{CI}^{o} \leq V_{\not\equiv M}^{o} \leq V_{\not\equiv M}^{L} ,$$

and obviously

$$V_{\not\equiv M}^{L} - V_{\not\equiv M}^{o} \leq V_{\not\equiv M}^{L} - V_{CI}^{L}$$

Thus we can derive an upper bound for the benefits of non-linear zero-memory optimal control by evaluating $V_{\not\equiv M}^{L} - V_{CI}^{o}$. Let us consider the problem of evaluating this measure.

7.4.1.1 Classical Information Pattern Ergodic Problem

The ergodic problem for a system with a classical information pattern is conceptually awkward. Instead we shall examine the average cost per period problem. We can solve this problem by considering the limiting behaviour of the finite time version of the problem, using the example of Chapter 3. However, we shall approach the problem using the classical information pattern analogue of Theorem 7.6 - the optimality equation. Recall that in Chapter 3, we have already established that the probability density function $p_n(x_n \mid I_n)$ is a suitable state vector. The development proceeds along similar lines to Theorem 7.6:-

Theorem 7.9

Suppose:-

i) - that there exists functionals $V^0(p(x_n \mid I_n))$ and $\lambda(p(x_n \mid I_n))$ such that; for any $p(x_n \mid I_n)$:-

$$\lambda(p_n(x_n|I_n)) + V^0(p_n(x_n|I_n)) = \inf_{\gamma \in \Gamma} \left\{ E(x^2 + \gamma^2(p_n(x_n|I_n)) + V^0(\mathcal{L}(p_n(x_n|I_n, \gamma, \check{I}_{n+1}) \mid I_n \right\}$$

where $p(x_{n+1} \mid I_{n+1}) = \mathcal{L}(p_n(x_n \mid I_n), \gamma, \check{I}_{n+1})$

$$= \frac{p(\check{I}_{n+1} \mid x_{n+1}) \int p(x_{n+1} \mid x_n, \gamma_n(p_n(x_n|I_n)) p_n(x_n|I_n)\, dx_n}{\iint p(\check{I}_{n+1} \mid x_{n+1})\, p(x_{n+1} \mid x_n, \gamma_n(p_n(x_n|I_n)) p_n(x_n|I_n)\, dx_n dx_{n+1}}$$

ii) - $\frac{E\, V^0(p_n(x_n(I_n))}{N} \longrightarrow 0$ **for any sequence** $\{p_n(x_n(I_n)\}$ generated by $p_{n+1}(x_{n+1} \mid I_{n+1}) = \mathcal{L}(p_n(x_n \mid I_n), \gamma, \check{I}_{n+1})$ and an admissable policy $(\gamma_0, \gamma_1, \text{---}) \in U$.

Then, if $\gamma^{\circ}(p(x_n \mid I_n))$ attains the minimum in the above equation for each $p(x_n \mid I_n)$, the stationary policy $\gamma^{\circ}(p(x_n \mid I_n))$ is optimal, and the optimal average cost per period is:-

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E(\lambda(p_i(x_i \mid I_i))$$

Proof

From the proposed optimality equation:-

$$\lambda(p_0(x_0 \mid I_0) + V^{\circ}(p_0(x_0 \mid I_0)) \leqslant E_{p_0/I_0}\left[x_0^2 + \gamma_0^2(p_0(x_0 \mid I_0)) + V^{\circ}(p_1(x_1 \mid I_1)) \mid I_0\right]$$

$$\lambda(p_1(x_1 \mid I_1) + V^{\circ}(p_1(x_1 \mid I_1)) \leqslant E_{p_1/I_1}\left[x_1^2 + \gamma_1^2(p_1(x_1 \mid I_1)) + V^{\circ}(p_2(x_2 \mid I_2)) \mid I_1\right]$$

$$\vdots$$

and by repeated application

$$\lambda_N(p_N(x_N \mid I_N) + V^{\circ}(p_N(x_N \mid I_N) \leqslant E_{p_N/I_N}\left[x_N^2 + \gamma_N^2(p_N(x_N \mid I_N)) + V^{\circ}(p_{N+1}(x_{N+1} \mid I_{N+1})) \mid I_N\right]$$

Taking expectations of each inequality, we obtain typically:-

$$E_{\gamma^{\circ}}[\lambda_i(p_i(x_i \mid I_i))] + E_{\gamma^{\circ}}[V^{\circ}(p_i(x_i \mid I_i))] \leqslant E_{\gamma^{\circ}}[x_i^2 + \gamma_i^2(p_i(x_i \mid I_i))] + E_{\gamma^{\circ}}[V^{\circ}(p_{i+1}(x_{i+1} \mid I_{i+1}))]$$

where the notation $E_{\{\gamma^0\}}$ means that the expectation is with respect to the probability distributions induced by $\left\{\gamma^0 \mid j = 0,1,\text{---}\ i\right\}$.

Thus:-

$$\frac{1}{N}\sum_{i=0}^{N} E\left(\lambda(p_i(x_i/I_i))\right) \leqslant \frac{1}{N}\sum_{i=0}^{N} E\left(x_i^2+\gamma_i^2(p_i(x_i/I_i))\right)$$

$$+ \frac{E\left[V^0(p_N(x_N/I_N))\right] - E\,V^0(p_0(x_0/I_0))}{N}$$

and in the limit (by condition (ii))

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N} E\left(\lambda(p_i(x_i/I_i))\right) \leqslant \lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N} E\left(x_i^2+\gamma_i^2(p_i(x_i/I_i))\right)$$

for all $\gamma_0,\gamma_1,\gamma_2$ ---. Further more equality holds when $\Gamma^0=(\gamma^0,\gamma^0,\text{---})$, and consequently

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N} E\left[\lambda(p_i(x_i/I_i))\right] = \inf_{\Gamma_0^0 \in \Omega} \lim_{N\to\infty} \frac{1}{N} E\left[\sum_{i=0}^{N}(x_i^2+\gamma_i^2(p_i(x_i/I_i)))\right]$$

It follows from this theorem that if, under the stationary policy γ^0, $\lambda(p_i(x_i \mid I_i))$ becomes a constant λ^* as $N \to \infty$, then the optimal average cost period is λ^*. We shall see this property displayed for our example here.

Returning to the problem:-

Suppose $\lambda(p_i(x_i \mid I_i)) = \sigma^2_{i/i} + c\,\sigma^2_w$, $c^2 - c + 1 = 0$,

and $V^o(p_i(x_i \mid I_i)) = c m^2_{i/i}$,

where $\sigma^2_{i/i}$ is the variance of $p_i(x_i \mid I_i)$ and $m^2_{i/i}$ is the second moment.
Let us verify that the above λ and V^o satisfy our optimality equation:-

$$\bar{V}(p_i(x_i \mid I_i)) \triangleq \inf_{\gamma} \left\{ m^2_{i/i} + \gamma^2 + c\,E\left[m^2_{i+1/i+1}(\gamma, p_i(x_i/I_i)) \middle| I_i \right] \right\}$$

$$= \inf_{\gamma} \left\{ m^2_{i/i} + \gamma^2 + c\,E\left[(x+\gamma+\omega)^2 / I_i \right] \right\}$$

$$= \inf_{\gamma} \left\{ (1+c)\, m^2_{i/i} + (1+c)\gamma^2 + c\sigma^2_\omega + 2c\gamma E[x_i/I_i] \right\}$$

This expression is minimised when:-

$$\gamma(p_i(x_i \mid I_i)) = -\frac{c}{c+1} E(x_i \mid I_i)$$

and the corresponding value of $\bar{V}(p_i(x_i \mid I_i)$ is given by:-

$$\overline{V}(p_i(x_i/I_i)) = (1+c)m_{i/i}^2 + \frac{(1+c)c^2(\hat{x}_{i/i})^2}{(1+c)^2} + c\sigma_w^2 - \frac{2c^2(\hat{x}_{i/i})^2}{1+c},$$

where $\hat{x}_{i/i} = E[x_i \mid I_i]$. Thus:-

$$\overline{V}(p_i(x_i/I_i)) = (1+c)m_{i/i}^2 + c\sigma_w^2 - \frac{c^2(\hat{x}_{i/i})^2}{1+c}$$

$$= c\sigma_w^2 + (1+c)m_{i/i}^2 - \frac{c^2(m_{i/i} - \sigma_{i/i}^2)}{1+c}$$

$$= c\sigma_w^2 + \frac{(1+2c)m_{i/i}^2}{1+c} + \frac{c^2\sigma_{i/i}^2}{1+c}$$

But according to the definition of c :-

i) $\dfrac{1+2c}{1+c} = c,$

and ii) $\dfrac{c^2}{1+c} = 1$

Thus $\overline{V}(p_i(x_i \mid I_i)) = c\,\sigma_w^2 + \sigma_{i/i}^2 + c\,m_{i/i}^2$

$$= \lambda + V^{o}(p_i(x_i \mid I_i))$$

The functions λ and V^0 as defined satisfy the optimality equation of Theorem 7.9.

Notice that the analysis so far does not in anyway depend on the Gaussian properties of x_0, $\{v_i\}$, and $\{w_i\}$. However to proceed further we must compute $\sigma^2_{i/i}$ and this requires Gaussian random variables. Recall from the arguments of Chapter 3 that if $p(x_n \mid I_n)$ is Gaussian with mean $\hat{x}_{n/n}$ and variance $\sigma^2_{n/n}$, then $p(x_{n+1} \mid I_{n+1})$ is Gaussian with mean $\hat{x}_{n+1/n+1}$ and variance $\sigma^2_{n+1/n+1}$, where:-

$$\hat{x}_{n+1/n+1} = \frac{y_{n+1}(\sigma^2_{n/n} + \sigma_w^2) + (\hat{x}_{n/n} + u_n)\sigma_w^2}{(\sigma_v^2 + \sigma_w^2 + \sigma^2_{n/n})}$$

and

$$\sigma^2_{n+1/n+1} = \frac{(\sigma^2_{n/n} + \sigma_w^2)\sigma_v^2}{(\sigma^2_{n/n} + \sigma_v^2 + \sigma_w^2)}$$

with $\sigma^2_{0/0} = \sigma_v^2 \sigma_0^2 / (\sigma_v^2 + \sigma_0^2)$. But, as we have remarked before, $\sigma^2_{n+1/n+1}$ is independent of the data I_{n+1}. That is:-

$$\lambda^* = \lim_{N \to \infty} \left\{ c\sigma_w^2 + \sigma^2_{N/N} \right\}$$

$$= c\sigma_w^2 + \sigma_\infty^2, \text{ where}$$

$$\sigma_\infty^2 = \frac{(\sigma_\infty^2 + \sigma_w^2)\,\sigma_v^2}{(\sigma_\infty^2 + \sigma_w^2 + \sigma_v^2)}$$

that is, $\sigma_\infty^4 + \sigma_\infty^2\,\sigma_w^2 - \sigma_w^2\,\sigma_v^2 = 0$

Our convergence condition (ii) will be satisfied if U is restricted to those policies which generate finite second moment in x_n. Thus the conditions of the theorem are satisfied.

In summary then, the optimal cost for the classical information pattern version of our problem is given by:-

$$\lambda^* = c\,\sigma_w^2 + \sigma_\infty^2,$$

where $\sigma_\infty^4 + \sigma_\infty^2\,\sigma_w^2 - \sigma_w^2\,\sigma_v^2 = 0,$

and $c^2 - c - 1 = 0$

Note also that the optimal law is linear in $E(x_n \mid I_n)$, which because of Gaussianity is linear in I_n. Let us next turn to the zero memory ergodic problem with static linear control laws as discussed in Section 3.5.

7.4.1.2 Zero Memory Ergodic Problem

Recall that in Section 3.5, the optimal static gain of our zero-memory problem was the (stable) solution of the following necessary condition:-

$$B^4 \sigma_v^2 + 4B^3 \sigma_v^2 - B^2(\sigma_v^2 + \sigma_w^2) + B\sigma_w^2 + \sigma_w^2 = 0$$

If the solution to this equation is B_0, then the optimal cost is given by:-

$$J_{ZML} = (1 + B_0^2)\sigma_X^2 + B_0^2 \sigma_v^2 = \sigma_X^2 + B_0^2(\sigma_X^2 + \sigma_v^2)$$

where $\sigma_X^2 = (1+B_0)^2 \sigma_X^2 + B_0^2\sigma_v^2 + \sigma_w^2$,

that is $$\sigma_X^2 = -\frac{\left[B_0^2\sigma_v^2 + \sigma_w^2\right]}{1 - (1+B_0)^2}$$

Thus $$J_{ZML} = \sigma_X^2 - 2B_0\sigma_X^2 - \sigma_w^2 = \sigma_w^2 \left\{ \frac{\sigma_X^2}{\sigma_w^2} - 2B_0 \frac{\sigma_X^2}{\sigma_w^2} - 1 \right\}$$

7.4.1.3 Comparison of Zero-Memory and Classical Information Pattern Problem

To proceed further, it is useful to introduce the variable ρ, where:-

$$\rho \triangleq \frac{\sigma_v^2}{\sigma_w^2}$$

Let us denote the optimal cost for the classical information pattern problem by J_{CIP}. Now

$$J_{CIP} = C\sigma_w^2 + \sigma_\infty^2 , \quad = \sigma_w^2 \left\{ C + \frac{\sigma_\infty^2}{\sigma_w^2} \right\},$$

where $C^2 - C - 1 = 0,$

and $\sigma_\infty^4 + \sigma_\infty^2 \sigma_w^2 - \sigma_w^2 \sigma_v^2 = 0,$

that is

$$\left(\frac{\sigma_\infty^2}{\sigma_w^2}\right)^2 + \left(\frac{\sigma_\infty^2}{\sigma_w^2}\right) - \rho = 0$$

Thus

$$J_{CIP} = \sigma_w^2 \left[\left\{ \frac{1+\sqrt{1+4}}{2} \right\} + \left\{ \frac{-1+\sqrt{1+4\rho}}{2} \right\} \right]$$

The positive roots in the above expressions are taken to ensure stable gain and positive variance.

Thus

$$J_{CLP} = \sigma_\omega^2 \left\{ \frac{\sqrt{5} + \sqrt{1+4\rho}}{2} \right\}$$

Finally, we have, for the ratio R of optimal costs,

$$R = \frac{J_{ZML}}{J_{CLP}} = \frac{2(\sigma^2 - 2B_0^2\sigma^2 - 1)}{\sqrt{5} + \sqrt{1+4\rho}}$$

where:-

$$B_0^4 \rho + 4B_0^3 \rho - B_0^2(1+\rho) + B_0 + 1 = 0.$$

and:-

$$\sigma^2 = - \frac{\left[B_0^2 \rho + 1\right]}{1 - (1+B_0)^2}$$

Further simplification of the above expression for R is not possible. It can however be easily evaluated by computer for various values of ρ. Some results are shown in Table 7.1 and Table 7.2. Two batches of results have been produced. In the first ρ is incremented in steps of 0·25 up to a final value of 12.500. It will be seen that the ratio of linear zero-memory costs to Kalman filter costs is, in this range of ρ, very close to unity. Qualitatively, we never do much better than 3% by using a Kalman filter instead of a simple static linear zero-memory controller.

-ARISON OF LINEAR STATIONARY PROBLEM WITH CLASSICAL PROBLEM

OF OSI AS A FUNCTION OF RATIO OF SV/SW

VARIANCES ROE	RATIO OF COSTS	OPTIMAL GAIN	STATIONARY V
500	1.00251607	-.52658676	1.37820737
	488	-.47434646	1.53726708
7		-	
n		-	
		-	
		-	9310483
7	1.01646699	-.35745928	2.08401238
	1.01773825	-.34459777	2.16933957
	1.01884348	-.33333333	2.25000000
	1.01981158	-.32334289	2.32668284
	1.02066517	-.31439019	2.39992307
	1.02142214	-.30629753	2.47014578
500	1.02209684	-.29892811	2.53769538
3.5000	1.02270092	-.29217452	2.60285540
	1.02324392	-.28595089	2.66586284
	1.02373375	-.28018771	2.72691781
500	1.02417699	-.27482787	2.78619133
5000	1.02457921	-.26982394	2.84383087
500	1.02494511	-.26513615	2.89996464
	1.02527872	-.26073083	2.95470493
.2500	1.02558349	-.25657924	3.00815076
.5000	1.02586240	-.25265666	3.06038995
500	1.02611805	-.24894172	3.11150086
	1.02635270	-.24541579	3.16155370
2500	1.02656834	-.24206255	3.21061170
	1.02676672	-.23886765	3.25873206
500	1.02694938	-.23581838	3.30596668
	1.02711769	-.23290345	3.35236285
500	1.02727287	-.23011279	3.39796381
	1.02741602	-.22743739	3.44280921
	1.02754812	-.22486912	3.48693550
	1.02767003	-.22240069	3.53037631
8.2500	1.02778256	-.22002546	3.57316270
	1.02788641	-.21773744	3.61532349
	1.02798224	-.21553115	3.65688542
	1.02807064	-.21340162	3.69787339
500	1.02815213	-.21134428	3.73831061
	1.02822721	-.20935496	3.77821879
500	1.02829632	-.20742980	3.81761823
	1.02835987	-.20556526	3.85652797
500	1.02841824	-.20375809	3.89496589
5000	1.02847176	-.20200527	3.93294882
500	1.02852075	-.20030399	3.97049259
.0000	1.02856550	-.19865167	4.00761215
.2500	1.02860629	-.19704588	4.04432162
0	1.02864335	-.19548440	4.08063434
1.7500	1.02867693	-.19396512	4.11656297
000	1.02870722	-.19248609	4.15211947
500	1.02873443	-.19104548	4.18731523
5		-	6106

Table 7.1 : Classical Performance versus Zero-Memory Performance

($\rho \leq 12.5$)

COMPARISON OF LINEAR STATIONARY PROBLEM WITH CLASSICAL PROBLEM

RATIO OF COST AS A FUNCTION OF RATIO OF SV/SW

RATIO OF VARIANCES ROE	RATIO OF COSTS	OPTIMAL GAIN	STATIONARY
.25000E+00	.10025E+01	-.52659E+00	.13782E+01
.10000E+01	.10112E+01	-.41145E+00	.17890E+01
.40000E+01	.10237E+01	-.28019E+00	.27269E+01
.16000E+02	.10289E+01	-.17309E+00	.46782E+01
.64000E+02	.10246E+01	-.10002E+00	.86313E+01
.25600E+03	.10167E+01	-.54988E-01	.16587E+02
.10240E+04	.10099E+01	-.29138E-01	.32553E+02
.40960E+04	.10054E+01	-.15060E-01	.64529E+02
.16384E+05	.10028E+01	-.76658E-02	.12852E+03
.65536E+05	.10015E+01	-.38701E-02	.25651E+03
.26214E+06	.10007E+01	-.19443E-02	.51250E+03
.10486E+07	.10004E+01	-.97449E-03	.10245E+04
.41943E+07	.10002E+01	-.48784E-03	.20485E+04
.16777E+08	.10004E+01	-.25011E-03	.40977E+0
.67109E+08	.10286E+01	-.15486E-03	.84256E+04
.26844E+09	.10741E+01	-.89469E-04	.17598E+
.10737E+10	.11239E+01	-.49944E-04	.36826E+05
.42950E+10	.11742E+01	-.27305E-04	.76949E+0
.17180E+11	.12238E+01	-.14719E-04	.16040E+06
.68719E+11	.12723E+01	-.78539E-05	.33352E+0
.27488E+12	.13195E+01	-.41586E-05	.69179E+06
.10995E+13	.13654E+01	-.21887E-05	.14317E+
.43980E+13	.14100E+01	-.11463E-05	.29569E+07
.17592E+14	.14533E+01	-.59794E-06	.60957E+07
.70369E+14	.14956E+01	-.31085E-06	.12546E+08
.28147E+15	.15367E+01	-.16114E-06	.25781E+08
.11259E+16	.15768E+01	-.83326E-07	.52909E+08
.45036E+16	.16160E+01	-.42995E-07	.10845E+09
.18014E+17	.16542E+01	-.22143E-07	.22203E+09
.72058E+17	.16917E+01	-.11385E-07	.45411E+0
.28823E+18	.17283E+01	-.58449E-08	.92788E+09
.11529E+19	.17642E+01	-.29966E-08	.18943E+1
.46117E+19	.17994E+01	-.15345E-08	.38641E+10
.18447E+20	.18339E+01	-.78491E-09	.78766E+1
.73787E+20	.18678E+01	-.40109E-09	.16044E+11
.29515E+21	.19011E+01	-.20477E-09	.32660E+11
.11806E+22	.19338E+01	-.10445E-09	.66445E+11
.47224E+22	.19660E+01	-.53240E-10	.13510E+12
.18889E+23	.19976E+01	-.27117E-10	.27455E+12
.75558E+23	.20288E+01	-.13803E-10	.55767E+12
.30223E+24	.20595E+01	-.70212E-11	.11322E+13
.12089E+25	.20897E+01	-.35695E-11	.22977E+13
.48357E+25	.21196E+01	-.18137E-11	.46610E+13
.19343E+26	.21490E+01	-.92111E-12	.94512E+13
.77371E+26	.21780E+01	-.46757E-12	.19158E+14
.30949E+27	.22066E+01	-.23724E-12	.38819E+14
.12379E+28	.22348E+01	-.12032E-12	.78631E+14
.49518E+28	.22627E+01	-.61000E-13	.15923E+15
.19807E+29	.22903E+01	-.30914E-13	.32233E+15
.79228E+29	.23175E+01	-.15661E-13	.65233E+15

Table 7.2 : Classical Performance versus Zero-Memory Performance

($\rho \leq 4^{13}$)

In the next set of results ρ is multiplied by 4.0 at each step (that is, σ_v/σ_w is doubled). This produces two significant observations.

(i) A local maximum appears to exist round about $\rho = 16$ with cost ratio 1.028.

(ii) For values of ρ which approach infinity, the magnitude of the cost ratio becomes larger. However, for $\rho \leq 4^{13}$, the cost ratio is less than 1.023.

A sketch of the cost ratio against ρ is given below:-

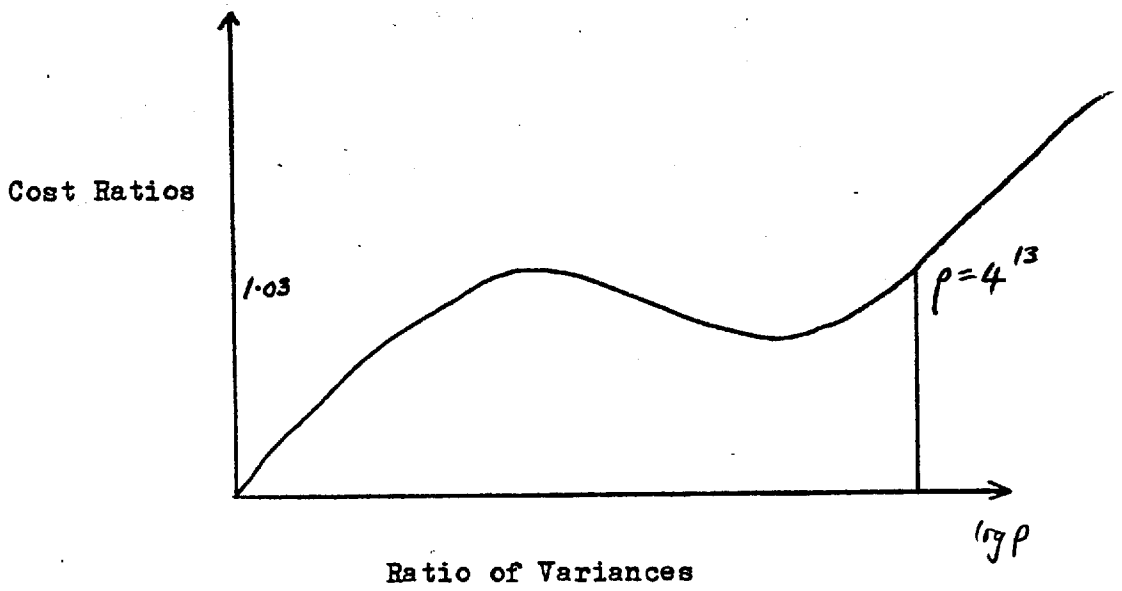


Fig. 7.1 Classical Performance versus Zero-Memory Performance.

7.4.1.5 Conclusions

The above numerical comparison of zero memory performance with the performance of a controller with a classical information pattern is indeed interesting. It indicates that, for the ergodic problem, the designer, in implementing anything other than simple static linear control, will achieve nothing more than very small marginal benefits for a very broad range of ρ. Outside this range, the model of the system and the philosophy of control at all become of dubious value.

Notice that we have compared optimal zero-memory linear control with optimal classical information pattern control. Thus, we cannot identify the source of the improved performance of the classical information pattern problem; it could be due to:-

i) the value of the extra information associated with the classical information pattern problem.

ii) the superiority of the optimal non-linear control zero-memory control law over the simple static linear control law we have adopted here.

iii) A combination of items (i) and (ii), above.

For $\rho \leq 4^{13}$ the value of the extra information of the classical information pattern problem is less than 3%. Thus the designer should carefully evaluate whether the cost of collecting this information is worth such a marginal return when a much simpler alternative exists, the performance of which is almost as good.

The preceding analysis and the above comments all relate to our simple linear scalar system. In real applications, the designer will be confronted with a more complex multi-dimensional problem. In the next section we shall examine methods which provide solutions to such problems.

7.4.2 Design for Linear Multi-Dimensional Systems

In Section 3.5 we developed the necessary conditions for the optimal gain of our simple scalar zero-memory ergodic problem. Two possibilities for finding the optimal gain were presented. The first alternative involved finding the solution of a quartic equation involving the optimal gain B and the variances σ_v^2 and σ_w^2. A second alternative is to find the solution to the following set of simultaneous equations:-

$$1) \qquad \lambda = 1 + b^2 + \lambda(1+b)^2$$

$$2) \qquad \sigma_X^2 = (1+b)^2 \sigma_X^2 + b^2 \sigma_v^2 + \sigma_w^2$$

$$3) \qquad b = \frac{-\lambda}{1+\lambda} \cdot \frac{\sigma_X^2}{\sigma_X^2 + \sigma_v^2}$$

It is this alternative which will be examined here, since it appears to offer the most appropriate technique for solving larger more complex linear multi-dimensional systems. The technique which is proposed for finding solutions to the above set of equations is quite simple and bears considerable similarity with the method of policy iteration discussed earlier:-

(a) Guess some stable b_0 from (3)

(b) Let k = 0

(c) Find λ_k according to (1) and σ_k^2 according to (2)

(d) Let k = k+1; compute:-

$$\underline{b} = -\frac{\lambda_{k-1}}{1+\lambda_{k-1}} \cdot \frac{\sigma_{k-1}^2}{\sigma_{k-1}^2 + \sigma_v^2}, \text{ and } b_k = \alpha \underline{b} + (1-\alpha) b_{k-1} \quad (0 < \alpha < 1)$$

If $b_k \neq b_{k-1}$, then go to (c) and continue.

Otherwise, if $b_k = b_{k-1}$ (or almost), then stop; $b^o = b_{k-1}$, where b^o is the optimal gain.

We can show that the above algorithm will produce a decrease in cost per iteration.

Let

$$\Delta b_K = b_K - b_{K-1} = \alpha \underline{b} + (1-\alpha) b_{K-1} - b_{K-1}$$

$$= \alpha (\underline{b} - b_{K-1})$$

$$= \alpha \left\{ \frac{-\lambda_{K-1}}{1+\lambda_{K-1}} \cdot \frac{\sigma_{K-1}^2}{\sigma_{K-1}^2 + \sigma_V^2} - b_{K-1} \right\} \text{ from (d)}$$

$$= -\alpha \left\{ \frac{b_K (1+\lambda_{K-1})(\sigma_{K-1}^2 + \sigma_V^2) + \lambda_{K-1} \sigma_{K-1}^2}{(1+\lambda_{K-1})(\sigma_{K-1}^2 + \sigma_V^2)} \right\}$$

Now, by Taylor's Theorem,

$$J(b_{K-1} + \Delta b_K) - J(b_{K-1}) \simeq \Delta b_K \left. \frac{dJ(b)}{db} \right|_{b=b_{K-1}}$$

But

$$\frac{dJ(b)}{db} = \frac{d}{db}\left[(1+b^2)\sigma_x^2 + b^2 \sigma_V^2 \right]$$

$$= 2b\sigma_x^2 + 2b\sigma_V^2 + (1+b^2)\frac{d\sigma_x^2}{db}$$

$$= 2b\sigma_x^2 + 2b\sigma_V^2 + (1+b^2)\left\{ \frac{2(1+b)\sigma_x^2 + 2b\sigma_V^2}{1-(1+b)^2} \right\} \text{ from (2)}$$

Also, $\lambda = \dfrac{1+b^2}{1-(1+b)^2}$ from (1)

$$\therefore \quad \frac{dJ(b)}{db} = 2b\sigma_x^2 + 2b\sigma_v^2 + \lambda\left(2(1+b)\sigma_x^2 + 2b\sigma_v^2\right)$$

$$= 2b(\sigma_x^2+\sigma_v^2) + \lambda 2b(\sigma_x^2+\sigma_v^2) + 2\lambda\sigma_x^2$$

$$= 2b(1+\lambda)(\sigma_x^2+\sigma_v^2) + 2\lambda\sigma_x^2$$

Thus

$$J(b_{k-1}+\Delta b_k) - J(b_{k-1}) = -2\alpha\left\{\frac{\left(b_{k-1}(1+\lambda_{k-1})(\sigma_{k-1}^2+\sigma_v^2) + \lambda_{k-1}\sigma_{k-1}^2\right)^2}{(1+\lambda_{k-1})(\sigma_{k-1}^2+\sigma_v^2)}\right\}$$

This expression is always negative provided that $1 + \lambda_{k-1} > 0$.

Recall that $X_{k-1} = \dfrac{1 + b_{k-1}^2}{1-(1+b_{k-1})^2}$; if b_{k-1} is stable then $(1+b_{k-1})^2 < 1$.

Thus λ_{k-1} is positive and consequently so is $1 + \lambda_{k-1}$. Thus the algorithm proposed above will produce a reduction in cost at each iteration, provided that the stepping parameter α is small enough to justify

the first order Taylor approximation.

We have developed the above algorithm in terms of our simple scalar example. This approach was chosen in order to maintain continuity with all our previous analysis. Let us now see how the algorithm can be applied to a general linear multi-dimensional system. The algorithm for this case has been fully worked out and justified by S.D.G. Cumming[1]. Below we merely formulate the problem and state the algorithm for its solution.

Given the n-state, r-input, m-output system

$$x_{k+1} = Ax_k + Bu_k = W_k$$

$$y_k = Hx_k + v_k$$

$$u_k = Ky_k ,$$

it is desired to minimise with respect to K the cost functional

$$J = \lim_{N \to \infty} \frac{1}{N} \left\{ E \sum_{k=0}^{N} (x_k Q x_k + u_k R u_k) \right\} .$$

The noise in the system is described as follows:-

$E(w_k) = 0$, independent white noise

$E(v_k) = 0$, independent white noise

$E(w_k w_i^T) = N \delta(i,k)$, where δ is the Kronecker delta, and

1. S.D.G. Cumming "Optimal Feedback Gains for Linear Discrete Time Systems with Noise on Inputs and Measurements". Imperial College Control Systems Research Report No. 6918.(17)

$$E(v_k v_i^T) = M \delta(i,k)$$

Let $A_k \triangleq (A + B K H)$ and assume that A_k is stable. Next we define $V \triangleq N + B K M K^T B^T + A_k V A_k^T$, (C (2) above) (A)

and $P \triangleq Q + H^T K^T R K H + A_k P A_k^T$ (C (1) above) (B)

Then the optimal gain K_0 is given by

$$K_0 = -(R + B P B)^{-1} B^T P V H^T (H V H^T + M)^{-1} \text{ (of (3) above) (C)}$$

A brief examination of these equations will confirm that they reduce to those we derived earlier for the scalar problem. The derivation of Cumming differs from our analysis in that we have employed the concept of constrained optimisation, whereas as Cumming has used the concept of a cost to infinity matrix. It is interesting to note that in our approach the Lagrange multiplier λ obeys the same equations as the cost to infinity matrix P in Cumming's approach. As we remarked earlier, the constrained optimisation approach was selected because it is apparently easier to extend to non-linear controllers. The notion of a cost to infinity matrix P seems to be appropriate only for linear control laws.

The algorithm proposed by Cumming for solving the above equations (A,B, and C) for K_0 is completely analogous to ours. We simply start off with some guess at $K_0(0)$. We solve (A) and (B) for V(0) and P(0) respectively. Next we compute a matrix $\underline{K}$ according to (C), and generate a new gain matrix from $K_0(1) = \alpha \underline{K} + (1-\alpha)K(0)$. Again we solve (A) and (B) for V(1) and P(1) respectively. We find the new $\underline{K}$ from (C), and

generate a new gain matrix from $K_0(2) = \alpha \underline{K} + (1-\alpha)K(1)$. This process is repeated until $K_0(k)$ becomes virtually constant.

This concludes our discussion of the general multi-dimensional linear zero-memory ergodic problem. We have provided some justification for the proposed algorithm in terms of our scalar example. For precise details the reader is referred to the report cited.

An interesting problem which remains is that of finding out, under what conditions, the local minimum which the proposed algorithm determines is a global minimum. We have encountered conditions under which this is true in Section 7.3.5 - it would be useful to extend them to the multi-dimensional problem.

7.4.3 Optimality of the Static Linear Control Law

Prolonged efforts have been made to estalish the optimality of the static linear control law over the class of all admissable functions and not merely the linear class. These efforts have failed.

One approach was to find functionals λ and $V(p)$ which satisfy Theorem 7.6. In this case, the difficulty is that the optimality equation has to hold for all p reachable by arbitrary policies; in other words whilst for normal p we can find a $V(p)$ which satisfies the optimality equation, it is not possible to find the general form of $V(p)$.

Another approach considered involved the use of Lagrange multipliers to establish a necessary condition for optimality. Again, the necessary condition is satisfied by the static linear controller, but

it was not possible to establish the uniqueness of this solution.

In the following section we shall consider a number of approximation methods for the ergodic problem. In the moment formulation method we shall derive necessary conditions which are satisfied by the best linear controller. This also occurs when we study a perturbation approach. We appear to have a situation similar to the Witsenhausen problem in which the linear control law satisfies the necessary condition for overall optimality. Perhaps what is required to save further wasted effort is a counter example (with $\rho \leq 4^{\frac{1}{3}}$ see Section 7.4.1.4)

7.4.4 Approximation Methods

In the previous section, we have examined in some detail the linear solution for the zero-memory ergodic problem. In this section we shall consider a number of approximation methods which can be applied when either the best linear control law is not good enough or the properties of the system are significantly changed. Recall that we have so far assumed that:-

i) the process equation is linear.

ii) the observation equation is linear.

iii) the cost function is quadratic.

iv) the observation and process noises are Gaussian.

We shall now briefly examine the problem of finding approximations to the optimal ergodic control law when some of the above assumptions are relaxed. We shall consider a 'moment' formulation of the zero-memory ergodic problem and an ' ϵ-approximation' technique. Only an outline of these topics will be discussed. This is because conceptually both techniques are quite simple. However, analytically they lead to extremely complex sets of equations which offer little insight into the problem. As a numerical technique, the moment formulation, with some refinement, does appear to offer some assistance in finding optimal non-linear controls. Let us first consider a moment formulation of the Ergodic Problem:-

7.4.4.1 Moment Formulation

We view the Ergodic Problem as a constrained optimisation problem. The costfunctional must be minimised with respect to the control law and the state probability density function, subject to the stationarity constraint linking the control law and the probability density function. In order to reduce this functional minimisation problem to one involving minimisation with respect to scalar quantities, we parameterise as follows:-

(i) The control law is expressed as a power series expansion of the observation:-

$$u = \gamma(y) = \sum_{i=0}^{\infty} c_i y^i$$

(ii) The state probability density function is expressed in terms of its moments:-

$$m_i = \int x^i \, p(x) \, dx$$

Notice that $E\left[f(x)\right] = \int f(x) \, p(x) \, dx$

$$= \int \sum_{i=0}^{\infty} f_i \, x^i \, p(x) \, dx$$

$$= \sum_{i=0}^{\infty} f_i \, m_i,$$

where f_i are the coefficients of the Taylor Series expansion of $f(x)$ about $x = 0$.

Let us now formulate a general Ergodic Problem. Given process and observation equations:-

$$x_{k+1} = f(x_k, u_k, w_k)$$

$$y_k = g(x_k, v_k)$$

where, for zero-memory control:-

$$u_k = \gamma(y_k),$$

the problem is to minimise the cost functional J with respect to γ, where

$$J = E\left[L(x_k, u_k)\right],$$

the expectation above being with respect to the stationary distribution generated by γ. Thus, introducing our parameters $\{c_i\}$ and $\{m_i\}$, the problem is to minimise:-

$$J\left(\left\{c_i\right\},\left\{m_i\right\}\right) = E\left[L\left(x,\sum_i c_i\, y^i\right)\right]$$

subject to the constraints:-

$$m_j = E\left\{\left[f\left(x,\sum_i c_i\, y^i,w\right)\right]^j\right\} \quad , \; j = 1,2,3,\text{----}$$

Using the technique of Lagrange multipliers, we can introduce the augmented cost function:-

$$J^A\left(\left\{c_i\right\},\left\{m_i\right\},\left\{\lambda_i\right\}\right) = E\left[L\left(x_k,\sum_i c_i\, y^i\right)\right]$$

$$+ \sum_j \lambda_j\left(m_j - E\left\{f\left(x,\sum_i c_i\, y^i,w\right)^j\right\}\right)$$

Notice that the expectations in the above expression can all be expressed, albeit perhaps very intricately, in terms of the moments $\left\{m_i\right\}$. We can then determine the optimal $\left\{c_i\right\}$ and $\left\{m_i\right\}$ via the necessary conditions

$$\frac{\partial J^A}{\partial c_i} = 0, \; i = 1,2,3,\text{----}$$

$$\frac{\partial J^A}{\partial m_i} = 0, \; i = 1,2,3,\text{----}$$

$$\frac{\partial J^k}{\partial \lambda_j} = 0, \; j = 1,2,3\text{----}$$

In its most general form, this formulation is exceedingly complex. However, for particular problems the formulation can be very convenient. It was in fact used in Section 3 to determine the optimal linear control law for our simple problem. It is particularly simple for this problem since we are only trying to determine c_1 and m_2. We know that the stationary distribution under linear control will be Gaussian, and therefore the only constraint we are interested in is that involving the second moment m_2. We can also argue that m_1 must be zero for linear (proportional) control.

Before we proceed to give an example of the application of the technique notice the similarity of this formulation with that encountered in Theorem 7.6 for example. There we were concerned with the problem of minimising expressions such as:-

$$\left\{ E(L(x,u)) + V(p') \right\}$$

where $$p'(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p(\theta)\ d\theta\ dy$$

Here we are concerned with minimising the expression:-

$$\left\{ E(L(x,u)) + \sum_j \lambda_j . (E(f(x,u,w))^j - m_j \right\}$$

Thus there would appear to be some connection between $V(p')$ and $\sum_j \lambda_j \left\{ E(f(x,u,w))^j - m_j \right\}$. The problem with the first formulation however is that we don't know what $V(p)$ is. At least in the moment formulation we know what we are dealing with - even though it can be

very complicated. It is a matter of weighing this complexity against the feasibility of a successive approximation via a dynamic programming formulation.

As an example of the application of the technique we shall revert to our simple scalar system. We shall discard the assumption that $\{w_i\}$ is a Gaussian sequence and attempt to find the best cubic control law:- We obtain for the augmented cost function:-

$$J^{\wedge} = E\left(x^2 + (c_1(x+v) + c_3(x+v)^3)^2\right)$$

$$+ \sum_j \lambda_j \cdot E\left(x^j - (x + c_1(x+v) + c_3(x+v)^3 + w)^j\right)$$

$$= E\left(x^2 + c_1(x+v)^2 + 2c_1c_3(x+v)^4 + c_3^2(x+v)^6\right)$$

$$+ \sum_j \lambda_j \cdot \left(m_j - E(x + c_1(x+v) + c_3(x+v)^3 + w)^j\right)$$

Notice that we have assumed that the optimal control is an odd function and dispensed with the coefficients c_0 and c_2. We thus obtain the following necessary conditions for optimality:-

1) $$2c_1(m_2 + \sigma_v^2) + 2c_3 E(x+v)^4$$

$$- \sum_j \lambda_j \cdot j E\left((x+v)(x + c_1(x+v) + c_3(x+v)^3 + w)^{j-1}\right) = 0$$

2) $$2c_1 E(x+v)^4 + 2c_3 E(x+v)^6$$

$$- \sum_j \lambda_j \cdot j E\left((x+v)^3(x + c_1(x+v) + c_3(x+v)^3 + w)^{j-1}\right) = 0$$

3) $m_j - E\left((x + c_1(x+v) + c_3(x+v)^3 + \omega)^j\right) = 0 \quad j = 0, 1, 2$

4) For m_1 $\quad \lambda_1 (1 - (1 + c_1)) = 0 \quad \Longrightarrow \lambda_1 = 0$

5) For m_2 $\quad 1 + c_1^2 + 12 c_1 c_3 \sigma_V^2 + 45 c_3^2 \sigma_V^2$

$$+ \sum \lambda_j \frac{d}{dm_j} \left\{ m_j - E(x + c_1(x+v) + c_3(x+v)^3 + \omega)^j \right\} = 0$$

6) For m_3, m_4, m_5 ----- and so on, as above.

There may be more than one solution tc the above set of equations. Let us attempt to find one. Suppose $c_3 = 0$, with c_1 given by the optimal solutions to the linear control law problem. If $c_3 = 0$, and c_1 as for linear problem, then all the above equations are satisfied by inspection except equation (2). Thus $c_3 = 0$ will be a solution if we can show that equation (2) is satisfied with c_1 the optimal linear solution. Thus, we must demonstrate that:-

$$2\, c_1\, E\left[x+v\right]^4 - \sum_j \lambda_j\, E\left[(x + c_1(x+v) + w)^{j-1}(x+v)^3\right] = 0,$$

with λ_j as determined by linear control law problem. Now for linear

control law problem it is easy to demonstrate that $\lambda_j = 0 \quad j \neq 2$. Hence we only need to demonstrate that (equation (2)):-

$$2\, c_1\, E(x+v)^4 - 2\lambda_2 E\left[(x+c_1(x+v)+w)(x+v)^3\right] = 0$$

Recall that we have from (1)

$$2\, c_1(m_2+\sigma_v^2) - 2\lambda_2\, E\left[(x+c_1(x+v)+w)(x+v)\right] = 0,$$

which yields $c_1(m_2+\sigma_v^2) = \lambda_2(m_2(1+b)+ b\sigma_v^2)$

Now returning to (2), and assuming the gaussianity if x,

$$c_1\, E(x+v)^4 = 3c_1\,(m_2+\sigma_v^2)^2,$$

and

$$\lambda_2 E\left[(x+b(x+v)+w(x+v)^3\right] = 3\lambda_2(m_2+\sigma_v^2)\left\{(1+b)m_2+b\sigma_v^2\right\}$$

Thus

$$c_1\, E(x+v)^4 - \lambda_2\, E\left[(x+b(x+v)+w)(x+v)^3\right] = 3(m_2+\sigma_v^2)\left\{c_1(m_2+\sigma_v^2) - \lambda_2\right.$$

$$\left.\cdot(m_2(1+b)+b\sigma_v^2)\right\} = 0$$

Thus equation (2) is satisfied if $c_3 = 0$ and c_1 given by solution to optimal linear control problem. We have merely demonstrated that the necessary conditions for our ergodic problem are satisfied by the best linear controller.

This concludes our discussion of what we have called the moment formulation. The technique is analytical in nature because it involves differentiation. If the designer is prepared to suffer the labour, it can be applied to quite general systems. As before we mention the possibility of FORMAC to assist in this respect although in the author's experience the expressions involved will too quickly exceed the memory of the computer. In the next section we shall discuss a numerical approximation method which conveniently relegates the hard work to the computer.

7.4.4.2 Edgeworth's Series Approximation

We remarked in the previous section that the 'moment' formulation was essentially an analytic technique since it required differentiation to derive the necessary conditions for optimality. As an alternative to differentiation in numerical optimisation problems, we can employ so-called 'direct-search procedures'.[1]

Suppose we wish to find the optimum cubic control law. Assuming that $c_0 = c_2 = 0$, we have to find optimal coefficients c_1 and c_3:-

$$u = \gamma(y) = c_1 y + c_3 y^3$$

Let us denote the corresponding cost as $J(c_1, c_2)$.

1. Box, Davies and Swann - Non Linear Optimisation Techniques (10)

At various points in the direct search procedure we shall have the problem of finding $J(c_1, c_2)$ for particular values of c_1 and c_2. To evaluate $J(c_1, c_2)$ we need to compute the stationary probability density function corresponding to c_1 and c_2. This is given by $p_\gamma(x)$ where:-

$$p_\gamma(x) = \iint p\left\{x \mid \theta, \gamma(y)\right\}\ p(y \mid \theta)\ p_\gamma(\theta)\ d\theta\ dy$$

Note that density functions $p\left\{x \mid \theta, \gamma(y)\right\}$ and $p\left\{y \mid \theta\right\}$ are known. Hence we have the problem of solving the above integral equation at $\gamma(y) = c_1 y + c_2 y^3$ to determine the function $p_\gamma(x)$. In the previous section it was suggested that we could express $p_\gamma(x)$ in terms of its moments, and we could utilise the same device here to determine $J(c_1, c_2)$. How easy this is would depend upon the complexity of the problem and since we are seeking a general numerical procedure we shall not pursue the matter further. Instead consider the problem of finding a suitable approximation of p_γ.

The approximation of probability density functions is a subject which has a long and glorious history. A suitable introduction to the subject can be found in "Mathematical Methods of Statistics", by H. Cramer.[1] One method of approximating a probability density function is by what is called an Edgeworth Series. For p_γ this has the form:-

$$p_\gamma(x) \simeq \frac{\phi}{\sigma}\left\{\frac{x-m}{\sigma}\right\}\left\{1 + \frac{\mu_3}{3!\sigma^3} H_3\left(\frac{x-m}{\sigma}\right)\right.$$

$$\left. + \frac{1}{4!}\left(\frac{\mu_4}{\sigma^4} - 3\right) H_4\left(\frac{x-m}{\sigma}\right) + \frac{10}{6!}\left(\frac{\mu_3}{\sigma^3}\right)^2 H_6\left(\frac{x-m}{\sigma}\right)\right\}$$

1. H. Cramer Mathematical Methods of Statistics (16)

where m is the mean p_γ - $m = \int x\, p_\gamma(x)\, dx$,

σ^2 is the variance of p_γ - $\sigma^2 = \int x^2 p_\gamma(x)\, dx - m^2$,

and μ_v is the v^{th} central moment - $\mu_v = \int (x-m)^v p_\gamma(x)\, dx$,

and where H_m denotes Hermite polynomials of degree m. In particular:-

$$H_3(t) = t^3 - 3t$$

$$H_4(t) = t^4 - 6t^2 + 3$$

$$H_6(t) = t^6 - 15t^4 + 45t^2 - 15.$$

Also, $\Phi(t) = \exp\left[-t^2/2\right]$.

The advantage of the Edgeworth expansion is that it approximates $p_\gamma(x)$ by a Gaussian function, plus some corrections. As we have remarked previously, in an infinite horizon problem this is a desirable attribute. Also the Edgeworth series appears to offer the most efficient approximation in terms of the number of terms required (see Cramer).

Thus we propose to approximate p_γ by its Edgeworth expansion. This means that we express p_γ in terms of m, σ, μ_3 and μ_4. For a particular control law,

$$u = \gamma(y) = c_1 y + c_3 y^3$$

we can compute the stationary values of m, σ, μ_3 and μ_4 using an iterative procedure similar to Picard's method of solving differential equations. At the k^{th} iterative we have m_r, σ_r, μ_3^r, and μ_4^r. This enables us to

approximate p_{γ}^{r} by the Edgeworth expansion. We then use the state transition equation to compute new values - $m_{r+1}, \sigma_{r+1}, \mu_3^{r+1}$, and μ_4^{r+1}. For example,

$$m_{r+1} = \int x \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta) p_{\gamma}^{r}(\theta)\ d\theta\ dy\ dx$$

The procedure is repeated until stability in the parameters - m, σ, μ_3 and μ_4 is detected. It will be appreciated that evaluating triple integrals in R^3 of the type above represents an onerous task even for the computer.

Having computed the stationary p_{γ} corresponding to a particular $\{c_1\ c_3\}$, we can at last compute $J(c_1\ c_3)$:-

$$J(c_1\ c_3) = \iint L(x, \gamma(y))\ p(y \mid x)\ p_{\gamma}(x)\ dx\ dy$$

This calculation represents the evaluation of a double integral in R^3.

We now have all we need to calculate the optimum value of $(c_1\ c_3)$ numerically. For a particular $(c_1\ c_3)$ we can compute the corresponding $J(c_1, c_3)$. We can employ a direct search procedure to alternatively find improvements to c_1 and c_3. Notice that the direct search procedure will require numerous evaluations of J along a particular search direction in order to determine which way to proceed along the search direction, to find a suitable step size, to determine the interval of uncertainty and successively refine this interval by quadratic interpolation, and

to determine the final value of J along the search direction. This represents a considerable task of computation because of the four triple - and one double - integration involved in calculating J.

A computer program was written to examine the effectiveness of the technique. It was tested on our familiar linear scalar system. For $\sigma_v = \sigma_w = 1$, the control law

$$u = \gamma(y) = -0.411\, y + 0\, y^3$$

was computed, as we would expect! However, the significant point is that the numerical procedure is extremely expensive in computer time and this expense arises from the computation of J requiring the evaluation of a number of complex integrals. Thus if the procedure is to be useful as a design aid, an examination of the integration technique would be profitable. In the computer program mentioned, a simple Simpson procedure was used with the limits of integration set at -5σ, where σ is the variance of the appropriate probability density function - (ie σ_v, σ_w, or σ_γ). Finally notice that in evaluating the optimal control law for our scalar system we did not exploit any of the properties of the system such as Gaussianity, linearity, and so on. What is more, the control law actually determined was:-

$$u = \gamma(y) = c_0 + c_1 y + c_2 y^2 + c_3 y^3,$$

This concludes our examination of numerical procedures. In the next section we shall discuss ϵ- approximation techniques.

7.4.4.3 Approximation by Perturbation Analysis

In Chapter 6 the technique of perturbation analysis was discussed at some length. Recall that the essence of the technique, as applied to the problems we have here, is to introduce a small parameter ϵ into the observation equation:-

$$y = x + \epsilon v$$

Following this, we then express some necessary condition for optimality as a Taylor series expansion about $\epsilon = 0$. By comparing coefficients we then derive a sequence of necessary conditions, each one corresponding to a coefficient of ϵ^i - i = 0,1,2 ----.

Suppose we wish to apply Perturbation Analysis to our zero-memory problem. It will be appreciated from the general analysis presented previously that we have available a number of approaches:-

(i) We can attempt to find a solution via the Successive Approximation method of Dynamic Programming.

(ii) We can attempt to find a solution via an optimality Equation such as that of Theorem 7.6, namely

$$\lambda + V(p\ ) = \inf_{\gamma} \left\{ E(L(x, \gamma(y)) + V(p^{x}) \right\}$$

Where:-

$$p'(x) = \iint p(x \mid \theta, \gamma(y))\ p(y \mid \theta)\ p\ (\theta)\ d\theta\ dy$$

(iii) We can attempt to solve the problem directly, formulated as a constrained functional minimisation problem; namely:-

Find γ_0 such that:-

$$E_{P_{\gamma_0}} (L(x, \gamma_0(y))) = \min_{\gamma} \left\{ E_{P_{\gamma}} (L(x, \gamma(y))) \right\}$$

for all γ such that:-

$$p_{\gamma}(x) = \iint p(x \mid \theta, \gamma(y))\, p(y \mid \theta)\, p_{\gamma}(\theta)\, d\theta\, dy$$

Approach (i), successive approximation via dynamic programming is an attractive possibility. We have seen in Chapter 6 how Perturbation Analysis can be applied to a finite time zero-memory problem to determine $o(\epsilon^2)$ approximations to the optimal solution. The example considered in Chapter 6 was the Witsenhausen problem. With some increase in complexity we could extend the technique to a more general N-stage process and observe the behaviour of our approximate solution as $N \to \infty$. In view of the extensive presentation of the Perturbation Analysis of finite time problems already presented in Chapter 6, we shall not pursue this matter further here (but see Section 5 of this chapter). Let us remark that the method of successive approximation represents a technique which in theory is straightforward and such is appealing. Nevertheless we can expect that the technique will lead to some very complex algebra.

Approach (ii), the determination of a solution via an Optimality Equation suffers from the usual drawback to such an approach - that is, we do not know what the function $V(p)$ looks like. This however need

not necessarily be a disadvantage. Whether or not it is will depend on the form of the optimality equation when it is expanded about $\epsilon = 0$. Suppose $V(p, \epsilon)$ is expanded as $V_0(p) + \epsilon V_1(p) + \frac{2}{2!} V_2(p) + ---$. Then it may turn out that we obtain a set of necessary conditions such that $V_0(p)$ can be easily obtained, and from then on $V_k(p)$ can be easily found from $V_{k-1}(p)$. Indeed, this is the whole idea of Perturbation Analysis! **However, the optimality equation does not yield a necessary condition in a form ideally suited for perturbation analysis; the matter will not be pursued here.**

Finally approach (iii), solution of a constrained functional minimisation problem looks like perhaps the most straightforward approach of all. As we have seen in the previous sections, it is a convenient formulation from which to derive the optimal linear control law. However if we generalise our problem to that of finding the optimal control law over the class of admissable functions, perturbation analysis leads to a vast array of subproblems not much simpler than our original problem.

Consider our familiar linear scalar zero-memory problem. Further analysis of the Ergodic version of this problem is perhaps only of academic interest. It does nevertheless provide an opportunity to illustrate the difficulties of applying perturbation analysis to the Ergodic Problem.

Suppose that the optimal solution to the problem is characterised by the control law $\gamma^\circ(y, \epsilon)$ and the corresponding stationary probability density function $p^\circ(x, \epsilon)$, and further suppose that:-

$$\gamma^{\circ}(y, \epsilon) = \gamma_0(y) + \epsilon \gamma_1(y) + \epsilon^2 \gamma_2(y) + \text{-----}$$

$$p^{\circ}(x, \epsilon) = p_0(x) + \epsilon p_1(x) + \epsilon^2 p_2(x) + \text{-----}$$

Let us now expand the integral constraint equation:-

$$p^{\circ}(x,\epsilon) = \iint \frac{\phi}{\sigma_w}\left\{\frac{x-(\theta+\gamma^{\circ}(y,\epsilon))}{\sigma_w}\right\}\frac{\phi}{\sigma_v}\left\{\frac{y-\theta}{\sigma_v}\right\} p^{\circ}(\theta,\epsilon)\, d\theta\, dy$$

as a power series in ϵ. The algebra is a little messy but the technique itself is simple and has been fully explored in Chapter 6. We obtain, by comparing coefficients,

$$p_0(x) = \int \frac{\phi}{\sigma_w}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\} p_0(y)\, dy,$$

$$p_1(x) = \int \frac{\phi}{\sigma_w}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\}\frac{\gamma_1(y)}{\sigma_w}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\} p_0(y)\, dy$$

$$+ \int \frac{\phi}{\sigma_w}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\} p_1(y)\, dy$$

$$p_2(x) = \frac{1}{2}\int \frac{\phi}{\sigma_w}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\}\left[\left(\frac{x-y-\gamma_0(y)}{\sigma_w}\right)\frac{2\gamma_2(y)}{\sigma_w}\right.$$

$$\left. + \frac{\sigma_v^2+\gamma_1^2(y)}{\sigma_w^2} + \left(\frac{x-y-\gamma_0(y)}{\sigma_w^2}\right)^2\left(\frac{\sigma_v^2+\gamma_1^2(y)}{\sigma_w^2}\right)\right] p_0(y)\,dy$$

$$+ \int \frac{\phi}{\sigma_w}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\}\left\{\frac{\sigma_v^2 p_0'(y)}{\sigma_w} + \frac{\gamma_1(y)p_1(y)}{\sigma_w}\right\}dy$$

$$+ \frac{1}{2}\int \frac{\phi}{\sigma_w}\left\{\frac{x-y-\gamma_0(y)}{\sigma_w}\right\}\left\{\sigma_v^2 p_0''(y) + 2p_2(y)\right\}dy$$

Although superficially the above equations do not look much of an improvement over the original constraint equation, two properties should be noted:-

i) we have a single integral, instead of a double integral.

ii) to some extent the effects of the control function have been separated, that is:-

$$p_0(x) = F_0(x, \gamma_0, p_0)$$

$$p_1(x) = F_1(x, \gamma_1, \gamma_0, p_0) + G_1(x, \gamma_0, p_1)$$

$$p_2(x) = F_2(x, \gamma_2, \gamma_1, \gamma_0, p_0) + G_2(x, \gamma_1, \gamma_0, p_0, p_1)$$

$$+ H_3(x, \gamma_0, p_0, p_2), \text{ and so on.}$$

Whether these properties are useful or not depends on our necessary condition for optimality. Let us see how we can determine this necessary condition. As we have mentioned before, basically we have a constrained functional minimisation problem. Consequently we shall develop our necessary condition using a variational approach. The treatment will be purely formal. For a more rigorous treatment of the topic, the reader can consult one of the many textbooks on the subject of functional optimisation, for example Luenberger[1]. Our formal approach is adequate for our purposes here.

Recall that our problem is to find that function $\gamma^\circ(y)$ such that γ° minimises the functional $V(p, \gamma)$, where:-

$$V(p, \gamma) = E\left[x^2 + \gamma^2(y)\right] = \iint (x^2 + \gamma^2(y))\, p(y \mid x)\, p_\gamma(x)\, dx,$$

subject to the equality constraint:-

$$p_\gamma(x) = \iint p(x \mid \theta, \gamma(y))\, p(y \mid \theta)\, p_\gamma(\theta)\, d\theta\, dy$$

1. 'Optimisation by Vector Space Methods' D.G.Luenberger. J.Wiley 1969

This is simply an optimisation problem involving the minimisation of a functional $V(p,\gamma)$ with respect to functions p,γ, subject to the functional constraint $G(p,\gamma,x) = 0$, for all x.

Suppose the optimal solution is p_0, γ_0, and consider variations δp and $\delta\gamma$ about p_0 and γ_0 respectively. For optimality, we have

$$\frac{dV}{d\alpha}(p_0+\alpha\delta p, \gamma_0+\alpha\delta\gamma)\Big|_{\alpha=0}$$

$$= \frac{d}{d\alpha}\left\{\iint (x^2+(\gamma_0(y)+\alpha\delta\gamma(y))^2 p(y|x)\left\{p_0(x)+\alpha\delta p(x)\right\}dy\,dx\right|_{\alpha=0}$$

Thus we obtain:-

$$\int V_\gamma(p_0,\gamma_0,y)\,\delta\gamma(y)\,dy + \int V_p(\gamma_0,x)\,\delta p(x)\,dx = 0 \qquad \text{——————} \quad (1)$$

where the operators V_γ and V_p are as follows:-

$$V_\gamma(p_0,\gamma_0,y) = \int 2\gamma_0(y)\; p(y\,|\,x)\; p_0(x)\; dx$$

$$V_p(\gamma_0,x) \quad = x^2 + \int \gamma_0^2(y)\; p(y\,|\,x)\; dy$$

Next consider the constraint equation. Again we examine the effect of variations $\delta\gamma$ and δ_p, and we compute:-

$$\left.\frac{dG(\gamma_0+\alpha\delta\gamma,\ p_0+\alpha\delta p)}{d\alpha}\right|_{\alpha=0}$$

$$=\left.\frac{d}{d\alpha}\left\{\iint p(x|\theta,\ \gamma_1(y)+\delta\gamma(y))\,p(y|\theta)\,(p_0(x)+\alpha\delta p(x))\,dx\,dy-p(x)\right\}\right|_{\alpha=0}$$

$$=\left.\frac{d}{d\alpha}\left\{\iint\frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma_0(y)+\alpha\delta\gamma(y)}{\sigma_\omega}\right\}\frac{\phi}{\sigma_v}\left\{\frac{y-\theta}{\sigma_v}\right\}(p_0(x)+\alpha\delta p(x))\,dx\,dy-p(x)\right\}\right|_{\alpha=0}$$

$$=0$$

when we introduce the particular details of our problem. Thus we obtain the following variational equation:-

$$\int G_\gamma(x,y,\gamma_0,p_0)\,\delta\gamma(y)\,dy+\int G_p(x,\theta,\gamma_0)\,\delta p(\theta)\,d\theta=0, \quad -(2)$$

where the operators G_p and G_γ are defined as follows:-

$$G_\gamma(x,y,\gamma_0,p_0)=\int\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_\omega^2}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_\omega}\right\}\frac{\phi}{\sigma_v}\left\{\frac{y-\theta}{\sigma_v}\right\}p_0(\theta)\,d\theta$$

and

$$G_p(x,\theta,\gamma_o) = \int \frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma_o(y))}{\sigma_\omega}\right\}\frac{\phi}{\sigma_v}\left\{\frac{y-\theta}{\sigma_v}\right\}dy - \delta(x-\theta)$$

where $\delta(\cdot)$ is the Dirac delta-function. Suppose we define the function $\delta g(x)$ as follows; according to (2)

$$\delta g(x) = \int G_p(x,\theta,\gamma_o)\delta p(\theta)d\theta = -\int G_\gamma(x,y,p_o)\delta\gamma(y)dy$$

Thus $\delta(y) = -\int G_\gamma^{-1}(y,x,\gamma_0,p_0)\,\delta g(x)\,dx,$

and $\delta p(\theta) = \int G_p^{-1}(\theta,x,\gamma_0)\,\delta g(x)\,dx$

We can now re-write equation (1) in the form:-

$$\int V_\gamma(p_o,\gamma_o,y)\,G_\gamma^{-1}(y,x,\gamma_o,p_o)\,\delta g(x)\,dx\,dy$$

$$-\int V_p(\gamma_o,x)G_p^{-1}(x,\psi,\gamma_o)\delta g(\psi)\,d\psi\,dx = 0$$

Now this equation must be satisfied for any $\delta g(x)$, which implies that:-

$$\int V_\gamma(p_o,\gamma_o,y)G_\gamma^{-1}(y,x,\gamma_o,p_o)dy = \int V_p(\gamma_o,\theta)G_p^{-1}(\theta,x,\gamma_o)d\theta = -\lambda(x)\ ,\ \text{say}$$

Thus we have,

$$\int\int V_\gamma(p_o,\gamma_o,y)G_\gamma^{-1}(y,x,\gamma_o,p_o)dy\,G_\gamma(x,y',\gamma_o,p_o)dx$$

$$= -\int \lambda(x)\,G_\gamma(x,y',\gamma_o,p_o)dx$$

and

$$\int V_p(\gamma_o,\theta)\, G_p^{-1}(\theta,x,\gamma_o)\, d\theta\, G_p(x,\theta',\gamma_o)\, dx$$

$$= -\int \lambda(x)\, G_p(x,\theta',\gamma_o)\, dx$$

Now $\int G_\gamma^{-1} G_\gamma \, dx = \delta(y-y')$

and $\int G_p^{-1} G_p \, dx = \delta(\theta-\theta')$

Thus we obtain the equations:-

$$V_\gamma(p_o,\gamma_o,y') + \int \lambda(x)\, G_\gamma(x,y',\gamma_o,p_o)\, dx = 0 \quad \text{— (3)}$$

$$V_p(\gamma_o,\theta') + \int \lambda(x)\, G_p(x,\theta',\gamma_o)\, dx = 0 \quad \text{— (4)}$$

together with the original constraint equation:-

$$G(p_0,\gamma_0) = 0 \qquad -5$$

The above equations are the functional form of the well known Lagrange multiplier equations. We have the unknowns γ_0, p_0, and λ and three equations to solve. Notice that since we have a functional constraint, the multiplier λ is a function. Having derived the above necessary conditions, we now have the problem of finding functions λ, p_0, γ_0 which satisfy equations (3), (4), and (5). Let us write the equations in the form appropriate for our simple scalar zero memory problem.

Equation (3) above becomes:-

$$\int 2\gamma_0(y)\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-x}{\epsilon\sigma_v}\right\}p_0(x)dx + \iint \lambda(x)\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_w^2}\right\}\frac{\phi}{\sigma_w}\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_w}\right\}\cdot$$

$$\cdot\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-\theta}{\epsilon\sigma_v}\right\}p_0(\theta)\,d\theta\,dx = 0, \text{ for all } y \qquad \text{——— (6)}$$

Equation (4) above becomes:-

$$\theta^2 + \int \gamma_0^2(y)\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-\theta}{\epsilon\sigma_v}\right\}dy + \iint \lambda(x)\cdot\left\{\frac{\phi}{\sigma_w}\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_w}\right\}\cdot\right.$$

$$\left.\cdot\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-\theta}{\epsilon\sigma_v}\right\}dy - \delta(x-\theta)\right\}dx = 0, \text{ for all } x \qquad \text{——— (7)}$$

Finally the constraint equation (5), as we have seen is:-

$$\iint \frac{\phi}{\sigma_w}\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_w}\right\}\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-\theta}{\epsilon\sigma_v}\right\}p_0(\theta)\,d\theta\,dy - p_0(x) = 0,$$

$$\text{for all } x \qquad \text{——— (8)}$$

First notice that our optimal linear solution derived earlier in Section 3, is a solution of the above equations. This is demonstrated below.

Assume that the optimal control law is linear in structure, that is:-

$$\gamma^{o}(y) = b\,y,$$

and that the Lagrange multiplier function $\lambda(x)$ is a quadratic function, that is:-

$$\lambda(x) = \lambda' x^2 + \lambda_0$$

Under the assumption of a linear control, p_0 will be Gaussian, that is:-

$$p_0(x) = \frac{\Phi}{\sigma_0}\left\{\frac{x}{\sigma_0}\right\}$$

According to equation (8) above, we have:-

$$\sigma_0^2 = (1+b)^2\sigma_0^2 + b^2\epsilon^2\sigma_v^2 + \sigma_w^2 \qquad \text{——— (9)}$$

Next consider equation (6), from which we obtain:-

$$b = \frac{-\lambda'}{1+\lambda'}\,\frac{\sigma_0^2}{\sigma_0^2 + \epsilon^2\sigma_v^2} \qquad \text{——————— (10)}$$

Finally, consider equation (7), by comparing the coefficients of α^2, we obtain:-

$$\lambda' = (1+b^2) + \lambda'(1+b)^2 \qquad \text{——————— (11)}$$

But these are precisely the equations we derived earlier in Section 3.3. Some comment is required regarding the constant term λ_0. To satisfy equation (7), requires that $\lambda_0 = \infty$. This is not surprising, since inspection of equation (7) reveals that $\lambda(x)$ is assuming the role of the cost-to-infinity function, that is:-

$$E\left[\lambda(x)\right] = \sum_{i=0}^{\infty} (x_i^2 + u_i^2)$$

Our purpose here is to explore the application of perturbation analysis to the general ergodic problem (although for clarity and to ease comparison we are basing our development on the simple linear scalar example). With this end in mind, we have introduced the parameter ϵ into the equations (6), (7) and (8) above. The procedure now is quite straightforward. We parameterise the unknowns γ_0, p_0, and λ with the parameter ϵ and expand each of the equations (6), (7) and (8) as power series about $\epsilon = 0$. We have already expanded equation (8). The expansion of equations (6) and (7) up to terms involving ϵ^2 (say), results in the traditional tedious algebra but is nevertheless practical. We shall not present the expansions here because we are concerned mainly with the principle of the technique. As an illustration of the complexity of the expansions which are encountered, consider the expansion of G_γ. For clarity we shall drop the subscript '0' indicating optimality. Suppose that:-

$$\gamma(y,\epsilon) = \gamma_0(y) + \epsilon\gamma_1(y) + \epsilon^2\gamma_2(y) + o(\epsilon^2),$$

$$p(x,\epsilon) = p_0(x) + \epsilon p_1(x) + \epsilon^2 p_2(x) + o(\epsilon^2),$$

and $$\lambda(x,\epsilon) = \lambda_0(x) + \epsilon\lambda_1(x) + \epsilon^2\lambda_2(x) + o(\epsilon^2)$$

Now recall that:-

$$G_\gamma(x,y,p,\epsilon)$$

$$= \int \left\{\frac{x-(\theta+\gamma(y,\epsilon))}{\sigma_\omega^2}\right\} \frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma(y,\epsilon))}{\sigma_\omega}\right\} \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-\theta}{\epsilon\sigma_v}\right\} p(\theta,\epsilon)\, d\theta$$

If we make the substitution $\theta = y + \epsilon\sigma_v s$, we obtain:-

$$G_\gamma(x,y,p,\epsilon)$$

$$= \int \left\{\frac{x-(y+\epsilon\sigma_v s+\gamma(y,\epsilon))}{\sigma_\omega^2}\right\} \frac{\phi}{\sigma_\omega}\left\{\frac{x-(y+\epsilon\sigma_v s+\gamma(y,\epsilon))}{\sigma_\omega}\right\} \phi(s)\, p(y+\epsilon\sigma_v s,\epsilon)\, ds$$

Given this expression for G_γ, we now compute the coefficients of ϵ^i by computing the expression for $\left.\dfrac{d^i G_\gamma}{d\epsilon^i}\right|_{\epsilon=0}$

Below we present the expression for $\left.\frac{d^2 G_\gamma}{d\epsilon^2}\right|_{\epsilon=0}$. We obtain:-

$$\left.\frac{d^2 G_\gamma}{d\epsilon^2}\right|_{\epsilon=0} = \frac{2\gamma_2(y)}{\sigma_w^2}\frac{\phi}{\sigma_w}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_w}\right\}p_0(y)$$

$$+3\left(\frac{\sigma_v^2+\gamma_1^2(y)}{\sigma_w^2}\right)\left(\frac{x-(y+\gamma_0(y))}{\sigma_w^2}\right)\frac{\phi}{\sigma_w}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_w}\right\}p_0(y)$$

$$+2\left(\frac{\gamma_1(y)p_1(y)+\sigma_v^2 p_0'(y)}{\sigma_w^2}\right)\frac{\phi}{\sigma_w}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_w}\right\}$$

$$+2\left(\frac{x-(y+\gamma_0(y))}{\sigma_w^2}\right)^2\gamma_2(y)\frac{\phi}{\sigma_w}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_w}\right\}p_0(y)$$

$$+\frac{(x-(y+\gamma_0(y)))^3}{\sigma_w^4}\frac{(\sigma_v^2+\gamma_1^2(y))}{\sigma_w^2}\frac{\phi}{\sigma_w}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_w}\right\}p_0(y)$$

$$+2\frac{(x-(y+\gamma_0(y)))^2}{\sigma_w^4}(\sigma_v^2 p_0'(y)+\gamma_1(y)p_1(y))\frac{\phi}{\sigma_w}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_w}\right\}$$

$$+\left(\frac{x-(y+\gamma_0(y))}{\sigma_w^2}\right)(\sigma_v^2 p_0''(y)+2p_2(y))\frac{\phi}{\sigma_w}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_w}\right\}$$

An expression of almost similar complexity will be obtained for $\left.\frac{d^2 V_\gamma}{d\epsilon}\right|_{\epsilon=0}$ and indeed for all the quantities required to obtain an $o(\epsilon^2)$ approximation to the optimal control law. The point we wish to emphasise in presenting the above expression for $\left.\frac{d^2 G_\gamma}{d\epsilon^2}\right|_{\epsilon=0}$ is that the perturbation technique can involve some very tedious algebra. For this reason we have not gone into detail here. Nevertheless the technique is applicable and for certain problems, it may be the only analytical technique which is available to the system designer. Notice that essentially the technique 'linearises' the problem. In other words, the awkward item in the expression for G is the term:-

$$\frac{\Phi}{\sigma_w}\left\{\frac{x-(\theta+\gamma(y,\epsilon))}{\sigma_w}\right\}$$

The perturbation technique enables us to use instead the term:-

$$\frac{\Phi}{\sigma_w}\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_w}\right\}$$

where the control $\gamma_0(y)$ is the generating solution and can be found quite easily via a number of methods. For example, we can solve directly the perfectly observable version ($\epsilon = 0$) of our original problem. Alternatively, we can evaluate the necessary conditions derived above at

$\epsilon = 0$ and solve the resulting simultaneous equations. Let us briefly examine this latter approach in order to gain an impression of the problem of finding higher order terms in the optimal control - $(\gamma_1(y), \gamma_2(y)$, and so on.

According to our definitions of $V_\gamma, G_\gamma, V_p, G_p$, we have at $\epsilon = 0$:-

$$V_\gamma(p,\gamma,y)\Big|_{\epsilon=0} = \int 2\gamma(y,\epsilon)\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-x}{\epsilon\sigma_v}\right\}p(x,\epsilon)\,dx\Big|_{\epsilon=0} = 2\gamma_0(y)p_0(y)$$

$$V_p(\gamma,x)\Big|_{\epsilon=0} = x^2 + \int \gamma^2(y,\epsilon)\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-x}{\epsilon\sigma_v}\right\}dy\Big|_{\epsilon=0} = x^2+\gamma_0^2(x)$$

$$G_\gamma(x,y,p,\gamma)\Big|_{\epsilon=0} = \int \left\{\frac{x-(\theta+\gamma(y,\epsilon))}{\sigma_\omega^2}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma(y,\epsilon))}{\sigma_\omega}\right\}\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-\theta}{\epsilon\sigma_v}\right\}p(\theta,\epsilon)\,d\theta\Big|_{\epsilon=0}$$

$$= \left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega^2}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega}\right\}p_0(y)$$

$$G_p(x,\theta,\gamma) = \frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_\omega}\right\}\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-\theta}{\epsilon\sigma_v}\right\}dy - \delta(x-\theta)\Big|_{\epsilon=0}$$

$$= \frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma_0(\theta))}{\sigma_\omega}\right\} - \delta(x-\theta)$$

Thus, our problem is to solve for γ_0, p_0, λ_0 the following equations:-

$$V_\gamma + \int \lambda G_\gamma\, dx\Big|_{\epsilon=0}$$

$$= 2\gamma_0(y)p_0(y) + \int \lambda_0(x)\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega^2}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega}\right\}p_0(y)\,dx = 0$$

that is,

$$2\gamma_0(y) + \int \lambda_0(x)\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega^2}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma_0(\theta))}{\sigma_\omega}\right\}dx = 0 \quad \text{——— (12)}$$

Also:-

$$V_p + \int \lambda G_p\, dx\Big|_{\epsilon=0}$$

$$= \theta^2 + \gamma_0^2(\theta) + \int \lambda_0(x)\left\{\frac{\phi}{\sigma_\omega}\left(\frac{x-(\theta+\gamma_0(y))}{\sigma_\omega}\right) - \delta(x-\theta)\right\}dx = 0$$

that is,

$$\theta^2 + \gamma_0^2(\theta) + \int \lambda_0(x)\frac{\phi}{\sigma_\omega}\left\{\frac{x-(\theta+\gamma_0(y))}{\sigma_\omega}\right\}dx - \lambda_0(\theta) = 0 \quad \text{——— (13)}$$

Together with the constraint equation:-

$$G(\gamma(\epsilon),\rho(\epsilon),\epsilon)\Big|_{\epsilon=0}$$

$$= \int \frac{\phi}{\sigma_\omega}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega}\right\} p_0(y)\,dy - p_0(x) = 0 \qquad \text{——— (14)}$$

These equations can be solved as follows:-

i) Differentiate (13) with respect to θ giving (A)

ii) Multiply (12) by $1+\gamma_0'(y)$ giving (B)

iii) Put $\theta=y$ in (A) and subtract from (B)

This gives us:-

$$\gamma_0(y) = y - \frac{\lambda_0'(y)}{2} \qquad -(C)$$

iv) Use the identity:-

$$\int \lambda(x)\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega^2}\right\}\frac{\phi}{\sigma_\omega}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega}\right\}dx$$

$$= \int \lambda'(x)\frac{\phi}{\sigma_\omega}\left\{\frac{x-(y+\gamma_0(y))}{\sigma_\omega}\right\}dx$$

(integration by parts) in (12) and substitute (C) above.

This gives:-

$$\int \gamma_0(x) \frac{\phi}{\sigma_\omega} \left\{ \frac{x - (y + \gamma_0(y))}{\sigma_\omega} \right\} dx = y + 2\gamma_0(y) \qquad \text{—— (D)}$$

This equation is satisfied with $\gamma_0(y) = ay$, where a satisfies the equation:-

$$a^2 - a - 1 = 0 \qquad \text{— (E)}$$

In evaluating this quadratic, the negative square root should be taken to give a stable gain.

(v) Substituting the solution $\gamma_0(y) = ay$ into (C) and integrating we obtain,

$$\lambda_0(y) = y^2 \left[1 - a\right] + C,$$

where C is a constant of integration.

If we make the substitution b = 1-a, we obtain the following expression for $\lambda_0(y)$:-

$$\lambda_0(y) = by^2 + C,$$

where b satisfies the equation:-

$$b^2 - b - 1 = 0 \qquad \text{——— (F)}$$

In this case, in evaluating the quadratic the positive square root should be taken.

(vi) Finally, we obtain for $\gamma_0(y)$:-

$$\gamma_0(y) = y - by = (1-b)y$$

But, if b satisfies (F), we have $(1-b) = \frac{-b}{1+b}$ Thus:-

$$\gamma_0(y) = \frac{-b}{1+b}\, y, \qquad \text{——— (G)}$$

and $$\lambda_0(y) = b^2y^2 + C, \qquad \text{——— (H)}$$

where $b^2 - b - 1 = 0$

Notice that we have not used the constraint equation (14) in finding γ_0 and δ_0. This should not concern us since for $\epsilon = 0$, we would expect that $\gamma_0(y)$ is a function of the 'state' $y(=x)$ and not the hyperstate $p_0(x)$. The corresponding p_0 is Gaussian with variance σ_0^2 defined by:-

$$\sigma_0^2 = (1+a)^2\,\sigma_0^2 + \sigma_w^2 \qquad \text{——— (I)}$$

Notice also that:-

(i) The generating solution above (equations (G), (H) and (I)) corresponds to the linear solution which we proved earlier was optimal (equations (9), (10) and (11)). The generating solution, as we

would expect, can be derived from the optimal linear solution by simply putting $\epsilon = 0$.

(ii) The generating solution can also be derived quite simply by using the successive approximation method of dynamic programming. In fact the procedure is somewhat simpler than the one adopted here. As we remarked above however, the procedure adopted here was selected in order to illustrate the use of the Lagrange multiplier type necessary conditions (equations (6), (7) and (8)) in the context of a perturbation analysis of the problem.

Having found the generating solution for the problem, the next step is to compare the coefficients of ϵ in equations (6), (7) and (8) and thus obtain simultaneous equations to solve for γ_1, p_1, and λ_1. Although the task of writing down these equations might be quite laborious we would expect that their ultimate solution will be quite straightforward, relying heavily upon the generating solution. Similar remarks apply to determination of γ_2, p_2, λ_2 and so on. It is not proposed to pursue the matter further here. Our intention has been merely to examine the technique of perturbation analysis as it applies to the ergodic problem. For clarity and for ease of comparison, we have presented this examination in terms of our familiar linear scalar example.

Despite their complexity, our necessary conditions:-

$$V_\gamma + \int \lambda G \, dx = 0$$

$$V_p + \int \lambda G_p \, dx = 0$$

$$G = 0$$

are still appropriate and might prove useful in a particular problem. Indeed the author has never seen these necessary conditions proposed as an approach to the solution of any ergodic problem, zero-memory or otherwise. This is perhaps because the bulk of the literature concerned with stationary problems considers problems for which more convenient techniques exist - notably the successive approximation method of dynamic programming. For the zero-memory case however, we cannot solve the finite time version of the problem and hence we have to adopt a more direct approach to finding a solution. This concludes our discussion of the perturbation approach to the zero-memory stationary problem.

7.4.5 Summary

In this section we have restricted our attention to what we have called the ergodic problem. First of all in Section 4.1 we compared the performance of the best linear zero-memory controller with that of the optimal classical information pattern controller. It was demonstrated that, for our simple scalar problem, there is only a very insignificant difference in performance unless σ_v/σ_w becomes very

large. This is encouraging. Our model of zero-memory control is simulating a real world situation in which the transmission of information from one controller to the next (or one time instant to the next) is impossible or prohibitively expensive. In this case it is reassuring for the designer to know that his linear zero-memory controller performs almost as well as the perfect-memory version.

In Section 4.2 we reported a technique for finding the optimal linear solution for a general multivariable linear system. This technique was originally proposed by Cumming. The scalar development of Section 3.5 was arrived at independently of Cumming's work. Indeed the development there was based on Lagrange multiplier techniques rather than Cumming's cost-to-infinity matrices.

In Section 4.3 we discuss some fruitless attempts to show that the best linear controller for the zero-memory ergodic problem is as good as any other. In summary, it seems that every necessary condition for optimality which we propose will be satisfied by the best linear controller. The difficulty rests in showing that this is the only form of controller which satisfies the necessary condition. This situation is very similar to that which exists for the Witsenhausen problem, where the linear controller satisfies a necessary condition for optimality but nevertheless is outperformed by a non-linear controller for certain values of the problem parameters. Perhaps then we should search for a counter example which will refute once and for all claims that the optimal control for the zero-memory ergodic is linear.

A number of approximation methods are presented in Section 4.4 using a familiar scalar problem as an example. Three methods were considered; the first involved a moment formulation of the problem, the second involved an Edgeworth series approximation to the stationary probability density function; and the third involved perturbation analysis. All three methods are characterised by the tremendous intricacy of their application. Whether coping with this intricacy is justified will depend on the particular problem with which the designer is confronted. And there is always the possibility of automating the algebra in some way (FORMAC).

One final remark concerning this section is necessary. We have concentrated primarily on the ergodic problem. However, as we have seen, the ergodic problem is simply a special case of the average cost per period problem and our results can be extended to this latter problem (see Theorem 7.8) in some cases. In the next section we shall consider that other example of a stationary problem - the discount problem.

7.5 Further Examination of the Discount Problem

In Section 3.3 we attempted to find the optimal linear stationary control for the discount problem. We were unsuccessful and all that we could do was to present a necessary condition for the determination of the optimal static linear controller. We also remarked in Section 3.3 that we should not be too surprised that we had encountered some difficulty since the determination of the optimal linear controller for the finite time problem is also somewhat intractable. In this section we shall use the successive approximation method of dynamic programming to identify the nature of the difficulty and we shall then employ perturbation analysis to generate approximate linear stationary control laws. The application of the method of successive approximation has the advantage that it not only provides us with solutions to the discount problem, but it also gives us solutions to the finite time version of the problem (In Chapter 6 we only considered a perturbation approach to Witsenhausen's example).

7.5.1 Successive Approximation Method of Dynamic Programming

Recall that the discount problem we wish to solve is as follows:-

Given the process:-

$$x_{n+1} = x_n + u_n + w_n,$$

with observations

$$y_n = x_n + v_n,$$

find that sequence of control laws,

$$\gamma_0^{\circ}(y_0,p_0), \gamma_1^{\circ}(y_1,p_1), \gamma_2^{\circ}(y_2,p_2) \quad \text{---} \quad \text{and so on such that:-}$$

$$J(p_0,\gamma_0^{\circ},\gamma_1^{\circ},\gamma_2^{\circ} \text{---}) \leq J(p_0,\gamma_0,\gamma_1,\gamma_2 \text{---}) \text{ for all } \gamma_0,\gamma_1,\gamma_2,$$

where:-

$$J(p_0,\gamma_0,\gamma_1,\gamma_2) \triangleq \lim_{N\to\infty} \left[\sum_{i=0}^{N} \rho^i \, E_{p_i} (x_i^2 + \gamma_i^2(y_i,p_i)) \right]$$

The expectation is with respect to p_i - the probability density funtion of the state x_i at time i. As usual the random variables $\{w_n\}$, $\{v_n\}$, are independent Gaussian sequences with zero mean and variances σ_w^2 and σ_v^2 respectively. The random variable x_0 is Gaussian with zero mean and variance σ_0^2.

As we have seen, the successive approximation method of dynamic programming permits us to interchange the lim and inf in the above formulation and derive the optimal solution of the infinite horizon problem as the limit of the finite time problem. Thus, given some arbitrary functional $V_0^{\circ}(p_0)$, we wish to find funotionals $V_n^{\circ}(p_n)$ such that:-

$$V_n^{\circ}(p_n) = \inf_{\gamma_n} \left\{ E_{p_n} \left[x^2 + \gamma_n^2(y,p_n) \right] + \rho V_{n-1}^{\circ}(p_{n-1}) \right\}$$

where,

$$p_{n+1}(x) = \iint \frac{\phi}{\sigma_\omega} \left\{ \frac{x-(\theta+\gamma_n(y_n,p_n))}{\sigma_\omega} \right\} \frac{\phi}{\sigma_v} \left\{ \frac{y-\theta}{\sigma_v} \right\} d\theta dy$$

We can then assert that,

$$\lim_{n \to \infty} V_n^{\circ}(p_n)\Big|_{p_n=p_0} = J(p_0, \gamma_0^{\circ}, \gamma_1^{\circ}, \gamma_2^{\circ} \text{---}),$$

and we thus have the optimal solution to our zero-memory discount problem.

The details of the dynamic programming analysis are presented in Appendix 5.3 . As usual, we find the optimal control law for n = 1 with little difficulty - it is $\gamma_1^{\circ}(y) = -\rho/1+\rho \; E[x|y]$ as we would expect. It is when we come to evaluate the optimal linear control law for n = 2 that we experience some difficulty. We are lead to the problem of finding $V_2^{\circ}(p_2)$ where

$$V_2^{\circ}(p_2) = \min_k \Big\{ k^2(1+\sigma_z^2) + k^2\sigma_v^2 + \rho(1+\rho)\left[(1+k)^2\sigma_z^2 + k^2\sigma_v^2 + \sigma_w^2\right]$$

$$- \frac{\rho^3}{1+\rho} \cdot \frac{(\sigma_z^2(1+k)^2 + k^2\sigma_v^2 + \sigma_w^2)^2}{\sigma_z^2(1+k)^2 + (k^2+1)\sigma_v^2 + \sigma_w^2} \Big\} + \rho^2\sigma_w^2$$

Clearly it is difficult to find an analytic impression for k which provides the minimum. Expansion of the above expression will reveal that k is determined by a quartic equation. The same difficulty will be encountered if we attempt to find the best linear control law for the finite time problem (put $\rho = 1$ above). We have to resort to an approximation technique at this point. We introduce ϵ into the observation equation in the usual way:-

$$y = x + \epsilon v.$$

The perturbation analysis, which is messy, is presented in Appendix 5.3. It is shown that:-

$$V_n^o(p_n) = a_n\sigma_n^2 + b_n\epsilon^2\sigma_v^2 + \frac{c_n\epsilon^4\sigma_v^4}{\sigma_n^2} + \sum_{i=1}^{n-1}\left(\frac{\epsilon^4\sigma_v^4 f_{ni}}{\sigma_n^2 + \sigma_\omega^2 g_{ni}}\right)$$

$$+ d_n\sigma_\omega^2 + o(\epsilon^4)$$

where

$$a_{n+1} = \frac{1+2\rho a_n}{1+\rho a_n}, \quad a_1 = \frac{1+2\rho}{1+\rho}$$

$$b_{n+1} = \rho b_n + \frac{\rho^2 a_n^2}{1+\rho a_n}, \quad b_1 = \frac{\rho^2}{1+\rho}$$

$$c_{n+1} = \frac{-\rho^2 a_n}{(1+\rho a_n)} \quad , \quad c_1 = \frac{\rho^2}{(1+\rho)}$$

$$f_{n+1,1} = \rho c_n (1+\rho a_n^2)$$

$$f_{n+1,i} = \rho f_{n,i-1} (1+\rho a_n)^2, \quad i = 2, \cdots n$$

$$g_{n+1,1} = (1+\rho a_n)^2$$

$$g_{n+1,i} = (1+g_{n,i-1})(1+\rho a_n)^2, \quad i = 2, \cdots n$$

$$d_{n+1} = \rho(a_n + d_n) \, , \, d_1 = \rho$$

with minimising k given by:-

$$k_{n+1}^{\circ}(\epsilon) = \frac{-\rho a_n}{1+\rho a_n}\left\{1 - \frac{\epsilon^2 \sigma_v^2}{\sigma_{n+1}^2} + \frac{\epsilon^4 \sigma_v^4}{\sigma_{n+1}^4}\right\}$$

$$+ \sum_{i=1}^{n} \frac{\epsilon^4 \sigma_v^4 f_{n+1,i}}{(\sigma_{n+1}^2 + \sigma_w^2 g_{n+1,i})^2} + o(\epsilon^4)$$

Thus the appropriate optimal linear control law γ^{L}_{n+1} is given by:-

$$\gamma^{L}_{n+1}(y_{n+1}) = \frac{-\rho a_n}{1+\rho a_n} E[x_{n+1} | y_{n+1}] + \sum_{i=1}^{n} \frac{e^4 \sigma_v^4 f_{n+1,i}\, y_{n+1}}{(\sigma_{n+1}^2 + \sigma_w^2 g_{n+1,i})} + o(e^4)$$

Recall that for the corresponding classical information pattern problem, the optimal control is:-

$$\gamma^{o}_{n+1}(p(x_{n+1} | I_{n+1})) = \frac{-\rho a_n}{1+\rho a_n} E[x_{n+1} | y_0, y_1, \cdots y_{n+1}; u_0, u_1, \cdots u_n]$$

It would appear by comparison of these two expressions that the term:-

$$\sum_{i=1}^{n} \frac{\epsilon^4 \sigma_v^4 f_{n+1,i}}{(\sigma_{n+1}^2 + \sigma_w^2 g_{n+1,i})^2}$$

appearing in the zero-memory control represents some form of correction term. Unfortunately, a precise interpretation of this correction term is difficult. However, notice the corresponding correction term in the expression for $V_{n+1}^{o}(p_{n+1})$ can be expanded as follows:-

$$\sum_{i=1}^{n} \frac{\epsilon^4 \sigma_v^4 f_{n+1,i}}{\sigma_{n+1}^2 + \sigma_w^2 g_{n+1,i}} = \epsilon^4 \sigma_v^4 \left\{ \frac{\rho c_n}{\sigma_{n+1}^2 (1+\hat{k}_{n+1})^2 + \sigma_w^2} + \frac{\rho^2 c_{n-1}}{(\sigma_{n+1}^2 (1+\hat{k}_{n+1})^2 + \sigma_w^2)(1+\hat{k}_n)^2 + \sigma_w^2} \right.$$

$$\left. + \frac{\rho^3 c_{n-2}}{((\sigma_{n+1}^2 (1+\hat{k}_{n+1})^2 + \sigma_w^2)(1+\hat{k}_n)^2 + \sigma_w^2)(1+\hat{k}_{n-1})^2 + \sigma_w^2} + \cdots \right\} + o(\epsilon^4)$$

$$= \epsilon^4 \sigma_v^4 \left\{ \frac{\rho c_n}{\sigma_n^2} + \rho^2 \frac{c_{n-1}}{\sigma_{n-1}^2} + \rho^3 \frac{c_{n-2}}{\sigma_{n-2}^2} + \cdots \right\} + o(\epsilon^4)$$

where $\hat{k}_{n+1} = \dfrac{-\rho a_n}{1+\rho a_n}$

where the terms $\underline{\sigma}_i^2$ are the estimates made by the controller at n+1 of the variance at i given perfect observation ($\epsilon = 0$).

The algorithm we have presented above represents and $o(\epsilon^4)$ approximation to the optimal linear solution for a finite-time zero-memory discount problem. Clearly, the algorithm can be used without difficulty for the general finite problem simply by putting $\rho = 1$. Let us next examine the asymptotic behaviour of our solution.

7.5.2 Asymptotic Behaviour of Solution

To obtain the $o(\epsilon^4)$ approximation to the infinite-horizon zero-memory discount problem, we simply let $n \to \infty$ in the expressions derived in Appendix 5.3 for $V^{o}_{n+1}(p_{n+1})$ and k^{o}_{n+1}. We obtain:-

$$\lim_{n\to\infty} V^{o}_{n+1}(p_x) = a\sigma_x^2 + b\epsilon^2\sigma_v^2 + c\epsilon^4\frac{\sigma_v^4}{\sigma_x^2} + d\sigma_w^2$$

$$+\epsilon^4\sigma_v^4\rho c\left\{\frac{1}{\sigma_x^2(1+k^2)+\sigma_w^2} + \frac{\rho}{(\sigma_x^2(1+k^2)+\sigma_w^2)(1+k^2)+\sigma_w^2} + \cdots\right\} + o(\epsilon^4)$$

$$\lim_{n\to\infty} k_{n+1}(\epsilon) = \frac{-\rho a}{1+\rho a}\left\{1-\epsilon^2\sigma_v^2-\epsilon^4\frac{\sigma_v^4}{\sigma_x^2}\right\}$$

$$+\epsilon^2\sigma_v^2\rho c\left\{\frac{1}{\sigma_x^2(1+k)^2+\sigma_\omega^2}+\frac{\rho}{(\sigma_x^2(1+k)^2+\sigma_\omega^2)((1+k)^2+\sigma_\omega^2)}\right.$$

$$\left.+\ \cdots\right\}+o(\epsilon^4)$$

where a satisfies $\rho a^2 + a(1-2\rho)-1 = 0$,

$$b = \frac{\rho^2 a^2}{(1+\rho a)(1-\rho)},$$

$$c = \frac{-\rho^2 a^2}{(1+\rho a)},$$

$$d = \frac{+\rho a}{(1-\rho)},$$

and $$k = \frac{-\rho a}{1+\rho a}.$$

Again, it does not seem possible to find convenient expressions for the infinite sums in the above expansions.

7.5.3 Summary

This concludes our attempts to find an $O(\epsilon^4)$ approximation to the optimal linear control law for the zero memory discount problem. The results are typically disappointing. This is perhaps what we would expect, since for small ρ the discount problem will behave much like the finite zero memory problem and as we know from our examination of Witsenhausen's problem, this is a particularly intractable problem. The zero memory discount problem thus remains unresolved. It is not possible here to pursue the matter further but a number of alternative approaches to the problem are worthy of suggestion:-

(i) Expansion as a power series in ρ.

The idea of this approach is to perturb the problem about the discount factor rather than about the observation noise. Clearly, the generating solution is trivial ($k_0 = 0$). Also the coefficients in $k^o(\rho)$ of ρ and ρ^2 can be obtained without difficulty. However the problem of finding the coefficient of ρ^3 soon assumes unmanageable complexity and preliminary investigation of the approach indicates that it has little to offer.

(ii) Solution of the Optimality Equation

Recall that the optimality equation for our problem is:-

$$V^{o}(\sigma_x^2) = \min_k \left\{ k^2(\sigma_x^2 + \sigma_v^2) + \sigma_x^2 + \rho V^{o}(\sigma_x^2(1+k)^2 + k^2\sigma_v^2 + \sigma_w^2) \right\}$$

A necessary condition for optimality from this equation is

$$k = \frac{-\rho V^{o}_{\sigma_x^2}(\sigma_x^2(1+k)^2 + k^2\sigma_x^2 + \sigma_w^2)}{1 + \rho V^{o}_{\sigma_x^2}(\sigma_x^2(1+k)^2 + k^2\sigma_x^2 + \sigma_w^2)} \cdot \frac{\sigma_x^2}{(\sigma_x^2 + \sigma_v^2)}$$

This expression could be used as the basis for a numerical procedure:-

$$k_{r+1} = \frac{-\rho V^{r}_{\sigma_x^2}(\sigma_x^2(1+k_r)^2 + k_r^2\sigma_v^2 + \sigma_w^2)}{1 + \rho V^{r}_{\sigma_x^2}(\sigma_x^2(1+k_r)^2 + k_r^2\sigma_v^2 + \sigma_w^2)} \cdot \frac{\sigma_x^2}{(\sigma_x^2 + \sigma_v^2)},$$

with $V^{r}(\sigma_x^2) = \lim_{n \to \infty} V_n^{r}(\sigma_x^2)$, where

$$V_n^{r}(\sigma_x^2) = k_r^2(\sigma_x^2 + \sigma_v^2) + \sigma_x^2 + \rho V_{n-1}^{r}(\sigma_x^2(1+k_r)^2 + k^2\sigma_v^2 + \sigma_w^2),$$

with $V_0^{r}(\sigma_x^2) = \sigma_x^2$ say.

This procedure would require quantisation of the functions - $V^{r}(\sigma_x^2)$ and it would also require numerical differentiation. The basis of the technique could probably be justified along the lines of the policy iteration algorithm discussed in Section 4.

As an alternative to the above numerical procedure, it might be possible to develop a perturbation technique along lines similar to those developed for the ergodic problem in the preceding section. The problem here is somewhat simpler in that we do not have a functional minimisation problem. Nevertheless, the problem still quickly becomes unwieldy.

This concludes Section 7. As remarked earlier, the results are disappointing, particularly when we recall that all that we are trying to do is find a linear control. However, given that the discount problem can perhaps provide an acceptable model of the finite time problem, the problem is worth further examination. It might turn out, for example, that the optimal control law is linear and this property might then offer some insight into the elusive solution of the Witsenhausen problem. But at the moment, this is conjecture and it does not seem likely that the optimal control law is linear. And if we embark upon a functional minimisation approach, then we will encounter the same difficulties of second-guessing and so on that appeared in the Witsenhausen problem. Which is where we started.

Whittle and Rudge[1] have considered the problem of finding optimal linear control laws for the zero-memory problem and provided an interesting characterisation of the difficulty. This is discussed in Chapter 9 (See 9.5)

1. Whittle and Rudge "The Optimal Linear Solution of a Symmetric Team Control Problem" (54)

7.6 Conclusions

At last we have reached the end of a long and possibly tedious chapter. Has the effort been worth it? It is clearly the author's belief that the answer is 'yes', as testified by the length of this chapter. Let us then attempt to justify this faith.

First of all, we remarked in the introduction that the long term behaviour of a system might in many cases be of more interest to the designer than its short term (or finite time) behaviour. The validity of this remark will of course depend on the economics of the problem under consideration. We must remember that at best an infinite horizon problem is always an approximation to a real world problem. However given some process which must operate over a long time period, the designer must seek a balance between the cost of implementing some non-stationary control law and the cost of implementing a stationary control law. It is assumed that the non-stationary control law is relatively expensive to implement but will result in superior performance of the system. Conversely, it is assumed that stationary control is relatively cheap to implement but will result in somewhat inferior performance. Which approach shows the greater profit clearly depends on the problem and no attempt has been made to model implementation costs here. Our premis is that the designer has already decided that stationary control will do the job to his satisfaction. What then are the tools that are available?

Following the introduction to the chapter, the 'traditional' theory of stationary control is reviewed in Appendix 5.1. The purpose of this review is simply to set the subsequent work in the appropriate context and to introduce the somewhat specialised jargon of stationary control, such as optimality equation, stationarity equation, successive approximation method, policy iteration method, and so on. The 'traditional' theory of stationary control relates almost exclusively to the control of finite state Markov chains. Consequently the discussion of Appendix 5.1 is presented in terms of this class of problem, the benefit being that an elegant matrix notation can be introduced and a numer of very useful numerical techniques can be proposed.

It is then indicated in Appendix 5.2 how the class of problem can be extended to include the perfectly observable linear system with quadratic cost. It is demonstrated that the theory reduces to a study of the behaviour of the matrix Riccati equation. What is more, the theory can be extended even further to include the classical information pattern version of the problem.

The real effort of the Chapter began in Section 2 with a discussion of the infinite horizon zero memory problem. It was shown that the traditional theorems of stationary control could be applied without much difficulty to the zero memory problem. The crucial difference however is that now the 'state' of the system is the probability density function $p(x)$ of the process state x. And this is infinite dimensional, and does not evolve in a nice and orderly linear fashion.

As a means of avoiding the difficulties associated with infinite dimensionality, we restricted our attention in Section 3 to the problem of finding optimal linear control laws. Because the control law $\gamma(y)$ is linear, the state $p(x)$ is Gaussian, and therefore it can be summarised by its sufficient statistics - its means and variance. (Since we require that $\lim_{n \to \infty} E(x_n) = 0$, it often turns out that we can assume that the mean is zero and therefore we can safely restrict our attention to proportional controllers $\gamma(y) = ky$ where k is a functional of only the variance of $p(x)$). Three problems were examined in Section 3; the discount problem, the average cost per period problem, and the ergodic problem. Difficulty was encountered with all but the ergodic problem. In fact all that we could do was present necessary conditions for the optimality of what we termed static controllers - that is, control laws of the form $\gamma(y) = ky$, where k does not depend on $p(x)$ at all. Naturally, these are a cruder form of control law but for the ergodic problem (and the average cost per period problem) they are just as good as the stationary control law. This is not true for the discount problem.

Encouraged with our success at least in finding the optimal linear controller for the ergodic problem, we proceeded in Section 4 to examine the ergodic problem in greater depth. We compared the performance of the zero memory optimal controller with that of the optimal controller for the corresponding classical information pattern problem and were thus able to attach a value to the extra information that the classical infor-

mation pattern provides. It was shown numerically of how little value this information can be - the conclusion being that in most cases the designer will do just as well to stick to zero memory control (if indeed he has a choice). For completeness we presented a method of finding the optimal linear controller for a general multivariable linear system with quadratic costs. Finally we presented a number of approximation methods which might be applicable for less tractable versions of the ergodic problem. These methods involved a moment formulation of the problem, an approximation method based on Edgeworth series, and a perturbation approach.

In Section 5 we returned to the discount problem in an attempt to find an approximation to the best linear control law. The approach was based on a perturbation analysis of the successive approximation method of dynamic programming, and the results so generated can be applied to the finite term zero memory problem. Unfortunately, the results are not particularly satisfactory, and although they might be useful, they lack elegance. A numer of alternatives for further investigation were suggested.

CHAPTER 8

Further Examination of Witsenhausen's Problem

		Page
8.1	Introduction	8.1
8.2	Optimal Linear Control for Witsenhausen Problem	8.5
8.3	Necessary Condition for Witsenhausen Problem	8.11
8.4	Attempts to Show that T is a Contraction Mapping	8.15
8.5	Conclusions	8.32

8.1 Introduction

The Witsenhausen problem was first introduced in Chapter 2, as an interesting example of a zero-memory problem. Subsequently, various aspects of the Witsenhausen problem have been discussed at a number of points in the chapters which followed. In this chapter we shall examine a further aspect of the problem in some detail.

We shall be concerned with an attempt to show that, under certain conditions to be determined, the optimal control for the Witsenhausen problem is linear. Recall from our discussions in Chapter 2 that we reported on Witsenhausen's demonstration that the optimal control law for the problem can be non-linear. It was shown that the non linearity arises from the dual role that the controller must adopt - not only must he control, he must also signal subsequent controllers about his actions and for observations. Given this property of the optimal control law, what then motivates our attempt to show that the optimal control is linear under certain circumstances? In Chapter 6 we derived a necessary condition for the optimal control law for the problem. This was in the form of a Gateaux differential. Now it turns out that the optimal linear control law for the Witsenhausen problem satisfies this necessary condition (as we shall demonstrate in Section 3). Thus, if we can show that the necessary condition has a unique solution, then we have proved that the optimal linear control is the optimal overall control law. We shall attempt to prove this uniqueness property by demonstrating that the necessary condition can be expressed in the form of a contraction mapping. We can then

invoke the Fixed Point Theorem to assert that there can be only one function which satisfies the necessary condition.

In Section 4 we get down to the details of the problem and attempt to establish the appropriate contraction mapping properties. This work is characterised by two features. First, the analysis is extremely intricate. Second, it appears to be possible to establish uniqueness only for a restricted class of control laws.

For completeness, let us briefly re-introduce the Witsenhausen problem,

Process Equations $x_1 = x_0 + u_0$

$x_2 = x_1 - u_1$

Observation Equations $y_0 = x_0$

$y_1 = x_1 + v$

Cost Function $J' = k^2 u_0^2 + x_2^2,\ k^2 > 0$

Admissable Controllers $u_0 = \gamma_0(y_0)$

$u_1 = \gamma_1(y_1),$

where (γ_0, γ_1) are any pair of Borel measurable functions. (All the variables above are real scalars). The observation noise v is Gaussian with zero mean and variance σ_v^2. For our discussion here, we shall take x_0 as Gaussian with zero mean and variance σ^2. The objective is to find (γ_0, γ_1) such that $E(J')$ is minimised.

It is convenient to reformulate the problem somewhat and introduce the funotion f defined as follows:-

$$f(x) = x + \gamma_1(x).$$

The problem now becomes one of minimising the function J with respect to (f, γ_1), where

$$J(f,\gamma_1) = E\left\{k^2(x - f(x))^2 + (f(x) - \gamma_1(y_1))^2\right\}$$

$$= E\left\{k^2(x - f(x))^2 + (f(x) - \gamma_1(f(x)+v))^2\right\}$$

The optimal choice for γ_1 can be found easily. It is:-

$$\gamma_1^{\circ}(y_1) = + E\left[x_1 \mid y_1\right] = E\left[f(x) \mid y_1\right]$$

Thus we obtain

$$J^{\circ}(f) = J(f,\gamma_1^{\circ}) = E\left\{k^2(x - f(x))^2\right\} + E\left\{(f(x) - E\left[f(x) \mid y_1\right])^2\right\}$$

$$= E\left\{k^2(x - f(x))^2\right\} + E\left\{\text{var } f(x) \mid y\right\} .$$

Mathematically, the Witsenhausen problem has become one of finding that function f° such that the above functional $J^{\circ}(f)$ is minimised. As we saw in Chapter 2, despite its innocent appearance above, $J^{\circ}(f)$ is a very complex functional when written down in terms of the various probability density functions. In the following sections, we shall examine $J^{\circ}(f)$ and the problem of finding f° much more thoroughly.

8.2 Optimal Linear Control

In order to find the optimal linear control for the Witsenhausen problem, we put $f^L(x) = a + bx$ and attempt to determine the values of a and b which minimise $J^o(f^L)$. Notice that if $f(x)$ is linear, then x_1 is Gaussian and consequently $\gamma_1^o(y) = E\left[x_1 \mid y\right]$ is a linear function of y.

Also, it is worth restating a Lemma from Witsenhausen's paper (which is proved in Appendix 4.2.1.)

Lemma 8.1

For any admissable (f, γ_1), there exists (f', γ_1') such that

$$E\left[f'(x)\right] = 0,\ E\left[x - f'(x)\right]^2 \leq \sigma^2,\ E\left[f'(x)\right] \leq 4\sigma^2, \text{ and}$$

$$J(f', \gamma_1') \leq J(f, \gamma_1)$$

The important implication of this Lemma is that we can restrict our search to those $f(x)$ which have zero mean and variance less than or equal to $4\sigma^2$.

The above Lemma was proved for the case when an admissable control law is any Borel measurable function. However, returning to the problem of finding optimal linear control laws, we see that the Lemma can be trivially restricted to the class of affine control laws. (constant + proportional term). This is because the transformations we use in the

Lemma:-

$$f' = \gamma_1' = 0, \text{ or } f'(x) = f(x) - m$$

$$\gamma_1'(y) = \gamma_1(y + m) - m.$$

keep us in the class of affine controllers.

Thus, we have:-

$$E\left[f^{L}(x)\right] = a + b\,E\left[x\right] = 0,$$

which implies that we need only consider linear control laws with $a = 0$. This property can of course be derived directly.

Thus we have,

$$J^{o}(f^{L})\Big|_{a=0} = E\left\{k^2 x^2 (1-b)^2\right\} + E\left\{\text{var}\,(bx \mid y)\right\}$$

$$= k^2 \sigma^2 (1-b)^2 + \frac{b^2 \sigma^2 \sigma_v^2}{b^2 \sigma^2 + \sigma_v^2}$$

$$= \sigma_v^2 \left\{ k^2 \sigma_0^2 (1-b)^2 + \frac{b^2 \sigma_0^2}{b^2 \sigma_0^2 + \sigma_v^2} \right\}, \text{ where } \sigma_0^2 = \sigma^2/\sigma_v^2.$$

The optimal value for b is found via the necessary condition - $d J^{\circ}(f)/db = 0$. Putting $t = \sigma_0 b$, we obtain the optimal value of b via a real root of the following stationarity equation:-

$$(t - \sigma_0)(1 + t^2)^2 + t/k^2 = 0$$

Witsenhausen demonstrated a very interesting graphical interpretation of the above equation. Its solutions are the points of intersection of the curves.

$$y = t/(1+t^2)^2 \quad \text{and} \quad y = k^2(\sigma_0 - t)$$

These curves are sketchched below:-

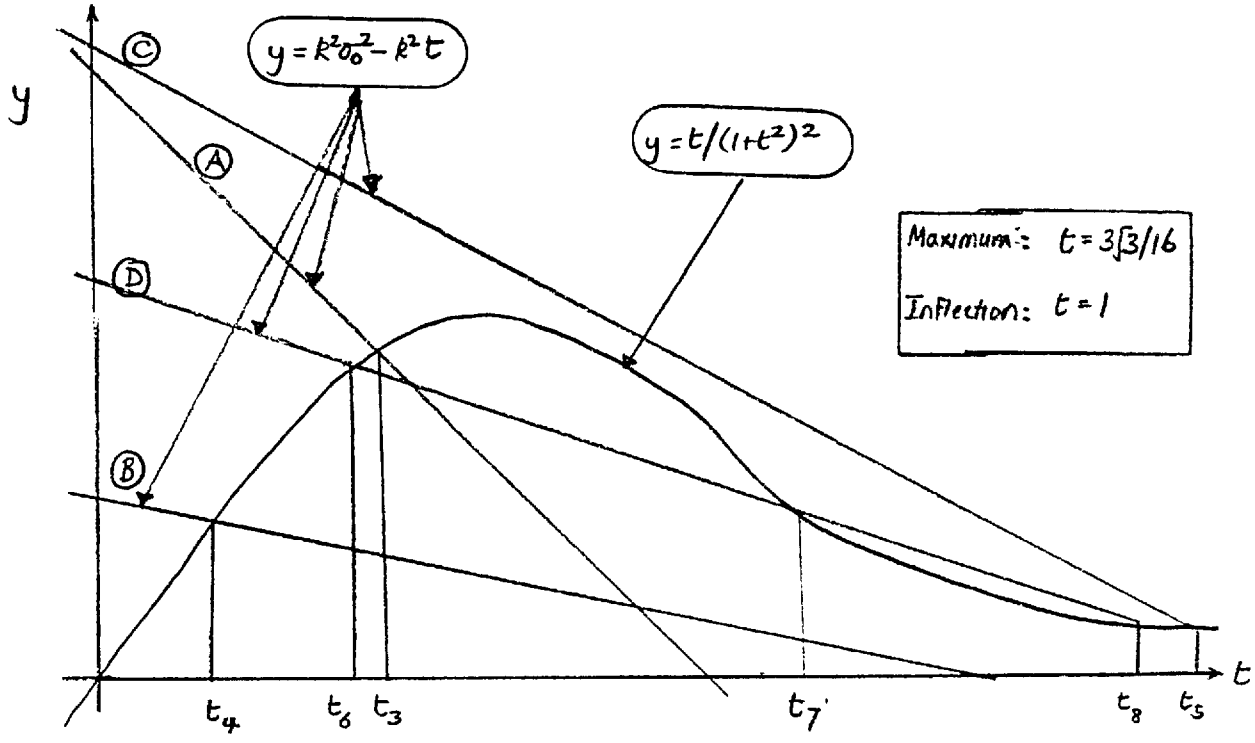


Fig. 8.1 Possible Solutions: Optimal Linear Control

In the diagram above, the curve has a point of inflection at $t = 1$, where it has a slope of $-\frac{1}{4}$. The points of intersection with the straight line $y = k^2 \sigma_o^2 - k^2 t$ will be determined by the values of k^2 and σ_0. Let us distinguish four cases:-

i) $k^2 \geq \frac{1}{4}$. In this case, whatever the value σ_0, there is only one real solution to the equation which therefore defines a unique optimum (line (A) t_3).

ii) $k^2 < \frac{1}{4}$; σ_0 sufficiently small. Here we have a unique solution with t small (line (B); t_4).

iii) $k^2 < \frac{1}{4}$; σ_0 sufficiently large. Again we have a unique solution, now with t large (line (C); t_5).

iv) $k^2 < \frac{1}{4}$; σ_0 intermediate. In this case there are three solutions, corresponding to two local minima and one local maximum. There is a value σ_c for σ_0 at which the two local minima are equal, and hence optimal. For $\sigma_0 < \sigma_c$, the lowest root is optimal; for $\sigma_0 > \sigma_c$, the largest root is optimal. (line (D); t_6, t_7, t_8)

The critical value σ_c is given by $k^2 \sigma_c^2 = 1$. This is demonstrated below. For $k^2 \sigma_c^2 = 1$, the stationarity equation becomes

$$(t-\sigma_c)(1+t^2)^2 + \sigma_c^2 t = 0$$

This equation can be factorised to give:-

$$(t^2 - \sigma_c t + 1)(t^3 + \sigma_c) = 0$$

The two roots of the quadratic are $t = \frac{1}{2}(\sigma_c \pm \sqrt{\sigma_c^2 - 4})$ and it can be checked that both these roots give the same value for $J^o(f^2)$, namely $\sigma_v^2(1 - k^2)$. Thus the critical condition is $k^2 \sigma_c^2 = 1$. (The real root of the cubic is the intermediate local maximum).

In his paper, Witsenhausen proceeds from a discussion of the optimal linear control law for the case of x_0 Gaussian to a disoussion of the overall optimal control with x_0 described by a two point symmetric distribution $P(x_0 = \sigma) = P(x_0 = -\sigma) = \frac{1}{2}$. By a priori considerations, if $f(\sigma) = a$, then $f(-\sigma) = -a$, and $a \geq 0$. Witsenhausen shows that any optimal value of a must satisfy the following transcendental equation:-

$$k^2(\sigma_0 - a) = -\frac{1}{2} h'(a),$$

where $h(a) = \sqrt{2\pi}\, a^2 \Phi(a) \int \frac{\Phi(y)}{\cosh(ay)}\, dy$

Now a plot of $-\frac{1}{2} h'(a)$ is similar to that of $t/(1 + t^2)^2$ sketched in Fig. 8.1. Thus the qualitative discussion of Fig. 8.1 applies in this case. However now we can attach a physical interpretation to the possile appearance of two local minima. As Witsenhausen says for small k^2 and appropriate σ_0 one policy is to bring a close to zero by means of f so that the second stage has little work to do; another policy is to make a larger than σ, creating a vast gap between a and - a, so that the second stage can almost infallibly separate these two values. This phenomenon was first discussed in Chapter 2. It has been repeated here to provide a possible physical interpretation of the appearance of two optimal linear control laws for the Witsenhausen problem.

We have now completed our discussion of optimal linear control. In the next section we shall consider the problem of finding overall optimal control laws.

8.3 A Necessary Condition for the Witsenhausen Problem

Previously in Section 6.4.2 we have derived a necessary condition for Witsenhausen's problem; namely the optimal control law $f(x)$ must satisfy the following equation:-

$$f(x) = x - \frac{\sigma_v^2}{2k^2}\int \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x)}{\sigma_v}\right\}\frac{D_f'(y)}{D_f(y)}\left\{\frac{2(y-f(x))^2}{\sigma_v} + \frac{D_f'(y)}{D_f(y)}(y-f(x)) - 2\right\}dy,$$

where $$D_f(y) = \int \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x)}{\sigma_v}\right\}\frac{\Phi}{\sigma}\left\{\frac{x}{\sigma}\right\}dx ,$$

$$D_f'(y) = \frac{d}{dy}\left[D_f(y)\right] = -\frac{y}{\sigma_v^2} + \int \frac{f(x)}{\sigma_v^2}\frac{\Phi}{\sigma_v}\left\{\frac{y-f(x)}{\sigma_v}\right\}\frac{\Phi}{\sigma}\frac{x}{\sigma}dx,$$

and $$\Phi(t) = \exp\left[-t^2/2\right]$$

We can express this equation in the form:-

$$f(x) = x - T(f)(x),$$

where:-

$$T(f)(x) \triangleq \frac{\sigma_v^2}{2k^2}\int_{\frac{\sigma_v}{\sigma_v}}^{\Phi}\left\{\frac{y-f(x)}{\sigma_v}\right\}\frac{D_f'(y)}{D_f(y)}\left\{2\left(\frac{y-f(x)}{\sigma_v}\right)^2+\frac{D_f'(y)}{D_f(y)}(y-f(x))-2\right\}dy$$

Notice that because of $D_f(y)$, T is a mapping of the entire function f and not merely its value $f(x)$. Thus we can write:-

$$f(x) = x - T(f)(x)$$

that is, $f(x) = T'(f)(x)$,

where $T'(f)(x) \triangleq I(x) - T(f)(x)$

The equation above, $f = T'f$ is simply a restatement of our necessary condition for optimality. In the previous chapter, the notion of a contraction operator was introduced to assist in the justification of the successive approximation method of dynamic programming. In this Chapter we are not particularly interested in the convergence properties of a contraction mapping - we are more concerned in establishing the uniqueness of a function f which satisfies $f = T'f$.

First of all, recall that the optimal linear solution to the Witsenhausen problem derived in the previous section is $f^L(x) = \frac{\sigma_v}{\sigma} t\,x$, where t satisfies:-

$$(t - \sigma_0)(1 + t^2)^2 + t/k^2 = 0$$

Now this solution $f^L(x)$ satisfies the above neceeasry condition for optimality, that is $f^L(x) = T'(f^L)(x)$. This can easily be verified by subsitution. If we put $f^L(x) = bx$, then:-

$$T'(f^L)(x) = x - \frac{bx}{k^2} \quad \frac{1}{(1+b^2\,\bar{\sigma}_0^2)^2} \quad , \text{ where } \bar{\sigma}_0 = \sigma / \bar{\sigma}_v$$

$$= x - \frac{\bar{\sigma}_v\, t\, x}{k^2\, \sigma} \quad \frac{1}{(1+t^2)^2} \quad , \text{ by substituting } b = \bar{\sigma}_v t/\sigma$$

$$= x + (t - \bar{\sigma}_0)\, x\, \frac{\bar{\sigma}_v}{\bar{\sigma}} \quad , \text{ by the optimality equation}$$

$$= x\, \frac{t\bar{\sigma}_v}{\bar{\sigma}} = \frac{bx}{}$$

Thus the optimal solution satisfies the overall necessary condition for optimality. That is, it is a solution of the functional equation $f = T'f$. If we could prove that under certain circumstances the solution to this equation is unique, then we could conclude that under the same circumstances the linear control law is optimal. But how can we show that the solution to the functional equation $f = T'f$ is unique? The approach which we shall adopt here is to attempt to show that T' is a contraction mapping. Let us then review the notion of a contraction mapping and its applications (a limited discussion of contraction mappings was presented in the previous chapter).

Given a Banach space E and an open subset $U \subset E$, suppose there exists a mapping T which maps every open subset U of E into U. Given any x_1, x_2 in U, suppose also that there exists a constant K, where $0 < K < 1$, such that,

$$\| T(x_1) - T(x_2) \| \leq K \| x_1 - x_2 \| ,$$

in other words T satisfies a Lipschitz condition with $K < 1$. Then T is a contraction mapping. The analytical value of a contraction mapping is provided by the following theorem (see for example reference 1 below).

Theorem 8.2 (Fixed Point Theorem)

With E, U and T as above, suppose that there exists a sphere $S = \{ x: \| x - x_0 \| \leq a \}$ that is contained in U such that

$$\| T(x_0) - x_0) \| < (1-K)a$$

Then there is a unique point $x^* \in S$ such that $T(x^*) = x^*$.

Proof

See Reference 1 cited below.

1. Saaty and Bram - "Non Linear Mathematics". (45)

8.4 Attempts to Show That T' is a Contraction Mapping

Our problem is to prove that, for arbitrary f_1 and f_2,:-

$$\left\| T'(f_1)(x) - T'(f_2)(x) \right\| < K \left\| f_1(x) - f_2(x) \right\|, \text{ where } 0<K<1,$$

and where $\| \cdot \|$ denotes some appropriate norm. A convenient method of handling such problems is to introduce a parameter ϵ, $0 \leqslant \epsilon \leqslant 1$, and define a function $f(x,\epsilon)$ as follows:-

$$f(x,\epsilon) \triangleq \epsilon f_2(x) + (1-\epsilon) f_1(x)$$

We now consider the mapping $T'(f)(x,\epsilon)$, and proceed as follows:-

$$\left\| T'(f_1)(x) - T'(f_2)(x) \right\| = \left\| T(f_2)(x) - T(f_1)(x) \right\| = \left\| \int_0^1 \frac{\partial Tf(x,\epsilon)\, d\epsilon}{\partial \epsilon} \right\|$$

Suppose we can show that

$$\frac{\partial T(f)(x,\epsilon)}{\partial \epsilon} \leqslant K \frac{\partial f(x,\epsilon)}{\partial \epsilon} \quad \text{for all } x, \epsilon \text{ with } 0<K<1. \quad \text{——(1)}$$

Then

$$\left\| T'(f_1)(x) - T'(f_2)(x) \right\| \leqslant K \left\| \int_0^1 \frac{\partial f(x,\epsilon)}{\partial \epsilon}\, d\epsilon \right\|$$

$$= K \left\| f_1(x) - f_2(x) \right\|$$

From this, it follows that the contraction mapping property of T' will be assured if we can demonstrate that (1) holds. This is easier said

than done, as will be seen in the sequel.

Notice that $\partial f(x, \epsilon)/\partial \epsilon = f_2(x) - f_1(x)$. Hence we can interchange the norm and integration with respect to ϵ operations in the above development, to obtain a somewhat more convenient form of condition (1) above; namely

$$\left\| \frac{\partial T(f)\,(x, \epsilon)}{\partial \epsilon} \right\| \leqslant K \left\| \frac{\partial f(x, \epsilon)}{\partial \epsilon} \right\| \quad \text{with } 0<K<1 \quad \text{——(2)}$$

if the particular norm function permits (see 8.4.2).

So far we have not mentioned the nature of the norm function. The choices are many. Perhaps the simplest approach is to attempt to demonstrate pointwise uniqueness. In other words, our condition above becomes

$$\left| \frac{\partial T(f)(x, \epsilon)}{\partial \epsilon} \right| \leqslant K \left| \frac{\partial f(x, \epsilon)}{\partial \epsilon} \right| \quad \text{for each } x, \epsilon \text{ with } 0<K<1. \text{ -(3)}$$

Suppose wecan show that this condition is satisfied for every x. Thus, by contraction mapping arguments, if two functions f_1 and f_2 satisfy the equation $f = T^{\dagger} f$, then $f_1(x) = f_2(x)$. Despite the simplicity of this approach, we shall see that it is inappropriate for our problem here:-

8.4.1 Pointwise Uniqueness

From the previous section, we have:-

$$T(f)(x,\ ) = c^2 \int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \left\{ 2\left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\}^2 + \frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} (y-f(x,\epsilon)) - 2 \right\} \cdot dy, \quad (1)$$

$$= c^2 \left\{ I_1 + I_2 + I_3 \right\}$$

where $c^2 \triangleq \sigma_v^2/k^2$, and:-

$$I_1(x,\epsilon) \triangleq 2 \int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \left\{ \frac{y-f(x,\epsilon)}{\sigma v} \right\}^2 dy,$$

$$I_2(x,\epsilon) \triangleq \int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \left(\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \right)^2 (y-f(x,\epsilon)) \, dy,$$

$$I_3(x,\epsilon) = -2 \int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \, dy$$

From this, we require $\dfrac{\partial T(f)(x,\epsilon)}{\partial \epsilon}$. This is given by:-

$$\frac{\partial T(f)(x,\epsilon)}{\partial \epsilon} = c^2 \left\{ \frac{\partial I_1}{\partial \epsilon} + \frac{\partial I_2}{\partial \epsilon} + \frac{\partial I_3}{\partial \epsilon} \right\}$$

$$\triangleq c^2 \left\{ J_{11}+J_{12}+J_{13}+J_{21}+J_{22}+J_{23}+J_{31}+J_{32} \right\},$$

where $\frac{\partial I_i}{\partial \epsilon^i} = \sum_j J_{ij}$. Let us examine some typical "J" terms from this expression, say J_{11} and J_{12}. We have:-

$$J_{11} = +2\int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\}^3 \frac{\partial f(x,\epsilon)}{\partial \epsilon} \frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \, dy$$

$$J_{12} = 2\int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\}^2 \frac{d}{d\epsilon} \left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \right] dy$$

Evaluation of bounds for J_{11} and J_{12} is complicated by the appearance of terms such as,

$$\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \quad \text{and} \quad \frac{d}{d\epsilon} \left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \right].$$

In order to put the zero-memory Witsenhausen problem into perspective, compare it for a moment with the classical information pattern version of the same problem. We would expect for the classical information pattern version that a similar necessary condition to the equation $f = T'f$ would

be obtained; however, recall that for the zero-memory problem,

$$\frac{D_f'(y, \epsilon)}{D_f(y, \epsilon)} = \frac{-y}{\sigma_v^2} + \frac{E\left[f(x, \epsilon) \middle| y\right]}{\sigma_v^2}$$

Thus for the classical information pattern version, we would expect terms such as:-

$$\frac{"D_f'(\,, \epsilon)"}{D_f(y, \epsilon)} = \frac{-y}{\sigma_v^2} + \frac{E\left[f(x, \epsilon) \middle| y, x\right]}{\sigma_v^2} = -\left\{\frac{y - f(x, \epsilon)}{\sigma_v^2}\right\}$$

For such $"D_f'(y, \epsilon) \;/\; D_f(y, \epsilon)"$, we obtain very conveniently upon substitution in (1) above, the result:-

$$T(f)(x, \epsilon) = 0,$$

from which it follows that:-

$$f(x) = x,$$

and this is indeed the optimal solution for the classical information pattern version of Witsenhausen's problem ($\gamma_0(x) = x - f(x) = 0$ - do nothing at the 'expensive' first stage and leave all corrections to the 'free' second stage). However, for the zero memory problem, the term - $D_f'(y, \epsilon) \;(D_f(y, \epsilon)$ does not behave in such a friendly manner.

In our expression for $\partial T(f)(x, \epsilon)/\partial \epsilon$, terms such as J_{11} which involve the 'undifferentiated' $D_f'(y, \epsilon)/D_f(y, \epsilon)$ can be handled with the aid of the following two Lemmas. (8.3 and 8.4).

Lemma 8.3

Let X be a random variable with second moment σ^2. Let Z be and independent random variable with symmetric unimodal probability density Φ.

Suppose $Y = X + Z$. Then:-

$$\min(-\sigma, -\sigma - 2y) \leqslant E\left[X \mid Y = y\right] \leqslant \max(\sigma, \sigma + 2y)$$

Proof

The above Lemma appears an an example in a report entitled, "Tchebycheff Inequalities for Conditional Moments", by J.M.C. Clark[1].

In the context of Witsenhausen's problem; we have:-

$$y = f(x) + v$$

and $E\left[f^2(x)\right] \leqslant 4\sigma^2$. Thus

$$\min(2\sigma, -2\sigma + 2y) \leqslant E\left[f(x) \mid Y = y\right] \leqslant \max(2\sigma, 2\sigma + 2y)$$

We can now use this bound on $E\left[f(x) \mid Y = y\right]$ to determine bounds on $D_f'(y, \epsilon)/D_f(y, \epsilon) = -(y - E\left[f(x) \mid Y = y\right]/\sigma_v^2$.

1. "Tchenycheff Inequalities for Conditional Moments". J.M.C. Clark Publication 72/74 Department of Computing and Control, Imperial College, London, SW7- August 1972. (14)

$$\frac{1}{\sigma_v^2}\left\{-y + \min(-2\sigma,\ -2\sigma + 2y)\right\} \leqslant \frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \leqslant \frac{1}{\sigma_v^2}\left\{-y + \max(2\sigma,\ 2\sigma + 2y)\right\}$$

The admissable region of $D_f'(y,\epsilon)/D_f(y,\epsilon)$ is illustrated below:-

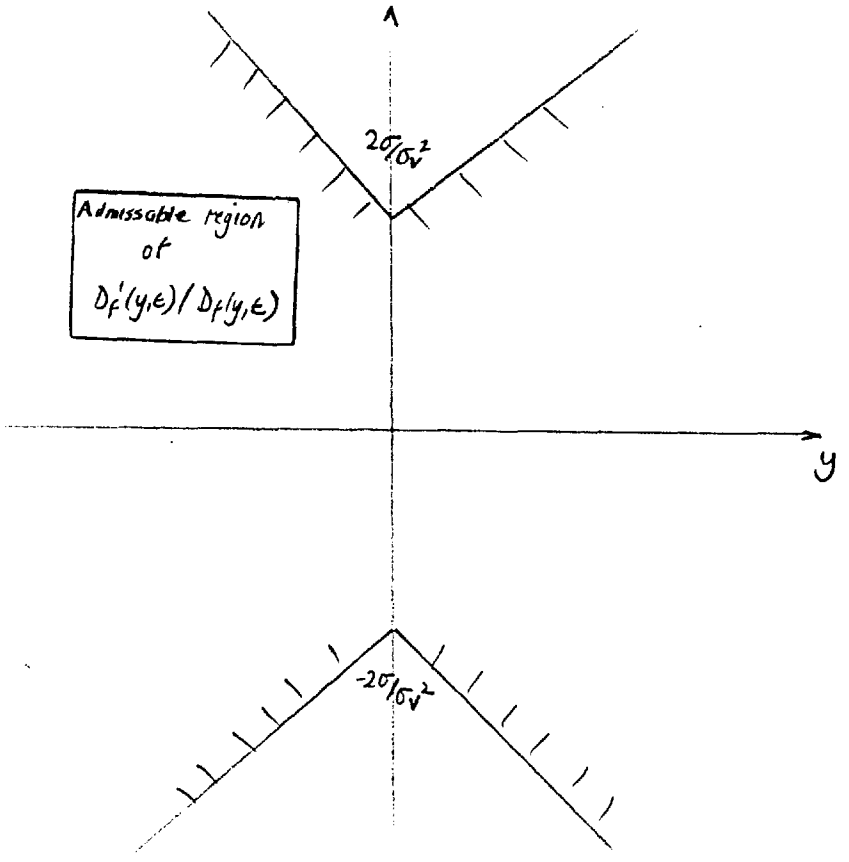


Fig. 8.2 Bounds for $D_f'(y,\epsilon)/D_f(y,\epsilon)$

Notice that $\left| \frac{D_f'(y, \epsilon)}{D_f(y, \epsilon)} \right| \leqslant \frac{2\sigma}{\sigma_v^2} + \frac{|y|}{\sigma_v^2}$

Next we require a bound on $f(x)$. This is obtained next:-

Lemma 8.4

Suppose $E\left[f^2(x)\right] \leqslant 4\sigma^2$, ~~Then:~~ and f is non-decreasing. Then:-

$$\frac{-2\sigma}{\sqrt{\int_{-\infty}^{x} p(x)\,dx}} \leqslant f(x) \leqslant \frac{2\sigma}{\sqrt{\int_{x}^{\infty} p(x)\,dx}}$$

Proof

Suppose $f(x) > 2\sigma / \sqrt{\int_x^\infty p(x)\,dx}$; then

$$\int_{-\infty}^{+\infty} f^2(x)p(x)\,dx > \int_x^\infty f^2(\theta)p(\theta)\,d\theta > \left(\int_x^\infty p(\theta)\,d\theta\right) f^2(x)$$

$$> 4\sigma^2, \text{ which is a contradiction.}$$

Thus we have bounds on $f(x)$. It is depicted below (for $p(x) = \frac{\Phi}{\sigma}\left\{\frac{x}{\sigma}\right\}$)

It has been established by Witsenhausen (56) that the optimal f is non-decreasing and we shall restrict our attention to such functions in what follows.

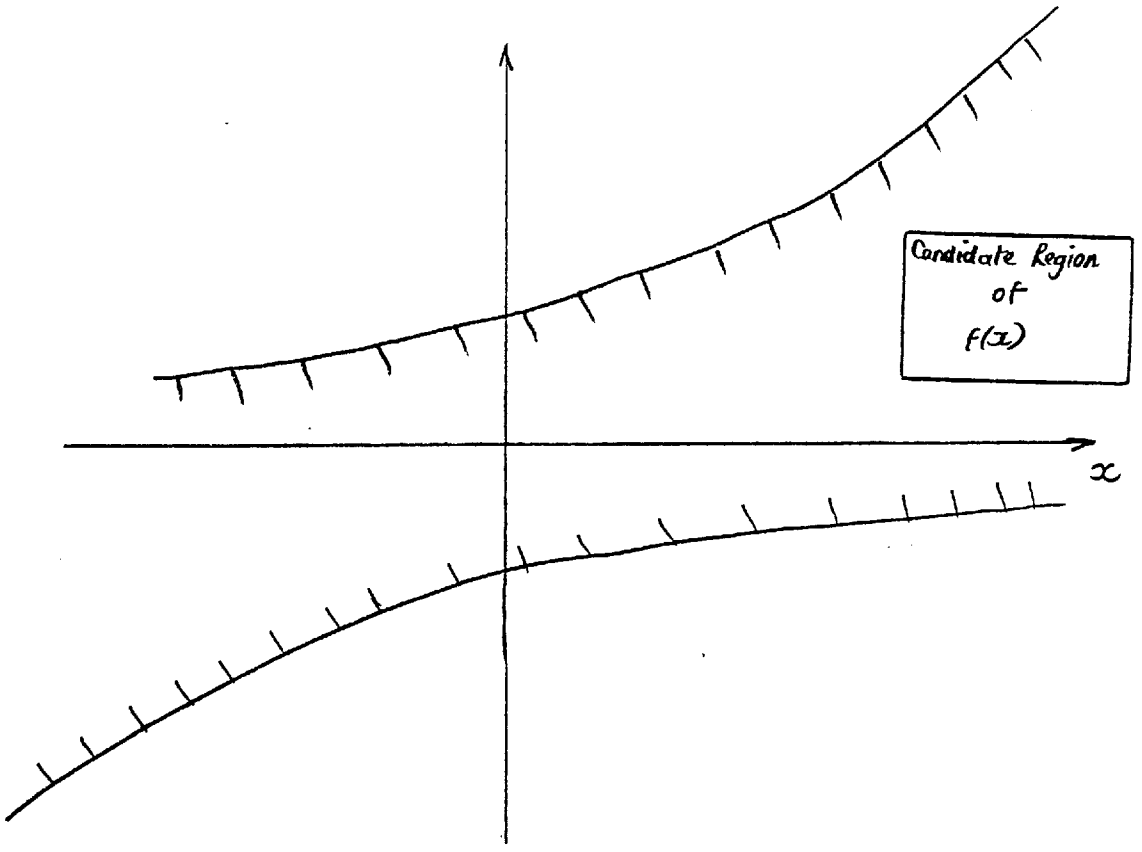


Fig. 8.3 Candidate region of f(x)

In this case $\left|f(x)\right| \leqslant \dfrac{2\sigma}{\int_{|x|}^{\infty} p(x)\,dx} \triangleq c(x)$.

Let us attempt to estimate J_{11}. We have:-

$$\left| J_{11} \right| = 2 \left| \int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \left\{ \frac{y-f(x,\ )}{\sigma_v} \right\}^3 \frac{\partial f(x,\epsilon)}{\partial \epsilon} \frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)} \, dy \right|$$

$$\leqslant 2 \left| \frac{\partial f}{\partial \epsilon} \right| \int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x,\epsilon)}{\sigma_v} \right\} \left| \frac{y-f(x,\epsilon)}{\sigma_v} \right|^3 \left[\frac{2\sigma}{\sigma_v^2} + \frac{|y|}{\sigma_v^2} \right] dy$$

This expression can be simplified to obtain:-

$$\left| J_{11} \right| \leqslant 2 \left| \frac{df}{\partial \epsilon} \right| \int \Phi(s) \, |s|^3 \left\{ \frac{2\sigma}{\sigma_v^2} + \frac{|s+f(x,\epsilon)|}{\sigma_v^2} \right\} ds,$$

which ultimately leads to:-

$$\left| J_{11} \right| \leqslant 2 \left| \frac{df}{d\epsilon} \right| \left\{ m_3 \frac{2\sigma}{\sigma_v^2} + \frac{m_4}{\sigma_v} + \frac{m_3 \, c(x)}{\sigma_v^2} \right\},$$

where $m_3 \triangleq \int \Phi(s) \, |s|^3 \, ds,$

$m_4 \triangleq \int \Phi(s) \, s^4 \, ds,$

and $c(x)$ is as defined above.

A similar technique can be employed to evaluate bounds for certain other terms $\{J_{ij}\}$. However, we will encounter difficulties in evaluating bounds for these terms which contain $\frac{\partial}{\partial \epsilon}\left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)}\right]$. Consider the term J_{12} defined earlier:-

$$J_{12} = 2\int \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\}^2 \frac{\partial}{\partial \epsilon}\left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)}\right] dy$$

Now the term $\frac{\partial}{\partial \epsilon}\left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)}\right]$ has the form indicated below:-

$$\frac{\partial}{\partial \epsilon}\left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)}\right] = \frac{1}{\sigma_v^2}\left[\frac{\partial}{\partial \epsilon}\left\{\frac{\int f(x,\,)\ \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} p(x)\ dx}{\int \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} p(x)\ dx}\right\}\right]$$

$$= \frac{1}{\sigma_v^2}\left[\frac{\int \frac{df(x,\epsilon)}{d\epsilon}\ \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} p(x)\ dx}{\int \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} p(x)\ dx}\right.$$

$$\left. + \int \frac{df(x,\epsilon)}{d\epsilon}\left\{\frac{y-f(x,\epsilon)}{\sigma_v^2}\right\} \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} p(x)\,dx\right.$$

$$- \int f(x,\epsilon) \; \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} \; p(x)\,dx \cdot \int \frac{df(x,\epsilon)}{d\epsilon} \; \left\{\frac{y-f(x,\epsilon)}{\sigma_v^2}\right\} \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} p(x)\,dx$$

$$\Big/ \left(\int \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} \; p(x)\; dx\right)^2$$

Thus suppose, $J_{12} \triangleq K_{121} + K_{122} + K_{123}$. Then typically we have:-

$$K_{121} = \frac{2}{\sigma_v^2} \iint \frac{\Phi}{\sigma_v}\left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\} \left\{\frac{y-f(x,\epsilon)}{\sigma_v}\right\}^2 \int \frac{df(s,\epsilon)}{d\epsilon} \; \frac{\Phi}{\sigma_v}\left\{\frac{y-f(s,\epsilon)}{\sigma_v}\right\} p(s)\,ds\,dy$$

$$\Big/ \int \frac{\Phi}{\sigma_v}\left\{\frac{y-f(r,\epsilon)}{\sigma_v}\; p(r)\right\} dr$$

This expression illustrates the intrinsic difficulty in our attempt to show pointwise uniqueness. Recall that we are trying to demonstrate that:-

$$\left|\frac{\partial T(f)(x,\epsilon)}{\partial \epsilon}\right| \leqslant K \left|\frac{\partial f(x,\epsilon)}{\partial \epsilon}\right| \quad \text{for each } x$$

But on the right hand side of the expression above for K_{121} we are obtaining a component $\frac{\partial f(s,\epsilon)}{\partial \epsilon}$, which is not what we want at all. What-

ever tricks we employ, we can never reduce K_{121} to an inequality of the form:-

$$\left| K_{121} \right| \leqslant F(x) \left| \frac{\partial f(x, \epsilon)}{\partial \epsilon} \right|$$

because the component $\frac{\partial f(x, \epsilon)}{\partial \epsilon}$ just does not arise. Hence we must abandon our attempt to show pointwise uniqueness. Instead, we have to consider uniqueness in function space. This is considered next.

8.4.2 Uniqueness in Function Space

A number of norm functions have been examined. The most promising one is perhaps:-

$$\| f \|_{L_1} = \int \left| f(x) \right| p(x) dx$$

This norm is attractive because we obtain $\left\| \frac{\partial f}{\partial \epsilon} \right\|$ from $\int \left| \frac{\partial f(s, \epsilon)}{\partial \epsilon} \right| p(s) ds$ rather than $\int \left| \frac{\partial f(x, \epsilon)}{\partial \epsilon} \right| p(x) dx$ (see discussion above). However, we shall see that this optimism is unfounded.

$$\| K_{121} \| \leqslant \frac{2}{\sigma_v^2} \int \left[\frac{\int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(x, \epsilon)}{\sigma_v} \right\} \left| \frac{y-f(x, \epsilon)}{\sigma_v} \right|^2 p(x) dx \int \left| \frac{df(s, \epsilon)}{d} \right| \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(s, \epsilon)}{\sigma_v} \right\} p(s) ds}{\int \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(r, \epsilon)}{\sigma_v} \right\} p(r) dr} \right] dy$$

$$\leqslant \frac{2}{\sigma_v^4} \iint \left\{ \left|y\right| + E\left[f^2(x,\epsilon)\middle| y\right] + 2\left|y\right| \left|E\left[f(x,\epsilon)\middle| y\right]\right| \right\} \frac{\Phi}{\sigma_v} \left\{ \frac{y-f(s,\epsilon)}{\sigma_v} \right\} dy \left| \frac{df(s,\epsilon)}{d\epsilon} \right| p(s)\, ds$$

A bound on the term $E\left[f(x,\epsilon)\middle| y\right]$ has already been presented (Lemma 8.3). By a similar technique we can derive bounds for $E\left[f^2(x,\epsilon)\middle| y\right]$. Ultimately, we should obtain an expression which contains sums of terms such as $L_{121\ i}$, where,

$$\left\| L_{121\ i} \right\| \leqslant \frac{2}{\sigma_v^4} \int a \left| f(s,\epsilon) \right|^n \left| \frac{df(s,\epsilon)}{d\epsilon} \right| p(s)\, ds, \text{ where a is}$$

a known constant.

Although we have $\left| f(s,\epsilon) \right| \leqslant c(s)$ (Lemma 8.4), this inequality is of little use in the above expression. Thus the L_1 norm is of no value in estimating terms containing $\frac{\partial}{\partial\epsilon}\left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)}\right]$. It should also be pointed out that we are led to a similar problem in evaluating bounds for terms such as $\left|J_{11}\right|$ (which is derived from a component which does not involve $\frac{\partial}{\partial\epsilon}\left[\frac{D_f'(y,\epsilon)}{D_f(y,\epsilon)}\right]$). Thus:-

$$\left\| J_{11} \right\| \leqslant 2 \int \frac{df}{d\epsilon} \left\{ m_3 \frac{2\sigma}{\sigma_v^2} + \frac{m_4}{\sigma_v} + m_3 \frac{c(x)}{\sigma_v^2} \right\} p(x)\, dx$$

$$= \left\{ \frac{4m_3\sigma}{\sigma_v^2} + \frac{2m_4}{\sigma_v} \right\} \left\| \frac{df}{d\epsilon} \right\| + \frac{2m_3}{\sigma_v^2} \int \left| \frac{df(x,\epsilon)}{d\epsilon} \right| c(x)\, p(x)\, dx$$

$$\leqslant \left\{ \frac{4m_3\sigma}{\sigma_v^2} + \frac{2m_4}{\sigma_v} \right\} \left\| \frac{df}{d\epsilon} \right\| + \frac{2m_3}{\sigma_v^2} \int \left\{ \frac{\left| \frac{df(x,\epsilon)}{d\epsilon} \right| 2\sigma p(x)}{\int_{(x)}^{\infty} p(s)\, ds} \right\} dx$$

It does not appear to be easy to convert the last term in this expression to one involving $\left\| \frac{df}{d\epsilon} \right\|$ as desired.

The results of sections 4.1 and 4.2 both lead to the unfortunate conclusion that we cannot demonstrate either pointwise uniqueness or uniqueness in function space without an immense amount of difficulty. The stumbling blocks are components such as:-

$$\int \left\{ \frac{\left| \frac{df}{d\epsilon}(x,\epsilon) \right| p(x)}{\left(\int_{(x)}^{\infty} p(s)\, ds \right)^n} \right\} dx$$

If we could estimate such terms in the form $K \left\| \frac{df}{d\epsilon} \right\|$, then we could, at least in theory, proceed to demonstrate uniqueness in function space, that is we could demonstrate the validity of the inequality:-

$$\left\| \frac{\partial T(f)(x,\epsilon)}{\partial \epsilon} \right\| < K \left\| \frac{\partial F(x,\epsilon)}{\partial \epsilon} \right\|$$

But the task of finding the expression for K in terms of the parameters of the problem is still colossal and it would take an act of extreme fanaticism to work K out.

8.4.3 Restricted Uniqueness Theorem

The author has been unsuccessful in developing the above analysis any further. However, a limited form of uniqueness result has been established. This requires the additional assumption that $\left|f(x)\right| \leqslant M$. Even a drastic assumption such as this by no means reduces the problem to the level of trivia - the algebraic effort is still of gargantuan proportions. We shall spare the reader the details and merely quote the result.

Theorem 8.3

For the Witsenhausen problem, if $\left|(f(x))\right| \leqslant M$, then

$$\left\| \frac{\partial T}{\partial \epsilon}(f)(x, \epsilon) \right\|_{L_1} \leqslant K \left\| \frac{\partial f}{\partial \epsilon} \right\|_{L_1},$$

where

$$K = \frac{\sigma_v^2}{2k^2} \left\{ 10 + \frac{4n_1 + M(6 + 40n_1 + 32n_3)}{\sigma_v} + \frac{M^2(124 + 8n_1)}{\sigma_v} \right.$$

$$\left. + \frac{M^3(120n_1 + 16)}{\sigma_v^3} + \frac{M + 47}{\sigma_v^4} \right\},$$

with $n_1 = \int_0^\infty x\, \Phi(x)\, dx$ and $n_3 = \int_0^\infty x^3\, \Phi(x)\, dx$

Thus the optimal solution is unique in L_1 if the parameters are such that $K < 1$.

Proof

The analysis proceeds as outlined in Section 4.2, making early use of the inequality $|f(x)| < K$ to reduce the algebraic complexity somewhat.

Notice that the above result is valid for any form of initial probability density function $p(x)$. Nevertheless, the assumption that $|f(x)| \leqslant M$ is still very restrictive and difficult to justify on any grounds other than analytical convenience. We could of course introduce $|f(x)| \leqslant M$ as a constraint in the problem. In this case however, the simple Gateaux differential necessary condition would not be appropriate and we would be forced to employ a Lagrangian formulation. And this would complicate the analysis even further.

8.5 Conclusions

We have failed to prove our objective; that under certain conditions the Witsenhausen problem has an optimal solution which is linear. This is disappointing, especially from the point of view of the author who has spent a long time examining the problem. All that we have managed is a somewhat restricted and, perhaps unrealistic uniqueness result, namely : the solution is unique if $|f(x)| < M$ for all x. And as we remarked above, establishing even this result invcIves an immense amount of unpleasant algebra.

So, are we mistaken in attempting to show that the linear solution is optimal? The answer is still 'no'. There remain strong grounds for believing in the validity of what we have been attempting to prove. First, we can establish contraction mapping properties for a re-arranged version of the stationarity equation for linear control:-

$$t = \sigma_0 - \frac{t}{k^2(1+t^2)^2}$$

This demonstrates that a unique solution exists for k^2 large enough. Admittedly, this tell us nothing new, since we have already demonstrated this via the discussion surrounding Fig. 8.1. But at least we have established some contraction mapping properties for the problem, and this does tend to support the approach taken in Section 4.2.

Second, if we could derive a contraction mapping property, it would be of the form:-

$$\left\| \frac{\partial T}{\partial \epsilon} \right\| < K \frac{\sigma_v^2}{k^2} \left\| \frac{\partial f}{\partial \epsilon} \right\|$$

Thus we would be able to claim uniqueness for $\frac{\sigma_v^2}{k^2}$ small enough.

Now in the parlance of communications theory, this is equivalent to stating that the signal to noise ratio should be small. But communications theory tells us that for small signal to noise ratio, linear modulation should be employed. And we have certainly gained enough insight into the problem to assert that $f(x)$ is part modulation of x and part control of x. If no modulation (no information) were required from $f(x)$, the control component would be linear. So if the modulation were also linear, we could reasonably expect that $f(x)$ is linear. This brings us back to the dual nature of the control function $f(x)$ - its control role and its communications role. It is a feature we have noted on several previous occasions. The contribution of communications theory to the study of non-classical information pattern problems is one of the topics which will be discussed in the next chapter.

CHAPTER 9

Some Comments on Recent Developments

		Page
9.1	Introduction	9.1
9.2	Separation and Estimation	9.9
9.3	Information Structures	9.20
9.4	Delayed Sharing Information Patterns	9.29
9.5	Linear Solutions in Non-Classical Problems	9.33
9.6	Communication and Control	9.35
9.7	Large Scale Systems	9.40
9.8	Conclusions	9.43

9.1 Introduction

In the preceding chapters of this thesis, we have been primarily concerned with an examination of the zero memory stochastic optimal control problem. The practical importance of this class of problem has been fully explained in Chapter 1. In our treatment, a number of interesting features of the problem have become apparent. For example, we have seen that the controller must adopt two perhaps contradictory roles; that of control and communication. We have seen how even linear controllers are difficult to design. We have remarked on the phenomenon of second-guessing, and so on.

In the present chapter we shall broaden our horizons somewhat and build on the concepts and ideas which we have now established. This will be achieved via a discussion of some selected literature from the field. The question of the information pattern of a stochastic optimal control problem had received very scant attention from researchers until the publication in 1968 of Witsenhausen's (56) epoch-making counter-example. In fact it would appear that not many people even realised that there was a problem to solve. And the sheer intractability of Witsenhausen's problem possibly delayed the publication of any articles which directly addressed the question of the information pattern. Our discussion of 'recent' literature on the subject will therefore be based on literature which has been published since 1971. This fits in nicely with our scheme of things. With the exception of Chapter 8, the work presented in the previous chap-

ters was performed before October 1971, and we have already put this in the context of 1971 via our introductory remarks in Chapter 1. Thus we can view the present chapter as an attempt to bring matters up to date and relate our work to that of a more recent origin.

We shall discover that whilst new insights into the problem have been gained and something approaching a new nomenclature has evolved, the non-classical information pattern problem remains as frustrating as ever. The problem still remains one of resolving the dilemma of communication versus control. Our discussion will not be restricted to non-classical information pattern problems. We have seen in Chapter 1 how non-classical information pattern problems are associated with decentralised controllers and how, for one reason or another, decentralised control is a feature of large scale system design. Thus we shall include some brief comments on this interesting topic. We shall also include some comments on classical information pattern problems since, as we asserted in Chapter 3, despite the universal acceptance of the LQG solution the basis of this solution is still not fully appreciated in all quarters.

Our discussion of recent literature is neither a survey nor a bibliography of the field. This can be found elsewhere - see for example Sandell et al (48) for a very comprehensive treatment. Our intention is to proceed in the same spirit as preceding chapters and discuss various topics informally wherever possible or relevant by means of a simple scalar example. Our discussion will in no way be exhaustive since the topics to

be treated have been selected according to personal taste, where it is felt that they support or extend notions already introduced.

In presenting a chapter such as this, the major difficulty is one of organisation. We could, for example, discuss the various contributions from various authors in chronological order; and the result would be chaos. Instead we shall attempt to group the contributions under a suitable topic heading and then present each topic in more or less evolutionary order. One further comment - we shall dispense with our custom of providing the title of a cited contribution in a footnote. Otherwise the chapter would be all footnotes.

We shall begin with a discussion of Radner's result on the optimisation of static teams.

The Optimisation of a Static Team

Team theory is a term which was first coined to describe a situation in game theory where the players in the game co-operate with each other rather than compete. This has obvious and immediate relevance to control theory, since we can review the control problem as a game in which each player (controller) makes his move (applies a control) in co-operation with other players (controllers) to minimise a single cost function. A static team is one where the order in which the players make their move is irrelevant, since the move of one player cannot influence the information (and hence the move) of another player. A dynamic team is one in which the order of play is relevant because the action of the first player might

affect the information available to the second player and his action, and so on. Thus a dynamic team problem corresponds to our conventional idea of a control problem in which the controller at each instant in time can be viewed as a player in the game. The notion of team theory has already been introduced in Chapter 1.

In this section we propose to discuss a major result of static team theory due to Radner (42). Inclusion of this discussion here is something of an anachronism in the light of our comments in Section 1. However, Radner's result is exploited on several occasions in what follows and is therefore worthy of more than passing reference.

Suppose we have the situation depicted below (single stage control problem):-

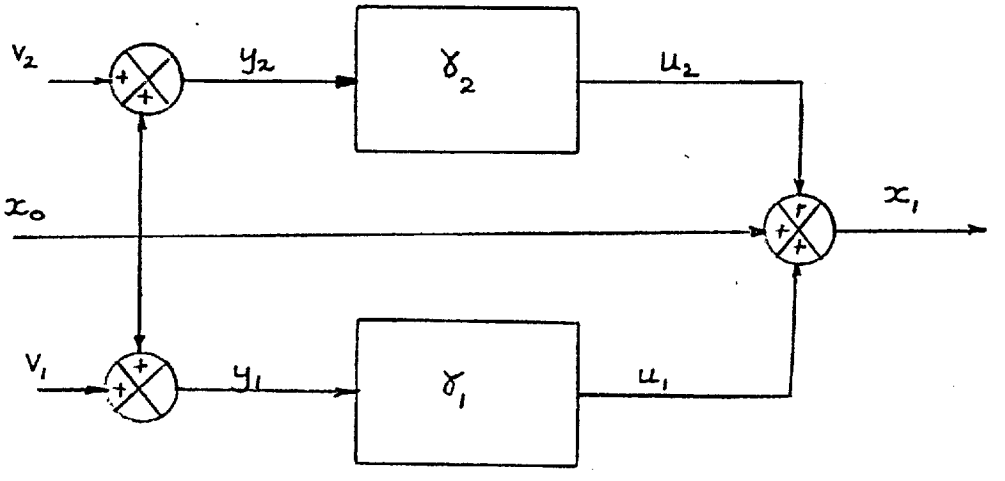


Fig. 9.1 Static Team Problem

Essentially, the state vector x_0 is observed noisily by the two controllers. They must determine control functions γ_i such that $u_i = \gamma_i(y_i)$ results in minimal expected cost J, where $J = E(x_1^2)$.

Clearly,

$$J = E\left\{x_1^2\right\}$$

$$= E\left\{x_0^2 + \gamma_1^2(y_1) + \gamma_2^2(y_2) + 2x_0\gamma_1(y_1) + 2x_0\gamma_1(y) + 2\gamma_1(y_1)\gamma_2(y_2)\right\}$$

$$= J_c + E\left\{\gamma_1^2(y) + \gamma_2^2(y) + 2x_0\gamma_1(y) + 2x_0\gamma_1(y) + 2\gamma_1(y_1)\gamma_2(y_2)\right\}$$

$$= J_c + J_a(\gamma_1, \gamma_2, p(x_0)).$$

Here we have separated out J into two components, one J_c is independent of (γ_1, γ_2) and hence for our purposes constant, the second J_a does depend on (γ_1, γ_2) and is sometimes called the active component of the cost. If γ_1^o and γ_2^o are the optimal controls then we have the following necessary condition for optimality for γ_1^o for example:-

$$J_a(\gamma_1^o, \gamma_2^o) \leq J_a(\gamma_1, \gamma_2^o) \text{ for all } \gamma_1$$

This type of condition is often referred to as person by person optimality.

Thus,

$$J_a(\gamma_1^\circ, \gamma_2^\circ, p_0) = \min_{\gamma_1} J_a(\gamma_1, \gamma_2^\circ, p_0)$$

$$= \min_{\gamma_1} E\{\gamma_1^2(y_1) + \gamma_2^{\circ 2}(y_2) + 2x_0\gamma_1(y_1) + 2x_0\gamma_2^\circ(y_2) + 2\gamma_1(y_1)\gamma_2^\circ(y_2)$$

$$= \min_{\gamma_1} E\Big[E\Big[\gamma_1^2(y_1) + \gamma_2^{\circ 2}(y_2) + 2x_0\gamma_1(y_1) + 2x_0\gamma_2^\circ(y_2) + 2\gamma_1(y_1)\gamma_2^\circ(y_2)\Big] \mid y_1\Big]$$

$$= E\Big[\min_{u_1}\Big[E\Big[u_1^2 + \gamma_2^{\circ 2}(y_2) + 2x_0u_1 + 2x_0\gamma_2^\circ(y_2) + 2u_1\gamma_2^\circ(y_2) \mid y_1\Big]\Big]\Big]$$

This is minimised when:-

$$u_1 + E[x_0 \mid y_1] + E[\gamma_2^\circ(y_2) \mid y_1] = 0.$$

and by symmetry we have

$$u_2 + E[x_0 \mid y_2] + E[\gamma_1^\circ(y_1) \mid y_2] = 0.$$

Thus we have two simultaneous equations to solve to determine the optimal γ_1° and γ_2°. Notice that this is not an easy problem in general, they

are functional equations in the sense that knowledge of γ_2^0 for all y_2 is required in order to compute u_1 for a particular y_1^2. However, suppose that x_0 and v_i $i = 1,2$ are gaussian random variables with zero means and variances σ_x^2 and σ_v^2. Further, suppose that $\gamma_i(y_i) = a_i y_i + b_i, i = 1,2$. Then:-

$$i) \quad E\left[x_0 \mid y_i\right] = \frac{\sigma_x^2 y_i}{\sigma_x^2 + \sigma_v^2},$$

$$2) \quad E\left[\gamma_i^0(y_i) \mid y_j\right] = \frac{a_i \sigma_x^2 y_j + b_i}{\sigma_x^2 + \sigma_v^2}, \; i \neq j.$$

Substituting these expressions into our necessary conditions above, we find that a_i and b_i can be determined uniquely such that the necessary conditions are satisfied. Radners contribution was to show that this argument can be extended to a more general multi-dimensional case subject to the traditional restrictions on the matrices of the quadratic cost funotion. Before we leave the static team problem, note that:-

1) Radner's result fails if x_0 and v_i are not both gaussian.

2) We have the phenomenon of second guessing appearing in the necessary condition. That is, we have to evaluate controller 1's estimate of controller 2's action.

We shall return to the static team problem in Section 9.3 when we discuss so-called partially nested information patterns. Next however we shall examine the topic of separation.

9.2 The Separation of Estimation and Control

The Witsenhausen counter-example has been referred to and discussed on many occasions previously. Indeed, it was a stimulus for the work in the first place. It is not surprising that following his identifica- of the importance of information patterns, Witsenhausen published a definitive paper on this topic which has to some extent directed the course of subsequent research (57). This paper contained a number of important definitions and concepts which will be discussed below. It did not however contain any proofs but instead a number of assertions, for which in some cases proofs were not then available. Some of these proofs have only recently been provided (see Kurturan (37) for example). It is not our intention to discuss Witsenhausen's paper in detail. Indeed, the paper bristles with an abundance of ideas which are worthy of far greater amplification and comment than is possible here. Instead we propose to simply mention some of Witsenhausen's terminology and attempts to link it to the development of Chapter 3.

9.2.1 Information Pattern

The information patterns considered throughout this thesis have generally been of the single controller, single observer variety. Witsenhausen's treatment encompasses the possibility of K control stations acting at a single time instant, with M observation posts taking measurements at each time instant. In particular, Witsenhausen discusses the important class of so-called delayed sharing patterns. Such a system can be said to consist of K combined control and observation stations, each

with perfect memory of its own observations and actions. In other words, a station receives the outputs from its observation post immediately and it remembers the control it applies. The stations share their data with a delay of n steps. We shall have more to say about delayed sharing patterns later.

9.2.2 Field Bases and Conditioning

Recall that in Chapter 3, we were concerned with emphasising:-

1) That until the control laws of a stochastic optimal control problem are specified, the problem is not generally well defined. We have merely quantities that are 'random variables to be'.

2) That depending on the information pattern, information about some control laws can perhaps be withheld. For example, in the classical information pattern case, $p(x_k \mid I_k)$ is independent of the control laws. Conversely, for the zero memory problem $p(x_k \mid I_k)$ cannot be computed until $\gamma_0, \gamma_1, \gamma_2 \cdots \gamma_k$ have been specified.

In Chapter 3 these questions were answered on an ad-hoc basis. Witsenhausen however proposes a more general (but more abstract) approach by introducing the notions of a field basis and a conditioning basis.

9.2.3 Equivalences and Substitutions

Again this topic is a formalisation of ideas we have already encountered. If $\underline{w}$ denotes the primitive random variables, then any system variable z can be expressed as a function of the primitive random variables and $\underline{u}$, the set of all control variables. That is,

$$z = f(u,w)$$

However for any design $\Gamma = \gamma_0, \gamma_1, \gamma_2 \cdots \gamma_N$, the dependence on u can be eliminated by recursive substitution in the system equations, as illustrated in Chapter 3. Essentially, this turns the control variables into random variables by a substitution:-

$$u = S(w, \Gamma)$$

Using this solution process, every system variable $z = f(u,w)$ becomes a parametric family of random variables

$$z = f(w, S(w, \Gamma)) = g(w, \Gamma),$$

parameterised by the design Γ. This idea now permits us to discuss the equivalence between two designs. Two designs Γ and Γ^* are called equivalent when $S(w, \Gamma) = S(w, \Gamma^*)$ for almost all w. Thus for any system variable z, we have almost surely $g(w, \Gamma) = g(w, \Gamma^*)$. Following this establishment of equivalence between designs, we can then proceed to define the equivalence between information patterns, and between conditioning bases.

The ideas of field bases, conditioning bases, equivalences, and substitutions are at the heart of separation and estimation in stochastic optimal control problems and they provide Witsenhausen in his paper with a powerful overview of the field. However, given a particular problem the ad hoc approach of Chapter 3 is probably the most straightforward.

9.2.4 Separation for Classical Information Patterns

In view of its importance, we review below the separation properties of various kinds of classical information pattern problems. Most of these properties were established at various points in Chapter 3; they are repeated here collectively, together with some appropriate remarks.

1. General System, Noise Immaterial, Cost Immaterial

We have seen that the optimal control law can be expressed as funcional of the 'state' $p(x_k \mid I_k)$. That is:-

$$\gamma_k^o(I_k) = \Phi_k(p(x_k \mid I_k))$$

In theory, Φ_k can be determined by backward induction.

2. Linear System, Gaussian Noise, Cost Immaterial

In this case the 'state' $p(x_k \mid I_k)$ can be summarised by its sufficient statistics - its mean (which is a linear function $\mathcal{L}_k$ of the control data I_k) and its variance (which is independent of I_k).

Thus we have:-

$$\gamma_k^o(I_k) = \Phi_k(p(x_k \mid I_k) = \Phi_k^*(\mathcal{L}_k(I_k))$$

Again Φ_k (or Φ_k^*) can be determined by backward induction. Note that Φ_k^* and consequently γ_k^o are not generally affine.

The property in the above solution which is crucial is that the variance is independent of the data. This property has been called neutrality in the literature (see for example Bar-Shalom and Edison (6)) where its appearance has been associated with the certainty equivalence principle (see below). This, and a similar property called quadraticity have more recently been discussed by Aldrich and Peterson (1)).

3. Linear System, Noise Immaterial, Cost Quadratic

Because we do not have gaussian noise, $p(x_k \mid I_k)$ is no longer Gaussian and the mean value is no longer affine in the data. Let us denote this mean by $F_k(I_k)$. The interesting property of the problem now is that:-

$$\gamma_k^o(I_k) = \Phi_k(p_k \mid I_k) = \bar{\Phi}_k(F_k(I_k))$$

The function $\bar{\Phi}_k$ is the feedback form of the optimal control for the deterministic system obtained by fixing the primitive random variables at their mean values. In this case $\bar{\Phi}_k$ is affine and we have

$$\gamma_k^o(I_k) = \bar{\Phi}_k \cdot E\left[x_k \mid I_k\right] = E\left[\bar{\Phi}_k(x_k) \mid I_k\right] .$$

This is an example of the Certainty Equivalence Principle. The fact that $\bar{\Phi}_k$ for our problem can be determined by solution of the corresponding deterministic problem is not a property which was discussed in Chapter 3, since we adopted the fairly traditional approach to the problem via the Kalman filtering equations, which require gaussianity. This approach is the one which permeates most of the literature on the optimal controlof linear stochastic system. That this should be so is understandable because it was not until 1969 that it was shown by Root (44) that the requirement of gaussianity could be relaxed. Unfortunately the introduction of the filtering equations has tended to add a complexity to the problem which later work has shown unnecessary.

We have already mentioned the work of Bar-Shalom and Edison (6) in connection with the Certainty Equivalence Principle. Their work appears to have removed much of the confusion on the topic and we shall briefly review some of their contributions:-

Dual Effect

This is a notion we have already encountered. Let us provide a somewhat formal definition. Let:-

$$m^{r}_{k,i} \triangleq E\left[\left(x_{k,i} - E\left[x_{k,i} \mid Y_k, U_{k-1}\right]\right)^{r} \mid Y_k, U_{k-1}\right]$$ be the r^{th} central moment of $x_{k,i}$ conditioned on Y_k, U_{k-1} (where $x_{k,i}$ is the i^{th} component of the state at time k). Then the control is said to have no dual effect of order r($r \geq 2$) if -

$$E\left[m^{r}_{k,i} \mid Y_j, U_{k-1}\right] = E\left[m^{r}_{k,i} \mid \tilde{Y}_j\right], \text{ for all } U_{k-1}, i, \text{ and } j \leqslant k,$$

where $\tilde{Y}_j$ are the measurements of the corresponding autonomous system - that is, $\tilde{Y}_j$ are the measurements if $U_{j-1} = 0$.

The lack of dual effect is sometimes called neutrality.

Certainty Equivalence Principle

Based on the above definition, Bar-Shalom and Edison prove a very powerful result on the application of the certainty equivalence principle. It is as follows:-

Consider a multi-dimensional process with linear dynamics and additive white, but not necessarily Gaussian, noise given by

$$x_{k+1} = A_k x_k + B_k u_k + w_k,$$

with observation equation

$$y_k = h_k(x_k, v_k),$$

where v_k is the measurement noise with known but arbitrary statistics. The only restrictive assumption on the measurement noise sequence is that it is independent of the process noise. The usual quadratic cost function is assumed.

Bar-Shalom and Edison then show that the optimal stochastic control for the above system has the certainty equivalence property if and only if

the control has no dual effect of second order, that is the updated covariance is not a function of the past control sequence.[1]

This is consistent with what we would expect from our dynamic programming arguments of Chapter 3. To establish the absence of dual effects we employ an analytic trick, as follows.

Establishing Neutrality

Consider a system with linear observations:-

$$y_k = H_k x_k + v_k,$$

with no restrictions on v_k. We now divide the system into autonomous and forced components as follows:-

$$x_k \triangleq \bar{x}_k + \underline{x}_k$$

where $\bar{x}_k$ is the state of autonomous part of the system, that is:-

$$\bar{x}_{k} = A_k \bar{x}_k + w_k,$$

with corresponding observation component $\bar{y}_k$, where:-

$$\bar{y}_k = H_k \bar{x}_k + v_k,$$

and where $\underline{x}_k$ is the state of the forced noiseless part of the system:-

$$\underline{x}_{k+1} = A_k \underline{x}_k + B_k u_k$$

1. The sufficiency of neutrality for certainty equivalence had already been suggested by Patchell and Jacobs (31).

with:-

$$\underline{y}_k = H_k \underline{x}_k$$

The initial state of the forced system is taken as:-

$$\underline{x}_0 = 0$$

Now, by superposition because everything is linear, $\underline{x}_k$ satisfies the process equation, and:-

$$\underline{y}_k = \bar{\underline{y}}_k + \underline{y}_k$$

Notice that at each k, the quantities $\underline{x}_k$ and $\underline{y}_k$ are known because they depend only on u_{k-1}. Therefore

$$E\left[\underline{x}_k \mid Y_k, U_{k-1}\right] = E\left[\bar{\underline{x}}_k \mid Y_k, U_{k-1}\right] + \underline{x}_k$$

$$= E\left[\bar{\underline{x}}_k \mid \tilde{Y}_k\right] + \underline{x}_k,$$

and $$\mathrm{cov}\left[\underline{x}_k \mid Y_k, U_{k-1}\right] = \mathrm{cov}\left[\bar{\underline{x}}_k \mid \tilde{Y}_k\right]$$

In other words, the control is neutral and certainty equivalence holds. Bar-Shalom and Edison show that even for certain classes of non-linear measurements, the same technique can be applied.

To be able to apply the certainty equivalence property in a particular problem is a great advantage since we reduce the difficulties to those of filtering, which could still be difficult but not as difficult as the overall optimal control problem. It is unfortunate that the benefits of neutrality have been so long in being widely appreciated. In two

'standard' textbooks on stochastic optimal control, Astrom (5) and Kushner (37) the requirement of gaussianity is still emphasised, apparently exclusively. More reprehensible, perhaps, is the manner in which the classical information pattern example in Chapter 4 is treated. The system there is linear and quadratic - consequently certainty equivalence could have been applied to reduce the computational burden. The example has however been retained in its primeval form to demonstrate the general technique which can be applied, in principle, to any classical information pattern problem. Similar apologies must be offered for the perturbation analysis in Chapter 6. Here our example was a linear system with quadratic costs; the only difference between it and the standard LQG problem, we asserted, was in that we would allow an arbitrary initial probability density function $p_0(x_0)$. By exploiting certainty equivalence, we see that this is not a particularly generous gesture anyway, and we could have handled the same problem just as well with a much reduced algebraic overhead. Thus, whilst the analysis for arbitrary $p_0(x_0)$ is perhaps of some limited academic interest it is of no practical significance whatsoever. Again, we can only defend our presentation in Chapter 6 on the grounds that it was our intention to discuss a general technique for classical information pattern problems and not one which relates to problems with a specific structure.

4. Linear Systems, Gaussian Noise, Quadratic Cost

Here, all the properties in 2,3, and 4 above apply and we have:-

$$\gamma_k^{\circ}(I_k) = \bar{\Phi}_k \cdot \mathcal{L}_k(I_k),$$

and since both $\bar{\Phi}_k$ and $\mathcal{L}_k$ are affine, we see that $\mathcal{L}_k^{\circ}$ is affine.

In addition to his survey of separation and estimation (which largely applies to classical information pattern problems), Witsenhausen's paper included a number of other topics, such as the difficulty of finding linear solutions to non-classical problems. These topics will be dealt with separately below.

9.3 Information Structures

The topic of the information structure in a dynamic multi-person control problem has received attention from a number of authors (Ho and Chu (24), (25), (13), Witsenhausen (57), Cole and Sage (15)). Cole and Sage in fact present a very readable appraisal of the topic and the discussion below is based on their paper, (although its content is due largely to Ho and Chu).

Let us first attempt to define information structure. We shall present this in the parlance of team theory, and in what follows we have plagiarised freely the work of Cole and Sage.

9.3.1 Information Functions

Simply stated, the information structure tells 'who knows what' in a particular problem. In other words, the information received by one player tells him certain things about the information available to and decisions made by other players. In addition this information may be perfect or noise-corrupted.

Suppose we have a team of N members indexed by the set $I = 1,2,\ldots,N$. The i^{th} member receives information z_i and on the basis of this information must make his decision u_i, which is some function of the information - $u_i = \gamma_i(z_i)$. The control law γ_i is chosen from the set of decision rules Γ_i. If the common performance index is denoted by $J(\gamma_1, \gamma_2, \ldots, \gamma_N)$, then the problem may be stated as: determine $\gamma_i^* \in \Gamma_i$ for all $i \in I$ such that $J(\gamma_1, \gamma_2, \ldots, \gamma_N)$ is minimised, or

$$\min_{\gamma_1 \ldots, \gamma_N} J(\gamma_1, \ldots, \gamma_N)$$

We shall distinguish between two quite different types of information. The first is the a priori information sometimes called the primitive information, and is represented by the symbol $\S \in R^n$, where R^n is an n-dimensional Euclidean space. Note that $\S$ is independent of any of the decisions. All of the uncertainty and randomness of the problem is embedded in this vector . We assume that $\S$ is gaussianly distributed with covariance $X > 0$, and, without loss of generality, $E(\S) = 0$. Thus $\S \sim N(0,X)$. In a typical problem the components of $\S$ may be random initial conditions and the various plant and measurement noises.

The second type of information which we shall treat is called individual information. This is the information available to a decision maker which is not shared by all other members of the team. An example of this might be information about the decisions made by other members of the team. Thus each member receives certain data concerning the a priori information and the decisions made by other players. For the i^{th} member we may write

$$z_i = \eta_i(\S, u_j;\ j = 1,2,\ldots,N,\ j \neq i)$$

where $z_i \in R^{q_i}$ and $u_j \in R^{k_j}$ and

$$\eta_i : R^n \times R^{k_j} \to R^{q_i}$$

A specification of η_i for every i is the information structure of a problem.

9.3.2 Causality

Given the definitions of γ_i and η_i above we can write:-

$$u_i = \gamma_i \left[\eta_i (\xi, u_j;\ j = 1,2,\ldots,N,\ j \neq i) \right]$$

Hence we have a closed-loop equation for u_i but we do not as yet know if it has a solution or if that solution, if it exists, is unique. Witsenhausen (54) has shown that the condition of causality is sufficient to guarantee the existence of a unique solution. By causality we mean that actions which have not yet occurred cannot influence a decision maker's information, or briefly, the future cannot influence the past.

9.3.3 Precedence Diagrams

We shall assume that z_i is a linear function of ξ and the u_j.

$$z_i = H_i \xi + \sum_{j \neq i}^{N} D_{ij} u_j \quad \text{for all } i.$$

One consequence of causality is that in the above equation if $D_{ij} \neq 0$, then $D_{ji} = 0$ for all i and j. In other words, if the decision of j affects the information received by i, then the decision of i cannot affect the information received by j.

In order to define more rigorously the role of causality in information structures we must first describe the concept of precedence. Consider the matrices D_{ij} (for all i,j). If and only if $D_{ij} \neq 0$, then j is said to be related to i, written jRi. Next j is said to be a precedent of i, written $j \{ i$, if and only if one of the following is true: (1) jRi,

or (2) there exists integers x,y,..., z $\in$ I such that jRx, xRy,... and zRi. The precedence diagram is a graphical means of portraying the information structure of a problem. Each decision maker is represented on the preference diagram by a node. If $j \{ i$, then the node corresponding to j is placed above the node corresponding to i. If jRi, then there is a solid directed line drawn from node j to node i. There is a path from node j to node i if and only if $j \{ i$. The starting decision makers are those whose decsions depend only on the a priori information ξ. The terminating decision makers are those whose decisions cannot influence the information received by any other decision maker.

Often it turns out that the information received by one decision maker j is embedded in the information received by another decision maker i when $j \{ i$, (j precedes i). If this is true for all i and j such that $j \{ i$, then the information structure is said to be partially nested. What this means is that in a partially nested information structure any decision maker knows exactly what his precedents have known. Since all the control laws are deterministic functions $u_i = \gamma_i(z_i)$ of the information, if a member knows what his precedents have known he can also determine the decisions made by his precedents. This perfect transmission information from j to i in an information structure is represented on the precedence diagram by a broken line parallel to the corresponding solid line from j to i. A typical precedence diagram for a partial nested information pattern is shown overleaf in Fig. 9.2.

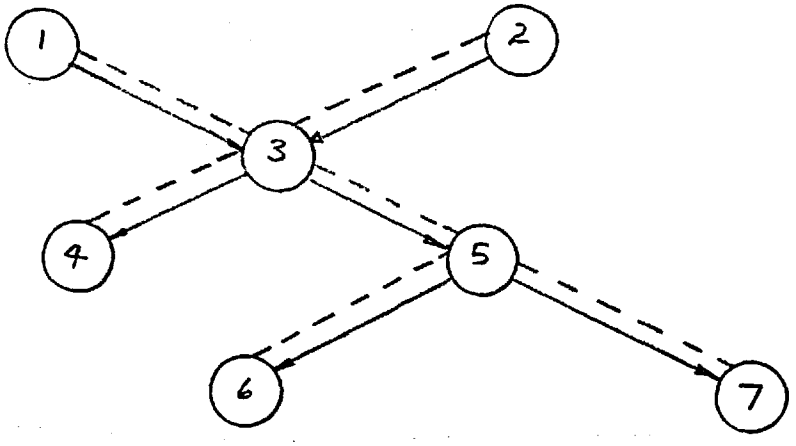


Fig. 9.2 Typical Partially Nested Precedence Diagram

The reason why are introducing the concept of a precedence diagram, particularly one for a partially nested information structure is that Ho and Chu use the device to provide a very interesting analysis of the multi-stage linear quadratic gaussian control problem. It is shown that this problem (which is termed a dynamic team with partially nested information structure) is equivalent to a static team problem.

Thus if we are given a dynamic team decision problem with a partially nested information structure, we can transform it into an equivalent static problem. Then Radner's(42) results described in 9.1 can be applied directly. Therefore a dynamic team decision problem with a partially nested information structure possesses a unique optimal solution which is affine in the information z_i. This is a very elegant result since it gives us a very quick way of showing the LQG classical information pattern problem has a linear solution in $(y_1, y_2 \ldots y_i;\ u_1, u_2 \ldots u_{i-1})$.

9.3.4 Application to Non-Classical Information Patterns

Surprisingly, the partial nesting technique can be directly applied quite successfully to a number of non-classical information pattern problems. In(24) for example, it is applied to the delayed sharing information pattern problem. However, since the technique does not give explicit solutions, we shall postpone a discussion of the delayed information sharing pattern based on dynamic programming principles until the next section. In(13), Chu extends the ideas of partial nesting to quite general information structures - the basis of his technique is presented next.

Suppose problem Ω has an information structure $z_i = \eta_i(\xi, u_j, j \subset i) = H_i\xi + \sum_{j \subset i} D_{ij}u_j$ which is not partially nested. Chu constructs an auxiliary problem denoted by $\tilde{\Omega}$ with the partially nested information structure $\tilde{z}_i = \tilde{\eta}_i(\xi, u_j; j \; i)$ obtained from the original structure by augmenting each z_i with the information received by each of its precedents. Each information variable z_i is now a vector with the information of its precedents $z_j (j \subset i)$ embedded as components of z_i. Thus the information structure $\tilde{z}_i$ is 'more informative' than z_i in the sense that knowledge of $\tilde{z}_i$ is sufficient to determine z_i.

Since the auxiliary problem $\tilde{\Omega}$ possesses a partially nested information structure it can be converted to an equivalent static problem and the optimal controls $\tilde{\gamma}_i^*$ which are linear in the information z_i can be determined. Also the cost function J^* for these controls can be evaluated. Since the amount of information contained in the set $\{\tilde{z}_i;\ i = 1,2,\ldots,N\}$ is at least as great as the amount of information contained in the set

$\{z_i : i = 1,2,\ldots,N\}$, the optimal cost of $\tilde{J}^*$ of the auxiliary problem can be no greater than the optimal cost J^* of the original problem. Hence $\tilde{J}^* \leq J^*$, i.e. the optimal cost obtained using the controls γ_i^* (which we know how to determine) provides a lower bound on the optimal cost function of the original problem. Recall that a similar technique was employed in the zero-memory ergodic problem in Chapter 7 to compute a lower bound.

The question naturally arises as to whether we can use the optimal controls $\tilde{\gamma}_i^*$ for the auxiliary problem $\bar{\Omega}$ to determine the optimal controls γ_i^* for the original problem Ω. To answer this question we must first define the concept of a composite function. We have:-

$$u_i = \gamma_i(\eta_i(\xi, u_j ; j \{ i))$$

Suppose we define p_i as follows, where N_1 denotes the set of starting decision makers:-

$$p_i(\xi) = \gamma_i(\eta_i(\xi)), \quad i \in N_1$$

and

$$p_i(\xi) = \gamma_i(\eta_i(\xi, p_j(\xi) ; j \{ i)), \quad i \notin N_1$$

We desire to find a set of controls $(\gamma_1, \gamma_2, \ldots, \gamma_N)$ such that the corresponding composite functions $p_i = \gamma_i \eta_i$ are equal for all i to $p_i^* = \tilde{\gamma}_i^* \eta_i$, where $\tilde{\gamma}_i^*$ for all i are the optimal controls for the auxiliary problem $\bar{\Omega}$. But under the controls γ_i such that $\gamma_i \eta_i = p_i = \tilde{p}_i^*$, we obtain a cost J such that $J = \tilde{J}^* \leq J^*$. Thus, $J = J^*$ - the optimal cost

Suppose for agiven team problem Ω we define an auxiliary problem $\tilde{\Omega}$ with a partially nested information structure and determine the optimal controls γ_i^* and the corresponding composite functions $\tilde{p}_i^*$ for all i for the auxiliary problem $\tilde{\Omega}$. We define the functions

$$g_i(\S) = \eta_i(\S, \tilde{p}_j^*(\S)) \; ; \; j \{ i) \quad \text{for all } i$$

This represents the information received by the decision makers in problem Ω as a result of applying the composite optimal decisions $\tilde{p}_i^*$ of problem $\tilde{\Omega}$.

It is shown by Chu that the optimal controls for the original problem are given by:-

$$\gamma_i = p_i^* g_i^{-1} \quad (i = 1,2,\ldots,N)$$

The weakness in the above approach is in the optimistic hope that we shall find functions g_i which are invertible. In fact, the inversion of g_i seems to correspond to an inversion of the one-zero permutation matrices which were discussed in Chapter 3, (and called information functions). Certainly, we would be lucky to encounter problems in which our B_i matrices can be inverted. Nevertheless, Chu (13) does present some very interesting examples of the application of the technique. To cope with non invertibility in g_i, it is also suggested that linear controllers can be prescribed to determine an upper bound for the problem. This suggestion should be viewed with scepticism however since, as we have seen in Chapter 7, and as we shall discuss later, linear solutions are not necessarily easy to find.

Despite the above reservations, the work of Ho and Chu is of considerable significance since it gives us a consistent approach to linear dynamic team problems with any kind of information pattern, involving various hierarchial structures for example. In addition, qualitative statements about the value of communication links in the network can be made.

This concludes our discussion of information structures.

9.4 Delayed Sharing Information Pattern Problems

In this chapter we have mentioned the above class of problem on a number of occasions. In this section we shall discuss the topic in more detail because it represents one area of decentralised control where some positive results can be obtained. A number of workers have made contributions here, notably Ho and Chu (25), Sandell and Athans (46), Kurturan (35), (36), and (37), Yoshikawa (60), and Toda and Aoki (52). Kurturan has gone further than most authors in that he has considered the n-step delayed sharing pattern. Consequently we shall base our discussion on his work (37).

Consider a typical N-step discrete time system as below, in which observations are made from two measurement posts and inputs are applied from two control stations:-

$$x_{k+1} = f_K(x_K, u_K^1, u_K^2, w_K)$$

$$y_k^1 = g_K^1(x_k, w_k^1)$$

$$y_k^2 = g_K^2(x_k, w_k^2),$$

where x_0, $\{w_K, v_k\}$ are independent random variables with given probability distributions. A quite general cost function can be defined provided it is additive. The information pattern for each control station is I_k^i where:-

$$I_k^i = \left\{ y_k^i, y_{k-1}^i \ldots, y_{k-n+1}^i, \delta_k^n \right\}, \quad i = 1,2$$

and δ_k^n denotes the common information if the two control posts at time is:-

$$\delta_k^n = \Phi \text{ (empty)} \quad \text{for } 0 \leq k \leq n,$$

and $$\delta_k^n = (Y_{k-n}) \text{ for } k > n$$

Using dynamic programming arguments, Kurturan demonstrates that for the n-step delay sharing problem, there is no loss if one restricts designs to be of the form:-

$$u_k^i = \Phi_k^i (y_k^i, y_{k-1}^i, \ldots y_{k-n+1}^i, p(x_{k-n+1} | \delta_k^n))$$

It is gratifying to note that this result is established by very similar techniques to those used to establish Theorem 3.2 (and which are equally as intricate). The 'state' for the dynamic programming formulation is $p(x_{k-n+1} | \delta_k^n)$. A useful looking separation theorem can be established by restricting attention to linear gaussian systems (cost immaterial). In this case, there is no loss if one restricts designs to be of the form:-

$$u_k^i = \Phi_k^i (y_k^i, y_{k-1}^i, \ldots y_{k-n+1}^i, \hat{x}_{k-n+1}),$$

where:-

$$\hat{x}_{k-n+1} = E\left[x_{k-n+1} | \delta_k^n \right],$$

and can be generated by a Kalman filter.

To obtain an appreciation of the above results, it is useful to consider the one step delay sharing pattern. This can be viewed as the usual dynamic programming result, but in which the minimisation at each point in the backward induction process is essentially a static team problem (see 9.2). By extension, an n-step delay sharing pattern would appear to consist of a backward induction process which involves a dynamic team minimisation problem at each stage (with no lateral sharing of information), (see for example Yoshikawa (60)).

In general the functions Φ_k^i are non linear in their arguments. However for the LQG problem with a one step delay sharing pattern, we know from Radner's theorem (42) that Φ_k is linear. In this case it is therefore possible to derive explicit equations for the optimal feedback gains.

In conclusion it is interesting to note that the delayed sharing information pattern problem has been approached in various directions by different authors. Thus Ho and Chu (24) use the concept of partial nesting. Sandell and Athans (46) use the matrix maximum principle, Yoshikawa (60) uses dynamic programming. In one paper Kurturan (36) uses an identity from the calculus of variations to transform the cost function into a more useful form (see for example Åström (5) page 278).

Finally, let us mention the notion of 'infinite second-guessing'. Although the concept was introduced by Toda and Aoki (52) in the context

of the one-step delay sharing pattern problem, it is really concerned with static optimisation problems. Recall that we obtained the following necessary conditions for a two-person static team problem in Section 9.2:-

$$u_1 + E\left[x_0 \mid y_1\right] + E\left[\gamma_2^{\circ}(y_2) \mid y_1\right] = 0,$$

and $$u_2 + E\left[x_0 \mid y_2\right] + E\left[\gamma_1^{\circ}(y_1) \mid y_2\right] = 0.$$

Infinite second guessing is a formulation of an iterative procedure which is proposed for the solution of the above equation. Simply, guess γ_1° - denote it by $\gamma_1^{\circ}(0)$. Use this to compute $\gamma_2^{\circ}(0)$. Use this to compute $\gamma_1^{\circ}(1)$. Use this to compute $\gamma_2^{\circ}(1)$, and so on. It can be shown under suitable assumptions, the procedure will converge to the optimal solution.

9.5 Linear Solutions in Non-Classical Problems

In Chapter 7 we discovered that it was difficult to determine even optimal linear controls for our zero memory problem. The Witsenhausen problem demonstrated a similar awkwardness in yielding its optimal linear control laws (see Chapter 8). The purpose of the present section is to mention an analysis of a dynamic team problem by Whittle and Rudge (54) which provides some insight into this phenomenon.

A very simple situation is considered (called a symmetric team) with process equation:-

$$x_{k+1} = a\, x_k + b \sum_{i=1}^{n} u_{ik} + w_k,$$

and observation equations for each team member i:-

$$y_{ik} = x_k + v_{ik}$$

The cost function is of the form

$$v = E\left[x_k^2 + \sum_{i=1}^{n} u_{ik}^2 \right],$$

the expectations being taken on the assumption that $\{ x_k, u_{1k}, u_{2k} \cdots u_{nk} \}$ are jointly stationary.

All team members are assumed to use a control law of the form

$$u_{ik} = \sum_{s=0}^{\infty} h_s\, y_{ik-s} = H(T)\, y_{ik},$$

where $H(z) = \sum_{s=0}^{\infty} h_s z^s$ is the generating function of the coefficients h_s, and T is the backward translation operator T, which reduces to t to t-1.

Under these conditions, Whittle and Rudge show that the optimal $H(z)$ is irrational. This indicates that the recursion relations associated with effective state estimation are no longer of finite order.

9.6 Communication and Control

A recurring theme throughout this thesis has been concerned with the dual nature of the controller - it must control and in addition it must communicate. A number of workers have published material concerned with this aspect of decentralisation (Sandell and Athans (46), Bismut (9), and Whittle and Rudge (55)).

An interesting explanation of why non-linear control laws arise in non-classical information pattern problems is provided by examination of what is called the control sharing problem (Sandell and Athans (46), and Bismut (9). The problem is identical to the LQG one step delay pattern considered in Section 9.4, except in its information pattern. However the admissable control functions $\gamma = (\gamma^1, \gamma^2)$ are functions such that each controller remembers his own observations and has knowledge of all controls previously applied by both controllers.

A single stage version was originally examined by Bismut (9) who showed that the set of control laws admissable for the control sharing information pattern is dense in the set of control laws admissable for the one step delay sharing problem. Bismut's approach uses advanced and abstract mathematics; fortunately his result has been extended and proved in a much simpler manner by Sandell and Athans (46).

Intuitively, their result says that a control law γ can be modified slightly to encode information about another variable y. Notice however that the controller's problem is not merely to encode two real numbers

into one, but to encode his observation onto a second without disturbing the second. This task can only be accomplished to within ϵ . As an example (Sandell et al (48)), suppose that at time t controller i would apply the optimal control

$$u_i^o(t) = 0.3333 \ldots$$

if the information were classical. Suppose that his observation at time t is

$$y_i(t) = 0.6666 \ldots$$

Consider the uniqueness of the choice of control

$$u_i(t) = 0.3333 \ldots . \; 336666$$

Clearly the control specified above is an attempt by controller i to signal the value of his observation to the other controllers which in turn can decode $y_i(t)$ from $u_i(t)$. Moreover, the control $u_i(t)$ can be made arbitrarily close to $u_i^o(t)$ so that the substitution of $u_i(t)$ for $u_i^o(t)$ affects the system responses by an infintesimally small amount. Thus the cost can be made arbitrarily close to that of the classical information pattern problem. These signalling control laws are quite nonlinear. Notice that the control must be communicated exactly without noise for the scheme to work. It would seem also that rational control laws and observations would permit exact matching of classical information pattern performance.

A different examination of the communication aspects of control was presented by Whittle and Rudge (55). They construct an example in which both communication and control are both required but the controls operate on separate parts of the system, so that joint control is not required. The setting is that of conventional communications theory with arbitrarily long sequences from stationary processes. Consequently the control laws (which are generally non-linear) are not realisable. The situation is one in which controller 1 can observe a time sequence $\S$ and on the basis of this he chooses an action u which affects his state variable x. Controller 2 observes a noise corrupted version y of x, and on the basis of this he must estimate $\S$. Controller 1 incurs costs by applying u; Controller 2 incurs costs corresponding to the error in his estimate of $\S$.

The transformation $u \rightarrow y$ can be regarded as defining a statistical communication channel with input $\{u_t\}$ and output $\{y_t\}$. Corresponding to a fixed cost arising from the application of u, we can attach a fixed capacity C to the channel. We then have the problem of transmitting through a channel of capacity C in such a way that the processed output $\{\hat{\S}_t\}$ is a minimally distorted version of $\{\S_t\}$. This is a problem in source encoding which is solvable by methods well established in communications theory. By varying C, we can determine the minimal total cost of the estimation problem.

The example of Whittle and Rudge described above is one of the rare instances where communications theory has been applied to a control pro-

blem. The difficulty in general seems to be that information theory does not concern itself with the use to which transmitted information will be put. It is concerned with information as an end in itself, which is measured in terms of its 'surprise value'. The subsequent question to which control theory requires an answer is 'so what?' It would therefore seem that current information theory will be of little value in addressing the problems of non-classical information patterns in stochastic control.

Perhaps an alternative to information theory is provided by the pragmatic approach of Jacobs and Patchell (29), who report on some interesting experimentation for non-LQG problems with classical information patterns. Here it is postulated that the optimum control law will have three components:-

$$u^o = u_c + u_{caution} + u_{probing},$$

where u^o is the optimal control for the problem (which is usually impossible to find)

u_c is the certainty equivalent control which would be obtained if there were no uncertainty in the problem,

$u_{caution}$ is a component which reflects the accuracy of the estimates in the control law, and

$u_{probing}$ is a component which reflects the possibility of affecting the rate of reduction of uncertainty.

A similar experimental approach to non-classical problems might prove to be very enlightening. Perhaps the only extra component we would need is one which reflects the second-guessing phenomenon. It might be possible to synthesise the Witsenhausen problem in such a way that a control law can be derived which at least might outperform the best linear control law and even the non linear control law which Witsenhausen demonstrates is sometimes better. We might even get a clue to the identity of the optimal control law.

9.7 Large-Scale Systems

In the Introductory Chapter we established the place of de-centalised control in the wider context of large-scale system control (although this is not necessarily the sole domain of decentralised control). In this section we wish to simply introduce a few topics from the interesting area of large-scale control sustems. A comprehensive survey of the entire subject is given by Sandell et al (48).

Stability

Stability is a subject which we have largely ignored in previous chapters. It is nevertheless important for obvious reasons. The problem is that the restriction of information in a non-classical information pattern might induce an instability which would be absent in the case of classical information patterns (for example, Aoki (3)). Conversely, it has also been demonstrated that the restriction of information can induce stability (Sundareshan (51)). Could the reason why international exchange rates fluctuate so irrationally be that technology is providing too much information too quickly?

Hierarchial Control

In a hierarchial system, the optimisation divided into two levels. The lower level problem consists of independent optimisation problems which are dependent on a co-ordinating signal. The higher level problem is that of the co-ordinator who chooses the co-ordinating signal. A typical configuration is shown below:-

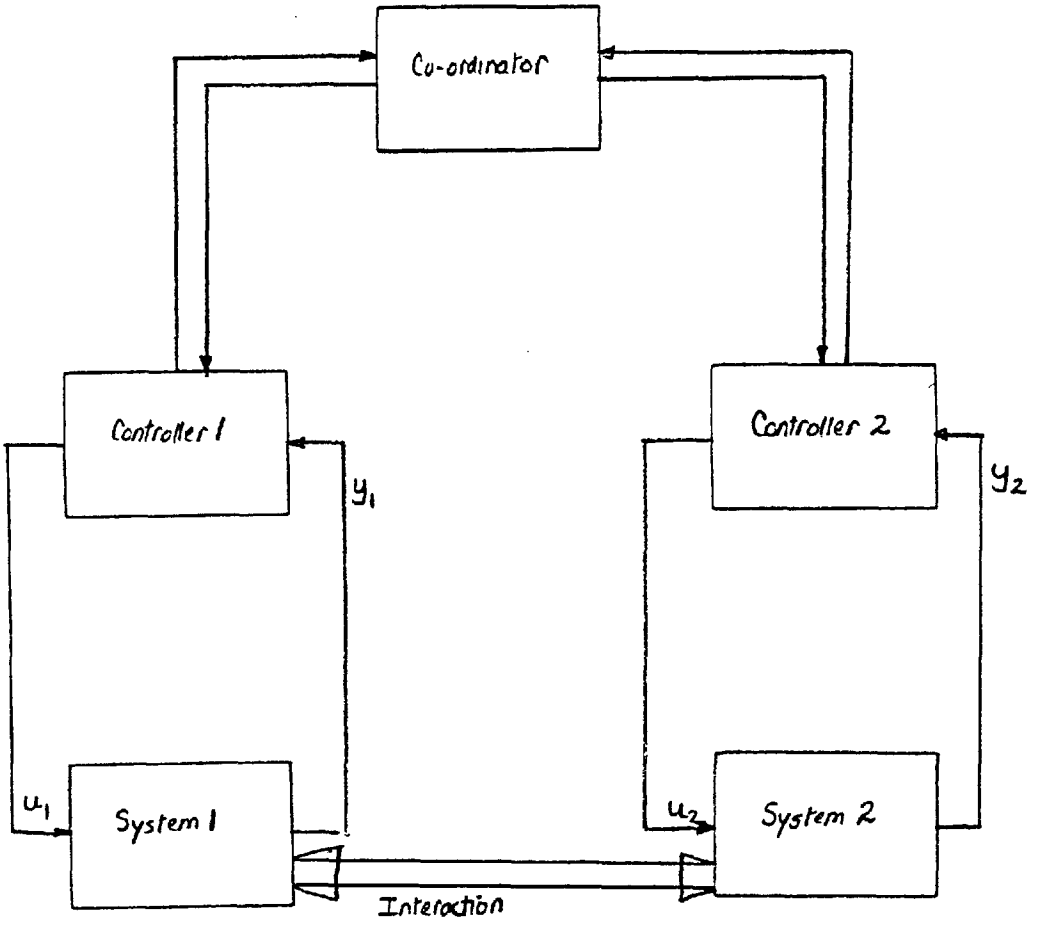


Fig. 9.5 Two Level Hierarchial Control Structure

Notice that the sub-systems 1 and 2 are not independent. The action of system 1 will affect system 2 and vice versa. Notice that the co-ordinator cannot have complete knowledge of the system (structure, observations, and so on). Otherwise we could dispense with the two controllers and leave it all up to the co-ordinator. To be realistic, we must im-

pose some restrictions on what the role of co-ordinator should be. For example, the co-ordinator could be told simply what the controllers have done, not what they have observed (control-sharing). But, as we have seen this requires noiseless communication, and perhaps this is unrealistic. Alternatively, we can consider perodic co-ordination in which the co-ordinator issues his co-ordinating signals less frequently (say monthly) than his sub system controllers who perhaps issue their control actions on a daily basis. This topic is discussed by Chong and Athans (12).

Whatever role is selected for the co-ordinator (in terms of the share of information between co-ordinator and control) it is necessary to model the interactions in the system by means of a constraint equation (as in mathematical programming). In this connection two principles have been developed - the so called 'interaction balance principle' and the interaction-prediction principle (see for example Sandell et al (48)). This concludes our brief discussion of large scale systems. One final note is of interest. In a 1976 paper, Basar (7) warns the control fraternity that linear deterministic systems (with x_0 random) may not have optimal linear solutions. And the example he uses is the Witsenhausen example.

9.8 Conclusions

In this chapter we have presented some comments on recent developments in decentralised (and centralised) stochastic optimal control which it is hoped will set the presentation of the preceding chapters in a contemporary context. We have examined the issue of separation and certainty equivalence. We have looked at the technique pioneered by Ho and Chu of Partial Nesting together with the more general concept of information structure. Some positive results for an important class of non-classical information pattern - the delay sharing information pattern problem - have been discussed. Next we discussed the often overlooked difficulty of finding even linear solutions to problems with non-classical information patterns. Some comments were made on the dual role of the control law in simultaneously controlling and communicating, and we concluded with some isolated remarks on large scale systems.

In considering current literature from the field of stochastic optimal control, the impression arises that interest in non-classical information pattern has been absorbed by the more practically significant problems of large scale systems. However the crucial dilemma of information versus control still requires to be resolved.

CHAPTER 10

Epilogue

		Page
10.1	Introduction	10.1
10.2	Review of Thesis	10.3
10.3	Conclusions	10.7

10.1 Introduction

This entire thesis was motivated by the sole desire to find a solution to the Witsenhausen problem. In this objective we have been unsuccessful. The Witsenhausen problem still remains an unresolved enigma. However, since its publication in 1968 the problem has performed a valuable service, it has focussed attention on what had been a neglected aspect of control theory - the information pattern. It has been realised that the consideration of information structure is the key to expanding the class of problems which can be examined by control theory to include a host of problems of very great practical significance. Conversely, an examination of singular information structures, such as zero-memory, has perhaps given us a more profound understanding of the classical information pattern problem which it had previously been assumed was well understood. This was certainly the author's experience and it is the reason why most chapters contain developments of both zero-memory and classical information pattern analysis. This impression has been confirmed by at least two correspondence items in the recent literature which comment upon conceptual errors which have appeared in published material. For example, Sandell and Athans (47) show that a proposed solution for a class of stochastic team problem is in fact incorrect. Similarly, Wernerson (53) argues that a proposed solution for classical information pattern problem with interrupted observations is incorrect. Without going into details here, it would seem that both of these errors could have been avoided had the authors adopted the admittedly ponderous but safe approach followed in Chapter 3.

In this concluding chapter we shall briefly review the contents of this thesis and summarise our conclusions. Finally we shall again emphasise the importance of the class of problems which we have been examining and we shall suggest areas for further investigation.

10.2 Review of Thesis

The simple example presented in Chapter 2 demonstrated most of the principles upon which subsequent chapters were based. We saw the need to carefully define the problem facing each controller (the direct approach or the indirect approach). We saw the need to work with control functions and not control values and we illustrated the substitution principle. But the most portentious feature of the example was the ease with which a solution for the perfect memory and classical information pattern versions could be obtained, whilst for the zero-memory problem an exceedingly complex functional minimisation was involved. We saw for the first time the appearance of the second guessing phenomenon. We also introduced our constant companion throughout - the Witsenhausen problem - and we illustrated the fundamental dilemma facing the controller; to communicate or to control.

In Chapter 3 we presented a much more general formulation of the stochastic optimal control problem in which a definition of information pattern was presented. The technique of Dynamic Programming was examined and we were able to produce, via an extremely cautious development, optimality equations for the zero-memory and perfect memory versions of our problem. We then proceeded to apply our dynamic programming necessary conditions to a simple linear system with quadratic cost and gaussian noise. In the zero memory problem, as we would expect from Chapter 2, we encountered a highly complex sub-problem involving minimisation with respect to an entire function. To solve the perfect memory problem we had to introduce the Kalman filtering equations in order to demonstrate a crucial pro-

perty of the LQG problem - that the conditional variance is independent of the control actions. At the same time we introduced the separation principle, the certainty equivalence principle, and the notion of a sufficient statistic.

The general principles developed in Chapter 3 were applied in Chapter 4 to a simple discrete state problem. Solutions were derived for both the zero-memory and classical information pattern versions of the problem. Besides illustrating how solutions can be obtained (with difficulty), it is believed that our example in Chapter 4 highlighted some important points regarding the information pattern of a problem. For example, the benefit of the extra information provided by the classical information pattern can be of very marginal value. Also, even for the classical information pattern problem it might not be necessary to remember all past observations and controls. Only certain observations are critical for achieving optimality - all the rest can be ignored. This raises the possibility of only transmitting the critical observations (if, of course, we can determine what they are).

Chapter 5 was devoted to a study of certain sub-optimal techniques and a discussion of a number of pitfalls which can arise in the pursuit of solutions to non-classical pattern problems. For the classical information pattern problem, we discussed such topics as dual control and open-loop optimal feedback control. For the zero-memory problem, the danger of adopting an approach such as the neo classical approach was discussed, and it was shown how, in the absence of realistic test examples, such an approach might lead to the wrong conclusions. We also discussed some sub-

optimal approaches to the zero-memory problem.

One important sub-optimal technique is provided by so called perturbation analysis and the whole of Chapter 6 was devoted to this topic. For the classical problem, the technique was applied for demonstration purposes to the familiar LQG problem. The Witsenhausen problem was also examined by perturbation analysis and an $o(\varepsilon^4)$ approximation was determined. The resulting control law had the interesting property that the correction term appears to represent a communication component. However, we were forced to conclude that perturbation analysis is not very useful as a design aid because of the complexity of the algebra which is involved.

Optimal stationary control was discussed in Chapter 7. We considered both the discount problem and the average cost per period problem (and the ergodic problem). A number of theorems were presented as zero-memory versions of those which arise in the traditional theory of stationary control, as applied to Markov chains and LQG systems. We also examined a number of approximation techniques, and introduced the notion of a static (as opposed to stationary) controller for the discount and average cost per period problem. The application of Edgeworth approximation, moment approximation, and perturbation analysis to the ergodic problem was demonstrated. For the discount problem, an approximation for the best linear control law was proposed, although the recusion relations which were derived were far from elegant and of dubious value.

In Chapter 8 we reported on the topic which has been the major source of delay in completing this thesis. We discussed the problem of showing that under certain circumstances the optimal control for the Witsenhausen problem is linear. Since a necessary condition for optimality in the Witsenhausen problem is satisfied by the linear solution, we attempted to show that the necessary condition will have a unique solution under certain circumstances. This proved impossible and all that we could do was prove uniqueness if the control law was bounded - which prohibits a linear control law.

The bulk of the work of the preceding chapters is dated circa 1971. Chapter 9 was an attempt to set this work in the context of more recent developments in the field. Topics such as the certainty equivalence principle, partial nesting, information structures, delayed information sharing patterns, simultaneous communication and control, optimal linear controls, and large scale system theory were mentioned.

Notice that it is not claimed that this thesis is either a survey or a bibliography. For a start, in 1972 there was little to survey of direct relevance anyway. What there was has been cited where appropriate. Secondly, very extensive surveys of the topics considered here have appeared recently in the literature (for example, Sandell et al (44)). Our prime concern has been to present a self contained introduction to the role of the information pattern of a problem and to offer some approximation techniques should the designer require them.

10.3 Conclusions

Decentralised control is a topic of immense practical significance as we argued in Chapter 1. What we have shown in the intervening chapters is that the traditional tools developed for the study of centralised control are not generally applicable to the de-centralised case. It is true that our traditional techniques can be used to derive approximate solutions to non classical problems. But throughout the entire spectrum of problems nowhere were we able to find exact optimal solutions. And the message of the Witsenhausen problem is unmistakeable - traditional techniques are simply inadequate. The Witsenhausen problem is simplicity itself in construction and yet its optimal solution can only be expressed as a solution to an exceedingly complicated necessary condition. The difficulty is apparently due to the dual role of the controller in that it must not only control but also communicate information to facilitate subsequent control. The classical theory is only well developed for these cases in which this dual effect is absent - when the communication of information is a feature of the problem then the classical techniques are apparently inappropriate. Only if the controller at some time instant gains nothing from a knowledge of the control actions of preceding controllers can classical techniques be applied. In general of course this singular situation will not prevail.

Non-classical information patterns can arise in many problems in the more traditional control engineering field. Most simply, all we need to do is visualise a not unrealistic situation in which the controller memory

is limited. For large scale systems and for systems comprised of mobile sub-systems, communication links between controllers are subject to delay and noise and should be regarded as part of the controlled system. Such systems arise for example in air-traffic control. If we broaden our horizon even further we realise that communication problems are control problems and what is more, that non classical information patterns will always occur since we shall always have at least two controllers (transmitter and receiver) which by definition will never have access to the same information. Proceeding further still, non classical information patterns arise in the management structure of most organisations. For example one management structure known as bureaucracy (not a term used in its popular sense) is characterised by specialisation and hierarchial control. No manager (controller) knows the intricate details of the entire enterprise. His role is the co-ordination of the activities of his sub-ordinates. Such structures appear to have arisen as a result of the finite capacity of the human mind (memory and intellect). Centralised control of the entire enterprise is beyond the capability of any one individual. Consequently decentralised control (in the form of delegation) has been introduced to cope with the situation. Clearly, the notion of non-classical information patterns can be introduced into every field of human activity. Civilisation has progressed through man's realisation that it is better to co-operate rather than to compete. But his ability to co-operate depends on his ability to communicate and in this respect his facilities are very sophisticated by comparison with the other animal forms. Nevertheless his capabilities are still limited and centralised control is out of the question; specialisation and communication are what makes it all work.

In stressing the ubiquity of non-classical information patterns the point we are making is that the state-of-the-art in decentralised control is well in advance of the theory. In fact, a study of these state-of-the-art systems reveals issues which are perhaps critical to the development of a theory of decentralised control. A technical example of this is provided by the current move towards de-centralised, or distributed, computing. Economics of scale arguments have led to the adoption of a centralised computing policy in many organisations. Thus a single centralised complex powerful general purpose computer is utilised for the processing of such diverse applications as payroll, stock control, invoicing, and so on. However it is becoming increasingly apparent that this approach does not provide a service which is totally satisfactory. Viewing the organisation as a hierarchial control structure (which it generally is), the problem seems to be that the computer has become an integral part of the control system. Management decisions (controls) are based on information (observations) provided by the computer. But management is decentralised whilst information is centralised - communication links are required to complete the loop. And these communication links are subject to delay, noise, and cost. Management's ability to effectively control operations is, it is argued, seriously hindered by the centralisation of information processing in what is essentially a decentralised control system. The solution which is proposed is to decentralise the information processing function. In other words, to implement a communications structure which reflects the control structure of the organisation. It is interesting to list some of the advantages which are claimed for decentralised computing:-

(i) reduced communication costs

(ii) increased reliability

(iii) reduced hardware and software costs

(iv) more effective management information

Each of the items above should be included in any theory of de-centralised control. That is, in addition to the usual cost function, such as manufacturing cost, additional costs should be introduced to represent the cost of communications, the value of reliability, the cost of system complexity, and the value of the information pattern (such as the cost of delays or errors). When a de-centralised computing approach has been implemented in an organisation it has usually been found to be highly beneficial. It is interesting to note that this has generally arisen because hardware (and software) has displayed a very human characteristic. This is concerned with the fact that a number of simple computers are generally much cheaper than one very sophisticated general purpose computer. Thus whilst we can construct 'superhuman' control computers of immense memory and intellect, we do much better to imitate the human trait of limited memory and intellect coupled with specialisation.

The current tools of control theory are unlikely to be able to handle the kind of multiple cost-function proposed above. Perhaps a few steps back are required, certainly the problem of de-centralised computing is a long way removed from the LQG problem. Perhaps new concepts of optimality are required. Perhaps new measures of information are also required. Perhaps we need to model the effects of noise in a different manner. For ex-

ample, in a digital communication network, channel noise causes retransmission to occur and this introduces delay and extra cost. These questions are not of academic interest alone; answers are required urgently for the benefit of society as a whole. The portents for the future are not reassuring. It is argued in some quarters that the introduction of microprocessor technology will lead to a society divided firmly in two halves - one half wealthy and employed and the other half poor and unemployed. Such predictions appear to hinge on the assumption that government will pursue a policy of maximising the single objective of economic growth. Other issues are of a lesser priority which can only be resolved within the limits imposed by the maximisation of economic growth. What is apparently required is a methodology for the simultaneous maximisation of several objective functions. Conversely, the information technologists argue that society will shortly be caught in the 'Entropy Trap' where the entire energy of one generation is spent in passing on information to the next generation. Here the assumption is that all information is of equal value and the possibilities of specialisation are not considered. These two contradictory predictions illustrate the central importance of topics such as decentralised control, the value of information and so on. We have shown in this thesis how classical techniques of control theory can only be of limited use in dealing with these topics. What is required is a new theory.

REFERENCES

1. ALDRICH C. 'Quadraticity and Neutrality in Discrete Time Stochastic Linear Quadratic Control'. Automatica, Vol. 13, pp 307-312, 1977.

2. AOKI M. 'Optimal Control of Partially Observable Markovian Systems'. Journal of the Franklin Institute, Vol. 280 No. 5, pp 367-386, November 1965.

3. AOKI M. 'Some Control Problems Associated with Decentralised Dynamic Systems'. IEEE Transactions on Automatic Control, October 1971.

4. ASTRÖM K.J. 'Optimal Control of Markov Processes with Incomplete State Information'. Journal of Math. Analysis and its Applications (10), pp 174-205, 1965.

5. ASTRÖM K.J. 'Introduction to Stochastic Control Theory'. Academic Press, 1970.

6. BAR-SHALOM Y. EDISON T. 'Dual Effect, Certainty Equivalence, and Separation in Stochastic Control'. IEEE Transactions on Automatic Control, Vol. AC-19, No.5, pp 494-500, October 1974.

7. BASAR T. 'On the Optimality of Non-Linear Systems in the Control of Linear Decentralised Systems'. IEEE Transactions on Automatic Control, Vol. AC-21, pp 797, October 1976.

8. BERTSEKAS D.P. Dynamic Programming and Stochastic Control'. Academic Press, 1976.

9. BISMUT J.M. 'An Example of the Interaction Between Information and Control - The Transparency of a Game'. Joint Harvard University - Imperial College Conference on Information Structure and Extensive Games, June 1972.

10. BOX
DAVIES
SWANN
'Non-Linear Optimisation Techniques'.
Oliver and Boyd, Edinburgh 1969.

11. CHIA C.W.
'Determination of Optimal Controls for Partially Observable Stochastic Systems with Non-Classical Information Patterns'.
Computing and Control Dept., Imperial College, London, S.W.7., Report No. 26/70, Sept. 1970.

12. CHONG C.Y.
ATHANS M.
'On the Periodic Co-ordination of Linear Stochastic Systems'.
Automatica Vol. 12, pp 321-335, 1976.

13. CHU K.C.
'Team Decision Theory and Information Structures in Optimal Control Problems - Part II
IEEE Transactions on Automatic Control, Vol. AC-17, No.1, February 1972.

14. CLARK J.M.C.
'Tchebycheff Inequalities for Conditional Moments'.
Department of Computing and Control, Imperial College, London, S.W.7., Publication 72/74, 1972.

15. COLE J.D.
SAGE A.P.
'Multi-person Decision Analysis in Large Scale Hierarchical Systems - Team Decision Theory'.
International Journal of Control, Vol. 22, No.1, pp 1-28, 1975.

16. CRAMER H.
'Mathematical Methods of Statistics'.
Princeton University Press, 1946.

17. CUMMING S.D.G.
'Optimal Feedback Gains for Linear Discrete Time Systems with Noise on Inputs and Measurements'.
Computing and Control Dept., Imperial College, London, S.W.7., Report 69/8 , 1969.

18. CURRY R.E. 'A New Algorithm for Suboptimal Stochastic Control'.

IEEE Transactions, Vol. AC-14, pp 533-536, October, 1969.

19. DREYFUS S.E. 'Some Types of Optimal Control of Stochastic Systems'.

SIAM J. Control, Vol. 2, pp 120-134, 1964.

20. FELDBAUM A.A. 'Optimal Control Systems'.

Academic Press, 1965.

21. FLEMING W.H. 'Stochastic Control for Small Noise Intensities'.

Division of Applied Mathematics, Brown University April, 1970.

22. HADLEY G. 'Linear Programming'.

Addison-Wesley Publishing Co., Inc., 1960.

23. HO Y.C. 'Differential Games, Dynamic Optimisation, and Generalised Control Theory'.

Journal of Optimisation Theory and Applications, Vol. 6, No.3, September 1970.

24. HO Y.C.
CHU K.C. 'Team Decision Theory and Information Structures in Optimal Control Problems - Part I'.

IEEE Transactions on Automatic Control Vol. AC-17, No.1, February 1972.

25. HO Y.C.
CHU K.C. 'Information Structure in Dynamic Multi-Person Control Problems'.

Automatica Vol. 10, pp 341-351, 1974.

26. HO Y.C.
BRYSON A.E. 'Applied Optimal Control'.

Hallsted Press 1969, (Revised 1975).

27. HOWARD R. 'Dynamic Programming and Markov Processes'.

MIT Press, Cambridge, Massachusetts, 1960.

28. JACOBS O.L.R. 'Introduction to Dynamic Programming'
Oxford University Press.

29. JACOBS O.L.R.
PATCHELL J.W. 'Caution and Probing in Stochastic Control'
International Journal of Control, Vol.16, No. 1, pp 189-199, 1972.

30. JACOBS O.L.R. 'On the Functional Recurrence Relation of Optimal Discrete Time Stochastic Control'.
International Journal of Control, Vol.26, No.5, pp 763-768, 1977.

31. PATCHELL J.W.
JACOBS O.L.R. 'Separability, Neutrality, and Certainty Equivalence'.
International Journal of Control, Vol. 13, pp 337-342, 1971.

32. JAVDAN M.R. 'A Unified Theory of Optimal Multi-Level Control'.
International Journal of Control, Vol. 22, No.4, pp 517-524, 1975.

33. KALMAN R.E. 'A New Approach to Linear Filtering and Prediction Problems'.
Transactions of ASME, Journal of Basic Engineering, pp 35-45, March 1960.

34. KINGMAN J.F.C.
TAYLOR S.J. 'Introduction to Measure and Probability'.
Cambridge University Press, 1966.

35. KURTURAN B.Z.
SIVAN R. 'Linear Quadratic Gaussian Control with One Step Delay Sharing Pattern'.
IEEE Transactions on Automatic Control, AC-19, pp 571-574, October, 1974.

36. KURTURAN B.Z. 'A Concise Derivation of the LQG One-Step-Delay Sharing Problem Solution'.
IEEE Transactions on Automatic Control, Vol. AC-20, pp 808-810, December 1975.

37. KURTURAN B.Z. 'Decentralised Stochastic Control with Delayed Sharing Information Pattern'. IEEE Transactions on Automatic Control, Vol. AC-21, pp 576-581, August 1976.

38. KUSHNER H.J. 'Introduction to Stochastic Control'. Holt, Reinhart and Winston Inc., 1971.

39. KUSHNER H.J. 'Near Optimal Control in the Presence of Small Stochastic Perturbations'. Transactions of ASME, Journal of Basic Engineering, pp 103-108, March 1965.

40. LUENBERGER D.G. 'Optimisation by Vector Space Methods'. Wiley, 1969.

41. MEIER L. LARSON R. TETHER A.J. 'Dynamic Programming for Stochastic Control of Discrete Systems'. IEEE Transactions, AC-16, pp 767-775, 1971.

42. RADNER R. 'Team Decision Problems'. Annals of Mathematical Statistics, Vol.33, No.3, pp 857, September 1962.

43. RADNER R. 'The Evaluation of Information in Organisations'. Proceedings: 4th Berkeley Symposium, pp 491-533, University of California Press, 1961.

44. ROOT J.G. 'Optimum Control of Non-Gaussian Linear Stochastic Systems with Inaccessible State Variables'. SIAM J. CONTROL, Vol.7, pp 317-323, 1969.

45. SAATY L. BRAM J. 'Non-Linear Mathematics'. McGraw Hill, New York, 1964

46. SANDELL N.R. ATHANS M. 'Solution of Some Non-Classical LQG Stochastic Decision Problems'. IEEE Transactions on Automatic Control, Vol. AC-19, No.2, pp 108-116, April 1974.

47. SANDELL N.R.
ATHANS M.
'Comments on "A Solution for a Class of Stochastic Co-operative Games" '.
International Journal of Control, Vol. 20, pp 169-170, 1974.

48. SANDELL N.R.
VARAIYA P.
ATHANS M.
SAFONOV M.G.
'Survey of Decentralised Control Methods for Large Scale Systems'.
IEEE Transactions on Automatic Control, Vol. AC-23, No.2, April 1978.

49. STRIEBEL C.
'Sufficient Statistics in the Optimum Control of Stochastic System'.
J. Mathematical Anal. Appl. 12, pp 576-592, 1965.

50. STRIEBEL C.
'Optimal Control of Discrete Time Stochastic Systems'.
Lecture Notes in Economics and Mathematical Systems, Vol.110, Springer, New York, 1975.

51. SUNDARESHAN M.K.
'Generation of Multilevel Control and Estimation Schemes for Large Scale Systems - A Perturbational Approach'.
IEEE Transactions on Systems, Man., and Cybernetics, Vol.SMC-7, No.3, pp 144-152, 1977.

52. TODA M.
AOKI M.
'Second-Guessing Technique for Stochastic Linear Regulator Problems with Delayed Information Sharing'.
IEEE Transactions on Automatic Control, Vol. AC-20, pp 260-262, April 1975.

53. WERNERSSON A.
'Comments on "Optimal Stochastic Control for Discrete Time Linear Systems with Interrupted Observations" '.
Automatica, Vol.10, pp 113-115, 1974.

54. WHITTLE P.
RUDGE J.F.
'The Optimal Linear Solution of a Symmetric Team Control Problem'.
J. Appl. Prob. 11, pp 377-381, 1974.

55. WHITTLE P.
RUDGE J.F.
'Simultaneous Communication and Control'.
Adv. Appl. Prob. 8, pp 365-384, 1976.

56. WITSENHAUSEN H.S. 'A Counter-Example in Stochastic Optimum Control'.

SIAM J. Control, Vol. 6, No.1, pp 131-147, 1968.

57. WITSENHAUSEN H.S. 'Separation of Estimation and Control for Discrete Time Systems'.

Proceedings of IEEE, Vol. 59, No.11, pp 1557-1566, November 1971.

58. WITSENHAUSEN H.S. 'Information Structures, Feedback, and Causality'.

SIAM Journal of Control, 9, pp 149-159, 1971.

59. WONHAM W.M. 'Optimal Stochastic Control'.

Automatica, Vol. 5, pp 113-118, 1969.

60. YOSHIKAWA T. 'Dynamic Programming Approach to Decentralised Stochastic Control Problems'.

IEEE Transactions on Automatic Control, AC-20, pp 796-797, December 1975.

APPENDIX 1

List of Frequently Occuring Symbols

APPENDIX 1

List of Frequently Occuring Symbols

The symbols listed below are followed by a brief statement of their meaning. Every attempt has been made to ensure that symbols are defined as they are introduced, or even re-introduced. In other cases the interpretation of a symbol should be apparent from its context. When a chapter number is specified it implies that the corresponding symbol is used only in that Chapter.

$p_n(x)$	probability density function of random variable x at time n.
$p_n(x_n \mid I_n)$	conditional probability density function of x_n given information I_n.
$p_\infty(x)$	asymptotic value of $p_n(x)$ as n approaches infinity. (Chap 7)
u_n	control value applied at time n.
v_n	noise in observation equation at time n.
w_n	noise in process equation at time n.
x_n	state variable in process at time n.
y_n	observation of process at time n.
B_n	information function (structure) at time n - (Chap 3)
C_n	space of control values at time n.
E_p	expectation operator (with respect to probability density p).

F_n	process function at time n ($x_{n+1} = F_n(x_n, u_n, w_n)$.
G_n	observation function at time n ($y_n = G_n(x_n, v_n)$).
$H_0(\Gamma_0, p_0)$	infinite horizon, average cost per period under policy Γ_0. (Chap 7)
$H_0^{\circ}(p_0)$	optimal value of $H_0(\Gamma_0, p_0)$ (Chap 7)
I_n	information available at time n.
$\breve{I}_{n+1}$	extra information at time n+1 $I_{n+1} = \breve{I}_{n+1} \cup I_n$.
J	sample cost function
$J(\rho, p, \Gamma_0)$	infinite horizon, discounted cost under policy Γ_0. (Chap 7)
$J_0(\rho, p)$	optimal value of $J(\rho, p, \Gamma_0)$. (Chap 7)
$L_n(X_n, U_n)$	cost incurred at time n.
O_n	space of observations at time n.
P_n^x	probability distribution of random variable x at time n.
P_n^v	probability distribution of random variable v at time n.
P_n^w	probability distribution of random variable w at time n.
S_n	state space of process variable x at time n.
S_K	space of probability density functions ($p / \int x^2 p(x)\,dx < K$) (Chap 7)
T	contraction mapping (optimality equation - Chap 7).
T_γ	contraction mapping (stationarity equation - Chap 7)
U^{N-1}	the sequence $u_0, u_1, u_2 \cdots u_{n-1}$.
U_K	space of admissable control functions ($\gamma / p \in S_k \Rightarrow p' = \mathcal{L}_\gamma\ p \in S_k$). (Chap 7)
U	arbitrary space of admissable control functions (Chap 7)

V	expected value of J.
$V_n^{\circ}(p_n)$	optimal cost-to-go from time n.
$V_0(\Gamma_0, p_0)$	infinite horizon, discounted cost under policy Γ_0. (Chap 7)
$V_0^{\circ}(p_0)$	optimal value of $V_0(\ _0, p_0)$. (Chap 7)
$\bar{V}(\gamma^*, p)$	infinite horizon discounted cost under policy (γ^*, γ^*---) (Chap 7)
$\hat{V}_N(\Gamma_0^N, p_0)$	N-stage discounted cost under policy Γ_0^N. (Chap 7)
$\hat{V}_N^{\circ}(p_0)$	optimal value of $V_N(\ _0^N, p_0)$. (Chap 7)
$V_{\infty}^{\circ}(p_0)$	limit as $N \to \infty$ of $V_N^{\circ}(p_0)$. (Chap 7)
$V(p_{\infty}, \gamma)$	asymptotic average cost per period under $(\gamma, \gamma, \ldots)$. (Chap 7)
$\bar{V}$	optimal value of $V(p_{\infty}, \gamma)$, where $p_{\infty} = L_{\gamma} p_{\infty}$. (Chap 7)
X	used to denote a random variable (as opposed to its value). (Chap 3)
Y	used to denote a random variable (as opposed to its value). (Chap 3)
Y^n	denotes the sequence $(y_0, y_1, ----y_n)$.
γ_n	control function or law at time n.
$\delta(x)$	Dirac-delta function.
δ	scalar multiplier.
ϵ	scalar multiplier.
Φ	gaussian function $\Phi(x) = 1/\sqrt{2\pi} \exp(-x^2/2)$.
λ	optimal average cost per period. (Chap 7)

λ	Lagrange multiplier.
ρ	discount factor. (Chap 7)
σ^2	variance.
σ_w^2	variance of random variable w.
σ_v^2	variance of random variable v.
σ_0^2	variance of random variable x_0.
$\sigma_{n/n}^2$	variance of random variable x_n given information I_n.
Γ_0	the sequence $(\gamma_0, \gamma_1, \gamma_2 \text{---})$.
Γ_0^N	the sequence $(\gamma_0, \gamma_1, \gamma_2 \text{---} \gamma_{N-1})$.
$\textcircled{H}$	the space of admissable control functions.
$\mathcal{L}_\gamma$	denotes 'state' transition operator $p_{n+1}(x) = \mathcal{L}_\gamma p_n(x)$, or
$\mathcal{L}$	denotes state transition operator $p_{n+1} = \mathcal{L}(p_n, \gamma)$.

APPENDIX 2

Analysis of the Linear Quadratic Gaussian Problem with Nested Information

		Page
A.2.1	Introduction	A2.1
A.2.2	Two Stage Example - Nested Observation Case	A2.1
A.2.3	Two Stage Example - Classical Information Pattern Problem	A2.4
A.2.4	Derivation of Filtering Equations	A2.7
A.2.5	Sufficient Statistics	A2.12
A.2.6	Comments	A2.17

A.2.1 Introduction

In this appendix we employ Theorem 3.2 to derive the solution of the traditional LQG problem. We discover that Theorem 3.2 is not enough; we have to develop the Kalman filtering equations to introduce the notions of neutrality and a sufficient statistic.

A.2.2 Two Stage Example - Nested Observation Case

Before we make any general comments regarding the dynamic programming formulation of the nested information case, we shall consider the application of thetechnique to a simple example.

Consider the following simple scalar system, defined for $n = 0,1,2$

$$X_{n+1} = X_n + U_n + W_n, \quad p(W_n) = \frac{\Phi}{\sigma_w}\left\{\frac{W_n}{\sigma_w}\right\}; \quad p(X_0) = \frac{\Phi}{\sigma_0}\left\{\frac{X_n}{\sigma_0}\right\}$$

$$Y_n = X_n + V_n, \qquad p(V_n) = \frac{\Phi}{\sigma_v}\left\{\frac{V_n}{\sigma_v}\right\}$$

$$J = \sum_{n=0}^{1}\left(X_n^2 + U_n^2\right) + X_2^2, \text{ and } I_0 = (Y_0), \quad I_1 = (Y_0\ Y_1)$$

Proceeding as suggested by Theorem 3.2, we obtain

1) $$V_2^{\circ}(p(X_2|I_2)) = \int X_2^2 p(X_2|I_2)\,dX_2$$

2) $$V_1^{\circ}(p(X_1|I_1)) = \inf_{\gamma_1(I_1)} \left\{\int (X_1^2 + \gamma_1^2(I_1))p(X_1|I_1)\,dX_1 + \frac{\int X_2^2\, p(Y_2|X_2)\int p(X_2|X_1,I_1)p(X_1|I_1)\,dX_1\; p(Y_2|I_1)\,dY_2}{\int p(Y_2|I_1)\,dY_2}\right.$$

$$= \inf_{\gamma_1(I_1)} \left\{ \int X_1^2 + \gamma_1^2(I_1) p(X_1|I_1)\, dX_1 \right.$$

$$\left. + \int X_2^2 \int \frac{\Phi}{\sigma_w} \left\{ \frac{X_2 - (X_1 + \gamma_1(I_1))}{\sigma_w} \right\} p(X_1|I_1)\, dX_1 dX_2 \right\}$$

$$= \inf_{\gamma_1(I_1)} \left\{ \int (X_1^2 + \gamma_1^2(I_1) + \sigma_w^2 + X_1^2 + \gamma_1^2(I_1) + 2X_1\gamma_1(I_1)) p(X_1|I_1)\, dX_1 \right\}$$

$$= \inf_{\gamma_1(I_1)} \left\{ \int (2X_1^2 + 2\gamma_1^2(I_1) + 2X_1\gamma_1(I_1)) p(X_1|I_1)\, dX_1 + \sigma_w^2 \right\}$$

The minimising $\gamma_1(I_1)$ is $\gamma_1^{\circ}(I_1) = -\frac{1}{2}\int X_1 p(X_1|I_1)\, dX_1$, and

$$V_1^{\circ}(p(X_1|I_1)) = \int 2X_1^2\, p(X_1|I_1)\, dX_1 - \tfrac{1}{2}\left(\int X_1 p(X_1|I_1)\, dX_1\right)^2 + \sigma_w^2$$

3) $$V_0^{\circ}(p(X_0|I_0) = \inf_{\gamma_0(I_0)} \left\{ \int (X_0^2 + \gamma_0^2(I_0))\, p(X_0|I_0)\, dX_0 + \sigma_w^2 \right.$$

$$\left. + \frac{\int\int 2X_1^2 p(Y_1|X_1) \int p(X_1|X_0, I_0) p(X_0|I_0)\, dX_0 dX_1 \int p(Y_1|X_1) \int p(X_1|X_0, I_0) p(X_0|I_0)\, dX_0 dX_1 dY_1}{\int p(Y_1|X_1) \int p(X_1|X_0, I_0) p(X_0|I_0)\, dX_0 dX_1} \right.$$

$$\left. - \frac{1}{2}\int \left(\frac{\int X_1 p(Y_1|X_1) \int p(X_1|X_0, I_0) p(X_0|I_0)\, dX_0 dX_1}{\int p(Y_1|X_1) \int p(X_1|X_0, I_0) p(X_0|I_0)\, dX_0 dX_1} \right)^2 \int p(Y_1|X_1) \int p(X_1|X_0, I_0) p(X_0|I_0) \right.$$

$$\left. \cdot dX_0 dX_1 dY_1 \right\}$$

$$= \inf_{\gamma_0(I_0)} \left\{ \int (X_0^2 + \gamma_0^2(I_0)) p(X_0|I_0)\, dX_0 + \sigma_w^2 \right.$$

$$\left. + \int 2(X_0^2 + \gamma_0^2(I_0) + 2X_0\gamma_0(I_0)) p(X_0(I_0)\, dX_0 \right.$$

$$- \frac{1}{2}\int \left(\frac{\int X_1 \frac{\Phi}{\sigma_v}\left\{\frac{Y_1 - X_1}{\sigma_v}\right\} \int \frac{\Phi}{\sigma_w}\left\{\frac{X_1 - (X_0 + \gamma_0(I_0))}{\sigma_w}\right\} p(X_0|I_0)\, dX_0\, dX_1}{\int p(Y_1|X_1) \int p(X_1|X_0, I_0) p(X_0|I_0)\, dX_0\, dX_1} \right)^2 p(Y_1|I_0)\, dY_1$$

$$= \inf_{\gamma_0(I_0)} \left\{ \int (3X_0^2 + 3\gamma_0^2(I_0) + 4X_0\gamma_0(I_0)) p(X_0|I_0)\, dX_0 + \sigma_w^2 \right.$$

$$- \frac{1}{2}\int \left(\frac{\iint X_1 \frac{\Phi}{\sigma_m}\left\{\frac{X_1 - Y_1\sigma_w^2 + (X_0 + \gamma_0(I_0))\sigma_v^2}{\sigma_w^2 + \sigma_v^2}\right\} \frac{\Phi}{\sigma_n}\left\{\frac{Y_1 - (X_0 + \gamma_0(I_0))}{\sigma_n}\right\} p(X_0|I_0)\, dX_0\, dX_1}{\int \frac{\Phi}{\sigma_m}\left\{\frac{X_1 - Y_1\sigma_w^2 + (X_0 + \gamma_0(I_0))\sigma_v^2}{\sigma_w^2 + \sigma_v^2}\right\} \frac{\Phi}{\sigma_n}\left\{\frac{Y_1 - (X_0 + \gamma_0(I_0))}{\sigma_n}\right\} p(X_0|I_0)\, dX_0\, dX_1} \right)^2$$

$$\left. \int \frac{\Phi}{\sigma_m}\left\{\frac{X_1 - Y_1\sigma_w^2 + (X_0 + \gamma_0(I_0))\sigma_v^2}{\sigma_w^2 + \sigma_v^2}\right\} \frac{\Phi}{\sigma_n}\left\{\frac{Y_1 - (X_0 + \gamma_0(I_0))}{\sigma_n}\right\} p(X_0|I_0)\, dX_0\, dX_1\, dY_1 \right\}$$

where $\sigma_m^2 = \sigma_w^2\, \sigma_v^2 / (\sigma_w^2 + \sigma_v^2)$

and $\sigma_n^2 = \sigma_v^2 + \sigma_w^2$

Thus,

$$V_0^{\circ}(p(X_0|I_0)) = \inf_{\gamma_0(I_0) + \sigma_w^2} \left\{ \int (3X_0^2 + 3\gamma_0^2(I_0) + 4X_0\gamma_0(I_0)) p(X_0|I_0)\, dX_0 \right.$$

$$\left. - \frac{1}{2} \int \left(\frac{\int \left[\frac{Y_1\sigma_w^2 + (X_0 + \gamma_0^{\circ}(I_0))\sigma_v^2}{\sigma_w^2 + \sigma_v^2} \right] \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - (X_0 + \gamma_0(I_0))}{\sigma_n} \right\} p(X_0|I_0)\, dX_0}{\int \frac{\Phi}{\sigma_n} \frac{Y_1 - (X_0 + \gamma_0(I_0))}{\sigma_n} p(X_0|I_0)\, dX_0} \right)^2 \right.$$

$$\left. \cdot \int \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - (X_0 + \gamma_0(I_0))}{\sigma_n} \right\} p(X_0|I_0)\, dX_0\, dY_1 \right.$$

At this point we wish to determine the infimum of this expression with respect to $\gamma_0(I_0)$, obtaining the required functional $\mathcal{F}_0$, where

$$\gamma_0^{\circ}(I_0) = \mathcal{F}_0(p(X_0|I_0))$$

But finding $\mathcal{F}_0$ from the above expression looks like a distinctly unpromising analytical prospect. Let us next examine the classical information pattern problem.

A.2.3 Two Stage Example - Classical Information Pattern Problem

Here we consider as an example a problem identical to that considered above, with the exception that,

$$I_0 = (Y_0)$$

$$I_1 = (Y_0, Y_1, U_0)$$

The analysis will proceed exactly as above but now our state transition equation will be a slightly different expression. Thus:-

$$1)\quad V_2^{\circ}(p(X_2(I_2)) = \int X_2^2 p(X_2|I_2)dX_2$$

$$2)\quad V_1^{\circ}(p(X_1|I_1)) = \inf_{\gamma_1(I_1)} \left\{ \int (X_1^2 + \gamma_1^2(I_1))\, p(X_1|I_1)dX_1 + \frac{\int X_2^2 p(Y_2U_1|X_2,I_1) \int p(X_2|X_1,I_1)p(X_0|I_0)dX_0dX_1)p(Y_2,U_1|I_1)dY_2dU_1}{p(Y_2,U_1|I_1)} \right\}$$

In this case,

$$p(Y_2,U_1|X_2|I_1) = p(Y_2|X_2)\ p\ (U_1|I_1)$$

$$= p(Y_2|X_2)\ \delta\ (U_1 - \gamma_1(I_1))$$

Thus we obtain $\gamma_1(I_1) = -\frac{1}{2}\int X_1 p(X_1|I_1)dX_1$, and

$$V_1^{\circ}(p(X_1|I_1) = \int 2X_1^2 p(X_1|I_1)dX_1 - \frac{1}{2}\left(\int X_1 p(X_1|I_1)dX_1\right)^2 + \sigma_w^2$$

$$3)\quad V_0^{\circ}(p(X_0|I_0) = \inf_{\gamma_0^{\circ}(I_0)} \left\{ \int (3X_0^2 + 3\gamma_0^2(I_0) + 4X_0\gamma_0(I_0))p(X_0|I_0)dX_0 + \sigma_w^2 \right.$$

$$\left. - \frac{1}{2}\int \left(\frac{\int X_1 p(Y_1|X_1)p(U_0|I_0) \int p(X_1|X_0,I_0)p(X_0|I_0)dX_0dX_1}{\int p(Y_1|X_1)p(U_0|I_0)\int p(X_1|X_0,I_0)pX_0|I_0)dX_0dX_1} \right)^2 \right.$$

$$\left. \cdot \int p(Y_1|X_1)p(U_0|I_0)\int p(X_1|X_0,I_0)p(X_0|I_0)dX_0dX_1dY_1dU_0 \right\}$$

$$= \inf_{\gamma_0(I_0)} \left\{ \int (3X_0^2 + 3\gamma_0^2(I_0) + 4X_0\gamma_0(I_0)) p(X_0|I_0)\, dX_0 \right.$$

$$+ \sigma_w^2$$

$$- \frac{1}{2} \int \delta(U_0 - \gamma_0(I_0)) \left\{ \frac{\int \left[\frac{Y_1\sigma_w^2 + (X_0+\gamma_0(I_0))\sigma_v^2}{\sigma_w^2 + \sigma_v^2} \right] \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - (X_0+\gamma_0(I_0))}{\sigma_n} \right\} p(X_0|I_0)\, dX_0}{\int \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - (X_0+\gamma_0(I_0))}{\sigma_n} \right\} p(X_0|I_0)\, dX_0} \right\}^2$$

$$\left. \cdot \int \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - (X_0+\gamma_0(I_0))}{\sigma_n} \right\} p(X_0|I_0)\, dX_0\, dY_1\, dU_0 \right\}$$

$$= \inf_{\gamma_0(I_0)} \left\{ \int (3X_0^2 + 3\gamma_0^2(I_0) + 4X_0\gamma_0(I_0)) p(X_0|I_0)\, dX_0 \right.$$

$$+ \sigma_w^2$$

$$- \frac{1}{2} \int \left\{ \frac{\int \left(\frac{Y_1\sigma_w^2 + (X_0+\gamma_0(I_0))\sigma_v^2}{(\sigma_w^2 + \sigma_v^2)} \right) \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - X_0 + \gamma_0(I_0)}{\sigma_n} \right\} p(X_0|I_0)\, dX_0}{\int \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - (X_0+\gamma_0(I_0))}{\sigma_n} \right\} p(X_0|I_0)\, dX_0} \right\}^2$$

$$\left. \cdot \int \frac{\Phi}{\sigma_n} \left\{ \frac{Y_1 - (X_0+\gamma_0(I_0))}{\sigma_n} \right\} p(X_0|I_0)\, dX_0\, dY_1 \right\}$$

But this, rather obviously, is the same problem which confronted us when we examined the nested information problem previously. Again, the task of finding $\gamma_0(I_0)$ looks difficult.

All this is somewhat surprising; as was demonstrated in Chapter 2, a solution to the linear, quadratic, gaussian problem with classical information pattern can be obtained with little difficulty. The solution of the corresponding nested observation problem can also be obtained with only a little more difficulty. How then can we make our dynamic programming formulation 'work' for a linear, quadratic, gaussian (LQG) problem?

In our example we could, of course, substitute for $p(X_0|I_0)$, using the expression,

$$p(X_0|I_0) = \frac{p(Y_0|X_0)p(X_0)}{\int p(Y_0|X_0)p(X_0)\,dX_0}$$

$$= \frac{\frac{\Phi}{\sigma_v}\left\{\frac{Y_0 - X_0}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{X_0}{\sigma_0}\right\}}{\int \frac{\Phi}{\sigma_v}\left\{\frac{Y_0 - X_0}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{X_0}{\sigma_0}\right\}dX_0}$$

But if we did make this substitution before finding the infimum, we would be deviating from our plan of obtaining $\gamma_0(I_0)$ as a functional $\mathcal{J}_0$ of $p(X_0|I_0)$. And although this approach would be acceptable for a two stage process for any system of a longer duration it would not be possible. The suggestion, however, does give us a clue as to how we can make dynamic programming work for the LQG problem.

A.2.4 Derivation of Filtering Equations

Consider first the classical information pattern and let us examine the form of $p(X_{n+1} \mid I_{n+1})$ for the LQG problem. We have for the classical information pattern, the state transition equation:-

$$p(X_{n+1}|Y^{n+1},U^{n}) = \mathcal{L}_n(p(X_n|I_n), \gamma_n(I_n), \overset{\vee}{I}_{n+1})$$

Thus,

$$p(X_{n+1}|Y^{n+1},U^{n}) = \frac{p(Y_{n+1}|X_{n+1})p(U_n|Y^n,U^{n-1})\int p(X_{n+1}|X_n,Y^n,U^{n-1})p(X_n|Y^n,U^{n-1})dX_n}{\int p(Y_{n+1}|X_{n+1})p(U_n|Y^n,U_{n-1})\int p(X_{n+1}|X_n,Y^n,U^{n-1})p(X_n|Y_n,U^{n-1})dX_n dX_{n+1}}$$

$$= \frac{p(Y_{n+1}|X_{n+1})\delta(U_n-\gamma_n(Y^n,U^{n-1}))\int p(X_{n+1}|X_n,\gamma_n(Y^n,U^{n-1}))p(X_n|Y^n,U^{n-1})dX_n}{\int p(Y_{n+1}|X_{n+1})\delta(U_n-\gamma_n(Y^n,U^{n-1}))\int p(X_{n+1}|X_n,\gamma_n(Y^n,U^{n-1}))p(X_n|Y^n,U^{n-1})dX_n dX_{n+1}}$$

Let us start at n=0, with the density function $p(X_0|Y_0)$.

Then,

$$p(X_1|Y_1,U_0) = \frac{p(Y_1|X_1)\delta(U_0-\gamma_0(Y_0))\int p(X_1|X_0,\gamma_0(Y_0))p(X_0|Y_0)dX_0}{\int p(Y_1|X_1)\delta(U_0-\gamma_0(Y_0))\int p(X_1|X_0,\gamma_0(Y_0))p(X_0|Y_0)dX_0 dX_1}$$

Considering our simple LQG problem, we have,

$$p(X_0|Y_0) = \frac{p(Y_0|X_0)p(X_0)}{\int p(Y_0|X_0)p(X_0)dX_0} = \frac{\frac{\Phi}{\sigma_v}\left\{\frac{Y_0-X_0}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{X_0}{\sigma_0}\right\}}{\int \frac{\Phi}{\sigma_v}\left\{\frac{Y_0-X_0}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{X_0}{\sigma_0}\right\}dX_0}$$

$$= \frac{\frac{\Phi}{\sigma_{0/0}}\left\{\frac{X_0-Y_0\sigma_0^2/(\sigma_0^2+\sigma_v^2)}{\sigma_{0/0}}\right\}\frac{\Phi}{\sigma_Y}\left\{\frac{Y_0}{\sigma_Y}\right\}}{\int \frac{\Phi}{\sigma_{0/0}}\left\{\frac{X_0-Y_0\sigma_0^2/(\sigma_0^2+\sigma_v^2)}{\sigma_{0/0}}\right\}\frac{\Phi}{\sigma_Y}\left\{\frac{Y_0}{\sigma_Y}\right\}dX_0},$$

where $\sigma^2_{0/0} = \sigma_v^2\sigma_0^2/(\sigma_v^2+\sigma_0^2)$

and $\sigma_Y^2 = \sigma_v^2+\sigma_0^2$

Thus $p(X_0|Y_0) = \frac{\Phi}{\sigma_{0/0}}\left\{\frac{X_0-Y_0\sigma_0^2/(\sigma_0^2+\sigma_v^2)}{\sigma_{0/0}}\right\}$

$= \frac{\Phi}{\sigma_{0/0}}\left\{\frac{X_0-\bar{X}_{0/0}}{\sigma_{0/0}}\right\}$

is where $\bar{X}_{0/0} = Y_0\sigma_0^2/(\sigma_0^2+\sigma_v^2)$

Using the 'state' transition equation, we obtain:-

$$p(X_1|Y_1,Y_0,U_0) =$$

$$\frac{\frac{\Phi}{\sigma_v}\left\{\frac{Y_1-X_1}{\sigma_v}\right\}\int\frac{\Phi}{\sigma_m}\left\{\frac{X_0-(X_1-\gamma_0(Y_0))\sigma_{0/0}^2+\bar{X}_{0/0}\sigma_w^2}{\sigma_{0/0}^2+\sigma_w^2}\right\}\frac{\Phi}{\sigma_n}\left\{\frac{X_1-\gamma_0(Y_0)-\bar{X}_{0/0}}{\sigma_n}\right\}dX_0}{\int\frac{\Phi}{\sigma_v}\left\{\frac{Y_1-X_1}{\sigma_v}\right\}\int\frac{\Phi}{\sigma_m}\left\{\frac{X_0-(X_1-\gamma_0(Y_0))\sigma_{0/0}^2+\bar{X}_{0/0}\sigma_w^2}{\sigma_{0/0}^2+\sigma_w^2}\right\}\frac{\Phi}{\sigma_n}\left\{\frac{X_1-\gamma_0(Y_0)-\bar{X}_{0/0}}{\sigma_n}\right\}dX_0\,dX_1}$$

$$= \frac{\frac{\Phi}{\sigma_v}\left\{\frac{Y_1-X_1}{\sigma_v}\right\}\frac{\Phi}{\sigma_n}\left\{\frac{X_1-\gamma_0(Y_0)-\bar{X}_{0/0}}{\sigma_n}\right\}}{\int\frac{\Phi}{\sigma_v}\left\{\frac{Y_1-X_1}{\sigma_v}\right\}\frac{\Phi}{\sigma_n}\left\{\frac{X_1-\gamma_0(Y_0)-\bar{X}_{0/0}}{\sigma_n}\right\}dX_1}$$

where $\sigma_n^2 = \sigma_{0/0}^2 + \sigma_w^2$

Thus,

$$p(X_1 | Y_1, Y_0, U_0) = \frac{\Phi}{\sigma_{1/1}} \left\{ \frac{X_1 - Y_1 \frac{\sigma_n^2 + (\gamma_0(Y_0) + \bar{X}_{0/0})\sigma_v^2}{\sigma_n^2 + \sigma_v^2}}{\sigma_{1/1}} \right\} \frac{\Phi}{\sigma_n'} \left\{ \frac{Y_1 - (\gamma_0(Y_0) + \bar{X}_{0/0})}{\sigma_n'} \right\}$$

$$\int \frac{\Phi}{\sigma_{1/1}} \left\{ \frac{X_1 - Y_1 \frac{\sigma_n^2 + (\gamma_0(Y_0) + \bar{X}_{0/0})\sigma_v^2}{\sigma_n^2 + \sigma_v^2}}{\sigma_{1/1}} \right\} \frac{\Phi}{\sigma_n'} \left\{ \frac{Y_1 - \gamma_0(Y_0) + \bar{X}_{0/0}}{\sigma_n'} \right\} dX_1$$

$$= \frac{\Phi}{\sigma_{1/1}} \left\{ \frac{X_1 - \bar{X}_{1/1}}{\sigma_{1/1}} \right\}$$

where $\bar{X}_{1/1} = \dfrac{Y_1(\sigma_{0/0}^2 + \sigma_w^2) + (\gamma_0(Y_0) + \bar{X}_{0/0})\sigma_v^2}{\sigma_{0/0}^2 + \sigma_w^2 + \sigma_v^2}$

$$\sigma_{1/1}^2 = \frac{(\sigma_{0/0}^2 + \sigma_w^2)\sigma_v^2}{\sigma_{0/0}^2 + \sigma_w^2 + \sigma_v^2}$$

This procedure can clearly be extended to derive $p(X_n | Y^n, U^{n-1})$, obtaining the well known recursive relations of the Kalman Filter.[1]

In general we will have:-

$$p(X_n | Y^n, U^{n-1}) = \frac{\Phi}{\sigma_{n/n}} \left\{ \frac{X_n - \bar{X}_{n/n}}{\sigma_{n/n}} \right\}$$

1. Kalman R.E. A New Approach to Linear Filtering and Prediction Problems. 1960 (33)

where $\bar{X}_{n/n}$ and $\sigma^2_{n/n}$ are defined by the recursive relations:-

$$\bar{X}_{n/n} = \frac{Y_n(\sigma^2_{n/n} + \sigma^2_w) + (\bar{X}_{n-1/n-1} + \gamma_{n-1}(I_{n-1}))\sigma^2_v}{\sigma^2_{n/n} + \sigma^2_w + \sigma^2_v}$$

and $$\sigma^2_{n/n} = \frac{(\sigma^2_{n-1/n-1} + \sigma^2_w)\sigma^2_v}{\sigma^2_{n-1/n-1} + \sigma^2_w + \sigma^2_v}$$

with $$\sigma^2_{0/0} = \sigma^2_v \sigma^2_0 / (\sigma^2_v + \sigma^2_0)$$

and $$\bar{X}_{0/0} = Y_0 \sigma^2_0 / (\sigma^2_0 + \sigma^2_v)$$

Various crucial comments can be made regarding the density function $p(X_n | Y^n, U^{n-1})$ for our LQG problem:-

i) it is gaussian

ii) its mean value is a linear function of the observations $\{Y_1, Y_2, Y_3 \cdots Y_n\}$ and the control values $\{\gamma_0(I_0), \gamma_1(I_1), \cdots \gamma_{n-1}(I_{n-1})\}$

iii) its variance is independent of all dontrols and observations. It is merely a constant.

Notice that we have generated $p(X_n | Y^n, U^{n-1})$ for the classical information pattern problem using our 'state' transition equation:-

$$p(Y_n | Y^n, U^{n-1}) = \frac{p(Y_n, U_{n-1} X_n, I_{n-1}) \int p(X_n X_{n-1}, I_{n-1}) p(X_{n-1} | I_{n-1}) dX_{n-1}}{\int p(Y_n, U_{n-1} | X_n, I_{n-1}) \int p(X_n | X_{n-1}, I_{n-1}) p(X_{n-1} | I_{n-1}) dX_{n-1} dX_n}$$

Thus our expressions for $\bar{X}_{n/n}$ and $\sigma^2_{n/n}$ involve the control law value $\gamma_i(I_i) i = 0,1,---n-1$ and not the control data quantities - $u_i, i = 0,1,---n-1$. Also notice that precisely the same expressions will be obtained for the nested observation case.

It is the fact that the variance $\sigma^2_{n/n}$ is a predetermined constant which permits a colossal simplification of the dynamic programming formulation for the LQG problem.

A.2.5 Sufficient Statistics

What we have demonstrated above is that our infinite dimensional state vector $p(X_n | Y^n, U^{n-1})$ can be summarised by two statistics $\bar{X}_{n/n}$ and $\sigma^2_{n/n}$ for a linear system with gaussian noises. Moreover only $\bar{X}_{n/n}$ actually depends on the control data (Y^n, U^{n-1}). Striebel[1] has defined $\bar{X}_{n/n}$ as an equivalent statistic in the sense that knowledge of the value of $\bar{X}_{n/n}$ for a particular (Y^n, U^{n-1}) is sufficient to enable - $p(X_n | Y^n, U^{n-1})$ to be computed and conversely from a knowledge of the function $p(X_n | Y^n, U^{n-1})$ for a particular (Y^n, U^{n-1}) the corresponding value of $\bar{X}_{n/n}$ can be computed. In general of course $\bar{X}_{n/n}$ is a function of (Y^n, U^{n-1}) as is $p(X_n | Y^n, U^{n-1})$.

As far as our dynamic programming formulation is concerned we can evidently use $\bar{X}_{n/n}$ as our state variable. This leads us to the following particular version of Theorem 3.2.

Theorem 3.3

For the linear system with gaussian noises, considered above, suppose:-

$$V_{k+1}^{\circ}(\bar{X}_{k+1/k+1}(I_{k+1})) \triangleq \inf V_{k+1}(\bar{X}_{k+1/k+1}(I_{k+1}), \gamma_{k+1}(I_{k+1}), \Gamma_{k+2}^{N-1})$$

$$\gamma_k(I_{k+1}), \Gamma_{k+2}^{N-1}(I_{k+1})$$

Then V_k° satisfies the following iterative relationship,

$$V_k^{\circ}(\bar{X}_{k/k}(I_k)) = \inf_{\gamma_k(I_k)} \left\{ \int (X_k^2 + \gamma_k^2(I_k)) \frac{\Phi}{\sigma_{k/k}} \left\{ \frac{X_k - \bar{X}_{k/k}}{\sigma_{k/k}} \right\} dX_k \right.$$

$$\left. + \int V_{k+1}^{\circ}(\bar{X}_{k+1/k+1}(I_{k+1})) p(\underset{\vee}{I}_{k+1} | I_k) d\underset{\vee}{I}_{k+1} \right\}$$

where $\bar{X}_{k+1/k+1}(I_{k+1}) = \dfrac{Y_{k+1}(\sigma_{k/k}^2 + \sigma_w^2) + (\bar{X}_{k/k}(I_k) + \gamma_k(I_k))\sigma_v^2}{\sigma_{k/k}^2 + \sigma_w^2 + \sigma_v^2}$

This result is a somewhat trivial consequence of Theorem 3.2 for our particular example. Let us apply it to the task of finding the optimal control sequence for the nested observation case. For convenience the argument I_{k+1} has been dropped from $\bar{X}_{k+1/k+1}$.

1) $$V_2^{\circ}(\bar{X}_{2/2}) \triangleq \int X_2^2 \frac{\Phi}{\sigma_{2/2}} \left\{ \frac{X_2 - X_{2/2}}{\sigma_{2/2}} \right\} dX_2$$

$$= \bar{X}_{2/2}^2 + \sigma_{2/2}^2$$

2) $$V_1^{\circ}(\bar{X}_{1/1}) = \inf \left\{ \int X_1^2 + \gamma_1^2(I_1) \frac{\Phi}{\sigma_{1/1}} \left\{ \frac{X_1 - \bar{X}_{1/1}}{\sigma_{1/1}} \right\} dX_1 \right.$$

$$+ \sigma_{2/2}^2 + \int \left[\frac{Y_2(\sigma_{1/1}^2+\sigma_w^2)+(\bar{X}_{1/1}+\gamma_1(I_1))\sigma_v^2}{\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2} \right]^2 \quad p(Y_2|I_1)dY_2$$

where $p(Y_2|I_1) = p(Y_2|X_2) \int p(X_2|X_1,\gamma_1(I_1))p(X_1|I_1)dX_1dX_2$

$$= \frac{\Phi}{\sigma_v}\left\{\frac{Y_2-X_2}{\sigma_v}\right\} \int \frac{\Phi}{\sigma_w}\left\{\frac{X_2-(X_1+\gamma_1(I_1))}{\sigma_w}\right\} \frac{\Phi}{\sigma_{1/1}}\left\{\frac{X_1-\bar{X}_{1/1}}{\sigma_{1/1}}\right\} \quad dX_1dX_2$$

$$= \frac{\Phi}{\sqrt{\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2}}\left\{\frac{Y_2-(\bar{X}_{1/1}+\gamma_1(I_1))}{\sqrt{\sigma_v^2+\sigma_w^2+\sigma_{1/1}^2}}\right\}$$

Thus $V_1^{\circ}(\bar{X}_{1/1}) = \inf_{\gamma_1(I_1)} \left\{ \bar{X}_{1/1}^2 + \sigma_{1/1}^2 + \gamma_1^2(I_1) + \sigma_{2/2}^2 \right.$

$$\frac{+(\sigma_{1/1}^2+\sigma_w^2)^2\left[(\bar{X}_{1/1}+\gamma_1(I_1))^2-(\sigma_v^2+\sigma_w^2+\sigma_{1/1}^2)\right]}{(\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2)^2} + \frac{(\bar{X}_{1/1}+\gamma_1(I_1))^2\sigma_v^4}{(\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2)^2}$$

$$\left. + \frac{2(\sigma_{1/1}^2+\sigma_w^2)\sigma_v^2}{(\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2)^2} \quad (\bar{X}_{1/1}+\gamma_1(I_1))^2 \right\}$$

$$V_1^{\circ}(\bar{X}_{1/1}) = \inf_{\gamma_1(I_1)} \left\{ \bar{X}_{1/1}^2+\sigma_{1/1}^2+\gamma_1^2(I_1) + \sigma_{2/2}^2 \right.$$

$$\left. + (\bar{X}_{1/1}+\gamma_1(I_1))^2 - \frac{(\sigma_{1/1}^2+\sigma_w^2)^2}{(\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2)} \right\}$$

The minimising $\gamma_1(I_1)$ is $\gamma_1^{\circ}(I_1) = -\frac{1}{2}\bar{X}_{1/1}$, which is what we would expect from our earlier attempt. Thus:-

$$V_1^{\circ}(\bar{X}_{1/1}) = \left\{ \frac{3}{2}\ \bar{X}_{1/1}^2 + \sigma_{1/1}^2 + \sigma_{2/2}^2 \ - \frac{(\sigma_{1/1}^2 + \sigma_w^2)^2}{(\sigma_{1/1}^2 + \sigma_w^2 + \sigma_v^2)} \right\}$$

3) $$V_0^{\circ}(\bar{X}_{0/0}) = \inf_{\gamma_0(I_0)} \left\{ X_{0/0}^2 + o_{0/0}^2 + \gamma_0^2(I_0) + \sigma_{1/1}^2 + \sigma_{2/2}^2 \right.$$

$$\left. - \frac{(\sigma_{1/1}^2 + \sigma_w^2)^2}{(\sigma_{1/1}^2 + \sigma_w^2 + \sigma_v^2)} \right.$$

$$\left. + \frac{3}{2} \int \left\{ \frac{Y_1(\sigma_{0/0}^2 + \sigma_w^2) + (\bar{X}_{0/0} + \gamma_0(I_0))\sigma_v^2}{(\sigma_{0/0}^2 + \sigma_w^2 + \sigma_v^2)} \right\}^2 p(Y_1 | I_0)\ dY_1 \right\}$$

where $$p(Y_1 | I_0) = \Phi \left\{ \frac{Y_1 - (\bar{X}_{0/0} + \gamma_0(I_0))}{(\sigma_{0/0}^2 + \sigma_w^2 + \sigma_v^2)} \right\}$$

Thus,

$$V_0^{\circ}(\bar{X}_{0/0}) = \inf_{\gamma_0(I_0)} \left\{ \bar{X}_{0/0}^2 + \sigma_{0/0}^2 + \gamma_0^2(I_0) + \sigma_{1/1}^2 + \sigma_{2/2} - \frac{(\sigma_{1/1}^2 + \sigma_w^2)}{\sigma_{1/1}^2 + \sigma_w^2 + \sigma_v^2} \right.$$

$$\left. + \frac{3}{2} \left\{ (\bar{X}_{0/0} + \gamma_0(I_0))^2 - \frac{(\sigma_{0/0}^2 + \sigma_w^2)^2}{(\sigma_{0/0}^2 + \sigma_w^2 + \sigma_v^2)} \right\} \right\}$$

The minimising $\gamma_0(I_0)$ is $\gamma_0^{\circ}(I_0) = -\frac{3}{5}\bar{X}_{0/0}$, and

$$V_0^{\circ}(\bar{X}_{0/0}) = \frac{8}{5}\bar{X}_{0/0}^2 + \sigma_{2/2}^2 + \sigma_{1/1}^2 - \frac{(\sigma_{1/1}^2+\sigma_w^2)^2}{(\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2)} + \sigma_{0/0}^2 - \frac{(\sigma_{0/0}^2+\sigma_w^2)^2}{\sigma_{0/0}^2+\sigma_w^2+\sigma_v^2}$$

At this point, we have essentially solved our problem. All that remains to be done is to use our state transition equation to eliminate the dependence of $\gamma_i^{\circ}(I_i)$ on $\bar{X}_{i/i}$. Thus we obtain,

$$\gamma_0^{\circ}(I_0) = -\frac{3}{5}\bar{X}_{0/0} = -\frac{3}{5}\left\{Y_0\sigma_0^2 \,/\, \sigma_0^2+\sigma_v^2\right\}$$

$$\gamma_1^{\circ}(I_1) = -\frac{1}{2}X_{1/1} = -\frac{1}{2}\left\{\frac{Y_1(\sigma_{1/1}^2+\sigma_w^2)}{\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2} + \frac{\frac{2}{5}\bar{X}_{0/0}\,\sigma_v^2}{\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2}\right\}$$

$$= -\frac{1}{2}\left\{\frac{Y_1(\sigma_{1/1}^2+\sigma_w^2)}{\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2} + \frac{2}{5}\,\frac{Y_0\sigma_0^2\sigma_v^2/(\sigma_0^2+\sigma_v^2)}{(\sigma_{1/1}^2+\sigma_w^2+\sigma_v^2)}\right\}$$

The solution to the classical information pattern example can be derived in exactly the same fashion as the solution to the nested observation example derived above. In fact it is interesting to note that precisely the same solution will be obtained for $\gamma_i^{\circ}(I_i)$ in terms of $\bar{X}_{i/i}$. Moreover, because of the state transition equation for $\bar{X}_{i/i}$, upon eliminating $\bar{X}_{i/i}$, we shall find that the control value terms U^i do not appear in the final form of the control law. We can always assume of course, when eliminating $\bar{X}_{i+1/i+1}$ using the expression -

$$\bar{X}_{i+1/i+1} = \frac{Y_{i+1}(\sigma_{i/i}^2 + \sigma_w^2) + (\bar{X}_{i/i} + \gamma_i^{\circ}(I_i))\sigma_v^2}{\sigma_{i/i}^2 + \sigma_w^2 + \sigma_v^2}$$

that the term $\gamma_i^{\circ}(I_i) = U_i$, thus obtaining $\gamma_{i+1}^{\circ}(I_{i+1})$ in terms of Y^{i+1} and U^i. This in fact is the correct assumption to make, and $\gamma_{i+1}^{\circ}(I_{i+1})$ will be optimal even if $\gamma_k^{\circ}(I_k) \neq U_k$ for some values of $k \leq i$. More comments on this point are made at the end of Section 3.3.3 when we make some concluding remarks regarding the general classical information pattern problem.

A.2.6 Comments

i) Dynamic programming in the strictest sense of the word was not sufficient to solve our problem. To make it work we had to impose an additional constraint into the analysis - that the initial distribution is gaussian and the consequence that all 'states' will be gaussian with, moreover, a variance which is independent of the data and controls. We saw rather trivially then that the problem thus becomes equivalent to one in which the state is the mean value of our original state - the conditional distribution of X_n given the control data.

ii) The optimal controls are linear in the state and, since the state is linear in the control data, the optimal controls are linear in the control data also. Notice that this observation has nothing to do with the fact that $p(X_n | I_n)$ is Gaussian. The state will be Gaussian whatever the form of the control law, since that state

depends on only the control values $\gamma_i(I_i)$ and not the control laws γ_i.

iii) Our examples have been for the scalar system. The same techniques and arguments can be extended for n-dimensional systems with little increase in complexity. In addition the algebra that was encountered can be considerably streamlined in a number of ways by introducing various recursion formulae. These can be found elsewhere, they are not presented here because we are primarily interested in concepts.

iv) Precisely the same optimal control functions will be obtained if we consider the same problems with all primitive random variables set at their mean values. The arguments will now be X_k instead of $\bar{X}_{k/k}$. This is an example of the certainty equivalence.principle.

APPENDIX 3

Solution of Test Example of Chapters 4 and 5

		Page
A.3.1	Solution for Zero-Memory Case	A3.1
A.3.2	Solution for Classical Information Pattern Case	A3.18
A.3.3	Elimination of $p(x_1 \mid I_1)$ from γ_1^o	A3.34
A.3.4	Neo-Classical Solution of Discrete State Problem	A3.39
A.3.5	Neo-Classical Solution of Witsenhausen's Problem	A3.48

A.3.1 Solution for Zero-Memory Case

In this problem, the information I_n available to the controller at time n is merely the current observation y_n.

Adopting the notation of Theorem 1 in Chapter 3, we have,

$$V_1^0(p_1) = \inf_{\gamma_1} \int x_2^2 \, p(x_2) \, dx_2$$

where:-

$$p(x_2) = \int\int p(x_2 | x_1, y_1; \gamma_1) \, p(y_1 | x_1) \, p(x_1) \, dx_1 \, dx_2$$

Thus

$$V_1^0(p_1) = \inf_{\gamma_1} \int\int\int x_2^2 \, p(x_2 | x_1, y_1; \gamma_1) \, p(y_1 | x_1) \, p(x_1) \, dx_1 \, dy_1 \, dx_2$$

$$= \inf_{\gamma_1} \int\int x_2^2 \, p(x_2 | x_1, y_1; \gamma_1) \, p(x_1 | y_1) \, dx_1 \, p(y_1) \, dy_1 \, dx_2$$

Now $p(y_1) = \int p(y_1 | x_1) \, p(x_1) \, dx_1$, which is independent of γ_1.

Then,

$$V_1^{\circ}(p_1) = \inf_{\gamma_1} \int \left\{ \int\int x_2^2 p(x_2 | x_1, y_1; \gamma_1)\, p(x_1 | y_1)\, dx_2\, dx_1 \right\} p(y_1)\, dy_1$$

$$= \int \inf_{\gamma_1(y_1)} \left\{ \int\int x_2^2 p(x_2 | x_1, y_1; \gamma_1)\, p(x_1 | y_1)\, dx_2\, dx_1 \right\} p(y_1)\, dy_1$$

$$= \sum_r \left\{ \inf_{\gamma_1(y_1)} \sum_k \sum_j X_k^2\, p_{kjr}^{\gamma_1}\, p_{jr}\, p_r \right.$$

where $p_{kjr} = p(x_2 = X_k \mid x_1 = X_j;\ y_1 = Y_r;\ \gamma_1)$

$p_{jr} = p(x_1 = X_j \mid y_1 = Y_r)$

and $p_r = p(y_1 = Y_r)$

Now $\sum X_k^2\, p_{kjr} = E[x_2^2]$ for fixed $x_1 = X_j$ and $y_1 = Y_r$

But for fixed x_1 and y_1, x_2^2 is a function of w_1. Thus we have,

$$E\left[x_2^2 \,\middle|\, x_1 = X_j,\ y_1 = Y_r;\ \gamma_1\right]$$

$$= E\left[x_1^2 + \gamma_1^2(y_1) + w_1^2 + 2x_1 w_1 + 2\gamma_1(y_1)\, w_1 + 2x_1 \gamma_1(y_1) \,\middle|\, x_1 = X_j, y_1 = Y_r; \gamma_1\right]$$

$$= X_j^2 + \gamma_1^2(Y_r) + \sigma_w^2 + 2X_j \gamma_1(Y_r), \text{ where } \sigma_w^2 = 2/3$$

This leads to the following expression for $V_1^{\circ}(p_1)$:-

$$V_1^{\circ}(p_1) = \sum_r \inf_{\gamma_1(Y_r)} \sum_j \left(X_j^2 + \gamma_1^2(Y_r) + 2X_j \gamma_1(Y_r) + 2/3 \right) p_{jr}\, p_r$$

$$= \sum_r \inf_{\gamma_1(Y_r)} \left(\sum_j X_j^2\, p_{jr} + 2/3 + \gamma_1^2(Y_r) + 2\gamma_1(Y_r) \sum_j X_j\, p_{jr} \right) p_r$$

which gives us the required optimal $\gamma_1(Y_r) = -\text{sgn}\left\{\sum_j X_j\, p_{jr}\right\}$.

Thus

$$V_1^{\circ}(p_1) = \sum_r \Big(\sum_j X_j^2\, p_{jr} + 2/3 + 1 - 2\,\text{sgn}\Big(\sum_j X_j\, p_{jr} \Big) \cdot \sum_j X_j\, p_{jr} \Big) p_r$$

In other words,

$$V_1^{\circ}(p_1) - 5/3 = \sum_j X_j^2\, p_j - 2 \sum_r \Big(\text{sgn} \sum_j X_j\, p_{jr} \Big) \sum_j X_j\, p_{jr} \cdot p_r$$

To express V_1° as we wish as a functional of p_1, we make the following substitutions:-

i)

$$p_{jr} = p(x_1 = X_j \mid y_1 = Y_r)$$

$$= \frac{p(y_1 = Y_r \mid x_1 = X_j)\, p(x_1 = X_j)}{\sum_j p(y_1 = Y_r \mid x_1 = X_j)\, p(x_1 = X_j)}$$

$$= \frac{p_{rj}\, p_j}{\sum_j p_{rj} p_j}$$

ii)

$$p_r = p(y_1 = Y_r)$$

$$= \sum_j p(y_1 = Y_r \mid x_1 = X_j)\, p(x_1 = X_j)$$

$$= \sum_j p_{rj}\, p_j$$

Thus

$$V_1^{\circ}(p_1) - 5/3 = \sum_j X_j^2 p_j - 2 \sum_r \operatorname{sgn}\left(\sum_j X_j p_{rj} p_j\right) \sum_j X_j p_{rj} p_j$$

Let us write $V_1^{\circ}(p_1)$ solely in term of p_1. First, we have:-

$$\sum_j X_j p_{rj} p_j =$$

r	Y_r	
1	4	p_1
2	3	$p_1 + \frac{2}{3} p_2$
3	2	$p_1 + \frac{2}{3} p_2 + \frac{1}{3} p_3$
4	1	$\frac{2}{3} p_2 + \frac{1}{3} p_3$
5	0	$\frac{1}{3} p_3 - \frac{1}{3} p_5$
6	-1	$-\frac{1}{3} p_5 - \frac{2}{3} p_6$
7	-2	$-\frac{1}{3} p_5 - \frac{2}{3} p_6 - p_7$
8	-3	$-\frac{2}{3} p_6 - p_7$
9	-4	$- p_7$

From this it is clear that:-

$$\operatorname{sgn} \sum_j X_j p_{rj} p_j =$$

r	Y_r	
1	4	+
2	3	+
3	2	+
4	1	+
5	0	$\operatorname{sgn}(p_3 - p_5)$
6	-1	−
7	-2	−
8	-3	−
9	-4	−

Also $\sum_j X_j^2 p_j = 9p_1 + 4p_2 + p_3 + 4p_5 + 9p_7$

Thus

$$V_1^o(p_1) - 5/3 = \sum_j X_j^2 p_j - 2\Big\{ p_1 + (p_1 + 2/3\, p_2) + (p_1 + 2/3 p_2 + 1/3\, p_3)$$

$$+ 2/3\, p_2 + 1/3\, p_3 + \mathrm{sgn}(p_3 - p_5)(1/3\, p_3 - 1/3\, p_5)$$

$$+ (1/3\, p_5 + 2/3\, p_6) + (1/3 p_5 + 2/3\, p_6 + p_7)$$

$$+ (2/3\, p_6 + p_7) + p_7 \Big\}$$

that is :-

$$V_1^o(p_1) - 5/3 = 3p_1 - 1/3\, p_3 - \frac{2\,\mathrm{sgn}(p_3 - p_5)}{3} - 1/3\, p_5 + 3p_7$$

The next step in the analysis is to substitute for p_1 in this expression using the state transition equation:-

$$p(X_1) = \iint p(x_1 \mid x_0, y_0; \gamma_0)\; p(y_0 \mid x_0)\; p(x_0)\; dx_0\; dy_0$$

that is

$$p_k^{\gamma_0} = \sum_r \sum_j p_{kjr}^{\gamma_0}\, p_{rj}\, p_j \qquad \text{—— (i)}$$

or

$$p_k^{\gamma_0} = \sum_r \sum_j p_{kjr}\, p_{jr}\, p_r \qquad \text{—— (ii)}$$

Now $V_0^{\circ}(p_0) = \inf_{\gamma_0} V_1^{\circ}\mathcal{L}_1(p_0, \gamma_0))$,

where the notation $\mathcal{L}_1(p_0, \gamma_0)$ is used to denote the above substitutions for p_1. To find this infimum implies evaluating $V_1^{\circ}\mathcal{L}_1(p_0, \gamma_0))$ for $2^5 = 32$ different functions γ_0. This is a daunting task and we are left to wonder if there is any simpler manner in which the optimal γ_0° can be obtained. The first temptation is to seek a solution using the same device as was used to determine γ_1°; that is, we substitute for p_k^{γ} using equation (ii) and not (i). Let us examine this approach first:-

$$V_0^{\circ}(p_0) - 5/3 = \inf_{\gamma_0} \Big\{ 3 \sum_r \sum_j p_{1jr}^{\gamma} p_{jr} p_r - 1/3 \sum_r \sum_j p_{3jr}^{\gamma} p_{jr} p_r$$

$$- 2/3 \operatorname{sgn}\Big\{ \sum_r \sum_j (p_{3jr}^{\gamma} - p_{5jr}^{\gamma}) p_{jr} p_r \Big\} \sum_r \sum_j (p_{3jr}^{\gamma} - p_{5jr}^{\gamma}) \cdot$$

$$\cdot p_{jr} p_r - 1/3 \sum_r \sum_j p_{5jr}^{\gamma} p_{jr} p_r$$

$$+ 3 \sum_r \sum_j p_{7jr}^{\gamma} p_{jr} p_r \Big\}$$

$$= \sum_r \inf_{\gamma_0} \Big\{ 3 \sum p^{\gamma}_{1jr} p_{jr} - \tfrac{1}{3} \sum p^{\gamma}_{3jr} p_{jr}$$

$$- \tfrac{2}{3} \operatorname{sgn} \Big[\sum_r \sum_j (p^{\gamma}_{3jr} - p^{\gamma}_{5jr}) p_{jr} p_r \Big] \sum (p^{\gamma}_{3jr} - p^{\gamma}_{5jr}) p_{jr}$$

$$- \tfrac{1}{3} \sum_j p^{\gamma}_{5jr} p_{jr} + 3 \sum p^{\gamma}_{7jr} p_{jr} \Big\} p_r$$

We would like to replace the infimum with respect to γ_0 with an infimum with respect to $\gamma_0(Y_r)$. Unfortunately, this cannot be done, since to evaluate the expression within $[\quad]$ still requires knowledge of $\gamma_0(Y_r)$ for every Y_r. Thus we must abandon this approach.

An alternative approach which is equivalent is the evaluating of all possible 2^5 functions γ_0 is to introduce function values u_m, where $u_m = \pm 1$. By introducing these function values explicitly into our expression for $V_1^{\circ}\mathcal{L}_1(p_0, \gamma_0)$ it is hoped that we might be able to reduce the number of candidates functions γ_0° for which $V_1^{\circ}\mathcal{L}_1(p_0, \gamma_0)$ must be evaluated. Note that there is still a set of 2^5 values of u_r; r= 1,2, ---5. This approach is basically an attempt to exploit the fact that the cardinality of the range of γ_0 is less than its domain.

Consider the following probabilities:-

$$p\left\{ x_1 = X_k \,\middle|\, x_0 = X_j, y_0 = Y_r \; \gamma_0 \right\} = p^{\gamma}_{kjr}$$

For each γ, $\left\{ p^{\gamma}_{kjr} \right\}$ is an element of a (7 x 3 x 5) array, and in theory we would have to evaluate $\left\{ p^{\gamma}_{kjr} \right\}$ for all 2^5 possible γ_0. Alternatively,

we can introduce a random variable u_0 which takes values U_m and which represents the control applied; then we have,

$$p(x_1 = X_k \mid x_o = X_j , y_o = Y_r ; \gamma_o)$$

$$= \sum_m p(x_1 = X_k \mid x_o = X_j , y_o = Y_r , u_o = U_m ; \gamma_o) p(u_o = U_m \mid y_o = Y_r ; \gamma_o)$$

$$= \sum_m p_{krm} p_{mr}^{\gamma_o}$$

The elements p_{krm} are independent of the control function γ, which now merely influences $p(u_0 = U_m \mid y_0 = Y_r; \gamma)$. For non-randomised control laws, $p(u_0 = U_m \mid y_0 = Y_r; \gamma_0) = \delta_{mr}^{\gamma}$ where $\delta_{mr}^{\gamma} = 0,1$ depending on γ. (Here U_1 is defined as 1 and U_2 is defined as -1). Each δ_{mr} is an element of a (2 x 5) array. Thus we can write state transition equation (i) as follows:-

$$p_k^{\gamma} = \sum_r \sum_j \sum_m p_{kjm} \delta_{mr}^{\gamma} p_{rj} p_j$$

Substituting this into our expression for $V_1^{\circ}(p_1)$ gives:-

$$V_1^{\circ}(\mathcal{L}_1(p_0,\gamma_0)) - 5/3 = 3\sum_r \sum_j \sum_m p_{1jm}\delta_{mr}^{\gamma} p_{rj} p_j$$

$$-1/3 \sum_r \sum_j \sum_m p_{3jm}\delta_{mr}^{\gamma} p_{rj} p_j$$

$$-1/3 \sum_r \sum_j \sum_m p_{5jm}\delta_{mr}^{\gamma} p_{rj} p_j$$

$$+3 \sum_r \sum_j \sum_m p_{7jm}\delta_{mr}^{\gamma} p_{rj} p_j$$

$$-2/3\, sgn\left\{\sum_r \sum_j \sum_m (p_{3jm}-p_{5jm})\delta_{mr}^{\gamma} p_{rj} p_j\right\} \sum_r \sum_j \sum_m (p_{3jm}-p_{5jm})\delta_{mr}^{\gamma} p_{rj} p_j$$

$$= \sum_r \sum_j \sum_m \left(3p_{1jm} - \frac{p_{3jm}}{3} - \frac{p_{5jm}}{3} + 3p_{7jm}\right)\delta_{mr}^{\gamma} p_{rj} p_j$$

$$-2/3\, sgn\left\{\sum_r \sum_j \sum_m (p_{3jm}-p_{5jm})\delta_{mr}^{\gamma} p_{rj} p_j\right\} \sum_r \sum_j \sum_m (p_{3jm}-p_{5jm})\delta_{mr}^{\gamma} p_{rj} p_j$$

We must find the elements δ_{mr}^{γ} for m = 1,2; r = 1,2,---5 where $\delta_{mr} = 0$, 1, $\sum_m \delta_{mr}^{\gamma} = 1$, such that this expression is minimised. Since once δ_{1r} is fixed, we know $\delta_{2r} = 1 - \delta_{1r}$ the problem is equivalent to finding the optimal sequence δ_{1r} r = 1,2,---5 where $\delta_{1r} = 0,1$. There are 2^5 such sequences so at the moment we have gained nothing over our original formula-

tion. The hope remains, however, that we might be able to discard some sequences $\{\delta_{1r}\}$ by inspection and perhaps even deduce the optimal sequence. For each r, we shall denote the elements of δ^{γ}_{mr} as δ_r and $1-\delta_r$ (i.e. $\delta^{\gamma}_{1r} = \delta_r$ and $\delta^{\gamma}_{2r} = 1 - \delta_r$)

Let us now turn to the task of finding $V_1^{\circ}(\mathcal{L}_1(p_0, \gamma_0))$.

(i) $\sum_r \delta^{\gamma}_{mr} p_{rj} = F^{\gamma}_{mj}$; where

$$F^{\gamma}_{mj} = \begin{bmatrix} \frac{1}{3}(\delta_1+\delta_2+\delta_3) & \frac{1}{3}(\delta_2+\delta_3+\delta_4) & \frac{1}{3}(\delta_3+\delta_4+\delta_5) \\ 1-\frac{1}{3}(\delta_1+\delta_2+\delta_3) & 1-\frac{1}{3}(\delta_2+\delta_3+\delta_4) & 1-\frac{1}{3}(\delta_1+\delta_4+\delta_5) \end{bmatrix}$$

(ii) $\sum_m p_{kjm} F^{\gamma}_{mj} \triangleq G^{\gamma}_{kj}$; where $G^{\gamma}_{kj} =$

		j = 1	2	3
		X_j = 1	0	-1
1	3	$\frac{1}{9}(\delta_1+\delta_2+\delta_3)$	0	0
2	2	$\frac{1}{9}(\delta_1+\delta_2+\delta_3)$	$\frac{1}{9}(\delta_2+\delta_3+\delta_4)$	0
3	1	$\frac{1}{3}$	$\frac{1}{9}(\delta_2+\delta_3+\delta_4)$	$\frac{1}{9}(\delta_3+\delta_4+\delta_5)$
4	0	$\frac{1}{3}-\frac{1}{9}(\delta_1+\delta_2+\delta_3)$	$\frac{1}{3}$	$\frac{1}{9}(\delta_3+\delta_4+\delta_5)$
5	-1	$\frac{1}{3}-\frac{1}{9}(\delta_1+\delta_2+\delta_3)$	$\frac{1}{3}-\frac{1}{9}(\delta_2+\delta_3+\delta_4)$	$\frac{1}{3}$
6	-2	0	$\frac{1}{3}-\frac{1}{9}(\delta_2+\delta_3+\delta_4)$	$\frac{1}{3}-\frac{1}{9}(\delta_3+\delta_4+\delta_5)$
7	-3	0	0	$\frac{1}{3}-\frac{1}{9}(\delta_3+\delta_4+\delta_5)$
k	X_k			

(iii) $\sum_j G_{kj} p_j = p_k^{\gamma}$, where

p_k^{γ} =

k	X_k	
1	3	$1/9\, p_1(\delta_1+\delta_2+\delta_3)$
2	2	$1/9\, p_1(\delta_1+\delta_2+\delta_3)+1/9\, p_2(\delta_2+\delta_3+\delta_4)$
3	1	$1/3\, p_1+1/9\, p_2(\delta_2+\delta_3+\delta_4)+1/9\, p_3(\delta_3+\delta_4+\delta_5)$
4	0	$1/3\, p_1-1/9\, p_1(\delta_1+\delta_2+\delta_3)+1/3\, p_2+1/9\, p_3(\delta_3+\delta_4+\delta_5)$
5	-1	$1/3\, p_1-1/9\, p_1(\delta_1+\delta_2+\delta_3)+1/3\, p_2-1/9\, p_2(\delta_2+\delta_3+\delta_5)+1/3\, p_3$
6	-2	$1/3\, p_2-1/9\, p_2(\delta_2+\delta_3+\delta_4)+1/3\, p_3-1/9\, p_3(\delta_3+\delta_4+\delta_5)$
7	-3	$1/3\, p_3-1/9(\delta_3+\delta_4+\delta_5)$

Thus we can express $V_1^{\circ}(\mathcal{L}_0(p_0,\gamma_0))$ as follows:-

$$V_1^{\circ}(\mathcal{L}_0(p_0,\gamma_0))-5/3 = \Big\{1/3\, p_1(\delta_1+\delta_2+\delta_3)-1/9\, p_1-1/27\, p_2(\delta_2+\delta_3+\delta_4)$$

$$-1/27\, p_3(\delta_3+\delta_4+\delta_5)-1/9\, p_1+1/27\, p_1(\delta_1+\delta_2+\delta_3)-1/9\, p_2$$

$$+1/27\, p_2(\delta_2+\delta_3+\delta_4)-1/9\, p_3+p_3-1/3\, p_3(\delta_3+\delta_4+\delta_5)$$

$$-2/3\, \mathrm{sgn}(T)T, \text{ where}$$

$$T = \tfrac{1}{3}p_1 + \tfrac{1}{9}p_2(\delta_2+\delta_3+\delta_4) + \tfrac{1}{9}p_3(\delta_3+\delta_4+\delta_5)$$

$$-\tfrac{1}{3}p_1 + \tfrac{1}{9}p_1(\delta_1+\delta_2+\delta_3) - \tfrac{1}{3}p_2 + \tfrac{1}{9}p_2(\delta_2+\delta_3+\delta_4) - \tfrac{1}{3}p_3$$

Consider the influence which the value of δ_1 has upon the value of this expression. The positive contribution to V_1° made by terms involving δ_1 is,

$$\left\{\tfrac{1}{3}p_1 + \tfrac{1}{27}p_1 \pm \tfrac{2}{3}p_1\right\}\delta_1$$

It is clear that $\delta_1 = 0$ is always the optimal selection for δ_1; that is

$$\gamma_0(Y_1) = U_2 = -1$$

By a similar argument we can deduce the value of δ_5. The positive contribution to V_1° made by terms involving δ_5 is,

$$\left\{-1/27 p_3 - 1/3 p_3 \pm 1/9 p_3\right\}\delta_5$$

Thus in this case $\delta_5 = 1$ is always the optimal selection δ_5^0; that is

$$\gamma_{0\,5}(Y) = U_1 = +1$$

At this point, all that remains to do is determine $\gamma_0(Y_r)$ for r = 2,3,4. It does not appear that this can be done by inspection and in this case all that can be done is evaluate V_1° for all possible functions defined by sequences $\{\delta\} = (0\ \delta_2\ \delta_3\ \delta_4\ 1)$. Since each of $\delta_2^0, \delta_3^0, \delta_4^0$ can take only one of two values 0 or 1, we have 2^3 possibilities to consider. This is a considerable improvement over our initial problem involving 2^5 possibilities. What is more, evaluation of these possibiliies is made very easy by the above expression involving the unknown δ. Let us evaluate V_1° for all possible remaining choices for γ.

Simplifying the expression for T and V gives:-

$$T = 1/9 p_1(\delta_1+\delta_2+\delta_3) - 1/9 p_2(3-2(\delta_2+\delta_3+\delta_4)) - 1/9(3-2(\delta_3+\delta_4+\delta_5))p_3$$

$$V_1^{\circ}(\mathcal{L}(p_0,\gamma_0)) = 1/27 p_1(-6+10(\delta_1+\delta_2+\delta_3)) - 1/9 p_2 + 1/27 p_3(24-10(\delta_3+\delta_4+\delta_5))$$
$$-2/3\,\mathrm{sgn}(T)\,T$$

Thus, we can construct the following table.

	$\delta_1+\delta_2+\delta_3$	$\delta_2+\delta_3+\delta_4$	$\delta_3+\delta_4+\delta_5$	T	sgnT	$V_1^0 - 5/3 + 2/3\ \mathrm{sgn}(T)\,T$
$\chi_0(00001)$	0	0	1	$-1/3\,p_2 - 2/9\,p_3$	−	$-6/27\,p_1 - 1/9\,p_2 + 14/27\,p_3$
$\chi_0(01001)$	1	1	1	$1/9\,p_1 - 1/9\,p_2 - 2/9\,p_3$	sgnT	$4/27\,p_1 - 1/9\,p_2 + 14/27\,p_3$
$\chi_0(00101)$	1	1	2	$1/9\,p_1 - 1/9\,p_2 - 1/9\,p_3$	sgnT	$4/27\,p_1 - 1/9\,p_2 + 4/27\,p_3$
$\chi_0(01101)$	2	2	2	$2/9\,p_1 + 1/9\,p_2 - 1/9\,p_3$	sgnT	$14/27\,p_1 - 1/9\,p_2 + 4/27\,p_3$
$\chi_0(00011)$	0	1	2	$-1/9\,p_2 - 1/9\,p_3$	−	$-6/27\,p_1 - 1/9\,p_2 + 4/27\,p_3$
$\chi_0(01011)$	1	2	2	$1/9\,p_1 + 1/9\,p_2 - 1/3\,p_3$	sgnT	$4/27\,p_1 - 1/9\,p_2 + 4/27\,p_3$
$\chi_0(00111)$	1	2	3	$1/9\,p_1 + 1/9\,p_2$	+	$4/27\,p_1 - 1/9\,p_2 - 6/27\,p_3$
$\chi_0(01111)$	2	3	3	$1/9\,p_1 + 2/9\,p_2$	+	$14/27\,p_1 - 1/9\,p_2 - 6/27\,p_3$

Given any values for p_1, p_2, p_3 we can compute T and J for each control function γ_0 and then select that control function which gives the minimum.

In our example, we require $p_1 = p_2 = p_3 = \frac{1}{3}$. Then we have,

Control	T	sgn T	$V_1^0(\mathcal{L}(p_0, \gamma_0))$
$\gamma_0(00001)$	$-5/27$	$-$	$130/81$
$\gamma_0(01001)$	$-2/27$	$-$	$146/81$
$\gamma_0(00101)$	$-1/27$	$-$	$138/81$
$\gamma_0(01101)$	$2/27$	$+$	$146/81$
$\gamma_0(00011)$	$-2/27$	$-$	$126/81$ *
$\gamma_0(01011)$	$1/27$	$+$	$138/81$
$\gamma_0(00111)$	$2/27$	$+$	$126/81$ *
$\gamma_0(01111)$	$5/27$	$+$	$130/81$

The optimal costs have been marked with an asterisk. Note that the optimal control is not unique. It does not appear to matter what we do when $y_1 = 0$ is observed. At this point we have essentially solved the zero-memory problem. All that we need to do is eliminate the dependence of γ_1° on $p(x_1)$ and the problem will be completely solved.

A.3.2 Solution for Classical Information Pattern Case

For this version of the problem, the information I_n available to the controller at each time instant n consists of the current observation y_n and all previous observations and controls $(y_i, u_i \mid i = 0, \cdots n-1)$.

Adopting the notation developed in Chapter 3, we have

$$V_2^\circ(p(x_2 \mid I_2)) \triangleq \int x_2^2 p(x_2 \mid I_2)\, dx_2$$

and

$$V_1^\circ(p(x_1 \mid I_1)) = \inf_{\gamma_1(I_1)} \left\{ \int V_2^\circ\, p(x_2 \mid I_2)\, p(\check{I}_2 \mid I_1)\, d\check{I}_2 \right\}$$

where

$$p(x_2 \mid I_2) = \frac{p(\check{I}_2 \mid x_2, I_1) \int p(x_2 \mid x_1, \gamma_1(I_1)) p(x_1 \mid I_1)\, dx_1}{\int p(\check{I}_2 \mid x_2, I_1) \int p(x_2 \mid x_1, \gamma_1(I_1)) p(x_1 \mid I_1)\, dx_1\, dx_2}$$

and

$$p(\check{I}_2 \mid I_1) = \int p(\check{I}_2 \mid x_2) \int p(x_2 \mid x_1, \gamma_1(I_1))\, p(x_1 \mid I_1)\, dx_1\, dx_2$$

Thus

$$V_1^{\circ}(p(x_1|I_1))$$

$$= \inf_{\gamma_1(I_1)} \left\{ \frac{\int\int x_2^2\, p(\overset{v}{I}_2|x_2) \int p(x_2|x_1,\gamma_1(I_1)) p(x_1|I_1)\, dx_1\, dx_2}{\int p(\overset{v}{I}_2|x_2) \int p(x_2|x_1,\gamma_1(I_1))\, p(x_1|I_1)\, dx_1\, dx_2} \right.$$

$$\left. \cdot \int p(\overset{v}{I}_2|x_2) \int p(x_2|x_1,\gamma_1(I_1)) p(x_1|I_1)\, dx_1\, dx_2\, d\overset{v}{I}_2 \right\}$$

$$= \inf_{\gamma_1(I_1)} \left\{ \int\int \int x_2^2\, p(\overset{v}{I}_2|x_2)\, p(x_2|x_1,\gamma_1(I_1)\, p(x_1|I_1)\, dx_1\, dx_2\, d\overset{v}{I}_2 \right\}$$

$$= \inf_{\gamma_1(I_1)} \left\{ \int\int x_2^2\, p(x_2|x_1,\gamma_1(I_1)) p(x_1|I_1) dx_1 dx_2 \right\}$$

$$= \inf_{\gamma_1(I_1)} \left\{ \sum_j \sum_k X_k\, p^{\gamma}_{kjr}\, p_{jr} \right\},$$

where $p^{\gamma}_{kjr} = p(x_2 = X_k \mid x_1 = X_j, I_1 = \bar{I}_r; \gamma_1)$,

and $p_{jr} = p(x_2 = X_j \mid I_1 = \bar{I}_r)$

Now $\sum_k X_k^2 \, p^{\gamma}_{kjr} = E\left[x_2^2\right]$ for fixed $x_1 = X_j$ and $I_1 = \bar{I}_r$.

But for fixed x_1 and I_1, x_2^2 is a function of w_1. Thus we have,

$$\sum_k X_k^2 p^{\gamma}_{kjr} = E\left[x_2^2 \,\middle|\, x_1 = X_j, I_1 = \bar{I}_r \,;\, \gamma\right]$$

$$= E\left[x_1^2 + \gamma_1^2(I_1) + w_1^2 + 2x_1 w_1 + 2\gamma_1(I_1) w_1 + 2x_1 \gamma_1(I_1) \,\middle|\, x_1 = X_j, I_1 = \bar{I}_r \,;\, \gamma_1\right]$$

$$= X_j^2 + \gamma_1^2(I_1) + \sigma_w^2 + 2X_j \gamma_1(\bar{I}_r), \text{ where } \sigma_w^2 = 2/3$$

Thus:-

$$V_1^{\circ}(p(x_1|I_1)) = \inf_{\gamma_1(\bar{I}_r)} \left\{ \sum_j \left(X_j^2 + \gamma_1^2(\bar{I}_r) + \sigma_w^2 + 2X_j \gamma_1(\bar{I}_r) \right) p_{jr} \right\}$$

$$= \inf_{\gamma_1(\bar{I}_r)} \left\{ \sum_j X_j^2 p_{jr} + \gamma_1^2(\bar{I}_r) + \sigma_w^2 + 2\gamma_1(\bar{I}_r) \sum_j X_j p_{jr} \right\}$$

This is minimised when $\gamma_1(\bar{I}_r) = -\operatorname{sgn}(\sum_j X_j p_{jr})$. Notice the similarity between this control and that found for the zero memory problem.

Thus we have,

$$V_1^\circ(p(x_1|I_1)) = \sum_j X_j^2 p_{jr} + 1 + 2/3 - 2\operatorname{sgn}(\Sigma X_j p_{jr}) \Sigma X_j p_{jr}$$

The next step is to compute:-

$$V_0^\circ(p(x_0|I_0)) = \inf_{\gamma_0(I_0)} \left\{ \int V_1^\circ(p(x_1|I_1)) p(\check{I}_1|I_0)\, d\check{I}_1 \right\}$$

where

$$p(x_1|I_1) = \frac{p(\check{I}_1|x_1)\int p(x_1|x_0,\gamma_0(I_0))p(x_0|I_0)\,dx_0}{\int p(\check{I}_1|x_1)\int p(x_1|x_0,\gamma_0(I_0))p(x_0|I_0)\,dx_0\,dx_1}$$

and

$$p(\check{I}_1|I_0) = \int p(\check{I}_1|x_1,I_0)\int p(x_1|x_0,\gamma_0(I_0))p(x_0|I_0)\,dx_0\,dx_1$$

Explicitly, we have

$$p(x_1|I_1) = \{p(x_1 = X_j | I_1 = \bar{I}_r)\}$$

Now $p_{jr} \triangleq p(x_1 = X_j \mid I_1 = \bar{I}_r)$

$$= p(x_1 = X_j \mid y_1 = Y_n, u_0 = U_m, y_0 = Y_s)$$

$$= \frac{p(y_1 = Y_n, u_0 = U_m \mid x_1 = X_j, y_0 = Y_s) \sum_p p(x_1 = X_j \mid x_0 = X_p, y_0 = Y_s; \gamma)}{\sum_j p(y_1 = Y_n, u_0 = U_m \mid x_1 = X_j, y_0 = Y_s) \sum_p p(x_1 = X_j \mid x_0 = X_p, y_0 = Y_s; \gamma)} \cdots$$

$$\cdots \frac{p(x_0 = X_p \mid y_0 = Y_s)}{p(x_0 = X_p \mid y_0 = Y_s)}$$

$$\triangleq \frac{p_{nmjs} \sum_p p^{\gamma}_{jps} p_{ps}}{\sum_j p_{nmjs} \sum_p p_{jps} p_{ps}}$$

and

$$p(\check{I}_1 \mid I_0) = p(y_1 = Y_n, u_0 = U_m \mid y_0 = Y_s)$$

$$\triangleq p_{nms}$$

$$= \sum_j p_{nmjs} \sum_p p^{\gamma}_{jps} p_{ps}$$

Thus

$$V_0^0(p(x_0|I_0)) = \inf_{\gamma_0(Y_s)} \left(\sum_n \sum_m \left\{ \frac{\sum_j X_j^2 p_{nmjs} \sum_p p_{jps}^{\gamma} p_{ps}}{\sum_p p_{nmjs} \sum p_{jps}^{\gamma} p_{ps}} + 5/3 \right. \right.$$

$$\left. \left. - 2\,\mathrm{sgn}(T)\, T \right\} \sum_j p_{nmjs} \sum_p p_{jps}^{\gamma} p_{ps} \right)$$

where

$$T = \frac{\sum_j x_j p_{nj} \sum_p p_{jps}^{\gamma} p_{ps}}{\sum_j p_{nmjs} \sum_p p_{jps}^{\gamma} p_{ps}}$$

The first term in this expression can be greatly simplified, leading to the following expression for $V_0^0(p(x_0 \mid I_0))$:-

$$V_0^0(p(x_0|I_0)) = \inf_{\gamma_0(Y_s)} \left\{ \sum_j X_j^2 \sum_p p_{jps} p_{ps} + 5/3 \right.$$

$$\left. - 2 \sum_n \sum_m \mathrm{sgn}(T) \sum x_j p_{nmjs} \sum p_{jps}^{\gamma} p_{ps} \right\}$$

Let us examine the problem of computing p_{nmjs}. We have p_{nmjs} defined as follows:-

$$p_{nmjs} = p(y_1 = Y_n, u_o = U_m \mid x_1 = X_j, y_o = Y_s)$$

$$= p(u_o = U_m \mid y_o = Y_n)\, p(y_1 = Y_n \mid x_1 = X_j)$$

$$= p^{\gamma}_{ms}\, p_{nj}$$

Thus, for T, we have

$$T = \frac{\sum_j X_j p_{nj} \sum_p p^{\gamma}_{jps} p_{ps}}{\sum p_{nj} \sum p^{\gamma}_{jps} p_{ps}}$$

and for $V_0^{\circ}(p(X_0 \mid I_0))$ we have

$$V_o^{\circ}(p(x_o \mid I_o)) = \inf_{\gamma_o(Y_s)} \left\{ \sum_j X_j^2 \sum_p p_{jps}\, p_{ps} + 5/3 \right.$$

$$\left. -2 \sum_n \operatorname{sgn} \left\{ \sum X_j p_{nj} \sum p^{\gamma}_{jps} p_{ps} \right\} \sum X_j p_{nj} \sum p^{\gamma}_{jps} p_{ps} \right\}$$

Unlike the zero-memory problem, all that we have to do here is find the infimum of this expression with repsect to the two possible function values $\gamma_0(Y_s) = +1$ and $\gamma_0(Y_s) = -1$. This will yield $\gamma_0(Y_s)$ as some function of p_{ps}.

Let us compute $\sum_p p_{jps}^{\gamma} p_{ps} = G_{js}\Big|_{\gamma_0(Y_s) = \substack{+1 \\ -1}}$, where

G_{js} for $\gamma_0(Y_s) = 1$

1	3	$\frac{1}{3} p_{1s}$
2	2	$\frac{1}{3}(p_{1s} + p_{2s})$
3	1	$\frac{1}{3}(p_{1s} + p_{2s} + p_{3s})$
4	0	$\frac{1}{3}(p_{2s} + p_{3s})$
5	-1	$\frac{1}{3} p_{3s}$
6	-2	0
7	-3	0
j	X_j	

G_{js} for $\gamma_0(Y_s) = -1$

1	3	0
2	2	0
3	1	$\frac{1}{3} p_{1s}$
4	0	$\frac{1}{3}(p_{1s} + p_{2s})$
5	-1	$\frac{1}{3}(p_{1s} + p_{2s} + p_{3s})$
6	-2	$\frac{1}{3}(p_{2s} + p_{3s})$
7	-3	$\frac{1}{3} p_{3s}$
j	X_j	

Also $p_{nj} \triangleq p(y_1 = Y_n \mid x_1 = X_j)$, and :-

		1	2	3	4	5	6	7	j
		3	2	1	0	-1	-2	-3	X_j
1	4	1/3	0	0	0	0	0	0	
2	3	1/3	1/3	0	0	0	0	0	
3	2	1/3	1/3	1/3	0	0	0	0	
4	1	0	1/3	1/3	1/3	0	0	0	
5	0	0	0	1/3	1/3	1/3	0	0	
6	-1	0	0	0	1/3	1/3	1/3	0	
7	-2	0	0	0	0	1/3	1/3	1/3	
8	-3	0	0	0	0	0	1/3	1/3	
9	-4	0	0	0	0	0	0	1/3	
n	Y_n								

$p_{nj} =$

Also

$$I_{ns} \Big|_{\gamma_0(Y_s) = +1} \triangleq \sum_j X_j \, p_{nj} \, G_{js},$$

Thus:-

$$I_{ns}\Big|_{\gamma_0(Y_s)=1} =$$

1	4	$3/9\, p_{1s}$
2	3	$3/9\, p_{1s} + 2/9(p_{1s}+p_{2s})$
3	2	$3/9\, p_{1s} + 2/9\,(p_{1s}+p_{2s}) + 1/9\,(p_{1s}+p_{2s}+p_{3s})$
4	1	$2/9\,(p_{1s}+p_{2s}) + 1/9\,(p_{1s}+p_{2s}+p_{3s})$
5	0	$1/9\,(p_{1s}+p_{2s}+p_{3s}) - 1/9\, p_{3s}$
6	-1	$-1/9\, p_{3s}$
7	-2	$-1/9\, p_{3s}$
8	-3	0
9	-4	0
n	Y_n	

and :-

$$I_{ns}\Big|_{\gamma_0(Y_s)=-1} =$$

1	4	0
2	3	0
3	2	$1/9\, p_{1s}$
4	1	$1/9\, p_{1s}$
5	0	$1/9\, p_{1s} - 1/9\,(p_{1s}+p_{2s}+p_{3s})$
6	-1	$-1/9\,(p_{1s}+p_{2s}+p_{3s}) - 2/9\,(p_{2s}+p_{3s})$
7	-2	$-1/9\,(p_{1s}+p_{2s}+p_{3s}) - 2/9\,(p_{2s}+p_{3s}) - 3/9\, p_{3s}$
8	-3	$-2/9\,(p_{2s}+p_{3s}) - 3/9\, p_{3s}$
9	-4	$-3/9\, p_{3s}$
n	Y_n	

Next

$$\operatorname{sgn} I_{ns}\Big|_{\gamma_0(Y_s) = +1} = \begin{bmatrix} + \\ + \\ + \\ + \\ + \\ - \\ - \\ 0 \\ 0 \end{bmatrix}$$

and

$$\operatorname{sgn} I_{ns}\Big|_{\gamma_0(Y_s) = -1} = \begin{bmatrix} 0 \\ 0 \\ + \\ + \\ - \\ - \\ - \\ - \\ - \end{bmatrix}$$

Thus $\sum_n (\operatorname{sgn}(I_{ns})\ I_{ns})\Big|_{\gamma_0(Y_s) = +1}$

$$= 3/9\, p_{1s} + 1/9\,(5p_{1s} + 2p_{2s}) + 1/9\,(6p_{1s} + 3p_{2s} + p_{3s})$$

$$+ 1/9\,(3p_{1s} + 3p_{2s} + p_{3s}) + 1/9\,(p_{1s} + p_{2s}) + 2/9\, p_{3s}$$

$$= \underline{2p_{1s} + p_{2s} + 4/9\, p_{3s}}$$

And, similarly,

$$\sum_n \left\{ \frac{\text{sgn}(I_{ns})\ I_{ns}}{} \right\} \Bigg| \gamma_0(Y_s) = -1$$

$$= 2/9\, p_{1s} + 1/9\,(p_{2s} + p_{3s}) + 1/9\,(p_{1s} + 3p_{2s} + 3p_{3s})$$

$$+ 1/9\,(p_{1s} + 3p_{2s} + 6p_{3s}) + 1/9\,(2p_{2s} + 5p_{3s}) + 3/9\, p_{3s}$$

$$= \underline{4/9\, p_{1s} + p_{2s} + .2p_{3s}}$$

Notice that we still need to compute $\sum_p \sum_j X_j^2\, p_{jps}\, \gamma_{ps}$ - this, as before, is simply:-

$$\sum_p (X_p^2 + \gamma_0^2(Y_s) + 2X_p\gamma_0(Y_s) + \sigma_w^2)\, p_{ps}$$

$$= 5/3 + \sum_p X_p^2 p_{ps} + 2\gamma_0(Y_s) \sum_p X_p p_{ps}$$

$$= 5/3 + (p_{1s} + p_{3s}) + 2\gamma_0(Y_s)(p_{1s} - p_{3s})$$

Thus we see that at last we have an expression which, in theory we can use for the determination of the optimal $\gamma_0(Y_s)$ for every s.

Thus we have,

$$V_0^\circ(\mathcal{L}(p(x_0/I_0), \gamma_0(I_0)$$

$$= 10/3 + (p_{1S} + p_{3S}) + 2(p_{1S} - p_{3S}) - 2(2p_{1S} + p_{2S} + 4/9\, p_{3S}) \Big|_{\gamma_0(Y_s)=+1}$$

$$= 10/3 + (p_{1S} + p_{3S}) - 2(p_{1S} - p_{3S}) - 2(4/9\, p_{1S} + p_{2S} + 2p_{3S}) \Big|_{\gamma_0(Y_s)=-1}$$

$$= 10/3 - p_{1S} - 2p_{2S} - 17/9\, p_{3S} \Big|_{\gamma_0(Y_s)=+1}$$

$$= 10/3 - 17/9\, p_{1S} - 2p_{2S} - p_{3S} \Big|_{\gamma_0(Y_s)=-1}$$

From this it is clear that the optimal control value $\gamma_0(Y_s)$ at Y_s depends only on the relative magnitudes of p_{1S} and p_{3S}. Denoting the two values of $V_0^\circ(\mathcal{L}(p(X_0 \mid I_0), \gamma_0(I_0))$ under $\gamma_0(Y_s) = +1$ and $\gamma_0(Y_s) = -1$ as V_+ and V_-, we obtain,

$$\Delta V \triangleq V_+ - V_- = \frac{8}{9}\, p_{1S} - \frac{8}{9}\, p_{3S}$$

If ΔV is positive, we will select $\gamma_0(Y_s) = -1$; and if ΔV is negative, we will select $\gamma_0{}^{o}(Y_s) = +1$. In other words,

$$\gamma_0{}^{o}(Y_s) = -\operatorname{sgn}(p_{1s} - p_{3s})$$

To reduce the algebraic complexity, let us set $p\left\{\begin{matrix} X_0 = 1 \\ 0 \\ -1 \end{matrix}\right\} = \frac{1}{3}$. Then,

$$p(x_o = X_p \mid y_o = Y_s) = \frac{p(y_o = Y_s \mid x_o = X_p)\, p(x_o = X_p)}{\Sigma\, p(y_o = Y_s \mid x_o = X_p)\, p(x_o = X_p)}$$

Now $p(y_o = Y_s)$ =

1	2	1/9
2	1	2/9
3	0	3/9
4	-1	2/9
5	-2	1/9
s	Y_s	

Thus

$p(x_0 = X_p \mid y_0 = Y_s) = p_{ps} =$

		1	2	3	4	5	s
		2	1	0	-1	-2	Y_s
1	1	1	1/2	1/3	0	0	
2	0	0	1/2	1/3	1/2	0	
3	-1	0	0	1/3	1/2	1	
p	X_p						

Thus for $p\left\{\begin{matrix} x_0 = 1 \\ 0 \\ -1 \end{matrix}\right\} = \frac{1}{3}$, we obtain

s	1	2	3	4	5
Y_s	2	1	0	-1	-2
$p_{1s} - p_{3s}$	1	1/2	0	-1/2	-1
$\gamma_0(Y_s)$	-1	-1	±1	+1	+1
V_0°	13/9	25/18	46/27	25/18	13/9

From this, the expected optimal cost is,

$$\frac{13}{9}\cdot\frac{1}{9} + \frac{25}{18}\cdot\frac{2}{9} + \frac{46}{27}\cdot\frac{3}{9} + \frac{25}{18}\cdot\frac{2}{9} + \frac{13}{9}\cdot\frac{1}{9} = \frac{122}{81}$$

At this point, we have essentially solved the classical information pattern version of the problem. As for the zero memory solution, all that remains to be done is to express γ_1° explicitly in terms of the control data I_1 rather than the 'state' $p(x_1 \mid I_1)$. This will be done in the next section.

A.3.3 Elimination of $p(x_1|I_1)$ from γ_1^0

First we require P_{jps}^{γ}, where

$$P_{jps}^{\gamma} = P\{x_1 = X_j | x_0 = X_p, y_0 = Y_s, \gamma_0^0\}$$

We have:-

		s	1	2	3	4	5
		Y_s	2	1	0	-1	-2
	j	X_j						$\gamma_0^0(Y_s)$
$P_{1ps}^{\gamma} =$	1	1	0	0	0	1/3	1/3
	2	0	0	0	0	0	0
	3	-1	0	0	0	0	0
$P_{2ps}^{\gamma} =$	1	1	0	0	0	1/3	1/3
	2	0	0	0	0	1/3	1/3
	3	-1	0	0	0	0	0
$P_{3ps}^{\gamma} =$	1	1	1/3	1/3	1/3	1/3	1/3
	2	0	0	0	0	1/3	1/3
	3	-1	0	0	0	1/3	1/3
$P_{4ps}^{\gamma} =$	1	1	1/3	1/3	1/3	0	0
	2	0	1/3	1/3	1/3	1/3	1/3
	3	-1	0	0	0	1/3	1/3
$P_{5ps}^{\gamma} =$	1	1	1/3	1/3	1/3	0	0
	2	0	1/3	1/3	1/3	0	0
	3	-1	1/3	1/3	1/3	1/3	1/3
$P_{6ps}^{\gamma} =$	1	1	0	0	0	0	0
	2	0	1/3	1/3	1/3	0	0
	3	-1	1/3	1/3	1/3	1/3	1/3
$P_{7ps}^{\gamma} =$	1	1	0	0	0	0	0
	2	0	0	0	0	0	0
	3	-1	1/3	1/3	1/3	0	0

Recall that p_{ps} =

1	1/2	1/3	0	0
0	1/2	1/3	1/2	0
0	0	1/3	1/2	1

Thus, if $\sum_p p_{jps} \, p_{ps} \triangleq \bar{p}_{js}$, then

$\bar{P}_{js}$ =

		1	2	3	4	5	s
		2	1	0	-1	-2	Y_s
1	3	0	0	0	0	0	
2	2	0	0	0	1/6	0	
3	1	1/3	1/6	1/9	1/3	1/3	
4	0	1/3	1/3	2/9	1/3	1/3	
5	-1	1/3	1/3	1/3	1/6	1/3	
6	-2	0	1/6	2/9	0	0	
7	-3	0	0	1/9	0	0	
j	X_j						

Also, we require $p_{nj} = p\left\{ y_1 = Y_n \mid x_1 = X_j \right\}$

$= p\left\{ v_1 = Y_n - X_j \right\}$, that is -

p_{nj} =

		1	2	3	4	5	6	7	j
		3	2	1	0	-1	-2	-3	X_j
1	4	1/3	0	0	0	0	0	0	
2	3	1/3	1/3	0	0	0	0	0	
3	2	1/3	1/3	1/3	0	0	0	0	
4	1	0	1/3	1/3	1/3	0	0	0	
5	0	0	0	1/3	1/3	1/3	0	0	
6	-1	0	0	0	1/3	1/3	1/3	0	
7	-2	0	0	0	0	1/3	1/3	1/3	
8	-3	0	0	0	0	0	1/3	1/3	
9	-4	0	0	0	0	0	0	1/3	
n	Y_n								

Finally, we can compute the expression $\Sigma_j X_j p_{nj} \Sigma_{jps} p_{ps}$ as follows on the next page:-

$\Sigma X_j P_{nj} \bar{P}_{js} =$

		1	2	3	4	5	s
		+2	+1	0	-1	-2	Y_s
1	4	0	0	0	0	0	
2	3	0	0	0	1/9	0	
3	2	1/9	1/18	1/27	1/6	1/9	
4	1	1/9	1/18	1/27	1/6	1/9	
5	0	0	-1/18	-2/27	+1/18	0	
6	-1	-1/9	-1/6	-7/27	-1/18	-1/9	
7	-2	-1/9	-1/6	-16/27	-1/18	-1/9	
8	-3	0	-1/9	-7/27	0	0	
9	-4	0	0	-1/9	0	0	
n	Y_n						

The zero terms in the top and bottom rows of this array correspond to events with zero probability. Zero elements along the row corresponding to the observation $y_1 = 0$ are 'genuine'. Using this array, we can now at last determine the optimal control law at time n=1 for the classical information pattern problem:-

$$\chi_{1}^{\circ}(I_1) = -\operatorname{sgn} \sum_{j} X_j \, p_{nj} \, \bar{p}_{js} =$$

		1	2	3	4	5	s
		2	1	0	-1	-2	Y_s
1	4						
2	3				-1		
3	2	-1	-1	-1	-1	-1	
4	1	-1	-1	-1	-1	-1	
5	0	±1	+1	+1	(-1)	±1	
6	-1	+1	+1	+1	+1	+1	
7	-2	+1	+1	+1	+1	+1	
8	-3		+1	+1			
9	-4			+1			
n	Y_n						

In this array, only those controls have been entered which correspond to control data which can actually occur.

A.3.4 Neo-Classical Solution of Discrete State Problem

Adopting the notation of Theorem 5.1, we have:-

$$V_2^{\circ}(y_2,p_2) \triangleq E\left[x_2^2 \,\middle|\, y_2\right] = \frac{\int x_2^2\, p(y_2|x_2)\,p_2(x_2)\,dx_2}{\int p(y_2|x_2)\,p_2(x_2)\,dx_2}$$

By application of Theorem 4.1,

$$V_1^{\circ}(y_1,p_1) = \min_{u_1}\left\{E\left[\frac{\int x_2^2\, p(y_2\,|\,x_2)\,p_2(x_2)\,dx_2}{\int p(y_2\,|\,x_2)\,p_2(x_2)\,dx_2}\;\middle|\;y_1\right]\right\}$$

where $p_2(x_2) = \dfrac{\int p(x_2\,|\,x_1,u_1)\,p(y_1\,|\,x_1)\,p_1(x_1)\,dx_1}{\int p(y_1\,|\,x_1)\,p_1(x_1)\,dx_1}$

and $p(y_2\,|\,y_1) = \dfrac{\int p(y_2\,|\,x_2)\int p(x_2\,|\,x_1,u_1)\,p(y_1\,|\,x_1)\,p_1(x_1)\,dx_1\,dx_2}{\int p(y_1\,|\,x_1)\,p_1(x_1)\,dx_1}$

As we expect, thanks to some convenient cancelling out, we obtain:-

$$V_1^{\circ}(y_1,p_1) = \min_{u_1}\left\{\frac{\int\int x_2^2\, p(x_2\,|\,x_1,u_1)\,p(y_1\,|\,x_1)\,p_1(x_1)\,dx_1\,dx_2}{\int p(y_1\,|\,x_1)\,p_1(x_1)\,dx_1}\right\}$$

For our problem, we define the following probabilities:-

$$p_{jkm} \overset{\Delta}{=} p\left\{x_2 = X_j \,\middle|\, x_1 = X_k, u_1 = U_m\right\} = p\left\{w_1 = X_j - (X_k + U_m)\right\}$$

$$p_{rk} \overset{\Delta}{=} p\left\{y_2 = Y_r \,\middle|\, x_1 = X_k\right\} = p\left\{v_1 = Y_r - X_k\right\}$$

$$p_k \overset{\Delta}{=} p\left\{x_1 = X_k\right\} \quad \text{- the state variable.}$$

Thus,

$$V_1^{\circ}(y_1, p_1) = V_1^{\circ}(r, p_k) = \min_{u_m} V_1(r', p_k, u_m)$$

$$= \min_{u_m} \left\{ \frac{\sum_j \sum_k X_j^2 \, p_{jkm} \, p_{rk} \, p_k}{\sum_k p_{rk} \, p_k} \right\}$$

where $j = 1,2, \text{---} 1_i (X_j = 5,4, \text{---} 4,-5)$,

and $k = 1,2, \text{---} 7 (X_k = 3,2, \text{---} 2,-3)$.

Evaluating this term for each r and possible control, we obtain:-

Y_r	r	$V_1(r, p_k, u_m)\|_{u_m=1}$	$V_1(r, p_k, u_m)\|_{u_m=-1}$	$V_1^\circ(r, p_k)$
4	1	$50/3$	$14/3$	$14/3$
3	2	$\frac{50p_1 + 29p_2}{3(p_1+p_2)}$	$\frac{14p_1 + 5p_2}{3(p_1+p_2)}$	$\frac{14p_1 + 5p_2}{3(p_1+p_2)}$
2	3	$\frac{50p_1 + 29p_2 + 14p_3}{3(p_1+p_2+p_3)}$	$\frac{14p_1 + 5p_2 + 2p_3}{3(p_1+p_2+p_3)}$	$\frac{14p_1 + 5p_2 + 2p_3}{3(p_1+p_2+p_3)}$
1	4	$\frac{29p_2 + 14p_3 + 5p_4}{3(p_2+p_3+p_4)}$	$\frac{5p_2 + 2p_3 + 5p_4}{3(p_2+p_3+p_4)}$	$\frac{5p_2 + 2p_3 + 5p_4}{3(p_2+p_3+p_4)}$
0	5	$\frac{14p_3 + 5p_2 + 2p_5}{3(p_3+p_4+p_5)}$	$\frac{2p_3 + 5p_4 + 14p_5}{3(p_3+p_4+p_5)}$	*
-1	6	$\frac{5p_4 + 2p_5 + 5p_6}{3(p_4+p_5+p_6)}$	$\frac{5p_4 + 14p_5 + 29p_6}{3(p_4+p_5+p_6)}$	$\frac{5p_4 + 2p_5 + 5p_6}{3(p_4+p_5+p_6)}$
-2	7	$\frac{2p_5 + 5p_6 + 14p_7}{3(p_5+p_6+p_7)}$	$\frac{14p_5 + 29p_6 + 50p_7}{3(p_5+p_6+p_7)}$	$\frac{2p_5 + 5p_6 + 14p_7}{3(p_5+p_6+p_7)}$
-3	8	$\frac{5p_6 + 14p_7}{3(p_6+p_7)}$	$\frac{29p_6 + 50p_7}{3(p_6+p_7)}$	$\frac{5p_6 + 14p_7}{3(p_6+p_7)}$
-4	9	$14/3$	$50/3$	$14/3$

* $\begin{cases} \dfrac{14p_3 + 5p_4 + 2p_5}{3(p_3+p_4+p_5)} & \text{if } p_3 < p_5 \\ \dfrac{2p_3 + 5p_4 + 14p_5}{3(p_3+p_4+p_5)} & \text{if } p_5 > p_3 \end{cases}$

Knowing the state transition equation as above, we can now eliminate p_k from our expression for $V_1^{\circ}(r,p_k)$, obtaining $V_1^{\circ}(r,\mathcal{L}_0(p_0,u_0,y_0))$. This is computed below:-

$V_1^{\circ}(y_1,\mathcal{L}_0(p_0,u_0,y_0))\Big|_{u_0=+1}$ =

		Y_n: 2	1	0	-1	-2
		n: 1	2	3	4	5
4	1	14/3	14/3	14/3	14/3	14/3
3	2	19/6	24/9	24/9	5/3	—
2	3	21/9	28/15	30/18	9/9	2/3
1	4	7/6	19/15	26/21	19/15	7/6
0	5	2/3	9/9	30/18	28/15	21/9
-1	6	—	5/3	12/9	12/9	7/6
-2	7	—	—	2/3	2/3	2/3
-3	8	—	—	—	—	—
-4	9	14/3	14/3	14/3	14/3	14/3
Y_r	r					

$V_1^{\circ}(y_1,\mathcal{L}_0(p_0,u_0,y_0))\Big|_{u_0=-1}$ =

4	1	14/3	14/3	14/3	14/3	14/3
3	2	—	—	—	—	—
2	3	2/3	2/3	2/3	—	—
1	4	7/6	12/9	12/9	5/3	—
0	5	21/9	28/15	30/18	9/9	2/3
-1	6	7/6	19/15	26/21	19/15	7/6
-2	7	2/3	9/9	30/18	28/15	21/9
-3	8	—	5/3	24/9	24/9	11/6
-4	9	14/3	14/3	14/3	14/3	14/3
Y_r	r					

Finally, we need to compute:-

$$p(y_1|y_0) = \frac{\int p(y_1|x_1) \int p(x_1|x_0,u_0) p(y_0|x_0) p_0(x_0)\, dx_0\, dx_1}{\int p(y_0|x_0) p_0(x_0)\, dx_0}$$

$$= \frac{\sum_k p(y_1=Y_r|x_1=X_k) \sum_p p(x_1=X_k|x_0=X_p,u_0=U_m) p(y_0=Y_n|x_0=X_p) p(x_0=X_p)}{\sum_p p(y_0=Y_n|x_0=X_p) p(x_0=X_p)}$$

that is:-

$$p_{rn} = \frac{\sum_k p_{rk} \sum_p p_{kpm} p_{np} p_p}{\sum_p p_{np} p_p}$$

$$= \sum_k p_{rk} p_k,$$

where p_k has already been computed to obtain the state transition equation.

Thus, for p_{rn}, we obtain:-

$$p_{rn}\Big|_{u_m=+1} =$$

		2	1	0	-1	-2	Y_n
		1	2	3	4	5	n
4	1	1/9	1/18	1/27	0	0	
3	2	2/9	1/6	1/9	1/18	0	
2	3	1/3	5/18	2/9	1/6	1/9	
1	4	2/9	5/18	7/27	5/18	2/9	
0	5	1/9	1/6	2/9	5/18	1/3	
-1	6	0	1/18	1/9	1/6	2/9	
-2	7	0	0	1/27	1/18	1/9	
-3	8	0	0	0	0	0	
-4	9	0	0	0	0	0	
Y_r	r						

$$p_{rn}\Big|_{u_m=-1} =$$

		2	1	0	-1	-2	Y_n
		1	2	3	4	5	n
4	1	0	0	0	0	0	
3	2	0	0	0	0	0	
2	3	1/9	1/18	1/27	0	0	
1	4	2/9	1/6	1/9	1/18	0	
0	5	1/3	5/18	2/9	1/6	1/9	
-1	6	2/9	5/18	7/27	5/18	2/9	
-2	7	1/9	1/6	2/9	5/18	1/3	
-3	8	0	1/18	1/9	1/6	2/9	
-4	9	0	0	1/27	1/18	1/9	
Y_r	r						

At least we are now in a position to compute $E\left[V_1^{\circ}(y_1, L_0(p_0, u_0, y_0)) \,\middle|\, y_0\right]$, as required by Theorem 5.1. We obtain:-

$E\left[V_1^{\circ}(y_1, L_0(p_0, u_0, y_0)) \,\middle|\, y_0\right] =$

Y_n	n	$u_0 = +1$	$u_0 = -1$	$V_0^{\circ}(y_0, p_0)$
2	1	7/3	13/9	13/9
1	2	10/6	25/18	25/18
0	3	46/27	46/27	46/27
-1	4	25/18	10/6	25/18
-2	5	13/9	7/3	13/9

At this stage we have solved our problem using the neo-classical approach. As we might have expected, the optimal costs we have computed are identical with those computed in Chapter 4 using Theorem 3.2. As we saw in Chapter 4, the probabilities of the observation y_0 are given by:-

$$P\begin{bmatrix} y_0 = 2 \\ y_0 = 1 \\ y_0 = 0 \\ y_0 = -1 \\ y_0 = 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} \\ \frac{2}{9} \\ \frac{3}{9} \\ \frac{2}{9} \\ \frac{1}{9} \end{bmatrix}$$

This yields an expected optimal cost of $\frac{122}{81}$, as obtained in Chapter 4. (Recall that for the zero-memory problem, the expected optimal cost is $\frac{126}{81}$, obtained in Chapter 4 by direct enumeration.)

A.3.5 Neo-Classical Solution of Witsenhausen's Problem

Applying Theorem 5.1, we obtain:-

$$V_1^{\circ}(y_1,p_1) = \min_{u_1} E\left[x_2^2 \mid y_1\right]$$

$$= \min_{u_1} E\left[(x_1 - u_1)^2 \mid y_1\right]$$

$$= \min_{u_1} E\left[x_1^2+u_1^2 - 2x_1 u_1 \mid y_1\right]$$

$$= \min_{u_1} \left\{u_1^2 - 2u_1 E\left[x_1 \mid y_1\right] + E\left[x_1^2 \mid y_1\right]\right\}$$

This is minimised when $u_1 = E\left[x_1 \mid y_1\right]$, giving

$$V_1^{\circ}(y_1,p_1) = E^2\left[x_1 \mid y_1\right] - 2E^2\left[x_1 \mid y_1\right] + E\left[x_1^2 \mid y_1\right]$$

$$= E\left[x_1^2 \mid y_1\right] - E^2\left[x_1 \mid y_1\right]$$

$$= \frac{\int x_1^2 p(y_1|x_1)p_1(x_1)dx_1}{\int p(y_1|x_1)p_1(x_1)dx_1} - \left(\frac{\int x_1 p(y_1|x_1)p_1(x_1)dx_1}{\int p(y_1|x_1)p_1(x_1)dx_1}\right)^2$$

Now the state transition equation is given by:-

$$p_1(x_1) = \frac{\int p(x_1|x_0,y_0;\gamma_0)p(y_0\ x_0)p_0(x_0)dx_0}{\int p(y_0|x_0)p_0(x_0)dx_0}$$

and also,

$$p(y_1|y_0) = \int \frac{p(y_1|x_1) \int p(x_1|x_0,y_0;\gamma_0)\, p(y_0|x_0)\, p_0(x_0)\, dx_0}{\int p(y_0|x_0)\, p_0(x_0)\, dx_0}\, dx,$$

Now $p(x_1|x_0,y_0;\gamma_0) = \delta(x_1 - (x_0 + \gamma_0(y_0)))$

Now $p(y_0|x_0) = \delta(y_0 - x_0)$,

$$p(y_1|x_1) = \frac{\Phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\},$$

and $$p_0(x_0) = \frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\},$$

where $$\Phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-t^2}{2}\right]$$

Thus,

$$p_1(x_1) = \frac{\int \delta(x_1-(x_0+\gamma_0(y_0)))\ \delta(y_0-x_0)\ \frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\} dx_0}{\int \delta(y_0-x_0)\ \frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\} dx_0}$$

$$= \frac{\delta(x_1-(y_0+\gamma_0(y_0)))\ \frac{\Phi}{\sigma_0}\left\{\frac{y_0}{\sigma_0}\right\}}{\frac{\Phi}{\sigma_0}\left\{\frac{y_0}{\sigma_0}\right\}}$$

$$= \delta(x_1 - (y_0 + \gamma_0(y_0))),$$

and,

$$p(y_1|x_1) = \frac{\Phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (y_0 + \gamma_0(y_0)))$$

Making the substitution for p_1 in our expression for V_1°, we obtain:-

$$V_1^{\circ}(y_1, L_0(p_0, y_0, u_0)) = \frac{\int x_1^2 \frac{\Phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (y_0 + u_0))\,dx_1}{\int \frac{\Phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (y_0 + u_0))\,dx_1}$$

$$- \left(\frac{\int x_1 \frac{\Phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (y_0 + u_0))\,dx_1}{\int \frac{\Phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (y_0 + u_0))\,dx_1} \right)^2$$

$$= (y_0 + u_0)^2 - (y_0 + u_0)^2 = 0$$

By further application of Theorem 5.1, we obtain:-

$$V_0^{\circ}(y_0, p_0) = \min_{u_0} E\left[k^2 u_0^2 + V_1^{\circ}(y_1, \mathcal{L}_0(y_0, p_0, \gamma_0(y_0))) \,\middle|\, y_0\right]$$

$$= \min_{u_0} \left\{k^2 u_0^2\right\}$$

$$= \underline{0} \quad \text{when } u_0 = 0 \text{ for all } y_0$$

Thus we see that the optimal control at time t = 0 is zero - in other words at t = 0 the controller does nothing (since to do something would incur a cost) and leaves all the control work until t = 1. At time t = 1, if we apply the proposed 'state transition equation':-

$$p_1(x_1) = \iint p(x_1\ x_0, \gamma_0(y_0))p(y_0|x_0)p_0(x_0)dx_0dy_0$$

$$= \iint \delta(x_1-x_0)\ \delta(y_0-x_0)\ \frac{\Phi}{\sigma_0}\left\{\frac{x_0}{\sigma_0}\right\}\ dx_0dy_0$$

$$= \frac{\Phi}{\sigma_0}\left\{\frac{x_1}{\sigma_0}\right\},$$

the control is given by:-

$$u_1 = \gamma_1(y_1) = E\left[x_1|y_1\right]$$

$$= \frac{\int x_1 p(y_1|x_1)p_1(x_1)dx_1}{\int p(y_1|x_1)dx_1}$$

$$= \frac{\int x_1\ \frac{\Phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\}\ \frac{\Phi}{\sigma_0}\left\{\frac{x_1}{\sigma_0}\right\}\ dx_1}{\int \frac{\Phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\}\frac{\Phi}{\sigma_0}\left\{\frac{x_1}{\sigma_0}\right\}\ dx_1}$$

$$= \frac{\int x_1 \frac{\Phi}{\sigma_m} \left\{ \frac{x_1 - y_1 \frac{\sigma_0^2}{\sigma_0^2+\sigma_v^2}}{\sigma_m} \right\} \frac{\Phi}{\sigma_n} \left\{ y_1 \right\} dx_1}{\int \frac{\Phi}{\sigma_m} \left\{ \frac{x_1 - y_1 \frac{\sigma_0^2}{\sigma_0^2+\sigma_v^2}}{\sigma_m} \right\} \frac{\Phi}{\sigma_n} \left\{ y_1 \right\} dx_1} ,$$

where $\sigma_m^2 = \sigma_0^2 \sigma_v^2 / (\sigma_0^2 + \sigma_v^2)$,

and $\sigma_n^2 = \sigma_0^2 + \sigma_v^2$

Thus, $u_1^{\circ} = \gamma_1^{\circ}(y_1) = - y_1 \sigma_0^2 / (\sigma_0^2 + \sigma_v^2)$

Thus, in summary via the neo-classical approach we obtain the following optimal controls for the Witsenhausen problem:-

$$\gamma_0^{\circ}(y_0) = 0$$

$$\gamma_1^{\circ}(y_1) = -y_1 \sigma_0^2 / (\sigma_0^2 + \sigma_v^2)$$

APPENDIX 4

Perturbation Analysis

		Page
A.4.1	Perturbation Solution of Classical Information Pattern Problem	A4.1
	A.4.1.1 Optimal Control at Time N-1	A4.1
	A.4.1.2 Optimal Control at Time N-2	A4.4
	A.4.1.3 Optimal Control at Time k	A4.13
A.4.2	Perturbation Solution of Witsenhausen's Problem	A4.18
	A.4.2.1 Dynamic Programming Formulation	A4.18
	A.4.2.2 Determination of Approximation to Optimal Control Law	A4.26

A.4.1 Perturbation Solution of Classical Information Pattern Problem

A.4.1.1 Optimal Control at time N-1

We have,

$$V^{\circ}_{N-1}(y_{N-1}, p_{N-1}) = \min_{u_{N-1}} \left\{ \int x_{N-1}^2 p(x_{N-1}/y_{N-1})\, dx_{N-1} + u_{N-1}^2 \right.$$

$$+ \int \left\{ \frac{\int x_N^2 p(y_N/x_N) \int p(x_N/x_{N-1}, u_{N-1})\, p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}\, dx_N}{\int p(y_N/x_N) \int p(x_N/x_{N-1}, u_{N-1}) p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}\, dx_N} \right\}$$

$$\left. \cdot \frac{\int p(y_N/x_N) \int p(x_N/x_{N-1}, u_{N-1})\, p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}\, dx_N\, dy_N}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \right\}$$

$$= \min_{u_{N-1}} \left\{ \int x_{N-1}^2 p(x_{N-1}/y_{N-1})\, dx_{N-1} + u_{N-1}^2 \right.$$

$$\left. + \int \frac{\int x_N^2 p(y_N/x_N) \int p(x_N/x_{N-1}, u_{N-1})\, p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}\, dx_N\, dy_N}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \right\}$$

$$= \min_{u_{N-1}} \left\{ \frac{\int x_{N-1}^2 p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} + u_{N-1}^2 \right.$$

$$\left. + \frac{\int\int x_N^2 p(x_N/x_{N-1}, u_{N-1})\, dx_N\, p(y_{N-1}/x_{N-1}) p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \right\}$$

Now

$$p(x_N / x_{N-1}, u_{N-1}) = \frac{\phi}{\delta\sigma_w}\left\{\frac{x_N - (x_{N-1} + u_{N-1})}{\delta\sigma_w}\right\}$$

Thus

$$\int x_N^2\, p(x_N/x_{N-1}, u_{N-1})\, dx_N = (x_{N-1} + u_{N-1})^2 + \delta^2\sigma_w^2,$$

and

$$V_{N-1}^{\circ}(y_{N-1}, p_{N-1}) = \min_{u_{N-1}} \left\{ \int (x_{N-1}^2 + u_{N-1}^2 + (x_{N-1} + u_{N-1})^2 + \delta^2\sigma_w^2) \cdot \right.$$

$$\left. \cdot \frac{p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \right\}$$

$$= \min_{u_{N-1}} \left\{ \frac{\int 2x_{N-1}^2\, p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} + 2u_{N-1}^2 \right.$$

$$\left. + \delta^2\sigma_w^2 + \frac{2u_{N-1} \int x_{N-1}\, p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \right\}$$

This expression is minimised with respect to u_{N-1} at u_{N-1}° where u_{N-1}° is defined as follows:-

$$u_{N-1}^{\circ} = \gamma_{N-1}^{\circ}(y_{N-1}, p_{N-1}) = -\frac{1}{2} \cdot \frac{\int x_{N-1}\, p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}/x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}$$

and the corresponding optimal V_{N-1}° is as follows:-

$$V_{N-1}^{\circ}(y_{N-1}, p_{N-1}) = \left\{ \frac{\int 2x_{N-1}^2\, p(y_{N-1}|x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}|x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \right.$$

$$\left. -\frac{1}{2} \cdot \left\{ \frac{\int x_{N-1}\, p(y_{N-1}|x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}}{\int p(y_{N-1}|x_{N-1})\, p_{N-1}(x_{N-1})\, dx_{N-1}} \right\}^2 + \delta^2\sigma_w^2 \right\}$$

We can, of course, express u_{N-1}° and V_{N-1}° more meaningfully as follows,

$$u_{N-1}^{\circ} = \gamma_{N-1}^{\circ}(y_{N-1}, p_{N-1}) = -\frac{1}{2} E[x_{N-1}|y_{N-1}],$$

and $$V_{N-1}^{\circ}(y_{N-1}, p_{N-1}) = 2E[x_{N-1}^2|y_{N-1}] - \frac{1}{2}\left\{E[x_{N-1}|y_{N-1}]\right\}^2 + \delta^2\sigma_w^2$$

As usual, the analysis at time N-1 presents no difficulty and we have derived the optimal control u_{N-1}° as an exact explicit functional of the state (y_N, p_{N-1}). There is no need to resort to perturbation analysis. This will not be the case at time N-2.

A.4.1.2 Optimal Control at Time N-2

Applying Theorem 5.1 to the problem of finding γ_{N-2}°, leads to the following expression for V_{N-2}°:-

$$V_{N-2}^{\circ}(y_{N-2}, p_{N-2}) = \min_{u_{N-2}} E[x_{N-2}^2 + u_{N-2}^2 + V_{N-1}^{\circ}(y_{N-1}, p_{N-1}) \mid y_{N-2}]$$

where $p_{N-1}(x_{N-1})$ and $p(y_{N-1} \mid y_{N-2})$ are defined by equations similar to those presented above for $p_N(x_N)$ and $p(y_N \mid y_{N-1})$. Substituting for V_{N-1}°, we obtain:-

$$V_{N-2}^{\circ}(y_{N-2}, p_{N-2}) = \min_{u_{N-2}} E[x_{N-2}^2 + u_{N-2}^2 + 2E[x_{N-1}^2 \mid y_{N-1}] + \delta^2\sigma_\omega^2 - \tfrac{1}{2}(E[x_{N-1} \mid y_{N-1}])^2 \mid y_{N-2}]$$

The term $E\left[X_{N-1}^2 \mid y_{N-1}\right]$ in this expression represents no problem. If we follow the same sequence of substitutions as we did at time N-1, we shall find that the implicit nesting in the problem results in some very convenient simplifications, namely

$$E[E[x_{N-1}^2 \mid y_{N-1}] \mid y_{N-2}] = E[x_{N-1}^2 \mid y_{N-2}]$$

$$= E[(x_{N-2} + u_{N-2} + \delta\omega)^2 \mid y_{N-2}]$$

$$= u_{N-2}^2 + E[x_{N-2}^2 \mid y_{N-2}] + 2u_{N-2}E[x_{N-2} \mid y_{N-2}] + \delta^2\sigma_\omega^2$$

Thus we obtain for V^{o}_{N-2}, the expression:-

$$V^{o}_{N-2}(y_{N-2}, p_{N-2}) = \min_{u_{N-2}} \; 3u^2_{N-2} + 3E[x^2_{N-2}/y_{N-2}] + 4u_{N-2}E[x_{N-2}/y_{N-2}]$$
$$+ 3\delta^2\sigma_w^2 - \frac{1}{2}.\, E\left[\left(E[x_{N-1}/y_{N-1}]\right)^2 \middle| y_{N-2}\right]$$

It is the appearance of the term $E\left[\left(E\left[x_{N-1}\middle| y_{N-1}\right]\right)^2 \middle| y_{N-2}\right]$ in the above expression which makes difficult the determination of the optimal u_{N-2}. And it is to resolve this difficulty that a perturbation technique is introduced.

In Section 6.2.3, and $o(\epsilon^2)$ expansion of $E\left[x\middle| y\right]$ was developed. It is as follows:-

$$E[x/y](\epsilon) = y + \epsilon^2\sigma_v^2\frac{p_x'(y)}{p_x(y)} + o(\epsilon^2)$$

where p_X is the probability density function of the random variable x in the observation equation:-

$$y = x - \epsilon V$$

Thus, at time N-1, we require

$$\left(E[x_{N-1}/y_{N-1}](\epsilon)\right)^2 = y^2_{N-1} + 2\epsilon^2\sigma_v^2 y_{N-1}\frac{p'_{N-1}(y_{N-1})}{p_{N-1}(y_{N-1})} + o(\epsilon^2)$$

Similarly it can be shown that:-

$$p'_{N-1}(x_{N-1})\Big|_{\epsilon=0} = -\frac{\phi}{\delta\sigma_w}\left\{\frac{x_{N-1}-(y_{N-2}+u_{N-2})}{\delta\sigma_w}\right\}\left\{\frac{x_{N-1}-(y_{N-2}+u_{N-2})}{\delta\sigma_w}\right\}$$

Consequently, we have:-

$$\frac{p'_{N-1}(x_{N-1})}{p_{N-1}(x_{N-1})} = -\left\{\frac{x_{N-1}-(y_{N-2}+u_{N-2})}{\delta^2\sigma_w^2}\right\} + O(\epsilon)$$

Thus

$$\left(E[x_{N-1}/y_{N-1}](\epsilon)\right)^2 = \left\{ y_{N-1}^2 - 2\epsilon^2\sigma_v^2 y_{N-1}\left[\frac{y_{N-1}-(y_{N-2}+u_{N-2})}{\delta^2\sigma_w^2}\right] + O(\epsilon^2)\right\}$$

Notice that this expression contains y_{N-2} and u_{N-2}, which indicates that our notation of $\left(E\left[x_{N-1}\,\middle|\,y_{N-1}\right]\right)^2$ is somewhat inadequate. The reason is that we have introduced our state transition equation as Theorem 5.1 requires. And, as we saw in the previous Chapter, the state p_{N-1} is a functional of p_{N-2} and also y_{N-2} and u_{N-2} (or $\gamma_{N-2}(y_{N-2}, p_{N-2})$. Thus it is not surprising that y_{N-2} and u_{N-2} occur in our expression for $\left(E\left[x_{N-1}\,\middle|\,y_{N-1}\right]\right)^2$.

The final step in evaluating $E\left[\left(E\left[x_{N-1}\,\middle|\,y_{N-1}\right]\right)^2\,\middle|\,y_{N-2}\right]$ is to determine an expression for $p(y_{N-1}\,|\,y_{N-2})$. Now -

Applying Theorem 5.1, we must now express p_{N-1} in terms of p_{N-2} by using our state transition equation:-

$$p_{N-1}(x_{N-1}) = \frac{\int p(x_{N-1}/x_{N-2},u_{N-2})\,p(y_{N-2}/x_{N-2})\,p_{N-2}(x_{N-2})\,dx_{N-2}}{\int p(y_{N-2}/x_{N-2})\,p_{N-2}(x_{N-2})\,dx_{N-2}}$$

$$= \frac{\int \frac{\phi}{\delta\sigma_w}\left\{\frac{x_{N-1}-(x_{N-2}+u_{N-2})}{\delta\sigma_w}\right\}\frac{\phi}{\epsilon\sigma_v}\left\{\frac{y_{N-2}-x_{N-2}}{\epsilon\sigma_v}\right\}p_{N-2}(x_{N-2})\,dx_{N-2}}{\int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y_{N-2}-x_{N-2}}{\epsilon\sigma_v}\right\}p_{N-2}(x_{N-2})\,dx_{N-2}}$$

$$= \frac{\int \frac{\phi}{\delta\sigma_w}\left\{\frac{x_{N-1}-(y_{N-2}+u_{N-2}+\epsilon\sigma_v s)}{\delta\sigma_w}\right\}\phi(s)\,p_{N-2}(y_{N-2}+\epsilon\sigma_v s)\,ds}{\int p(s)\,p_{N-2}(y_{N-2}+\epsilon\sigma_v s)\,ds}$$

Notice that all we require for p_{N-1} is an $o(\epsilon)$ approximation. Thus:-

$$p_{N-1}(x_{N-1})\Big|_{\epsilon=0} = \frac{\phi}{\delta\sigma_w}\left\{\frac{x_{N-1}-(y_{N-2}+u_{N-2})}{\delta\sigma_w}\right\}$$

$$p(y_{N-2}|y_{N-1}) = \frac{\int p(y_{N-1}|x_{N-1}) \int p(x_{N-1}|x_{N-2},u_{N-2}) p(y_{N-2}|x_{N-2}) p_{N-2}(x_{N-2}) dx_{N-2} dx_{N-1}}{\int p(y_{N-2}|x_{N-2}) p_{N-2}(x_{N-2}) dx_{N-2}}$$

$$= \frac{\int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{x_{N-1}-y_{N-1}}{\epsilon\sigma_v}\right\} \int \frac{\phi}{\delta\sigma_w}\left\{\frac{x_{N-1}-(x_{N-2}+u_{N-2})}{\delta\sigma_w}\right\} \frac{\phi}{\epsilon\sigma_v}\left\{\frac{x_{N-2}-y_{N-2}}{\epsilon\sigma_v}\right\} p_{N-2}(x_{N-2}) dx_{N-2} dx_{N-1}}{\int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{x_{N-2}-y_{N-2}}{\epsilon\sigma_v}\right\} p_{N-2}(x_{N-2}) dx_{N-2}}$$

$$= \frac{\int \phi(t) \int \frac{\phi}{\delta\sigma_w}\left\{\frac{y_{N-1}+\epsilon\sigma_v(t-s)-(y_{N-2}+u_{N-2})}{\delta\sigma_w}\right\} \phi(s) p_{N-2}(y_{N-2}+\epsilon\sigma_v s) ds\, dt}{\int \phi(s) p_{N-2}(y_{N-2}+\epsilon\sigma_v s) ds}$$

Thus:-

$$p(y_{N-1}|y_{N-2})\Big|_{\epsilon=0} = \frac{\phi}{\delta\sigma_w}\left\{\frac{y_{N-1}-(y_{N-2}+u_{N-2})}{\delta\sigma_w}\right\},$$

$$\left\{\frac{d}{d\epsilon}\left(p(y_{N-1}|y_{N-2})\right)\right\}_{\epsilon=0} = 0,$$

and

$$\left\{ \frac{d^2}{d\epsilon^2} \left(p(y_{N-1} | y_{N-2}) \right) \right\}_{\epsilon=0} = \frac{\phi}{\delta\sigma_w} \left\{ \frac{y_{N-1} - (y_{N-2} + u_{N-2})}{\delta\sigma_w} \right\} \cdot$$

$$\cdot \frac{2\epsilon^2\sigma_v^2}{\delta^2\sigma_w^2} \left\{ \left(\frac{y_{N-1} - (y_{N-2} + u_{N-2})}{\delta\sigma_w} \right)^2 - 1 + (y_{N-1} - (y_{N-2} + u_{N-2})) \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} \right\}$$

Thus we have finally:-

$$p(y_{N-1} | y_{N-2}) = \frac{\phi}{\delta\sigma_w} \left\{ \frac{y_{N-1} - (y_{N-2} + u_{N-2})}{\delta\sigma_w} \right\} \cdot$$

$$\cdot \left\{ 1 + \frac{\epsilon^2\sigma_v^2}{\delta^2\sigma_w^2} \left\{ \left[\frac{y_{N-1} - (y_{N-2} + u_{N-2})}{\delta\sigma_w} \right]^2 - 1 + (y_{N-1} + (y_{N-2} + u_{N-2})) \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} \right\} \right\}$$

$$+ o(\epsilon^2)$$

At last we are in a position to compute $E\left[\left(E\left[x_{N-1} \mid y_{N-1} \right] \right)^2 \mid y_{N-2} \right]$.

We have,

$$E\left[\left(E\left[x_{N-1} | y_{N-1} \right](\epsilon) \right)^2 \mid y_{N-2} \right]$$

$$= \int \left\{ y_{N-1}^2 - 2y_{N-1}\, \epsilon^2\sigma_v^2 \left\{ \frac{y_{N-1} - (y_{N-2} + u_{N-2})}{\delta^2\sigma_\omega^2} \right\} \frac{\phi}{\delta\sigma_\omega} \left\{ \frac{y_{N-1} - (y_{N-2} + u_{N-2})}{\delta\sigma_\omega} \right\} \cdot \right.$$

$$\left. \cdot \left\{ 1 + \frac{\epsilon^2\sigma_v^2}{\delta^2\sigma_\omega^2} \left\{ \left[\frac{y_{N-1} - (y_{N-2} + u_{N-2})}{\delta\sigma_\omega} \right]^2 - 1 + (y_{N-1} - (y_{N-2} + u_{N-2})) \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} \right\} \right\} \right\} dy_{N-1}$$

$$+ o(\epsilon^2)$$

Putting $y_{N-1} = y_{N-2} + u_{N-2} + \delta\sigma_w s$, we obtain

$$E\left[\left(E\left[x_{N-1} \mid y_{N-1} \right] \right)^2 \mid y_{N-2} \right]$$

$$= \int \left\{ \left[y_{N-2} + u_{N-2} + \delta\sigma_\omega s \right]^2 - 2\left[y_{N-2} + u_{N-2} + \delta\sigma_\omega s \right] \frac{\epsilon^2\sigma_v^2 s}{\delta\sigma_\omega} \right\} \phi(s) \cdot$$

$$\cdot \left\{ 1 + \frac{\epsilon^2\sigma_v^2}{\delta^2\sigma_\omega^2} \left\{ s^2 - 1 + s\delta\sigma_\omega \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} \right\} \right\} ds + o(\epsilon^2)$$

$$= (y_{N-2} + u_{N-2})^2 + \delta^2\sigma_\omega^2 + 2\epsilon^2\sigma_v^2 (y_{N-2} + u_{N-2}) \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} + o(\epsilon^2)$$

Returning at last to our minimisation problem, we now have:-

$$V^{\circ}_{N-2}(y_{N-2}, p_{N-2}) = \min_{u_{N-2}} \Big\{ 3u^2_{N-2} + 3E[x^2_{N-2}/y_{N-2}] + 4u_{N-2}E[x_{N-2}|y_{N-2}]$$

$$+ 3\delta^2\sigma_w^2 - \frac{1}{2}E\Big[\big(E[x_{N-1}|y_{N-1}]\big)^2 \Big| y_{N-2}\Big] + o(\epsilon^2)\Big\}$$

$$= \min_{u_{N-2}} \Big\{ 3u^2_{N-2} + 3\Big(y^2_{N-2} + \epsilon^2\sigma_v^2\Big[1 + 2y_{N-2}\frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})}\Big]\Big)$$

$$+ 4u_{N-2}\Big[y_{N-2} + \epsilon^2\sigma_v^2\frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})}\Big] + 3\delta^2\sigma_w^2 + o(\epsilon^2)$$

$$- \frac{1}{2}\Big[(y_{N-2}+u_{N-2})^2 + \delta^2\sigma_w^2 + 2\epsilon^2\sigma_v^2(y_{N-2}+u_{N-2})\frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})}\Big]\Big\}$$

$$= \min_{u_{N-2}} \Big\{ \frac{5}{2}u^2_{N-2} + \frac{5}{2}y^2_{N-2} + \frac{5}{2}\delta^2\sigma_w^2 + 3u_{N-2}y_{N-2} + 3\epsilon^2\sigma_v^2$$

$$+ 5\epsilon^2\sigma_v^2 y_{N-2}\frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} + 3u_{N-2}\epsilon^2\sigma_v^2\frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} + o(\epsilon^2)\Big\}$$

Differentiating with respect to u_{N-2} and equating to zero, we obtain:-

$$5u_{N-2} + 3y_{N-2} + 3\epsilon^2\sigma_v^2 \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} + o(\epsilon^2) = 0$$

Thus the required solution is:-

$$u^{\circ}_{N-2} = \gamma^{\circ}_{N-2}(y_{N-2}, p_{N-2}, \epsilon) = -\frac{3}{5}\left\{ y_{N-2} + \epsilon^2\sigma_v^2 \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} \right\} + o(\epsilon^2)$$

Substituting this expression for u_{N-2}° into our definition of V_{N-2}°, we obtain:-

$$V^{\circ}_{N-2}(y_{N-2}, p_{N-2}) = \frac{8}{5} y^2_{N-2} + \frac{16}{5} y_{N-2} \epsilon^2\sigma_v^2 \frac{p'_{N-2}(y_{N-2})}{p_{N-2}(y_{N-2})} + \frac{5}{2}\delta^2\sigma_\omega^2 + 3\epsilon^2\sigma_v^2 + o(\epsilon^2)$$

At this stage, we have completed our computations for time N-2. We have obtained both the optimal control law $u_{N-2}^{\circ}(y_{N-2}, p_{N-2})$ and the corresponding $V_{N-2}^{\circ}(y_{N-2}, p_{N-2})$ as $o(\epsilon^2)$ approximations.

A.4.1.3 Optimal Control at time k

The basic steps in the analysis are presented below, although details of the intricate algebra have been omitted.

$$V_k^0(y_k, p_k) = \min_{u_k} \left\{ E\left[x_k^2 + u_k^2 + V_{k+1}^0 (y_{k+1}, p_{k+1}) \mid y_k \right] \right\}$$

First, we substitute for p_{k+1}. From the definition of V_{k+1}, all that we require is an $O(\epsilon)$ approximation for $\dfrac{p'_{k+1}(y_{k+1})}{p_{k+1}(y_{k+1})}$. This is simply,

$$\frac{p'_{k+1}(y_{k+1})}{p_{k+1}(y_{k+1})} = -\left\{ \frac{y_{k+1} - (y_k + u_k)}{\delta^2 \sigma_w^2} \right\} + O(\epsilon)$$

Next we need $p(y_{k+1} \mid y_k)$. The form of this probability density function has already been determined in the calculations at time N-2. We have:-

$$p(y_{k+1}/y_k) = \frac{\phi}{\delta\sigma_w}\left\{ \frac{y_{k+1} - (y_k + u_k)}{\delta\sigma_w} \right\} \left\{ 1 + \frac{\epsilon^2\sigma_v^2}{\delta^2\sigma_w^2} \left\{ \left[\frac{y_{k+1} - (y_k + u_k)}{\delta\sigma_w} \right]^2 - 1 + (y_{k+1} - (y_k + u_k)) \cdot \frac{p'_k(y_k)}{p_k(y_k)} \right\} \right\} + o(\epsilon^2)$$

Finally, all that we need is $E\left[x_k^2 \mid y_k \right]$, and this is simply:-

$$E[x_k^2/y_k] = y_k^2 + \epsilon^2\sigma_v^2 \left\{ 1 + 2y_k \frac{p'_k(y_k)}{p_k(y_k)} \right\} + o(\epsilon^2)$$

Making the appropriate substitutions and performing the necessary integration, we obtain:-

$$V_k^{\circ}(y_k, p_k) = \min_{u_k} \left\{ (1+a_{k+1}) y_k^2 + (1+a_{k+1}) u_k^2 + 2a_{k+1} y_k u_k \right.$$

$$+ 2\epsilon^2\sigma_v^2 y_k (1+a_{k+1}) \frac{p_k'(y_k)}{p_k(y_k)} + 2a_{k+1}\epsilon^2\sigma_v^2 u_k \frac{p_k'(y_k)}{p_k(y_k)}$$

$$\left. + (a_{k+1} + d_{k+1})\delta^2\sigma_w^2 + (1 + 2a_{k+1} - b_{k+1} + c_{k+1})\epsilon^2\sigma_v^2 + o(\epsilon^2) \right\}$$

Differentiating with respect to u_k, equating to zero, and so on, we obtain the following expression for u_k°:-

$$u_k^{\circ} = \gamma_k^{\circ}(y_k, p_k, \epsilon) = \frac{-a_{k+1}}{1+a_{k+1}} \left\{ y_k + \epsilon^2\sigma_v^2 \frac{p_k'(y_k)}{p_k(y_k)} \right\} + o(\epsilon^2)$$

When this expression for u_k° is substituted into our definition V_k°, we obtain:-

$$V_k^{\circ}(y_k, p_k) = \left\{ \frac{1+2a_{k+1}}{1+a_{k+1}} \right\} y_k^2 + 2\epsilon^2\sigma_v^2 \frac{(1+2a_{k+1})}{(1+a_{k+1})} y_k \frac{p_k'(y_k)}{p_k(y_k)}$$

$$+ (a_{k+1} + d_{k+1})\delta^2\sigma_w^2 + (1 + 2a_{k+1} - b_{k+1} + c_{k+1})\epsilon^2\sigma_v^2 + o(\epsilon^2)$$

Thus we see that the following recursive relative relations can be defined for the determination of V_k^o and u_k^o from V_{k+1}^o.

Suppose

$$V_{k+1}^o(y_{k+1}, p_{k+1}) = a_{k+1} y_{k+1}^2 + \epsilon^2 \sigma_v^2 b_{k+1} y_{k+1} \frac{p_{k+1}'(y_{k+1})}{p_{k+1}(y_{k+1})} + c_{k+1} \epsilon^2 \sigma_v^2 + d_{k+1} \delta^2 \sigma_w^2 + o(\epsilon^2)$$

Then

$$u_k^o = \gamma_k^o(y_k, p_k) = -e_k \left\{ y_k + \epsilon^2 \sigma_v^2 \frac{p_k'(y_k)}{p_k(y_k)} \right\} + o(\epsilon^2)$$

and

$$V_k^o(y_k, p_k) = a_k y_k^2 + \epsilon^2 \sigma_v^2 b_k y_k \frac{p_k'(y_k)}{p_k(y_k)} + c_k \epsilon^2 \sigma_v^2 + d_k \delta^2 \sigma_w^2 + o(\epsilon^2)$$

where $a_k = \dfrac{1 + 2a_{k+1}}{1 + a_{k+1}}$,

$b_k = \dfrac{2(1 + 2a_{k+1})}{(1 + a_{k+1})}$,

$c_k = 1 + 2a_{k+1} - b_{k+1} + c_{k+1} = 1 + c_{k+1}$, since $2a_{k+1} = b_{k+1}$

$d_k = a_{k+1} + d_{k+1}$,

and $e_k = \dfrac{a_{k+1}}{1 + a_{k+1}}$

To apply these recursion relations, the starting values obtained above for time N-2 can be used, namely:-

$$a_{N-2} = \frac{8}{5}$$

$$b_{N-2} = \frac{16}{5}$$

$$c_{N-2} = 3$$

$$d_{N-2} = \frac{5}{2}$$

Alternatively the starting values at time N-1 can be used. These have not been quoted here since the computations which we presented previously for V°_{N-1} and u°_{N-1} were not performed explicitly in terms of y_{N-1} and p_{N-1}. It is however a very simple matter to transform our earlier expression for V°_{N-1} into the form desired.

We have:-

$$V^{\circ}_{N-1}(y_{N-1}, p_{N-1}) = 2E[x_{N-1}^2 / y_{N-1}] - \tfrac{1}{2}\left(E[x_{N-1}/y_{N-1}]\right)^2 + \delta^2\sigma_w^2$$

Using our $O(\epsilon^2)$ expansions for $E\left[x_{N-1}^2 \mid y_{N-1}\right]$ and $E\left[x_{N-1} \mid y_{N-1}\right]$, we obtain:-

$$V^{\circ}_{N-1}(y_{N-1}, p_{N-1}) = \frac{3}{2}y_{N-1}^2 + 3\epsilon^2\sigma_v^2\frac{p'_{N-1}(y_{N-1})}{p_{N-1}(y_{N-1})} + 2\epsilon^2\sigma_v^2 + \delta^2\sigma_w^2 + O(\epsilon^2)$$

Thus we have the starting values at time N-1 as follows:-

$$a_{N-1} = \frac{3}{2},$$

$$b_{N-1} = 3,$$

$$c_{N-1} = 2$$

$$d_{N-1} = 1$$

Although of no practical importance, the above starting values do provide a valuable opportunity to check the accuracy of the intricate algebraic manipulations involved in deriving the various recursion relations for a_k, b_k, c_k, and d_k presented earlier.

A.4.2 Perturbation Solution of Witsenhausen's Problem

A.4.2.1 Dynamic Programming Formulation

Applying Theorem 3.1, we have:-

$$V_2^\circ(p_2) \triangleq \int x_2^2 p_2(x_2)\, dx_2$$

Thus

$$V_1^\circ(p_1) = \inf_{\gamma_1} \left\{ \int x_2^2 p_2(x_2)\, dx_2 \right\}, \text{ since } L_1 = 0 \text{ and}$$

where $p_2(x_2)$ is given by the state transition equation. This is as follows:-

$$p_2(x_2) = \iint p(x_2 | x_1, \gamma_1(y_1))\, p(y_1 | x_1)\, p_1(x_1)\, dx_1\, dy_1$$

$$= \iint \delta(x_2 - (x_1 + \gamma_1(y_1)))\, p(x_1 | y_1)\, p(y_1)\, dx_1 dy_1$$

and
$$V_1^\circ(p_1) = \inf_{\gamma_1} \left\{ \iint x_2^2\, \delta(x_2 - (x_1 + \gamma_1(y_1)))\, p(x_1 | y_1) p(y_1)\, dx_1\, dy_1\, dx_2 \right\}$$

$$= \inf_{\gamma_1} \left\{ \iint (x_1 + \gamma_1(y_1))^2 p(x_1 | y_1)\, p(y_1)\, dx_1\, dy_1 \right\}$$

$$= \int \inf_{\gamma_1} \left\{ \int (x_1 + \gamma_1(y_1))^2 p(x_1 | y_1)\, dx_1 \right\} p(y_1)\, dy_1$$

This is minimised when:-

$$\gamma_1^\circ(y_1) = -\int x_1 p(x_1|y_1)\,dx_1 = -E[x_1|y_1]$$

and the corresponding V_1° is as follows:-

$$V_1^\circ(p_1) = \int \left\{ \int x_1^2 p(x_1|y_1)\,dx_1 - \int x_1 p(x_1|y_1)\,dx_1 \right\}^2 p(y_1)\,dy_1$$

$$= \int x_1^2 p_1(x_1)\,dx_1 - \iint \left\{ \frac{\int x_1 \frac{\phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\} p_1(x_1)\,dx_1}{\int \frac{\phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\} p_1(x_1)\,dx_1} \right\}^2 \frac{\phi}{\sigma_v}\left\{\frac{y_1-x_1}{\sigma_v}\right\} \cdot p(x_1)\,dx_1\,dy_1$$

Next to compute V_0°, we must solve:-

$$V_0^\circ(p_0) = \inf_{\gamma^\circ} \left\{ E[k^2 \gamma_0^2(y_0)] + V_1^\circ(p_1) \right\}$$

where

$$p_1(x_1) = \iint p(x_1|x_0, \gamma_0(y_0))\, p(y_0|x_0)\, p_0(x_0)\,dx_0\,dy_0$$

$$= \iint \delta(x_1 - (x_0 + \gamma_0(y_0)))\, \delta(y_0 - x_0)\, p_0(x_0)\,dx_0\,dy_0$$

$$= \int \delta(x_1 - (x_0 + \gamma_0(x_0)))\, p_0(x_0)\,dx_0$$

Thus

$$V_0^0(p_0) = \inf_{\gamma_0} \left\{ \int k^2 \gamma_0^2(x_0) p_0(x_0)\, dx_0 + \iint x_1^2 \delta(x_1 - (x_0 + \gamma_0(x_0))) p_0(x_0)\, dx_0 \right.$$

$$- \iint \left\{ \frac{\iint x_1 \frac{\phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (x_0 + \gamma_0(x_0))) p_0(x_0)\, dx_0\, dx_1}{\iint \frac{\phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (x_0 + \gamma_0(x_0))) p_0(x_0)\, dx_0\, dx_1} \right\}^2 \cdot$$

$$\left. \cdot \frac{\phi}{\sigma_v}\left\{\frac{y_1 - x_1}{\sigma_v}\right\} \delta(x_1 - (x_0 + \gamma_0(x_0))) p_0(x_0)\, dx_0\, dx_1\, dy_1 \right\}$$

Next, make the substitution $f(x) = x + \gamma_0(x)$, and we obtain

$$V_0^0(p_0) = \inf_{f} \left\{ \int k^2 (f(x) - x)^2 p_0(x)\, dx + \int f^2(x) p_0(x)\, dx \right.$$

$$\left. - \iint \left\{ \frac{\int f(x) \frac{\phi}{\sigma_v}\left\{\frac{y_1 - f(x)}{\sigma_v}\right\} p_0(x)\, dx}{\int \frac{\phi}{\sigma_v}\left\{\frac{y_1 - f(x)}{\sigma_v}\right\} p_0(x)\, dx} \right\}^2 \frac{\phi}{\sigma_v}\left\{\frac{y_1 - f(x)}{\sigma_v}\right\} p_0(x)\, dx\, dy_1 \right\}$$

If we introduce the following definitions:-

$$N_f(y) = \int F(x) \frac{\phi}{\sigma_V}\left\{\frac{y-f(x)}{\sigma_V}\right\} p_o(x)\,dx \quad , \text{ and}$$

$$D_f(y) = \int \frac{\phi}{\sigma_V}\left\{\frac{y-F(x)}{\sigma_V}\right\} p_o(x)\,dx$$

we obtain:-

$$V_o^o(p_o) = \inf_f \left\{ \int k^2 (f(x)-x)^2 p_o(x)\,dx + \int f^2(x)\, p_o(x)\,dx \right.$$

$$\left. - \iint \left\{\frac{N_f(y)}{D_f(y)}\right\}^2 \frac{\phi}{\sigma_V}\left\{\frac{y-f(x)}{\sigma_V}\right\} p_o(x)\,dx\,dy \right\}$$

Notice that

$$\frac{d}{dy}[D_f(y)] = -\int \frac{\phi}{\sigma_V}\left\{\frac{y-f(x)}{\sigma_V}\right\}\left\{\frac{y-f(x)}{\sigma_V}\right\} p_o(x)\,dx$$

$$= \frac{-y}{\sigma_V^2} D_f(y) + \frac{N_f(y)}{\sigma_V^2}$$

Thus $$\frac{N_f(y)}{\sigma_V^2 D_f(y)} = \frac{y}{\sigma_V^2} + \frac{1}{D_f(y)} \frac{d}{dy}[D_f(y)]$$

Incorporating this identity into our expression for V_0°, we obtain:-

$$V_0^{\circ}(p_0) = \inf_{f} \left\{ E[k^2(f(x)-x)^2] + E[f^2(x)] \right.$$

$$\left. - \int \left\{ y + \frac{\sigma_v^2}{D_f(y)} \frac{d}{dy}[D_f(y)] \right\}^2 D_f(y)\, dy \right\}$$

Now

$$\int y^2 D_f(y)\, dy = E[f^2(x)] + \sigma_v^2$$

and

$$\int y \frac{d}{dy}[D_f(y)]\, dy = [y D_f(y)] - \int D_f(y)\, dy = -1$$

Thus

$$V_0^{\circ}(p_0) = \inf_{\gamma} \left\{ E[k^2(f(x)-x)^2] + \sigma_v^2 \right.$$

$$\left. - \int \frac{\sigma_v^4}{D_f(y)} \left\{ \frac{d}{dy}[D_f(y)] \right\}^2 dy \right\}$$

At this point it is convenient to consider certain properties of the optimal f(x). These properties were originally proved and stated as Lemma 1 by Witsenhausen in his paper. Those contents of this Lemma in which we are interested here are stated and proved below. (This Lemma will be referred to again in Chapter 8).

Lemma

Any f^o which provides an infimum for $v_0^o(p_0)$ has the following properties:-

i) $E\left[f^o(x)\right] = 0$

ii) $E\left[(f^o(x))^2\right] \leq 4\sigma_0^2$,

where σ_0^2 is the variance of the initial state x_0.

Proof

Suppose $g(y_2) \triangleq -\gamma_2(y)$, then the cost function for Witsenhausen's problem can be expressed as $J(f,g,p_0)$, where

$$J(f,g,p_0) = E[k^2u_0^2 + x_2^2], \text{ that is}$$

$$J(f,g,p_0) = E[k^2(f(x_0)-x_0)^2 + (f(x_0)-g(f(x_0)+v))^2]$$

Two cases must be distinguished. First suppose $E\left[(x_0 - f(x_0))^2\right] > \sigma_0^2$, Then,

$$J(f,g,p_0) \geq k^2 E[(f(x_0)-x_0)^2] > k^2\sigma_0^2$$

Consider then the functions $f' \equiv g' \equiv 0$. We have

$$J(f', g', p_0) = k^2 \sigma_0^2 < J(f, g_0, p_0),$$

and trivially, $E\left[f'(x_0)\right] = 0$,

and $$E\left[(f'(x_0))^2\right] \leqslant 4\sigma_0^2$$

Next suppose that $E\left[(x_0 - f(x_0))^2\right] \leqslant \sigma_0^2$. Then, by Schwarz's inequality, we can show that $E\left[(f(x_0))^2\right] \leqslant 4\sigma_0^2$. Consequently the mean $m = E\left[f(x)\right]$ exists. In this case, consider the functions,

$f'(x) = f(x) - m$ and $g'(y) = g(y + m) - m$. We have

$$J(f', g', p_0) = E\left[k^2(f(x_0) - x_0 - m)^2 + (f(x_0) - m - g(f(x_0) + v) + m)^2\right]$$

$$= J(f, g, p_0) - k^2 m^2 \leqslant J(f, g, p_0)$$

Again for this selection of f', we have:-

$$E\left[f'(x_0)\right] = 0$$

and $$E\left[(f'(x_0))^2\right] \leqslant 4\sigma_0^2.$$

Thus, for any $(f, g) \in \Gamma$, we can always define functions (f', g') such that

$$J(f', g', p_0) \leqslant J(f, g, p_0)$$

with the properties that,

$$E\left[f'(x)\right] = 0$$

and $$E\left[(f'(x))^2\right] \leq 4\sigma_0^2$$

In other words, we need only consider the class of Borel measurable functions for f which have the above properties. This completes the proof of the Lemma.

Returning now to our expression for V_0^o, we have,

$$V_o^o(p_o) = \inf_f \left\{ E\left[k^2(f(x)-x)^2\right] + \sigma_v^2 - \int \frac{\sigma_v^4}{D_f(y)} \left\{ \frac{d}{dy}\left[D_f(y)\right] \right\}^2 dy \right\}$$

In order to utilise the perturbation approach, we wish to introduce a small parameter ϵ into the observation equation as follows:-

$$y_1 = x_1 + \epsilon v$$

At the same time, it is convenient to define a cost functional J as follows:-

$$J(p_o, e, f) = E\left[k^2(f(x)-x)^2\right] + \epsilon^2\sigma_v^2 - \epsilon^4\sigma_v^4 \int \frac{d}{dy}\left[D_f(y)\right] \frac{dy}{D_f(y)}$$

This completes the dynamic programming formulation of Witsenhausen's problem.

A.4.2.2 Determination of Approximation to Optimal Control Law

Below we consider the problem of expanding $D_g(y, \epsilon)$ about $\epsilon = 0$. Introducing the specific form $\frac{\Phi}{\sigma_0}\left\{\frac{x}{\sigma_0}\right\}$ for $p_0(x_0)$, we have:-

$$D_g(y,\epsilon) = \int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-z}{\epsilon\sigma_v}\right\}\frac{\phi}{\sigma_0}\left\{\frac{q(z,\epsilon)}{\sigma_0}\right\}\frac{dq}{dz}(z,\epsilon)\,dz$$

Next we introduce the substitution $z = y + s\epsilon\sigma_v$, and we obtain:-

$$D_g(y,\epsilon) = \int \phi(s)\frac{\phi}{\sigma_0}\left\{\frac{q(y+\epsilon\sigma_v s,\epsilon}{\sigma_0}\right\}\frac{dq}{dz}(y+s\epsilon\sigma_v,\epsilon)\,ds$$

Evaluation of

$$D_g(y,\epsilon)\Big|_{\epsilon=0}, \quad \frac{\partial}{\partial\epsilon}\left\{D_g(y,\epsilon)\right\}\Big|_{\epsilon=0}, \text{ and } \frac{\partial^2}{\partial\epsilon^2}\left\{D_g(y,\epsilon)\right\}\Big|_{\epsilon=0}$$

involves some intricate but fairly straightforward algebra. In order to reduce the burden of this algebra, we can without loss of generality make the assumption that $\sigma_0 = 1$. This simply amounts to a re-scaling of x_0. We thus obtain the following expansion for $D_g(y, \epsilon)$:-

$$D_g(y,\epsilon) = \phi(y) + \frac{\epsilon^2\sigma_v^2}{2!}\left\{y^2\phi(y) - \phi(y)\right\} + o(\epsilon^2)$$

Thus

$$D_g'(y,\epsilon) = \frac{d}{dy}\left\{D_g'(y,\epsilon)\right\} = -y\,\phi(y) - \epsilon^2\sigma_v^2\left\{y^3\phi(y) - 3y\,\phi(y)\right\} + o(\epsilon^2)$$

and

$$\frac{D_g'(y,\epsilon)}{D_g(y,\epsilon)} = \frac{-y\,\phi(y) - \epsilon^2\sigma_v^2(y^3\phi(y) - 3y\,\phi(y)) + o(\epsilon^2)}{\phi(y) + \epsilon^2\sigma_v^2(y^2\phi(y) - \phi(y)) + o(\epsilon^2)}$$

$$= -y - \epsilon^2\sigma_v^2 y + o(\epsilon^2)$$

Let us now expand our necessary condition about $\epsilon = 0$. We obtain

$$2k^2\left(g_0(z) + \epsilon g_1(z) + \frac{\epsilon^2}{2!}g_2(z) + \frac{\epsilon^3}{3!}g_3(z) + \frac{\epsilon^4}{4!}g_4(z) + o(\epsilon^4) - z\right)$$

$$= \epsilon^2\sigma_v^2\int \frac{\phi}{\epsilon\sigma_v}\left\{\frac{y-z}{\epsilon\sigma_v}\right\}\left\{-y - \epsilon^2\sigma_v^2 y + o(\epsilon^2)\right\}\left\{2\left(\frac{y-z}{\epsilon\sigma_v}\right)^2 - \left\{\frac{y + \epsilon^2\sigma_v^2 y + o(\epsilon^2)}{\epsilon\sigma_v}\right\}(y-z) - 2\right\} dy$$

Making the substitution $y = z + \epsilon\sigma_v s$, we obtain for the RHS above:-

$$-\epsilon^2\sigma_v^2 \int \phi(s) \left\{(z+\epsilon\sigma_v s) + \epsilon^2\sigma_v^2 (z+\epsilon\sigma_v s)\right\} \left\{ 2s^2 - \left((z+\epsilon\sigma_v s) + \epsilon^2\sigma_v^2 (z+\epsilon\sigma_v s)\right).s\epsilon\sigma_v - 2 \right\} ds + o(\epsilon^4)$$

$$\triangleq -\epsilon^2\sigma_v^2 F(z,\epsilon)$$

Thus function $F(z,\epsilon)$ can be expanded as a function of ϵ. We obtain:-

$$F(z,\epsilon) = 0 + 0 + \epsilon^2\sigma_v^2 z$$

The necessary condition can now be expressed as:-

$$2k^2\left(g_0(z) + \epsilon g_1(z) + \frac{\epsilon^2}{2!} g_2(z) + \frac{\epsilon^3}{3!} g_3(z) + \frac{\epsilon^4}{4!} g_4(z) - z + o(\epsilon^4)\right)$$

$$= -\epsilon^4\sigma_v^4 z$$

By comparing coefficients we obtain:-

$$g_0(z) = z$$

$$g_1(z) = g_2(z) = g_3(z) = 0$$

$$g_4(z) = \frac{-12\sigma_v^4 z}{k^2}$$

APPENDIX 5

Analysis of the Stationary Problem

		Page
A.5.1	Some Popular Results for Stationary Problems	A5.1
	A.5.1.1 The Discount Problem	A5.1
	A.5.1.2 The Average Return per Period Problem	A5.15
A.5.2	Linear System, Quadratic Cost Stationary Problem	A5.27
	A.5.2.1 The Discount Problem	A5.29
	A.5.2.2 The Average Cost per Period Problem	A5.33
A.5.3	Approximate Optimal Linear Controller for Zero-Memory Discount Problem	A5.37

A.5.1 Some Popular Results for Stationary Problems

In this section we shall present some well known results regarding the optimal control policies of a stationary problem. These results will provide the framework for a discussion of the zero memory example in the next section. Various types of stationary problem have been examined in the literature. These can be broadly categorised as follows:-

i) the 'discount' problem

ii) the 'average cost per period' problem

We shall summarise each of these problems in turn. In each case, we shall discuss the general, perfectly observable version of the stochastic control problem.

A.5.1.1 The Discount Problem

A.5.1.1.1 Problem Statement

Consider the following process equation:-

$$x_{k+1} = f(x_k, u_k, w_k) \quad k = 0,1,2,$$

where $\{w_k\}$ is a sequence of identically distributed random variables. Given an initial state x_0, the problem is to find a sequence of control functions $\Pi_0 = \{\gamma_k \mid k = 0,1,2,\cdots\}$ which minimise the cost functional:-

$$V_0(\Pi_0, x_0) = \lim_{N \to \infty} \underset{\{w_k\}}{E} \left\{ \sum_{k=0}^{N-1} \rho^k L(x_k, \gamma_k(x_k)) \right\} ,$$

where the expectation is simply with respect to the primitive random variables $\left\{ w_k \,\middle|\, k = 0 \text{ --- } N-2 \right\}$.

Notice that in this case the control values u_k are functions of the current state (or observation) x_k. The quantity ρ is the so-called discount factor, which satisfies $0 < \rho < 1$. The discount factor is introduced primarily as a mathematical convenience to guarantee the finiteness of the cost functional. Nevertheless, the discount factor can have a useful economic interpretation when the cost of system operation is expressed in money terms; for example, if α is the rate of inflation and $\rho = \frac{1}{1+\alpha}$, then the cost functional represents the net-present value of the total future costs.

An alternative formulation of the discount problem can be obtained by interchanging the expectation and limit operations in the definition of the cost functional. Mathematically, this represents a much more delicate problem but under fairly general conditions the problems are equivalent.

Suppose that the optimal cost $V_0^{\circ}(x_0)$ is defined as follows:-

$$V_0^{\circ}(x_0) = \inf_{\Gamma_0} V_0(\Gamma_0, x_0) \text{ for all } x_0.$$

In this definition the infimum is with respect to all sequences of control functions $\Gamma_0 = \left\{ \gamma_k \,\middle|\, k = 0,1,2,\text{---} \right\}$ which are admissable. It is interesting to refine the problem even further and restrict our attention to all sequences of stationary control functions $\Gamma_0 = \left\{ \gamma \,\middle|\, k = 0,1,2,\text{---} \right\}$.

In other words, the same control law is applied at every stage. In case we can redefine the cost functional to obtain:-

$$\bar{V}_0(\gamma, x_0) = \lim_{N \to \infty} \; \underset{\{w_k\}}{E} \left\{ \sum_{k=0}^{N-1} \rho^k L(x_k, \gamma(x_k)) \right\}$$

and $$\bar{V}_0^{\,o}(x_0) = \inf_{\gamma} \bar{V}_0(\gamma, x_0)$$

Intuitively, the problem facing the controller at time 0 is not much different to the problem he faces at time 1 - they both have the same infinite horizon problem to solve (give or take a discount factor). Thus it is not surprising that quite frequently the solution of the discount problem when restricted to consideration of stationary control policies solves the more general problem when the γ_k are allowed to vary with k. When this occurs the benefits are considerable; not only are the control laws easy to implement, they are also easier to find. In what follows we shall call a stationary sequence of control laws optimal if,

$$V_0^{\,o}(x) = \bar{V}_0(\gamma, x) \quad \text{for all } x.$$

In other words, the sequence $\{\gamma\}$ is optimal if we can do no better using a non-stationary sequence of control laws.

Having introduced the discount problem, let us briefly review some of the related results.

A.5.1.1.2 Some Properties of the Optimal Solution

The principle of dynamic programming has already been discussed at length in previous chapters, particularly Chapter 3. All of this preceding work has been concerned with a finite horizon (N-stage) problem. Let us consider a N-stage version of the discount problem; find that sequence of control laws $\Gamma_0^N = \{\gamma_k \mid k = 0,1,---N-1\}$ which minimises the cost functional:-

$$\hat{V}_N(\Gamma_0^N, x_0) \triangleq E_{\{w_k\}} \left\{ \sum_{k=0}^{N-1} \rho^k L(x_k, \gamma_k(x_k)) \right\} ,$$

where it is assumed that the process equation is as defined above in A.5.1.1.1. For this finite horizon version of the problem, a dynamic programming approach yields the following algorithm:-

$$\hat{V}_0^o(\bar{x}_0) = 0 \text{ for all } x$$

$$\hat{V}_k^o(\bar{x}_k) = \inf_{\bar{\gamma}_k(\bar{x}_k)} E_{\bar{w}_k} \left\{ L(\bar{x}_k, \bar{\gamma}_k(\bar{x}_k)) + \rho \hat{V}_{k-1}^o(f(\bar{x}_k, \bar{\gamma}_k(\bar{x}_k), \bar{w}_k)) \right\} ,$$

for all $\bar{x}_k$ and $k = 1,2,---N$. Notice that we have adopted a somewhat different indexing scheme from that employed previously; we are proceeding from lower to higher values of k. In terms of our earlier notation (Chapter 3), we have:-

$$\bar{x}_k = x_{N-k},$$

$$\hat{V}^{o}_{k} = V^{o}_{N-k},$$

$$\bar{\gamma}_k = \gamma_{N-k}, \text{ and}$$

$$\bar{w}_k = w_{N-k}.$$

What is more, inspection of the above equation for $\hat{V}^{o}_{k}$ reveals that the time indices on $\bar{x}_k, \bar{w}_k$, and $\bar{\gamma}_k$ are redundant anyway due to the stationarity of the problem. Thus we can drop the time indices and for convenience we might as well dispense with the 'bar' in the notation. Thus we obtain the revised dynamic programming algorithm:-

$$\hat{V}^{o}_{0}(x) = 0 \quad \text{for all } x$$

$$\text{and } V^{o}_{k}(x) = \inf_{\gamma_k(x)} E_w \left\{ L(x,\gamma(x)) + \rho \hat{V}^{o}_{k-1}(f(x,\gamma(x),w)) \right\}$$

for k = 1,2,---N. We know from our finite horizon dynamic programming arguments that, for any N:-

$$\hat{V}^{o}_{N}(x) = \inf_{\Gamma^N_0} \hat{V}_N(\Gamma^N_0, x_0)$$

As we might expect, when we turn to the infinite horizon version it can be shown that:-

$$V_0^{\circ}(x) = \lim_{N \to \infty} \hat{V}_N^{\circ}(x) = \lim_{N \to \infty} \inf_{\Gamma_0^N} \hat{V}_N(\Gamma_0^N, x)$$

This property is sometimes referred to as the 'Convergence of the Dynamic Programming Algorithm'. It simply tells us that the optimal value function for the infinite horizon discount problem is the limit as N tends to infinity of the optimal value function of the N stage discount problem. Expressed another way, we can interchange the 'lim' and 'inf' operations. Furthermore, under fairly acceptable conditions, it can be shown that convergence is assured whatever function we select for V_0° - it is not necessary that $V_0^{\circ} = 0$.

Of course, it may not be desirable to apply the dynamic programming algorithm ad infinitum to solve the infinite horizon problem discount problem. Instead, we might seek a more direct method based on the asymptotic behaviour of the dynamic programming approach. This behaviour is expressed in the so-called optimality equation - the optimal value function $\hat{V}_0^{\circ}(x)$ satisfies:-

$$V_0^{\circ}(x) = \inf_{\gamma(x)} E_w \left\{ L(x, \gamma(x)) + \rho V_0^{\circ}(f(x, \gamma(x), w)) \right\}$$

Under certain conditions, V_0° is the unique solution of the above functional equation. What is perhaps more surprising, however, is that if $\gamma^{\circ}(x)$ is such that $\gamma^{\circ}(x)$ attains the infimum in the right hand side

of the above equation for each x, then the sequence of stationary control laws $\{\gamma^{\circ}(x)\}$ is optimal. Conversely, if $\{\gamma^{\circ}\}$ is an optimal stationary policy, then $\gamma^{\circ}(x)$ attains the infimum in the right hand side of the optimality equation. Thus we see that the optimality equation gives necessary and sufficient conditions for optimality of a stationary control law sequence.

An interesting corollary of the optimality equation concerns the value of the cost function for a specified stationary control law. Given any admissable stationary policy $\{\gamma\}$, then $\bar{V}_0(\gamma,x)$ is the unique solution for all x of the the equation:-

$$\bar{V}_0(\gamma,x) = \underset{w}{E}\left[L(x,\gamma(x)) + \rho\bar{V}_0(\gamma,f(x,\gamma(x),w))\right]$$

We shall call this equation the stationarity equation.

It should be emphasised that the optimality equation yields an optimal stationary policy provided the infimum in its right hand side is attained. If this is not the case, then we can get as close as we like (within $\epsilon > 0$) to V_0° by using an ϵ-optimal stationary policy.

One further result concerning the convergence of the dynamic programming algorithm is worthy of mention. Starting from any (bounded) V_0°, the rate of convergence is at least as fast as the rate of a convergent geometric progression:-

$$\sup_x \left|\hat{V}_k^{\circ}(x) - V_0^{\circ}(x)\right| \leq \rho^k \sup_x \left|\hat{V}_0^{\circ} - V_0^{\circ}(x)\right|$$

A.5.1.1.3 Some Computational Procedures

Above we have discussed some of the properties of the solution of the infinite discount problem. Here, we shall quickly examine a number of computational techniques which are available for the determination of the optimal control laws. Notice that these techniques are generally numerical and consequently require that the state space and control value space are finite sets. When these sets are not finite it is necessary to replace them with finite sets by means of a discretisation procedure. Having thus restricted ourselves to finite spaces, the problems then become those of controlling a finite state Markov chain. Notationally, and computationally, it then becomes convenient to express our problems and their associated algorithms in matrix form. For completeness, this topic is discussed below.

Matrix Notation

Let the state space S consist of n states denoted by 1,2,---n. That is:-

$$S = \{1,2,---n\},$$

The transition probabilities will be denoted by $p_{ij}(\gamma)$, where

$$p_{ij}(\gamma) = P\left\{ x_{k+1} = j \;\middle|\; x_k = i,\ u_k = \gamma(x_k) \right\} \text{ , for all } i,j \in S.$$

Sometimes the transition probabilities are given directly in the statement of the problem; on other occasions they can be computed from the process equation and the known probability distribution of w_k as we have seen in previous chapters (see Chapter 3).

Let us now examine more closely the basic expression used in the previous section:-

$$\underset{w}{E}\left[L(x_k, \gamma_k(x_k)) + \rho V(f(x_k, \gamma_k(x_k), w_k))\right]$$

$$= L(x_k, \gamma_k(x_k)) + \rho \int V(f(x_k, \gamma_k(x_k), w_k)\ p(w_k)\ dw_k$$

$$= L(x_k, \gamma_k(x_k)) + \rho \int V(x_{k+1})\ p(x_{k+1} \mid x_k, \gamma_k(x_k))\ dx_{k+1}$$

Expressing the relevant value functions as n dimensional vectors, we obtain the following matrix equations:-

i)

$$\hat{V}{}^{o}_{k}(i) = \inf_{\gamma_i} \left\{L(i, \gamma_i) + \rho \sum_j \hat{V}{}^{o}_{k-1}(j)\ p_{ij}(\gamma_i)\right\}$$

$$\underline{\hat{V}{}^{o}_{k}} = \inf_{\gamma} \left\{\underline{L}\ (\gamma) + \rho P(\gamma)\ \hat{V}{}^{o}_{k-1}\right\}$$ -dynamic programming algorithm

---

ii)

$$V_0^{\circ}(i) = \inf_{\gamma_i} \left\{ L(i,\gamma_i) + \rho \sum_j V_0^{\circ}(j)\, p_{ij}(\gamma_i) \right\}$$

$$\underline{V}_0^{\circ} = \inf_{\gamma} \left\{ \underline{L}(\gamma) + \rho P(\gamma)\, \underline{V}_0^{\circ} \right\} \quad \text{- optimality equation}$$

---

iii)

$$\bar{V}_0(\gamma_i, i) = L(i,\gamma_i) + \rho \sum_j \bar{V}_0(j)\, p_{ij}(\gamma_i)$$

$$\underline{\bar{V}}_0(\gamma) = \underline{L}(\gamma) + \rho P(\gamma)\, \underline{\bar{V}}_0(\gamma) \quad \text{- stationarity equation}$$

The above notation now provides a convenient framework for discussing the computational techniques for the solution of the infinite horizon discount problem. In the following discussion, the practice of underlining to denote vectors has been dropped.

The Method of Succesive Approximation

This technique is simply an application of the dynamic programming algorithm. We have, for any $\hat{V}_0^{\circ}$,

$$V_0^{\circ} = \lim_{k \to \infty} \hat{V}_k^{\circ}$$

$$= \lim_{k \to \infty} \inf_{\gamma} \left\{ L(\gamma) + \rho P(\gamma) \hat{V}_{k-1}^{\circ} \right\}$$

Analysis can show that when the space of w_k is finite, this method will yield an optimal policy after a finite number of iterations. Also, a number of refinements to the technique can be made which will improve the convergence of the procedure. In general however the technique only converges in an infinite number of iterations, since it does not restrict its search to stationary policies.

The Method of Policy Iteration

This method sets out to find the optimal stationary policy directly by using the stationarity equation and the following result.

$$\text{If } L(\gamma^{\circ}) + \rho P(\gamma^{\circ}) \bar{V}_0(\gamma') = \min_{\gamma} \left\{ L(\gamma) + \rho P(\gamma) \bar{V}_0(\gamma') \right\},$$

$$\text{then } \bar{V}_0(\gamma^{\circ}) \leq \bar{V}_0(\gamma') \text{ for any } \gamma'.$$

What is more, if γ° is not the optimal stationary policy then strict inequality will hold above for at least one state i. The policy **iteration**

algorithm is as follows:-

a) Guess an initial stationary policy γ^0.

b) For k = 0,1,-----, compute the value function $\bar{V}_0(\gamma^k)$. This can be achieved using the stationarity equation or by using a technique similar to the method of successive approximation, namely

$$\bar{V}_0(\gamma^k) = \lim_{t \to \infty} \tilde{V}_t(\gamma^k)$$

$$= \lim_{t \to \infty} \left[L(\gamma^k) + \rho \tilde{V}_{t-1}(\gamma^k) \right]$$

where $\tilde{V}_0$ is any starting vector.

c) Obtain a new stationary policy γ^{k+1} by finding that γ^{k+1} which satisfies:-

$$L(\gamma^{k+1}) + \rho P(\gamma^{k+1}) \bar{V}_0(\gamma^k) = \min_{\gamma} \left[L(\gamma) + \rho P(\gamma) \bar{V}_0(\gamma^k) \right]$$

d) Increment k by one, and return to step (b).

The procedure is terminated at the end of step (b) if it is found that $\bar{V}_0(\gamma^k) = \bar{V}_0(\gamma^{k+1})$. Since the collection of all possible stationary policies is a finite collection (since the state space and control value space are both finite), and since the method generates an improved policy at every iteration, it follows that the algorithm will always find an optimal policy in a finite number of iterations (of k). If the state space is of small dimension then finding $\bar{V}_0(\gamma^k)$ by solving the stationary equation is quite feasible, it involves the solution of a system of linear equations:-

$$\bar{V}_0(\gamma^k) = \left[I - \rho P(\gamma^k)\right]^{-1} L(\gamma^k)$$

When the dimension of the state space is very large, finding a solution of the above equation may not be so attractive and we might have to utilise the proposed iterative (on t) technique to find $\bar{V}_0(\gamma^k)$. Since we are now iterating on both t and k, the whole technique loses some of its elegance and appeal.

The Method of Linear Programming[1]

Another technique which will provide the optimal solution in a finite number of steps is based on a formulation of the problem in a linear programming framework.

Consider the optimality equation. Suppose γ is any policy, not necessarily optimal. Then

$$V_0^{\circ}(i) \geq L(i,\gamma_i) + \rho \sum_j V_0^{\circ}(i)\, p_{ij}(\gamma_i)$$

There will be one such inequality for each possible value of γ_i, of which let us suppose there are m values. Equality will be achieved for the optimal value of γ_i. Additionally there will be one such group of inequalities for each state i, giving a total of n x m inequalities.

Consider now the following linear programming problem:- maximise

$$\sum_{i=1}^{n} \lambda_i,$$

subject to the constraints $\lambda_i \geq 0$, for i = 1---n

1. See for example Bertsekas Page 248 (8)

and $\lambda_i \leq L(i, \gamma_i^k) + \rho \sum p_{ij}(\mu^k) \lambda_j$, for i=1---n, k=1---m

For those constraints in the basis matrix of the solution to the linear programming problem, strict equality in the constraints will hold. For these constraints therefore the corresponding γ_i^k and λ_i are solutions to the optimality equation.

Notice that the linear programming problem has n variables - $(\lambda_i \mid i=1---n)$ and n x m constraints. As n increases the technique will become impractical. Some relief can be obtained however by solving the dual linear programming problem. We then have a problem with n x m variables and n constraints, which is more manageable computationally. To the authors knowledge this possibility has not been fully explored. There is a strong suspicion that the dual problem will look very much like the policy iteration method. A general description of Linear Programming can be found in Hadley, cited below.[1]

This completes our discussion of the discounted infinite horizon problem. Before examining another type of stationary problem, we shall mention that the discount factor is introduced as a mathematical convenience to ensure convergence. Occasionally this can be guaranteed by other properties of the problem and we can study the undiscounted infinte horizon problem. Very similar results to those discussed above will be obtained. Next let us examine a second class of stationary problem - the so-called average return per period problem.

1. G. Hadley Linear Programming **Addison** Wesley 1961 **(22)**

A.5.1.2 The Average Return per Period Problem

A.5.1.2.1 Problem Statement

As in the previous section, we consider the process equation:-

$$x_{k+1} = f(x_k, u_k, w_k), \quad k = 0,1,2....$$

where $\{w_k\}$ is a sequence of identically distributed random variables. Given an initial state x_0, the problem is to find a sequence of control functions $\Gamma_0 = \{\gamma_k \mid k = 0,1,2---\}$ which minimise the cost functional:-

$$V_0(\Gamma_0, x_0) = \lim_{N \to \infty} \frac{1}{N} E_{\{w_k\}} \left\{ \sum_{k=0}^{N-1} L(x_k, \gamma_k(x_k)) \right\} ,$$

where the expectation is simply with respect to the primitive random variables $\{w_k \mid k = 0,1,---N-2\}$.

Clearly the quantity $V_0(\Gamma_0, x_0)$ represents some kind of average cost of operation per period. Compare this with the Discount Problem of the previous section. To have a meaningful Discount Problem, the average return per period must be zero (at present prices). If the discount rate ρ is unity however, it might turn out that the long term expected total cost is infinite. Nevertheless, it might well be that the average cost per period $V_0(\Gamma_0, x_0)$ as defined above is finite for some Γ_0 and we can still formulate a minimisation problem

In the previous section, a matrix notation was introduced to describe the finite state Markov chains which are encountered when computational

procedures are examined. In this section, we shall utilise this notation from the outset; in other words, the problems themselves will be finite state Markov chains rather than the quantised versions we introduce for computational purposes. This is principally because the finite state case is where the study of the average cost per period problem originates (see Howard[1]). It is also the case about which most is known.

A.5.1.2.1 Some Properties of the Optimal Solution

As in the previous section, the discussion which follows is purely descriptive. The aim is simply to report the main results and provide background material for the later work on the stationary zero memory problem.

Let us first make some broad remarks about our problem. We have previously defined the transition probability matrix $P(\gamma)$ as follows:-

$$\left[P(\gamma)\right]_{ij} = p_{ij}(\gamma) = p\left(x_{k+1} = j \,\middle|\, x_k = i, u_k = \gamma_k(x_k)\right)$$

Next let us compute:-

$$p(x_{k+2} = j \mid x_k = i,\ u_k = \gamma_k(x_k))$$

$$= \sum_k p(x_{k+2} = j \mid x_{k+1} = k,\ u_{k+1} = \gamma_{k+1}(x_{k+1}))\, p(x_{k+1} = k \mid x_k = i, u_k = \gamma_k(x_k))$$

But we are studying a stationary system, hence the time subscript k is irrelevant. Also, we shall assume a stationary control law sequence $\{\gamma\}$.

1. R. Howard – ' Dynamic Programming and Markov Processes' MIT Press 1960 (27)

Thus:-

$$p(x_{k+2}=j \mid x_k=i,\ u_k=\gamma_k(x_k)) = \sum_k p_{ik}(\gamma)\, p_{kj}(\gamma)$$

$$= \left[P(\gamma).\ P(\gamma)\right]_{ij}$$

$$= \left[P^2(\gamma)\right]_{ij}$$

By extending the above argument we can show that:-

$$p(x_n=j \mid x_0=i,\ u_0=\gamma_0(x_0)) = \left[P^n(\gamma)\right]_{ij}$$

Suppose we define a vector of probabilities of state occupancy at time n as follows:-

$$p_n = \begin{bmatrix} p(x_n = 1) \\ p(x_n = 2) \\ p(x_n = n) \end{bmatrix}$$

Then clearly,

$$p_n = p_0\, P^n(\gamma).$$

Depending on the structure of the Markov chain, the matrix $P^n(\gamma)$ has some very interesting properties as $n \to \infty$. For a very broad class of problem, it turns out that p_n as n tends to infinity is independent of the initial probability density p_0. This implies that $\lim_{n \to \infty} P^n(\gamma)$ has identical rows; in fact, each row of this matrix is denoted by a

row vector p^*_γ, where p^*_γ is a solution of the equation

$$p^*_\gamma = p^*_\gamma P(\gamma)$$

we shall call p^* the stationary distribution of the state.

For more general Markov chains it is usual to study the behaviour of $\frac{1}{n}\sum_{i=0}^{n} P^i(\gamma)$ rather than P^n as n tends to infinity. For our purposes, we shall assume that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n} p^i(\gamma) = \lim_{n\to\infty} P^n(\gamma) = P^*(\gamma)$$

Let us now examine the average cost per period in more detail. Suppose we restrict ourselves to a stationary sequence of control laws $\{\gamma\}$, and let us introduce the notation $\bar{V}_0(\gamma, x_0)$, where

$$\bar{V}_0(\gamma, x_0) = V_0(\Gamma_0, x_0) = \lim_{N\to\infty} \frac{1}{N} \mathop{E}_{\{w_k\}} \left[\sum_{k=0}^{N-1} L(x_k, \gamma_k(x_k)\right]$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \int L(x_k, \gamma_k(x_k)) p(x_k \mid x_0, \gamma)\, dx_k,$$

by a similar argument to that used for the discount problem. Introducing a vector notation, we obtain

$$\bar{V}_0(\gamma) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k(\gamma)\, L(\gamma)$$

$$= P^*(\gamma)\, L(\gamma)$$

Consider for a moment what the above expression means; for the class of problem we consider here, the average cost per period (where the average is over time) is equal to the mathematical expectation with respect to the stationary distribution p^*; that is:-

$$\bar{V}_0(\gamma) = P^*(\gamma)L(\gamma) = \begin{bmatrix} p^*_{\gamma}, L(\gamma) \\ p^*_{\gamma}, L(\gamma) \\ | \\ | \\ p^*_{\gamma}, L(\gamma) \end{bmatrix}$$

$$= E^* \left[L(\gamma) \right] ,$$

where the notation E^* denotes expectation with respect to the stationary distribution p^*. Notice that each component of the cost $\bar{V}_0(\gamma)$ is identical: that is, the average cost per period does not depend on the initial state x_0.

We can now derive an equation for the cost $\bar{V}_0(\gamma)$ which is analogous to the stationarity equation obtained for the discount problem.

Consider the vector $q(\gamma)$ defined as follows:-

$$q(\gamma) = \left[I - P(\gamma) + P^*(\gamma) \right]^{-1} \left[I - P^*(\gamma) \right] L(\gamma).$$

By utilising the special properties of $P^*(\gamma)$ and $P(\gamma)$, we can express the cost $\bar{V}_0(\gamma)$ as follows:-

$$\bar{V}_0(\gamma) + q(\gamma) = L(\gamma) + P(\gamma)\, q(\gamma)$$

Again, we shall call this equation the stationarity equation. It is satisfied by every admissable stationary control sequence $\{\gamma\}$. Notice that for the discount problem, the corresponding equation is:-

$$q(\gamma) = L(\gamma) + \rho P(\gamma)\, q(\gamma).$$

For the discount problem, the average cost per stage $\bar{V}_0(\gamma)$ is zero.

As it stands the equation:-

$$\bar{V}_0(\gamma) + q(\gamma) = L(\gamma) + P(\gamma) q(\gamma)$$

is of little use in calculating $\bar{V}_0(\gamma)$ for a particular stationary policy γ. We have already remarked that for a wide class of problem $\bar{V}_0(\gamma)$ is independent of the initial state. Thus we can denote $\bar{V}_0(\gamma)$ by a scalar $c(\gamma)$ multiplied by a unit vector - we obtain -

$$c(\gamma)e + q(\gamma) = L(\gamma) + P(\gamma)\, q(\gamma)$$

The quantity $q(\gamma)$ does depend on the initial condition however. Thus we have a system of n linear equation (n is the number of states) but with (n + 1) unknowns $(c, q_1, q_2 \cdots q_n)$. The difficulty can be resolved by requiring that $q_s = 0$ for some s. Under certain assumptions, this will ensure that $c(\gamma)$ is unique and represents the quantity we are interested in - the average cost per period. The justification for putting $q_s = 0$ for some s is provided by the easily verified observation that, if

c, q solve the stationarity equation, then so does $c, q + e\lambda$ where λ is any scalar. The solution with $q_s = 0$ for some s is sometimes called the normalised cost.

Having introduced the stationarity equation, let us next state an analogue for the optimality equation. If all states in the Markov chain can be reached (eventually) from all others, then there exists a constant c and a vector q such that,

$$c\, e + q = \min_{\gamma} \left\{ L(\gamma) + P(\gamma)\, q \right\}.$$

Furthermore the optimal value of the cost functional $\bar{V}_0(\Gamma_0, x_0)$ is equal to c for every x_0, that is

$$c = \inf_{\Gamma_0} V_0(\Gamma_0, x_0) \text{ for all } x_0.$$

In addition, if $\gamma^\circ(x_i)$ attains the minimum in the right hand side of the optimality equation for every x_i, then the stationary policy $\Gamma_0 = (\gamma^\circ, \gamma^\circ, \text{---})$ is optimal.

Finally, it is interesting to establish a connection between the average cost per period problem and the discount problem.

Suppose $\Gamma_0 = \left\{ \gamma_1, \gamma_2, \text{---} \right\}$, and

$$V_0^\rho(\Gamma_0, x_0) = \lim_{N \to \infty} \underset{\{w_k\}}{E} \left\{ \sum^{k} \rho^k L_k(x_k, \gamma_k(x_k)) \right\}$$

Also, suppose

$$\hat{V}{}_0^{\rho}(x_0) = \inf_{\Gamma_0} V_0^{\rho}(\Gamma, x_0).$$

In other words $\hat{V}{}_0^{\rho}(x_0)$ is the optimal cost for the discount problem. Then the scalar c satisfying the optimality equation above is given by:-

$$c = \lim_{\rho \to 1} (1-\rho)\, V_0^{\rho}(x_0) \text{ for any } x_0$$

What is more, any policy γ which minimises $c(\gamma)$ also minimises $V_0^{\rho}(x_0)$ near $\rho = 1$.

A.5.1.2.3 Some Computational Procedures

As we should expect, it turns out that the computational techniques of the discount problem can be applied with some modification to the average cost per period problem. We shall briefly state two procedures; the method of successive approximation, and the method of policy iteration.

The Method of Successive Approximation

The successive approximation method for the average cost per period problem is nothing more than a slight reformulation of the corresponding algorithm for the discount problem. All that we do is put $\rho = 1$ and normalise the costs.

Again, we do not restrict our attention to stationary policies. All that we require is that there exists an $\epsilon > 0$ and a positive integer m

such that for some state s and all admissable policies $\Gamma_0 = \{\gamma_0, \gamma_1, \cdots\}$, we have

$$\left[P(\gamma_m).P(\gamma_{m-1})\cdots.P(\gamma_0)\right]_{is} \geq \epsilon > 0 \text{ for all } i,$$

where $[\quad]_{is}$ denotes the element of the i the row and s^{th} of the matrix $[\quad]$. In other words, whatever policy we select, we could reach state s from any other starting state i.

Consider now the following algorithm:-

$$H_{k+1} = \min_{\gamma} \left\{L(\gamma) + P(\gamma)\, q_k\right\} \quad ; k = 0,1,2\cdots$$

$$q_{k+1}(i) = H_{k+1}(i) - H_{k+1}(s). \text{ for all } i;\ k = 0,1,2\cdots$$

where q_0 is any function. The notation $q_{k+1}(i)$ is used to denote the i^{th} component of the vector q_{k+1}. As k increases we will have:-

$$q(i) = \lim_{k\to\infty} q_k(i),$$

and $c+q(i) = \lim_{k\to\infty} H_k(i).$

Clearly,

$$c = \lim_{k\to\infty} H_k(s) = \inf_{\Gamma_0} \bar{V}_0(\Gamma_0, x_0) \text{ for all } i.$$

Also the c and q which we compute are solutions of the optimality equation:-

$$ce + q = \min_{\gamma} \left\{ L(\gamma) + P(\gamma)\, q \right\} .$$

The above is the successive approximation algorithm for the average cost per period problem. A number of refinements are available to improve the performance of the algorithm but these will not be discussed here. Next let us examine the policy iteration method.

The Method of Policy Iteration

Again, this method is quite similar to the method discussed earlier for the discount problem. Given a stationary policy, one obtains an improved policy by means of a minimisation process until no further improvement is possible. If we have an irreducible Markov chain, then the technique will work. The algorithm is as follows:-

a) Guess an initial stationary policy $\Gamma_0^{\circ} = \left\{ \gamma^{\circ}, \gamma^{\circ}, \gamma^{\circ}, \text{----} \right\}$

b) For k = 0,1,2,---, compute the average cost per stage $c(\gamma^k)$ by solving the stationarity equation:-

$$c(\gamma^k)\, e + q(\gamma^k) = L(\gamma^k) + P(\gamma^k)\, q(\gamma^k),$$

with $q(\gamma^k)_s = 0$

c) Obtain a new stationary policy γ^{k+1} by finding that γ^{k+1} which satisfies:-

$$L(\gamma^{k+1}) + P(\gamma^{k+1})\, q(\gamma^k) = \min_{\gamma} \left\{ L(\gamma) + P(\gamma)\, q(\gamma^k) \right\}$$

Any γ^{k+1} which satisfies the above equation will give rise to a new set of values for $c(\gamma^{k+1})$ and $g(\gamma^{k+1})$ by solving the stationary equation. It can be shown that

$$c(\gamma^{k+1}) \leq c(\gamma^{k}).$$

d) Increment k by 1 and return to step (b).

The procedure is terminated at the end of step (b) if it is found that:-

$$c(\gamma^{k+1}) = c(\gamma^{k}),$$

and the optimal stationary policy will be:-

$$\gamma^{*} = \gamma^{k}$$

Again convergence in a finite number of steps is guaranteed because the set of all stationary policies (for finite state space) is finite.

This concludes our general discussion of the stationary problem. Our treatment has been primarily concerned with discrete (and finite) state systems which are perfectly observable. As we remarked earlier, this is because the study of the stationary problem, particularly the average cost per period problem, originated with this type of system, and it is the type of system about which most is known. When we extend our interest to infinite state or continuous state systems, the problem generally becomes considerable more complicated. The exception to this

statement is provided by the special case of a linear system with quadratic cost and gaussian noise. Results for this class of problem are well known, both for the perfectly observable and classical information pattern versions of the problem. To the author's knowledge little has been reported regarding any non-classical information pattern version of the problem. This is a problem which is considered in Section 7.2. In appendix 5.2 we review the traditional theory of the linear system, quadratic cost stationary problem.

A.5.2 Linear System: Quadratic Cost Stationary Problem

In the preceding section we considered a fairly general formulation of the stationary problem. The process equation was non-linear and the nature of the cost function was not specified. The results which were described are appropriate however only when quite severe restrictions are imposed on the cost function and on the cardinality of the state space, control space, and disturbance space. Generally speaking the analysis of infinite horizon optimisation problems involving an infinite state space presents considerable difficulties and there does not appear to exist much in the way of a complete and powerful theory. However, certain special cases can be satisfactorily analysed and one such case is the linear system with quadratic cost. We have already encountered this type of problem in previous chapters. At the risk of some repetition let us **emphasise** why this class of problem is so important:-

i) It naturally represents many engineering problems.

ii) It can be analysed.

iii) Its solutions can be easily implemented.

As in the previous section, our discussion will be purely descriptive. Also it will be divided into an examination of the linear-quadratic discount problem and then an examination of the linear-quadratic cost per period problem. The whole subject area is characterised by an abundance of fine points. It is beyond the scope of this discussion even

to mention these matters, let alone clarify them. We are merely concerned with providing background material for Section 2.

Before proceeding with our descriptions of the discount problem and the average cost per period problem, there is one important result which is common to a treatment of both problems. It concerns the asymptotic behaviour of the discrete-time Riccati equation which we have already encountered in Chapter 3. Let A be an (n x n) matrix, B an (n x m) matrix, Q an (n x n) symmetric positive definite matrix, and R an (m x m) symmetric positive definite matrix. Assume that (A,B) is controllable and that Q may be written as CC^T where (A,C^T) is observable. Consider the discrete time matrix Riccati equation,

$$P_{k+1} = A^T \left[P_k - P_k B(B^T P_k B + R)^{-1} B^T P_k \right] A + Q$$

for k = 0,1,2---, and where P_0 is an arbitrary positive semi-definite matrix. Then, there exists a positive definite matrix P such that:-

$$\lim_{k \to \infty} P_k = P,$$

where P is the unique solution of the algebraic matrix Riccati equation:-

$$P = A^T \left[P - PB(B^T PB + R)^{-1} B^T P^T \right] A + Q$$

Furthermore if $D \triangleq A + BL$, where

and $$L \triangleq -(B^T PB + R)^{-1} B^T PA, \text{ then:-}$$

$$\lim_{k\to\infty} L^k x = 0 \quad \text{for every } x \in R^n.$$

This last result implies stability of any resulting optimal stationary control system we might derive from the Riccati equation. The notions of controllability and observability have been introduced above, although they will be of little concern to us in Section 7.2 since we shall be dealing with a scalar system.

Having introduced the algebraic matrix Riccati equation, we can now proceed to discuss the discount problem and the average cost per period problem.

A5.2.1 The Discount Problem

Consider the following process equation:-

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad k = 0,1,.....,$$

where A and B are as specified above. Also, consider the cost function

$$V_0(\Gamma_0, x_0) = \lim_{N\to\infty} E_{\{w_k\}} \left\{ \sum_{k=0}^{N-1} \rho^k \left\{ (x^T_k Q x_k) + \gamma^T_k(x_k) R \gamma_k(x_k) \right\} \right\},$$

where Q and R are as specified above. The problem is then to minimise V_0 with respect to the sequence of control laws $\Gamma_0 = (\gamma_0, \gamma_1, ---)$:-

$$V^0_0(x_0) = \inf_{\Gamma_0} V_0(\Gamma_0, x_0).$$

The analysis proceeds as follows:-

i) First of all, employ the dynamic programming algorithm described in the preceding section to derive the functions

$$\left\{ \hat{V}_k^{\,o}(x) \;\middle|\; k = 0,1,\text{---} \right\}, \text{ where } V_0^{\,o}(x) = 0. \text{ It turns out that:-}$$

$$\hat{V}_k^{\,o}(x) = x^T P_k x + \sum_{m=0}^{k-1} \rho^{k-m} E_w \left\{ w^T P_m w \right\},$$

where the matrices $\{P_k\}$ are given recursively by:-

$$P_{k+1} = A^T \left[\rho P_k - \rho^2 P_k B (B^T P_k B + R)^{-1} B^T P_k \right] A + Q.$$

ii) If we define $\bar{R}$ as R/ρ and $\bar{A}$ as $\sqrt{\rho}\, A$, we can write the preceding equation as:-

$$P_{k+1} = \bar{A}^T \left[P_k - P_k B (B^T P B + \bar{R})^{-1} B^T P_k \right] \bar{A} + Q$$

Assuming the right properties for $\bar{A}$ (or A), B, and Q we can use our result concerning the algebraic matrix Riccati equation to assert that

$$\lim_{k \to \infty} P_k = P,$$

where,

$$P = A^T \left[\rho P - \rho^2 PB (B^T PB + R)^{-1} B^T P^T \right] + Q$$

Notice that P is the unique solution of this equation within the class of positive semi-definite symmetric matrices.

iii) The implication of the above analysis is that

$$\hat{V}_{\infty}^{\circ}(x) \triangleq \lim_{k \to \infty} V_k^{\circ}(x) = x^T P x + c,$$

where $$c = \lim_{k \to \infty} \sum_{m=0}^{k-1} \rho^{k-m} E_w \left\{ w^T P_m w^T \right\}$$

$$= \left(\frac{\rho}{1-\rho} \right) E_w \left\{ w^T P w \right\}$$

iv) Having derived the pointwise limit function $\hat{V}_{\infty}^{\circ}(x)$, we can show that it satisfies the optimality equation:-

$$\hat{V}_{\infty}^{\circ}(x) = \min_{\gamma} E_w \left[x^T Q x + \gamma^T(x) R \gamma(x) + \rho V_{\infty}^{\circ}(Ax + B\gamma(x) + w) \right]$$

and for our problem this is sufficient to ensure that

$$\hat{V}_{\infty}^{\circ}(x) = V_0^{\circ}(x)$$

v) Finally, for our problem, the minimising $\gamma^*(x)$ used to satisfy the optimality equation in (iv) above is also the optimal stationary control law $\gamma^{\circ}(x)$, since we were able to replace 'inf' with 'min'. This ensures that γ° satisfies both the stationarity equation and the optimality equation, which is a necessary and sufficient condition for the optimality of γ°.

The resulting γ° is of the form:-

$$\gamma^{\circ}(x) = -\rho(\rho B^T P B + R)^{-1} B^T P A x$$

Its linearity and stationarity make it extremely attractive for engineering applications. The result can, moreover be extended to the classical information pattern problem. The analysis can proceed in precisely the same manner as that presented above for the perfectly observable case. All that is different is that the perfectly observable dynamic programming algorithm is replaced by the one corresponding to the classical information pattern case (with a state variable $p(x_k \mid I_k)$ - see Chapter 3). For a linear system with quadratic costs and gaussian noise, we have seen how the classical information pattern dynamic programming algorithm can be reduced to one involving a pair of sufficient statistics, the conditional mean and variance. The structure of this problem is identical to that for the perfectly observable case. And we can use the same analytical arguments to derive the optimal solutions; in other words,

$$\gamma^{o}(\hat{x}) = -\rho(\rho B^{T}PB + R)^{-1}B^{T}PA\,\hat{x},$$

where $\hat{x}$ is the conditional expectation of the state. Note that $\hat{x}$ is the expected value of the state given all the information available. In an infinite horizon problem, this is a lot of information. Clearly the above result only makes sense if $\hat{x}$ is generated by a Kalman fitter type of equation:-

$$\hat{x}_{k+1} = A\hat{x}_k + B\gamma_k(\hat{x}_k) + V_k, \qquad k = 0,1,\text{----}$$

where x_0 is easily calculated from the initial observation y_0 and the a priori distribution of x_0. The variable V_k in the above (perfectly observable) state transition equation. A discussion of the extension of

the finite time classical information pattern problem to a discounted infinite horizon version is provided by

Let us now examine the linear quadratic average cost per period problem. We shall see that results can be obtained just as easily as they were for the discount problem.

A.5.2.2 The Average Cost per Period Problem

In this case we again consider the system,

$$x_{k+1} = Ax_k + Bu_k + w_k \ , \quad k = 0,1,\text{----}.$$

but this time with the cost functional

$$V(x_0, \Gamma_0) = \lim_{N \to \infty} \frac{1}{N} \underset{\{w_k\}}{E} \left[\sum_{k=0}^{N-1} x_k^T Q x_k + \gamma_k(x_k)^T R \gamma_k(x_k) \right]$$

Let us define $V^0(x_0)$ as follows:-

$$V^0(x_0) = \inf_{\Gamma_0} V(x_0, \gamma_0)$$

Again we must assume the appropriate properties for A,B, and Q. Let us first examine the N-stage problem. The optimal solution can be found via a dynamic programming approach

Suppose

$$\bar{V}_N(\Gamma_0^N, x_0) = \frac{1}{N} \underset{\{w_k\}}{E} \left[\sum_{k=0}^{N-1} x_k^T Q x_k + \gamma_k^T(x_k) R \gamma_k(x_k) \right] ,$$

[1] H. Kushner' Introduction to Stochastic Control' Holt, Rinehart and Wilson 1971. (38)

then

$$\bar{V}_N^{\,o}(x_o) = \inf_{\Gamma_0^N} \bar{V}_N(\Gamma_0^N, x_o)$$

$$= \frac{1}{N}\left[x_o^T P_N x_o + \sum_{k=0}^{N-1} E\, w^T R_k w \right],$$

where P_k satisfies the discrete time Riccati equation:-

$$P_{k+1} = \left[A^T P_k - P_k B(B^T P_k + R)^{-1} B^T P_k \right] A + Q$$

$$P_o = 0$$

Now our result concerning the algebraic matrix Riccati equation states that,

$$\lim_{k \to \infty} P_k = P,$$

where P is the unique solution (within the class of positive semi-definite matrices) of the equation:-

$$P = A^T \left[P - P B(B^T P B + R)^{-1} B^T P \right] A + Q$$

Thus

$$\lim_{N \to \infty} \bar{V}_N^{\,o}(x_o) = \lim_{N \to \infty} \left\{ \frac{1}{N} \left[x_o P_N x_o + \sum_{k=0}^{N-1} E(w^T P_k w) \right] \right\}$$

$$= \lim_{N \to \infty} \frac{1}{N} \left[\sum_{k=0}^{N-1} E(w^T P_k w) \right]$$

$$= E(w^T P w)$$

$$\triangleq C$$

In other words, the optimal finite horizon cost tends in the limit to C. Also, the optimal finite horizon optimal policy in its initial stages tends to the stationary policy:-

$$\gamma^\circ(x) = -(B^T P B + R)^{-1} B^T P A x$$

At this point, we have defined the quantities $\gamma^\circ(x)$, C, and P. It is easy to verify that these quantities satisfy the ' optimality ' equation for the average cost per period problem:-

$$C + x^T P x = \min_u E\left[x^T Q x + u^T R u + (Ax + Bu + w) K^T (Ax + Bu = w)\right] ,$$

where the minimum in the right hand side is obtained at $\gamma^\circ(x)$ as specified above. It can be shown that this is indeed an optimality equation. It turns out that:-

$$C \leq V(x_0, \Gamma_0) \text{ for all } x_0, \Gamma_0,$$

with equality if $\Gamma_0 = (\gamma^\circ, \gamma^\circ, \gamma^\circ \text{---})$. A restriction on the optimality of γ° is that γ° is optimal only over all policies Γ_0 for which $E\left[x_N^T P x_N\right]$ is bounded **uniformly over N.**

Notice that the optimal control law is exactly the same as that obtained for the discount problem (with $\rho = 1$). For the case of the classical information pattern version of the problem, we can again reformulate our problem in terms of a perfectly observable sufficient statistic and then employ the same analysis as above to derive the optimal con-

trol laws. Not surprisingly it turns out that,

$$\gamma^{\circ}(\hat{x}) = -(B^T P B + R)^{-1} B^T P A \hat{x}$$

is the optimal control law, where the conditional mean x is generated from the information available using a Kalman fitter.

A.5.3 Approximate Optimal Linear Controller for Zero-Memory Discount Problem

Below we present details of the successive approximation method applied to the problem of finding the best linear control law for the zero-memory problem.

Since the starting functional $V_0^{\circ}(p_0)$ is arbitrary we shall take it as $E_{p_0}(x^2)$.

(i) Evaluation of $V_1^{\circ}(p_1)$

$$V_1^{\circ}(p_1) = \min_{\gamma} \left\{ E_{p_1}\left[x^2 + \gamma_1^2(y_1, p_1)\right] + \rho V_0(p_0) \right\}$$

$$= \min_{\gamma} \left\{ E_{p_1}\left[x^2 + \gamma_1^2(y_1, p_1)\right] + \rho E_{p_1}\left[x^2 + \gamma_1^2(y_1, p_1) + 2x\gamma_1(y_1, p_1)\right] + \rho\sigma_\omega^2 \right.$$

This is minimised when:-

$$\gamma_1(y, p_1) = \frac{\rho}{1+\rho} E_{p_1}\left[x \mid y\right]$$

and the resulting $V_1^{\circ}(p_1)$ is as follows:-

$$V_1^{\circ}(p_1) = (1+\rho) \underset{p_1}{E}\left[x^2\right] - \frac{\rho^2}{1+\rho} \underset{p_1}{E}\left[\left(\underset{p_1}{E}\left[x \mid y\right]\right)^2\right] + \rho\sigma_w^2$$

Given that we are seeking optimal linear control, we know that p_1 is Gaussian and hence $\gamma_1(y,p_1)$ is linear in y as we require. Now for p_1 Gaussian with variance σ_1^2, we have:-

$$\underset{p_1}{E}\left[x \mid y\right] = \frac{\sigma_1^2 y}{\sigma_1^2 + \sigma_v^2}$$

where we have assumed that p_1 has zero mean. Thus for $V_1^{\circ}(p_1)$ we have:-

$$V_1^{\circ}(p_1) = (1+\rho)\sigma_1^2 - \frac{\rho^2}{1+\rho}\left(\frac{\sigma_1^2}{\sigma_1^2+\sigma_v^2}\right)^2(\sigma_1^2+\sigma_v^2) + \rho\sigma_w^2$$

$$\text{or} \quad V_1^{\circ}(p_1) = (1+\rho)\sigma_1^2 - \frac{\rho^2}{1+\rho}\,\frac{\sigma_1^4}{(\sigma_1^2+\sigma_v^2)} + \rho\sigma_w^2$$

(ii) Evaluation of $V_2(p_2)$

It is at this point that the determination of the best linear control law becomes difficult. According to our algorithm, we must evaluate:-

$$V_2^{\circ}(p_2) = \min_{k(p_2)} \left\{ E\left[x^2 + (k(p_2)y)^2\right] + \rho V_1^{\circ}(p_1) \right\}$$

Under linear control, we know the probability density functions $\{p_i\}$ are Gaussian and hence summarised by their variances $\{\sigma_i^2\}$. Let us write $V_2^{\circ}(p_2)$ explicitly:-

$$V_2^{\circ}(p_2) = \min_k \left\{ k^2(1+\sigma_2^2) + k^2\sigma_v^2 + \rho(1+\rho)\sigma_1^2 - \frac{\rho^3}{1+\rho} \frac{\sigma_1^4}{\sigma_1^2+\sigma_v^2} \right\} + \rho^2\sigma_\omega^2$$

where $\sigma_1^2 = (1+k)^2\sigma_2^2 + k^2\sigma_v^2 + \sigma_w^2$

Sunstituting for σ_1^2, we obtain:-

$$V_2^{\circ}(p_2) = \min_k \left\{ k^2(1+\sigma_2^2) + k^2\sigma_v^2 + \rho(1+\rho)\left[(1+k)^2\sigma_2^2 + k^2\sigma_v^2 + \sigma_\omega^2\right] \right.$$

$$\left. \frac{-\rho^3}{(1+\rho)} \frac{\left(\sigma_2^2(1+k)^2 + k^2\sigma_v^2 + \sigma_\omega^2\right)^2}{\sigma_2^2(1+k)^2 + (k^2+1)\sigma_v^2 + \sigma_\omega^2} \right\} + \rho^2\sigma_\omega^2$$

Clearly it is difficult to find an analytic expression for k which provides the minimum of the above expression. Expansion of the above expression will reveal that k is determined by a quartic equation. The same difficulty will be encountered if we attempt to find the best linear control law for the finite time problem (put $\rho = 1$ above). We have to resort to an approximation technique at this point. Suppose we introduce the parameter ϵ into the observation equation; that is:-

$$y = x + \epsilon v$$

Let us now retrace our step somewhat and write $V_1^0(p_1)$ as follows:-

$$V_1^0(p_1) = (1+\rho)\sigma_1^2 - \frac{\rho^2}{1+\rho}\,\frac{\sigma_1^4}{\sigma_1^2 + \epsilon^2\sigma_v^2} + \rho\sigma_\omega^2$$

$$= (1+\rho)\sigma_1^2 - \frac{\rho^2}{1+\rho}\,\sigma_1^2\left\{1 + \epsilon^2\frac{\sigma_v^2}{\sigma_1^2}\right\}^{-1} + \rho\sigma_\omega^2$$

$$= (1+\rho)\sigma_1^2 - \frac{\rho^2}{1+\rho}\,\sigma_1^2\left\{1 - \epsilon^2\frac{\sigma_v^2}{\sigma_1^2} + \epsilon^4\frac{\sigma_v^4}{\sigma_1^4}\right\} + \rho\sigma_\omega^2 + O(\epsilon^4)$$

that is,

$$V_1^{\circ}(p_1) = \left(\frac{1+2\rho}{1+\rho}\right)\sigma_1^2 + \frac{\rho^2}{(1+\rho)}\epsilon^2\sigma_v^2\left\{1 - \frac{\epsilon^2\sigma_v^2}{\sigma_1^2}\right\} + \rho\sigma_\omega^2 + o(\epsilon^4)$$

We now propose to find an $o(\epsilon^4)$ approximation to the optimal linear control law for the discount problem.

Let us first write $V_1^{\circ}(p_1)$ as follows:-

$$V_1^{\circ}(p_1) = a_1\sigma_1^2 + b_1\epsilon^2\sigma_v^2 + c_1\epsilon^4\frac{\sigma_v^4}{\sigma_1^2} + d\sigma_\omega^2 + o(\epsilon^4)$$

where

$$a_1 = \frac{(1+2\rho)}{(1+\rho)}$$

$$b_1 = \frac{\rho^2}{(1+\rho)}$$

$$c_1 = \frac{-\rho^2}{(1+\rho)}$$

$$d_1 = \rho$$

Now

$$V_2^o(p_2) = \min_k \left\{ \sigma_2^2(1+k^2) + k^2\epsilon^2\sigma_v^2 + \rho a_1 \left(\sigma_2^2(1+k)^2 + k^2\epsilon^2\sigma_v^2 + \sigma_w^2\right) \right.$$

$$\left. + \rho b_1 \epsilon^2\sigma_v^2 + \rho c_1 \frac{\epsilon^4\sigma_v^4}{\left(\sigma_2^2(1+k)^2 + k^2\epsilon^2\sigma_v^2 + \sigma_w^2\right)} \right\}$$

Differentiating with respect to k, we obtain the following necessary condition for optimality:-

$$k\left\{1+\rho a_1 - \frac{\rho c_1 \epsilon^4\sigma_v^4}{\left(\sigma_2^2(1+k)^2 + k^2\epsilon^2\sigma_v^2 + \sigma_w^2\right)}\right\}\left(\sigma_2^2 + \epsilon^2\sigma_v^2\right)$$

$$+\rho a_1\sigma_2^2 - \frac{\rho c_1\epsilon^4\sigma_v^4\sigma_2^2}{\left(\sigma_2^2(1+k)^2 + k^2\epsilon^2\sigma_v^2 + \sigma_w^2\right.} + o(\epsilon^4) = 0$$

Suppose that

$$F(k,\epsilon) \triangleq \frac{1}{\left(\sigma_2^2(1+k)^2 + k^2\epsilon^2\sigma_v^2 + \sigma_w^2\right)^2}$$

then

$$F(k,\epsilon)\Big|_{\epsilon=0} = \frac{1}{\left(\sigma_z^2(1+k)^2+\sigma_w^2\right)^2}$$

$$\frac{dF(k,\epsilon)}{d\epsilon}\Big|_{\epsilon=0} = 0$$

and

$$\frac{d^2F(k,\epsilon)}{d\epsilon^2}\Big|_{\epsilon=0} = \frac{-4k^2\sigma_v^2}{\left(\sigma_z^2(1+k)^2+\sigma_w^2\right)^3}$$

Thus

$$F(k,\epsilon) = \frac{1}{\left(\sigma_z^2(1+k)^2+\sigma_w^2\right)^2} - \frac{\epsilon^2}{2!}\frac{4k^2\sigma_v^2}{\left(\sigma_z^2(1+k)^2+\sigma_w^2\right)^3} + o(\epsilon^4)$$

Our necessary condition now becomes:-

$$k(\epsilon)\left\{1+\rho a_1 - \frac{\rho c_1 \epsilon^4 \sigma_v^4}{\left(\sigma_z^2(1+k_0)^2+\sigma_w^2\right)^2} + o(\epsilon^4)\right\}\left(\sigma_z^2+\epsilon^2\sigma_v^2\right)$$

$$+\rho a_1\sigma_z^2 - \frac{\rho c_1\epsilon^4\sigma_v^4}{\left(\sigma_z^2(1+k_0)^2+\sigma_w^2\right)^2} + o(\epsilon^4) = 0$$

where $k_0 = k(\epsilon)\Big|_{\epsilon=0}$. Comparing coefficients, we obtain

$$k_0 = -\frac{\rho a_1}{1+\rho a_1},$$

$$k_1 = 0,$$

$$k_2 = -k_0 \frac{\sigma_v^2}{\sigma_2^2},$$

$$k_3 = 0$$

and

$$k_4 = -k_2 \frac{\sigma_v^2}{\sigma_2^2} - \frac{\rho c_1 \sigma_v^4 (1+\rho a_1)^2}{(\sigma_2^2 + \sigma_w^2 (1+\rho a_1)^2)^2}$$

Thus

$$k(\epsilon) = -\frac{\rho a_1}{(1+\rho a_1)} \left\{ 1 - \epsilon^2 \frac{\sigma_v^2}{\sigma_2^2} + \epsilon^4 \frac{\sigma_v^4}{\sigma_2^2} \right\} + \frac{\rho c_1 \epsilon^4 \sigma_v^4 (1+\rho a_1)^2}{(\sigma_2^2 + \sigma_w^2 (1+\rho a_1)^2)^2} + O(\epsilon^4)$$

Let us now use this approximation for $k(\epsilon)$ to claculate $V_2^{\circ}(p_2)$

$$V_2^{\circ}(p_2) = \sigma_2^2(1+k^2(\epsilon)) + k^2(\epsilon)\epsilon^2\sigma_v^2 + \rho a_1(\sigma_2^2(1+k(\epsilon))^2 + k^2(\epsilon)\epsilon^2\sigma_v^2 + \sigma_w^2) + \rho b_1\epsilon^2\sigma_v^2 + \frac{\rho c_1\epsilon^4\sigma_v^4}{\sigma_2^2(1+k(\epsilon))^2 + k^2(\epsilon)\epsilon^2\sigma_v^2 + \sigma_w^2} + \rho d_1\sigma_w^2 + o(\epsilon^4)$$

$$= \frac{\sigma_2^2(1+2\rho a_1)}{(1+\rho a_1)} + \epsilon^2\sigma_v^2\left\{\rho b_1 + \frac{\rho a_1^2}{1+\rho a_1}\right\}$$

$$-\epsilon^4\sigma_v^4\left\{\frac{\rho^2 a_1^2}{\sigma_2^2(1+\rho a_1)} - \frac{\rho c_1}{(1+k_0)^2 + \frac{\sigma_w^2}{\sigma_2^2}}\right\} + \rho(a_1+d_1)\sigma_w^2 + o(\epsilon^4)$$

If we compare this expression with that which we defined for $V_1^{\circ}(p_1)$, we can see that the term which will give us difficulty is -

$$\frac{\epsilon^4\sigma_v^4\rho c_1}{\sigma_2^2(1+k_0)^2 + \sigma_w^2}$$

since it apparently upsets any plans we have for formulating a convenient algorithm for generating V_n in terms of V_{n-1}. Suppose we write $V_2^{\circ}(p_2)$ as follows:-

$$V_2^0(p_2) = a_2\sigma_2^2 + b_2\epsilon^2\sigma_v^2 + c_2\epsilon^4\frac{\sigma_v^4}{\sigma_2^2} + \frac{f_{21}\epsilon^4\sigma_v^4}{\sigma_2^2+\sigma_w^2 g_{21}} + d_2\sigma_w^2$$

where $a_2 = \dfrac{1+2\rho a_1}{1+\rho a_1}$,

$b_2 = \rho b_1 + \dfrac{\rho^2 a_1^2}{1+\rho a_1}$,

$c_2 = \dfrac{-\rho^2 a_1^2}{1+\rho a_1}$

$d_2 = \rho(a_1+d_1)$,

$f_{21} = (1+\rho a_1)^2\rho c_1$,

$g_{21} = (1+\rho a_1)^2$.

(iii) Evaluation of $V_3^0(p_3^0)$

The next step is to compute $V_3^0(p_3)$. As before we make the substitution $\sigma_3^2 = (1+k(\epsilon))^2\sigma_2^2 + k^2(\epsilon)\epsilon^2\sigma_v^2 + \sigma_w^2$. Repeating the procedure we used above to determine the optimal gain at 'time' 2, we obtain:-

$$k(\epsilon)(\sigma_3^2+\epsilon^2\sigma_v^2)\left\{1+\rho a_1-\frac{\rho c_2\epsilon^4\sigma_v^4}{(\sigma_3^2(1+k(\epsilon))^2+k^2(\epsilon)\sigma_v^2\epsilon^2+\sigma_\omega^2)^2}\right.$$

$$\left.-\frac{\rho f_2\epsilon^4\sigma_v^4}{(\sigma_3^2(1+k(\epsilon))^2+k^2(\epsilon)\epsilon^2\sigma_v^2+\sigma_\omega^2(1+g_2))^2}\right\}+\rho a_1\sigma_3^2$$

$$-\frac{\rho c_2\epsilon^4\sigma_v^4\sigma_3^2}{(\sigma_3^2(1+k(\epsilon))^2+k^2(\epsilon)\epsilon^2\sigma_v^2+\sigma_\omega^2)^2}$$

$$-\frac{\rho f_2\epsilon^4\sigma_v^4\sigma_3^2}{(\sigma_3^2(1+k(\epsilon))^2+k^2(\epsilon)\epsilon^2\sigma_v^2+\sigma_\omega^2(1+g_2))^2}+o(\epsilon^4)$$

The $o(\epsilon^4)$ approximation to k which satisfies this equation is:-

$$k(\epsilon)=\frac{-\rho a_2}{(1+\rho a_2)}\left\{1-\frac{\epsilon^2\sigma_v^2}{\sigma_3^2}+\frac{\epsilon^4\sigma_v^4}{\sigma_3^4}\right\}+\frac{\rho c_2\epsilon^4\sigma_v^4(1+\rho a_2)^2}{(\sigma_3^2+\sigma_\omega^2(1+\rho a_2)^2)^2}$$

$$+\frac{\rho f_2\epsilon^4\sigma_v^4(1+\rho a_2)^2}{(\sigma_3^2+\sigma_\omega^2(1+g_2)(1+\rho a_2)^2)^2}+o(\epsilon^4)$$

and the corresponding $V_3^o(p_3)$ is given by:-

$$V_3^o(p_3)=\left\{\frac{1+2\rho a_2}{1+\rho a_2}\right\}\sigma_3^2+\epsilon^2\sigma_v^2\left\{\rho b_2+\frac{\rho^2 a_2^2}{1+\rho a_2}\right\}-\epsilon^4\sigma_v^4$$

$$+\frac{\rho c_2\epsilon^4\sigma_v^4(1+\rho a_2)^2}{(\sigma_3^2+\sigma_\omega^2(1+\rho a_2)^2)}+\frac{\rho f_2\epsilon^4\sigma_v^4(1+\rho a_2)^2}{(\sigma_3^2+\sigma_\omega^2(1+g_2)(1+\rho a_2)^2)}$$

$$+\rho(a_2+d_2)\sigma_\omega^2+o(\epsilon^4)$$

iv) Evaluation of $V^{o}_{n+1}(p_{n+1})$

At this point we can see a pattern emerging, although not a very attractive one. Suppose we define $V^{o}_{n}(p_n)$ as:-

$$V^{o}_{n}(p_n) = a_n\sigma_n^2 + b_n\epsilon^2\sigma_v^2 + c_n\epsilon^4\frac{\sigma_v^4}{\sigma_n^2} + \sum_{i=1}^{n-1}\left(\frac{\epsilon^4\sigma_v^4 f_{ni}}{\sigma_n^2+\sigma_w^2 g_{ni}}\right) + d_n\sigma_w^2 + o(\epsilon^4)$$

Then:-

$$V^{o}_{n+1}(p_{n+1}) = a_{n+1}\sigma_{n+1}^2 + b_{n+1}\epsilon^2\sigma_v^2 + c_{n+1}\frac{\epsilon^4\sigma_v^4}{\sigma_{n+1}^2} + \sum_{i=1}^{n}\left(\frac{\epsilon^4\sigma_v^4 f_{n+1,i}}{\sigma_{n+1}^2+\sigma_w^2 g_{n+1,i}}\right) + d_{n+1}\sigma_w^2 + o(\epsilon^4)$$

where

$$a_{n+1} = \frac{1+2\rho a_n}{(1+\rho a_n)}$$

$$b_{n+1} = \rho b_n + \frac{\rho^2 a_n^2}{(1+\rho a_n)}$$

$$c_{n+1} = \frac{-\rho^2 a_n}{(1+\rho a_n)}$$

$$f_{n+1,1} = \rho c_n (1+\rho a_n)^2$$

$$f_{n+1,i} = \rho f_{n,i-1} (1+\rho a_n)^2, \quad i = 2, \cdots n$$

$$g_{n+1,1} = (1+\rho a_n)^2$$

$$g_{n+1,i} = (1+g_{n,i-1})(1+\rho a_n)^2, \quad i = 2, \cdots, n$$

$$d_{n+1} = \rho(a_n + d_n)$$

with minimising $k^o(\epsilon)$ given by:-

$$k^o_{n+1}(\epsilon) = \hat{k}_{n+1}\left\{1 - \frac{\epsilon^2\sigma_v^2}{\sigma_{n+1}^2} + \frac{\epsilon^4\sigma_v^4}{\sigma_{n+1}^4}\right\} + \sum_{i=1}^{n} \frac{\epsilon^4\sigma_v^4 f_{n+1,i}}{(\sigma_{n+1}^2 + \sigma_w^2 g_{n+1,i})^2} + o(\epsilon^4)$$

where $$\hat{k}_{n+1} = -\frac{\rho a_n}{1+\rho a_n}$$

The starting values for the above algorithm have been presented earlier; they are:-

$$a_1 = \frac{1+2\rho}{1+\rho} \quad , \quad b_1 = \frac{\rho^2}{1+\rho} \quad , \quad c_1 = -\frac{\rho^2}{1+\rho} \quad , \quad d_1 = \rho$$

It is worthwhile to consider for a moment the structure of the above approximation to the optimal control linear control law for the zero memory discount problem. Note that $E\left[x \mid y\right]$, where $y = x + \epsilon v$ and $p(x)$ is Gaussian with zero mean and variance σ_x^2, is given by

$$E[x/y] = \frac{\sigma_x^2 y}{\sigma_x^2+\sigma_v^2} = y\left\{1 - \epsilon^2\frac{\sigma_v^2}{\sigma_x^2} + \epsilon^4\frac{\sigma_v^4}{\sigma_x^4}\right\} + o(\epsilon^2)$$

Thus the optimal linear control law is of the form

$$k_{n+1}(\epsilon)y_{n+1} = \frac{-\rho a_n}{1+\rho a_n} E[x_{n+1} \mid y_{n+1}] + \sum_{i=1}^{n} \frac{\epsilon^4\sigma_v^4 f_{n+1,i}\, y_{n+1}}{\sigma_{n+1}^2 + \sigma_w^2 g_{n+1,i}} + o(\epsilon^4)$$

Compare this control law with the optimal control law for the corresponding classical information pattern discount problem:-

$$\gamma_{n+1}^{o}(y_{n+1}) = \frac{-\rho a_n}{1+\rho a_n} E\left[x_{n+1} \mid y_0, y_1, y_2, y_{n+1}\ u_0, u_1, \text{---} u_n\right]$$