THE OPTIONAL SAMPLING THEOREM FOR PARTIALLY ORDERED TIME PROCESSES AND MULTIPARAMETER STOCHASTIC CALCULUS

by

Robert Buchanan Washburn, Jr.
B.S., Yale University
1973

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February, 1979

Signature of Author.....
Department of Mathematics, January 5, 1979

Certified by.....
Thesis Supervisor

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

THE OPTIONAL SAMPLING THEOREM FOR PARTIALLY ORDERED TIME PROCESSES AND MULTIPARAMETER STOCHASTIC CALCULUS

by Robert Buchanan Washburn, Jr.

Submitted to the Department of Mathematics on January 5, 1979 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The first part of this thesis investigates the optional sampling theorem for submartingales indexed by a partially ordered time set. Although the optional sampling theorem is true for martingales indexed by partially ordered time under very general circumstances, the submartingale version requires one to restrict the stopping times to a special subclass of general stopping times. These special stopping times, which we call reachable, are defined in terms of a sequential stopping problem in the partially ordered time set. We show that the optional sampling theorem is generally true for submartingales and reachable stopping times. Conversely, if the optional sampling theorem is true for a given pair of stopping times and all submartingales, then these stopping times must satisfy our reachibility condition. Thus, we are able to characterize the stopping times which make the optional sampling theorem true for submartingales.

The second part of this thesis derives a generalization of Ito's stochastic differentiation formula for a class of multiparameter processes including the multiparameter Wiener process. We develop simple stochastic differentiation rules to obtain this formula, and we show that the formula is a natural consequence of the ordinary deterministic multiparameter differentiation formula and the one-parameter Ito stochastic differentiation formula. In the two-parameter case our result agrees with those of Wong-Zakai and Cairoli-Walsh. We then apply the formula to represent multiparameter square integrable martingales which are measurable with respect to the multiparameter Wiener process.

Thesis supervisor: Alan Willsky
Title: Associate Professor of Electrical Engineering

ACKNOWLEDGEMENTS

The author first wishes to express grateful thanks to his supervisor, Professor A. S. Willsky of the Department of Electrical Engineering. His encouragements, guidance, comments and infinite patience were all greatly appreciated throughout the long haul. His prodding finally got me to the end.

The author also wishes to thank Professors R. M. Dudley and H. Chernoff of the Department of Mathematics for their comments and suggestions and for agreeing to be readers of such a massive thesis. In addition, I would like to thank Professors T. G. Kurtz (University of Wisconsin - Madison) and E. Wong (University of California - Berkeley) for providing me with their interesting and valuable papers on optional sampling and stochastic calculus. Thanks also to Dr. Steven Barta (Bell Labs) for many long discussions on decentralized control problems and to Professor S. K. Mitter of the Department of Electrical Engineering for much encouragement and many interesting discussions on our great common interest, statistical mechanics, and on the philosophy of mathematics.

Special thanks go to Ms. Phyllis Ruby for single-handedly typing this monster and to Ms. Ruby and Ms. Nancy Toscano

for kindly helping me through the official requirements of the graduate degree program.

Greatest thanks are due to my parents, whose love and encouragement have supported me through many years of school, and to Laura who was with me the last and longest mile.

Finally, the author gratefully acknowledges the financial support from the Department of Mathematics of the Massachusetts Institute of Technology.

CONTENTS

INTRODUCTION

1. Background for Random Fields

A stochastic process is a collection of random variables indexed by a single real or integer parameter. That is, the parameter space is a subset of the real line and one usually thinks of the parameter as a time. However, in some applications it is more appropriate to consider collections of random variables indexed by more general parameter sets. Such a collection of random variables with a general parameter set is called a random field and sometimes, a stochastic process with multidimensiontal time or equivalently, a multiparameter stochastic process.

Random problems are common in the natural sciences, particularly in those sciences which investigate continuum phenomena -- for example, fluid mechanics and electrodynamics. However, although the problems are common, treatment of these problems as multiparameter stochastic phenomena is a relatively recent development. Considering the extreme difficulty of such treatments, this late development is not surprising. Indeed, at the present time random field models generally fall into two extreme categories: exact models which are intractable to

calculation and computation, and computationally tractable models which do not model realistic phenomena. It was the goal of this thesis to define new models of random fields which are general enough to cover some realistic phenomena, yet which possess enough structure to give some hope of computational results.

This thesis investigates two mathematical models for such random fields. In Part I we study random fields indexed by partially ordered sets and we develop martingale and optional sampling theory for such partially-ordered time processes. In Part II we study random fields indexed by R^n which can be represented by stochastic integrals over R^n. In section 2 of this introduction we discuss the contents of the thesis in more detail. In the remainder of the present section we briefly survey a range of random field problems and a corresponding range of random field mathematical models.

Perhaps the oldest and one of the most difficult applications of stochastic methods to field problems is in turbulence theory. Here the random fields are random three dimensional velocity vectors parameterized by four dimensional space-time. Taylor (1938) first modelled the phenomenon of turbulence in fluid dynamics as a

spatially stationary random field which may be described by its spectral characteristics. Batchelor (1953) gives a good discussion of this approach with emphasis on using this model to understand the physical phenomenon of turbulence. The recent encyclopedic treatment by Monin and Yaglom (1971, 1975) provides a detailed treatment of the mathematical basis of this statistical model for turbulence as well as the physics. For a concise discussion of the mathematical tools and techniques alone, see Lumley (1970).

The statistical approach to turbulence is based on Fourier and correlation analysis of the random velocity field, and it uses the hydrodynamic equations to derive relationships between the various correlation coefficients. Unfortunately, the resulting set of equations is infinite and they must be solved simultaneously to obtain any one correlation coefficient. Thus, in order to compute low order correlation coefficients one must assume that the correlation coefficients vanish or have a fixed form above some order so that one can truncate the infinite set of equations and obtain a finite set of equations for the desired low order correlations. This problem of truncating the infinite set of equations is known as the closure problem and is discussed, for example, in

Monin and Yaglom (1975).

Despite the extreme difficulty of obtaining computational results from the statistical turbulence model, this approach is widely taken in related sciences. For example, Csanady (1973) uses some of the simpler results in his study of turbulent diffusion of pollutants in the atmosphere. Epstein (1969) also takes this general approach for his study of the statistical treatment of dynamic meteorology. A similar spectral model is popular for large scale meteorological fields. See, for example, Tatarskaya (1965) or more recently, Fraedrich and Bottiger (1978).

A problem related to the turbulence problem which uses similar mathematical techniques is the problem of wave propagation through random media. For example, one might have electromagnetic waves passing through a medium whose refractive index is represented as a random scalar field -- see Dence and Spence (1973). Similarly, one has sound waves propagating underwater (see Beseieris and Kohler (1978)) and seismic waves propagating through the ground (see Cameron and Hannan (1978)). In each of these cases the practically important problem is to reconstruct the most probable signal from the received wave which has been distorted by the random medium. Frisch (1968) gives

a more general, detailed discussion of these problems.

A different problem which also uses spectral techniques is the problem of describing anomalies in the earth's gravitational field. The problem is roughly similar to the turbulence problem in that there is a basic spatially stationary random field which must satisfy additional equations (principally Laplace's equation in this case rather than the Navier-Stokes equation of fluid dynamics). Note that the random field in this case is naturally parameterized by a sphere. Consequently, spherical harmonic functions are used in the spectral analysis of the global statistical properties. Cartesian spectral analysis is possible for the spectral analysis of local statistical properties. The paper by Nash and Jordan (1978) surveys current results and provides an extensive bibliography.

Spatially stationary random fields are also used to describe phenomena for which there are no additional mathematical relationships to satisfy. The principal example is that of two dimensional image processing. Here the problem is to describe the noise in the image by a spatially stationary random field model and then estimate the uncorrupted image in a statistically optimal way. Rosenfeld and Kak (1976) provide an excellent survey

of the general problem of digital image processing, discussing both the statistical and nonstatistical methods. Larrimore (1977) surveys the problem of statistical inference for such problems. Note that the problem of computer processing of images has inspired much work on two-parameter recursive estimation techniques. We will discuss this work later.

Problems similar to the image processing problem occur for geology and hydrology. Again, spatially stationary random fields are used to describe phenomena for which there are no additional mathematical relationships known. The typical problem is to use a spatially stationary random field model to describe statistically a spatial pattern for which only a few measurements are available. For example, Rhenals, Rodriguez and Schaake (1974) use two parameter spatially stationary random fields to describe rainfall patterns. The practical problem here is to obtain a statistical estimate of rainfall and run-off patterns from relatively few rain gauges. A similar approach is taken in statistical geology, for example, as discussed in Merriam (1970). Note that Matheron (1965) has developed a correlation technique called kriging which can be used to describe certain nonstationary fields. The problem in geology is to

estimate statistically mineral deposits in a given region on the basis of very few measurements (mines, wells) or on the basis of indirect measurements (surface topography, faults). One uses observations to construct a probability map of mineral deposits and then one explores the region most likely to contain deposits. See Harbaugh, Doveton and Davis (1977) for a survey of this approach to oil exploration. Delfiner and Delhomme (1975) discuss the general application of kriging to this problem. A specific application of spectral techniques to real mineral exploration in South Africa is discussed in Longshaw and Gilbertson (1976).

So far we have discussed mainly the spatially stationary random field models. Indeed, these are the most popular models and they possess a rich structure determined by the geometric structure of the parameter set. Generally speaking, the structure of the parameter set determines what assumptions one can make about the mutual dependence of the parameterized collection of random variables which constitute the random field. If the parameter set has a group of transformations associated with it, it is possible to generalize the concept of stationary time series to the so-called second-order homogeneous random field. One does this by

requiring the covariance function of the random field to be invariant under the transformations of the parameters. For example, a random field which is parameterized by R^3 and which has a covariance function invariant with respect to rotations and translations in R^3 is called isotropic. The general group theoretic case is treated by Yaglom (1961). An earlier treatment of the special case of random fields parameterized by R^2 is given by Whittle (1954) who pays special attention to problems of statistical inference for such homogeneous random fields. The statistical problem for more general homogeneous random fields is discussed in Hannan (1965, 1967, 1969).

Homogeneous random field models suffer the serious problem that few realistic phenomena are actually homogeneous.

But while the stationary theory of random fields has become well-established and richly developed, the development of nonstationary theories of random fields has lagged behind. Much of the difficulty stems from the fact that the usual one-parameter theory of dynamic processes depends heavily on the specific properties of the real line. Particularly important is the ordering

of the real line which permits its interpretation as a time parameter for a dynamic process. Unfortunately, for all practical purposes the total linear ordering of the real line is unique to that parameter set. This fact severely complicates the attempt to model non-stationary random fields by drawing analogies with one-parameter processes. Nevertheless, some progress has been made.

For example, mathematicians interested in statistical mechanics models have generalized the notion of Markov property to random fields. Kemeny, Snell and Knapp (1976) give a brief introduction to the mathematical theory of Markov random fields with a denumerable parameter set. Spitzer (1971) and Preston (1974) introduce the Markov random field in connection with the theoretical study of phase transitions in statistical mechanics.

Bartlett (1975) has used nearest neighbor Markov models to represent patterns of species distribution in ecology.

Earlier, Lévy (1956) generalized the concept of Markov property to random fields with continuous parameter space R^n. Wong (1971) gives a brief introduction to this topic, and in his paper Wong (1969) gives covariance

conditions for a homogeneous Markov field with parameter space R^n. These covariance conditions are very restrictive and indeed, they show that nontrivial homogeneous Markov fields are possible only for certain dimensions n. Kallianpur and Mandrekar (1974) obtain a wider class of Markov fields by defining generalized random fields analogous to generalized functions (see also Gelfand and Vilenkin (1964)).

By generalizing the order relations of the real line, one can develop another model of nonstationary random fields and this is the model we develop in this thesis. Instead of assuming a total linear ordering of the parameter set, one assumes only a partial ordering of the parameter set. Chapter 1 of part I presents the formal definition of partial order, but as an example, the order relation between sets defined by set inclusion is prototypical. With this partial order structure on the parameter set it is possible to define naturally martingale, stopping time, increasing process and related concepts for random fields, and several authors have done this. Bochner (1955) presented martingale theory with a general partially ordered time set. In this paper Bochner attempted to unite several ideas

from the theory of stochastic processes under a general theory of martingales indexed by a partially ordered parameter set. In addition he stated (without proof) the general version of the submartingale convergence theorem and the optional sampling theorem for submartingales. Unfortunately, both results are not generally true when the parameter set is not totally linearly ordered. Krickeberg (1956) discussed fully the convergence question when the parameter set is partially ordered and directed (so that every two parameters have a common upper bound in the parameter set). Later Chow (1960) noted that the optional sampling theorem for submartingales was generally false, and he proved it true for martingales parameterized by directed sets. Chow (1960) also considered convergence of these martingales. Recently, Kurtz (1977) has proved the optional sampling theorem for martingales indexed by a general partially ordered parameter set with topological structure. In Part I we discuss in detail the optional sampling theorem for submartingales: this result is surprisingly different from the martingale result.

Specific partially ordered parameter sets have

received much more attention than the general case. The most studied example is the parameter set R^n with a coordinate-wise ordering. That is, if $\underline{s}$ and $\underline{t}$ are vectors in R^n, one defines the order relation $\underline{t} \leq \underline{s}$ to mean that $t_i \leq s_i$ for each of the n coordinates s_i and t_i of $\underline{s}$ and $\underline{t}$, respectively. Such a partially ordered parameter is known as multidimensional time.

The additional structure of the multidimensional time parameter set allows one to obtain more specific results. For example, Cairoli (1970) has extended the martingale inequalities of Doob (1953) to a class of multiparameter martingales. Cairoli (1971) has also extended the submartingale decomposition theorem but the extension applies to a multiparameter process called an S-process and not to multiparameter submartingales. In Part I we present a counterexample which shows that the general multiparameter submargingale can not be decomposed as the ordinary one-parameter submartingale.

Multidimensional time has also provided a convenient setting for extending the Wiener and Ito stochastic integrals which are closely connected with the

one-parameter martingale theory. Ito (1951) introduced the multiple Wiener integral. In a series of recent papers Wong and Zakai (1974, 1975a, 1975b) have extended the one-parameter results of Ito (1961) to two-parameter processes. Among other results, Wong and Zakai extended the celebrated stochastic differentiation formula of Ito to two-parameter processes and used it to represent two-parameter martingales by means of stochastic integrals over R^2 and $R^2 \times R^2$. Cairoli and Walsh (1975, 1977) present similar results and Yor (1975) generalizes the representation theorem to n-parameter martingales.

Multiparameter processes have also been applied in information processing problems in recent years. See Bose (1977) for a general survey of current work in multiparameter system theory. For example, two-parameter processes are widely used to model the filtering and processing of noisy two-dimensional images. In this problem it is particularly desirable to obtain efficient computational procedures for processing large arrays of data. In this regard much work has been done on designing recursive algorithms for which the recursion is defined with respect to the two-parameter partial

order. We discuss this in more detail in Chapter 2 of Part I. Larrimore (1977) gives a general survey of the area. Woods (1976) and Woods and Radewan (1977) discuss recursive estimators in the discrete parameter case. Wong (1968, 1976, 1978) applies the stochastic integration theory of Wong and Zakai (1974, 1975) to obtain recursive algorithms for the continuous parameter case.

2. Summary of Thesis

The following is a brief synopsis of the contents of the chapters in this thesis.

Part I. The Optional Sampling Theorem for Partially Ordered Time Processes.

Chapter I

Partial order relations are formally defined and then the partially ordered time stochastic process, martingale, stopping time and other concepts are defined. Multiparameter processes are introduced and some basic examples, particularly the multiparameter Wiener process, are presented and discussed.

We review Kurtz's (1977) results on optional sampling for martingales and show the results do not

generalize to submartingales. However, we do prove the submartingale theorem for a class of two-parameter submartingales and present a counterexample for a similar class of three-parameter submartingales.

Chapter 2

Given an a priori distribution of a signal source in space and time, a moving sensor must intercept the radiating signal in minimum expected time. This simple problem illustrates how the issues of causality (of the information structure) and recursion (of the computational solution) arise in the formulation and solution of a partially ordered time problem. This problem is generalized in Chapter 3.

Chapter 3

We extend the results of Chapter 2 to a general classof stopping problems with partially nested information structure. After formulating the problem precisely, we show how to solve it by means of a dynamic programming equation. We find that many of the usual one-parameter results still hold. For example, the dynamic program has a unique optimal cost function associated with it. This cost function can be computed

by means of a backward recursive computation with respect to the partial order relation on the time set.

Chapter 4

We define a new type of stopping time, called reachable, which is included in the general class of stopping times defined in Chapter 1. We show that the optional sampling theorem is true for submartingales if the stopping times are reachable. Conversely, using the dynamic programming results of Chpater 3, we show that if the optional sampling theorem is true for all submartingales given a fixed pair of stopping times, then the stopping times must be reachable.

Chapter 5

We conclude Part I by discussing the significance of the results for random field theory and by indicating possible extensions and directions for further research.

Part II. Multiparameter Stochastic Calculus

Chapter 1. Introduction

We discuss briefly the background of the stochastic calculus.

Chapter 2. Definitions and Notation

We present the necessary definitions and notation we need to develop the multiparameter stochastic calculus.

Chapter 3. Two-Parameter Stochastic Differentiation

We discuss the two-parameter stochastic differentiation formula and the various new types of stochastic integrals required to interpret the formula. We compare our integrals with the Wong-Zakai (1974) and Cairoli-Walsh (1975) integrals, thus establishing the equivalence of their formulas and our own. We also introduce a stochastic partial differential operator that obeys very simple differentiation rules and allows a simple formal derivation of the stochastic differentiation formula. This stochastic partial differentiation operator is the same as the "stochastic partial" which appears in Cairoli-Walsh (1975).

Chapter 4. Multiparameter Stochastic Differentiation

This chapter discusses the n-parameter stochastic differentiation formula and the n-parameter integrals necessary to interpret it rigorously. Here we emphasize the concept of a stochastic differential. This point of view permits a simple formal derivation of the multiparameter stochastic differentiation formula and

emphasizes the multiparameter formula's close relation to the deterministic formula and the one-parameter Ito formula. The derivation up to this point is formal in nature, as we utilize differentiation formulas and differentials without rigorous proof of their properties.

Chapter 5. Stochastic Measure and Stochastic Integral

This chapter contains the first rigorous results. Here we define the concept of stochastic measure and its corresponding integral in order to treat all types of stochastic integrals together. At the same time we ofer some geometric insight into the stochastic calculus.

Chapter 6. Mixed Integrals and Products of Stochastic Measures

Here we define mixed stochastic-deterministic integrals and the product of stochastic measures in preparation for the main results in the next chapter.

Chapter 7. Product Differentiation Rule

We present the rigorous proof of the product differentiation rule, which is the central technical result of this second part. This justifies the formal

manipulations of chapter 2 and chapter 3.

Chapter 8. Representation of Multiparameter Martingales

We first prove the multiparameter stochastic differentiation formula for vector-valued processes. Using this result we show that all Wiener functionals with finite variance can be represented by multiparameter stochastic integrals. From this it follows that all square integrable multiparameter martingales can be represented by multiparameter stochastic integrals.

Chapter 9. Conclusion

We conclude by discussing possible generalization and extensions of the multiparameter stochastic differentiation formula to more general types of processes.

PART I: THE OPTIONAL SAMPLING THEOREM FOR PARTIALLY ORDERED TIME PROCESSES

CHAPTER 1

OPTIONAL SAMPLING OF STOCHASTIC FIELDS

1. Introduction

It was recongized early by Bochner (1955) that it is possible to extend the idea of stochastic process to processes with time parameters taking values from a partially ordered set* rather than the usual totally ordered parameter set such as the integers or the real numbers. With respect to the partial ordering, there exists a natural concept of submartingale, martingale and stopping time, and one can formulate the corresponding optional sampling theorem. Bochner (1955) stated the optional sampling theorem for martingales and submartingales indexed by a special kind of p.o. set called a directed set (see section 2 for definitions and notation). Unfortunately, Bochner omitted the proof of these results and in fact the result is false for submartingales in general (see the second counterexample in section 3). Later Chow (1960) noted that Bochner's result was not true for submartingales

*We will abbreviate "partially ordered" as "p.o." in the sequel. Thus, we have "p.o. set" for "partially ordered set" and "p.o. time" for "partially ordered time."

(the counterexample in section 3 is essentially the same as Chow's counterexample), but Chow was still able to prove the optional sampling theorem for martingales indexed by directed sets. However, in his paper Chow assumed an unnatural restriction on stopping times defined on p.o. time sets. We will explain this restriction in section 2 when we define stopping times on p.o. time sets. Recently Kurtz (1977) has been able to prove the optional sampling theorem for martingales indexed by directed sets using the natural definition of stopping time which we also assume. For the case of discrete valued stopping times, the proof of Kurtz is similar to that of Chow, but without the unnatural restriction on the stopping time that Chow makes. In addition, Kurtz extends the optional sampling theorem to the case of martingales indexed by a topological lattice.

However, none of these proofs of the optional sampling theorem for martingales indexed by directed sets extends to the submartingale case. Indeed, as we mentioned earlier, simple counterexamples show that the submartingale version of the optional sampling theorem is generally false. Nevertheless, we prove the optional sampling theorem for two-parameter submartingales measurable with respect to the two-parameter Wiener process defined in Park (1970).

More generally, the result is proved true if the two-parameter family of σ-fields satisfies the conditional independence hypothesis of Cairoli and Walsh (1975). A simple counterexample shows that the optional sampling theorem for two-parameter submartingales can be false if this conditional independence hypothesis is not satisfied. Moreover, the corresponding optional sampling theorem for n-parameter submartingales can be false for $n > 2$, even if one assumes the n-parameter analogy of the conditional independence hypothesis.

This chapter is organized as follows. Section 2 presents basic definitions for p.o. time sets. Thus, we define martingales, submartingales and stopping times with respect to p.o. time, and we state what the optional sampling theorem should be for p.o. time submartingales. Section 3 presents the simple proof of the optional sampling theorem for martingales indexed by finite directed sets. We then discuss why this proof doesn't work if the time set is not directed or if martingales are replaced by submartingales. In the case where the p.o. time set is not directed but where the family of σ-fields satisfies the conditional independence hypothesis, we show that the optional sampling theorem is still true for martingales.

In section 4 we assume that the time set is Z_+^2 and

that the probability structure satisfies the conditional independence hypothesis stated in Cairoli and Walsh (1975). With these assumptions we then prove the optional sampling theorem for two-parameter submartingales when the stopping times take finitely many values. We also present a counter-example to show that analogous results do not hold for the time set R^n when $n > 2$. Section 5 treats the case when the stopping times take countably many values but are not necessarily bounded in Z_+^2. This case requires a uniform integrability type condition on the submartingale similar to the condition for optional sampling on the time set of integers. Section 6 extends the result of section 4 to stopping times which are bounded in R^2 but may take a continuum of values. In this case we must assume the submartingales are "right continuous" in an appropriate sense with respect to the partial ordering of R^2. Finally, section 7 concludes the chapter by discussing the counter-examples and the significance of the failure of the optional sampling theorem for submartingales. This section indicates how the negative results concerning the optional sampling theorem for general p.o. times suggest a new approach to optional sampling which we develop in the succeeding chapters of part I.

1.2 Definitions and Notation

Although we are most concerned with the time set R^2 in this chapter, for future use we will present the definitions for more general partially ordered (p.o.) time sets. Recall first the definition of partially ordered set:[1]

A partially ordered set $(T,\leq)$ is a set T together with a binary relation $\leq$ defined on T and satisfying the following properties for all r,s and t in T:

(2.1) $t \leq t$ (reflexive property),

(2.2) $t \leq s$ and $s \leq t$ imply $s = t$ (antisymmetric property),

(2.3) $r \leq s$ and $s \leq t$ imply $r \leq t$ (transitive property).

What makes the partial order partial is the absence of the fourth property,

(2.4) $s \leq t$ or $t \leq s$ $\forall$ $s, t \in T$.

If a partial order also obeys (2.4), one says it is a total order. Thus, for example, the usual ordering of the real numbers or integers is total, but the ordering of sets defined by set inclusion is only partial.

---

[1] See, for example, Birkhoff and MacLane (1967).

[2] The same symbol $\leq$ will denote all order relations, including the usual order for integers or real numbers. We will indicate what order relation $\leq$ represents whenever there is danger of confusion.

A special kind of p.o. set which we will discuss in section 1.3 is the directed set. A set $(T,\leq)$ is directed if $(T,\leq)$ is partially ordered and if any two elements of T have a common upper bound - i.e. for any r, s in T there is a t in T such that $r \leq t$ and $s \leq t$.

With this definition of p.o. set the notions of increasing σ-fields, martingales and stopping times for partially ordered time sets are natural generalizations of these concepts for the usual linear time sets. For example, Bauer (1971) discusses the general case in his text on probability theory when he defines martingales. An early paper which takes this general viewpoint of partially ordered time is Bochner (1955). For convenience, the definitions are repeated here.

Let $(T,\leq)$ be a partially ordered set and let (P,Ω,F) be a probability space. A collection $\{F_s : s\in T\}$ of sub-σ-fields of F is increasing if $s \leq t$ in T implies that $F_s \subset F_t$. We will assume that each F_s contains all the null events, that is, all subsets of zero probability sets.

A random function $X : T \times \Omega \to R$ is adapted to the increasing family $\{F_s : s\in T\}$ if for each t in T, the map $\omega \to X(t,\omega)$ is F_t-measurable. Keeping with the usual

notation for processes, we will sometimes write $X(t,\omega)$ as $X_t(\omega)$ and the random variable $\omega \to X(t,\omega)$ as X_t. Note that for general parameter sets T the mapping $t \to X_t$ is a random field as described in the general introduction to this thesis, but one with the special property that it is adapted to an increasing family of σ-fields.

A submartingale with respect to the increasing family $\{F_t : t\varepsilon T\}$ of σ-fields is a collection $\{M_t : t\varepsilon T\}$ of real-valued random variables such that for each t in T the random variable M_t is F_t-measurable, the expectation $E(|M_t|)$ is finite and for all s in T such that $s \leq t$,

$$E(M_t|F_s) \geq M_s. \tag{2.5}$$

Similarly, a supermartingale is defined by reversing the inequality in (2.5) and a martingale is defined by replacing inequality with equality in (2.5).

As an example of a martingale we present Ito's normal random measure. Suppose $(X,\mathcal{B},m)$ is a measure space and define T as the set

$$T = \{E : m(E) < \infty, E \varepsilon \mathcal{B}\},$$

partially ordered by set inclusion. Ito (1951) constructed a zero-mean, Gaussian family $\{\beta(E) : E \in T\}$ with the property that for all E and E' in T

$$(2.6) \qquad E(\beta(E)\beta(E')) = m(E \cap E')$$

The property (2.6) implies that for disjoint sets $\{E_n : n \geq 1\}$ in T such that $\bigcup_{n\geq 1} E_n$ belongs to T.

$$(2.7) \qquad \beta(\bigcup_{n\geq 1} E_n) = \sum_{n\geq 1} \beta(E_n) \quad \text{a.s.}$$

where the infinite sum is defined as a mean square limit. Due to (2.7), one calls β a random measure.

If F_E is the smallest σ-field which makes $\beta(E')$ measurable for all E' in $\mathcal{B}$ for which $E' \subset E$, then $\{F_E : E \in T\}$ is an increasing family of σ-fields with respect to the set inclusion order on T. The normal random measure β is a random function on T adapted to $\{F_E : E \in T\}$ as is easily seen. Moreover, β is a martingale. To see this note that if $E' \subset E$ for E, E' in T, then $\beta(E-E')$ and $\beta(E')$ are independent of each other and

$$E(\beta(E)\,|\,\mathcal{F}_{E'}) = E(\beta(E-E')+\beta(E')\,|\,\mathcal{F}_{E'})$$

$$= E(\beta(E-E')) + \beta(E')$$

$$= \beta(E').$$

An important example of a partially ordered time martingale is the multiparameter Wiener process discussed for example in Park (1970). Although they are particularly concerned with stochastic integration, the papers by Cairoli and Walsh (1975) and Wong and Zakai (1974) and part II of this thesis also contain reference material on the two-parameter Wiener process. One can derive the multiparameter Wiener process from the normal random measure presented above as follows.

Let T be the set of t in R^n with nonnegative coordinates t_i. Define the multiparameter order $\leq$ on T so that $s \leq t$ for s and t in T if and only if $s_1 \leq t_i$ (the order relation on R) for each coordinate, $1 \leq i \leq n$. For t in T let $[0,t]$ denote the closed rectangle in R^n defined as

$$[0,t] = [0,t_1] \times [0,t_2] \times \ldots \times [0,t_n].$$

Suppose $X = R^n$. $\mathcal{B}$ is the Borel σ-field of R^n and

m is Lebesgue measure on R^n, and form the normal random measure β on R^n and the corresponding σ-field $\mathcal{F}_E$ as above. For each t in T, define

$$W_t = \beta([0,t]), \tag{2.8}$$

$$\mathcal{F}_t = \mathcal{F}_{[0,t]}. \tag{2.9}$$

Then it is easy to see that $\{W_t : t \in T\}$ is a martingale with respect to the σ-fields $\{\mathcal{F}_t : t \in T\}$ and the multiparameter partial order. Part II will treat the multiparameter Wiener process on R^n in more detail and define stochastic integrals with respect to it. We will see there that all square integrable multiparameter martingales with respect to the σ-fields of (2.9) can be represented as stochastic integrals.

Returning to the case of general p.o. time, let us define stopping times. A <u>stopping time</u> τ with respect to the p.o. time set $(T,\leq)$ and the corresponding increasing family of σ-fields $\{\mathcal{F}_t : t \in T\}$ is defined as a T-valued function of the underlying probability space Ω such that the set $\{\omega : \tau(\omega) \leq t\}$ is in $\mathcal{F}_t$ for each t in T. This seems to be the natural extension of stopping times to p.o. time sets and this definition is the one given by several authors including Bochner (1955), Krickeberg (1956),

Bauer (1971) and Kurtz (1977). Chow (1960), on the other hand, imposes the extra condition that $\{\omega : \tau(\omega) \geq t\}$ is in $\mathcal{F}_t$ as well as $\{\omega : \tau(\omega) \leq t\}$. For one-parameter time sets which are discrete, Chow's extra condition follows from the first more natural condition that $\{\omega : \tau(\omega) \leq t\}$ is $\mathcal{F}_t$-measurable. For general p.o. time sets this need not be the case as figure 2.1 illustrates and as we now explain. In either the one-parameter or p.o. time case, the set $\{\omega : \tau(\omega) < t\}$ is $\mathcal{F}_t$-measurable for a stopping time τ. In the one-parameter case, however, $\tau(\omega) \not< t$ is equivalent to $\tau(\omega) \geq t$ and thus, the set $\{\omega : \tau(\omega) \geq t\}$ is the complement of $\{\omega : \tau(\omega) < t\}$ and must be $\mathcal{F}_t$-measurable. On the other hand, in the general p.o. time case (for example, the two-parameter case of figure 2.1), the relation $\tau(\omega) \not< t$ need not imply the relation $\tau(\omega) \geq t$. Thus, in the general p.o. time case the $\mathcal{F}_t$-measurability of $\{\tau \leq t\}$ need not imply the $\mathcal{F}_t$-measurability of $\{\tau \geq t\}$.

Kurtz (1977) showed that Chow's extra condition that $\{\tau \geq t\} \in \mathcal{F}_t$ is unnecessary to prove the optional sampling theorem for martingales indexed by directed sets. This extra condition is also insufficient to obtain the optional sampling theorem for martingales indexed by non-directed sets or by submartingales indexed by directed sets. Indeed, the two counterexamples of section 1.3 both use stopping

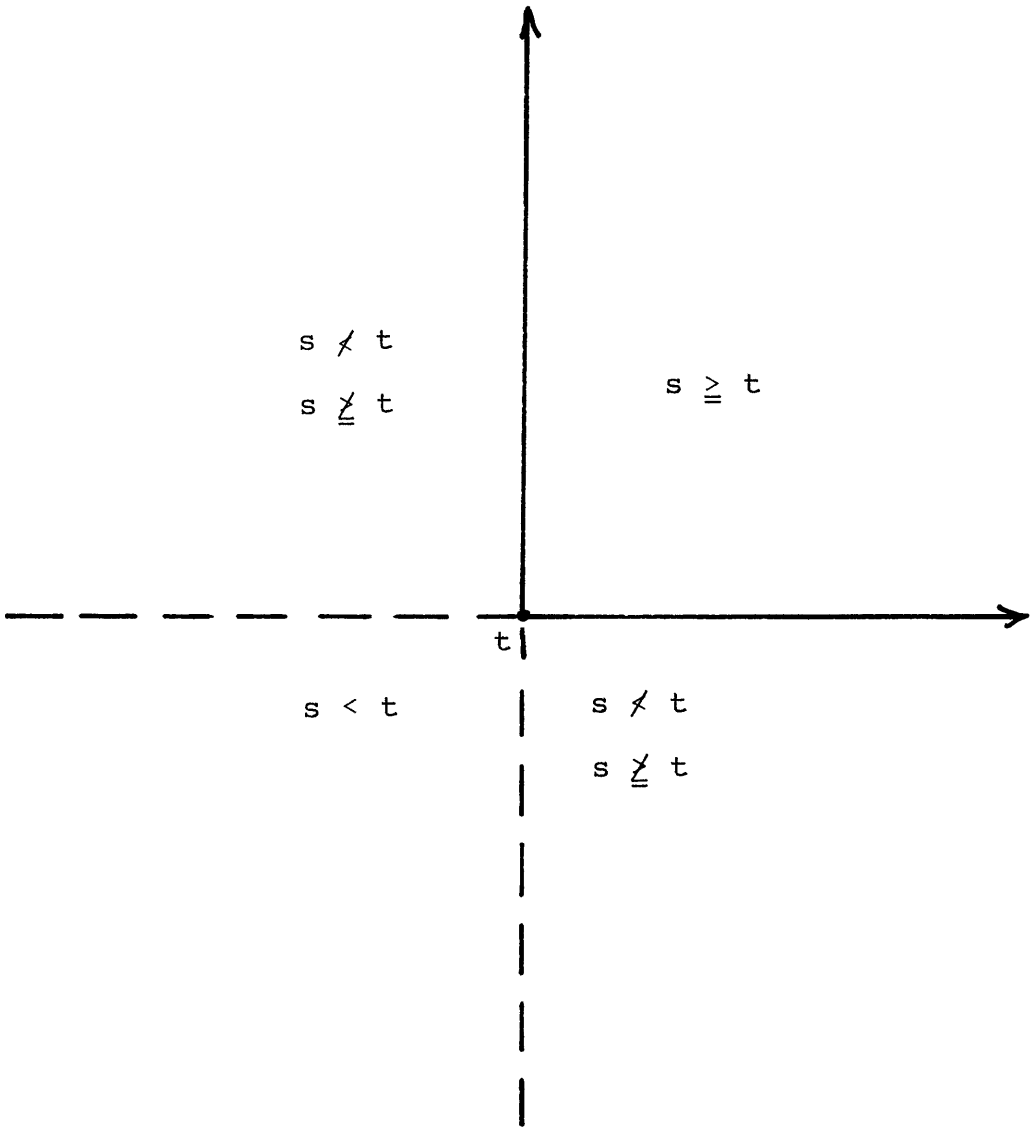


Figure 2.1

times which satisfy Chow's more restrictive definition. Therefore, we do not use Chow's definition of stopping time, but we restrict our attention to the first natural definition.

If τ is a stopping time, there is a corresponding σ-field denoted $\mathcal{F}_\tau$ defined as the collection of all $\mathcal{F}$-measurable sets A such that the set $A \cap \{\tau \leq t\}$ is in $\mathcal{F}_t$ for each t in T.

The optional sampling theorem for general p.o. sets essentially asserts that under appropriate conditions, if τ and σ are stopping times on T such that $\sigma \leq \tau$, and if $\{M_t : t \in T\}$ is a submartingale (martingale) with respect to $\{\mathcal{F}_t : t \in T\}$, then $E(M_\tau | \mathcal{F}_\sigma) \geq (=) M_\sigma$. As we see in succeeding chapters, the distinction between the martingale and submartingale versions of the optional sampling theorem is crucial. Whereas the martingale version is true in very general cases, the submartingale version is true only in special cases - some of which we discuss in this chapter and in chapter 4 (section 4.1).

In proving the optional sampling theorem for the special time sets $T = Z_+^2$ or $T = R_+^2$ we will assume that the σ-fields $\{\mathcal{F}_t : t \in T\}$ obey the following conditional independence hypothesis. Let $s \wedge t$ denote the point

$$(\min\{s_1,t_1\}, \min\{s_2,t_2\})$$

as illustrated in figure 2.2.

Conditional independence hypothesis:

For each s and t in the time set T, the σ-fields $\mathcal{F}_s$ and $\mathcal{F}_t$ are conditionally independent given the σ-field $\mathcal{F}_{s\wedge t}$.

This is condition (F4), p. 113, in Cairoli and Walsh (1975). Note that this condition is satisfied if $\{\mathcal{F}_t : t \in T\}$ is generated by the two-parameter Wiener process, as it comes from Ito's normal random measure. Indeed, if we let $A = [0,t] - [0,t\wedge s]$, $B = [0,t\wedge s]$, $C = [0,s] - [0,t\wedge s]$ as in figure 2.2, then the σ-field $\mathcal{F}_t$ is generated by $\mathcal{F}_A$ and $\mathcal{F}_B$ and the σ-field $\mathcal{F}_s$ is generated by $\mathcal{F}_C$ and $\mathcal{F}_B$. Moreover, $\mathcal{F}_B = \mathcal{F}_{t\wedge s}$ and all three σ-fields $\mathcal{F}_A$, $\mathcal{F}_B$, $\mathcal{F}_C$ are independent, since A, B, C are mutually disjoint. It follows that given $\mathcal{F}_{t\wedge s}$ $(= \mathcal{F}_B)$ the two σ-fields $\mathcal{F}_t$ and $\mathcal{F}_s$ must be conditionally independent.

The conditional independence hypothesis generalizes easily to processes on Z_+^n or R_+^n. Indeed, this hypothesis makes sense for any p.o. set $(T,\leq)$ such that for any two elements t,s in T there is a greatest lower bound $t\wedge s$ also in T. That is, $r \leq t$ and $r \leq s$ imply that

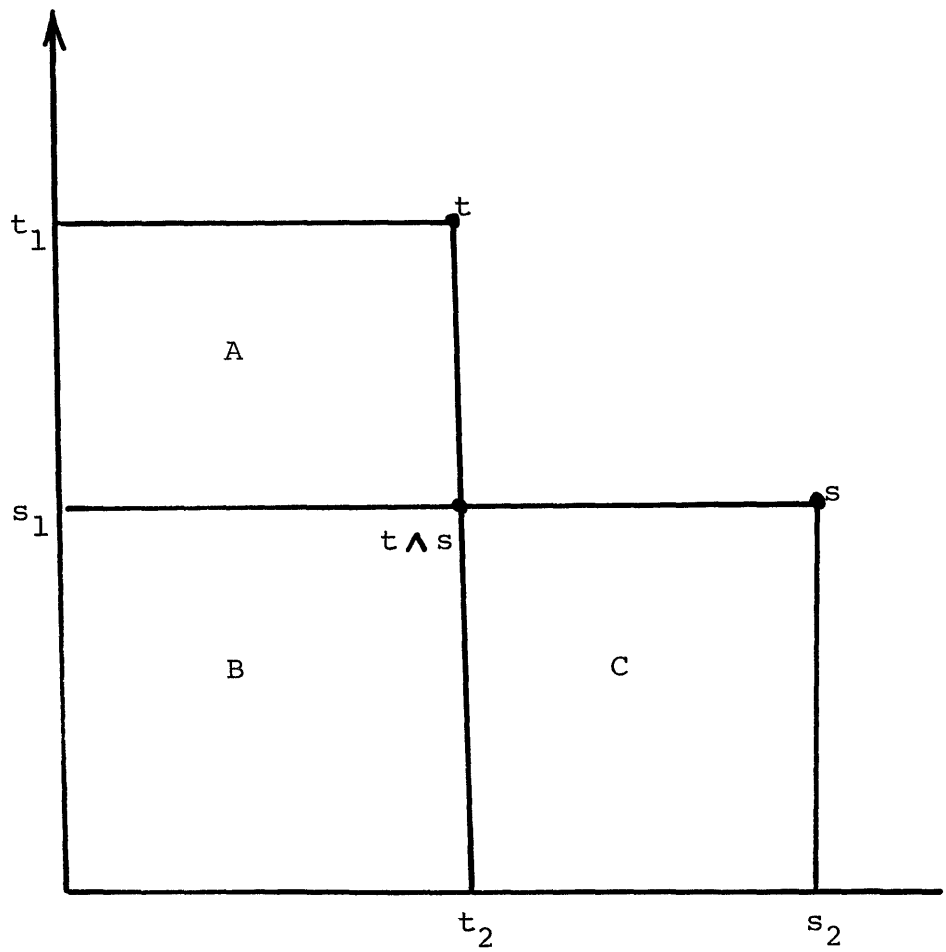


Figure 2.2

$r \leq t \wedge s$.

In the case $T = Z_+^n$ or R_+^n, the conditional independence hypothesis is true if the σ-fields are generated by the n-parameter Wiener process on R_+^n.

To complete this section we note the following continuity condition that will be needed in the case $T = R_+^2$. A process $\{M_t : t \in R_+^2\}$ is <u>right continuous</u> if for all ω except for a set of probability zero, for all t in R_+^2, we have

$$\lim_{\substack{s \to t \\ t \leq s}} M_s(\omega) = M_t(\omega).$$

1.3 Optional Sampling for Martingales and Counterexamples

Bochner (1955) originally stated the optional sampling theorem for both martingales and submartingales indexed by a directed set, but he omitted the proof of either case. Although Bochner's theorem is true for martingales, as Kurtz (1977) has recently shown, it is not generally true for submartingales as we will show by counterexample in this section. This failure of the optional sampling for submartingales has nothing to do with lack of uniform integrability type conditions - the examples we present have finite time sets and finite sample spaces so that

such conditions are trivially satisfied. The failure for submartingales is essentially due to the lack of a total ordering of the p.o. time set.

In this section we present the optional sampling theorem for martingales indexed by directed sets due to Kurtz (1977).* Then we show by means of a very simple counterexample that this result can be false if the p.o. time set is not directed. Next we present a simple counterexample that shows the submartingale version of the optional sampling theorem can be false even if the p.o. time set is directed. Finally, we show that the martingale version of the optional sampling theorem is true for finite p.o. time sets which are not directed if the conditional independence hypothesis is true.

Theorem 3.1 (optional sampling for martingales, directed time sets)

Suppose $(T,\leq)$ is a finite directed p.o. set and let σ and τ be stopping times with respect to the increasing family of σ-fields $\{\mathcal{F}_t : t \in T\}$ such that $\sigma \leq \tau$. If $\{M_t : t \in T\}$ is a martingale with respect

*Kurtz (1977) considers infinite time sets with topological structure, but we confine ourselves to finite p.o. time sets.

to $\{F_t : t \in T\}$, then $E(M_\tau | F_\sigma) = M_\sigma$.

Proof:

To prove $E(M_\tau | F_\sigma) = M_\sigma$ we must prove for each $A \in F_\sigma$ that

$$(3.1) \qquad E(M_\tau \cdot I_A) = E(M_\sigma \cdot I_A)$$

where I_A is the indicator function for the set A. Let t_1 be the greatest element of T (which must exist since T is finite and directed). We prove (3.1) by proving that

$$(3.2) \qquad E(M_{t_1} \cdot I_A) = E(M_\sigma \cdot I_A)$$

and then proving that

$$(3.3) \qquad E(M_{t_1} \cdot I_A) = E(M_\tau \cdot I_A).$$

Since for any s in T we must have $s \leq t_1$, the martingale property of $\{M_t : t \in T\}$ implies that

$$(3.4) \qquad E(M_{t_1} | F_s) = M_s.$$

Note that by definition of F_σ, if $A \in F_\sigma$ then

$A \cap \{\sigma = s\}$ is F_s-measurable. Multiplying (3.4) by $I_{A \cap \{\sigma=s\}}$ and taking expectations yields

$$(3.5) \quad E(M_{t_1} \cdot I_{A \cap \{\sigma=s\}}) = E(M_s \cdot I_{A \cap \{\sigma=s\}}).$$

Since clearly $M_s \cdot I_{A \cap \{\sigma=s\}} = M_\sigma \cdot I_{A \cap \{\sigma=s\}}$, we must have from (3.5) that

$$(3.6) \quad E(M_{t_1} \cdot I_{A \cap \{\sigma=s\}}) = E(M_\sigma \cdot I_{A \cap \{\sigma=s\}}).$$

Since T is finite, we may sum over all s in T to obtain (3.2).

To prove (3.2) we used only the fact that σ was a stopping time and $A \in F_\sigma$. Thus, if we can show that $A \in F_\sigma$ implies $A \in F_\tau$ we will also have (3.3) by the same result by replacing σ with τ in (3.2). Suppose $A \in F_\sigma$. To show $A \in F_\tau$ we must show that

$$(3.7) \qquad A \cap \{\tau \leq t\} \in F_t$$

for each t in T. Since we assume $\sigma \leq \tau$, we have that $\tau \leq t$ implies $\sigma \leq t$. Thus, $\{\sigma \leq t\} \subset \{\tau \leq t\}$ and

(3.8) $\quad A \cap \{\sigma \leq t\} \cap \{\tau \leq t\} = A \cap \{\tau \leq t\}.$

But since $A \in F_\sigma$, we must have $A \cap \{\sigma \leq t\} \in F_t$ by definition of F_σ. Moreover, $\{\tau \leq t\} \in F_t$ since τ is a stopping time. Hence, the intersection $A \cap \{\sigma \leq t\} \cap \{\tau \leq t\}$ is in F_t and (3.8) implies (3.7). This shows that $A \in F_\sigma$ implies $A \in F_\tau$. The result (3.3) follows and the proof is complete. ///

The proof of theorem 3.1 does not work if the p.o. time set T is not directed, nor does it work if the process $\{M_t : t \in T\}$ is assumed to be a submartingale and not a martingale. If T is not directed, there is no greatest element t_1 and clearly the proof we have given for the theorem cannot be used. If T is directed, but $\{M_t : t \in T\}$ is only a submartingale, then we can only prove the inequalities

(3.9) $\quad E(M_{t_1} \cdot I_A) \geq E(M_\sigma \cdot I_A)$

and

(3.10) $\quad E(M_{t_1} \cdot I_A) \geq E(M_\tau \cdot I_A)$

in place of (3.2) and (3.3). Unfortunately, this allows us no deductions about the relationship of $E(M_\sigma \cdot I_A)$ and $E(M_\tau \cdot I_A)$.

The following two counterexamples show that theorem 3.1 does not extend to general non-directed p.o. sets T nor to submartingales $\{M_t : t \in T\}$.

Example 1. (Martingale with non-directed time set)

Let $T = \{(0,0), (0,1), (1,0)\}$ with the partial ordering $(0,0) \leq (0,1)$ and $(0,0) \leq (1,0)$ as illustrated in figure 3.1. Let $\Omega = \{\omega_1, \omega_2\}$, $F = 2^\Omega$, and define $P(\{\omega_1\}) = P(\{\omega_2\}) = \frac{1}{2}$. Then define $F_{(0,0)} = \{\emptyset, \Omega\}$ and let $F_{(0,1)} = F_{(1,0)} = F$. Define the martingale $\{M_t : t \in T\}$ as follows:

$$M_{(0,0)}(\omega_1) = M_{(0,0)}(\omega_2) = 0,$$

$$M_{(1,0)}(\omega_1) = M_{(0,1)}(\omega_2) = -1,$$

$$M_{(1,0)}(\omega_2) = M_{(0,1)}(\omega_1) = 1.$$

Define stopping times σ and τ as follows:

$$\tau(\omega_2) = (0,1), \ \tau(\omega_1) = (1,0),$$

$$\sigma(\omega_1) = \sigma(\omega_2) = (0,0).$$

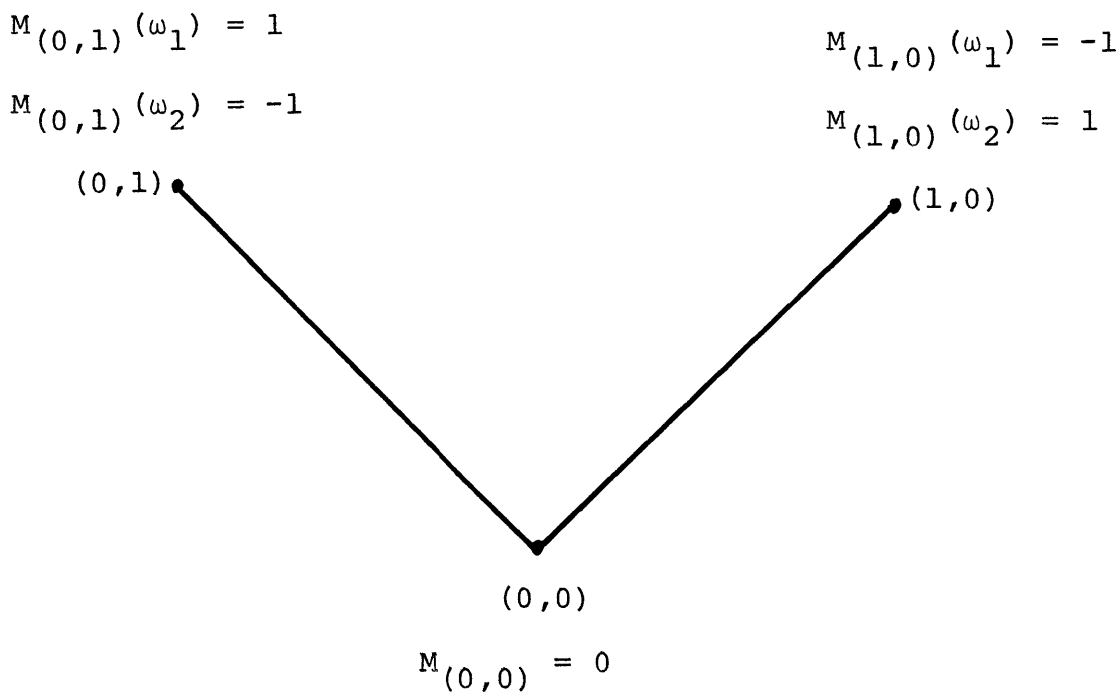

$M_{(0,1)}(\omega_1) = 1$
$M_{(0,1)}(\omega_2) = -1$
(0,1)
$M_{(1,0)}(\omega_1) = -1$
$M_{(1,0)}(\omega_2) = 1$
(1,0)
(0,0)
$M_{(0,0)} = 0$

Figure 3.1

Then it follows that $\sigma \leq \tau$ but

$$E(M_\tau | F_\sigma) = -1 \neq 0 = M_\sigma.$$

Example 2. (Submartingale with directed time set, due to Chow)

Define $T = \{(0,0), (0,1), (1,0), (1,1)\}$ with the partial order relation $(0,0) \leq (0,1)$, $(0,0) \leq (1,0)$, $(0,1) \leq (1,1)$, $(1,0) \leq (1,1)$ and $(0,0) \leq (1,1)$ as shown in figure 3.2. Define F, $F_{(0,0)}$, $F_{(0,1)}$, $F_{(1,0)}$, $M_{(0,0)}$, $M_{(0,1)}$, $M_{(1,0)}$, P, Ω, σ and τ as in example 1. In addition, let $F_{(1,1)} = F$ and define $M_{(1,1)}$ as

$$M_{(1,1)}(\omega_1) = M_{(1,1)}(\omega_2) = 2.$$

It is clear that $\{M_t : t \in T\}$ is a submartingale. However, when τ and σ are defined as in example 1 we have the same result $E(M_\tau | F_\sigma) = -1$ and $M_\sigma = 0$ so that $E(M_\tau | F_\sigma) \not\geq M_\sigma$ even though $\sigma \leq \tau$.

The reason that the optional sampling theorem fails in these two examples is that the events occurring at the two incomparable times $(0,1)$ and $(1,0)$ are dependent. Indeed, the corresponding σ-fields $F_{(1,0)}$ and $F_{(0,1)}$ are equal in these examples. In the next result for martingales we rule out this possibility by assuming

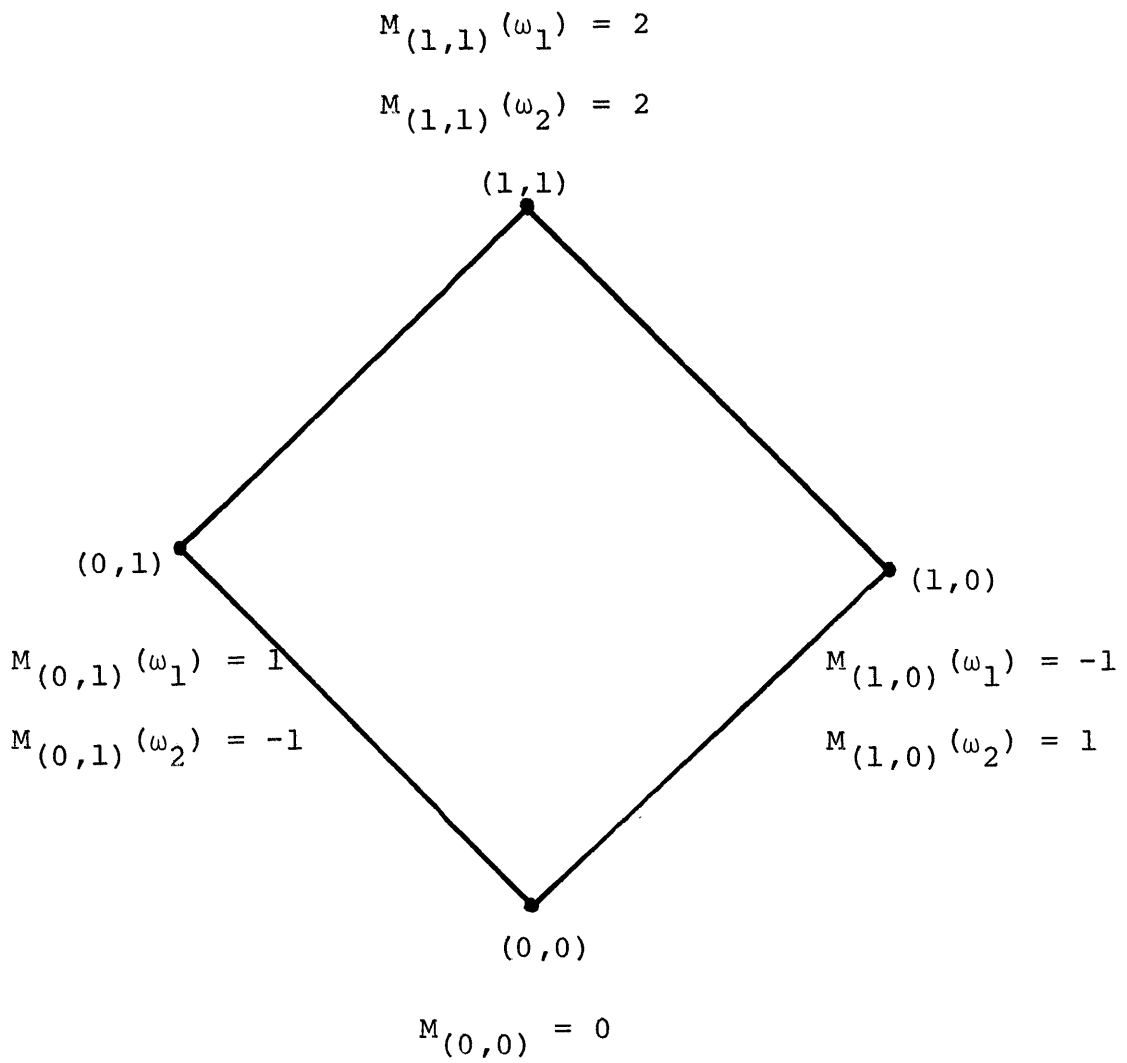

$M_{(1,1)}(\omega_1) = 2$
$M_{(1,1)}(\omega_2) = 2$
(1,1)
(0,1)
(1,0)
$M_{(0,1)}(\omega_1) = 1$
$M_{(0,1)}(\omega_2) = -1$
$M_{(1,0)}(\omega_1) = -1$
$M_{(1,0)}(\omega_2) = 1$
(0,0)
$M_{(0,0)} = 0$

Figure 3.2

that $F_{(1,0)}$ and $F_{(0,1)}$ are conditionally independent given $F_{(0,0)}$.

Theorem 3.2 (optional sampling for martingales, non-directed sets with conditional independence hypothesis)

Suppose that $(T,\leq)$ is a p.o. set with the property that for each t and s in T there exists a greatest lower bound $t \wedge s$. Let $\{F_t : t \in T\}$ be an increasing family of σ-fields indexed by T and suppose that for each t and s in T, the σ-fields F_t and F_s are conditionally independent given $F_{s\wedge t}$. If $\{M_t : t \in T\}$ is a martingale with respect to $\{F_t : t \in T\}$, and if σ and τ are stopping times with $\sigma \leq \tau$, then $E(M_\tau | F_\sigma) = M_\sigma$.

Proof:

Essentially, we show that $(T,\leq)$ may be imbedded in a directed set (T^*,s^*), and that M can be extended to T^* so that the martingale property is preserved. The result then follows from theorem 3.1.

Suppose $1 \notin T$ and define $T^* = T \cup \{1\}$. Define $\leq^*$ so that $t \leq^* 1$ for all t in T^* and so that for $t,s \in T$ we have $t \leq^* s$ if and only if $t \leq s$. Thus, $(T,\leq)$ is imbedded in the directed set $(T^*,\leq^*)$. Define F_1 as the σ-field generated by $\{F_t : t \in T\}$. Thus,

$\{F_t : t \in T^*\}$ is an increasing family of σ-fields. Let $\{t_1, t_2, \ldots t_n\}$ be the maximal elements of T with respect to $\leq$. That is, the t_i are elements such that there is no t in T with $t_i < t$. Define the random variable M_1 by the formula

$$(3.11) \qquad M_1 = \sum_{m=1}^{n} (-1)^{m-1} \{ \sum_{1 \leq i_1 < \ldots < i_m \leq n} M_{t_{i_1} \wedge t_{i_2} \ldots \wedge t_{i_m}} \}.$$

In (3.11), $t_{i_1} \wedge \ldots \wedge t_{i_m}$ denotes the greatest lower bound of $\{t_{i_1}, \ldots, t_{i_m}\}$ which must exist because of our assumptions about T.

To show $\{M_t : t \in T^*\}$ is a martingale with respect to $\{F_t : t \in T^*\}$ it suffices to show that for each i

$$(3.12) \qquad E(M_1 | F_{t_i}) = M_{t_i}.$$

Then, if $t \in T$ we must have $t \leq t_i$ for some i, and because $F_t \subset F_{t_i}$ we would have from (3.12) that

$$(3.13) \qquad E(M_1 | F_t) = M_t.$$

To prove (3.12) we must first derive the fact that for all t, s in T we have

(3.14) $\quad E(M_t|\mathcal{F}_s) = M_{t\wedge s}.$

To see this let A be any $\mathcal{F}_s$ measurable set. Then the random functions I_A and M_t are conditionally independent given $\mathcal{F}_{t\wedge s}$. Thus, we have

(3.15) $\quad E(I_A|\mathcal{F}_{t\wedge s})\cdot E(M_t|\mathcal{F}_{t\wedge s}) = E(I_A M_t|\mathcal{F}_{t\wedge s}).$

The martingale property tells us that $E(M_t|\mathcal{F}_{t\wedge s}) = M_{t\wedge s}$ since $t\wedge s \leq t$. Moreover, since $M_{t\wedge s}$ is $\mathcal{F}_{t\wedge s}$-measurable, we have that

(3.16) $\quad E(I_A|\mathcal{F}_{t\wedge s})M_{t\wedge s} = E(I_A M_{t\wedge s}|\mathcal{F}_{t\wedge s}).$

Substituting (3.16) in (3.15) gives

(3.17) $\quad E(I_A M_{t\wedge s}|\mathcal{F}_{t\wedge s}) = E(I_A M_t|\mathcal{F}_{t\wedge s})$

and taking expectations in (3.17) yields

(3.18) $\quad E(I_A M_{t\wedge s}) = E(I_A M_t).$

Since $A \in \mathcal{F}_s$ in (3.18) was arbitrary and since $M_{t\wedge s}$ is $\mathcal{F}_s$-measurable because $t\wedge s \leq s$, it follows from (3.18) that

(3.14) is true.

Using the result (3.14) we can now prove (3.12) by direct calculation using the formula (3.11) for M_1. To facilitate notation we show (3.12) in the case $i = 1$; the other cases are exactly similar. First rewrite (3.11) as follows:

$$(3.19)\quad M_1 = M_{t_1} + \sum_{m=1}^{n-1} (-1)^{m-1} \Big\{ \sum_{1<i_1<\ldots<i_m\le n} M_{t_{i_1}} \cdots t_{i_m} - \sum_{1<i_2<\ldots<i_{m+1}\le n} M_{t_1} t_{i_2} \cdots t_{i_{m+1}} \Big\}$$

Conditioning (3.19) with respect to $\mathcal{F}_{t_1}$ and using (3.14) gives

$$(3.20)\quad E(M_1|\mathcal{F}_{t_1}) = M_{t_1} + \sum_{m=1}^{n-1} (-1)^{m-1} \Big\{ \sum_{1<i_1<\ldots<i_m\le n} M_{t_1} t_{i_1} \cdots t_{i_m} - \sum_{1<i_2<\ldots<i_{m+1}\le n} M_{t_1} t_{i_2} \cdots t_{i_{m+1}} \Big\}.$$

Noting that the terms in brackets in (3.20) vanish

identically, we see that (3.12) is true. This completes the proof of the theorem. ///

To illustrate the theorem in a simple case, consider the case when T is the two-parameter time set as shown in figure 3.1:

$$T = \{(1,0),\ (0,1),\ (0,0)\}.$$

Then the random M_1 is defined by (3.11) as

$$M_1 = M_{(1,0)} + M_{(0,1)} - M_{(0,0)}.$$

Taking the conditional expectation of M_1 with respect to $F_{(1,0)}$ we obtain from (3.14) that

$$\begin{aligned} E(M_1 | F_{(1,0)}) &= M_{(1,0)} + M_{(0,1)\wedge(1,0)} - M_{(0,0)\wedge(1,0)} \\ &= M_{(1,0)} + M_{(0,0)} - M_{(0,0)} \\ &= M_{(1,0)}. \end{aligned}$$

Note that theorem 3.2 asserts that the martingale version of optional sampling is true with the conditional independence hypothesis because we can imbed the original time set in a directed time set and apply theorem 3.1.

In the next section we see that the conditional independence hypothesis implies that optional sampling is true for submartingales when the parameter set is Z_+^2 or R_+^2. In the submartingale case, however, the results are much more restricted than in the martingale case - even when the conditional independence hypothesis is satisfied, the optional sampling theorem can fail for three-parameter submartingales as we show with a counterexample in the next section.

1.4 Optional Sampling for Two-Parameter Submartingales: Finite Valued Stopping Times

Theorem (Optional sampling)

Suppose that $\{F_t : t \in Z_+^2\}$ is an increasing family of σ-fields satisfying the conditional independence hypothesis. Let σ and τ be stopping times taking finitely many values in Z_+^2 and such that $\sigma \leq \tau$. If $\{M_t : t \in T\}$ is a submartingale with respect to $\{F_t : t \in T\}$, then $M_\sigma \leq E(M_\tau | F_\sigma)$.

The proof of the theorem requires the following lemma about stopping times defined on Z_+^2.

Lemma

Suppose that $\{F_t : t \in Z_+^2\}$ satisfies the conditional

independence hypothesis and let τ be a stopping time with respect to $\{\mathcal{F}_t : t \in Z_+^2\}$. For each $s = (s_1,s_2)$ in Z_+^2 there exist A_s^1 and A_s^2 in $\mathcal{F}_s$ such that $A_s^1 \cap A_s^2 = \emptyset$, $A_s^1 \cup A_s^2 = \Omega$, and also

$$A_s^1 \cap \{\omega : s < \tau(\omega)\} \subset \{\omega : (s_1+1,s_2) \leqq \tau(\omega)\}$$

and

$$A_s^2 \cap \{\omega : s < \tau(\omega)\} \subset \{\omega : (s_1,s_2+1) \leqq \tau(\omega)\}$$

The set $S = \{t : s < t\}$ is divided into the sets $S^1 = \{t : (s_1+1,s_2) \leq t\}$ and $S^2 = \{t : (s_1,s_2+1)\}$ as illustrated in figure 4.1. The lemma says that if $\tau \in S$, then one can determine whether $\tau \in S^1$ or $\tau \in S^2$ on the basis of information in $\mathcal{F}_s$ alone. Thus, on the basis of $\mathcal{F}_s$ alone we can pick a point t which is either (s_1+1,s_2) or (s_1,s_2+1) such that $s < t \leq \tau$. In later chapters we will generalize the property described by this lemma, and in chapter 4 we will define a special class of stopping times (called reachable stopping times) in terms of this property.

Proof:

It suffices to prove the result for $s = (0,0)$; the

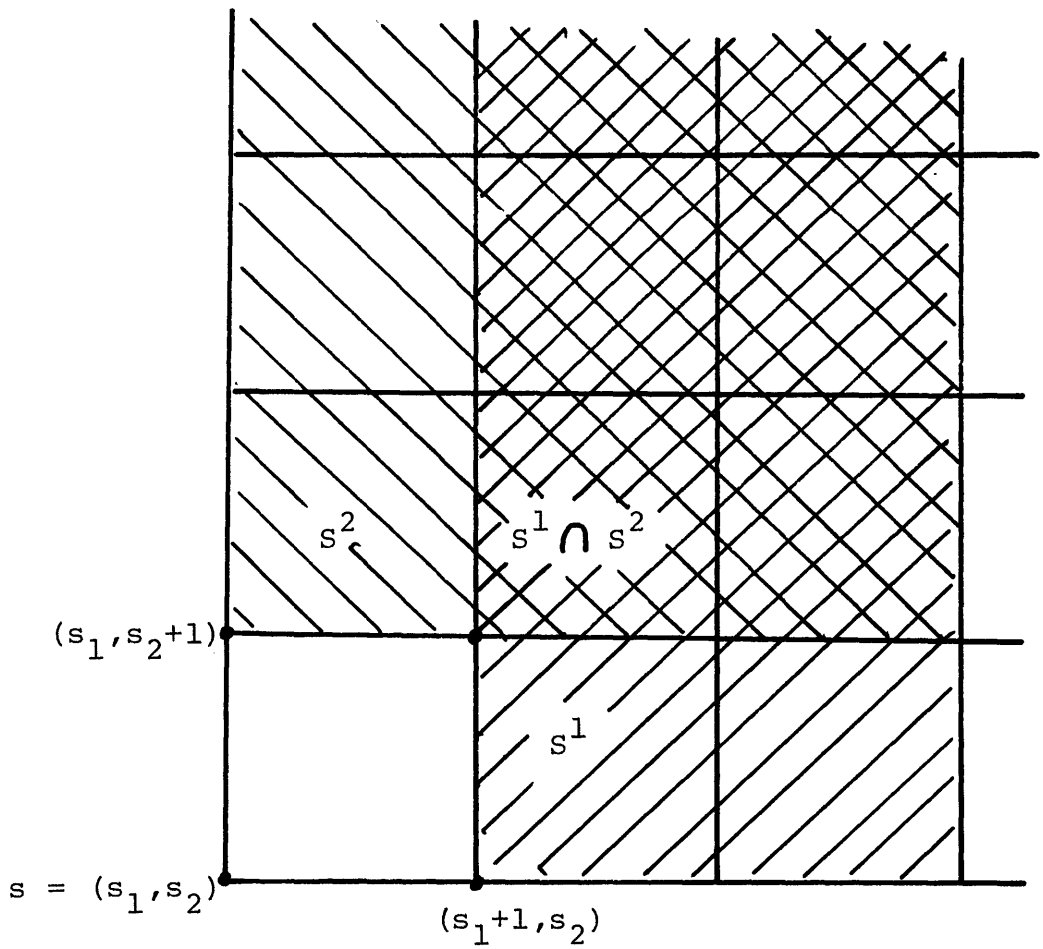


Figure 4.1

other cases are exactly similar. The conditional independence hypothesis implies that $\mathcal{F}_{(0,n)}$ and $\mathcal{F}_{(m,0)}$ are conditionally independent given $\mathcal{F}_{(0,0)}$ for any $m,n \geq 0$. Since $\{\tau = (0,n)\}$ lies in $\mathcal{F}_{(0,n)}$ and $\{\tau = (m,0)\}$ lies in $\mathcal{F}_{(m,0)}$, and since $\{\tau = (0,n)\} \cap \{\tau = (m,0)\} = \emptyset$ if $m > 0$ and $n > 0$, it follows that

$$(4.1) \quad P(\{\tau=(0,n)\}|\mathcal{F}_{(0,0)}) \cdot P(\{\tau=(m,0)\}|\mathcal{F}_{(0,0)}) = 0 \text{ a.s.}$$

for all $n,m > 0$.

Define two sets K_1 and K_2 in such that

$$K_1 = \bigcup_{n>0} \{\tau = (0,n)\} \quad \text{and} \quad K_2 = \bigcup_{m>0} \{\tau = (m,0)\}.$$

Adding up the equations (4.1) for all $m,n > 0$ gives

$$(4.2) \quad P(K_1|\mathcal{F}_{(0,0)}) \cdot P(K_2|\mathcal{F}_{(0,0)}) = 0 \text{ a.s.}$$

Define the set E to be $\{\omega : P(K_1|\mathcal{F}_{(0,0)}) > 0\}$. Since $P(K_1|\mathcal{F}_{(0,0)})$ is an $\mathcal{F}_{(0,0)}$-measurable function, E must be an $\mathcal{F}_{(0,0)}$-measurable set. It follows from (4.2) that for all ω in E, except for a subset of probability zero, $P(K_2|\mathcal{F}_{(0,0)}) = 0$. Thus, one must have

$$P(E \cap K_2) = \int_E P(K_2 | \mathcal{F}_{(0,0)}) dP = 0.$$

Furthermore, it is easy to see that

$$P(E \cap K_1) = \int_E P(K_1 | \mathcal{F}_{(0,0)}) dP = P(K_1).$$

Hence, it must be true that both $E \cap K_2$ and K_1/E have zero probability. Assuming, as we always do, that $\mathcal{F}_{(0,0)}$ contains all null events, then we see that $N = E \cap K_2$ and $N' = K_1/E$ are both in $\mathcal{F}_{(0,0)}$. Define $A^2_{(0,0)}$ as the set $(E/N) \cup N'$ and $A^2_{(0,0)}$ as the set $\Omega/A^2_{(0,0)}$. Since E, N and N' are each in $\mathcal{F}_{(0,0)}$, both $A^1_{(0,0)}$ and $A^2_{(0,0)}$ are in $\mathcal{F}_{(0,0)}$. Moreover, one can rewrite $A^1_{(0,0)}$ and $A^2_{(0,0)}$ in terms of E, K_1 and K_2 as follows:

$$(4.3) \quad A^2_{(0,0)} = (E/K_2) \cup (K_1/E) \subset (\Omega/K_2) \cup K_1,$$

$$(4.4) \quad A^1_{(0,0)} = [((\Omega/K_1) \cap \Omega/(E/K_2)) \cup (E \cap K_2)] \subset (\Omega/K_1) \cup K_2.$$

Recall how K_1 and K_2 were defined above. If ω lies in $A^2_{(0,0)}$, then (4.3) tells us that either $\omega \in K_1$ or $\omega \notin K_2$. In the former case, $\tau(\omega) = (0,n)$ for some $n > 0$, and thus, $(0,1) \leq \tau(\omega)$ in this case. If $\omega \notin K_2$ and in addition $\tau(\omega) \neq (0,0)$, then $\tau(\omega) \neq (m,0)$ for any

$m > 0$ and $m = 0$. Consequently, $(0,1) \leqq \tau(\omega)$ is true in this case also and we have proved

$$A^2_{(0,0)} \cap \{(0,0) < \tau\} \subset \{(0,1) \leqq \tau\}.$$

A similar argument using (4.4) and switching the roles of K_1 and K_2 in the above argument gives the other result,

$$A^1_{(0,0)} \cap \{(0,0) < \tau\} \subset \{(1,0) \leqq \tau\}.$$

This completes the proof of the lemma.

Before proving the optional sampling theorem, let us discuss the problems involved in extending the one-parameter proof to the two-parameter situation. Suppose that σ and τ are stopping times in Z^2_+ and $\sigma \leqq \tau$. Let A be in $\mathcal{F}_\sigma$ and let $B = A \cap \{\sigma=s\}$ so that B lies in $\mathcal{F}_s$ and $s \leqq \tau(\omega)$ for ω in B. We want to prove that

$$(4.5) \quad \int_B M_s \, dP \leqq \int_B M_\tau \, dP.$$

To understand the two-parameter proof, first consider the one-parameter case. For the moment suppose in (4.5)

that s is in Z_+ instead of Z_+^2 and suppose that τ is a one-parameter stopping time in Z_+. To prove (4.5) in this case, rewrite the left hand side as

$$(4.6) \quad \int_B M_s \, dP = \int_{B \cap \{s=\tau\}} M_\tau \, dP + \int_{B \cap \{s+1 \leq \tau\}} M_s \, dP.$$

Since $\{s+1 \leq \tau\}$ lies in $\mathcal{F}_s$, the submartingale property implies that

$$(4.7) \quad \int_{B \cap \{s+1 \leq \tau\}} M_s \, dP \leq \int_{B \cap \{s+1 \leq \tau\}} M_{s+1} \, dP \ ,$$

and hence,

$$(4.8) \quad \int_B M_s \, dP \leq \int_{B \cap \{s=\tau\}} M_\tau \, dP + \int_{B \cap \{s+1 \leq \tau\}} M_{s+1} \, dP \ .$$

Proceeding by induction, one can then show that

$$(4.9) \quad \int_B M_s \, dP \leq \int_{B \cap \{s \leq \tau < s+n\}} M_\tau \, dP + \int_{B \cap \{s+n+1 \leq \tau\}} M_{s+n+1} \, dP$$

for all integers $n \geq 0$. By letting $n \to \infty$ one obtains (4.5).

Unlike the one-parameter situation, (4.7) is not true for two-parameters and thus, one cannot go from (4.6) to (4.8) and by induction, to (4.9). The problem is that for one-parameter stopping times if $\tau \not\leq s$ then necessarily

$s+1 \leqq \tau$ so that $\{s+1 \leqq \tau\}$ is always $\mathcal{F}_s$-measurable, and thus, (4.7) follows. But for two-parameter times if $s \neq \tau$ (assuming $s \leqq \tau$ to begin with) then either $(s_1,s_2+1) \leqq \tau$ or $(s_1+1,s_2) \leqq \tau$. Neither $\{(s_1+1,s_2) \leqq \tau\}$ nor $\{(s_1,s_2+1) \leqq \tau\}$ is necessarily $\mathcal{F}_s$-measurable and (4.7) is not generally true. The counterexample in section 3 is an example for which this happens.

However, if one assumes the conditional independence hypothesis, then the preceding lemma allows one to replace $\{(s_1+1,s_2) \leqq \tau\}$ by A_s^1 and $\{(s_1,s_2+1) \leqq \tau\}$ by A_s^2 in the two-parameter version of (4.7). Since A_s^1 and A_s^2 are $\mathcal{F}_s$-measurable, the argument can proceed.

Proof of the optional sampling theorem:

Let B and s be fixed the same as in (4.5), which equation we shall prove. For each t in Z_+^2 let $|t|$ denote the nonnegative integer t_1+t_2. We now construct sets B_t in $\mathcal{F}$ for each t for which $s \leqq t$ such that the B_t have the following properties:

$$(4.10) \quad B_s = B \cap \{s \neq \tau\} = B \cap \{s < \tau\}$$

$$(4.11) \quad B_t \cap B_{t'} = \emptyset \quad \text{if} \quad |t-s| = |t'-s| \quad \text{and} \quad |t \neq t'|$$

where both $s \leqq t$ and $s \leqq t'$

(4.12) $B_t \subset \{t < \tau\}$ for each t

(4.13) B_t is $\mathcal{F}_t$-measurable

(4.14) $B \cap \{|\tau-s| > n\} = \bigcup\{B_t : |t-s| = n,\ s \leq t\}$

(4.15) $$\int_B M_s \, dP \leq \int_{B \cap \{|\tau-s| \leq n\}} M_\tau \, dP + \sum_{\substack{|t-s|=n \\ s \leq t}} \int_{B_t} M_t \, dP$$

where $n \geq 0$ in (4.14) and (4.15). Note that (4.15) is the two-parameter version of (4.9). Once we have (4.14) and (4.15), the optional sampling theorem is finished. For then, since τ takes only finitely many values, the set $\{|\tau-s| > n\} = \emptyset$ for some n sufficiently large. Thus $B_t = \emptyset$ for all t such that $|t-s| = n$. Likewise, B is contained in $\{|\tau-s| \leq n\}$. Then equation (4.15) becomes (4.5), which is the same as

$$\int_{A \cap \{\sigma=s\}} M_\sigma dP \leq \int_{A \cap \{\sigma=s\}} M_\tau dP.$$

Since σ takes only finitely many values, there is no problem summing over s in Z_+^2 and obtaining the answer

$$\int_A M_\sigma dP \leq \int_A M_\tau dP$$

for all A in $\mathcal{F}_\sigma$. In other words, $M_\sigma \leq E(M_\tau | \mathcal{F}_\sigma)$.

To prove the existence of the sets $\{B_t : s \leq t\}$, proceed by induction on $n = |t-s|$. To start, the set B_s is defined as $B \cap \{s \neq \tau\}$, which clearly satisfies (4.10) through (4.15). This is the n=0 step of the induction. Suppose that we have found B_t for all t such that $|t-s| \leq n$ and we want to find B_r for r such that $|r-s| = n+1$. For each t such that $s \leq t$ and $|t-s| = n$ define the sets B_t^1 and B_t^2 as $B_t^1 = B_t \cap A_t^1$ and $B_t^2 = B_t \cap A_t^2$. If $s \not\leq t$ but $|t-s| = n$, define $B_t^1 = B_t^2 = \emptyset$. The sets A_t^1 and A_t^2 are the ones appearing in the lemma and depend on the stopping time τ. The sets B_t^1 and B_t^2 inherit the following properties from A_t^1 and A_t^2: $B_t^1 \quad B_t^2 = \emptyset$; $B_t^1 \quad B_t^2 = B_t$; $t + (1,0) \leq \tau(\omega)$ for ω in B_t^1 and $t + (0,1) \leq \tau(\omega)$ for ω in B_t^2; B_t^1 and B_t^2 belong to F_t. In order to reduce the notation in the proof let α denote $(1,0)$ and let β denote $(0,1)$ in Z_+^2.

For each r such that $s \leq r$ and $|r-s| = n+1$ define B_r as $B_r = (B_{r-\alpha}^1 \cup B_{r-\beta}^2)/\{\tau=r\}$. Now check that these B_r satisfy (4.11) through (4.15).

Suppose $s \leq r,r'$ and $|r-s| = |r'-s| = n$ but $r \neq r'$. From the definition of B_r and $B_{r'}$ it is easy to see that $B_r \subset B_{r-\alpha}^1 \cup B_{r-\beta}^2$ and $B_{r'} \subset B_{r'-\alpha}^1 \cup B_{r'-\beta}^2$. Since

(3.11) is true if $|t-s| = |t'-s| = n$, it follows that $B_t^i \cap B_{t'}^j = \emptyset$ if $t \neq t'$ for $i,j = 1,2$. If $t = t'$, then $B_t^1 \cap B_{t'}^2 = B_t^1 \cap B_t^2 = \emptyset$ since B_t^1 and $B_{t'}^2$ were constructed to be mutually disjoint. Let t,t' take the values $r-\alpha$, $r-\beta$, $r'-\alpha$, $r'-\beta$ we find that $(B^1_{r-\alpha} \cup B^2_{r-\beta}) \cap (B^1_{r'-\alpha} \cup B^2_{r'-\beta}) = \emptyset$ and hence, $B_r \cap B_{r'} = \emptyset$. This proves (4.11) for the induction step.

Since (4.12) is true if $|t-s| = n$ and since $|r-\alpha-s| = |r'-\beta-s| = n$ if $|r-s| = n+1$, it must happen that $r \leq \tau(\omega)$ for ω in $B^1_{r-\alpha}$ or $B^2_{r-\beta}$. It easily follows that $r < \tau(\omega)$ for ω in B_r. This proves (4.12).

It is easy to see that

$$B \cap \{|\tau-s| > n+1\} \supset \bigcup \{B_r : s \leq r, |r-s| = n+1\}.$$

We must now show the inclusion the other way to prove equation (4.14). If ω lies in $B \cap \{|\tau-s| > n+1\}$ then ω lies in $B \cap \{|\tau-s| > n\}$ also. Since (4.14) is true for n,

$$B \cap \{|\tau-s| > n\} = \bigcup \{B_t : |t-s| = n, s \leq t\}$$

and hence, ω lies in B_t for some t for which $s \leq t$ and $|t-s| = n$. From (4.12) it follows that $t < \tau(\omega)$ so that either $t+\alpha \leq \tau(\omega)$ or $t+\beta \leq \tau(\omega)$. However, $|\tau(\omega) - s| > n+1$ by assumption so that equality cannot occur in either case -- that is, either $t+\alpha < \tau(\omega)$ or $t+\beta < \tau(\omega)$. Thus, either ω is in $B_t^1/\{\tau = t+\alpha\}$ or it is in $B_t^2/\{\tau = t+\beta\}$ and hence ω lies in either $B_{t+\alpha}$ or $B_{t+\beta}$. Therefore,

$$B \cap \{|\tau-s| > n+1\} \subset \bigcup\{B_r : s \leq r, |r-s| = n+1\}$$

and (4.14) is true for the induction step.

Finally, to show (4.15) note that if $s \leq t$ and $|t-s| = n$, then

$$\int_{B_t} M_t \, dP = \int_{B_t^1} M_t \, dP + \int_{B_t^2} M_t \, dP$$

$$\leq \int_{B_t^1} M_{t+\alpha} \, dP + \int_{B_t^2} M_{t+\beta} \, dP.$$

The last inequality is a consequence of the submartingale property. Rewrite the last two integrals as

$$\int_{B_t^1} M_{t+\alpha} dP = \int_{B_t^1 \cap \{\tau = t+\alpha\}} M_\tau dP + \int_{B_t^1 / \{\tau = t+\alpha\}} M_{t+\alpha} dP$$

$$\int_{B_t^2} M_{t+\beta} dP = \int_{B_t^2 \cap \{\tau = t+\beta\}} M_\tau dP + \int_{B_t^2 / \{\tau = t+\beta\}} M_{t+\beta} dP.$$

The set $B_t^1 \cap \{\tau = t+\alpha\}$ is the same as $B_t^1 \cap \{|\tau - s| = n+1\}$ and likewise $B_t^2 \cap \{\tau = t+\beta\}$ is $B_t^2 \cap \{|\tau - s| = n+1\}$. Putting these facts together one finds that

$$\int_{B_t} M_t \, dP \leq \int_{(B_t^1 \cup B_t^2) \cap \{|\tau - s| = n+1\}} M_\tau \, dP$$

$$+ \int_{B_t^1 / \{\tau = t+\alpha\}} M_{t+\alpha} \, dP + \int_{B_t^2 / \{\tau = t+\beta\}} M_{t+\beta} dP$$

or

$$\int_{B_t} M_t \, dP \leq \int_{B_t \cap \{|\tau - s| = n+1\}} M_\tau \, dP$$

$$+ \int_{B_t^1 / \{\tau = t+\alpha\}} M_{t+\alpha} dP + \int_{B_t^2 / \{\tau = t+\beta\}} M_{t+\beta} dP.$$

Adding the inequality over all t such that $s \leq t$ and $|t-s| = n$ gives

$$\sum_{\substack{|t-s|=n \\ s \leqq t}} \int_{B_t} M_t \, dP = \int_{B \cap \{|\tau-s|>n\} \cap \{|\tau-s|=n+1\}} M_\tau \, dP$$

$$+ \sum_{\substack{|t-s|=n \\ s \leqq t}} \int_{B_t} M_t \, dP.$$

Since $\{|\tau-s| > n\} \cap \{|\tau-s| = n+1\} = \{|\tau-s| = n+1\}$ and since $(B \cap \{|\tau-s| \leqq n\}) \cup (B \cap \{|\tau-s| = n+1\}$ is equal to $B \cap \{|\tau-s| \leqq n+1\}$ we obtain (4.15) for $n+1$ by adding this last inequality to (4.15) for n. This completes the induction argument for constructing B_t for all t such that $s \leqq t$, and the proof is finished.

QED

Although one can easily extend the conditional independence hypothesis to R^n time sets and even to more general p.o. time sets as described in section 1.2, the optional sampling theorem is not true in these more general sitatuions. The following example shows that the result fails for three-parameter submartingales even when conditional independence holds.

Example.

We construct an example for the time set T defined as the Cartesian product $\{0,1\} \times \{0,1\} \times \{0,1\}$, a subset of Z_+^3, namely the vertices of the unit cube. One could

easily extend this example to one on all of Z_+^3 or to Z_+^n for some $n > 3$. Let T be partially ordered with the coordinate-wise ordering of Z_+^3 analogous to the partial ordering we defined on Z_+^2.

Let $\Omega = \{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6,\omega_7,\omega_8\}$ with $P(\{\omega_i\}) = \frac{1}{8}$ for each i, and let $\mathcal{F}$ be the collection of all subsets of Ω. $\mathcal{F}_{(0,0)}$ is taken to be $\{\emptyset,\Omega\}$ and the other σ-fields are defined by their atoms.

$\mathcal{F}_{(1,0,0)}$ is generated by the sets

$\{\omega_1,\omega_2,\omega_3,\omega_4\}$, $\{\omega_5,\omega_6,\omega_7,\omega_8\}$.

$\mathcal{F}_{(0,1,0)}$ is generated by

$\{\omega_1,\omega_2,\omega_5,\omega_6\}$, $\{\omega_3,\omega_4,\omega_7,\omega_8\}$.

$\mathcal{F}_{(0,0,1)}$ is generated by

$\{\omega_1,\omega_3,\omega_5,\omega_7\}$, $\{\omega_2,\omega_4,\omega_6,\omega_8\}$.

$\mathcal{F}_{(1,1,0)}$ is generated by

$\{\omega_1,\omega_2\}$, $\{\omega_3,\omega_4\}$, $\{\omega_5,\omega_6\}$, $\{\omega_7,\omega_8\}$.

$\mathcal{F}_{(1,0,1)}$ is generated by

$\{\omega_1,\omega_3\}$, $\{\omega_2,\omega_4\}$, $\{\omega_5,\omega_7\}$, $\{\omega_6,\omega_8\}$.

$\mathcal{F}_{(0,1,1)}$ is generated by

$\{\omega_1,\omega_5\}$, $\{\omega_2,\omega_6\}$, $\{\omega_3,\omega_7\}$, $\{\omega_4,\omega_8\}$.

Finally, $\mathcal{F}_{(1,1,1)}$ is the same as $\mathcal{F}$.

One can think of these σ-fields being generated by three independent random variables taking on the values 0 or 1 with equal probability. Thus, define the random variables α,β,γ as follows:

$$\alpha(\omega_1) = \alpha(\omega_2) = \alpha(\omega_3) = \alpha(\omega_4) = 0,$$

$$\alpha(\omega_5) = \alpha(\omega_6) = \alpha(\omega_7) = \alpha(\omega_8) = 1,$$

$$\beta(\omega_1) = \beta(\omega_2) = \beta(\omega_5) = \beta(\omega_6) = 0,$$

$$\beta(\omega_3) = \beta(\omega_4) = \beta(\omega_7) = \beta(\omega_8) = 1,$$

$$\gamma(\omega_1) = \gamma(\omega_3) = \gamma(\omega_5) = \gamma(\omega_7) = 0,$$

$$\gamma(\omega_2) = \gamma(\omega_4) = \gamma(\omega_6) = \gamma(\omega_8) = 1.$$

Then $\mathcal{F}_{(1,0,0)}$ is generated by α, $\mathcal{F}_{(0,1,0)}$ is generated by β, $\mathcal{F}_{(0,0,1)}$ is generated by γ, $\mathcal{F}_{(1,1,0)}$ is generated by α and β, $\mathcal{F}_{(1,0,1)}$ is generated by α and γ, $\mathcal{F}_{(0,1,1)}$ is generated by β and γ, and $\mathcal{F}_{(1,1,1)}$ is generated by all three random

variables. Since α, β and γ are independent, it is easy to check that $\{\mathcal{F}_t : t \in T\}$ satisfies the conditional independence hypothesis for Z_+^3.

Define $M_{(1,1,1)}(\omega_3) = M_{(1,1,1)}(\omega_6) = -1$ and let $M_{(1,1,1)}(\omega_i) = 1$ for $i \neq 3,6$. Define $M_t = 0$ if $t \neq (1,1,1)$. To check that $\{M_t : t \in T\}$ is a submartingale it suffices to show that $E(M_{(1,1,1)} | \mathcal{F}_t) \geq 0$, for $t = (1,1,0), (1,0,1), (0,1,1)$. A simple calculation shows that $E(M_{(1,1,1)} | \mathcal{F}_{(1,1,0)})$ is equal to 1 if $\omega = \omega_1, \omega_2, \omega_7, \omega_8$ and it is equal to 0 if $\omega = \omega_3, \omega_4, \omega_5, \omega_6$. The other conditional expectations are similar. Finally, define $\tau(\omega) = (1,1,0)$ if $\omega = \omega_1, \omega_2$, let $\tau(\omega) = (1,0,1)$ if $\omega = \omega_5, \omega_7$, let $\tau(\omega) = (0,1,1)$ if $\omega = \omega_4, \omega_8$, and let $\tau(\omega) = (1,1,1)$ if $\omega = \omega_3, \omega_6$. It is easy to check that τ is a stopping time on T. Let $\sigma = (0,0,0)$, then

$$E(M_\tau | \mathcal{F}_\sigma) = E(M_\tau) = (-1)\cdot\frac{1}{4} + (0)\cdot\frac{3}{4} = -\frac{1}{4} \not\geq 0 = M_\sigma.$$

Note that the 3-parameter analogue of the lemma of section 3 is not true in this example. Since $\mathcal{F}_{(0,0,0)}$ is $\{\emptyset, \Omega\}$ there cannot be three sets in $\mathcal{F}_{(0,0,0)}$ denoted $A^1_{(0,0,0)}$, $A^2_{(0,0,0)}$, $A^3_{(0,0,0)}$ which partition Ω

and such that

$$A^1_{(0,0,0)} \subset \{(1,0,0) \leqq \tau\},$$

$$A^2_{(0,0,0)} \subset \{(0,1,0) \leqq \tau\},$$

$$A^3_{(0,0,0)} \subset \{(0,0,1) \leqq \tau\}.$$

This cannot occur because one of the sets $A^i_{(0,0,0)}$ must be Ω and none of the sets $\{t \leqq \tau\}$ is Ω for $(0,0,0) < t$. Thus, the optional sampling theorem is not true even though the σ-fields satisfy the conditional independence hypothesis for Z^3_+.

1.5 Optional Sampling: Stopping Times in Z^2_+

Most of the work is already done since the construction of the sets B_t in the previous theorem does not require that σ or τ have finitely many values. However, other conditions are necessary to make sure that (4.15) converges to the result

$$\int_B M_s \, dP \leqq \int_B M_\tau \, dP$$

as $n \to \infty$.

Theorem (optional sampling for countable times)

Suppose $\{F_t : t \in Z^2_+\}$ satisfies the conditional

independence hypothesis and suppose that σ and τ are stopping times on Z_+^2 with $\sigma \leq \tau$. Suppose that $\{M_t : t \in Z_+^2\}$ is a submartingale with respect to $\{F_t : t \in Z_+^2\}$. For each integer $n \geq 0$, define M_n to be the random variable $\max \{M_t : t \in Z_+^2, |t| = n\}$. Assume that the following conditions hold:

(5.1) $E(|M_\sigma|) < \infty$ and $E(|M_\tau|) < \infty$,

(5.2) $$\liminf_{n \to \infty} \int_{\{|\tau| > n\}} M_n \, dP = 0.$$

Then it is true that $E(M_\tau | F_\sigma) \geq M_\sigma$.

Proof:

Suppose that A is in F_σ and s is in Z_+^2 and define B as the set $A \cap \{\sigma = s\}$. As in the theorem for finite stopping times, there are sets B_t for each t for which $s \leq t$ such that (4.10) through (4.15) are true. In particular, (4.15) states that

$$\int_B M_s \, dP \leq \int_{B \cap \{|\tau - s| \leq n\}} M_\tau \, dP + \sum_{\substack{|t-s|=n \\ s \leq t}} \int_{B_t} M_t \, dP$$

for each integer $n \geq 0$. The set $\{|\tau - s| \leq n\} \cap B$ increases to B as $n \to \infty$. Since $E(|M_\tau|)$ is finite,

dominated convergence implies that

$$(5.3)\qquad \lim_{n \to \infty} \int_{B \cap \{|\tau - s| \leq n\}} M_\tau \, dP = \int_B M_\tau \, dP.$$

Suppose that $|t-s| = n$ and $s \leq t$. Then $|t| = n + |s|$ and by definition of $M_{n+|s|}$,

$$M_t \leq M_{n+|s|}$$

for all these t. Thus, it follows that

$$\sum_{\substack{|t-s|=n \\ s \leq t}} \int_{B_t} M_t \, dP \leq \int_{B \cap \{|\tau - s| > n\}} M_{n+|s|} \, dP.$$

To obtain the right side, use (4.11) and (4.14). For all ω in B, it happens that $s \leq \tau(\omega)$ so that we have $B \cap \{|\tau - s| > n\} = B \cap \{|\tau| > |s| + n\}$. Condition (5.2) implies that for some subsequence $n(k)$ of the positive integers,

$$\lim_{k \to \infty} \int_{\{|\tau| > n(k)\}} |M_{n(k)}| \, dP = 0.$$

Thus, if $n'(k) = n(k) - |s|$, it follows that

$$\lim_{k \to \infty} \int_{B \cap \{|\tau| > |s| + n'(k)\}} M_{|s|+n'(k)} \, dP = 0.$$

The inequality in (4.15) is bounded above as follows:

$$\int_B M_s dP \leq \int_{B \cap \{|\tau - s| \leq n\}} M_\tau \, dP + \int_{B \cap \{|\tau| > n + |s|\}} M_{n+|s|} \, dP.$$

If n approaches infinite through the subsequence $n'(k)$, then this inequality together with (5.3) implies the result,

$$\int_B M_s \, dP \leq \int_B M_\tau \, dP$$

or

$$\int_{A \cap \{\sigma = s\}} M_\sigma \, dP \leq \int_{A \cap \{\tau = s\}} M_\tau \, dP.$$

Since $E(|M_\sigma|)$ and $E(|M_\tau|)$ are finite, dominated convergence allows one to add over all s in Z_+^2 to finish the proof. QED

Note that the conditions (5.1) and (5.2) are similar to the usual one-parameter assumptions. In particular, in the one-parameter case, if $t = n$, then

$$M_t = \text{Max } \{M_s : s \in Z_+^1, |s| = n\}$$

and condition (5.2) is

$$\liminf_{t \to \infty} \int_{\{\tau > t\}} M_t \, dP = 0.$$

1.6 Optional Sampling: Stopping Times in R_+^2

One proceeds more or less as for R_+^1 -- that is, by assuming right continuity and taking limits of stopping times taking a finite number of values. We consider here only the continuous version of section 4. Similarly, the continuous version of section 5 would follow from the discrete version in section 5 after assuming condition (5.1) and condition (5.2) for all positive real numbers n.

Theorem (optional sampling for R_+^2)

Suppose that $\{F_t : t \in R_+^2\}$ satisfies the conditional independence hypothesis. Let σ and τ be stopping times on R_+^2 with $\sigma \leqq \tau \leqq t_0$ for some t_0 in R_+^2. If $\{M_t : t \in R_+^2\}$ is a right continuous submartingale with respect to $\{F_t : t \in R_+^2\}$ and $E(|M_\sigma|)$ and $E(|M_\tau|)$ are finite, then $E(M_\tau | F_\sigma) \geqq M_\sigma$.

Proof:

The proof follows the one-parameter case closely. Let M_t^a be defined as $\max\{M_t, a\}$ for all a in R and all t in R_+^2. For any t in R_+^2 define $[t]_n^+$ and $[t]_n^-$ as

$$[t]_n^+ = \min\{q2^{-n} : q \in Z_+^2,\ t \leqq q2^{-n}\}$$

and

$$[t]_n^- = \max \{q2^{-n} : q \in Z_n^+,\ q2^{-n} \leq t\}.$$

The minimum and maximum are taken with respect to the partial ordering of R_+^2, and it is not hard to see that both exist in this case. If $s \leq t$, then $[s]_n^+ \leq [t]_n^+$. Also, $[s]_{n+1}^+ \leq [s]_n^+$ for all s in R_+^2, and for all s and t in R_+^2, $[s]_n^+ \leq t$ if and only if $s \leq [t]_n^-$.

Let $\tau_n(\omega) = [\tau(\omega)]_n^+$ and $\sigma_n(\omega) = [\sigma(\omega)]_n^+$ and assume without loss of generality that t_0 is in $Z_+^2 \cap R_+^2$ so that $[t_0]_n^+ = t_0$ for all n. For each n, $\sigma_n \leq \tau_n \leq t_0$, and for all ω as $n \to \infty$, both $\sigma_n(\omega)$ and $\tau_n(\omega)$ decrease monotonically to $\sigma(\omega)$ and $\tau(\omega)$ respectively. Furthermore, σ_n and τ_n are stopping times. To see this, note that

$$\{\sigma_n \leq s\} = \{[\sigma]_n^+ \leq s\} = \{\sigma \leq [s]_n^-\} \in F_{[s]_n^-} \subset F_s.$$

Next, note that $\{M_t^a : t \in R_+^2\}$ is a submartingale with respect to $\{F_t : t \in R_+^2\}$. If A is in F_s and $s \leq t$, then

$$\int_A M_s^a \, dP = \int_{A \cap \{M_s < a\}} a \, dP + \int_{A \cap \{M_s \geq a\}} M_s \, dP.$$

Since $A \cap \{M_s \geq a\}$ lies in $\mathcal{F}_s$, we have

$$\int_{A \cap \{M_s \geq a\}} M_s \, dP \leq \int_{A \cap \{M_s \geq a\}} M_t \, dP \leq \int_{A \cap \{M_s \geq a\}} M_t^a \, dP.$$

Since it is also clear that

$$\int_{A \cap \{M_s < a\}} a \, dP \leq \int_{A \cap \{M_s < a\}} M_t^a \, dP$$

it follows that

$$\int_A M_s^a \, dP \leq \int_A M_t^a \, dP.$$

Applying the optional sampling theorem of section 4 to σ_n and τ_n, we have that for A in $\mathcal{F}_\sigma$, A is also in $\mathcal{F}_{\sigma_n}$ (since $\sigma \leq \sigma_n$), and hence,

$$\int_A M_{\sigma_n}^a \, dP \leq \int_A M_{\tau_n}^a \, dP. \tag{6.1}$$

To obtain the desired result, one takes the limit in (6.1) as $n \to \infty$, but in order to do this, one must first show that $\{M_{\sigma_n}^a\}$ and $\{M_{\tau_n}^a\}$ are uniformly integrable.

Let $\mathcal{T}$ be the class of all stopping times ρ with respect to $\{\mathcal{F}_t : t \in R_+^2\}$ such that $\rho \leq t_0$ and

ρ takes finitely many values. We show for each real a that $\{M_\rho^a : \rho \in T\}$ is uniformly integrable. If $\{M_t\}$ is is a submartingale, then so is $\{M_t - a\}$. Since M_t^a is just $a + \max\{M_t - a, 0\}$, it suffices to show that $\{M_\rho^a : \rho \in T\}$ is uniformly integrable for $a = 0$. As usual, let M_t^+ denote the maximum of M_t and 0. Since $\{M_t^+ : t \in R_+^2\}$ is a submartingale, the optional sampling theorem for finite stopping times asserts that for all ρ in T and all A in $\mathcal{F}_\rho$,

$$\int_A M_\rho^+ \, dP \leq \int_A M_{t_0}^+ \, dP.$$

If $A = \{M_\rho > c\}$ for a real number c, then the following inequality results:

$$c\, P(\{M_\rho^+ > c\}) \leq \int_{\{M_\rho^+ > c\}} M_\rho \, dP \leq \int_{\{M_\rho^+ > c\}} M_{t_0}^+ \, dP.$$

Thus, we have

$$\lim_{c \to \infty} (\sup \{P(\{M_\rho^+ > c\}) : \rho \in T\}) = 0$$

and since M_{t_0} and hence, $M_{t_0}^+$ is integrable,

$$\lim_{c \to \infty} (\sup \{\int_{\{M_\rho^+ > c\}} M_{t_0}^+ \, dP : \rho \in T\}) = 0;$$

It follows that

$$\lim_{c \to \infty} \left(\sup \left\{ \int_{\{M_\rho > c\}} M_\rho^+ \, dP : \rho \in T \right\}\right) = 0$$

and hence, $\{M_\rho^+ : \rho \in T\}$ is uniformly integrable. Consequently, $\{M_\rho^a : \rho \in T\}$ will be uniformly integrable for any real number a.

Because $t \to M_t$, and hence $t \to M_t^a$ is right continuous, and since $\sigma_n \downarrow \sigma$ and $\tau_n \downarrow \tau$, it is clear that

$$\lim_{n \to \infty} M_{\sigma_n}^a = M_\sigma^a \quad \text{a.s.}$$

and

$$\lim_{n \to \infty} M_{\tau_n}^a = M_\tau^a \quad \text{a.s.}$$

The uniform integrability allows one to conclude from this fact and (6.1) that

$$\int_A M_\sigma^a \, dP \leq \int_A M_\tau^a \, dP$$

for all A in F_σ and all a in R. Noting that M_σ and M_τ are integrable and using dominated convergence as $a \to -\infty$, one obtains the result

$$\int_A M_\sigma \, dP \leq \int_A M_\tau \, dP.$$

QED

Section 7. Significance of the failure of the optional sampling theorem

We have shown by simple counterexamples that the optional sampling theorem is not true generally for p.o. time sets. However, in the special case of two-parameter submartingales which have a special underlying probability structure (e.g. such as that generated by the two-parameter Wiener process), the theorem is true. Nevertheless, the optional sampling theorem breaks down when the number of parameters increases to three, despite the assumption of conditional independence.

The failure of the optional sampling theorem indicates that the stopping time and submartingale theory of p.o. time sets is radically different from the one-parameter theory. For example, the failure of the optional sampling theorem implies the lack of a Doob decomposition[1] of submartingales indexed by p.o. time. Suppose that $X_t = M_t + A_t$ where M_t is a martingale and A_t is an <u>increasing process</u> in

[1] See Doob (1953) for the discrete time theorem, or Meyer (1966) for the continuous time case.

the sense that $A_t \geq A_s$ if $s \leq t$ in the p.o. time set; that is, suppose X_t has a Doob decomposition. Then, assuming that the p.o. time set is directed, we see that Kurtz's results imply that $E(M_\tau | \mathcal{F}_\sigma) = M_\sigma$ for for stopping times τ and σ such that $\sigma \leq \tau$. It is clear that $A_\tau \geq A_\sigma$ almost surely, and hence, $E(A_\tau | \mathcal{F}_\sigma) \geq A_\sigma$. Thus, $E(X_\tau | \mathcal{F}_\sigma) \geq X_\sigma$, and the optional sampling theorem is true for X_t if X_t has a Doob decomposition. Consequently, if the optional sampling theorem is not true for a submartingale X_t, then X_t cannot have a Doob decomposition.

The remaining chapters approach the problem of extending the optional sampling theorem by restricting the class of stopping times. The motivation for this particular restriction of the class of stopping times comes from two different directions:

(1) it is possible to restrict stopping times so that the lemma of section 1.4 is true for more general p.o. time sets;

(2) for a class of stopping problems for processes evolving on p.o. sets, it is more "realistic" to consider a restricted class of stopping times.

We define the new class of stopping times (called <u>reachable</u>) in terms of a <u>decision function</u> which we will

define generally in section 3.2. As we show in chapter 3, our decision function is the same as Haggstrom's (1966) control variable when the p.o. time set is a special kind of p.o. set called a tree. In chapter 4 we formally define reachable stopping times and we show that the submartingale version of the optional sampling theorem is true for this class of stopping times. We also discover that in the case of two-parameter σ-fields satisfying the conditional independence hypothesis, all stopping times are reachable. Thus, the general result of chapter 4 contains the result of section 1.4 as a special case. Finally, we show that in a certain sense reachable stopping times are the only stopping times that satisfy the submartingale version of the optional sampling theorem. Namely, we show that if τ,σ are fixed stopping times with respect to $\{F_t : t \in T\}$ and such that $\tau \geq \sigma$, then $E(M_\tau | F_\sigma) \geq M_\sigma$ for all submartingales $\{M_t : t \in T\}$ with respect to $\{F_t : t \in T\}$ if and only if τ is reachable.

CHAPTER 2

SIGNAL INTERCEPTION EXAMPLE

2.1 Introduction

In this chapter and the succeeding ones, chapters 3 and 4, we are going to approach the optional sampling theorem from a different standpoint than in chapter 1. Instead of considering arbitrary stopping times and σ-fields with a special structure, we are going to consider arbitrary increasing families of σ-fields and a special class of stopping times. We will call these stopping times reachable and define them formally in chapter 4. Moreover, we will show there that the optional sampling theorem is true for all submartingales if the stopping times are reachable, and conversely, if the optional sampling theorem is true for all submartingales for some fixed stopping times, then these stopping times must be reachable.*

To motivate the definition of reachable stopping time presented in chapter 4, we use the present chapter and chapter 3 to introduce a large class of stopping problems for sequential processes evolving in p.o. sets. In this chapter we consider only a simple example of the

---

*See chapter 4 generally and section 4.5 in particular for the precise statement.

general class of stopping problems in order to indicate how the p.o. set framework can model physical constraints imposed on a problem. The next chapter will consider the general class of stopping problems and it will indicate how many different problems fit into the general framework.

In the course of developing the general class of stopping problems, we will also show how the p.o. set formulation gives meaning to the multiparameter generalization of causality, recursive computation and dynamic programming. For one-parameter systems the concept of casuality, or nonanticipation as it is also called, is a powerful one. In solving optimization problems it permits the use of Bellman's (1957) dynamic programming; in calculation it often permits efficient recursive computation. For multiparameter systems there is generally no concept of causality, nor are methods of dynamic programming and recursive computation generally applicable in multiparameter problems. However, in this chapter and the next we find that the class of stopping problems considered naturally defines a generalization of the one-parameter concepts of causality, recursive computation and dynamic programming to the multiparameter case (in fact, to the case of p.o. parameter systems).

We briefly summarize the contents of chapter 2.

In words, the simple example considered is the following. A transmitter (the signal source) is located at an unknown position represented by ξ. It begins transmitting at an unknown time represented by θ. From this time on the transmitter radiates its signal continuously in all directions and the signal travels at a constant speed denoted by c. A mobile receiver starts from a known location in space at a given starting time and travels with speed at most c. If the prior joint distribution of ξ and θ is known, how should the receiver travel to intercept the signal in the least expected amount of time?

Section 2 formulates this example (the signal interception problem) in precise mathematical terms. Section 3 derives the dynamic programming equation and discusses its solution for a given terminal condition. This section also introduces the idea of computing the solution of the dynamic programming equation in a "multi-parameter recursive" manner. Section 4 discusses the solution of the dynamic program for the specific case of one spatial dimension. The solution is also carried out for a specific numerical example. Finally, in section 4 we conclude and indicate how this simple example fits into the general p.o. set framework presented in chapter 3.

2.2 Mathematical Formulation of the Signal Interception Problem

In order to avoid technical complications which tend to be obscure the essence of the problem, let us assume that the time variable takes as values only multiples of a fixed positive time unit denoted δ. In chapter 5, which concludes part I, we will discuss the mathematical difficulties of using continuous time and we will suggest specific directions for further research on this problem.

Mathematically, ξ and Θ are random variables defined with respect to an underlying probability space (P,Ω,F). The random variable ξ takes its values in R^n and Θ takes its values in δZ, the set of integer multiples of the positive real number δ. The problem is to find the R^n-valued random velocity process V such that:

(1) the Euclidean norm of $V(t,\omega)$ satisfies the inequality

$$||V(t,\omega)|| \leq c \qquad \forall t, \ \forall \omega \in \Omega;$$

(2) if the receiver starts from the spatial location X_0 in R^n at time t_0 in δZ, the R^n-valued random trajectory X defined by

$$X(t,\omega) = x_0 + \sum_{s=t_0+1}^{t} V(s,\omega)\cdot\delta$$

for all t in δZ for which $t > t_0$, must intercept the signal in the least expected amount of time. In order to complete the mathematical formulation of the problem we must define explicitly what it means to intercept the signal. This is easy enough to do, and the condition that $X(t,\omega)$ has intercepted the signal by time t is that the following inequalities hold:

$$\Theta(\omega) \leq t,$$

$$||X(t,\omega)-\xi(\omega)|| \leq c|t-\Theta(\omega)|.$$

We will discuss what this means shortly, but before we do, there is another specification for this problem that we must make. Beside the constraint on its norm, the velocity process V must satisfy the following causality constraint: at each point of the space-time trajectory of the receiver, the velocity can depend only on the past information available to the receiver. Conventionally, one could specify an increasing family of σ-fields, $\{\mathcal{F}_t : t_0 \leq t\}$, such that $\mathcal{F}_t$ represents the past information up to time t. Then the causality

requirement for V would say that $V(t,\cdot)$ is $\mathcal{F}_t$-measurable for each t. However, the present problem is far from conventional in this respect. There is no past information, $\mathcal{F}_t$, depending only on the time t. The most important aspect of the signal interception problem is that the finite signal speed defines a <u>past</u> for each point in space-time. Moreover, <u>the past is different for different points in space as well as time</u>. It is this dependence of the past on both space and time together that distinguishes this problem from a conventional stochastic optimal control or stopping problem.

To make the notion of past precise, let (x,t) denote a space-time point with spatial coordinate x in R^n and temporal coordinate t in R. We say that the space-time point (y,s) is <u>before</u> (x,t) and write $(y,s) \leq (x,t)$ if the following two inequalities hold:

$$(2.1) \qquad \begin{aligned} s &\leq t, \\ ||x-y|| &\leq c|t-s|. \end{aligned}$$

In the theory of special relativity (where c is then the speed of light) this is precisely the condition that determines when a signal can travel (at a speed at

most the speed of light) from the space-time point (y,s) to the space-time point (x,t). The set $K^-(x,t)$ of all space-time points before (x,t) constitutes the <u>past</u> of (x,t). Likewise, the set $K^+(x,t)$ of all space-time points (y,s) such that (x,t) is before (y,s) constitutes the <u>future</u> of (x,t). In the theory of relativity $K^-(x,t)$ and $K^+(x,t)$ are called the <u>backward</u> and <u>forward</u> <u>light</u> <u>cones</u>, respectively. Indeed, they are circular cones in $R^n \times R$ with a common axis parallel to the time-axis. For one spatial dimension the sets $K^-(x,t)$ and $K^+(x,t)$ are pictured in figure 2.1. It is important to note that in the space-time case the past and future do not inlcude all of space-time -- there are points (y,s) which are in neither $K^-(x,t)$ nor $K^+(x,t)$. In the language of p.o. sets, this means that the space-time order relation defines a partial order but not a total order .

In terms of the space-time order $\leq$, the condition that the trajectory $X(\cdot,\omega)$ <u>intercepts</u> the signal at time t is just that

$$(2.2) \qquad (\xi(\omega),\Theta(\omega)) \leq (X(t,\omega),t).$$

Thus, what we mean by "signal interception" is that the

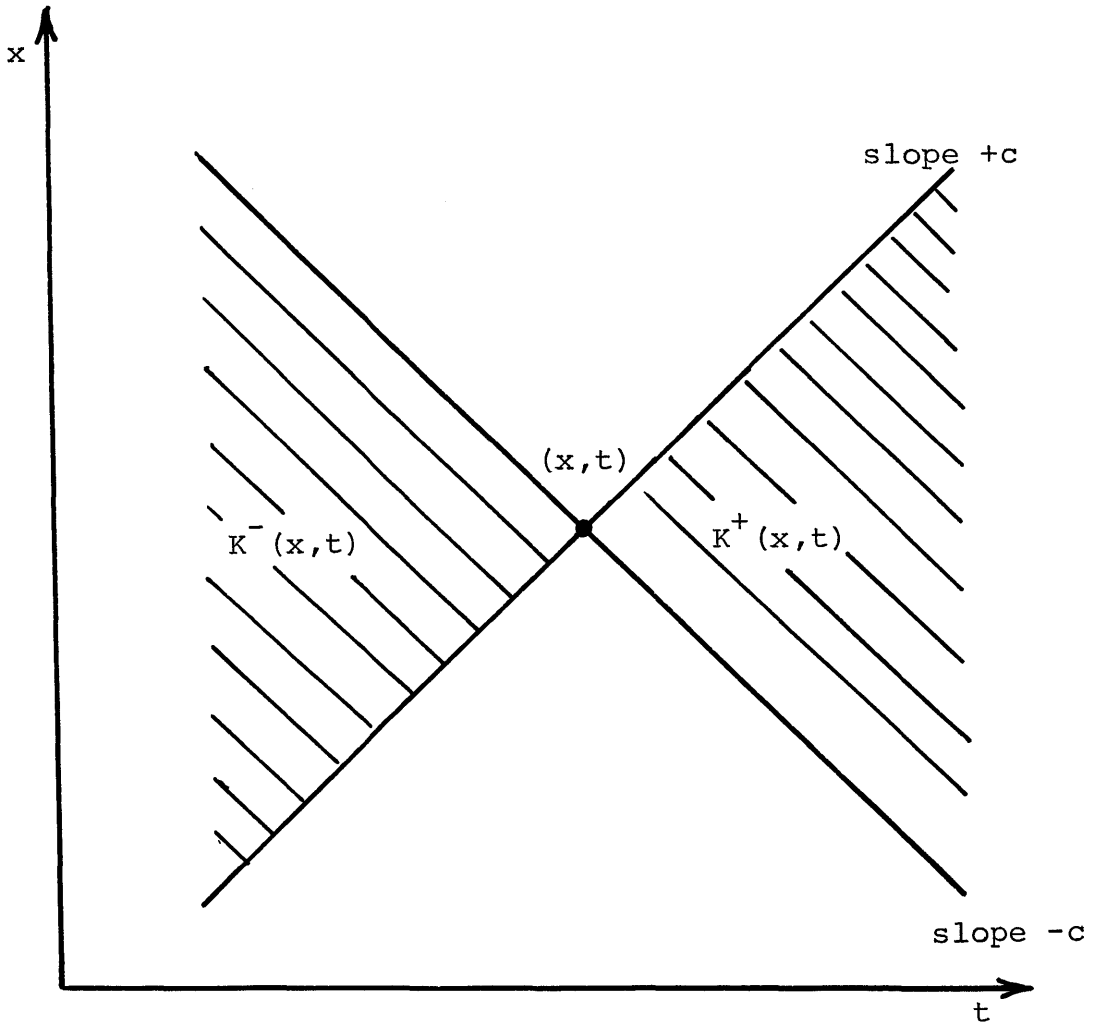


Figure 2.1

receiver at space-time point $(X(t,\omega),t)$ will receive a signal which was sent from spatial location $\xi(\omega)$ at some time after, $\Theta(\omega)$ and has reached the spatial location $X(t,\omega)$ by time t. To illustrate the situation, figure 2.2 shows an example for one spatial dimension when ξ and Θ are fixed numbers. The trajectory pictured is the minimum time trajectory from (x_0,t_0) in this case. It is clear in this trivial deterministic case that it is optimal to choose V to be the constant equal to

$$c\,\frac{\xi-x_0}{||\xi-x_0||}$$

if $x_0 \neq \xi$ and anything such that $||V|| \leq c$ if $x_0 = \xi$.

In the stochastic problem when ξ and Θ are random variables, the <u>past information</u> of the space-time point (x,t) is a σ-field $\mathcal{F}_{(x,t)}$. In this problem the only information allowed the receiver comes from knowing whether or not the receiver has intercepted the signal from (ξ,Θ). Thus, the appropriate $\mathcal{F}_{(x,t)}$ is is the smallest σ-field containing all the probabilistic events that the signal was intercepted in the space-time region $\bar{K}(x,t)$ -- that is, $\mathcal{F}_{(x,t)}$ is the smallest

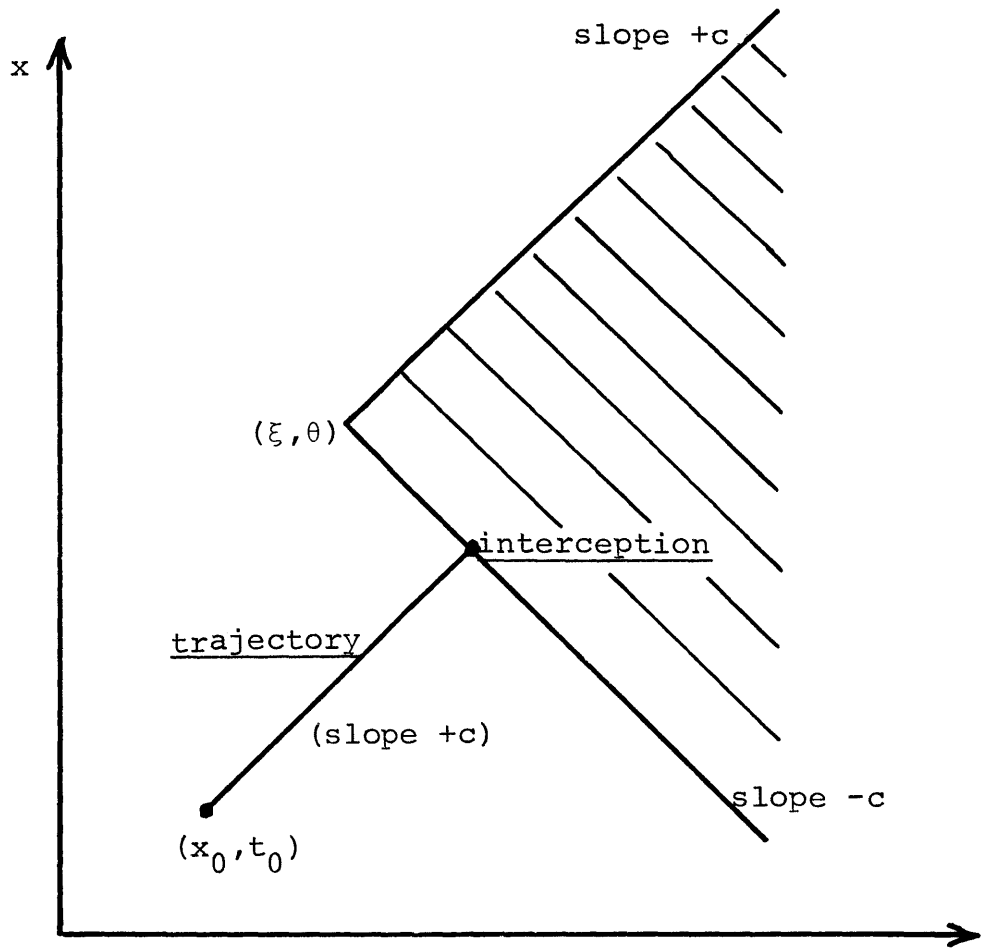

x
slope +c
(ξ,θ)
interception
trajectory
(slope +c)
slope -c
(x_0,t_0)

Figure 2.2

σ-field containing the sets $\{\omega : (\xi(\omega),\Theta(\omega)) \leq (y,s)\}$ for all space-time points (y,s) which are before (x,t). Note that the family $\{\mathcal{F}_{(x,t)}\}$ is <u>increasing</u> in the sense that $\mathcal{F}_{(y,s)} \subseteq \mathcal{F}_{(x,t)}$ if $(y,s) \leq (x,t)$. Chapter 3 will treat such properties in more generality and detail; for the present problem we do not need to say more about the σ-fields $\mathcal{F}_{(x,t)}$.

Now we can formulate the <u>causality constraint</u> on V precisely. To do this it is best to consider the velocity as a random vector field $(x,t,\omega) \to V(x,t,\omega)$ over space-time. Thus, the random trajectory corresponding to V will be the solution X of the random difference equation

$$X(t+\delta,\omega) = X(t,\omega) + V(X(t,\omega),t,\omega)\cdot\delta \tag{2.3}$$

with the initial condition $X(t_0,\omega) = x_0$. The norm constraint on V is the same as before, namely,

$$||V(x,t,\omega)|| \leq c \tag{2.4}$$

for all x, t and ω. The <u>causality constraint</u> requires that $V(x,t,\cdot)$ is $\mathcal{F}_{(x,t)}$-measurable for each (x,t) in space-time. This means that the receiver, having arrived at the space-time point (x,t), must calculate its new

velocity $V(x,t,\omega)$ only on the basis of information available at the point (x,t) -- namely, the past information $\mathcal{F}_{(x,t)}$ of the space-time point (x,t).

Let $C(x_0,t_0)$ denote the set of all receiver trajectories which are unique solutions of (2.3) for some causally constrained velocity field satisfying (2.4). That is, $C(x_0,t_0)$ is the class of admissable trajectories for the problem. For such a trajectory X define the interception time as

$$(2.5) \quad \tau(X,\omega) = \inf \{t : (\xi(\omega),\Theta(\omega)) \leq (X(t,\omega),t)\}$$

or $+\infty$ if the infimum is taken over an empty set. If $E(\cdot)$ is the expectation operation associated with the probability space $(P,\Omega,\mathcal{F})$ then our task is to find an optimal trajectory X^* in $C(x_0,t_0)$, or better, a corresponding velocity field V^* that gives X^* via (2.3), such that

$$(2.6) \qquad E(\tau(X^*)) \leq E(\tau(X))$$

for all trajectories X in $C(x_0,t_0)$. In (2.6) we have have suppressed the ω dependence of $\tau(X^*)$ and $\tau(X)$ as usual in denoting random variables.

This completes the mathematical formulation of the signal interception problem. The precise formulation presented in this section will help clarify the generalization to p.o. sets presented in chapter 3.

1.3 Dynamic Programming Equation

Despite the severely mathematical formulation of the problem in the previous section, the actual solution presented in this section will follow easily from some simple heuristic arguments. Moreover, these arguments will help clarify the interception problem. In the more general case presented in chapter 3 a fully rigorous mathematical derivation is necessary, but in the present relatively simple problem the rigorous derivation obscures the simplicity of the solution. Therefore, we leave a detailed mathematical development until chapter 3, and in that chapter we will indicate how the results of this section fit into the general mathematical framework.

This section determines the solution of the interception problem by means of conventional dynamic programming. However, this dynamic program is notable in that it is better interpreted as an unconventional multi-parameter recursive equation than as the usual one-parameter recursive dynamic programming equation. We

briefly discuss the multiparameter recursion at the end of this chapter but leave the detailed discussion to the next chapter. In this section we also establish the dynamic programming equation and discuss the existence and uniqueness of solutions to that equation. Note that in this section all time variables, usually denoted by t, will take values in δZ unless otherwise stated.

Before deriving the dynamic programming equation we must define some new notation in addition to that of section 2.2. Let $p(x,t)$ denote the probability distribution

$$p(x,t) = P(\{\omega : (\xi(\omega),\theta(\omega)) \leq (x,t)\}).$$

That is, p is the prior distribution of (ξ,θ) which completely describes the statistics of the problem. For future application let us define $q(x,t)$ as

$$q(x,t) = 1 - p(x,t).$$

Our equation will have a simpler appearance in terms of q. Next define $\tau(x,t)$ as the least expected amount of time until interception, given that the receiver starts at spatial location x at time t and has not intercepted

the signal there. Now let us derive the dynamic programming equation for τ.

Let $p_1(x,t,v)$ denote the probability that the signal will be intercepted in the next unit δ of time if the receiver uses velocity v and the signal is not yet intercepted at (x,t). Let $q_1(x,t,v) = 1-p_1(x,t,v)$, the probability of interception not occurring in the next unit of time when the receiver uses velocity v and has not yet intercepted the signal. Then if the receiver uses velocity v at space-time point (x,t), the least expected amount of time until interception is

$$(3.1) \quad \delta \cdot p_1(x,t,v) + [\tau(x+v\cdot\delta, t+\delta)+\delta] \cdot q_1(x,t,v).$$

Note that $(x+v\delta, t+\delta)$ will be the next space-time location of the receiver if it uses velocity v at (x,t). Thus, $\tau(x+v\cdot\delta, t+\delta)$ is the least expected time until interception from $(x+v\delta, t+\delta)$ if no interception has occurred before $(x+v\delta, t+\delta)$. Since $\tau(x,t)$ is the least expected time until interception starting from (x,t), it must be the infimum of the expression (3.1) taken over permissable values of v. That is, $\tau(x,t)$ is given as

(3.2) $\tau(x,t) =$

$$\inf \{\delta \cdot p_1(x,t,v) + [\tau(x+v\delta, t+\delta)+\delta] \cdot q_1(x,t,v) : ||v|| \leq c\}.$$

Finally, let us write p_1 and q_1 in terms of p. By definition, $p_1(x,t,v)$ is given by

$$p_1(x,t,v) = P(\{(\xi,\Theta) \leq (x+\delta \cdot v, t+\delta)\} | \{(\xi,\Theta) \not\leq (x_0 t)\})$$

or in other words,

$$(3.3) \quad p_1(x,t,v) = \frac{P(\{(\xi,\Theta) \leq (x+\delta v, t+\delta), (\xi,\Theta) \not\leq (x,t)\})}{P(\{(\xi,\Theta) \not\leq (x,t)\})}$$

The probability $P(\{(\xi,\Theta) \not\leq (x,t)\})$ is simply $1 - p_1(x,t,v)$ or $q_1(x,t,v)$. The numerator is the is the probability that the random space-time variable (ξ,Θ) is contained in the backward cone $K^-(x+\delta v, t+\delta)$ but not contained in the backward cone $K^-(x,t)$. Figure 3.1 illustrates the situation. Since $||v|| \leq c$, the cone $K^-(x,t)$ is contained in $K^-(x+\delta v, t+\delta)$ and the numerator of (3.3) is the probability that (ξ,Θ) lies between the two cones, that is, in the set

$$K^-(x+\delta v, t+\delta) - K^-(x,t).$$

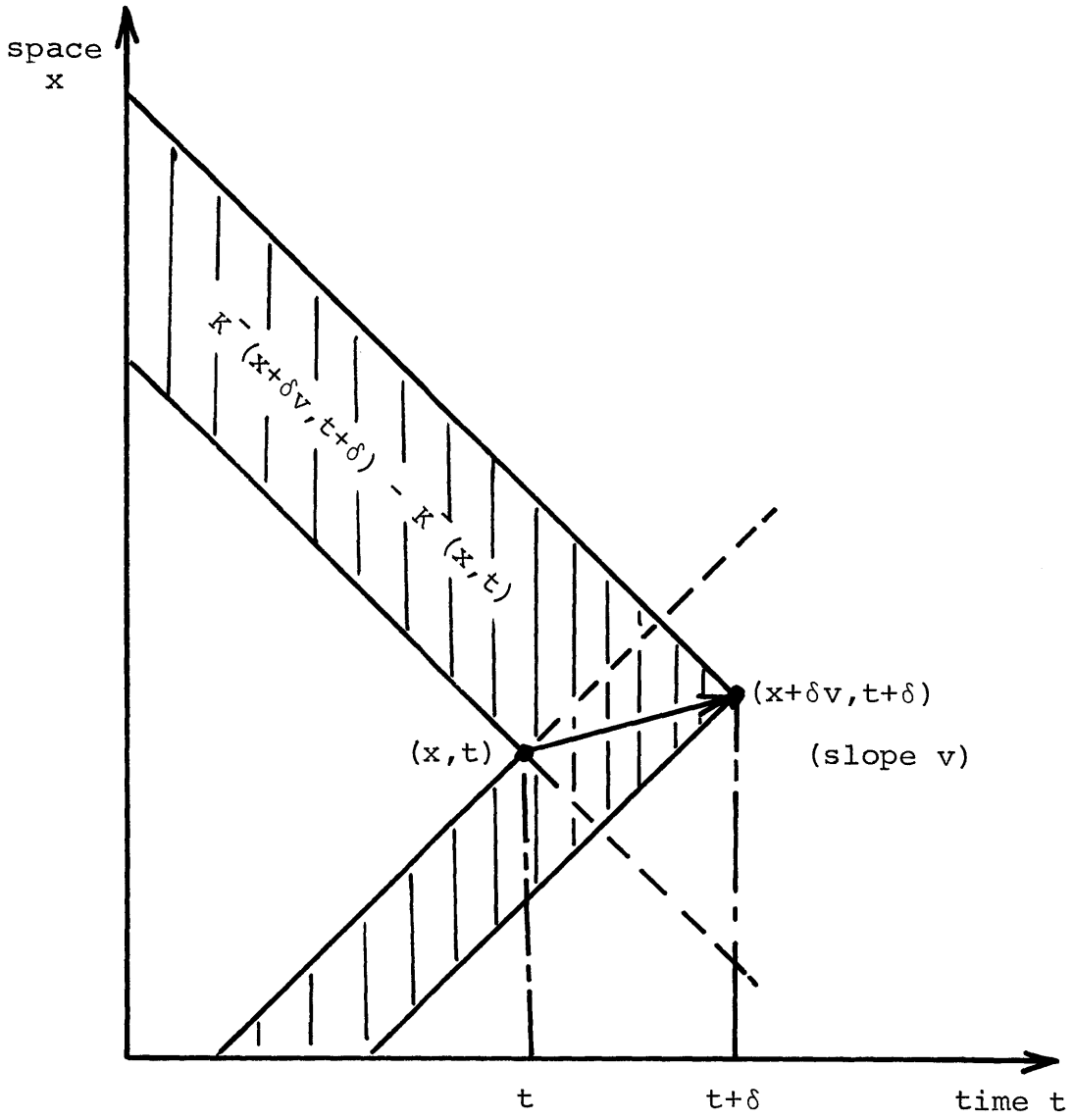


Figure 3.1

$P(\{(\xi,\theta) \leq (x+\delta v, t+\delta), (\xi,\theta) \not\leq (x,t)\})$ is probability (ξ,θ) lies in the shaded region.

Noting that $p(x,t)$ is the probability that (ξ,Θ) lies in $K^-(x,t)$ and $p(x+\delta v,t+\delta)$ is the probability that (ξ,Θ) lies in $K^-(x+\delta v,t+\delta)$, one sees that the numerator is $p(x+\delta v,t+\delta) - p(x,t)$ and thus, $p_1(x,t,v)$ is given by

$$p_1(x,t,v) = \frac{p(x+\delta v,t+\delta) - p(x,t)}{1 - p(x,t)}$$

or, in terms of q,

$$p_1(x,t,v) = \frac{q(x,t) - q(x+\delta v,t+\delta)}{q(x,t)} .$$

Likewise, $q_1(x,t,v)$ is given by

$$q_1(x,t,v) = \frac{q(x+\delta v,t+\delta)}{q(x,t)} .$$

Substituting these expressions in equation (3.2) yields the dynamic programming equation for τ, namely

$$(3.4) \quad \tau(x,t) = \inf \{\tau(x+\delta v,t+\delta)\frac{q(x+\delta v,t+\delta)}{q(x,t)} : ||v|| \leq c\} + \delta.$$

Note that if $q(x,t) = 0$, then $(\xi,\Theta) \leq (x,t)$ almost surely and one may set $\tau(x,t) = 0$, since the interception has already occurred.

Before studying equation (3.4) further, let us

transform it to a more convenient form. Define now the new function ψ for each (x,t) as

$$(3.5) \qquad \psi(x,t) = q(x,t)\tau(x,t).$$

Since q is assumed given, we can obtain ψ from τ if τ is known or τ from ψ if ψ is known (using the fact that $\tau(x,t) = 0$ when $q(x,t) = 0$ when necessary). For the function ψ equation (3.4) becomes

$$(3.6) \qquad \psi(x,t) = \inf \{\psi(x+\delta v, t+\delta) : ||v|| \leq c\} + \delta q(x,t).$$

We note in passing that a slight alteration of (3.6) will account for additional velocity constraints. For example, suppose $S(x,t)$ is a subset of R^n for each (x,t), which represents the possible choices of velocity at (x,t). As long as $||v|| \leq c$ for all v in $S(x,t)$, the derivation of (3.6) remains the same and the only difference is that (3.6) becomes

$$(3.7) \qquad \psi(x,t) = \inf \{\psi(x+\delta v, t+\delta) : v \in S(x,t)\} + \delta \cdot q(x,t).$$

Suppose that τ and hence, ψ exists for the interception problem. That is, suppose that the least

expected time until interception from (x,t) exists for each (x,t). For example, a sufficient condition for τ to exist is the following: for each (x,t) in space-time assume that there exists an admissable trajectory X in $C(x,t)$, which was defined in section 2.2, and a real number $r \geq t$ such that

$$(\xi(\omega),\theta(\omega)) \leq (X(r,\omega),r)$$

for almost all ω such that $(\xi(\omega),\theta(\omega)) \not\leq (x,t)$. In this case $\tau(x,t)$ will exist and $0 \leq \tau(x,t) \leq r-t$. Thus, ψ exists and it must necessarily satisfy equation (3.6). Moreover, if we know ψ, then we can determine an optimal velocity control from (3.6). To do this, assume that for each (x,t) there is a v_0 with $||v_0|| \leq c$ and such that

$$(3.8)\quad \psi(x+\delta v_0,t+\delta) = \inf\{\psi(x+\delta v,t+\delta) : ||v|| \leq c\}.$$

For each (x,t) let $V_0(x,t)$ be one choice of v_0 that satisfies equation (3.8). Then $V_0(x,t)$ is the velocity to use to minimize the expected time to interception from (x,t) if no interception has occurred before (x,t). Thus, an optimal velocity control for the problem is then

V^*, defined as

$$(3.9) \qquad V^*(x,t,\omega) = V_0(x,t)$$

if $(\xi(\omega),\theta(\omega)) \not\leq (x,t)$, and

$$V^*(x,t,\omega) = 0$$

if $(\xi(\omega),\theta(\omega)) \leq (x,t)$. The optimal trajectory X^* in $C(x_0,t_0)$ corresponding to V^* is the solution of the finite difference equation,

$$(3.10) \quad X^*(t+\delta,\omega) = \delta V^*(X(t,\omega),t,\omega) + X^*(t,\omega)$$

for $t \geq t_0$ with the initial condition $X^*(t_0,\omega) = x_0$. Note that the optimal velocity control V^* is not necessarily unique. Indeed, several different controls may each yield the minimum expected time.

Having derived an optimal velocity V^* from ψ, we now show how to obtain ψ from the dynamic programming equation. If ψ exists for the problem and if the dynamic programming equation (3.6) has a unique solution, then ψ must be that unique solution. Although the problem may

have several optimal velocity controls, we now show that the equation (3.6) has one and only one solution ψ if certain terminal conditions are specified, and we show that the solution can be found by computing backwards from the terminal conditions. First we do this in terms of conventional one-parameter dynamic programming, and then we show how the dynamic program can be interpreted as a multi-parameter dynamic program. In the multi-parameter interpretation one must understand the expressions "backwards" and "terminal" in terms of the space-time order $\leq$ as we will explain.

Suppose that the function $x \to \psi(x,t)$ (denoted $\psi(\cdot,t)$) was given for the time $t = t_1$. Given $\psi(\cdot,t)$, equation (3.6) determines $\psi(\cdot,t-s)$. Thus, starting from $\psi(\cdot,t_1)$, one can recursively compute the functions $\psi(\cdot,t)$ for all t such that $t < t_1$. Remember that the unit of time is δ so that both t_1 and t are integer multiples of δ. Thus, it is clear that the solution $\psi(x,t)$ is determined uniquely for all x in R^n and all times t such as that $t \leq t_1$, if $\psi(\cdot,t_1)$ is given as a terminal condition. So far we are viewing (3.6) as an ordinary recursive equation in the one parameter t with the one-parameter terminal condition $\psi(\cdot,t_1)$. However, it

is useful to reconsider (3.6) as a multiparameter recursive equation with the space-time parameter (x,t). The interpretation of (3.6) as a multiparameter recursion (or multiparameter dynamic program) stems from the fact that for a given (y,s) in space-time, $x(y,s)$ does not depend on $\psi(x,s+\delta)$ for all x but just those x such that $x = y + \delta \cdot v$ for some v with $||v|| \leq c$. In particular, the computation of ψ at (y,s) depends only on the value of ψ at (x,t) such that $(y,s) \lneq (x,t)$. If we only require to find $\psi(y,s)$ for (y,s) in some bounded region A, we only need to know $\psi(x,t)$ for (x,t) in some other bounded region B, and not for infinitely many spatial coordinates x. Conversely, if we are given the value of ψ at each (x,t) in a bounded region B of space-time, then the dynamic programming equation (3.6) determines the value of ψ at all (y,s) in a largest region A and at no other points. Note that A is also bounded. At this point it may be helpful to refer to the figures 3.2 and 3.3 which illustrate the sets B and the corresponding sets A in two different cases of B for one spatial dimension. Let us make this relationship between B and A more explicit.

Suppose that the restriction of the function ψ to B, denoted $\psi|B$, is given. Let $B_0 = B$ and for each integer

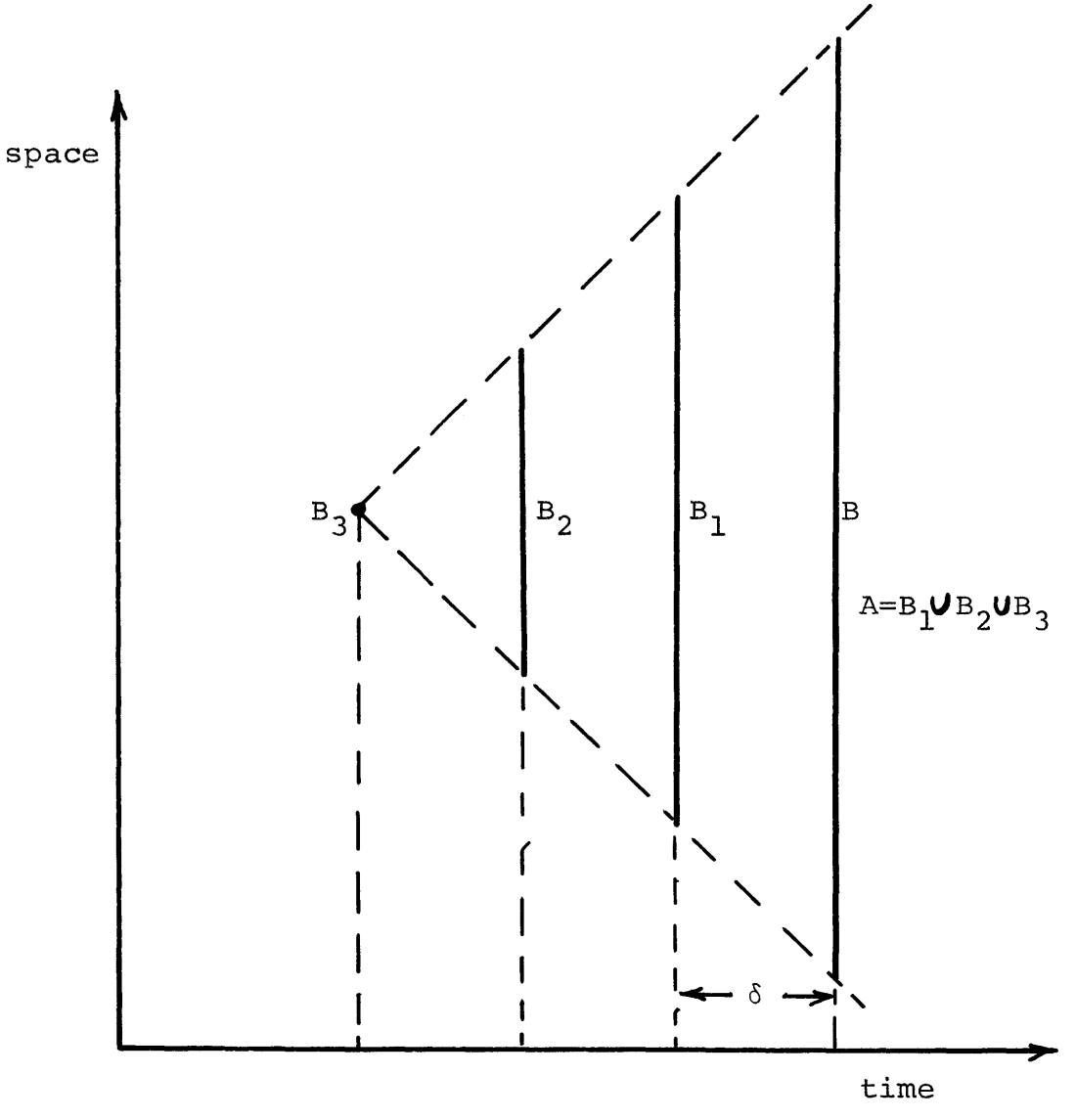


Figure 3.2

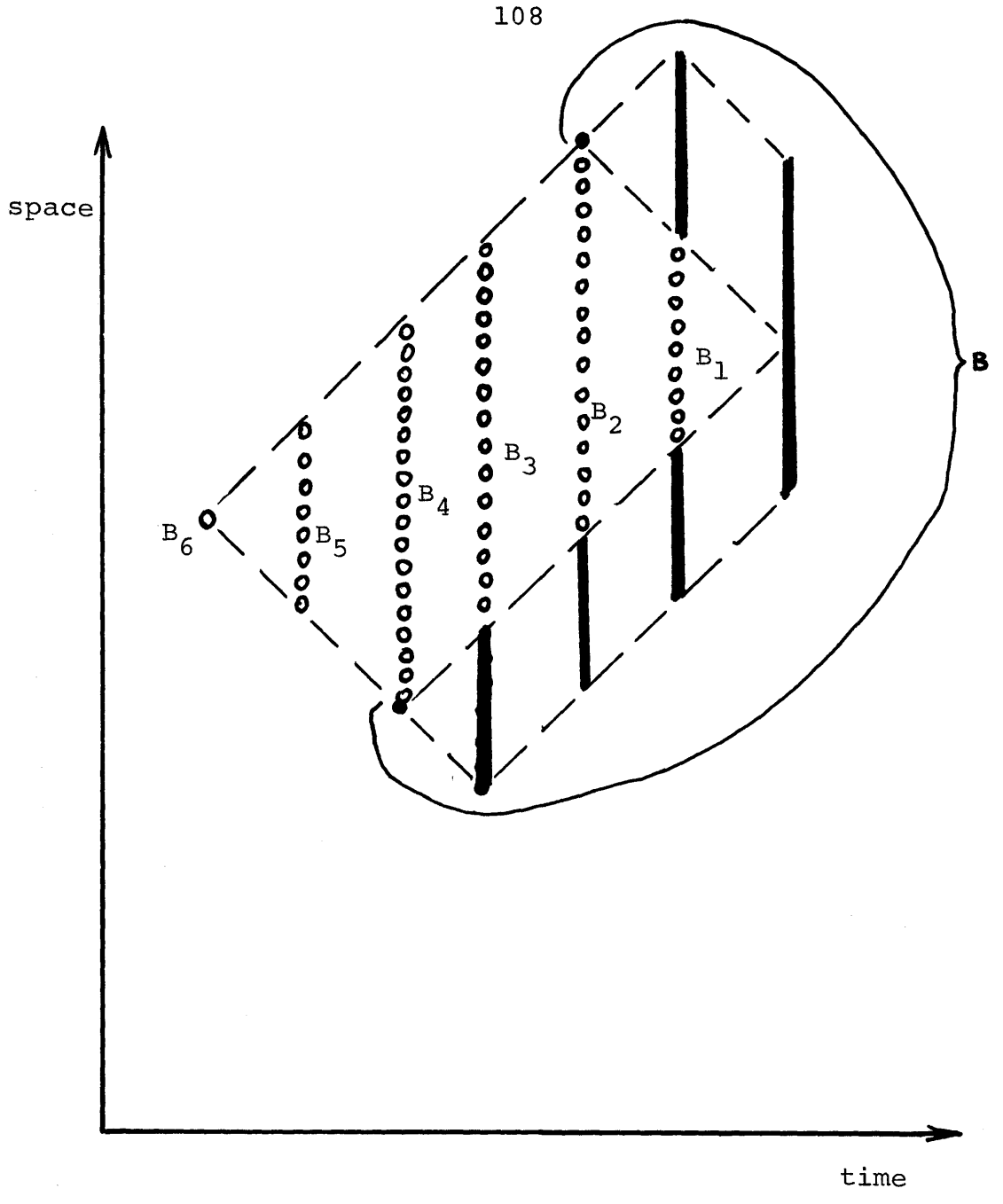


Figure 3.3

$n \geq 1$ let B_n be the set of (y,s) such that $(y+\delta v, s+\delta)$ lies in B_{n-1} for all v such that $||v|| \leq c$. Then from (3.6), as remarked above, $\psi|B_n$ depends just on $\psi|B_{n-1}$. Thus, given $\psi|B$ to start with, one can determine uniquely $\psi|A$, where $A = \bigcup_{n\geq 1} B_n$. Note that A is the largest region in space-time for which one can determine the value of Ψ starting from just $\psi|B$. Note also that if B is bounded, then so is A and also, $B_n = \emptyset$ for all n larger than some finite n_0.

The values of ψ given on a region B, namely $\psi|B$, is called a terminal condition for the dynamic program -- it is the multiparameter generalization of the terminal condition $\psi(\cdot,t_1)$, but B can contain space-time points at different times. Physically, B is a region in space-time at which we know the minimum expected time to go to interception, and hence, at which we know $\psi|B$. In the next section we demonstrate how to determine the terminal condition in the case of one spatial dimension.

Starting from a given terminal condition, one calculates backward with respect to the space-time order relation using the dynamic programming equation (3.6). Hence, we call this method of computation a multiparameter backward recursion. The following section investigates the multiparameter

recursion in more detail for the case of one spatial dimension, and by means of a particular example it illustrates more clearly the idea of a terminal set and a multiparameter backward recursion with respect to the space-time order relation. Chapter 3 generalizes these ideas to p.o. sets and shows that the dynamic program has a unique solution given a terminal condition.

2.4 Solution of Dynamic Programming Equation for One Spatial Dimension

In order to clarify further the nature of the multiparameter recursive equation (3.6) and to indicate the effectiveness of such a multiparameter recursion, this section investigates equation (3.6) for the case of one spatial dimension and solves a specific example in this case. For simplicity assume that $c = 1$ and $\delta = 1$, and allow the velocity to take only the values, +1, 0, -1. This last restriction makes the spatial variable discrete as well as the time variable. Thus, the parameter space is $Z^2 = Z \times Z$ instead of $R \times Z$, and it is natural to assume that (x_0,t_0), the initial location of the receiver, lies in Z^2 and that the random variable (ξ,θ) takes its values in Z^2.

To guarantee the existence of solutions of (3.6)

let us assume that $q(x_1,t_1) = 0$ for some space-time point (x_1,t_1) in Z^2. By definition of q this means that

$$P(\{(\xi,\Theta) \leq (x_1,t_1)\}) = 1$$

or in other words, $(\xi,\Theta) \leq (x_1,t_1)$ almost surely. Physically, this means that the signal will be sure to reach the spatial position x_1 by time t_1. Thus, if a receiver starts at the space-time point (x_0,t_0) it can be sure to intercept the signal by moving at maximum speed to the spatial position x_1 and waiting until time t_1. Using this policy, the receiver is guaranteed that its time of interception is no greater than $t_0 + |t_1-t_0| + |x_1-x_0|$. Consequently, the least expected time $\tau(x_0,t_0)$ to interception from (x_0,t_0) must exist and

$$\tau(x_0,t_0) \leq |t_1-t_0| + |x_1-x_0|.$$

It follows that $\psi(x_0,t_0) = q(x_0,t_0)\tau(x_0,t_0)$ exists, and the function ψ must satisfy (3.6) as we showed in the previous section.

Having shown the existence of ψ, we now show how to calculate a terminal condition for this ψ, which one can then use with equation (3.6) to calculate the function ψ everywhere. The terminal set B we shall use is the set defined by

$$(4.1) \qquad B = \{(y,s) : |y-x_1| = |s-t_1-u|,\ s \leq t_1,\ u = 0,1\}.$$

Actually, we find the terminal condition first for the larger set $B' = \{(y,s) : (y,s) \not\leq (x_1,t_1)\}$. The sets B and B' are illustrated by figure 4.1 and 4.2 respectively. Note that $B \subset B'$.

To calculate the correct values of ψ in the region B' we show that $q(x,t)$ and $\psi(x,t)$ only depend on one parameter and thus, we are able to calculate them with a simple one-parameter recursion. In the region B_I, defined as the set of points (x,t) such that $t+x \geq t_1+x_1$, the functions q and ψ depend only on $t-x$. Likewise, in the region B_{II}, defined as the set of (x,t) such that $t-x \geq t_1-x_1$, these functions depend only on $t+x$. Figure 4.3 illustrates the sets B_I and B_{II}; the set B' is the union $B_I \cup B_{II}$.

Intuitively, B_I and B_{II} are space-time regions

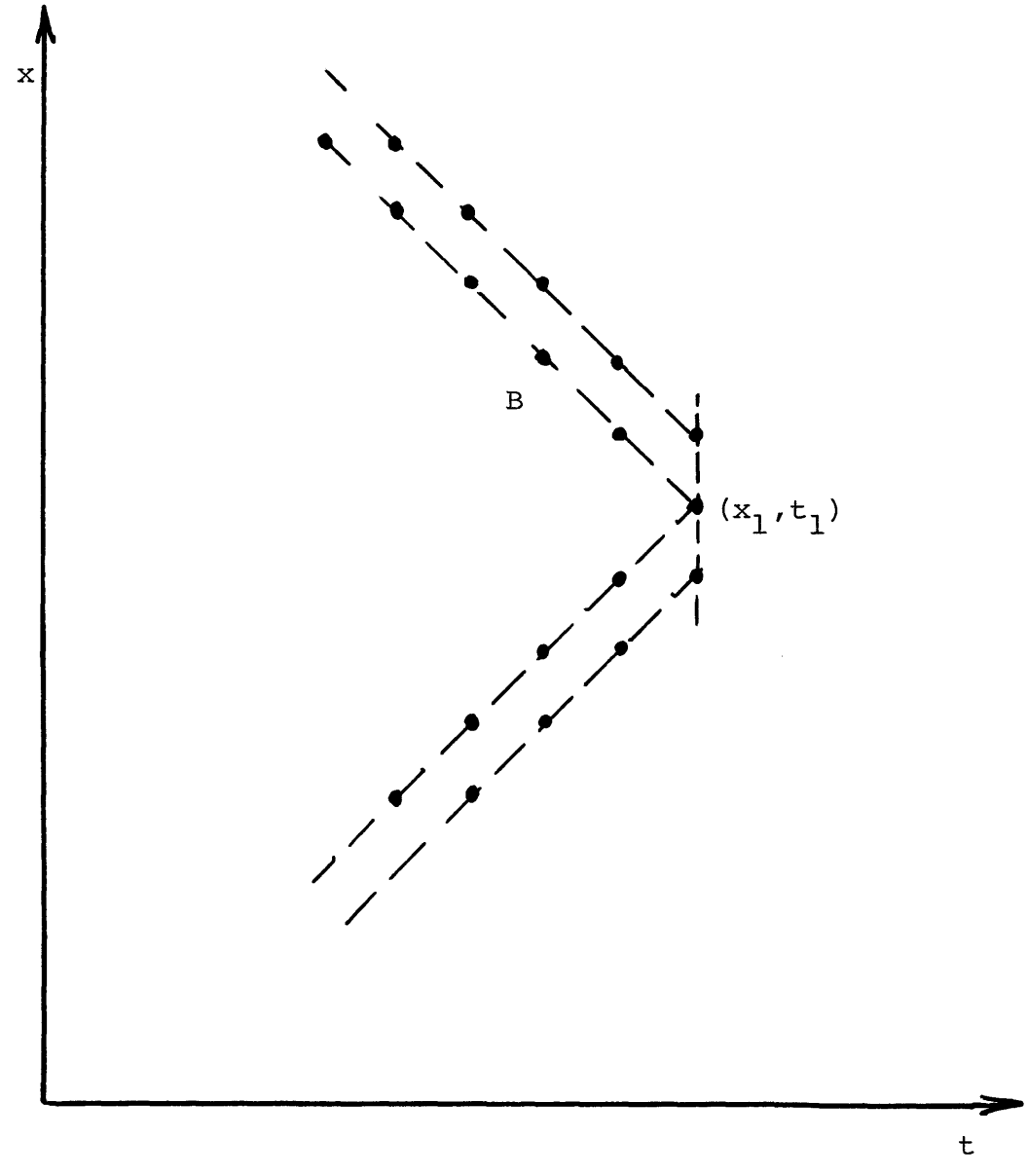


Figure 4.1

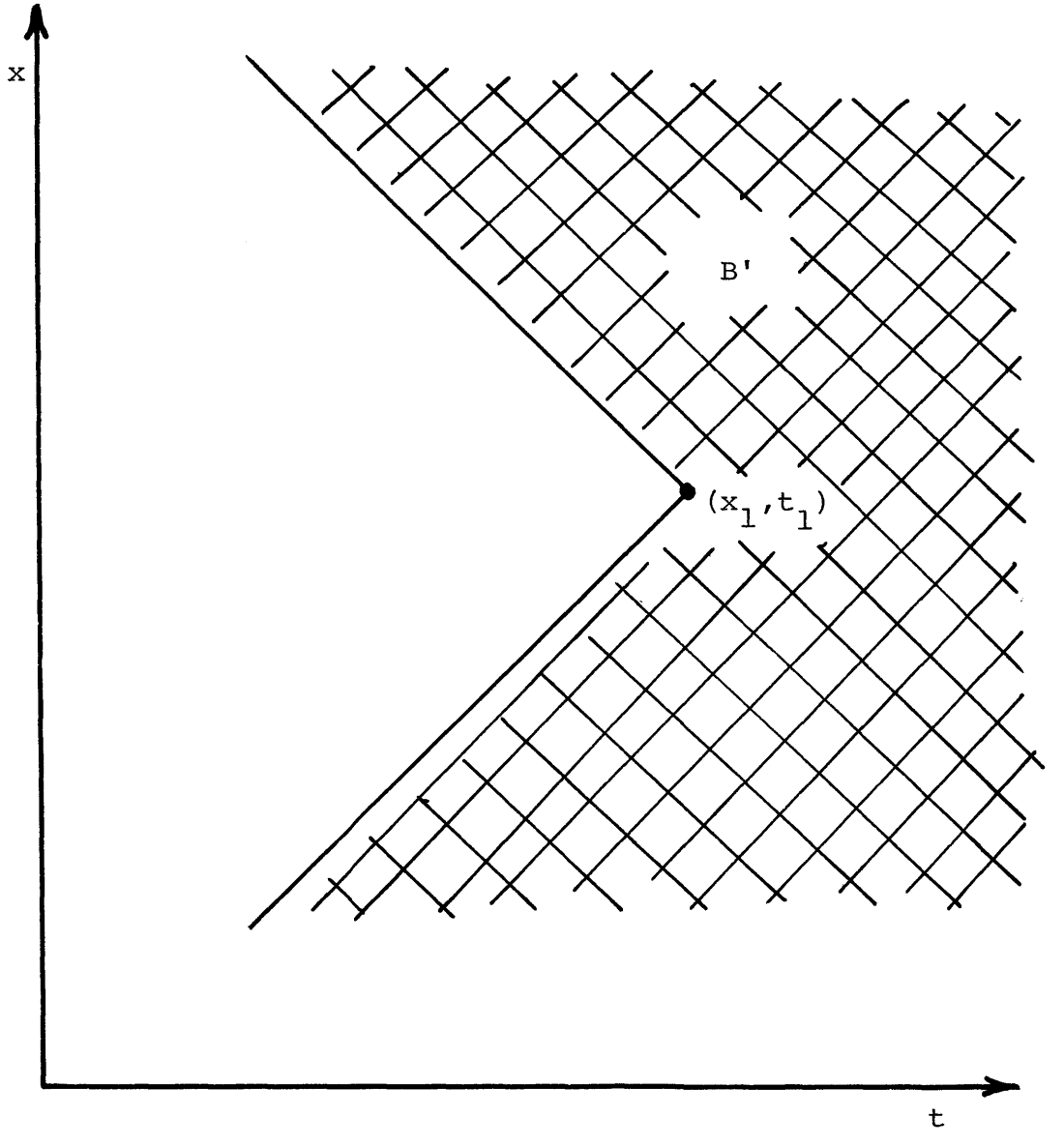


Figure 4.2

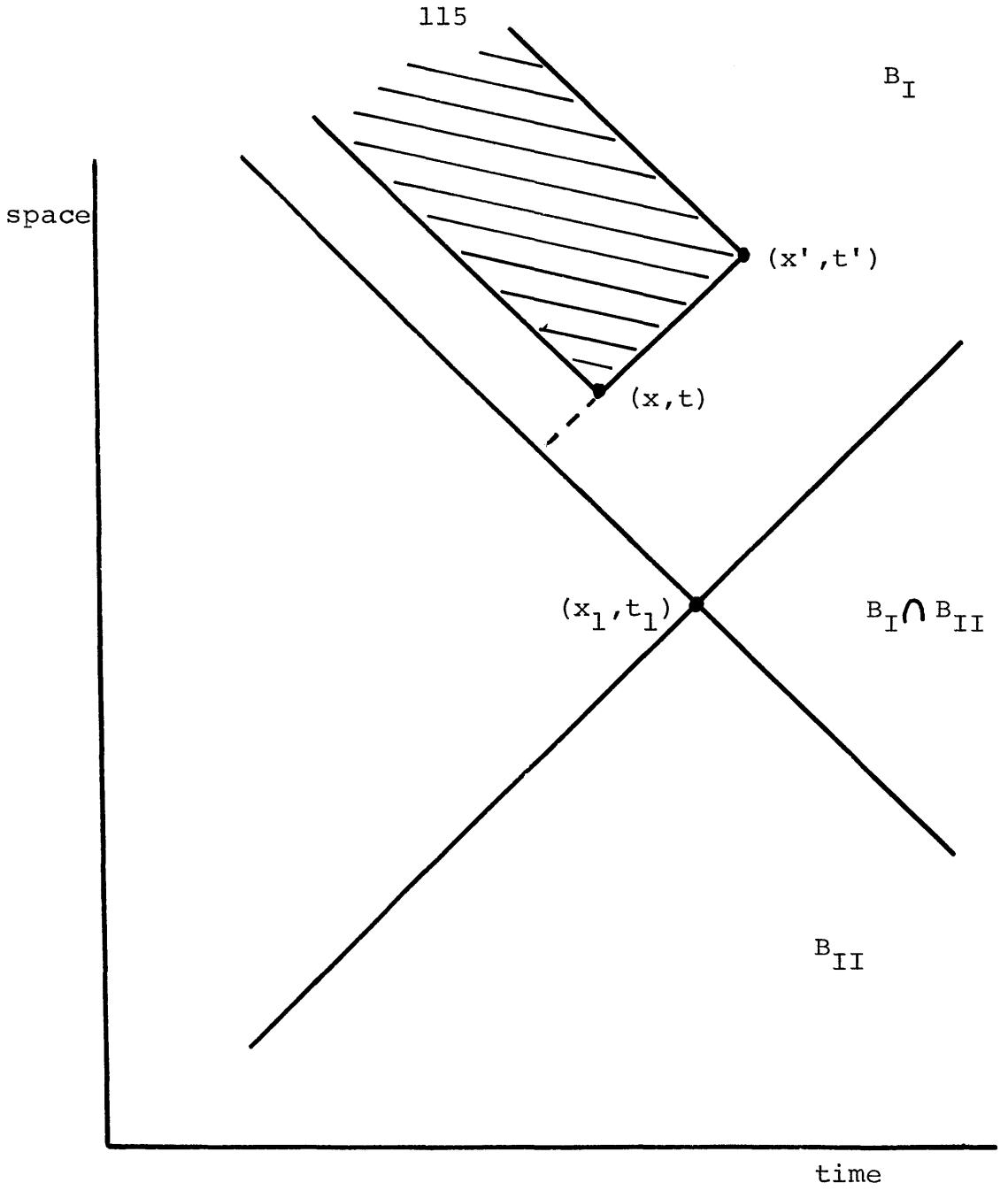


Figure 4.3

in which it is always an optimal policy for the receiver to head toward x_1 at maximum speed. In B_I one uses the velocity $v = -1$, in B_{II} one uses the velocity $v = +1$, and in the intersection $B_I \cap B_{II}$ one can use any velocity since the signal has already been intercepted in that region. Using these optimal control velocities in the dynamic programming equation (3.6), we obtain the recursive relation

$$\psi(x,t) = \psi(x-1,t+1) + q(x,t) \tag{4.2}$$

for (x,t) in region B_I, and the relation

$$\psi(x,t) = \psi(x+1,t+1) + q(x,t) \tag{4.3}$$

for (x,t) in B_{II}. In the intersection $B_I \cap B_{II}$ we must have $\psi = 0$. To turn these equations into one-parameter recursive equations, we must show that $q(x,t) = q(x',t')$ when $t-x = t'-x'$ for (x,t) and (x',t') in B_I or when $t+x = t'+x'$ for (x,t) and (x',t') in B_{II}. Suppose that (x,t) and (x',t') are in B_I and $t-x = t'-x'$. Without loss of generality, suppose that $x'+t' \geq x+t$, and hence, $(x,t) \leq (x',t')$ as illustrated in figure 4.3. Then $q(x,t)-q(x',t')$ is the probability

that $(\xi,\Theta) \leq (x',t')$ and $(\xi,\Theta) \not\leq (x,t)$. The region $\{(y,s) : (y,s) \leq (x',t'), (y,s) \not\leq (x,t)\}$ is illustrated in figure 4.3; note that it is contained in B_I and hence in B'. We have assumed $q(x_1,t_1) = 0$ so that $P(\{(\xi,\Theta) \in B'\}) = 0$. It follows that $P(\{(\xi,\Theta) \leq (x',t'), (\xi,\Theta) \not\leq (x,t)\})$ and hence that $q(x,t)-q(x',t') = 0$. Thus, $q(x,t)$ depends only on $t-x$ when (x,t) lies in B_I. Similarly, $q(x,t)$ depends only on $t+x$ when (x,t) lies in B_{II}. Note that for (x,t) in $B_I \cap B_{II}$ we have $q(x,t) = 0$.

Let $q_I(t-x) = q(x,t)$ for (x,t) in B_I and let $q_{II}(t+x) = q(x,t)$ in B_{II}. Then one can rewrite the equations (4.2) and (4.3) as

$$\psi(x,t) = \psi(x-1,t+1) + q_I(t-x) \tag{4.4}$$

for (x,t) in B_I, and

$$\psi(x,t) = \psi(x+1,t+1) + q_{II}(t+x) \tag{4.5}$$

for (x,t) in B_{II}. Equations (4.4) and (4.5) can be used to compute ψ recursively in regions B_I and B_{II} given that $\psi = 0$ in the intersection $B_I \cap B_{II}$. If ψ_I and ψ_{II} are given by the recursive equations

(4.6) $\psi_I(r-2) + \psi_I(r) + q_I(r)$

for $r \leq t_1 - x_1 + 1$, and

(4.7) $\psi_{II}(r-2) = \psi_{II}(r) + q_I(r)$

for $r \leq t_1 + x_1 + 1$, together with the terminal conditions $\psi_I(t_1-x_1) = \psi_I(t_1-x_1+1) = 0$ and $\psi_{II}(t_1+x_1) = \psi_{II}(t_1+x_1+1) = 0$, then it is not hard to see that $\psi(x,t) = \psi_I(t-x)$ and $\psi(x,t) = \psi_{II}(t+x)$ satisfy the equations (4.4) and (4.5). Clearly, (4.4) and (4.5) determine ψ uniquely in regions B_I and B_{II} given that $\psi = 0$ in the intersection. Thus, we can determine ψ uniquely in the region B' via the one-parameter equations (4.6) and (4.7). In this way we have determined the terminal condition for the dynamic program.

Instead of considering the terminal condition on the infinite set B' or even the infinite set B defined by (4.1), let us consider the terminal condition on the following finite subsets of B. Define the sets $B_{m,n}$ as

(4.8) $B_{m,n} = \{(y,s): t_1-x_1-m \leq s-y, t_1+x_1-n \leq s+y, (y,s) \in B\}$

for all nonnegative integers m and n. The set $B_{7,10}$ is illustrated in figure 4.4. To calculate the values $\psi(y,s)$ for (y,s) in $B_{m,n}$ one must use (4.6) to calculate m steps backward from $r = t_1 - x_1$ and one must use (4.7) to calculate n steps backward from $r = t_1 + x_1$. From the terminal condition $\psi|B_{m,n}$ one can use (3.6) and calculate $\psi(x,t)$ for all (x,t) such that

$$(4.9) \qquad t_1 - x_1 - m \leq t - x < t_1 - x_1,$$

$$t_1 + x_1 + n \leq t + x < t_1 + x_1.$$

Let $A_{m,n}$ denote the set of (x,t) described by (4.9) (figure 4.4 shows the set $A_{m,n}$ corresponding to $B_{m,n}$). With the aid of the sets $A_{m,n}$ and $B_{m,n}$ it is possible now to describe the two-parameter recursive calculation of $\psi(x,t)$ for (x,t) such that $(x,t) < (x_1,t_1)$.

Having already calculated $\psi|B_{m,n}$ above from (4.6) and (4.7), one may calculate $\psi|A_{m,n}$ by computing $\psi|B^1_{m,n}$, $\psi|B^2_{m,n}$, ..., $\psi|B^N_{m,n}$ sequentially from equation (3.6) such that $A_{m,n} = B^1_{m,n} \cup B^2_{m,n} \ldots \cup B^N_{m,n}$, as discussed in section 2.3 and illustrated in figure 3.3. However, another option is available once $\psi|A_{m,n}$ is

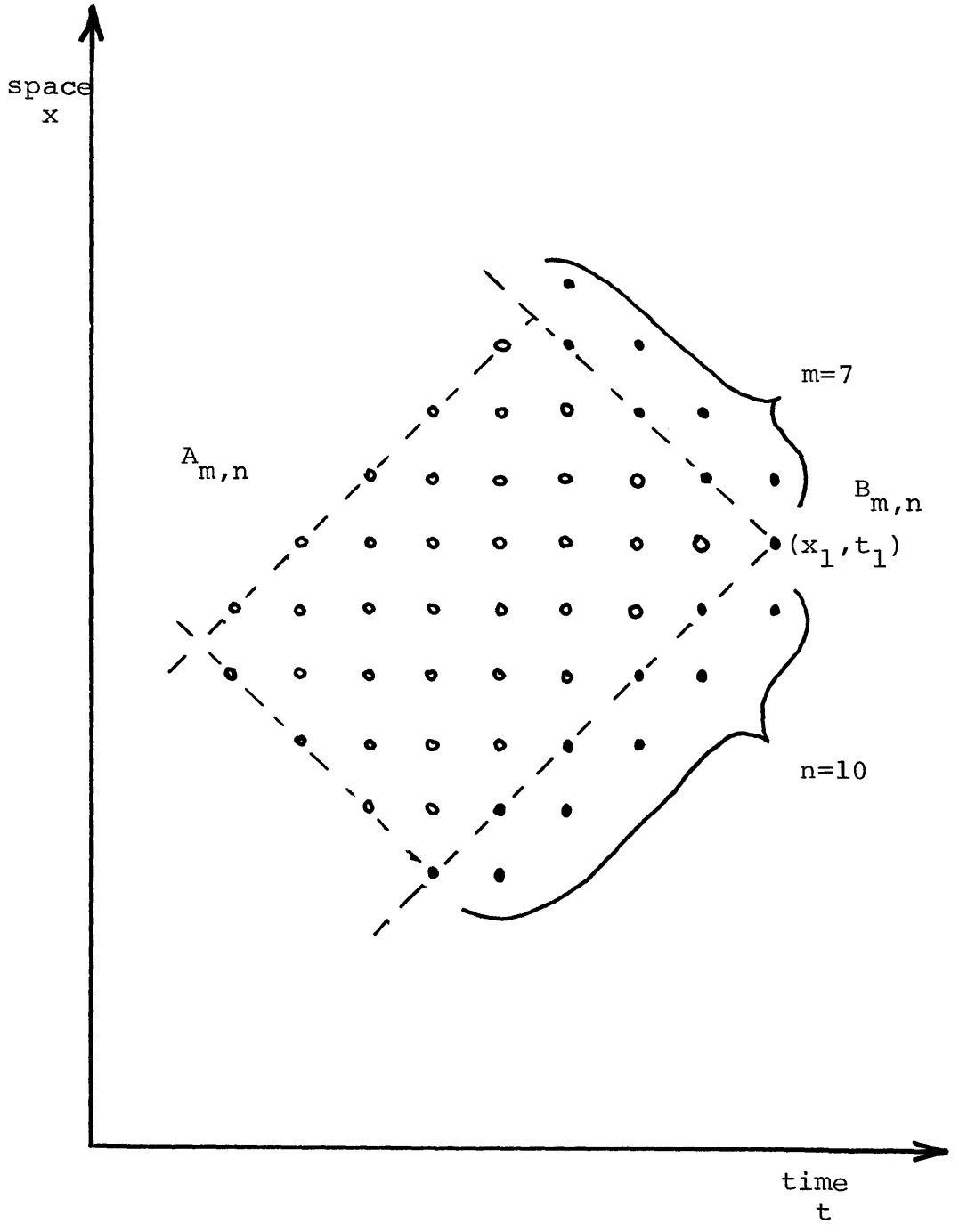


Figure 4.4

computed. (Refer to figure 4.5 in the following discussion.) Suppose one next wishes to calculate $\psi|A_{m+1,n}$. Then first use $\psi|B_{m,n}$ in equation (4.6) to find $\psi|B_{m+1,n}$. Note that to do this requires using the one-parameter recursion (4.6) only once, using only one value of $\psi(x,t)$ for (x,t) in $B_{m,n} - B_{m-2,n}$. Next, the calculation of $\psi|A_{m+1,n}-A_{m,n}$ from $\psi|B_{m+1,n}$ and $\psi|A_{m,n}$ requires only the one value of $\psi|B_{m+1,n}$, namely $\psi|B_{m+1,n}-B_{m,n}$, and the m values of $\psi|A_{m,n}$, namely $\psi|A_{m,n}-A_{m-1,n}$. This calculation requires using m times. Putting $\psi|A_{m+1,n}-A_{m,n}$ togehter with the previously calculated $\psi|A_{m,n}$, one has the new result $\psi|A_{m+1,n}$. In a similar fashion, one can calculate $\psi|A_{m,n+1}$ from $\psi|B_{m,n}$ and $\psi|A_{m,n}$ using (4.7) and (3.6). Thus, in calculating $\psi(x,t)$ for $(x,t) < (x_1,t_1)$ one is free to choose the sequence (m_ν,n_ν) for positive integers ν, and calculate $\psi|A_{m_\nu,n_\nu}$ from $\psi|A_{m_{\nu-1},n_{\nu-1}}$ successively for each ν. The sequence $\{(m_\nu,n_\nu) : \nu \geq 1\}$ is arbitrary except that $(m_1,n_1) = (1,1)$, and for each ν either $(m_{\nu+1},n_{\nu+1}) = (m_{\nu+1},n_\nu)$ or $(m_{\nu+1},n_{\nu+1}) = (m_\nu,n_{\nu+1})$.

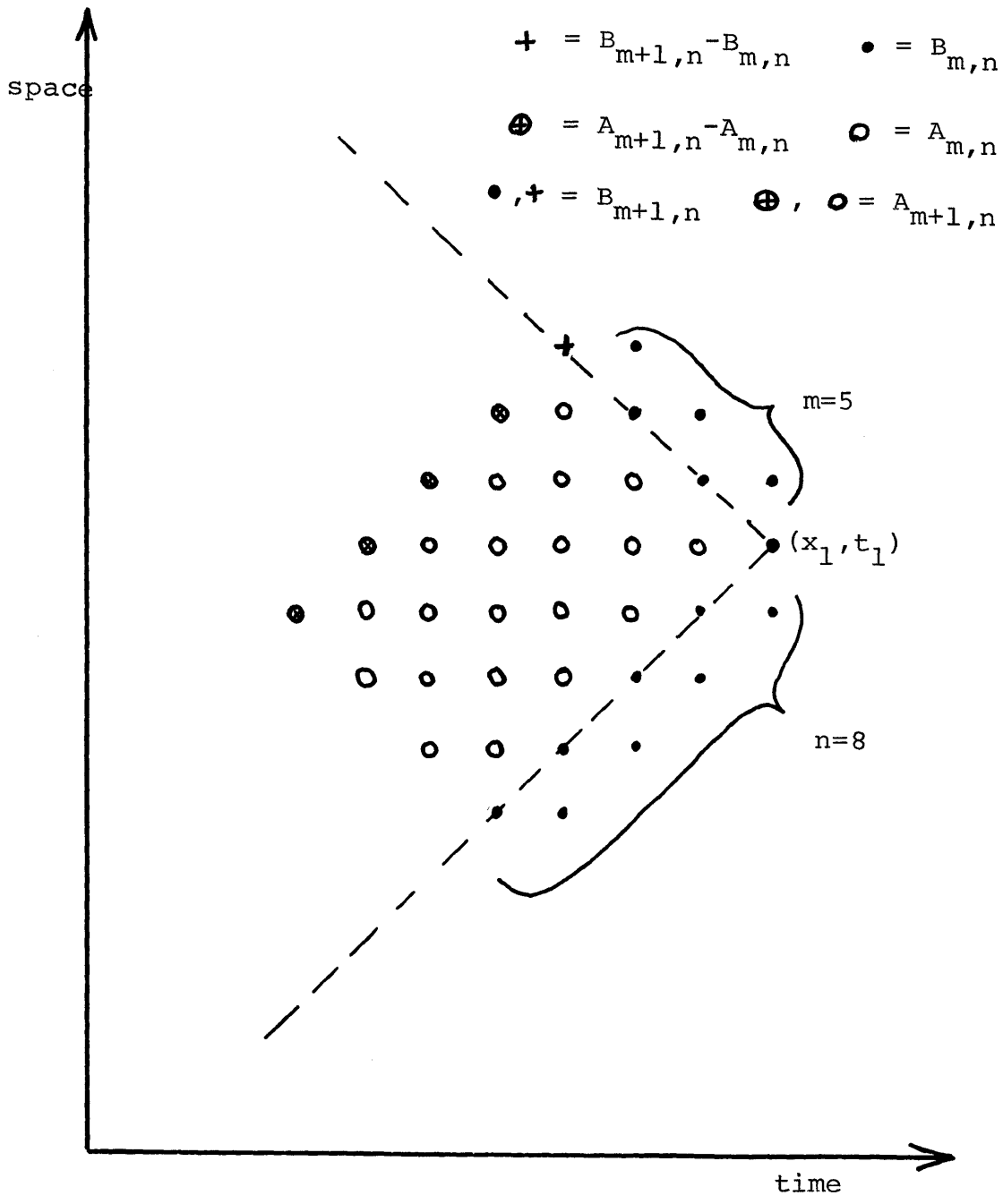
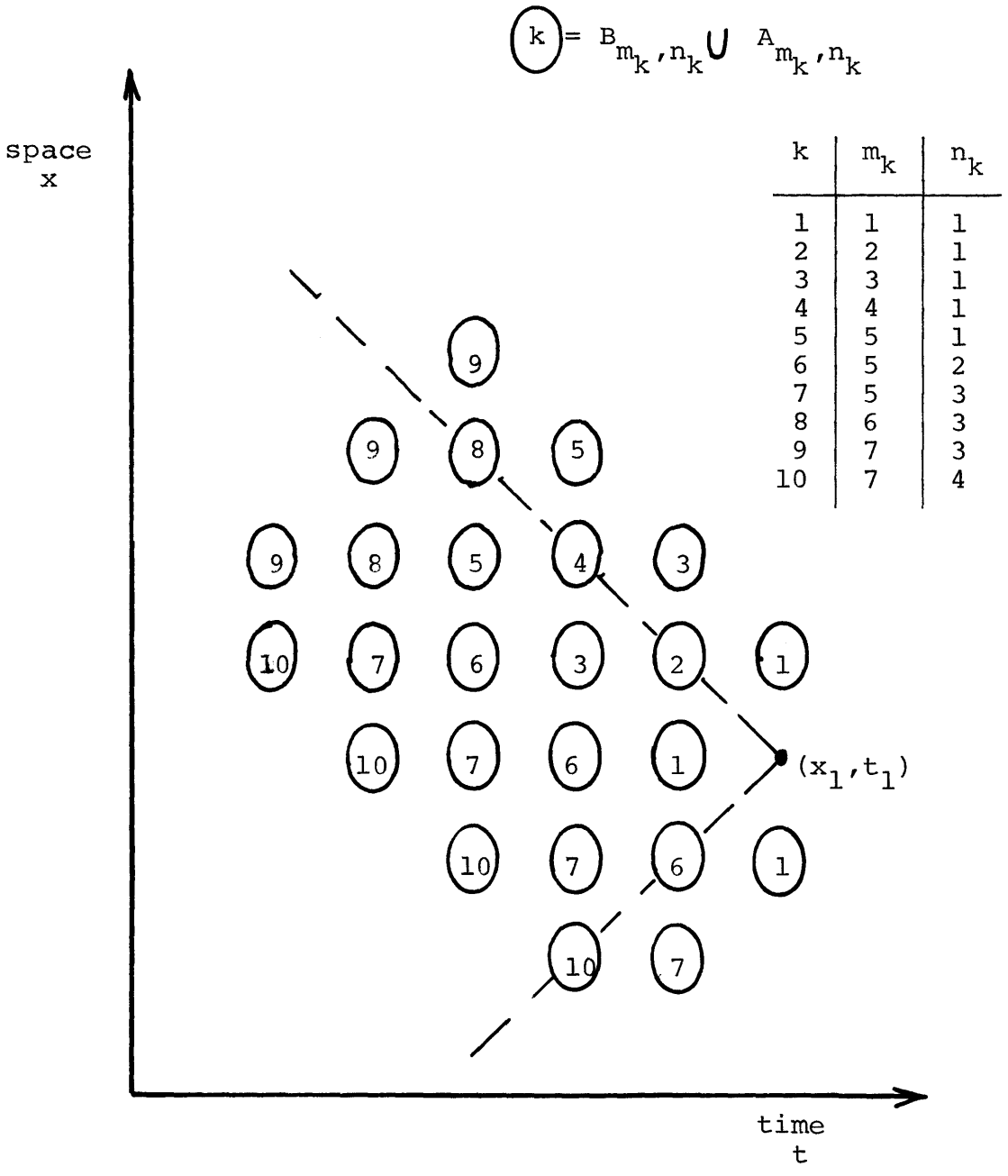


Figure 4.5

Figure 4.6 illustrates one such choice of sequence and the successive calculation of ψ. The freedom to choose how to carry out the recursive calculation is one major characteristic of multiparameter recursive calculations. The space-time order relation defines a precedence relation for the computation of ψ in the sense described by Chan (1976). That is, to compute ψ at the point (x,t) one must first compute ψ at the points (y,s) for which $(x,t) \leq (y,s)$.[1] Thus, if (x,t) and (y,s) are two space-time points, and if $(x,t) \leq (y,s)$, then it is necessary to compute $\psi(y,s)$ before computing $\psi(x,t)$. If neither $(x,t) \leq (y,s)$ nor $(x,t) \geq (y,s)$, then $\psi(x,t)$ and $\psi(y,s)$ can be computed simultaneously or in parallel. This precedence relation in computation and parallelism in computation is a general property of the more general problem described in chapter 3. Note that one advantage of a multiparameter point of view is that we may use the multiparameter structure to construct efficient algorithms.

After this lengthy discussion of the general calculation of ψ from (3.6) in the case of one spatial

[1] More exactly, at the points (y,s) for which $s = t+1$, and for which $y = x+1, x, x-1$.



k	m_k	n_k
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	5	2
7	5	3
8	6	3
9	7	3
10	7	4

Figure 4.6

dimension, let us finish by illustrating the calculation with a specific example. In addition to the assumptions made at the beginning of the chapter, assume that $\bar{q}(u,v)$ is given by

$$(4.10) \qquad \bar{q}(u,v) = 1 - v^2u^2(3v^2-4vu^2+2u^4)$$

and that $q(x,t)$ is given in terms of $\bar{q}$ as

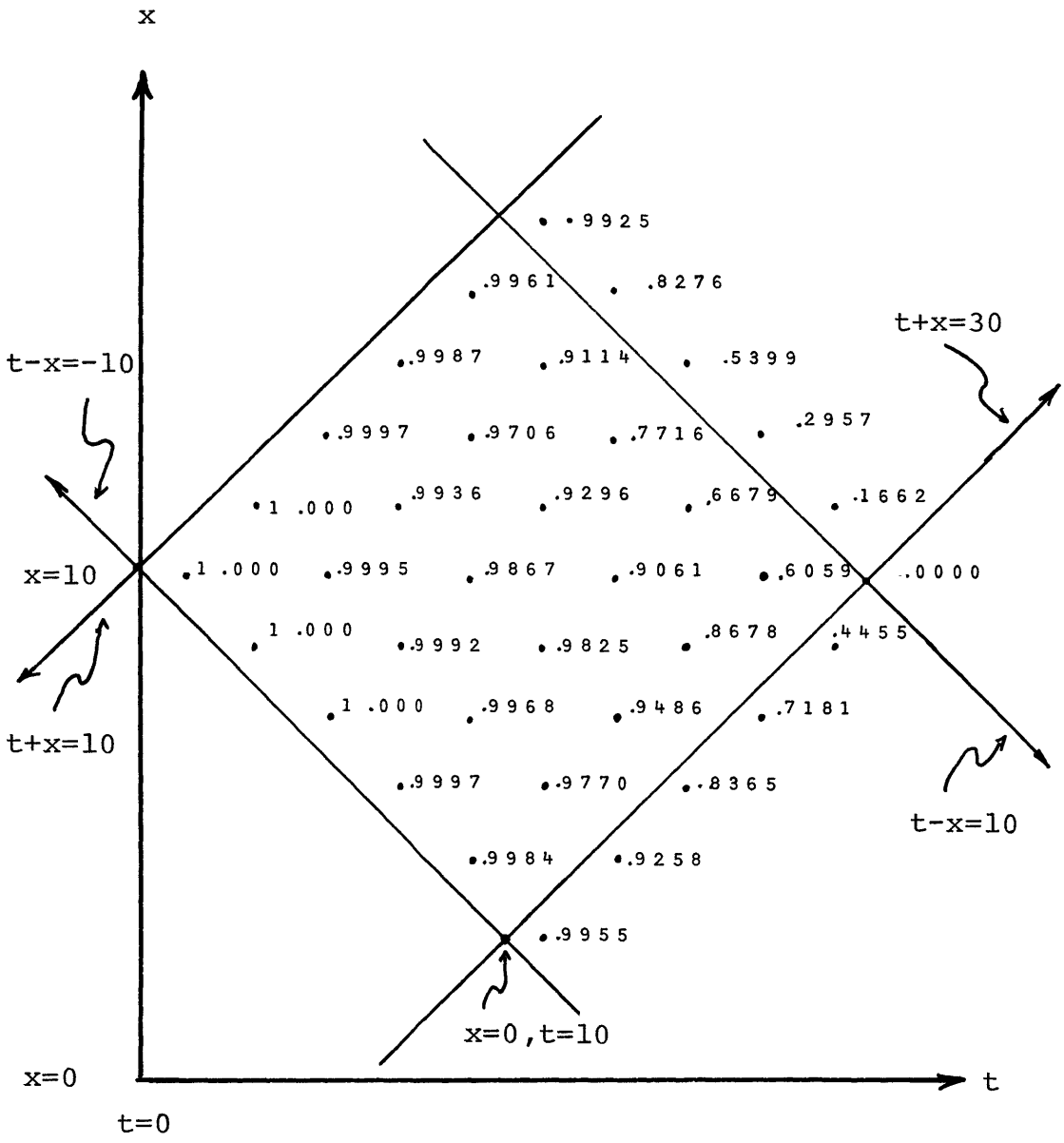
$$(4.11) \qquad q(x,t) = \bar{q}\left(\frac{t - x + 10}{20}\ ,\ \frac{t + x - 10}{20}\right)$$

for (x,t) lying in the square defined by

$$(4.12) \qquad -10 \leq t - x \leq 10$$

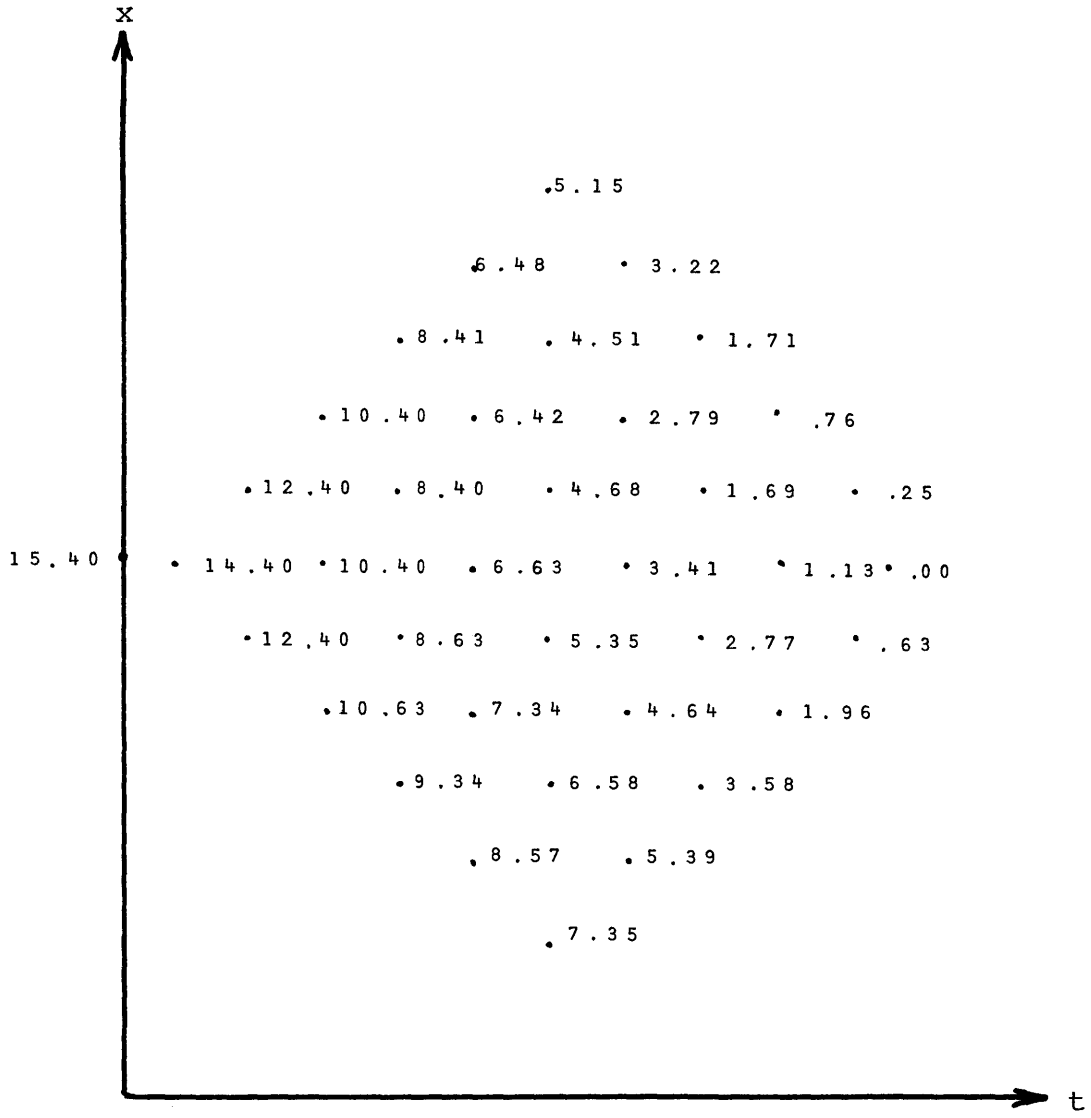
$$10 \leq t + x \leq 30$$

Note that $q(10,20) = 0$. See figure 4.7 for the tabulation of $q(x,t)$. In figure 4.8 we have tabluated the values of $\psi(x,t)$ in the region defined by (4.12). The set $B_{20,-20}$ is denoted by points marked "x" in figure 4.8; the remaining points form the set $A_{-20,-20}$. Finally, in figure 4.9 we have drawn the optimal velocity vector field that one obtains from solving (3.8) as described in



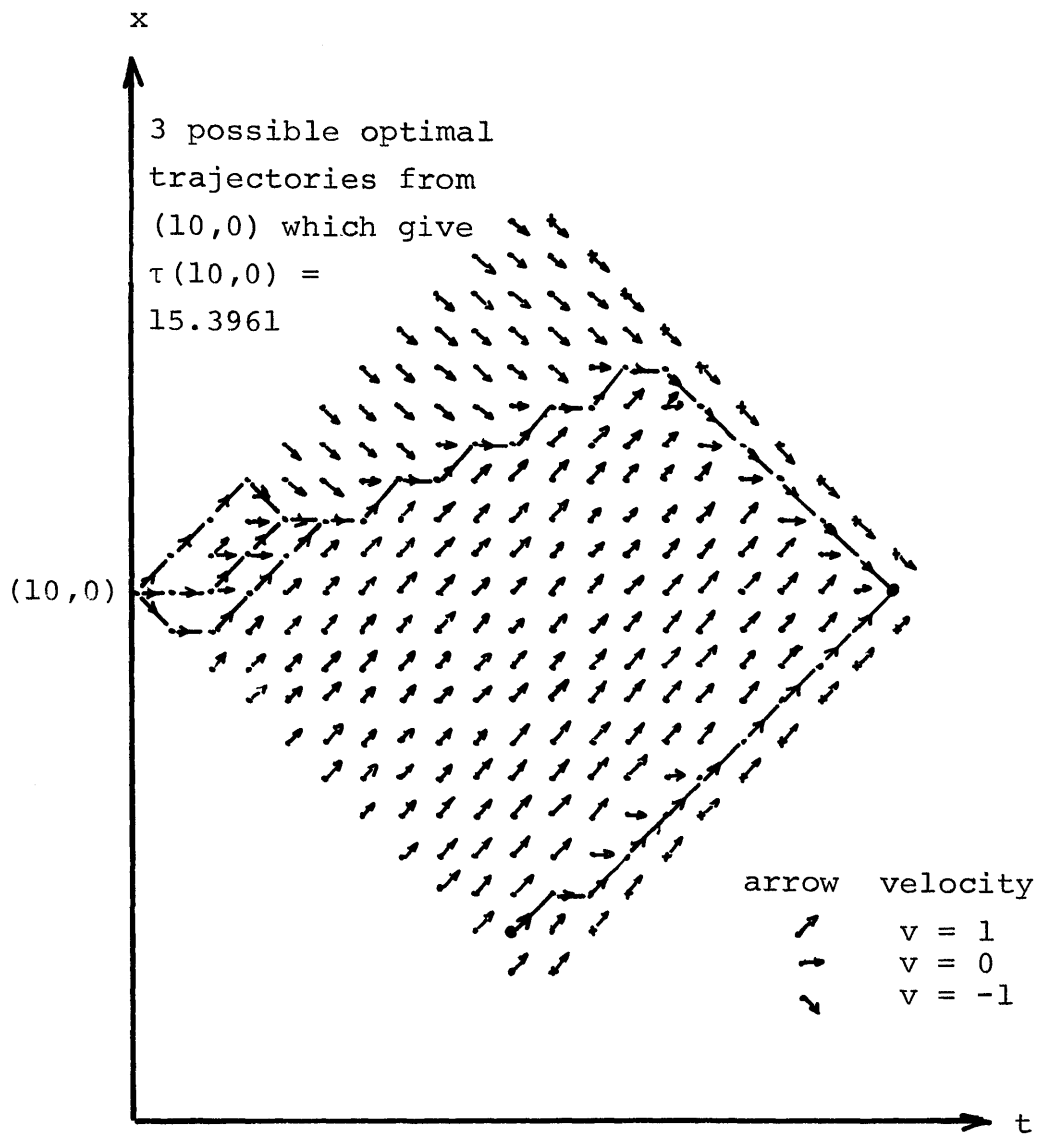
Tabulation of q(x,t)

Figure 4.7



Tabulation of $\psi(x,t)$

Figure 4.8



Optimal Velocity Vector Field and Optimal Trajectories

Figure 4.9

section 2.3. In addition, figure 4.9 shows some optimal trajectories from points within the region described by (4.12). The least expected time to interception appears beside each trajectory.

2.5 Conclusion to chapter 2

In this chapter we have seen how the signal interception problem, formulated originally as a one-parameter problem, can also be interpreted naturally as a two-parameter problem. Instead of specially distinguishing the time variable from the spatial variable, we treated both together as a two dimensional parameter, space-time. The success of this two-parameter treatment depends on the structure of the original problem -- in particular, the original problem must be able to fit into a partial order framework. In the next chapter we will describe this particular framework in more detail and generality, but for now let us note that the p.o. formulation has two basic requirements:

(1) one must be able to formulate the problem as a sequential stopping problem for sequential processes monotonically increasing in a p.o. set (e.g., this is automatically true for any trajectory in space-time which travels slower than the speed of the signal);

(2) one must be able to express the information available for decision-making as an increasing family of σ-fields indexed by the p.o. set (e.g., this is automatically true in space-time if one assumes that information cannot propagate faster than the given signal speed c).

In the next chapter we show that many different problems can be formulated according to (1). Likewise, the requirement (2) is reasonable for many problems, and we discuss what it means intuitively in chapter 3. Note that (2) defines exactly what Ho and Chu (1972) call a partially nested information structure in their study of team decision problems. We discuss their work in the next chapter.

Given the abstract partial order structure, we find that many of the features of the signal interception problem remain true in the general case. Specifically, there is a general version of the dynamic program (3.6) which can be solved uniquely given a terminal condition similar to the multiparameter terminal condition described in section 2.3. The solution is computed by backward recursion with respect to the partial order relation and we find that the partial order defines a precedence relation for computation at comparable points (comparable points with respect to the partial order) and that incomparable

points can be computed in parallel just as for the space-time example of this chapter.

On the other hand, the solution of section 2.4 does not extend to general p.o. structures. Indeed, we are not able to compute the terminal condition in the signal interception problem if the spatial dimension is greater than 1. The reason for this is the peculiar structure of the space-time partial order relation. For one spatial dimension, the space-time order is equivalent to a coordinate-wise partial order defined on a subset of R^2 as one can see by rotating the time and space axis 45^o. For n spatial dimensions where $n > 1$ the space-time order is not equivalent to a coordinate-wise ordering of R^{n+1} -- indeed, for $n > 1$ the partially ordered space-time is not even a lattice (as it would have to be if it were order isomorphic to the coordinate-wise ordering of R^{n+1}).

Although the solution of section 2.4 does not generalize to higher dimensional space-time problems, this solution does have an extension to problems which have the p.o. structure of coordinate-wise ordering of R^n. However, we will not discuss this generalization any further. Many researchers have found the special structure of the

coordinate-wise ordering of R^n attractive for defining and studying multiparameter systems. Much of this work has centered around the particular case $n = 2$ with the goal of developing efficient algorithms for processing images. For example, Woods and Radewan (1977) discuss some computationally efficient extensions of Kalman filtering, a one-parameter recursive estimation algorithm, to two-parameter random fields. Other work has been done on extending concepts from time series analysis (e.g. autoregressive moving average models) to recursive two-parameter random fields. See Larrimore (1977) for a general discussion.

For a general survey of work in multiparameter systems see Bose (1977). Willsky (1976) gives a general survey of recent work in two-parameter recursive systems for which recursion is defined in terms of the coordinate-wise ordering of R^2.

CHAPTER 3

STOPPING PROBLEMS WITH PARTIALLY NESTED INFORMATION

Section 3.1. Introduction

The signal interception problem of chapter 2 can be generalized greatly in terms of stochastic processes indexed by p.o. sets and adapted to a family of σ-fields increasing with respect to the p.o. relation on the index set.* This chapter presents such a generalization: stopping problems with partially nested information structure. In terms of the p.o. set terminology of section 1.2 the abstract stopping problem is to choose an increasing trajectory in a p.o. set in such a way that it minimizes a terminal cost exacted at the last point of the trajectory. What distinguishes this problem from conventional stopping problems is the causality constraint discussed in section 2.2, generalized to p.o. sets. In the general case this constraint requires that decisions made at a point of the trajectory depend on the information available at that point. What is novel is that the information varies from point to point in a particular way compatible with the structure of the partially ordered set.

---

*Refer to chapter 1, section 1.2 for the pertinent definitions.

Specifically, the available information is represented by an increasing family of σ-fields indexed by the p.o. set. Ho and Chu (1972) have coined the term partially nested information to describe this situation and we adhere to their terminology.

To avoid technical complications in the formulation of the stopping problem with partially nested information we consider only p.o. time sets which are discrete and bounded as we now explain.

A p.o. set $(T,\leq)$ is discrete if it is countable and if the set $\{s : r \leq s \leq t\}$ is finite for each r and t in T. The p.o. set $(T,\leq)$ is discrete and bounded above (or just discrete bounded) if the set $\{s : t \leq s\}$ is finite for each t in T.

A time t in T is called a terminal time if there is no s in T such that $t < s$. In other words, the terminal times are the maximal upper bounds with respect to the partial order. If $(T,\leq)$ is discrete bounded then for each s in T there is a terminal time t such that $s \leq t$. This property of discrete bounded time sets leads to the following useful induction principle.

Induction Principle for discrete bounded time sets

Suppose that $(T,\leq)$ is a discrete bounded partially

ordered set and suppose that $P(t)$ is a logical proposition for each t in T. Further suppose that

(1) $P(t)$ is true for all terminal times t in T;

(2) if $P(s)$ is true for all s in T such that $t < s$, then $P(t)$ is also true.

Then $P(t)$ is true for all t in T.

Proof

Call s an immediate successor of t and write $s <\cdot t$ if $s < t$ and there is no r such that $s < r < t$. Let T_0 denote the set of all terminal times in T and define T_{n+1} as the set of all immediate successors of T_n. That is,

$$T_{n+1} = \{s : s <\cdot t,\ t \in T_n\}$$

for $n \geq 0$. If T is discrete bounded then for every s in T there is a terminal time t such that $s \leq t$. Moreover, the set $\{r : s \leq r \leq t\}$ is finite; say it has n members. Then it is not hard to see that s belongs to T_n. Thus, $T = \bigcup \{T_n : n \geq 0\}$ if T is discrete and bounded.

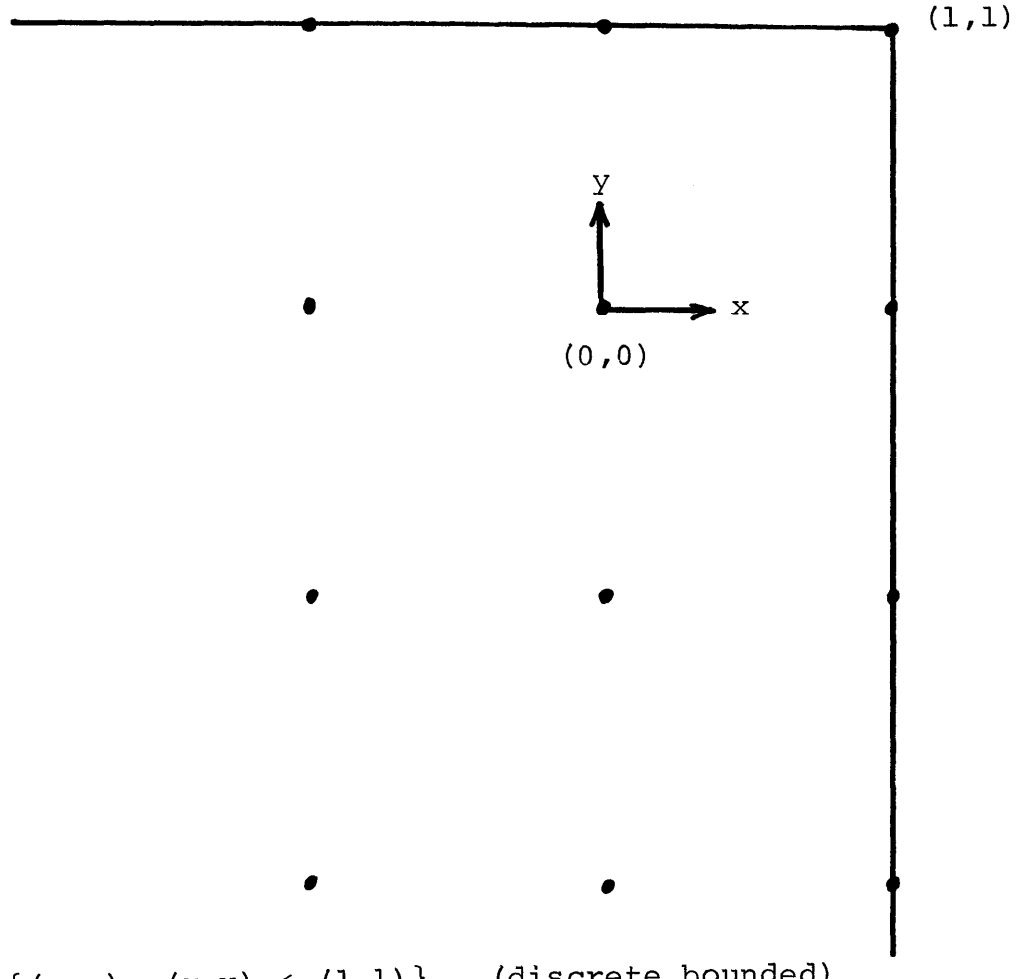
Let P_n be the proposition that $P(t)$ for all t

in T_n. Then (1) implies that P_0 is true, and (2) implies that P_{n+1} is true if P_n is true. According to the ordinary principle of mathematical induction, it follows that P_n is true for all $n \geq 0$. Consequently, $P(t)$ is true for all t in T. ///

Note that this inductive principle is a formalization of the multiparameter backward recursion described in sections 2.3 and 2.4.

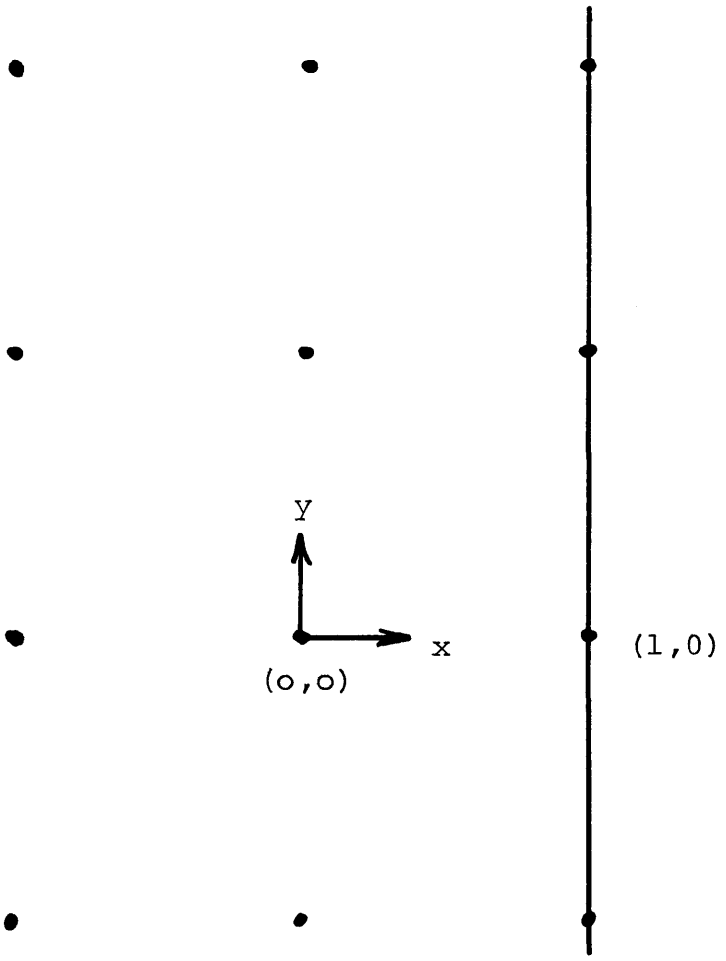
As an example of discrete bounded time sets, consider the multiparameter ordering of Z^2, the Cartesian product $Z \times Z$ of the integers Z. The set Z^2 is discrete but not bounded. The set $\{(x,y) : (x,y) \in Z^2, (x,y) \leq (1,1)\}$ is discrete bounded, but the set $\{(x,y) : (x,y) \in Z^2, x \leq 1\}$ is discrete but not bounded. Another set which is discrete bounded is $\{(x,y) : (x,y) \in Z^2, x+y \leq 1\}$. These sets are illustrated in figures 1.1, 1.2 and 1.3 respectively.

The case of unbounded (but discrete) time sets may be treated by adding a few reasonable assumptions to the stopping problem as we note in the concluding section 3.5. The case of nondiscrete (i.e. continuous parameter) time is more difficult and unsolved at the present time. In the concluding chapter 5 of part I we discuss the problems involved in the continuous case.



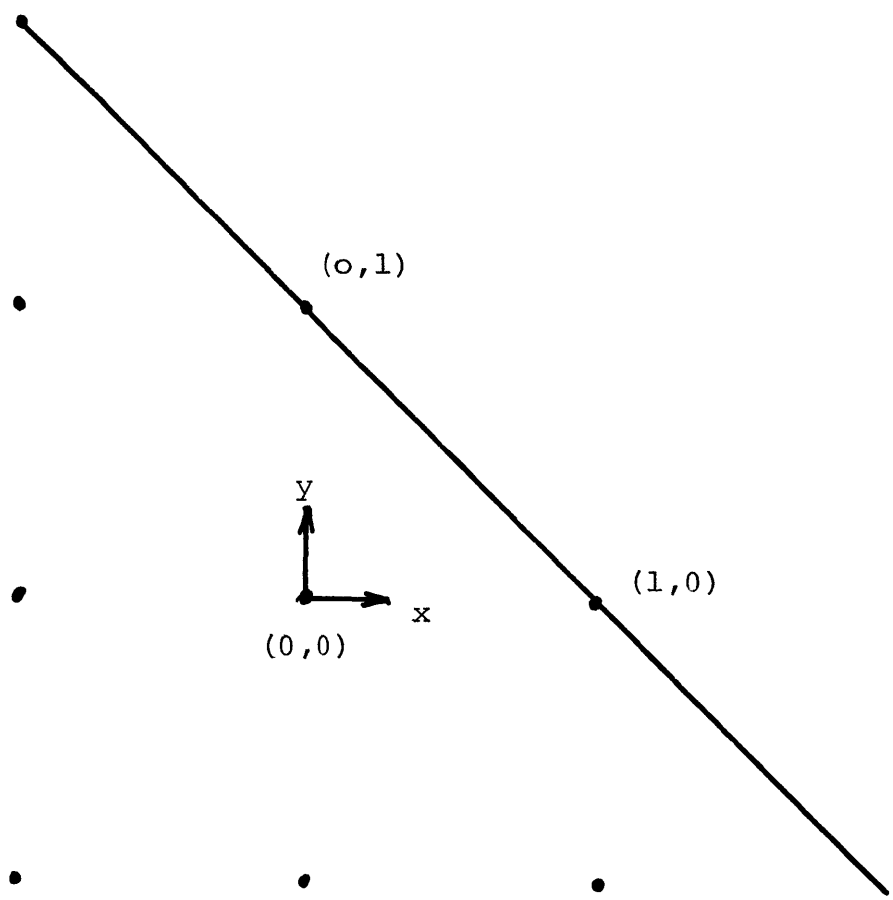
$\{(x,y): (x,y) \leq (1,1)\}$ (discrete bounded)

Figure 1.1



{(x,y): $x \leq 1$ } (discrete, but not bounded)

Figure 1.2



$\{(x,y): x + y \leq 1\}$ (discrete bounded)

Figure 1.3

After presenting the general abstract stopping problem with partially nested information in section 3.2, we discuss the intuitive meaning behind the various assumptions made in the abstract formulation in order to indicate what physical problems might fit into the framework of the general stopping problem. In this context we discuss the analogy between our stopping problem and an optimal control problem. Moreover, we note the dual control aspects of the stopping problem regarded as a dual control problem in the sense of Fel'dbaum (1960). To clarify the special information structure of our stopping problem we discuss the partially nested information structure defined by Ho and Chu (1972) in their study of team decision problems.

Next in section 3.2 we define a general space-time partial order relation and show how the signal interception problem of chapter 2 fits into the general framework of the stopping problem with partially nested information. This discussion of the space-time interception problem helps illustrate the special structure of the stopping problem. Although many examples fit into the general framework, some simple examples do not. As we discuss, simple sequential decision problems without perfect

memory do not satisfy the basic assumption of partially nested information. Witsenhausen (1968, 1971) has considered such counterexamples and the general difficulty presented by such "nonclassical" information structures as for example, imperfect memory.

After discussing the space-time interception problem, we present a quite different problem concerning the optimal exploration for oil or mineral resources. This problem fits easily into the general framework when the p.o. relation is defined as set inclusion on subsets of the region to be explored.

In section 3.3 we present the dynamic programming solution of the stopping problem with partially nested information. This dynamic program is the generalization of (2.3.6) to the p.o. set case. Although Bertele and Brioschi (1972) have considered nonserial dynamic programming -- dynamic programming in cases for which the "time" is not linearly ordered or serial, they do not consider the particular nonserial case of partially ordered time. The assumption of p.o. time and partially nested information creates a backward recursive structure which Bertele and Brioschi do not have in the cases they consider. Backward recursion is best described in terms of the induction principle for p.o. sets we presented above.

Starting from terminal times, one solves the dynamic programming equation by working backward with respect to the partial order. The backward recursive structure of the dynamic program for our problem facilitates efficient computation by indicating clearly what computations must precede others and what computations may be done in parallel. In our problem the p.o. relation on the time coincides with the precedence relation for computation as described by Chan (1976).

The proof of the necessity that the optimal cost function satisfies the dynamic program equation (Theorem 3.1) is very similar to the serial dynamic programming case as described in Bellman (1957) for example. The properties of the serial dynamic program generally hold for the p.o. dynamic program. Thus, we have a necessary and sufficient characterization of the optimal decision functions in terms of the dynamic program for the optimal cost (Corollary 3.2). Although there may be several different optimal decision functions which solve one stopping problem, for given terminal conditions there is only one optimal cost function which solves the dynamic program equation (Theorem 3.3). To complete section 3.3 we solve a general version of the signal interception problem of chapter 2 using the general

dynamic program.

In section 3.4 we discuss the work of Haggstrom (1966). He developed the stopping problem with partially nested information structure in the special case when the p.o. time set is a tree.* We briefly describe Haggstrom's work and then discuss its relation to our own. Finally, in section 3.5 we conclude this chapter and discuss possible extensions.

3.2 Stopping problem with partially nested information structure

Let us formally define the stopping problem with partially nested information structure before discussing its intuitive meaning. To avoid technical complications which are not essential to the basic development of our stopping problem, we assume that the partially ordered time set $(T,\leq)$ is discrete bounded. In the conclusion of part I we discuss the significance of the discrete bounded hypothesis and indicate how one might extend the theory to more general time sets, particularly to discrete unbounded time sets and time sets like R^n which have a topological structure.

---

*See section 3.4 for a precise definition.

Assume that $\{F_t : t \in T\}$ is an increasing family of σ-fields with respect to the underlying probability space (P,Ω,F) and the partially ordered set $(T,\leq)$. We call T the time set and $\{F_t : t \in T\}$ the information structure for the problem.

A cost function or terminal cost c is a mapping from $T \times \Omega$ into the extended real numbers $R^* = R \cup \{+\infty\}$ such that for each t in T the random variable $\omega \to c(t,\omega)$ is F_t-measurable. We often suppress the ω in $c(t,\omega)$ and let $c(t)$ denote the random variable $\omega \to c(t,\omega)$. We want to express expectations like $E(c(t))$ and yet allow for $c(t)$ to take the value $+\infty$. Therefore, let us assume that the negative part of $c(t)$ is absolutely integrable. That is,

$$E(|c(t)| \cdot 1_{c(t)<0}) < \infty \tag{2.13}$$

for each t in T. As always, 1_A denotes the indicator function of the subset A of Ω. With (2.13) we may always interpret $E(c(t))$ as an element of R^*.

A constraint set is a collection $\{C_t : t \in T\}$ of random functions mapping Ω into subsets of T such that for each t in T we have

(2.14) for all ω, t, $s \in C_t(\omega)$ implies $t \leq s$,

(2.15) for all ω, $C_t(\omega) \neq \emptyset$,

(2.16) for each subset C of T the set $\{\omega : C_t(\omega) = C\}$ is F_t-measurable.

Corresponding to $\{C_t : t \in T\}$ is the class D of admissable decision functions which is the set of all maps ϕ from $T \times \Omega$ into T such that

(2.17) for all t in T and ω in Ω, $\phi(t,\omega) \in C_t(\omega)$,

(2.18) for all t and s in T, $\{\omega : \phi(t,\omega) \leq s\}$ is F_t-measurable.

To state the stopping problem requires a preliminary definition. For ϕ in D define ϕ^0 as $\phi^0(t,\omega) = t$ for all t and ω. Then define ϕ^{n+1} in terms of ϕ^n as $\phi^{n+1}(t,\omega) = \phi(\phi^{n+1}(t,\omega),\omega)$. Properties (2.17) and (2.14) imply that $t \leq \phi(t,\omega)$ for all t and ω. Thus, the sequence $\{\phi^n(t,\omega) : n \geq 0\}$ is entirely contained in $\{s : t \leq s\}$ and clearly, $\phi^n(t,\omega) \leq \phi^{n+1}(t,\omega)$ for each $n \geq 0$. Since T is discrete bounded, the set $\{s : t \leq s\}$ and hence also $\{\phi^n(t,\omega) : n \geq 0\}$ is finite. It follows that we must have $\phi^{n+1}(t,\omega) = \phi^n(t,\omega)$ for some n depending only on t. If $k \geq n$, it is clear from the

definition of ϕ^k that $\phi^k(t,\omega) = \phi^n(t,\omega)$ for all $k \geq n$. Thus, we have for each t and ω that $\{\phi^n(t,\omega) : n \geq 0\}$ possesses a well-defined limit which we denote $\phi^\infty(t,\omega)$. To denote the random function $\omega \to \phi^\infty(t,\omega)$ we will suppress the ω and write $\phi^\infty(t)$. In section 4.1 of the next chapter we will see that the random function $\phi^\infty(t)$ is a stopping time on the p.o. set T in the sense of section 1.2. In fact, the function $\phi^\infty(t)$ has more structure than a general stopping time -- for example, the optional sampling theorem is true for this particular stopping time, as we also show in section 4.1. Indeed, in chapter 4 we define the class of reachable stopping times as those stopping times which can be written as $\phi^\infty(t)$ for some decision function ϕ. Remarkably, we find that all stopping times defined for the two-parameter p.o. set with the probability structure discussed in chapter 1 are reachable in this sense. Returning to the present chapter, we can now define the general stopping problem with partially nested information structure.

The stopping problem for the partially nested information structure $\{F_t : t \in T\}$ is to find for a given initial time Θ in T an optimal decision function ϕ_* in D such that

(2.19) $$E(c(\phi_*^\infty(\theta))\,|\,F_\theta) \leq E(c(\phi^\infty(\theta))\,|\,F_\theta)$$

for all ϕ in D.

The assumptions we have made about c, T and D guarantee that $c(\phi^\infty(t))$ is integrable for each t in T and ϕ in D, although the integral may be $+\infty$. To see this let $c^n(t,\ )$ denote $c(\phi^n(t,\omega),\omega)$. We defined $\phi^{n+1}(t,\omega) = \phi(\phi^n(t,\omega),\omega)$, but it should be clear that $\phi(\phi^n(t,\omega),\omega) = \phi^n(\phi(t,\omega),\omega)$. Thus, $c^{n+1}(t,\omega) = c^n(\phi(t,\omega),\omega)$ for $n \geq 0$, where c^0 is just c. For $n \geq 0$ we can express c^{n+1} as

(2.20) $$c^{n+1}(t,\omega) = \sum_{t\leq s} c^n(s,\omega)\,1_{\phi(t,\omega)=s}.$$

Note that (2.18) implies that $1_{\phi(t,\omega)=s}$ is F_t-measurable and hence, F-measurable for each t and s. In addition, the discrete boundedness of T implies that the sum over s in (2.20) is a finite sum. Thus, if $c^n(t)$ is F-measurable and if the negative part of $c^n(t)$ is absolutely integrable for each t in T, then the same holds true for c^{n+1}. Since $c^0(t) = c(t)$ is F-measurable and has an absolutely integrable negative part, (2.20) implies this is true of c^n for all $n \geq 0$.

For each t in T there is an integer $n \geq 0$ such that $\phi^{\infty}(t) = \phi^{n}(t)$. In that case, $c(\phi^{\infty}(t)) = c^{n}(t)$ and one sees that $c(\phi^{\infty}(t))$ is $\mathcal{F}$-measurable and the negative part of $c(\phi^{\infty}(t))$ is absolutely integrable. Thus, the expectation and conditional expectations of $c(\phi^{\infty}(t))$ are well-defined, although perhaps infinite.

Having defined mathematically the stopping problem with partially nested information structure, let us now explain the intuitive content of this definition.

To begin with, we examine the problem as a discrete time optimal control problem. For the sake of comparison, consider the following conventional discrete time optimal control problem. The integer variable n plays the role of time in this problem and $0 \leq n \leq N$. Suppose that c maps R^{p} into R^{1} and f maps $R^{p} \times R^{q}$ into R^{p}. For a given initial state x_0 in R^{p} the optimal control problem is to find a mapping γ from R^{p} into R^{q} such that $c(x(N))$ is minimum when

$$x(n+1) = f(x(n),u(n)), \quad 0 \leq n \leq N-1 \tag{2.21}$$

$$x(0) = x_0, \tag{2.22}$$

$$u(n) = \gamma(x(n)), \quad 0 \leq n \leq N-1 \tag{2.23}$$

The values $u(n)$ are called <u>controls</u> and γ is the <u>control law</u>. The function f represents the dynamical system and c is the <u>terminal cost</u> Thus, the problem is to steer the deterministic process $\{x(n) : 0 \leq n \leq N\}$ by means of the control law γ so as to minimize the terminal cost of the process, $c(x(N))$. If one neglects for a moment the probabilistic aspects of the stopping problem, one sees that the stopping problem is just such an optimal control problem. Indeed, the function ϕ in D serves as the control law steering the process $\{t(n) : 0 \leq n\}$ which is defined by

$$t(n+1) = u(n) \tag{2.24}$$

$$t(0) = \Theta \tag{2.25}$$

$$u(n) = \phi(t(n)). \tag{2.26}$$

We have suppressed the ω-dependence in writing $t(n)$ and $\phi(t(n))$ to emphasize the similarity to (2.21), (2.22) and (2.23). The initial time Θ corresponds to x_0 in (2.22). The decision function ϕ corresponds to the function $x \to \gamma(x)$ mapping the state space R^p into the control space R^q. Note that instead of requiring the process $\{t(n) : 0 \leq n\}$ to obey some particular dynamical relation as $\{x(n) : 0 \leq n \leq N\}$ obeys (2.21), we have required that

ϕ belong to the set D of admissable decision functions satisfying properties (2.17) and (2.18). The choice of a particular constraint set $\{C_t : t \in T\}$ corresponds to the choice of a particular dynamic relation (2.21).

As an example of the correspondence between the stopping problem and the optimal control problem, consider the one-parameter stopping time problem with T being $\{0,1,2,\ldots n\}$. In the usual formulation of the one-parameter stopping time problem, one looks for a stopping time τ on T such that the expected terminal cost $E(c(\tau))$ is minimized. By defining ϕ so that

$$(2.27) \qquad \begin{aligned} \phi(t,\omega) &= t \qquad \text{if} \quad \tau(\omega) \leq t \\ &= t+1 \qquad \text{if} \quad \tau(\omega) > t, \end{aligned}$$

one sees that the usual one-parameter stopping time problem can be formulated in the optimal control problem framework. The function ϕ in (2.27) is a control law that steers the times $t(n)$, defined as in (2.24), (2.26) to the optimum final time τ. Note that ϕ has the right measurability property (2.18) -- if τ is a stopping time with respect to $\{F_t : t \in T\}$, then $\omega \to \phi(t,\omega)$ is F_t-measurable for each t. Thus, the one-parameter stopping time problem is a stopping problem with partially nested

information as we have defined it. Note also that $\tau(\omega) = \phi^{\infty}(0,\omega)$.

The interpretation of the stopping problem as an optimal control problem is essential. For the case of one-parameter time the control interpretation is equivalent to the original stopping time problem. For general partially ordered time sets the optimal control interpretation is not equivalent to the usual stopping time interpretation for general stopping times on p.o. sets as defined in section 1.2. The reason for this non-equivalence is closely related to the failure of the optional sampling theorem for general stopping times defined on p.o. sets. In section 4.1 of the next chapter we investigate this problem further. We find that if we restrict ourselves to the class of reachable stopping times, briefly mentioned earlier in this section, then the problem of minimizing the expectation $E(c(\tau))$ for reachable stopping times τ such that $\theta \leq \tau$, is equivalent to the stopping problem with partially nested information that we have defined in this section. If we try to minimize $E(c(\tau))$ with respect to all stopping times τ defined on a p.o. set, then we find that the problem is much more difficult (e.g. we lose the dynamic programming solution). Moreover, it is possible that a general stopping time will

yield a smaller expected terminal cost than any reachable stopping time, or equivalently, than any decision function. A simple example of this is presented in section 4.1. In the latter half of this section we argue that in many circumstances the decision function formulation is the natural formulation for some optimization problems, and that in these cases one would not want to allow general stopping times on p.o. sets but only reachable stopping times which corresponded to decision functions. In the conclusion of part I in chapter 5 we discuss problems in which the more general stopping time formulation is desirable and for which decision functions and reachable stopping times are not sufficient.

Having shown how the stopping problem is like the optimal control problem, we must now show how it is different.

The difference arises in the particular probabilistic structure of the problem. Consider the optimal control problem set forth in (2.21), (2.22), (2.23) and the stopping problem set forth in (2.24), (2.25), (2.26). In both problems let us refer to the variables n as the <u>stage</u> so that we can talk about the n-th stage of either problem and avoid confusing partially ordered

time in the stopping problem with ordinary linear time in the optimal control problem. In the conventional stochastic optimal control problem one introduces an increasing family of σ-fields $\{F_n : 0 \leq n\}$ indexed by the stage, so that F_n represents the information available for the control decision at the n-th stage of the problem. That is, one requires that the control u(n), now a random variable, be F_n-measurable for each stage n. In our stopping problem, on the other hand, the information does not depend on the stage n but rather the position t(n) at stage n. In other words, the information available for the control decision at stage n is the σ-field $F_{t(n)}$. We will show in chapter 4 that t(n) defined by (2.24), (2.25), (2.26) in our stopping problem is a stopping time, and in fact, a reachable stopping time for any decision function $\phi \in D$ in (2.26). Moreover, $F_{t(n)}$ is defined in terms of the stopping time t(n) and the σ-fields $\{F_t : t \in T\}$ as in section 2.2. One might think to let the increasing σ-fields $\{F_{t(n)} : 0 \leq n\}$ in the optimal stopping problem correspond to $\{F_n : 0 \leq n\}$ in the optimal control problem, but there is an important distinction. The σ-fields $\{F_{t(n)} : 0 \leq n\}$ depend on the decision function ϕ whereas the σ-fields $\{F_n : 0 \leq n\}$ are independent of

the control law γ. Another way of saying this is the the following: going from stage n to $n+1$ in the optimal control problem, the new information at stage $n+1$ does not depend on what we do at stage n; but in the stopping problem, the new information at stage $n+1$ does depend on where in T we decide to go.

Although the stopping problem differs from the conventional optimal stochastic control problem with regard to information structure, it is similar to the so-called dual control problems introduced by Fel'dbaum (1960). The dual control problem is an optimal stochastic control problem in which the information available at stage n depends on the control decisions taken up to stage n. In this problem the control plays the dual role of optimally regulating the controlled process on the one hand, and obtaining information on the other. Often, the two roles conflict so that new information is bought at the expense of optimal control and conversely.

As we saw above, in the stopping problem at stage n information depends on the position $t(n)$ in T; and this position is a function of the control decision represented by ϕ in (2.24). Thus, the stopping problems are a subclass of dual control problems.

Although the information structure for the stopping problem is more complex than for the conventional optimal stochastic control problem, this information structure has special structure induced by the partially ordered time set. We now discuss this special structure, the so-called partially nested information structure.

Ho and Chu (1972) introduced the term partially nested information structure to define a class of solvable team decision problems. The team decision problem, described in Radner (1962) or Marschak and Radner (1972) for example, is to minimize an expected cost which depends on the decisions of several decision makers who do not necessarily have the same information on which to base their decisions. By having different information for different decision makers, the team problem is more challenging than the ordinary optimization problem, but by having only a single cost to optimize, it is less difficult than a game problem.* To facilitate the solution of their team decision problem, Ho and Chu assume a particular relationship between decision makers or agents as they are called and the information of each agent. Suppose A is a finite set of agents. Between

*Game in the sense of the von Neumann - Morgenstern (1944) theory.

each agent there is a precedence relation denoted $i \{ j$ for i and j in A and interpreted to mean that the decision of i affects the information of j. Ho and Chu assume that $\{$ has three properties: (1) $i \{ i$ never occurs for i in A; (2) if $i \{ j$ then one does not have $j \{ i$; (3) if $i \{ j$ and $j \{ k$, then $i \{ k$. It is clear that these properties are very close to those properties of the partially ordered set in (2.1), (2.2) and (2.3). In fact, if one defines $i \leq j$ as the relation "$i \{ j$ or $i=j$" then $(A, \leq)$ is a partially ordered set as defined in section 2.1.

Having established a precedence relation for the set of agents, Ho and Chu next assign a σ-field $\mathcal{F}_i$, the information set, to each agent i. The team problem has a <u>partially nested information structure</u> if $i \{ j$ implies $\mathcal{F}_i \subset \mathcal{F}_j$ for each agent i and j. Thus, a partially nested information structure is equivalent to saying that the σ-fields $\{\mathcal{F}_i : i \in A\}$ are increasing with respect to the partial order defined by the precedence relation. Intuitively, the information structure is partially nested if each agent i has the information of all those agents j whose decisions may affect the information of i.

The specific partially nested information structure and even the existence of such a structure depends on the problem at hand. In concluding this section, we present two examples of problems which naturally require the partially nested information structure. We first reconsider the space-time problem of chapter one and indicate how that problem fits into the general framework presented in this section.

A more general space-time order relation is possible for sets T which are finite subsets of $X \times R$ for which (X,d) is a metric space with distance function d. For x, y in X and t, s in R one generalizes (1.1) so that $(y,s) \leq (x,t)$ if

$$s \leq t, \tag{2.28}$$

$$d(x,y) \leq |t-s|.$$

Such a generalization can model signals propagating through an inhomogeneous medium for which $d(x,y)$ is the least time for a signal to travel between two points x and y in the medium. For example, consider light travelling in a medium whose index of refraction varies in space but not in time.

Associated with the partially ordered space-time $(X \times R, \leq)$ is an increasing family $\{F_{(x,t)}\}$ of σ-fields which naturally represent the information restrictions in a space-time problem. Intuitively, one can understand this as follows. Assume that (1) information can be sent from a space point x to another space point y in a minimum amount of time $d(x,y) > 0$, and (2) received information is not forgotten. The σ-field $F_{(x,t)}$ represents all information available at space point x up to time t. Thus, it includes information from all signals that arrive at x at times before t (since there is no forgetting). If $(x,t) \leq (y,s)$, then a signal sent from x at time t will arrive at y at time $t + d(x,y)$ which is before s. In this signal we send all the information available at x up to time t so that it becomes available to y up to time s. This last assertion is the intuitive statement of the mathematical relation $F_{(x,t)} \subset F_{(y,s)}$. Thus, the increasing property of the σ-fields follows from

(1) the speed limit on sending information

and (2) the perfect memory of the receiver.

The requirement of perfect memory may seem unimportant but it is an essential assumption without which the increasing property of the σ-fields may fail. This

happens even in the simple case when $(T,\leq)$ is totally ordered. For example, let T be the integers and suppose that $\{x_t : t \in T\}$ is a stochastic process. Define $\mathcal{F}_t$ to be the smallest σ-field for which $x_t, x_{t-1}, \ldots, x_{t-n}$ are measurable. The fixed finite integer n represents a finite memory capacity. The extreme case $n = 0$ is the memoryless situation, and the infinite case $n = \infty$ represents perfect memory. For finite memories it is clear that in general one can have $\mathcal{F}_t \not\subset \mathcal{F}_{t+1}$. Witsenhausen (1968, 1971) has discussed such counter-examples and the general difficulty presented by non-classical information structures in stochastic optimal control problems. Classical information structures are those represented by sequentially ordered, increasing families of σ-fields -- the perfect memory situation mentioned above. Thus, the partially ordered, but still increasing families of σ-fields, assumed by Ho and Chu (1972) and by our stopping problem are examples of nonclassical information structures. However, this class of nonclassical information structure (that is, partially nested information structures) is simpler to handle than say, the nonclassical structure represented by finite memory, which is not partially nested. Recently, nonclassical information considerations

have become important in systems theory, particularly in the context of decentralized control. The decentralized control problem is essentially a team problem in which the agents must act with only local information of the state of the entire system. Thus, the information restrictions on the problem are as important as the dynamic equations and the optimality criterion. In chapter 5 we will use the example of decentralized control problems to illuminate the difference between reachable stopping times and general stopping times on p.o. sets.

Returning to the space-time example, we note how the interception problem of chapter 2 fits into the general framework of the present section. We have already seen how the information structure is partially nested in the space-time example. Note that for the interception problem the information was particularly simple: at space-time point (x,t) one knows only whether or not $(\xi(\omega),\Theta(\omega)) \leq (u,s)$ for some (y,s) such that $(y,s) \leq (s,t)$.

One has a choice in setting up the cost function and the constraint set. For example, one choice is to define the cost function c as $c((x,t),\omega) = t$ for all t and ω. This choice of c gives expected time as the

optimization criterion in (2.19), but it does not give the expected time of interception unless we make a further restriction. The natural restriction is to define the constraint set $\{C_t : t \in T\}$ so that $C_{(x,t)}(\omega)$ includes (x,t) if and only if $(\xi(\omega),\theta(\omega)) \leq (x,t)$.
Thus, one defines

$$(2.29)\quad C_{(x,t)}(\omega) = \begin{cases} \{(y,s) : (x,t) \leq (y,s)\} & \text{if } (\xi(\omega),\theta(\omega)) \leq (x,t) \\ \{(y,s) : (x,t) < (y,s)\} & \text{if } (\xi(\omega),\theta(\omega)) \not\leq (x,t) \end{cases}$$

for each (x,t) and ω.

Alternatively, one may define the cost c so that

$$(2.30)\quad c((x,t),\omega) = t \text{ if } (\xi(\omega),\theta(\omega)) \leq (x,t)$$

$$\text{and } c((x,t),\omega) = +\infty \text{ if } (\xi(\omega),\theta(\omega)) \not\leq (x,t).$$

By assigning infinite cost to the decision to stop before interception, we achieve the same result as in defining $C_{(x,t)}$ by (2.29). With the cost (2.30), we can define the constraint $C_{(x,t)}$ more simply as

$$C_{(x,t)}(\omega) = \{(y,s) : (x,t) \leq (y,s)\}$$

for all (x,t) and ω.

The set of all decision functions D for the interception problem correspond to the admissable vector fields $V(x,t,\omega)$ described in section 2.2. The decision function ϕ corresponding to a given V is defined by

$$(2.31) \qquad \phi(x,t,\omega) = (x + V(x,t,\omega)\cdot\delta,\ t+\delta)$$

if $(\xi(\omega)),\theta(\omega)) \not\leq (x,t)$, and by

$$\phi(x,t,\omega) = (x,t)$$

if $(\xi(\omega),\theta(\omega)) \leq (x,t)$.

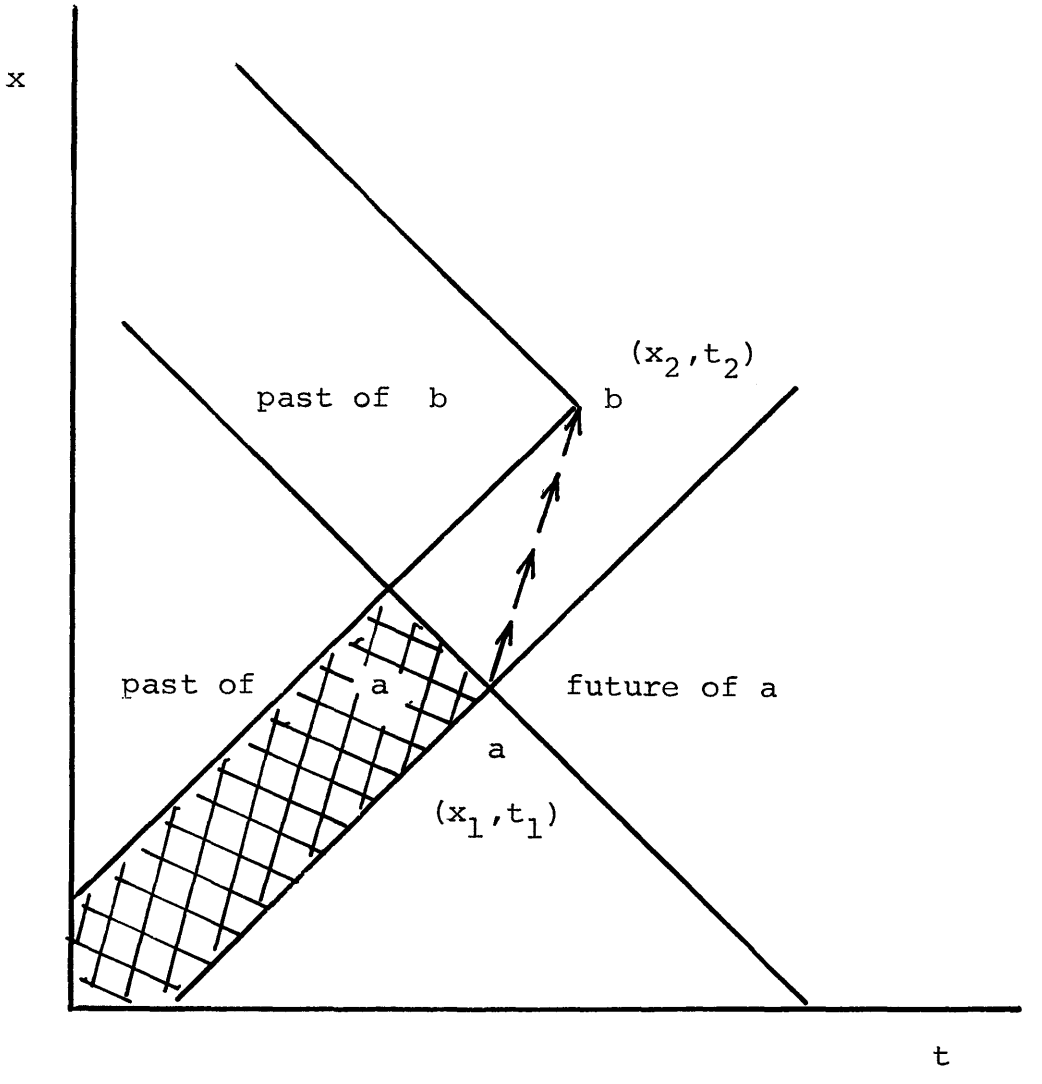
The first relation in (2.31) comes from (2.2.3) in section 2.2. The properties (2.17), (2.14) that ϕ must satisfy are together equivalent to the norm constraint on V in (2.2.4) -- that is, the magnitude of the velocity is less than or equal to the speed of signal propagation. Likewise, the measurability restriction (2.18) on ϕ is equivalent to the causality constraint on V discussed in section 2.2.

Before proceeding to the next example let us discuss

further the norm constraint on the velocity V, or more generally, the condition that the decision functions are increasing -- that is, $t \leq \phi(t)$ a.s. for any decision function ϕ and for any p.o. time t. This condition is implied by the original assumption (2.14) on the constraint sets. This increasing property of the decision functions is necessary to preserve the nested structure of the information as we now explain. If some decision function were not increasing, it might be possible to communicate information by means of this decision function from a time a to a time b where $a \not\leq b$ and where the information σ-field $\mathcal{F}_a$ is not contained in $\mathcal{F}_b$. Consider the space-time interception problem as a specific example. Suppose $a = (x_1,t_1)$ and $b = (x_2,t_2)$ and $a \not\leq b$ as illustrated in figure 2.1. Since $t_1 < t_2$, this means that $|x_1-x_2| > c|t_1-t_2|$ where c is the maximum signal speed. A decision function ϕ which yields $b = \phi(a)$ implies a moving interceptor that can travel faster than the signal speed c: that is, $V(x_1,t_1) > c$. By travelling faster than the signal speed, the interceptor can signal additional information to the space-time point b which is not contained in $\mathcal{F}_b$ as indicated in figure 2.1. If we tried to overcome this difficulty by defining a new information set

$\mathcal{F}'_b$ which included the information signalled to b by the interceptor, we would find that $\mathcal{F}'_b$ now depends on the trajectory of the interceptor up to the space-time point b as well as on the point b. For example, in the case of one space dimension, the extra information at b would depend on whether the "faster-than-light" interceptor came from the direction left or right of x_2 (compare figures 2.1 and 2.2). The dependence of the information σ-fields on the trajectory of the decision function destroys the partially nested structure and creates a much more difficult problem.

To summarize the above remarks, note that the requirement that the decision functions are monotonically increasing functions in the p.o. time set ensures that the information σ-fields $\{\mathcal{F}_t : t \in T\}$ (which depend only on points in T) include any possible information signalled by a decision function. Sandell and Athans (1975) have discussed the importance of considering the "real" information (messages) and the information conveyed by decisions or controls (protocol) together as one in the control of a communication network. Our signal interception problem is a special example of the class of problems that Sandell and Athans call relativistic stochastic control problems.



If $a \not\leq b$ and if an interceptor can travel from a to b, then b can obtain information not in its own past (shaded region) from the interceptor.

Figure 2.1

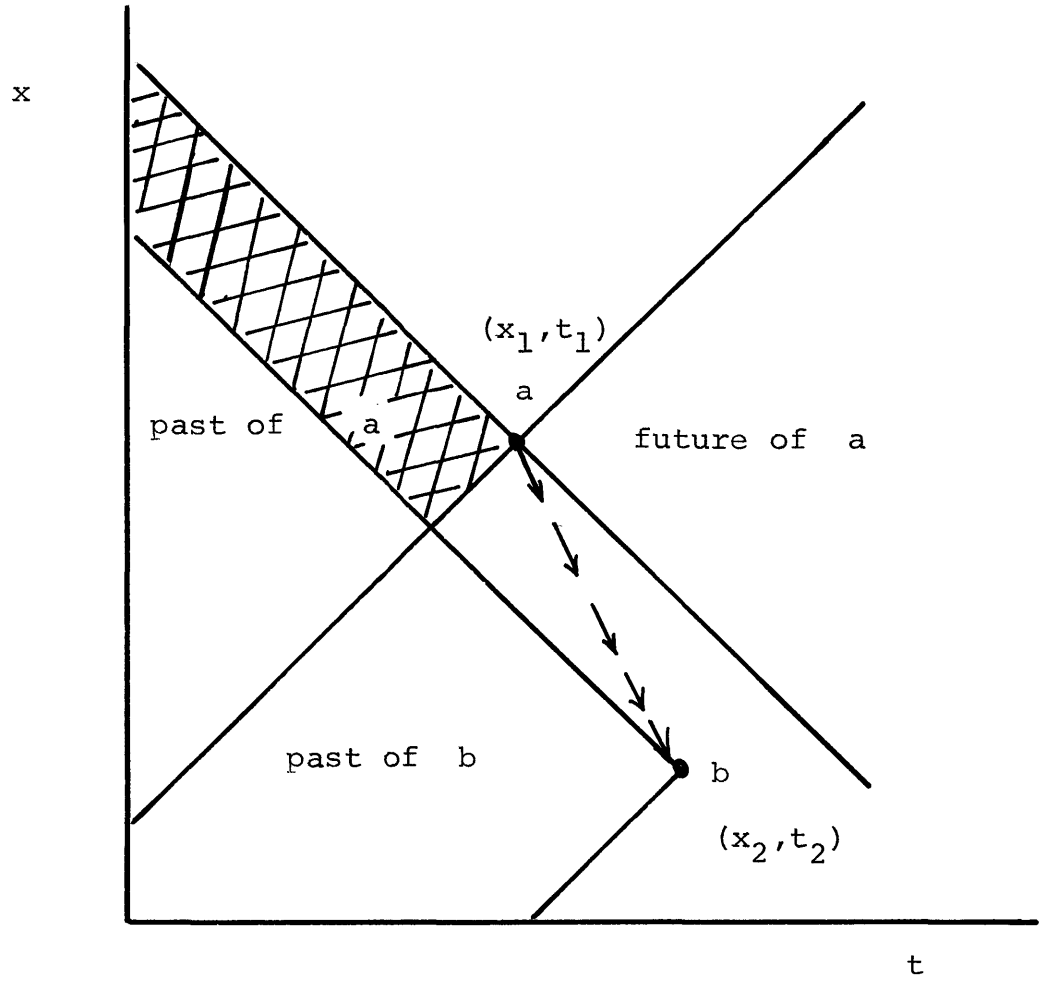


Figure 2.2

To conclude this section we discuss a quite different problem (optimal oil exploration) which fits naturally into the p.o. time framework of our stopping problem with partially nested information.

Recently much work has been done in applying probabilistic methods to the problem of oil and mineral exploration. Harbaugh (1977) surveys various techniques in practice and theory today and gives an extensive bibliography of recent work.

The basic problem of oil exploration is to plan the exploration to maximize the expected amount of oil discovered -- or more appropriately, as Harbaugh (1977) discusses, to maximize some expected utility function. The utility can take account not only of such things as cost of exploration and profit of success, but also more subtle factors such as the cost of risking capital in the exploration venture. We may formulate the general problem as a stopping problem on the p.o. set of subsets of a region of a plane. The terminal cost will be the above-mentioned utility; and we seek to compute a decision function that essentially tells us where to drill next.

To formulate the problem as a stopping problem on a p.o. set, let us assume that the plane region is

discretized into a finite set of points X. For each subset $S \subseteq X$ associate the cost

$$c(S,\omega) = \sum_{x \in S} c(x,\omega) \tag{2.32}$$

where $c(x,\omega)$ is the cost (negative utility) of drilling at location x when situation ω holds (ω is the probability space variable). The problem is to choose an optimal S in X to minimize the expected cost $c(S,\omega)$. But since the problem as we have stated it is probabilistic, we must further state what information is available for choosing the subset S.

Before any exploration is done, there may possibly be some prior information about the distribution of oil in X. The σ-field $\mathcal{F}_{\emptyset}$ denotes this information where $\emptyset$ denotes the empty subset of X. Let $\mathcal{F}_S$ represent the information available after exploration of the locations in S, together with the prior information. Today this information is represented typically by a computer-generated contour map of expected oil finds (or expected utility) in a region under exploration. See Harbaugh (1977) for a lengthy discussion of this "automatic surface contouring." Mathematically, given as input the information from exploring locations in S, the algorithm

generates the conditional expectation $E(c(x)|\mathcal{F}_S)$ in the form of contour lines (lines of constant expected utility or cost). Thus, summing over any new region A one can obtain the conditional expectation of the cost of drilling at locations in A:

$$(2.33) \qquad E(c(A)|\mathcal{F}_S) = \sum_{x\in A} E(c(x)|\mathcal{F}_S).$$

In the next section we will find that these conditional expectations are sufficient to compute an optimal decision function ϕ which subsets of X into subsets of X and gives the optimal next region to drill, or the order to stop drilling, given information from previous drillings.

The partially ordered time set T for the stopping problem is a collection of subregions of X partially ordered by set inclusion. By choosing T advantageously, one can introduce additional restrictions into the problem. For example, one can include only connected subregions in T, thus requiring all new drilling to be contiguous to previously exploration regions. Likewise, restricting the constraints $\{C_t : t \in T\}$ can also add new features to the problem. For example, the condition that $\phi(S,\omega) = S$ if $c(S,\omega) \geq L$ for ϕ in D implies that one stop drilling if the cost of exploration at any

stage ever exceeds a preset limit L. One can model this restriction by letting the constraint $C_S(\omega)$ be $\{S\}$ when $C(S,\omega) \geq L$.

In terms of the stopping problem with partially nested information the exploration problem is to find an optimal decision function ψ_* in D such that

$$(2.34) \qquad E(c(\psi_*^\infty(\emptyset)) \mid F_\emptyset) \leq E(c(\psi^\infty(\emptyset)) \mid F_\emptyset)$$

for all ψ in D. Then one successively explores the increasing region $\psi_*^1(\emptyset,\omega)$, $\psi_*^2(\emptyset,\omega)$, ... until the final region $\psi_*^\infty(\emptyset,\omega)$ is reached.

To conclude the example and the section, we remark that certain control problems can be formulated as optimal stopping problems with partially nested information. As an example, suppose that in the oil exploration problem the value of c at x in X depends not only on x but also on the amount of effort $u(x)$ spent drilling at x. For definiteness assume that $u(x)$ takes values in the finite set E. Thus, (2.32) is replaced by

$$(2.35) \qquad c(S,u_S,\omega) = \sum_{x \in S} c(x,u(x),\omega)$$

where u_S denotes a function from S into E. The problem now is to choose a region S to explore and the effort u_S to spend exploring S. We may formulate the problem in terms of the optimal stopping problem by defining a new partial order as follows. If A and B are subregions of X, and u_A and v_B are functions from A and B respectively into E, then define $(A,u_A) \leq (B,v_B)$ to mean that

(2.35) $$A \subset B,$$

$$u_A(x) = v_B(x) \quad \text{for all} \quad x \quad \text{in} \quad A.$$

We choose for the partially ordered time set T a collection of pairs (S,u_S) with the partial order $\leq$ defined by (2.35). As a simple first choice for the σ-fields $\mathcal{F}_{(S,u_S)}$, we assume that $\mathcal{F}_{(S,u_S)} = \mathcal{F}_S$ so that the information depends only on the region explored and not on the effort of exploration. In this case it is easy to check that the σ-fields $\mathcal{F}_{(S,u_S)}$ are partially nested with respect to $\leq$.

With this basic structure one can formulate the new control problem as a stopping problem. If ψ_* is an optimal decision function for the new problem, then

$\psi_*^{\infty}((\emptyset,u_{\emptyset}),\omega)$ gives the subregion in X and also the effort to spend exploring that subregion in order to minimize the expected cost defined by (2.35).

3.3 The Optimal Stopping Problem and Dynamic Programming for Partially Ordered Time Sets

In this section we solve the optimal stopping problem with partially nested information by means of a dynamic programming equation, the generalization of (2.3.6) in section 2.3. We discover that one ϕ_* satisfies (2.19) for all initial times θ in T and we show how to characterize ϕ_* in terms of the solution of the dynamic program. Next we prove that the optimal cost for the dynamic program has a unique solution although the optimal decision function ϕ_* is not unique. Finally, in order to illustrate the abstract results, we present the interception problem and show how the equation (2.3.6) corresponds to the more general dynamic program we derive here.

Assume the same restrictions for the problem as mentioned in the beginning of section 3.2. In particular, $(T,\leq)$ is discrete bounded. Define $\pi_*(t)$ as

$$(3.1)\quad \pi_*(t) = \text{ess inf}\,\{E(c(\phi^{\infty}(t))\,|\,F_t) : \phi \in D\}$$

where essinf is the <u>essential infimum</u>. Thus, $\pi_*(t)$ is an F_t-measurable random variable such that

$$(3.2)\qquad \pi_*(t) \leq E(c(\phi^{\infty}(t))\,|\,F_t), \qquad \phi \in D,$$

and such that for every other random variable $\pi_*(t)'$ satisfying (3.2) we also have $\pi_*(t) \geq \pi_*(t)'$ almost surely. Snell (1952) notes that the essential supremum of a collection of random variables always exists, and of course, the same is true for the essential infimum.

<u>Theorem 3.1</u> (dynamic programming)

For each t and s such that $t < s$, define $\rho(s,t)$ so that

$$(3.3)\qquad \rho(s,t) = E(\pi_*(s)\,|\,F_t).$$

Let $\rho(t,t)$ be defined as

$$(3.4)\qquad \rho(t,t) = c(t)$$

for all t in T.

If π_* is the function defined in (3.1), then for each t in T, it satisfies the dynamic programming equation

$$(3.5) \qquad \pi_*(t) = \min \{\rho(s,t) : s \in C_t\}$$

almost surely. Note that $\pi_*(t)$, $\rho(s,t)$ and C_t all may depend on ω.

Furthermore, there is an optimal decision function ϕ_* in D such that for each t in T one has

$$(3.6) \qquad \pi_*(t) = E(c(\phi_*^{\infty}(t)) \mid F_t).$$

Proof

We first construct ϕ_* and then prove (3.5) and (3.6) by means of induction on the discrete bounded set T.

Let $\rho(t)$ be defined by

$\rho(t,\omega) = \min \{\rho(s,t,\omega) : s \in C_t(\omega)\}$.

Let us show $\rho(t)$ is F_t-measurable. Note that each $\rho(s,t)$ in (3.3) is F_t-measurable.

We may write $\rho(t)$ as

$$(3.7) \qquad \rho(t) = \sum_C \min \{\rho(s,t) : s \in C\} \cdot 1_{C_t=C}$$

where C ranges over the subsets of T. The assumption that T is discrete bounded and the restriction (2.14) on C_t imply that the range of C_t is finite and hence the sum in (3.7) is finite. Moreover, (2.16) implies that $1_{C_t=C}$ is $\mathcal{F}_t$-measurable. Since we need only consider finite subsets C in the sum, the minimum $\min\{\rho(s,t) : s \in C\}$ is also $\mathcal{F}_t$-measurable. It follows from (3.7) that $\rho(t)$ is $\mathcal{F}_t$-measurable.

It is possible to choose disjoint sets $A(s,t,C)$ in $\mathcal{F}_t$ for each t in T, $C \subseteq \{r : t \leq r\}$ and s in C such that

$$\bigcup\{A(s,t,C) : s,\ C,\ s \in C\} = \Omega \tag{3.8}$$

$$\rho(t,\omega) = \rho(s,t,\omega) \quad \text{if} \quad C_t(\omega) = C,\ s \in C \tag{3.9}$$

and $\omega \in A(s,t,C)$.

For example, one way to do this is to enumerate C as $\{s_1,s_2,\ldots,s_n\}$ and define $A(s_i,t,C)$ for $1 \leq i \leq n$ as

$$A(s_i,t,C) = \{\omega : \rho(t) = \rho(s_i,t), \rho(t) < \rho(s_j,t),\ 1 \leq j < i;\ C_t(\omega) = C\}.$$

Define the function $\phi_* : T \times \Omega \to T$ so that for all t

in T, and all finite subsets C in the range $C_t(\Omega)$,

$$\phi_*(t,\omega) = s \tag{3.10}$$

when ω lies in $A(s,t,C)$.
It is not hard to see that ϕ_* satisfies (2.17) and (2.18), and hence, ϕ_* lies in D.

Suppose that t is a terminal time of T. Then from (2.14) and (2.15) it follows that $C_t(\omega) = \{t\}$ for all ω. Thus, the right hand side of (3.5) is $\rho(t,t) = c(t)$. Likewise, (2.17) implies that $\phi(t,\omega) = t$ for all ϕ in D. From the definition of ϕ^∞ it is clear that $\phi^\infty(t,\omega) = t$ also. Consequently, (3.1) gives $\pi_*(t) = E(C(t)|F_t) = c(t)$ since we assume $c(t)$ is F_t-measurable. Thus, (3.5) is true for all terminal times in T. Since $\phi^\infty(t,\omega) = t$ for all ϕ in D when t is a terminal time and since ϕ_* defined by (3.10) is in D, it follows from (3.1) that ϕ_* satisfies (3.6) for all terminal times t.

Having shown that (3.5) is true and that ϕ_* satisfies (3.6) for all terminal times, we now prove the inductive step. Suppose that (3.5) is true for all t such that $r < t$ and suppose that ϕ_* defined

in (3.10) satisfies (3.6) for all t such that $r < t$. We show that (3.5) is true and ϕ_* satisfies (3.6) at $t = r$.

For $r < t$, the inductive hypothesis for (3.6) implies that

$$\pi_*(t) = E(C(\phi_*^\infty(t))|F_t). \tag{3.11}$$

From the definition (3.3) and from (3.11) we have for $r < t$ that

$$\rho(t,r) = E(E(c(\phi_*^\infty(t))|F_t)|F_r).$$

Since $r < t$ implies $F_r \quad F_t$, we see that

$$\rho(t,r) = E(c(\phi_*^\infty(t))|F_r) \tag{3.12}$$

for $r < t$.

The relation (3.12) allows us to show that

$$\rho(r) = E(c(\phi_*^\infty(r))|F_r) \tag{3.13}$$

as follows. From the definition (3.10) of ϕ_* it

follows that $\phi_*(r) = t$ for ω in $A(t;r,C)$ for all t in C and all C in $C_r(\Omega)$. Consequently, $\phi_*^\infty(r) = \phi_*^\infty(\phi_*(r)) = \phi_*^\infty(t)$ for ω in $A(t,r,C)$. Thus, we have

$$(3.14) \qquad c(\phi_*^\infty(r)) = \sum_C \sum_{t\in C} c(\phi_*^\infty(t)) 1_{A(t,r,C)},$$

where C ranges in $C_r(\Omega)$.

If r is in C, then $\phi_*^\infty(r) = r$ for ω in $A(r,r,C)$ and thus,

$$\begin{aligned} c(\phi_*^\infty(r) 1_{A(r,r,C)} &= c(r)\cdot 1_{A(r,r,C)} \\ &= \rho(r,r)\cdot 1_{A(r,r,C)}. \end{aligned}$$

In this case we have

$$(3.15) \qquad E(c(\phi_*^\infty(r)) 1_{A(r,r,C)} | F_r) = \rho(r,r) 1_{A(r,r,C)}.$$

Since $A(t,r,C)$ is F_r-measurable for each t and C, it is clear from (3.12) that

$$(3.16) \qquad E(c(\phi_*^\infty(t)) 1_{A(r,r,C)} | F_r) = \rho(t,r) 1_{A(t,r,C)}$$

for t in C such that $t \neq r$.

Taling the conditional expectation of (3.14) with respect to $\mathcal{F}_r$ and applying (3.15), (3.16) yields

$$(3.17)\quad E(c(\phi_*^\infty(r))\,|\,\mathcal{F}_r) = \sum_C \sum_{t\in C} \rho(t,r)\,1_{A(t,r,C)}.$$

The construction in (3.9) implies that $\rho(t,r) = \rho(r)$ for ω in $A(t,r,C)$, and hence (3.17) becomes

$$(3.18)\quad E(c(\phi_*^\infty(r))\,|\,\mathcal{F}_r) = \rho(r) \sum_C \sum_{t\in C} 1_{A(t,r,C)}.$$

Since $A(t,r,C)$ are disjoint and satisfy (3.8), the right side of (3.18) is $\rho(r)$ and we have the desired result (3.13).

Note that (3.11) asserts that the right side of (3.5) and the right side of (3.6) are equal for $t = r$ with ϕ_* as in (3.10). If we can demonstrate that (3.5) is true for $t = r$, then (3.6) will follow and the theorem will be proved. Thus, we now show $\pi_*(r) = \rho(r)$.

Recalling the definition of $\pi_*(r)$ from (3.1) and the expression for $\rho(r)$ in (3.13), we see that $\pi_*(r) \leq \rho(r)$. Thus, to show $\pi_*(r) = \rho(r)$ it suffices to prove $\pi_*(r) \geq \rho(r)$.

Let ϕ be any member of D and let $B(t,r,C) = \{\omega : \phi(r,\omega) = t,\ C_t(\omega) = C\}$ for all t in C and C in $C_r(\Omega)$. The $B(t,r,C)$ are disjoint sets in F_r and

$$\{B(t,r,C) : t,C,t \in C\} = \Omega \tag{3.19}$$

Clearly, the $B(t,r,C)$ play the same role for ϕ that he $A(t,r,C)$ play for ϕ_*. Thus, we can show that

$$c(\psi^\infty(r)) = \sum_C \sum_{t\in C} c(\phi^\infty(t)) 1_{B(t,r,C)} \tag{3.20}$$

in the same way that we showed (3.14).

The definition (3.1) implies that

$$E(c(\phi^\infty(t)) \mid F_t) \geq \pi_*(t) \tag{3.21}$$

for all t.

Conditioning (3.21) with respect to F_r and using the definition (3.3) gives

$$E(c(\psi^\infty(t)) \mid F_r) \geq \rho(t,r) \tag{3.22}$$

for t such that $r < t$. Since the sets $B(t,r,C)$ are $\mathcal{F}_r$-measurable, we may multiply (3.22) by the indicator function $1_{B(t,r,C)}$ to obtain

$$(3.23) \qquad E(c(\phi^{\infty}(t))1_{B(t,r,C)}|\mathcal{F}_r) \geq \rho(t,r)1_{B(t,r,C)},$$

if $t \in C$ and $t \neq r$. If $r \in C$, then by definition $\phi(r) = r$ on $B(r,r,C)$ and so $\phi^{\infty}(r) = r$ on $B(r,r,C)$. Thus, it follows that $c(\phi^{\infty}(r))1_{B(r,r,C)} = c(r)\cdot 1_{B(r,r,C)}$. Definition (3.4) gives $c(r) = \rho(r,r)$ and hence,

$$(3.24) \quad E(c(\phi^{\infty}(r))1_{B(r,r,C)}|\mathcal{F}_r) = \rho(r,r)\cdot 1_{B(r,r,C)}.$$

Thus, (3.23) is true for all t in C.

By definition of $\rho(r)$ it is clear that $\rho(t,r) \geq \rho(r)$ for all t in $C_r(\omega)$. Thus, $\rho(t,r) \geq \rho(r)$ for all t in C if ω lies in $B(t,r,C)$. Thus, (3.23) and (3.24) imply that

$$(3.25) \quad E(c(\phi^{\infty}(t))1_{B(t,r,C)}|\mathcal{F}_r) \geq \rho(r)1_{B(t,r,C)}$$

for all t in C. Adding (3.25) over all t and C such that $t \in C$, and using (3.19) and (3.20) gives the result,

(3.26) $\quad E(c(\phi^{\infty}(r))\,|\,F_r) \geq \rho(r).$

In (3.26) the decision function ϕ is arbitrary. Taking the essential infimum of (3.26) over ϕ in D gives the desired result $\pi_*(r) \geq \rho(r)$.

This completes the proof of the inductive step and the proof is now complete. ///

The following corollary follows easily from the theorem, and it completely characterizes the optimal decision functions ϕ_* which satisfy (2.39).

Corollary 3.2 (optimal decision functions)

If ϕ_* is a member of D such that for each t in T and ω in Ω we have

(3.27) $\quad \rho(\phi_*(t,\omega),t) = \min\{\rho(s,t,\omega) : s \in C_t(\omega)\}$

where ρ is defined as in theorem 3.1 then ϕ_* satisfies (3.6) of theorem 3.1 for all t in T.

Conversely, if ϕ_* is an optimal decision function satisfying (3.6) for all t, then ϕ_* must satisfy (3.27) for all t.

Proof

The first part of the corollary follows from the proof of theorem 3.1. The only properties of ϕ_* we used in the proof of the theorem were that ϕ_* belonged to D and that

$$\rho(s,r) = \min \{\rho(t,r) : t \in C\}$$

if $\phi_*(r,\omega) = s$ and $C_r(\omega) = C$. This is exactly the statement (3.27).

We prove the converse as follows. If $t < s$, then by definition (3.3) of $\rho(s,t)$ we have

$$\rho(s,t) = E(\pi_*(s) | F_t). \tag{3.28}$$

Since (3.6) is assumed true, conditioning with respect to F_t gives

$$\rho(s,t) = E(c(\phi_*^{\infty}(s)) | F_t). \tag{3.29}$$

On the other hand, if $\phi_*(t) = t$, then $\phi_*^{\infty}(t) = t$ and by definition (3.4) of $\rho(t,t)$ we have

(3.30) $\rho(t,t) = E(c(\phi_*^\infty(t))|F_t)$ for $\omega \in \{\phi_*(t) = t\}$.

Noting that the events $\{\phi_*(t) = t\}$ and $\{\phi_*(t) \neq t\} = \{t < \phi_*(t)\}$ are F_t-measurable, we put (3.29) and (3.30) together to obtain

(3.31) $\rho(\phi_*(t),t) = E(c(\phi_*^\infty(\phi_*(t)))|F_t)$.

It is not hard to see that $\phi_*^\infty(\phi_*(t)) = \phi_*^\infty(t)$, so that (3.31) becomes

(3.32) $\rho(\phi_*(t),t) = E(c(\phi_*^\infty(t))|F_t)$.

Applying (3.6) to the right side of (3.32) and using the dynamic programming equation (3.5) gives the result (3.27). ///

The next theoretical result of this section is a uniqueness theorem for the dynamic programming equation.

Theorem 3.3 (uniqueness for dynamic programming)

Suppose that π_1 and π_2 are F-measurable random extended real-valued functions on T, and suppose that for $i = 1,2$ and each t in T we have that

(3.33) $\rho_i(s,t) = E(\pi_i(s) \mid F_t)$

for all s such that $t < s$,

(3.34) $\rho_i(t,t) = c(t)$

(3.35) $\pi_i(t) = \min \{\rho_i(s,t) : s \in C_t\}$

Then for each t in T, $\pi_1(t) = \pi_2(t)$ almost surely.

Proof

We use an induction argument as we did in proving theorem 3.1. Suppose t is a terminal time of T, then $C_t(\omega) = \{t\}$ for all ω, and (3.35) and (3.34) imply $\pi_i(t) = c(t)$ for $i = 1,2$. Thus, $\pi_1(t) = \pi_2(t)$ for all terminal times t.

Now suppose $\pi_1(s) = \pi_2(s)$ for all s such that $t < s$. Then from (3.33) it is clear that $\rho_1(s,t) = \rho_2(s,t)$ for all s such that $t < s$. Since (3.34) implies $\rho_1(t,t) = c(t) = \rho_2(t,t)$, we see that $\rho_1(s,t) = \rho_2(s,t)$ for all s such that $t \leq s$. In particular, for all s in $C_t(\omega)$ and hence, (3.32) implies $\pi_1(t) = \pi_2(t)$. This completes the proof. ///

Cost-to-go formulation

Conventional dynamic programs are often formulated in terms of the so-called cost-to-go. In our formulation the cost-to-go using decision function ϕ and starting at time t would be

$$c(\phi^{\infty}(t)) - c(t). \tag{3.36}$$

We may formulate the dynamic programming solution of the optimal stopping time in terms of the optimal cost-to-go, $\lambda_*(t)$, starting from t, which is defined as

$$\lambda_*(t) = \inf \{E(c(\phi^{\infty}(t)) - c(t) \mid F_t) : \phi \in D\}. \tag{3.37}$$

We assume that $c(t) < +\infty$ so that the subtraction of $c(t)$ is always unambiguous. Thus, we have

Theorem 3.4 (cost-to-go formulation of dynamic programming)

For each t and s such that $t < s$, define $\mu(s,t)$ so that

$$\mu(s,t) = E(\lambda_*(s) \mid F_t) + E(c(s) - c(t) \mid F_t). \tag{3.38}$$

Let $\mu(t,t)$ be defined as

(3.39) $\mu(t,t) = 0$

for all t in T.

If λ_* is the function defined in (3.37), then for each t in T, it satisfies the dynamic programming equation

(3.40) $\lambda_*(t) = \min \{\mu(s,t) : s \in C_t\}$

almost surely.

Furthermore, there is an optimal decision function ϕ_* in D such that for each t in T one has

(3.41) $\lambda_*(t) = E(c(\phi_*^\infty(t)) - c(t) \mid F_t)$.

It is clear that Corollary 3.2 and Theorem 3.3 have corresponding versions for Theorem 3.4. It should be clear that Theorems 3.1 and 3.4 are connected by the relations

(3.42) $\pi_*(t) - c(t) = \lambda_*(t)$,

(3.43) $\rho(s,t) - c(t) = \mu(s,t)$.

Moreover, an optimal decision function ϕ_* for one problem is an optimal decision function for the other problem also.

In some cases the cost-to-go formulation in Theorem 3.4 may be easier to work with than the original formulation in Theorem 3.1. In particular, this depends on whether one can express $E(c(s) - c(t)|\mathcal{F}_t)$ in a convenient form. For example, in the treasure hunt problem for the cost c defined by (2.32) we have for $A \subset B$ that

$$c(B) - c(A) = \sum_{x \in B-A} c(x)$$

Thus, knowing $E(c(x)|\mathcal{F}_A)$ for each x not in A gives $E(c(B) - c(A)|\mathcal{F}_A)$ for all subregions B containing A.

Example: signal interception problem

To help clarify the abstract results of this section we solve the interception problem using theorem 3.1. First we formulate a slightly more general version of the signal interception problem of chapter 2.

Let $(T, \leq)$ be a discrete bounded time set with a single terminal time t_1. Suppose that $\omega \to \tau(\omega)$ is a

T-valued random function whose statistics are completely determined by the prior distribution $q(t)$ defined as $q(t) = P(\tau \not\leq t)$ for each t in T.

For each t in T define F_t as the smallest σ-field containing $\{\tau = s\}$ for each s such that $s \leq t$. It is not hard to see that F_t is generated by the irreducible disjoint sets (atoms) $\{\tau = s\}$ for $s \leq t$ and $\{\tau \not\leq t\}$.

Let $c(t)$ be a deterministic finite real-valued cost function such that $c(s) \leq c(t)$ for $s \leq t$. Then the interception problem is to find a decision function ϕ shich minimizes $E(c(\phi^{\infty}(t))\,|\,F_t)$ subject to the interception constraint $\tau \leq \phi^{\infty}(t)$.

To formulate this interception problem in terms of a constraint set, let C_t be an F_t-measurable constraint set which satisfies (2.14), (2.15), (2.16) and also has the following two properties:

(3.44) $\qquad t \in C_t(\omega), \qquad \omega \in \{\tau \leq t\},$

(3.45) $\qquad t \notin C_t(\omega), \qquad \omega \notin \{\tau \leq t\}.$

If D is the class of admissable decision functions

corresponding to the constraint sets $\{C_t : t \in T\}$, then (3.44) and (3.45) force $t < \phi(t,\omega)$ when $\tau(\omega) \not\leq t$ for all ϕ in D, and they allow the decision $t = \phi(t,\omega)$ when $\tau(\omega) \leq t$. Thus, the decision to stop is possible if and only if interception has occurred -- if and only if $\tau \leq t$. Note that we require T to have a single terminal time so that the terminal constraint C_{t_1}, which must be $\{t_1\}$ from (2.14) and (2.15), would be consistent with the interception constraint $\tau(\omega) \leq t_1$.

Having formulated the interception problem as an optimal stopping problem with partially nested information, we see from Theorem 3.1 that the solution satisfies the dynamic program (3.5). Let us now find π_* in terms of the prior distribution q.

Since $\pi_*(t)$ is F_t-measurable and since $\{\tau \not\leq t\}$ is an atom of F_t we can write $\pi_*(t)$ as

$$(3.46) \qquad \pi_*(t,\omega) = \pi_*(t,\omega) 1_{\tau(\omega) \leq t} + f(t) 1_{\tau(\omega) \not\leq t}.$$

where $f(t)$ is a deterministic real function of t to be determined. Since $c(t) \leq c(s)$ for all s such that $t \leq s$ it follows that $c(t) \leq c(\phi^\infty(t))$ for all ϕ in D. If $\tau(\omega) \leq t$, then $t \in C_t(\omega)$ and we may

choose ϕ so that $\phi(t,\omega) = t$. Thus, for $\tau(\omega) \leq t$ we have from (3.1) that $\pi_*(t) = c(t)$. Hence, $\pi_*(t)$ has a particularly simple form, namely

$$(3.47) \qquad \pi_*(t) = c(t)1_{\tau \leq t} + f(t)1_{\tau \not\leq t}.$$

Thus, we need to know only $f(t)$ to determine $\pi_*(t)$.

For $\tau \not\leq t$, and $t < s$, the conditional expectation $E(\pi_*(s) \mid \mathcal{F}_t)$ is

$$E(c(s)1_{\tau \leq s} + f(s)1_{\tau \not\leq s} \mid \tau \not\leq t)$$

$$= c(s)P(\tau \leq s \mid \tau \not\leq t) + f(s)P(\tau \not\leq s \mid \tau \not\leq t).$$

The conditional probability $P(\tau \not\leq s \mid \tau \not\leq t)$ is

$$\frac{P(\tau \not\leq s,\ \tau \not\leq t)}{P(\tau \not\leq t)} = \frac{P(\tau \not\leq s)}{P(\tau \not\leq t)} = \frac{q(s)}{q(t)}$$

for $t \leq s$. Thus, we find that $E(\pi_*(s) \mid \mathcal{F}_t)$ or, in our earlier notation, $\rho(s,t)$ is given by

$$(3.48) \qquad \rho(s,t) = c(s) + (f(s) - c(s))\frac{q(s)}{q(t)}.$$

Since $\{\tau \not\leq t\}$ is an atom of $\mathcal{F}_t$, and since C_t

satisfies (2.16), there is at most a single set C in T such that $C_t(\omega) = C$ when $\tau(\omega) \not\leq t$. Let $\overline{C}_t$ denote this set when $P(\{\tau \not\leq t\}) > 0$. Using $\overline{C}_t$ with (3.46) we find that for $\tau(\omega) \not\leq t$ we have

$$(3.49)\quad \min\{\rho(s,t) : s\varepsilon C_t\} = \min\{c(s) + (f(s)-c(s))\frac{q(s)}{q(t)} : s\varepsilon \overline{C}_t\}.$$

Using the dynamic program (3.5) and the expression (3.47) when $\tau(\omega) \not\leq t$ we obtain the following equation for f:

$$(3.50)\quad f(t) = \min\{c(s) + f(s)-c(s))\frac{q(s)}{q(t)} : s \varepsilon \overline{C}_t\}$$

when $q(t) > 0$, and

$$f(t) = c(t)$$

when $q(t) = 0$.

Noting that $t \not\varepsilon \overline{C}_t$ when $q(t) > 0$, we find that (3.50) gives a deterministic dynamic program to calculate $f(t)$ recursively backwards from the terminal time t_1.

In terms of the function $\psi(t)$ defined as $\psi(t) = q(t)(f(t)-c(t))$ the dynamic program (3.50) becomes

(3.51) $\psi(t) = \min \{\psi(s) + q(t)(c(s)-c(t)) : s \in \overline{C}_t\}$

when $q(t) > 0$, and

$\psi(t) = 0$

when $q(t) = 0$.

Equation (3.51) is the exact general version of (2.3.6) in section 2.3. The functions ψ are the same in both cases. In terms of (x,t) space-time coordinates (3.51) should read

(3.52) $\psi(x,t) = \min \{\psi(y,s)+q(x,t)(t-s) : (y,s) \in \overline{C}_{(x,t)}\}$.

Since $\overline{C}_{(x,t)} = \{(y,s) : s = t+\delta, y = x+\delta v, |v| \leq c\}$, one sees that $t - s = \delta$ and (3.52) becomes exactly (2.3.6).

The optimal decision functions for the interception problem are easily obtained from (3.50) or (3.51). If ϕ_* is an optimal decision function then one can show

(3.53) $\phi_*(t,\omega) = t$ if $\tau(\omega) \leq t$

$\phi_*(t,\omega) = \overline{\phi}_*(t)$ if $\tau(\omega) \not\leq t$

where $\overline{\phi}_* : T \to T$ is a deterministic function such that $\overline{\phi}_*(t) \in \overline{C}_t$ and $s = \overline{\phi}_*(t)$ minimizes

$$(3.54) \qquad \psi(s) + q(t)c(s)$$

for s in $\overline{C}_t$ and $q(t) > 0$. In the space-time signal interception problem the expression (3.54) becomes

$$(3.55) \qquad \psi(x+\delta\nu, t+\delta) + q(x,t)(t+\delta)$$

It is clear that minimizing (3.55) for ν such that $|\nu| \leq c$ is the same as minimizing $\psi(x+\delta\nu, t+\delta)$. Thus, we obtain the same optimal decision function as given in section 2.3 in equations (2.3.8) and (2.3.9).

Section 3.4. Conclusions

In this chapter we have defined the class of stopping problems with partially nested information. Using the concept of partially ordered time we were able to generalize the two-parameter signal interception problem of chapter 2. The general stopping problem with partially nested information includes all one-parameter stopping time problems and allows one to formulate many new stopping problems such as the space-time interception problem or the oil exploration problem discussed in section 3.2.

The optimal control aspects of the general optimal stopping problem are significant as we have remarked in section 2.2. It is the control formulation of the stopping problem in terms of decision functions (the controls), rather than the stopping time formulation, that permits a reasonable solution of the general stopping problem.

We solved the stopping problem by means of a partially ordered time dynamic program in section 3.3. Two characteristic of the stopping problem make the dynamic program solution possible: (1) the control formulation of the problem in terms of decision functions as we discussed above; and (2) the partially nested information

structure of the problem. The partially nested information structure, originally defined by Ho and Chu (1972), derives from causality relations in a particular problem. The partially nested structure depends on (1) an antisymmetric and transitive precedence relation between cause (effecting agent) and effect (affected agent) and (2) the affected agent's perfect knowledge or memory of all things affecting it. This information structure may arise in problems as diverse as the space-time problem or the oil exploration problem. As we noted in section 3.2 it is important to include the signaling effect of the control decisions of agents in the total information structure. Thus, for the signal interception problem it is important that the interceptor speed is not greater than the signal speed.

The partially ordered dynamic program we obtain is nonserial in the sense of the nonserial dynamic programming of Bertele and Brioschi (1972), but the p.o. dynamic program possesses much more structure than any of the cases considered by Bertele and Brioschi and hence, we have been able to exploit this structure to advantage in a way not considered by these authors. To be specific the dynamic program of theorem 3.1 is backward recursive with respect to the partial order. Starting from the

terminal times, one solves the dynamic program by working backward with respect to the partial order. To solve the dynamic at a time t in the partially ordered set, one only need to know the solution at times s such that $t < s$. If t_1 and t_2 are incomparable times (neither $t_1 \leq t_2$ nor $t_2 \leq t_1$), then one need not solve the dynamic program first at t_1 and then at t_2, or first at t_2 and then at t_1. That is, one may solve the program at t_1 and t_2 simultaneously if t_1 and t_2 are incomparable. Thus, the partially ordered structure of the time set allows one to solve in parallel at distinct incomparable times. In the terminology of Chan (1976), the partial order relation on the time set is also the precedence relation for computing the optimal cost solution of the dynamic program.

Finally, let us discuss how to generalize the stopping problem for time sets which are not discrete bounded. In particular, we want to consider discrete time sets which are not bounded and time sets like R^n which are not discrete but have a topological structure. In addition to extending the results of this chapter for more general time sets, it may be possible to extend the results for cases where decisions are made continuously rather than sequentially (i.e., the decision trajectories

are continuous trajectories). We will consider the sequential case here and defer discussion of the continuous case until chapter 5.

The extension of the results to bounded (i.e. compact) continuous time sets is a problem of mathematical technique and does not add substantially new content to the problem. Note that we are not intending to let the stage n be continuous but only to let the time set T be continuous (i.e., the decisions where to go in T are made sequentially). Indeed, a topological structure on T is not necessary so much as a measure structure compatible with the partial order on T. Instead of (2.16) and (2.18) we require that the constraint sets and the decisions functions are measurable with respect to the measure structure on T. The cost function must be jointly measurable with respect to T and with respect to the σ-fields $\{\mathcal{F}_t : t \in T\}$. That is, the restriction ot the cost to $\{s : s \leq t\} \times \Omega$ should be jointly measurable when $\{s : s < t\}$ takes the measure structure of T and Ω takes the measure structure of $\mathcal{F}_t$. We may say that C is progressively measurable in analogy to the case of one-parameter time (see Dellacherie (1973)).

Together with the measurability conditions, one requires finally some condition either on the constraint

sets or on the cost function to ensure that the decision to stop occurs after a finite number of stages. Moreover, given a fixed initial point in the time set, the number of decisions (or stages) until stopping must be uniformly bounded from above. For example, stopping is guaranteed if T is a bounded subset of space-time and the constraint sets $C_{(x,t)}$ are such that

$$C_{(x,t)} \subseteq \{(y,s) : t + \delta \leq s\}$$

where δ is a fixed positive number.

With these conditions, the results of section 3.2 are probed much as for the case of discrete bounded time. One minor difference is that the minimum in (3.5) is now an infimum. Thus, the optimal cost will still exist and satisfy the dynamic program equation (3.5), but there may be no optimal decision function unless the infimum in (3.5) is a minimum. Nevertheless, we may find decision functions which obtain expected cost arbitrarily close to the optimum --- these are the so-called ε-minimizing solutions of Snell (1952).

The extension to discrete but unbounded time sets is more difficult than the previous extension to bounded,

nondiscrete time sets but the extension is still manageable. The essential new difficulty is that the sequence $\{\phi^n(t) : n \geq 1\}$ does not have to stop for an unbounded time set unless we further restrict the cost function or the constraint set. In this case, unlike the first case considered above, we want to require that the decision to stop occurs after a finite number of stages, but we do not want to require that this number is uniformly bounded. One reasonable condition that guarantees that the optimal decision to stop occurs after a finite number of stages is to assume that the cost function increases to $+\infty$ as the p.o. time t increases. In this case one still obtains Theorem 3.1 (that the optimal cost satisfies the dynamic program) but there is no terminal set from whicy to solve for the optimal cost by working bakcwards recursively, because the time set is unbounded. In some cases it may be possible to reduce the problem with unbounded time set to an equivalent problem with a bounded time set. For example, this is what was done with the space-time interception problem in chapter 2. Let us note that this difficulty with unbounded time sets is also a difficulty with one-parameter time problems and is not a special problem with multi-parameter dynamic programming.

CHAPTER 4

REACHABLE STOPPING TIMES AND THE OPTIONAL SAMPLING THEOREM

1. Introduction

In this chapter we will tie together the results of chapter 2 and 3 on the stopping problem with partially nested information with the results of chapter 1 on optional sampling. In section 2 we define a special class of stopping times in terms of the decision functions introduced in chapter 3. Specifically, we define a reachable stopping time, and we define the weaker concept that a stopping time τ is reachable from a stopping time σ. It is not hard to show that if τ is a reachable stopping time, then it is reachable from every stopping time σ such that $\sigma \leq \tau$. With this definition of reachable stopping time, we note that all one-parameter stopping times are reachable. Moreover, in the two-parameter case considered in chapter 1 all stopping times are reachable. Also, the control variable of Haggstrom (1966) discussed in chapter 3 is a reachable stopping time. The counter-examples in chapter 1 illustrate stopping times which are not reachable. This fact follows from the optional sampling theorem we prove in section 3. We show that if

τ and σ are stopping times on a finite time set and if τ is reachable from σ, then the optional sampling theorem is true with τ and σ for any submartingale on T. Since the stopping times defined in the counter-examples of chapter 1 do not satisfy the optional sampling theorem for a particular submartingale, then these stopping times cannot be reachable. Note that the optional sampling theorem for reachable stopping times includes the result of section 1.4 as a special case since all the two-parameter stopping times in 1.4 are reachable as we will see. In section 4 we show how the stopping problem with partially nested information can be stated in terms of reachable stopping times although we find this formulation less convenient than the decision function formulation of chapter 3. Finally, in section 5 we apply the dynamic programming results of section 3.3 to prove a converse of the optional sampling theorem. That is, we show that if τ and σ are fixed stopping times on a finite p.o. time set T with $\sigma \leq \tau$, and if the optional sampling theorem is true with τ and σ for all submartingales, then τ must be reachable from σ. Thus, reachable stopping times are the only stopping times which make the optional sampling theorem generally true for submartingales.

2. Reachable Stopping Times

We define reachable stopping times in terms of the decision functions of chapter 3, but we do not require the constraint sets C_t mentioned there. That is, one may assume that the set of constraints for the decision functions in this chapter are the weakest possible constraints, namely those defined by

$$(2.1) \qquad C_t(\omega) = \{s : t \leq s\} \qquad \omega \in \Omega.$$

Thus, ϕ is a decision function if ϕ maps $T \times \Omega$ into T and if both

$$(2.2) \qquad t \leq \phi(t,\omega) \qquad t \in T, \quad \omega \in \Omega$$

and

$$(2.3) \qquad \{\omega : \phi(t,\omega) \leq s\} \in \mathcal{F}_t \qquad t, s \in T.$$

As before, we will often suppress the ω and simply write $\phi(t)$ for the function $\omega \to \phi(t,\omega)$ mapping Ω into T. Having thus reviewed decision functions, we now define reachable stopping times. Let us assume that the p.o. time set $(T,\leq)$, the σ-fields $\{\mathcal{F}_t : t \in T\}$ and the probability space are fixed.

Definition 1.

Suppose τ and σ are two stopping times. We say τ is reachable from σ if there exists a decision function ϕ such that

$$\tau(\omega) = \phi^{\infty}(\sigma(\omega),\omega) \qquad \forall\ \omega \in \Omega. \tag{2.4}$$

In particular, note that $\sigma \leq \tau$.

Definition 2.

A stopping time τ is said to be reachable if there exists a decision function ϕ such that for all t

$$\tau(\omega) = \phi^{\infty}(t,\omega) \quad \forall\ \omega \in \{t \leq \tau\} \tag{2.5}$$

It is not hard to see that if τ is reachable and if $\sigma \leq \tau$, then τ is reachable from σ. For (2.5) implies that $\tau(\omega) = \phi^{\infty}(s,\omega)$ for $\omega \in \{s = \sigma\}$ whence follows (2.4).

All one-parameter stopping times are reachable in the sense of definition 2. The decision function ϕ corresponding to a one-parameter time τ was defined previously in (3.2.27) of section 3.2 as

$$(2.6)\qquad \phi(t,\omega) = t \quad \text{if} \quad \tau(\omega) \leq t,$$
$$= t+1 \quad \text{if} \quad \tau(\omega) \not\leq t.$$

We also find that the two-parameter stopping times defined in chapter 1 are reachable. We show this in the following theorem.

Theorem 2.1 (two-parameter stopping times are reachable)

Suppose that $\{F_t : t \in Z_+^2\}$ is an increasing family of σ-fields which satisfy the conditional independence hypothesis (defined in section 1.2). Let τ be a stopping time taking finitely many values in Z_+^2. Then τ is reachable.

Proof: The proof follows simply from the lemma of section 1.4. Define the random function $\phi : Z_+^2 \to Z_+^2$ as follows in terms of the sets A_t^1 and A_t^2 of the lemma:

$$\phi(t,\omega) = (t_1+1,t_2) \qquad \forall\, \omega \in A_t^1 - \{\tau=t\}$$
$$(2.7)\qquad \phi(t,\omega) = (t_1,t_2+1) \qquad \forall\, \omega \in A_t^2 - \{\tau=t\}$$
$$\phi(t,\omega) = (t_1, t_2) \qquad \forall\, \omega \in \{\tau=t\}$$

where $t = (t_1,t_2)$. The sets A_t^1 and A_t^2 were

constructed so that $A_t^1 \cap A_t^2 = \phi$, $A_t^1 \cup A_t^2 = \Omega$ and $A_t^1, A_t^2 \in \mathcal{F}_t$. Thus, (2.7) defines a decision function on Z_+^2. In addition, the sets A_t^1 and A_t^2 are such that

$$A_t^1 \cap \{\omega : t < \tau(\omega)\} \subset \{\omega : (t_1+1,t_2) \leq \tau(\omega)\},$$
$$A_t^2 \cap \{\omega : t < \tau(\omega)\} \subset \{\omega : (t_1,t_2+1) \leq \tau(\omega)\}. \tag{2.8}$$

From this, it follows that for ω such that $t \leq \tau(\omega)$ we must have $\phi^n(t,\omega) \leq \tau(\omega)$ for each $n \geq 0$. Moreover, we have equality $\phi^n(t,\omega) = \tau(\omega)$ if and only if $\phi^n(t,\omega) = \phi^{n+1}(t,\omega)$. Since τ takes only finitely many values in Z_+^2, we must have $\phi^n(t) = \phi^{n+1}(t)$ for some finite n if $t \leq \tau$. Hence, $\phi^\infty(t,\omega)$ exists and is equal to $\tau(\omega)$ when $t \leq \tau(\omega)$. ///

The existence of sets A_t^1 and A_t^2 with the property (2.8) is equivalent to the stopping time τ being reachable. This fact extends to more general cases and offers some added insight into the definition of reachable stopping times.

Suppose that T is a finite set. For t in T, an immediate successor of t is an element $s \in T$ such

that $t < s$ and such that there is no $r \in T$ such that $t < r < s$. We denote this relationship by $t <\cdot s$. Suppose that τ is a stopping time on T, and suppose for each $t \in T$ and each s such that $t <\cdot s$ there is a set $A_t^s \in F_t$ such that

$$(2.9) \qquad A_t^s \cap \{\omega : t < \tau(\omega)\} \subset \{\omega : s \leq \tau(\omega)\}.$$

In addition suppose that $A_t^s \cap A_t^{s'} = \emptyset$ if $t <\cdot s$, $t <\cdot s'$ and $s \neq s'$, and suppose that $\bigcup_{t<\cdot s} A_t^s = \Omega$. Then as in Theorem 4.1 we can show τ is reachable. The decision function ϕ is defined as in (2.7) so that:

$$(2.10) \qquad \begin{aligned} &\phi(t,\omega) = s \quad \forall\ \omega \in A_t^s - \{\tau=t\}, \quad s, t <\cdot s \\ &\phi(t,\omega) = t \quad \forall\ \omega \in \{\tau=t\}. \end{aligned}$$

Note that this result implies that Haggstrom's (1966) control variables are reachable stopping times. As we discussed in chapter 3, a control variable τ is a stopping time on the tree time set of finite sequences $t = t_1 t_2 \ldots t_n$ such that

(2.11) $$\{\tau \geq t\, t_{n+1}\} \in \mathcal{F}_t$$

for all sequences t in the tree and all t_{n+1}. Noting that the sequences $t\, t_{n+1}$ are precisely the immediate successors of t (see figure 2.1), we see that the sets $A_t^{t\, t_{n+1}}$ can be defined in terms of $\{\tau \geq t\, t_{n+1}\}$. In Haggstrom's problem it is clear that the reachability of τ means precisely that if one knows τ stops beyond the $n^{\underline{th}}$ node (so that $t_1 t_2 t_3 \ldots t_n \leq \tau$), then one can decide which branch τ takes from t_n on the basis of $\mathcal{F}_{t_n}$ alone. Thus, if $\omega \in A_t^{t\, t_{n+1}}$, one chooses the branch to t_{n+1}.

To see an example of a stopping time which is not reachable, consider the stopping time τ in examples 1 and 2 of section 1.3. In these examples $\mathcal{F}_{(0,0)}$ was the trivial σ-field and contained no information about which branch the stopping time τ takes from $(0,0)$. That is, the sets $A^1_{(0,0)}$ and $A^2_{(0,0)}$ are not in $\mathcal{F}_{(0,0)}$. In the next section we show that if a stopping time τ is reachable from σ, then the optional sampling theorem with τ and σ is true for all submartingales. This result then shows that in examples 1 and 2 of section 1.3 and in the example of section 1.4, the stopping time τ is not reachable from σ (and of course, τ is not reachable).

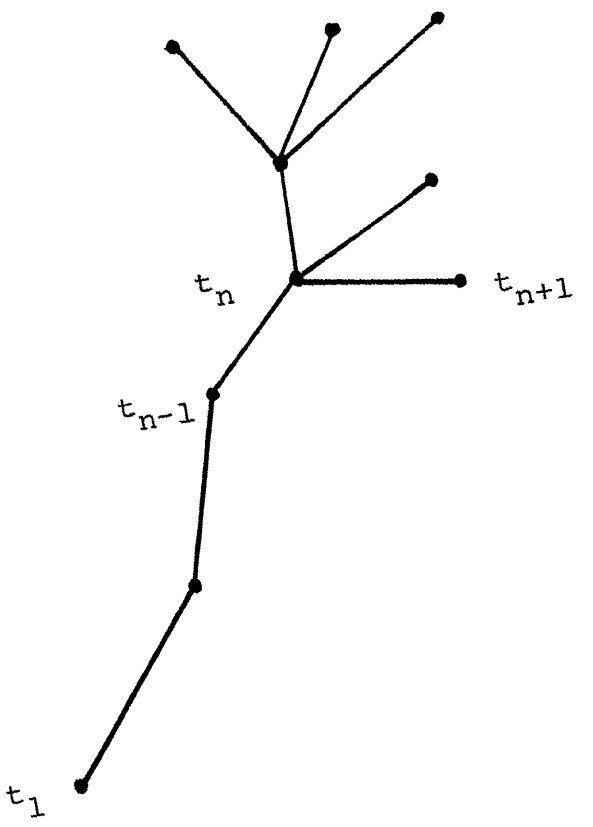


Figure 2.1

3. Optional Sampling for Reachable Stopping Times

The optional sampling result of section 1.3 is contained in the more general result which we now prove. We limit ourselves to p.o. time sets which are finite and afterwards indicate how to extend the results to more general time sets.

Theorem 3.1 (optional sampling for reachable stopping times)

Suppose that $\{F_t : t \in T\}$ is an increasing family of σ-fields indexed by the finite p.o. set $(T,\leq)$. Let τ and σ be stopping times with respect to $\{F_t : t \in T\}$ and suppose that τ is reachable from σ. Then for any submartingale $\{M_t : t \in T\}$ with respect to $\{F_t : t \in T\}$ we have $M_\sigma \leq E(M_\tau | F_\sigma)$.

Proof: Suppose that ϕ is a decision function such that $\tau = \phi^\infty(\sigma)$. Essentially, we show that the T-valued random functions $\phi^n(\sigma)$ are stopping times on T for integers $n \geq 0$. We then show that the one-parameter process $\{M_{\phi^n(\sigma)} : n \geq 0\}$ is a submartingale with respect to the increasing family of σ-fields $\{F_{\phi^n(\sigma)} : n \geq 0\}$ and obtain the conclusion:

Since $\phi^n(\sigma)$ is defined so that

$$\phi^n(\sigma(\omega),\omega) = \phi(\phi^{n-1}(\sigma(\omega),\omega),\omega),$$

it suffices to show that $\phi(\sigma)$ is a stopping time whenever σ is a stopping time in order to show all $\phi^n(\sigma)$ are stopping times. Since ϕ is a decision function, $\sigma \leq \phi(\sigma)$ a.s. for any random function σ. Thus, we have

$$\{\phi(\sigma) \leq t\} = \{\phi(\sigma) \leq t\} \quad \{\sigma \leq t\}$$

for any t, and we may write $\{\phi(\sigma) \leq t\}$ as

$$(3.1) \qquad \{\phi(\sigma) \leq t\} = \bigcup_{s \leq t} \{\phi(s) \leq t\} \cap \{\sigma = s\}.$$

If σ is a stopping time, then $\{\sigma = s\}$ is F_s-measurable and hence, also F_t-measurable. It follows that $\{\phi(\sigma) \leq t\}$ is F_t-measurable for all t and $\phi(\sigma)$ is a stopping time.

Thus, the σ-fields $F_{\phi^n(\sigma)}$ are well-defined since $\{\phi^n(\sigma)\}$ are stopping times. Moreover, since $\phi^n(\sigma) \leq \phi^{n+1}(\sigma)$, it is easy to see that $\{F_{\phi^n(\sigma)} : n \geq 0\}$ is an increasing one-parameter family of σ-fields. To show that $\{M_{\phi^n(\sigma)} : n \geq 0\}$ is a submartingale with respect to $\{F_{\phi^n(\sigma)} : n \geq 0\}$ it suffices to show that

(3.2) $\quad M_\sigma \leq E(M_{\phi(\sigma)} | F_\sigma)$

for an arbitrary stopping time σ. Note that F_σ and the conditional expectation $E(\cdot | F_\sigma)$ are defined so that $E(\cdot | F_\sigma) I_{\sigma=s} = E(\cdot | F_s) I_{\sigma=s}$. This fact allows us to rewrite $E(M_{\phi(\sigma)} | F_\sigma)$ as follows.

(3.3) $\quad E(M_{\phi(\sigma)} | F_\sigma) = \sum_s E(M_{\phi(s)} | F_s) I_{\sigma=s}$

where the summation is taken aver the finite set T. We may write $M_{\phi(s)}$ as

(3.4) $\quad M_{\phi(s)} = \sum_{t \geq s} M_t \cdot I_{\phi(s)=t}$

where again the summation is taken over the finite set of t such that $t \geq s$. Since ϕ is a decision function, the random variable $I_{\phi(s)=t}$ is F_s-measurable. Thus, we have

(3.5) $\quad E(M_{\phi(s)} | F_s) = \sum_{t \geq s} E(M_t | F_s) I_{\phi(s)=t}.$

Since $\{M_t : t \in T\}$ is a submartingale and since $t \geq s$ we have $E(M_t | F_s) \geq M_s$. Using this inequality

in (3.5), we obtain

$$(3.6) \qquad E(M_{\phi(s)} | F_s) \geq M_s \cdot \sum_{t \geq s} I_{\phi(s)=t}.$$

Since $\phi(s) \geq s$ (because ϕ is a decision function), we have $\sum_{t \geq s} I_{\phi(s)=t} = 1$ in (3.6) and we obtain

$$(3.7) \qquad E(M_{\phi(s)} | F_s) \geq M_s.$$

Substituting (3.7) into (3.3) yields

$$(3.8) \qquad E(M_{\phi(\sigma)} | F_\sigma) \geq \sum_s M_s \, I_{\sigma=s}$$

which is clearly the desired result (3.2).

To finish the proof we note that

$$(3.9) \qquad E(M_{\phi^n(\sigma)} | F_\sigma) \geq M_\sigma$$

for all $n \geq 0$, and since T is finite we must have $\phi^n(\sigma) = \phi^\infty(\sigma) = \tau$ for n sufficiently large. This completes the proof. ///

Extending this result to infinite time sets T requires additional hypotheses on the submartingale

$\{M_t : t \in T\}$ and the stopping times σ and τ. For example, even if T is assumed to be discrete, bounded it is possible for σ and τ to take infinitely many values. Nevertheless, on the set $\{\omega : \sigma(\omega) = s\}$ the stopping time τ can take only a finite number of values since T is discrete and bounded and $s \leq \tau(\omega)$ for ω inthis set. If we assume that $E(|M_\tau|)$ and $E(|M_\sigma|)$ are finite, then the proof of theorem 3.1 goes through with very little change. For fixed s we can use essentially the same argument to show that

$$(3.10) \qquad E(M_{\phi^n(s)} | \mathcal{F}_s) I_{\sigma=s} \geq M_s \cdot I_{\sigma=s}, \text{ for } n \geq 0.$$

Since T is discrete and bounded, we must have $\phi^n(s) = \tau$ on the set $\{\sigma=s\}$ for some finite n sufficiently large. Thus, from (3.10) we obtain

$$(3.11) \qquad E(M_\tau | \mathcal{F}_s) \cdot I_{\sigma=s} \geq M_\sigma \cdot I_{\sigma=s}.$$

Using the fact that T is countable and that $E(|M_\tau|)$, $E(|M_\sigma|)$ are finite, we can sum (3.11) over s in T to obtain the desired result.

The case of discrete, unbounded time requires more complicated hypotheses. If we assume that $E(|M_{\phi^n(\sigma)}|) < \infty$

for each $n \geq 0$, then the argument of theorem 3.1 can be used to show that $\{M_{\phi^n(\sigma)} : n \geq 0\}$ is a submartingale with respect to $\{F_{\phi^n(\sigma)} : n \geq 0\}$. Since $\phi^\infty(\sigma) = \tau$ we know that for each ω there is an integer $n \geq 0$ such that $\phi^\infty(\sigma(\omega),\omega) = \phi^n(\sigma(\omega),\omega)$, but in general there will be no single n unformly valid for all ω. However, we can define a one-parameter stopping time N as follows: let $N(\omega)$ be the smallest integer n such that $\phi^n(\sigma(\omega),\omega) = \phi^\infty(\sigma(\omega),\omega)$. Equivalently, $N(\omega)$ is the smallest integer n such that $\phi^{n+1}(\sigma(\omega),\omega) = \phi^n(\sigma(\omega),\omega)$ and for this reason, N is a one-parameter stopping time with respect to the increasing family of σ-fields $\{F_{\phi^n(\sigma)} : n \geq 0\}$. The optional sampling theorem for τ and σ will be true if

$$E(M_{\phi^N(\sigma)} \mid F_\sigma) \geq M_\sigma. \tag{3.12}$$

A sufficient condition that (3.12) is true can be obtained from the one-parameter theory presented in Doob (1953), for example. Thus, we have:

Theorem 3.2

Suppose that $(T,\leq)$ is a discrete p.o. set and let

τ and σ be stopping times with respect to the increasing family of σ-fields $\{\mathcal{F}_t : t \in T\}$. Suppose that $\{M_t : t \in T\}$ is a submartingale with respect to $\{\mathcal{F}_t : t \in T\}$. Furthermore, suppose that the following conditions are satisfied:

(3.13) τ is reachable from σ.

(3.14) $E(|M_{\phi^n(\sigma)}|) < \infty \quad n \geq 0, \; E(|M_\tau|) < \infty.$

(3.15) $\liminf_{n \to \infty} E(M_{\phi^n(\sigma)} \cdot I_{\{N>n\}} \; \{M_{\phi^n(\sigma)} > 0\}) = 0.$

Then $E(M_\tau | \mathcal{F}_\sigma) \geq M_\sigma$.

Proof:

Conditions (3.14) and (3.15) are sufficient conditions to show $E(M_N | \mathcal{F}_0) \geq M_0$ where $\{M_n : n \geq 0\}$ is the one-parameter submartingale defined $M_n = M_{\phi^n(\sigma)}$ with respect to the family $\{\mathcal{F}_n : n \geq 0\}$ defined $\mathcal{F}_n = \mathcal{F}_{\phi^n(\sigma)}$ and N is the one-parameter stopping time defined above. ///

It is unclear at present how to simplify these conditions in (3.14) and (3.15) to more direct conditions on σ, τ, ϕ and $\{M_t : t \in T\}$. Moreover, since several decision functions may yield the same stopping time τ, it would be preferable to have conditions directly in

terms of σ, τ and $\{M_t : t \in T\}$ without mention of a particular decision function ϕ. For the two-parameter case in section 1.5 we were able to state such conditions in (1.5.1) and (1.5.2).

An extension to a continuous, compact p.o. time set T requires hypotheses to enable one to obtain the continuous parameter reachable stopping time τ as a limit of discrete, bounded reachable stopping times τ_n so that the sequence $\{\tau_n\}$ decreases monotonically to τ with respect to the partial order on T. If $\{M_t : t \in T\}$ is also assumed right continuous with respect to the partial order, then we may prove the bounded, continuous time case as the limit of the finite case in theorem 3.1. This is what we did in the case $T = Z_+^2$ in section 1.6. However, in that case we required no elaborate <u>definition</u> of continuous parameter reachable stopping times as a limit of discrete parameter reachable stopping times, but rather we <u>deduced</u> that each continuous parameter stopping time was the limit of a decreasing sequence of discrete stopping times. However, in that case each stopping time was automatically reachable and we did not require an elaborate definition of reachable stopping time in terms of decision functions. We will not pursue the continuous parameter case any further.

4. Reachable Stopping Times and the Stopping Problem with Partially Nested Information

From Definition 1 of the previous section it should be clear that if ϕ is a decision function, then $\phi^{\infty}(t)$ is a stopping time which is reachable from the constant stopping time t. Thus, we could reformulate the stopping problem with partially nested information of section 3.2 in (3.2.19) in terms of reachable stopping times as follows: for a given initial time θ in T choose a stopping time τ_* reachable from θ such that

$$(4.1) \qquad E(C(\tau_*) \mid F_\theta) \leq E(C(\tau) \mid F_\theta)$$

for all stopping times τ reachable from θ. In section 3.2 we formulated the stopping problem with constraint sets $\{C_t : t \in T\}$, and it is also possible to introduce constraints in the reachable stopping time formulation. Let R_θ be the set of all stopping times reachable from θ and such that for $\tau \in R_\theta$ we have $\tau(\omega) = t$ only if $t \in C_t(\omega)$. Then the constrained stopping problem is to find τ_* in R_θ such that (4.1) is true for all τ in R_θ.

As we saw in chapter 3, Haggstrom (1966) formulated an optimal stopping problem on trees in terms of a control

variable which corresponds closely to our decision functions when the p.o. time set is a tree. We also saw in section 4.2 that control variables are reachable stopping times. Thus, one may consider Haggstrom's stopping problem in terms of the reachable stopping time formulation as well as the decision function formulation. Indeed, for p.o. time sets which are trees there is little distinction between decision functions and reachable stopping times. The reason for this is that in the case of trees, a reachable stopping time τ is actually a sequence $\tau_1\tau_2\tau_3\ldots$ which uniquely defines the decision function which reaches τ. Intuitively, one may thing of τ as a random function assigning a branch of the tree to each ω in the sample space Ω. This branch indicates the sequence of decisions which must be made to reach τ. Thus, in figure 4.1 if $\tau(\omega) = t_1t_2\ldots t_n$ then the decision function ϕ corresponding to τ makes the assignments $\phi(t_i,\omega) = t_{i+1}$ for $1 \leq i \leq n-1$ and $\phi(t_n,\omega) = t_n$.

For general partially ordered time sets the relationship between decision functions and reachable stopping times is not as close as in the case of trees. A stopping time τ which is reachable from σ may be reachable via different decision functions ϕ. For example,

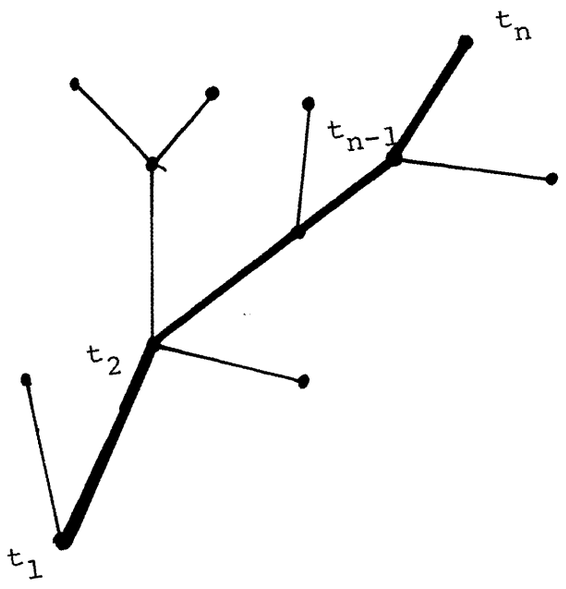


Figure 4.1

suppose T is the two-parameter time set $\{(0,0), (0,1), (1,0), (1,1)\}$ discussed in example 2 of section 1.3 and illustrated in figure 4.2. Define $\phi_1((0,0)) = (0,1)$, $\phi_1((0,1)) = (1,1)$, $\phi_1((1,0)) = (1,0)$ and $\phi_1((1,1)) = (1,1)$; define $\phi_2((0,0)) = (1,0)$, $\phi_2((1,0)) = (1,1)$, $\phi_2((1,1)) = (0,1)$. Then $\phi_1^\infty((0,0)) = \phi_2^\infty((0,0) = (1,1)$ but clearly $\phi_1 \neq \phi_2$. Both decision functions reach (1,1) from (0,0) but ϕ_1 reaches (1,1) by going through (1,0) and ϕ_2 reaches (1,1) by going through (0,1).

Note also that more than one reachable stopping time may correspond to the same decision function. In the example above we have that $\phi_1^\infty((1,0)) = (1,1)$, but $\phi_1^\infty((1,0)) = (1,0)$. Thus, ϕ_1 defines the constant stopping time (1,1) reachable from (0,0) and the different stopping time (1,0) reachable from (1,0).

Although it is possible to formulate the stopping problem with partially nested information in terms of reachable stopping times, it seems more advantageous to use the decision function formulation. The most important advantage is that the decision function formulation leads directly to the dynamic program equation which allows

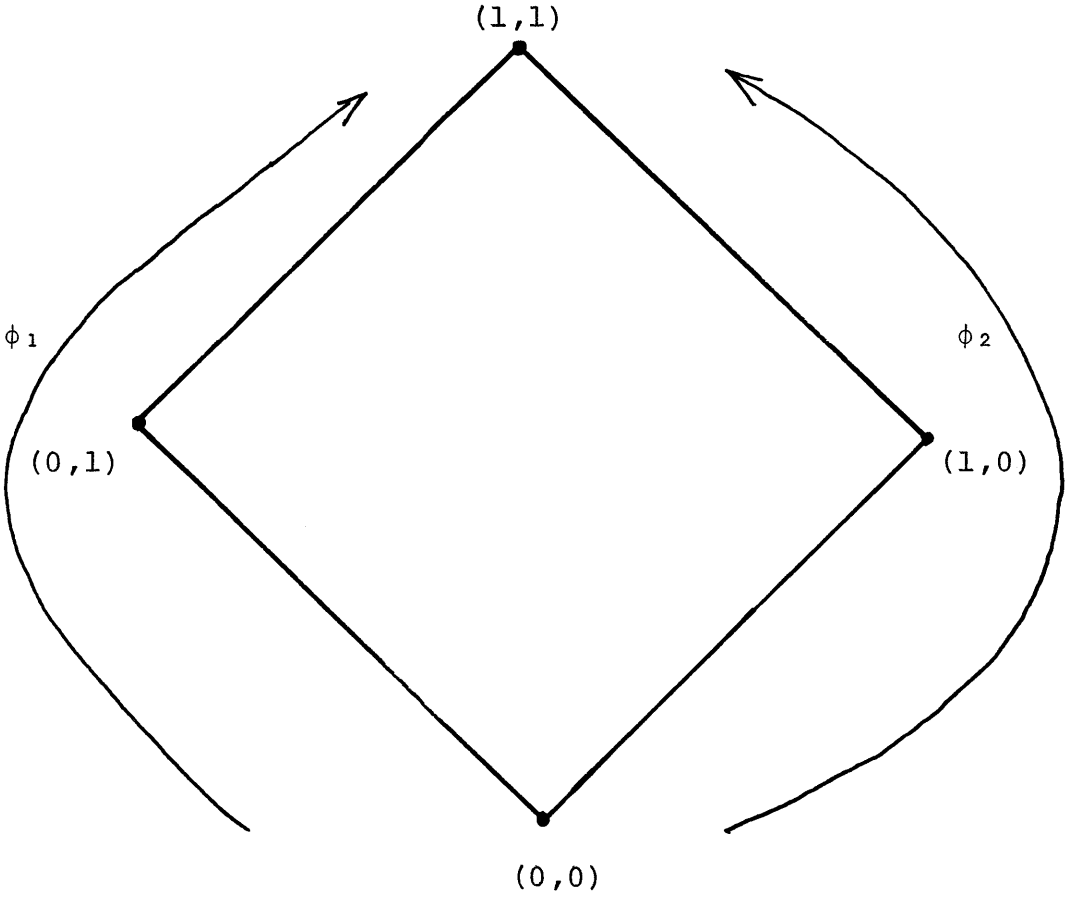


Figure 4.2

recursive computation of the optimal cost and an optimal decision function for the stopping problem. Moreover, except in the one-parameter and special two-parameter cases, reachable stopping times must be defined in terms of decision functions. We will discuss these points again in chapter 5.

5. Converse of the Optimal Sampling Theorem

In this section we apply the dynamic programming theorem 3.3.1 to show that, at least for finite time sets, reachable stopping times are the only stopping times which satisfy the optimal sampling theorem for arbitrary submartingales.

Theorem 5.1

Suppose that τ and σ are stopping times on the discrete, bounded p.o. set $(T,\leq)$ with respect to the increasing family of σ-fields $\{F_t : t \in T\}$ such that $\sigma \leq \tau$. Suppose that for any submartingale $\{M_t : t \in T\}$ with respect to $\{F_t : t \in T\}$ we have that $M_\sigma \leq E(M_\tau | F_\sigma)$. Then τ must be reachable from σ.

Proof:

Consider the following stopping problem on T with the partially nested information structure $\{F_t : t \in T\}$.

Define the cost C as

(5.1) $$C(t) = I_{\tau \neq t}.$$

Thus, $C(t)$ is $\mathcal{F}_t$-measurable and $C(t) = 0$ if and only if $\tau = t$, otherwise $C(t) = 1$. Define constraint sets $\{C_t : t \in T\}$ such that

(5.2) $$C_t = \{s : t \leq s\}.$$

Since T is discrete, bounded, there exists a unique optimal cost function π^* such that

(5.3) $$\pi^*(t) = \min \{\rho(s,t) : t \leq s\}$$

where $(s,t) = E(\pi^*(s)|\mathcal{F}_t)$ for $t < s$ and $\rho(t,t) = c(t)$. In particular, we see that

(5.4) $$\pi^*(t) \leq E(\pi^*(s)|\mathcal{F}_t)$$

for all s such that $t \leq s$, so that $\{\pi^*(t) : t \in T\}$ is a submartingale with respect to $\{\mathcal{F}_t : t \in T\}$. By definition, $\pi^*(t)$ is given by

$$(5.5) \qquad \pi^*(t) = \text{ess inf } \{E(C(\phi^\infty(t)) \mid \mathcal{F}_t) : \phi \in D\}$$

where D is the set of admissable decision functions for this problem. From (5.5) it should be clear that $\pi^*(\tau) = 0$ (use the decision function $\phi(t) \equiv t$). Moreover, it should be clear that $\pi^*(t) \geq 0$ for all t and hence $\pi^*(\sigma) \geq 0$.

By assumption, the optional sampling theorem with τ and σ is satisfied for all submartingales on T with respect to $\{\mathcal{F}_t : t \in T\}$. Thus, we have

$$(5.6) \qquad \pi^*(\sigma) \leq E(\pi^*(\tau) \mid \mathcal{F}_\sigma)$$

and it follows that $\pi^*(\sigma) = 0$.

Because the time set T is discrete and bounded, we can calculate an optimal decision function ϕ_* corresponding to π^* by dynamic programming so that

$$(5.7) \qquad \pi^*(t) = E(C(\phi_*^\infty(t)) \mid \mathcal{F}_t)$$

for all t in T. It follows that

$$(5.8) \qquad \pi^*(\sigma) = E(C(\phi_*^\infty(\sigma)) \mid \mathcal{F}_\sigma)$$

and since $\pi^*(\sigma) = 0$ and since $C \geq 0$ we must have (a.s.)

$C(\phi_*(\sigma)) = 0$. But this means that $\tau = \phi_*^{\infty}(\sigma)$ and thus, τ is reachable from σ with the decision function ϕ_*. ///

Theorems 3.1 and 5.1 provide a complete characterization of the stopping times for which the submartingale version of the optional sampling theorem is true. That is, all such stopping times are reachable (or more precisely, the optional sampling theorem is true for stopping times $\tau \geq \sigma$ if and only if τ is reachable from σ). It may happen that for some martingale $\{M_t : t \in T\}$ we have

$$E(M_\tau | F_\sigma) \geq M_\sigma \tag{5.9}$$

even though τ is <u>not</u> reachable from σ. We noted this fact in the conclusion of chapter 1 (section 1.7). For example, if T is a <u>directed</u> set and the submartingale $\{M_t : t \in T\}$ has a Doob decomposition as noted in section 1.7, then (5.9) is true for all stopping times, reachable or not. However, if τ is <u>not</u> reachable from σ, then Theorem 5.1 implies that for some submartingale $\{M_t : t \in T\}$ the relation (5.9) will be false.

CHAPTER 5

CONCLUSIONS

Part I of this thesis has consisted of an extensive investigation of the optional sampling theorem for submartingales indexed by a partially ordered index set. We started this investigation in chapter 1 by studying the particular case of two-parameter time sets. There we discovered that when the σ-fields satisfy a conditional independence hypothesis (as, for example, do the σ-fields generated by the two-parameter Wiener process), then the optional sampling theorem is true for all stopping times and all submartingales (assuming uniform integrability conditions where necessary). Unfortunately, this result depended crucially on the particular parameter set and probability structure. Changing either the parameter set (e.g., letting T be a three-parameter time set) or the probability structure (i.e. not assuming conditional independence of the σ-fields) allowed us to construct simple counterexamples to the optional sampling theorem.

After studying the optional sampling theorem for two-parameter processes we digressed in chapter 2 to study a simple stopping problem (the signal interception problem) for two-parameter processes. For one-parameter

processes there is a close relationship between the optional sampling theorem and stopping problems, as Snell (1952) first investigated. Thus, our digression to study stopping problems for two-parameter processes was in fact an indirect attack on the optional sampling theorem.

In chapter 3 we extended the two-parameter example to cases with more general partially ordered time sets. Here we found that the general stopping problem with partially nested information allows one to formulate naturally a wide variety of stopping problems which are difficult or impossible to formulate in the conventional one-parameter framework. We then showed that one can extend the dynamic programming solution of the one-parameter stopping problem to the more general case of partially ordered time. The resulting dynamic program is backward recursive with respect to the partial order, and this recursive property enables one to use the full structure of the partially ordered time set to calculate the solution to the stopping problem in an efficient manner. Note that the results of chapter 3 extend the work of Haggstrom (1966).

In chapter 4 we returned to consider the optional sampling theorem for general partially ordered time sets.

Here we defined the concept of a reachable stopping time, based on the concept of decision function introduced in chapter 3 to define the stopping problem with partially nested information. We showed that the optional sampling theorem is true for submartingales if the stopping times are reachable. We also showed that the two-parameter stopping times considered in chapter 1 are reachable. Finally, we showed that if σ and τ are stopping times such that $\sigma \leq \tau$ and

$$(1) \qquad E(M_\tau | \mathcal{F}_\sigma) \geq M_\sigma$$

for all submartingales $\{M_t\}$, then τ must be reachable from σ. The proof of this result was a simple corollary of the dynamic programming result of chapter 3. Thus, in chapter 4 we obtained a complete characterization of the stopping times σ, τ which allow (1) to be true for all submartingales $\{M_t\}$. We have also found that the optional sampling is intimately associated with sequential stopping problems -- namely, reachable stopping times are defined by sequential decision functions as described in chapter 3. Thus, in a sense the optional sampling theorem for partially ordered time sets is necessarily a one-parameter result. That is, the optional

sampling theorem is true for general submartingales if and only if the stopping times are reachable and hence, if and only if the theorem can be reduced to its one-parameter version.

In the remainder of this chapter we will discuss the possibility of extending these results on optional sampling and stopping problems to continuous parameter time sets. In the context of the stopping problem with partially nested information this means that we are going to consider cases where decisions are made continuously rather than sequentially (i.e., the decision trajectories are continuous trajectories). In the context of the optional sampling theorem this means we are going to consider cases where a stopping time τ can have a continuum of values in its range $\tau(\Omega)$.

Although the generalization of these results to the continuous parameter case described above is difficult, it seems worthwhile, particularly for study stopping problems with partially nested information. Let us briefly explain why this should be true.

To extend the results of chapter 3 to the case where decisions are made continuously, one must make sense of differential equations of the form

$$(2) \qquad \frac{d\tau}{ds} = \phi(\tau(s))$$

where ϕ is a decision function and $s \to \tau(s)$ is a continuous trajectory of stopping times. In (2) we assume that the time set is some linear space such as R^n. In order that the path $s \to \tau(s)$ should incrase we must assume that the components of ϕ are all nonnegative.

For a given stopping problem the optimal cost function will satisfy a differential equation version of the dynamic programming equation. To see what such an equation will look like, consider equation (2.3.6) in chapter 2, namely

$$(3) \quad \psi(x,t) = \inf \{\psi(x+\delta\nu, t+\delta) : ||\nu|| \leq e\} + \delta q(x,t)$$

where ν, x lie in R^n and t lies in R^1. Recall that the parameter set was R^{n+1} and the partial order relation was the space-time ordering. Assuming that ψ is continuously differentiable in x and t and assuming that we can interchange the infimum and limit $(\delta \to 0)$ operations, we obtain

$$(4) \quad 0 = \inf \{<\frac{\partial \psi}{\partial x}, \nu> + \frac{\partial \psi}{\partial t} : ||\nu|| \leq c\} + q(x,t)$$

from (3) by letting $\delta \to 0$. Note that we can solve the infimum problem in (4) and obtain the partial differential equation

$$(5) \qquad 0 = c \cdot \left|\left|\frac{\partial \psi}{\partial x}\right|\right| + \frac{\partial \psi}{\partial t} + q(x,t)$$

for ψ.

The reason for studying the continuous parameter stopping problem is that there is a better chance of obtaining exact solutions to a wide class of problems if we can formulate the problem as a differential equation as in (5). Moreover, there seems a better chance of finding asymptotic approximations of (5) than of the discrete version (3).

Now let us turn our attention to the possibility of extending reachable stopping times and the optional sampling theorem to continuous parameter stopping times. Note that we were able to do this in a special two-parameter case in chapter 1. However, this special case doesn't offer much help in the general case. In the special case of chapter 1 the optional sampling theorem was true for all continuous parameter stopping times. Hence, there was no difficulty of defining reachable

stopping times in the continuous parameter case. In general, however, not all continuous parameter stopping times will be reachable (i.e., will satisfy the optional sampling theorem) and we must somehow distinguish reachable, continuous parameter stopping times. The most natural approach is to say a stopping time τ is reachable if it is the limit

$$\tau = \lim_{s\to\infty} \tau(s) \tag{6}$$

where $\tau(s)$ satisfies a differential equation (2) for some decision function. This approach is feasible in R^n and other linear spaces, but it obviously fails if we cannot define differential equations in the partially ordered time set. For example, how would one define (2) in case the time set were the collection of all closed, convex sets in R^2? The alternative approach is to define a continuous parameter reachable stopping time as the limit of discrete parameter reachable stopping times. This approach will obviously work for general continuous parameter time sets, but the definition of reachable stopping time is rather crude and unwieldy. The continuous parameter reachable stopping time τ depends on a sequence of discrete parameter stopping times $\{\tau_n\}$, or equivalently

on the corresponding decision functions $\{\phi_n\}$. It would seem more elegant to define reachability of τ in terms of a single decision function ϕ, but this may be impossible.

The results of chapter 4 on the converse of the optional sampling theorem offer another approach that seems very effective. Instead of defining reachability first and then proving the optional sampling theorem, let us define reachability in terms of the optional sampling result (1) for submartingales. That is, define τ to be reachable from σ if (1) is true for all submartingales $\{M_t\}$. This definition of reachability is just as easy to make for continuous parameter τ, σ as for discrete parameter τ, σ. Having defined reachability in this way, the next step is to see if τ can be characterized in terms of a single decision function or a sequence of decision functions (i.e., as the limit of the corresponding stopping times). The best strategy in this investigation seems to be to follow the derivation of theorem 4.5.1, the converse of the optional sampling theorem. Using the discrete parameter dyanmic programming results of chapter 3, we can investigate solutions of the stopping problem with the cost function c defined by

(7) $$c(t) = 1_{\tau \neq t}$$

where τ is the continuous parameter stopping time we are studying. It seems reasonable that an ε-optimal solution τ_ε of this stopping problem will be an approximation to τ and that $\tau_\varepsilon \to \tau$ as $\varepsilon \to 0$.

The previous discussion has been carried out only at the heuristic level and all the conjectures discussed will require rigorous proof. In conclusion, let us note that Kurtz (1978) has investigated a class of continuous parameter stopping times which satisfy a differential equation of the form (1). Kurtz (1978) investigated these multiparameter stopping times while studying processes $X(t)$ (here $t \in R^1$) of the form

(8) $$X(t) = X(0) + \sum_{i=1}^{N} \alpha_i W_i\left(\int_0^t \beta_i(X(s))ds\right)$$

where W_i are independent Wiener processes. The i-th component of the stopping times $\tau(t)$ is given by

(9) $$\tau_i(t) = \int_0^t \beta_i(X(s))ds$$

Note that τ_i represents a random time change for the

Wiener process W_i. We conjecture that the stopping time τ defined by (9) will be reachable in our sense and hence, that the optional sampling theorem for submartingales is true for these stopping times. Note that Kurtz (1978) used the martingale version of theorem which he had proved earlier in Kurtz (1977).

PART II: MULTIPARAMETER STOCHASTIC CALCULUS

CHAPTER 1

INTRODUCTION TO MULTIPARAMETER STOCHASTIC DIFFERENTIATION FORMULA

Over the past few years increasing interest has developed in the study of random fields, that is, random "processes" with multiparameter "time" parameters. Several authors have discussed the multiparameter Wiener process and the closely related multiparameter generalization of Ito's stochastic integral. However, when extending Ito's stochastic differentiation formula to the multiparameter case, Wong and Zakai (1974, 1975) and Cairoli and Walsh (1975) discovered that the two-dimensional formula required a new type of stochastic integral and mixed stochastic-deterministic integrals different from the Ito-type multidimensional integral defined by Ito (1951).

In this part of the thesis we shed new light on the Wong-Zakai and Cairoli-Walsh results by showing that the stochastic differentiation formula is a natural consequence of the ordinary deterministic differentiation formula in several parameters and the one-parameter Ito stochastic differentiation formula. A brief description of the contents and organization of this part follows.

In chapter 2 we present some background material for multiparameter processes. Chapter 3 then discusses the two parameter formula and the various new types of integrals required to interpret the formula. We compare the integrals with the Wong-Zakai and Cairoli-Walsh integrals, thus establishing the equivalence of their formulas and our own. We also introduce a stochastic partial differential operator that obeys very simple differentiation rules and allows a simple formal derivation of the stochastic differentiation formula. This stochastic partial differentiation operator is the same as Cairoli-Walsh's stochastic partial. Chapter 4 discusses the n-parameter stochastic differentiation formula and the n-parameter integrals necessary to interpret it rigorously. Here we emphasize the concept of a stochastic differential. This point of view permits a simple formal derivation of the multiparameter stochastic differentiation formula and emphasizes the multiparameter formula's close relation to the deterministic formula and the one-parameter Ito formula. The derivation up to this point is formal in nature, as we utilize differentiation formulas and differentials without rigorous proof of their properties. In the next three chapters we rigorously interpret the

differential formulas of chapters 3 and 4 in terms of stochastic integrals and stochastic measures. In chapter 5 we define the concept of a stochastic measure and its integral so that we can treat all types of stochastic integrals together. Chapter 6 defines mixed integrals and products of stochastic measures. Chatper 7 presents the rigorous proof of the product differentiation rule, which is the central technical results of this paper. In chapter 8 we apply the multiparameter differentiation formula to represent multiparameter square integrable martingales by multiparameter stochastic integrals. Finally, in chapter 9 we conclude with a discussion of possible extensions of our results. The appendices to part II contain some proofs and technical details for chapters 3 through 7.

CHAPTER 2

PRELIMINARY BACKGROUND AND NOTATION

2.1 Introduction

Before proceeding to the main results of this paper it is first necessary to present some background material concerning multiparameter stochastic process, or random fields. Several authors have developed this material which has become fairly standard. The references given in this section are by no means exhaustive and we refer the reader to the introduction to this thesis for a more complete discussion. Cairoli and Walsh (1975) give a detailed discussion of the basic material and provide a detailed bibliography. Cairoli (1970, 1971) discusses multiparameter martingales in detail. Ito (1951) and McKean (1963) present the basic material on multiparameter Wiener processes and integrals. See also McKean's (1963) bibliography.

This paper treats multiparameter stochastic processes for which the parameter belongs to the unit cube $[0,1]^n$ in R^n. We use I, I_1, I_2, and so on to denote such unit cubes; the index serves only to distinguish different cubes and not to indicate dimensionalty. In section 2.2 we define the important concept of a partial ordering of

$[0,1]^n$. This allows us to generalize the notion of an increasing family of σ-fields, adapted process and martingale to multi-parameter processes. In 2.3 we introduce the Gaussian random measure and the corresponding multiparameter Wiener process. Section 2.4 defines partitions and partition processes of $[0,1]^n$ and 2.5 uses these definitions to construct the Ito-type stochastic integral from a Gaussian random measure. The reader may also refer to chapter 1 of part I for some of these basic definitions.

2.2 Partial Order

We write $x < y$ for x and y in I if $x_i \leq y_i$. With this notion of partial order we can define an increasing family of σ-fields parameterized on I. Denoting the basic probability space by (P,Ω,F), we say that $\{F_z | z\varepsilon I\}$ is an increasing family of σ-fields if F_z is a sub σ-field of F for each z in I and if $F_{z_1} \subset F_{z_2}$ whenever $z_1 < z_2$. We assume that (P,Ω,F) is complete and F_o contains all P-negligible sets of F.

For an increasing family $\{F_z | z\varepsilon I\}$ of σ-fields, a function ϕ mapping $\Omega \times I$ into R is adapted to $\{F_z | z\varepsilon I\}$ if ϕ_z is F_z-measurable for each z in I.

We say ϕ is a measurable process if it is $\mathcal{F}\times\mathcal{B}$-measurable as a mapping from $\Omega\times I$ into R. The σ-field $\mathcal{B}$ is the usual family of Borel sets of I. We often write ϕ as ϕ_z to emphasize its dependence on the parameter z. Given the framework developed to this point, it is natural to define a martingale ϕ as a measurable adapted process such that

$$(2.1) \qquad E\{\phi_{z_2} | \mathcal{F}_{z_1}\} = \phi_{z_1} \quad \text{a.s.}^{\dagger}$$

for each z_1 and z_2 in I such that $z_1 < z_2$. See Cairoli (1974,1975) and part I, chapter 1 for further material on multiparameter martingales.

2.3 Gaussian Random Measure and Multiparameter Wiener Process

Following Cairoli and Walsh (1975), we call Φ a Gaussian random measure if Φ assigns a zero mean Gaussian random variable, $\Phi(A)$, with variance $m(A)$ to each Borel set A in I (here m is Lebesgue measure). Furthermore, if A and B are disjoint Borel sets, then $\Phi(A)$

---

†We suppress "a.s." (almost surely) in the rest of the paper, but it is always understood in equations of random variables.

and $\Phi(B)$ are independent random variables and

$$\Phi(A \cap B) = \Phi(A) + \Phi(B)$$

Let $[0,z]$ denote the set of all ξ in I such that $\xi < z$, and define F_z as the smallest σ-field such that $\Phi(A)$ is F_z-measurable for all A included in $[0,z]$. Then $\{F_z | z \varepsilon I\}$ is an increasing family of σ-fields and we can naturally associate a martingale with the measure Φ and this increasing family of σ-fields. Thus, define W_z as

$$W_z = \Phi([0,z]).$$

It is easy to check that W_z is a martingale with respect to $\{F_z | z \varepsilon I\}$. In the one-dimensional case W_z is, in fact, the standard Wiener process. Thus, W_z is called the multiparameter Wiener process in the higher dimensional cases. Keeping the analogy with the one parameter case, we can define a stochastic integral in terms of the multiparameter Wiener process just as one defines the Ito stochastic integral for the one-parameter Wiener process. First, however, we must define a partition of the cube I and a partition process.

2.4 Partitions and Partition Processes

A set, P, of real numbers is a partition of the interval, $[0,1]$, if P is a finite subset of $[0,1]$ and if P contains the real numbers 0 and 1. A partition of a higher dimensional cube, $[0,1]^n$, is a subset, P, of $[0,1]^n$ which is the Cartesian product of partitions, $P_1, P_2, \ldots, P_n$, of the interval $[0,1]$. That is,

$$P = P_1 \times P_2 \times \ldots \times P_n.$$

If P is a partition of $[0,1]$ and z is in P but $z \neq 1$, then z^+ denotes the smallest element of P which is strictly greater than z.† If $z = 1$, then define $z^+ = 1$. If P is a partition of $[0,1]^n$ and z is in P, define z^+ coordinate-wise. That is, P is a product of one-dimensional partitions, $P_1, P_2, \ldots, P_n$, and we define the m-th coordinate of z^+ as z_m^+ where z_m is the m-th coordinate of z and z_m^+ is defined with respect to P_m as above. For example, a partition of $[0,1]^2$ is a rectangular lattice of points, and figure 5.1 of chapter 5 illustrates the relation between z and z^+.

Write $x << y$ for x and y in I if $x_i < y_i$

†Wong-Zakai (1974) use the z^+ notation.

for each coordinate and define

$$\delta z = \{\xi \varepsilon I \mid z << \xi < z^+\} = (z, z^+]^\dagger$$

for each z in P. Note that $\delta z \cap \delta z' = \phi$ if z and z' are in P and $z \neq z$; also note that $I = \bigcup_{z \varepsilon P} \delta z$. Sometimes we use the notation z^+ without referring to any partition; in those cases z^+ just denotes an arbitrary point of I such that $z < z^+$.

Define the norm, $|P|$, of a partition P as

$$|P| = \max_{z \varepsilon P} ||z^+ - z||,$$

where $||x|| = \max |x_i|$ for each x in I.

A multiparameter process ϕ on I is a partition process with respect to the increasing family $\{F_z \mid z \varepsilon I\}$ if there is a partition P of I such that

$$\phi = \sum_{z \varepsilon P} \phi_z \chi_{\delta z} \tag{2.2}$$

where ϕ_z is F_z-measurable for each z in P and

---

†Cairoli-Walsh (1975) denote δz by $(z, z^+]$ in [4].

$E\{\phi_z^2\}$ is finite for each z in P. When a partition process satisfies (2.2), we say it is defined on P. Thus, if P' is a partition of I and $P' \supset P$, then a partition process defined on P is also defined on P'. We write $L_{pp}^2(\Omega\times I)$ for the space of all partition processes on $\Omega\times I$. Similarly, let $L_a^2(\Omega\times I,\ell)$ denote the space of all measurable processes ϕ adapted to $\{F_z | z\varepsilon I\}$ such that

$$\int E\{\phi_z^2\}\ell(dz) < \infty. \tag{2.3}$$

The measure ℓ is a Borel measure on I. Using typical measure theory arguments, one can easily show that $L_{pp}^2(\Omega\times I)$ is a dense linear subspace of $L_a^2(\Omega\times I,\ell)$.

2.5 Ito-type Stochastic Integral

With this concept of partition and partition process we can easily define an Ito-type stochastic integral on I. Choose the increasing family $\{F_z | z\varepsilon I\}$ of σ-fields generated by the Gaussian random measure, Φ. Suppose ϕ is a partition process with respect to this increasing family and suppose ϕ is defined on the partition P. Then

$$(2.4) \qquad \int \phi_z \Phi(dz) = \sum_{z \in P} \phi_z \Phi(\delta z)$$

defines a linear mapping from $L_{pp}^2(\Omega \times I)$ into $L^2(\Omega)$, the space of random variables with finite variance. For each ϕ in $L_{pp}^2(\Omega \times I)$ we have that

$$(2.5) \qquad E\{(\int (\phi_z(dz))^2\} = \int E\{\phi_z^2\} dz,$$

and thus, the mapping defined by (2.4) has a unique continuous extension which is defined on all of $L_a^2(\Omega \times I, m)$. In addition, this extension satisfies (2.5) for all ϕ taken in $L_a^2(\Omega \times I, m)$. In the case that $I = [0,1]$, this integral is the Ito stochastic integral, and thus, we often call it the Ito-type multiparameter stochastic integral to distinguish it from other types of stochastic integrals defined later.

The multiparameter Ito-type integral generates martingales just as in the one-parameter case. Define a process M on I by

$$(2.6) \qquad M_z = \int \chi_{[0,z]}(\xi) \phi_\xi \Phi(d\xi)$$

for each z in I and a fixed ϕ in $L_z^2(\Omega \times I, m)$. Then

M is a martingale with respect to the σ-fields generated by the Gaussian random measure, Φ.

We could also define a Gaussian random measure for which m is an arbitrary finite Borel measure instead of Lebesgue measure as in section 2.3. We could then construct an Ito-type integral from this measure in exactly the same way as we have done, and this integral would also define a martingale as in (2.6). To distinguish between this more general Gaussian random measure and the measure defined in 2.3, we call the latter the <u>standard Gaussian random measure</u>. In this paper we are particularly interested in the Gaussian random measure defined

$$A \to \int \chi_A(\xi) p(\xi) \Phi(d\xi)$$

where $p \in L^{\infty}(I)$ and Φ is the standard Gaussian random measure on I. One can find further details and references about the multiparameter Wiener process and its integral in Cairoli (1971), Ito (1951) and McKean (1963).†

†In the literature $\int \phi_z \Phi(dz)$ is usually denoted $\int \phi_z dW_z$ or $\int \phi_z W(dz)$.

CHAPTER 3

TWO-PARAMETER STOCHASTIC DIFFERENTIAL FORMULA

3.1 Introduction

In this chapter we introduce the two-parameter stochastic differential formula. In section 3.2 we review the one-parameter differential formula and apply it to the multi-parameter process defined in chapter 2. In section 3.3 we present the stochastic partial differential operator and its differentiation rules.* This allows a simple formal derivation of the two-parameter stochastic differentiation formula. Section 3.4 discusses the various stochastic, mixed and deterministic integrals on the plane that are needed to define the stochastic differentiation formula. We correlate these integrals with those of Wong-Zakai (1974, 1975) and Cairoli-Walsh (1975, 1977) and thus, show the equivalence of their differentiation formulas with our own.

3.2 One-Parameter Stochastic Differentiation Formula

Suppose M_t is a continuous, square integrable one-parameter martingale. Let $f(u,t)$ be a real-valued function of t and u in R with continuous partial derivatives up to second order. Then Kunita and Watanabe (1967) have shown that $f(M_t,t)$ satisfies the following differentiation formula.

---

*The differentiation rules are proved rigorously in chapter 7.

$$(3.1) \qquad f(M_T,T) - f(M_S,S) = \int_S^T f_u(M_\tau,\tau)dM$$

$$+ \int_S^T [f_t(M_\tau,\tau) + \frac{1}{2}f_{uu}(M_\tau,\tau)V_t(\tau)]d\tau$$

In (3.1) f_u, f_{uu} and f_t denote partial derivatives of f and V_t is the partial derivatives of $V(t) = \langle M,M\rangle_t$, the increasing process associated with the martingale M_t.* The first integral in (3.1) is a stochastic integral, defined so that

$$(3.2) \qquad \int_S^T \phi_\tau dM_\tau = \sum_{\tau\in P} \phi_\tau(M_{\tau+} - M_\tau)$$

if ϕ is a partition process defined on a partition P of $[S,T]$. Thus, the stochastic integral in (3.1) is an Ito-type stochastic integral as in section 2.5. The second integral in (3.1) is an ordinary Lebesgue integral.

How should one generalize (3.1) to the multiparameter case? First of all, we limit our consideration to multiparameter martingales of the form

---

*See Kunita and Watanabe (1967) or Meyer (1968) for definition of the increasing process of a martingale.

$$M_z = \int \chi_{[0,z]}(\xi)p(\xi)\Phi_0(d\xi), \tag{3.3}$$

where I_0 is the standard Gaussian random measure on $[0,1]^n$ and p is in $L^\infty([0,1]^n,m)$[†]. In chapter 9 we discuss the extension of these results to more general martingales and multiparameter processes. Associated with M_z, is the deterministic increasing process

$$V(z) = \int_{[0,z]} p(\xi)^2 d\xi. \tag{3.4}$$

Note that the one-parameter process $z_i \to M_{z_1,z_2,\ldots,z_n}$ is a continuous square integrable martingale with increasing process $z_i \to V(z_1,z_2,\ldots,z_n)$. Thus, we may use (3.1) to write (in the case i = 1)

$$\begin{aligned} f(M_{T,z_2,\ldots,z_n},T,z_2,\ldots,z_n) - f(M_{S,z_2,\ldots,z_n},S,z_2,\ldots,z_n) \\ = \int_S^T f_u(M_{\tau,z_2,\ldots,z_n},\tau,z_2,\ldots,z_n)\partial_{z_1} M_{\tau,z_2,\ldots,z_n} \\ + \int_S^T [f_{z_1}(M_{\tau,z_2,\ldots,z_n},\tau,z_2,\ldots,z_n) + \tfrac{1}{2}f_{uu}(M_{\tau,z_2,\ldots,z_n}) \\ V_{z_1}(\tau,z_2,\ldots,z_n)]d\tau. \end{aligned} \tag{3.5}$$

[†]Here m is the Lebesgue measue on $[0,1]^n$. Refer to section 2.5 for the definition of the Guassian random measure and integral (3.3).

In this formula, the stochastic integral must agree with (3.2). That is,

$$(3.6)\quad \int_S^T \phi_{\tau,z_2,\ldots,z_n}\,\partial_{z_1} M_{\tau,z_2,\ldots,z_n} = \sum_{\tau\in P} \phi_{\tau,z_2,\ldots,z_n}(M_{\tau^+,z_2,\ldots,z_n} - M_{\tau,z_2,\ldots,z_n}),$$

where P is partition of $[S,T]$ and $z_1 \to \phi_{z_1,z_2,\ldots,z_n}$ is a partition process defined on P. It is more convenient to write (3.5) in differential form, and then we can write the general i case as easily as the special i = 1 case. Thus, we have that

$$(3.7)\quad \partial_i f(M_z,z) = f_u(M_z,z)\partial_i M_z + [f_i(M_z,z) + \tfrac{1}{2} f_{uu}(M_z,z)V_i(z)]dz_i,$$

where for notational convenience we have used the subscript i instead of z_i to indicate partial differentiation and the stochastic differentials $\partial_{z_i} M$. The differential expression (3.7) is basically a one-parameter result. We are seeking, however, a truly n-parameter stochastic differentiation formula that expresses $f(M_z,z)$ as an n-dimensional integral and not a one dimensional integral as in (3.6).

3.3 Differentiation Rules and the Two-Parameter Formula

If we can represent the process ϕ as an integral,

$$\phi_z = \int_{[0,z]} \psi_\xi \Phi(d\xi),$$

then for any rectangle $(z_1,z_2] = (x_1,x_2]x(y_1,y_2]$ in $[0,1]^2$ we have

$$(3.8) \qquad \phi_{x_2,y_2} - \phi_{x_1,y_2} - \phi_{x_1,y_1} + \phi_{x_1,y_1} = \int_{(z_1,z_2]} \psi_\xi \Phi(d\xi).$$

Thus, we are motivated to look for an integral representation of $\phi_{x_2,y_2} - \phi_{x_1,y_2} - \phi_{x_2,y_1} + \phi_{x_1,y_1}$ where $\phi_z = F(M_z)$, F is a differentiable function, and $z \to M_z$ is the two-parameter process defined in section 3.2[†]. If (3.8) holds we want to say that

$$\partial_1\partial_2\phi_z = \psi_z \Phi(dz)$$

in some sense. If $z \to M_z$ were a deterministic differentiable function, we could write

$$(3.9) \qquad \partial_1\partial_2 F(M_z) = \frac{\partial^2 F(M_z)}{\partial_{z_1}\partial_{z_2}} dz_1 dz_2$$

The point of this work is to make sense of expressions like (3.9) in the stochastic case. We now introduce the

---

[†]Wong-Zakai (.974) also sought a representation of the process $F(M_z)$ in terms of the double difference

$$F(M_{x_2,y_2}) - F(M_{x_1,y_2}) - F(M_{x_2,y_1}) + F(M_{x_1,y_1}).$$

stochastic partial differential operator ∂_i and present differentiation rules for applying it (these rules are proved in chapter 7). At the same time we use these rules to derive the stochastic differentiation formula for the two-parameter case. Note that the stochastic partial differential operator agrees with the stochastic partial defined in Cairoli and Walsh (1975).

Rule 1: Ito differentiation rule.

First, ∂_i must operate on a process of the form $z \to f(M_z,z)$ according to the Ito formula (3.7). Thus, for example, we must have

$$\partial_1 F(M_z) = F'(M_z)\partial_1 M_z + \tfrac{1}{2}F''(M_z)V_1(z)dz_1,$$

$$\partial_2 F'(M_z) = F''(M_z)\partial_2 M_z + \tfrac{1}{2}F'''(M_z)V_2(z)dz_2,$$

and

$$\partial_2(F''(M_z)V_1(z)) = F'''(M_z)V_1(z)\partial_2 M_z + [V_{12}(z)f''(M_z) + \tfrac{1}{2}F''''(M_z)V_2(z)V_1(z)]dz_2.$$

Rule 2: Product differentiation rule.

Second, ∂_i must obey the product differentiation rule of ordinary calculus when operating on the product of two differentials or the product of a process and a

differential. For example, this rule means that

$$\partial_2(F'(M_z)\ {}_1M_z) = \partial_2(F'(M_z))\partial_1 M_z + F'(M_z)\partial_2\partial_1 M_z$$

and

$$\partial_2(F''(M_z)V_1(z)dz_1) = \partial_2(F''(M_z)V_1(z))dz_1 + F''(M_z)V_1(z)\partial_2(dz_1)$$

Rule 3: Deterministic differentials and linearity.

We also require that ∂_i acts like the deterministic operator $\frac{\partial(\cdot)}{\partial z_i}dz_i$ when it operates on deterministic differentials. Thus, $\partial_2(dz_1) = 0$. Finally, we assume that the operator ∂_i is linear. Thus, the stochastic partial differential operator corresponds closely to the deterministic partial differential operator. The Ito one-parameter rule completely accounts for the difference between the stochastic and deterministic operators.

From these rules the two-parameter stochastic differentiation formula follows:

$$\begin{aligned}(3.10)\quad \partial_2\partial_1 F(M_z) &= F'(M_z)\partial_2\partial_1 M_z + F''(M_z)\partial_2 M_z\partial_1 M_z \\ &+ \tfrac{1}{2}F'''(M_z)V_1(z)\partial_2 M_z dz_1 + \tfrac{1}{2}F'''(M_z)V_2(z)\partial_1 M_z dz_2 \\ &+ [F''(M_z)V_{12}(z) + \tfrac{1}{4}F''''(M_z)V_1(z)V_2(z)]dz_1 dz_2.\end{aligned}$$

The deterministic differentiation formula corresponds to setting V identically equal to zero in (3.10). The terms which depend on V are the so-called correction terms needed for the stochastic case.

Generally speaking, the differential equations we write always involve mixed differentials - we never take a derivative twice with respect to the same varaible. Thus, in (3.10) we have only the differentials $\partial_2\partial_1 M_z$, $\partial_2 M_z \partial_1 M_z$ and $\partial_2 M_z dz_1$, but not $\partial_2\partial_2 M_z$ or $\partial_1 M_z \partial_1 M_z$. Therefore, we can always obtain the integral equation corresponding to a stochastic differential equation by integrating with respect to the differentiated variables in the original differential equation. For example, the integral version of (3.10) is

$$
\begin{aligned}
(3.11)\qquad & F(M_{x_2,y_2}) - F(M_{x_1,y_2}) - F(M_{x_2,y_1}) + F(M_{x_1,y_1}) \\
&= \int F'(M)\partial_2\partial_1 M_\xi + \int F''(M_\xi)\partial_2 M\, \partial_1 M_\xi \\
&+ \int \tfrac{1}{2}F'''(M_\xi)V_1(\xi)\partial_2 M_\xi d\xi_1 \\
&+ \int \tfrac{1}{2}F'''(M_\xi)V_2(\xi)\partial_1 M_\xi d\xi_2 \\
&+ \int [\tfrac{1}{4}F''''(M_\xi)V_1(\xi)V_2(\xi) + F''(M_\xi)V_{12}(\xi)]d\xi_1 d\xi_2
\end{aligned}
$$

where all integrables are taken over the rectangle $(x_1,x_2] \times (y_1,y_2]$ in $[0,1]^2$.

3.4 Two Dimensional Integrals

We need to define five types of integrals over the square $[0,1]^2$. Let us refer to these integrals as stochastic integrals, mixed stochastic-deterministic integrals and deterministic integrals as follows. The stochastic integrals are written

$$\int \phi_\xi \partial_1 \partial_2 M_\xi \tag{3.12}$$

and

$$\int \phi_\xi \partial_1 M_\xi \partial_2 M_\xi. \tag{3.13}$$

The mixed integrals are written

$$\int \phi_\xi \partial_1 M_\xi d\xi_2 \tag{3.14}$$

and

$$\int \phi_\xi \partial_2 M_\xi d\xi_1. \tag{3.15}$$

The deterministic integral is

$$\int \phi_\xi d\xi_1 d\xi_2 = \int \phi_\xi d\xi. \tag{3.16}$$

In this section we consider these integrals briefly, stating

their main properties but leaving details and proofs for chapters 5, 6 and 7. We define the stochastic integrals of (3.12) and (3.13) by means of forward-differences just as (3.6) is defined by the forward difference, $M_{\tau^+,z_2,\ldots,z_n} - M_{\tau,z_2,\ldots,z_n}$.* Define the forward difference operators Δ_i as follows. Suppose $z \to F_z$ is a real-valued function of z in $[0,1]^n$ and suppose P_i is a partition of [0,12. Then define the forward difference operator, Δ_i, with respect to P_i by

$$(3.17)\quad \Delta_i F_z = F_{z_1,\ldots,z_{i-1},\tau^+,z_{i+1},\ldots,z_n} - F_{z_1,\ldots,z_{i-1},\tau,z_{i+1},\ldots,z_n}$$

for all τ in P_i.* Thus, $\Delta_i F_z$ is defined as long as z_i, the i^{th} coordinate of z, is in P_i.

Now let us consider the stochastic integral (3.12). For a partition process ϕ defined on a partition $P=P_1 \times P_2$ of $[0,1]^2$, define

$$(3.18)\qquad \int_P \phi_\xi \partial_1 \partial_2 M_\xi = \sum_{z\in P} \phi_z \Delta_1 \Delta_2 M_z$$

where Δ_1 and Δ_2 are defined with respect to P_1 and P_2 respectively. It is easy to see that $\Delta_1\Delta_2 M_z = \int_{\delta z} p(\xi)\Phi_0(d\xi)$ for z in P, and hence, (3.18)

*See section 2.4 for the definition of τ^+ where τ is in P_i.

defines the Ito-type integral of section 2.5. That is,

$$\int \phi_\xi \partial_1 \partial_2 M_\xi = \int \phi_\xi p(\xi) \Phi_0(d\xi) \tag{3.19}$$

for all ϕ in $L_a^2(\Omega x[0,1]^2,m)$.

The second integral (3.13) is a new type of integral and is more difficult to define. For a partition process ϕ defined on a partition $P=P_1 \times P_2$ of $[0,1]^2$, define

$$\int_P \phi_\xi \partial_2 M_\xi \partial_1 M_\xi = \sum_{z\in P} \phi_z \Delta_2 M_z \Delta_1 M_z. \tag{3.20}$$

The right side of (3.20) depends on the partition P as well as the partition process ϕ, and it will differ for the same ϕ if we choose different partitions P. However, we prove in chapter 6 that if ϕ is a partition process defined on the partitions P^n of $[0,1]^2$ for each natural number n and if $|P^n| \to 0$ as $n \to \infty$, then $\int_{P^n} \phi_\xi \partial_2 M_\xi \partial_1 M_\xi$ converges in $L^2(\Omega)$. The limit depends only on the partition process ϕ and not on the sequence of partitions. Denote the limit by $\int \phi \partial_2 M \partial_1 M$. Then, as we prove in section 6, $\phi \to \int \phi \partial_2 M \partial_1 M$ is a continuous linear mapping which extends to all of $L_a^2(\Omega x[0,1]^2,m)$. Note that there is only one integral corresponding to $\partial_1 \partial_2 M$ and $\partial_2 \partial_1 M$, that is,

$$\int \phi \partial_1 \partial_2 M = \int \phi \partial_2 \partial_1 M$$

Similarly, we have that

$$\int \phi \partial_1 M \partial_2 M = \int \phi \partial_2 M \partial_1 M$$

Thus, the differentials $\partial_1 \partial_2 M$ and $\partial_2 \partial_1 M$, and $\partial_2 M \partial_1 M$ and $\partial_1 M \partial_2 M$ are respectively equal.

Wong and Zakai (1974, 1976) define a multiple Wiener integral

(3.21) $$\int \psi_{\xi,\xi'} dw_\xi dw_{\xi'}$$

where ξ and ξ' <u>both</u> range over the plane square $[0,1]^2$. In their differentiation formula, $\psi_{\xi,\xi'}$ has the form $\psi_{\xi\ \xi'} = \phi_{\xi v \xi'}$ where $\xi v \xi'$ is the least upper bound (with respect to the partial order $<$, of ξ and ξ' in $[0,1]^2$). After developing precisely the stochastic integrals in chapters 5, 6 and 7 we prove in appendix B that

(3.22) $$\int \phi_{\xi v \xi'} dw_\xi dw_{\xi'} = \int \phi_\xi \partial_1 w_\xi \partial_2 w_\xi .$$

Cairoli and Walsh (1975) define the same multiple Wiener integral but the also show that

$$\int \phi_{\xi v \xi'} dw_{\xi} dw_{\xi'} = \int \phi_{\xi} dJ_{w}(\xi)$$

where J_w is a measure corresponding to $\partial_1 w \partial_2 w$. The product measure $\partial_1 w \partial_2 w$ is the same there as it is in this paper. See chapter 6 for a general discussion of such product measures and see appendix B for further discussion of the Cairoli-Walsh integral and their differentiation formula.

The stochastic integrals have some interesting properties which we present now but whose proof we defer until chapters 5 and 6. With V defined as in (3.4), one has that

$$(3.23) \qquad E\{(\int \phi \partial_2 \partial_1 M)^2\} = \int E\{\phi^2\} \partial_2 \partial_1 V$$

where $\partial_2 \partial_1 V = V_{12}(\xi) d\xi$. Similarly, it is true that

$$(3.24) \qquad E\{(\int \phi \partial_2 M \partial_1 M^2\} = \int E\{\phi^2\} \partial_2 V \partial_1 V$$

The stochastic integrals have a martingale property similar to that described for the Ito-type integral in chapter 2.* Thus, we have for ϕ in $L_a^2(\Omega x [0,1]^2, m)$ that the process N defined by either

*See Wong-Zakai (1974, 1976) and Cairoli-Walsh (1975) and chapter 5.

(3.25) $$N_z = \int \chi_{[0,z]} \phi \partial_2 \partial_1 M$$

or

(3.26) $$M_z = \int \chi_{[0,z]} \phi \partial_2 M \partial_1 M$$

is a martingale on $[0,1]^2$. Finally, the two stochastic integrals on $[0,1]^2$ are orthogonal in the sense that

(3.27) $$E\{(\int \phi \partial_2 \partial_1 M)(\int \psi \partial_2 M \partial_1 M)\} = 0$$

for every ϕ and ψ in $L_a^2(\Omega x [0,1]^2, m)$.†

The mixed stochastic-deterministic integrals are defined as double integrals. If ϕ is in $L_a^2(\Omega x [0,1]^2, m)$, we prove later that

$$z_2 \to \int \phi_{\tau, z_2} \partial_1 M_{\tau, z_2}$$

is almost-surely Lebesgue integrable. Thus, we define (3.14) as

(3.28) $$\int \phi_\xi \partial_1 M_\xi d\xi_2 = \int \{\int \phi_{\tau,\sigma} \partial_1 M_{\tau,\sigma}\} d\sigma.$$

Then we can prove the inequality

---

†See Wong and Zakai (1974) and Cairoli and Walsh (1975).

$$(3.29) \qquad E\{(\int \phi_\xi \partial_1 M_\xi d\xi_2)^2\} \leq \int E\{\phi_\xi^2\}\partial_1 V(\xi) d\xi_2.$$

The other mixed integral (3.15) is defined similarly and satisfies an inequality analogous to (3.29).

This definition of mixed integral completely agrees with Cairoli and Walsh's definition of mixed integral. Wong and Zakai, however, employ mixed multiple integrals, written

$$\int \psi_{\xi,\xi'}\, d\xi dW_{\xi'}$$

and

$$\int \psi_{\xi,\xi'}\, dW_\xi d\xi'.$$

However, these mixed integrals are equivalent to ours if $\psi_{\xi,\xi'} = \phi_{\xi v\xi'}$, and we prove in appendix B that

$$\int \phi_{\xi v\xi'}\, d\xi dW_{\xi'} = \int \phi_\xi\ \xi_1 \partial_1 W_\xi d\xi_2$$

and

$$\int \phi_{\xi v\xi'}\, dW_\xi d\xi' = \int \phi_\xi \xi_2 \partial_2 W_\xi d\xi_1.$$

The deterministic integral is the Lebesgue integral on $[0,1]^2$, and it satisfies

(3.30) $$E\{(\int \phi_\xi d\xi)^2\} \leq \int E\{\phi_\xi^2\} d\xi,$$

for all ϕ in $L_a^2(\Omega x[0,1]^2, m)$.

Appendix B presents the Cairoli-Walsh and Wong-Zakai integrals in more detail, giving the precise definition of these integrals and proving their equivalence to the integrals defined in this paper. Once it is proved that the integrals are equivalent, it follows immediately that the differentiation formulas are also equivalent. We present the Wong-Zakai and Cairoli-Walsh differentiation formulas in appendix B also.

CHAPTER 4

MULTIPARAMETER INTEGRALS AND DIFFERENTIAL FORMULAS

4.1 Introduction

For three or more parameters the stochastic differential formula is too long to write down easily, but we can still apply the same differentiation rules of section 3.3 to calculate the stochastic differentiation formulas in the general multiparameter cases. In section 4.2 we introduce a notational scheme for denoting the various multiparameter differentials and their integrals. Section 4.3 states some properties of these integrals and section 4.4 describes how to interpret stochastic differential equations in terms of stochastic integral equations. In section 4.5 we state the main results of this paper, the general differentiation rules. We prove these results in the following chapters (5, 6, 7).

4.2 Multiparameter Differentials and Integrals

Let $z \to M_z$ be the multiparameter process on $[0,1]^n$ defined in (3.3). The general stochastic differential of M will have the form

$$(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_p} M)\cdot(\partial_{j_1}\partial_{j_2}\cdots\partial_{j_q} M)\cdot(\partial_{k_1}\partial_{k_2}\cdots\partial_{k_r} M)\ \cdots$$

where $i_1, i_2,\ldots,i_p, j_1, j_2,\ldots,j_q, k_1, k_2,\ldots,k_r,$ are

all distinct elements of $\{1,2,\ldots,n\}$[†]. Since the order of differentiation does not matter, as we show later, we can write

$$\partial_{i_1}\partial_{i_2}\cdots\partial_{i_p} M = \partial_S M$$

where $S = \{i_1,i_2,\ldots,i_p\}$. Similarly, the order of the factors $\partial_S M$ does not matter and we can compactly write the general differential as

$$\partial_{S_1} M\cdot\partial_{S_2} M\cdot\ldots\partial_{S_k} M = \partial_\Lambda M$$

where $\Lambda = \{S_1,S_2,\ldots,S_k\}$ and $S_1,S_2,\ldots,S_k$ are mutually disjoint subsets of $\{1,2,\ldots,n\}$. Let $\bigcup\Lambda$ denote $\bigcup_{i=1}^{k} S_i$. If S is a subset of $\{1,2,\ldots,n\}$ and $(\bigcup\Lambda)\bigcap S = \phi$, then we write the mixed differential

$$\partial_\Lambda M_z dz_S = \partial_\Lambda M_z dz_{i_1} dz_{i_2}\cdots dz_{i_q}$$

where $S = \{i_1,i_2,\ldots,i_q\}$. Whenever we write the differential $\partial_\Lambda M dz_S$, we assume that $(\bigcup\Lambda)\bigcap S = \phi$.

For each differential $\partial_\Lambda M_z dz_S$ there corresponds

---

[†]We only take mixed derivatives as noted in section 3.3.

an integral which we write

$$\int \phi_z \partial_\Lambda M_z dz_S$$

The integration is carried out over the variables z_i in $[0,1]$ for i in $(\cup\Lambda)\cup S$.

4.3 Properties of Higher Dimensional Integrals

First let us give a rough definition of the higher dimensional stochastic integrals in terms of forward difference operators. Define the finite difference operator Δ_Λ in the same way as ∂_Λ. Thus, if Δ_i is the forward difference operator with respect to partition P_i of $[0,1]$ (see sections 2.4 and 3.4), then define

$$\Delta_S M_z = \Delta_{i_1} \Delta_{i_2} \cdots \Delta_{i_p} M_z$$

for $S = \{i_1, i_2 \ldots i_p\} \subseteq \{1,2 \ldots n\}$. Similarly, if $\Lambda = \{S_1, S_2 \ldots S_k\}$ and $S_1, S_2 \ldots S_k$ are mutually disjoint subsets of $\{1,2,\ldots,n\}$, then

$$\Delta_\Lambda M_z = \Delta_{S_1} M_z \cdot \Delta_{S_2} M_z \cdot \ldots \cdot \Delta_{S_k} M_z .$$

Suppose $P = P_1 \times P_2 \times \ldots \times P_n$ is a partition of $[0,1]^n$

and ϕ is a partition process defined on P. Then define

$$(4.1) \qquad \int_P \phi_\xi \partial_\Lambda M_\xi = \sum_{z \in P} \phi_z \Delta_\Lambda M_z.$$

As $|P| \to 0$, this expression tends to a limit in mean square. We define the stochastic integral $\int \phi \partial_\Lambda M$ to be this limit. Note that $\int \phi \partial_\Lambda M$ is a process over the undifferentiated variables, i.e. those variables not integrated.

As in the two dimensional case, we have

$$(4.2) \qquad E\{(\int \phi_\xi \partial_\Lambda M_\xi)^2\} = \int E\{\phi_\xi^2\} \partial_\Lambda V(\xi)$$

where V is defined in (3.4) and ϕ is in $L_a^2(\Omega x[0,1]^n, m)$. The higher dimensional integrals also have the martingale and orthogonality properties of the plane integrals in chapter 3. If ϕ is in $L_a^2(\Omega, [0,1]^n, m)$ and $\Lambda = \{1,2,\ldots,n\}$, then

$$(4.3) \qquad M_z = \int \chi_{[0,z]}(\xi) \phi_\xi \partial_\Lambda M_\xi$$

defines a martingale on $[0,1]^n$.

Suppose ϕ and ψ are in $L_a^2(\Omega x[0,1]^n, m)$ and $\bigcup \Lambda_1 = \bigcup \Lambda_2$ but $\Lambda_1 \neq \Lambda_2$. Then

$$(4.4) \qquad E\{(\int \phi_\xi \partial_{\Lambda_1} M_\xi)(\int \psi_\xi \partial_{\Lambda_2} M_\xi)\} = 0.$$

That is, different stochastic integrals are orthogonal.

We define the mixed stochastic-deterministic integrals in the same way as in chapter 3. Suppose S is a subset of $\{1,2,\ldots,n\}$ disjoint from Λ. Then $\int \phi_\xi \partial_\Lambda M_\xi$ is almost surely integrable with respect to dz_S and we define the mixed integral by

$$(4.5) \qquad \int \phi_\xi \partial_\Lambda d\xi_S = \int\{\int \phi_\xi \partial_\Lambda M_\xi\} d\xi_S.$$

For all ϕ in L_a^2 we have the inequality

$$(4.6) \qquad E\{(\phi_\xi \partial_\Lambda M_\xi d\xi_S)^2\} \leq \int E\{\phi_\xi^2\} \partial_\Lambda V(\xi) d\xi_S.$$

4.4 Stochastic Differential Equations

We wish to interpret equations of the form

$$(4.7) \qquad \partial_S(F_z \partial_{\Lambda_0} M dz_{S_0}) = \sum_{\nu=1}^{N} \phi_z^\nu \partial_{\Lambda_\nu} M dz_{S_\nu}$$

where for each $\nu = 1,2,\ldots,N$, $(\cup\Lambda_\nu)\cup S_\nu = S \cup (\cup\Lambda_0)\cup S_0$. We assume that $S \cap (\cup\Lambda_0 \cup S_0) = \phi$ and that $\cup\Lambda_\nu \cap S_\nu = \phi$ for each $\nu = 0,1,2,\ldots,N$. The precise interpretation of (4.7) in terms of an integral equation is notationally

complex, but the basic idea is simple. We define the differential equation so that if we integrate it over any rectangle, the resulting integral equation will be true.

Let $S = \{i_1, i_2, \ldots, i_q\}$ and $(\bigcup \Lambda_0) \bigcup S_0 = \{i_{q+1}, \ldots, i_{q+p}\}$. Let P_i be a partition of $[0,1]$ for $i = 1,2,\ldots,q+p$. Then we interpret (4.7) so that for all $z \varepsilon P_1 \times \ldots \times P_q$ and for all $\bar{z} \varepsilon P_{q+1} \times \ldots \times P_{q+p}$ the following integral equation is true.

$$(4.8) \qquad \Delta_s \int_{\delta \bar{z}} F_\xi \partial_{\Lambda_0} M d\xi_{S_0} = \sum_{v=1}^{N} \int_{\delta z \times \delta \bar{z}} \phi_\xi^v \partial_{\Lambda_v} M d\xi_{S_v}.$$

We assume on the left side of (4.8) that $\xi_{i_j} = z_j$, for $1 \leq j \leq q$, and that Δ_S operates with respect to $P_1 \times P_2 \times \ldots \times P_q$. The integration over $\delta \bar{z}$ or $\delta z \times \delta \bar{z}$ in (4.8) indicates that ξ_{i_j} ranges over $(z_j, z_j+]$ for $1 \leq g \leq q$ and over $(\bar{z}_j, \bar{a}_j+]$ for $q+1 \leq q+p$ in the integration.

4.5 Differentiation Rules for Higher Dimensions

In chapter 7 we prove two theorems that generalize the differential rules of Section 3.3. We state these theorems now.

Theorem 4.1 If $S_1, S_2, \ldots, S_q$ are mutually disjoint

subsets of $\{1,2,\ldots,n\}$ and $i \notin S_1 \cup S_2 \cup \ldots \cup S_q$, then

$$(4.9) \qquad \partial_i(\partial_{S_1} M \ldots \partial_{S_q} M) = \sum_{j=1}^{q} \partial_{S_j \cup \{i\}} M \cdot \prod_{k \neq j} \partial_{S_k} M.$$

Theorem 4.2 If $\partial_i f_z = \phi_z \partial_i M + \psi_z dz_i$ and $i \notin \cup\Lambda \cup S$, then

$$(4.10) \quad \partial_i(f_z \partial_\Lambda M dz_S) = \phi_z(\partial_i M \partial_\Lambda M) dz_S + \psi_z \partial_\Lambda M(dz_S dz_i) + f_z \partial_i(\partial_\Lambda M) dz_S,$$

where $\partial_i(\partial_\Lambda M)$ is interpreted as in the right side of (4.9).

Note that (4.9) and (4.10) agree with the ordinary deterministic rule for differentiating products. By applying Ito's one dimensional rule to Theorem 4.2 we obtain stochastic differentiation formulas as follows. Suppose $f(u,z)$ has continuous partial derivatives in u and z up to second order where u is in $[0,1]$ and z is in $[0,1]^n$. Then for $i \notin \cup\Lambda \cup S$, Ito's rule and Theorem 4.2 imply that

$$(4.11) \qquad \partial_i(f(M_z,z) \partial_\Lambda M dz_S) = f_u(M_z,z) \partial_i M \partial_\Lambda M dz_S$$

$$+ [f_i(m_z,z) + \tfrac{1}{2} f_{uu}(M_z,z) V_i(z)] \partial_\Lambda M dz_S dz_i$$

$$+ f(M_z,z) \partial_i(\partial_\Lambda M) dz_S.$$

One can derive the n-dimensional stochastic differentiation formula by using (4.11) to write the mixed partial derivative

$$\partial_1\partial_2\cdots\partial_n\ [f(M_z,z)]$$

in terms of the differentials $\partial_\Lambda M dz_S$.

Note that for $n = 3$ there are 5 pure stochastic differentials and 9 mixed stochastic differentials. Thus, including the one deterministic differential there are 15 differentials. For $n = 4$ there are 52 differentials and for $n = 5$ there are 203 differentials. Consequently, even for three dimensions the stochastic differentiation formula is too cumbersome to write easily, and the formulas grow even longer for dimensions higher than three.

CHAPTER 5

STOCHASTIC MEASURES AND THEIR INTEGRALS

5.1 Introduction

In this section we present stochastic integrals from a general and abstract point of view. This abstraction is necessary in order to prove the assertions of chapter 4 for all the various stochastic integrals. Hopefully, it makes the stochastic integrals of chapter 4 easier to understand by emphasizing properties common to all stochastic integrals. The geometric interpretation of stochastic integrals in terms of measures will also make it easier to visualize the properties of stochastic integrals. We include several figures for the two-dimensional case to aid in this visualization.

In section 5.2 we define the stochastic measure and discuss its relation to the stochastic differentials of chapter 4. Section 5.3 shows that each stochastic measure generates a stochastic integral that obeys the martingale property of 4.2. We also discuss the converse problem of constructing stochastic measures from stochastic integrals with the martingale property. In 5.4 we define orthogonal stochastic measures and prove that their integrals are also orthogonal. This section prepares for a later proof of (4.4) in section 6.4.

5.2 Definition of Stochastic Measure and Examples

Define Φ to be a stochastic measure on the unit cube I in R^n if the following conditions are true.

(1) Let A denote the class of all finite unions of rectangles of the form $(z,z^+]$ in I. Then for each A in A, $\Phi(A)$ is a real-valued random variable of (P,Ω,F). Furthermore, $\Phi(A)$ has zero mean and a finite variance satisfying

$$E\{\Phi(A)^2\} \leq \ell(A) \tag{5.1}$$

where ℓ is a finite, nonnegative Borel measure on I.

(2) If $A_1, A_2, \ldots, A_N$ are mutually disjoint elements of A, then

$$\Phi\Big(\bigcup_{i=1}^{N} A_i\Big) = \sum_{i=1}^{N} \Phi(A_i).$$

(3) For all z and z^+ in I such that $z < z^+$, $\Phi((z,z^+])$ is F_{z^+}-measurable and independent of F_z. Furthermore, if $\bar{z} \not> z$, then

$$E\{\Phi((z,z^+])\,|\,F_{\bar{z}}\} = 0. \tag{5.2}$$

Here $\{F_z\}_{z\in I}$ is some prespecified increasing family of

σ-fields as in chapter 2.

Conditions (1) and (2) of the definition are clear. Equation (5.2) says that $\Phi((z,z^+])$ and $\mathcal{F}_{\overline{z}}$ are uncorrelated. That is, if ψ is an $\mathcal{F}_{\overline{z}}$-measurable bounded random variable, then $E\{\psi\Phi((z,z^+])\} = 0$. Note that $\overline{z} \not> z$ is equivalent to $[0,\overline{z}] \quad (z,z^+] \neq 0$. The latter statement has a clear geometrical interpretation as shown in figure 5.1.

Define the σ-field $\mathcal{F}_{i,z}$ as $\mathcal{F}_{z(i)}$, where $z(i)$ is the element of I with all coordinates equal to 1 except for the i^{th} coordinate which is the same as the i^{th} coordinate of z. For each z in $I = [0,1]^n$ there are n such σ-fields. From figure 5.1 it should be clear that requiring (5.2) is equivalent to requiring

$$E\{\Phi((z,z^+]) \,|\, \mathcal{F}_{i,z}\} = 0 \tag{5.3}$$

for each $i = 1,2,\ldots,n$.

The Gaussian random measure of chapter 2 is a stochastic measure as defined here, but the notion of stochastic measure is much broader. For example, each stochastic integral of 4.2 defines a stochastic measure Φ_Λ as follows. If $\Lambda = \{1,2,\ldots,n\}$, define the measure Φ_Λ as

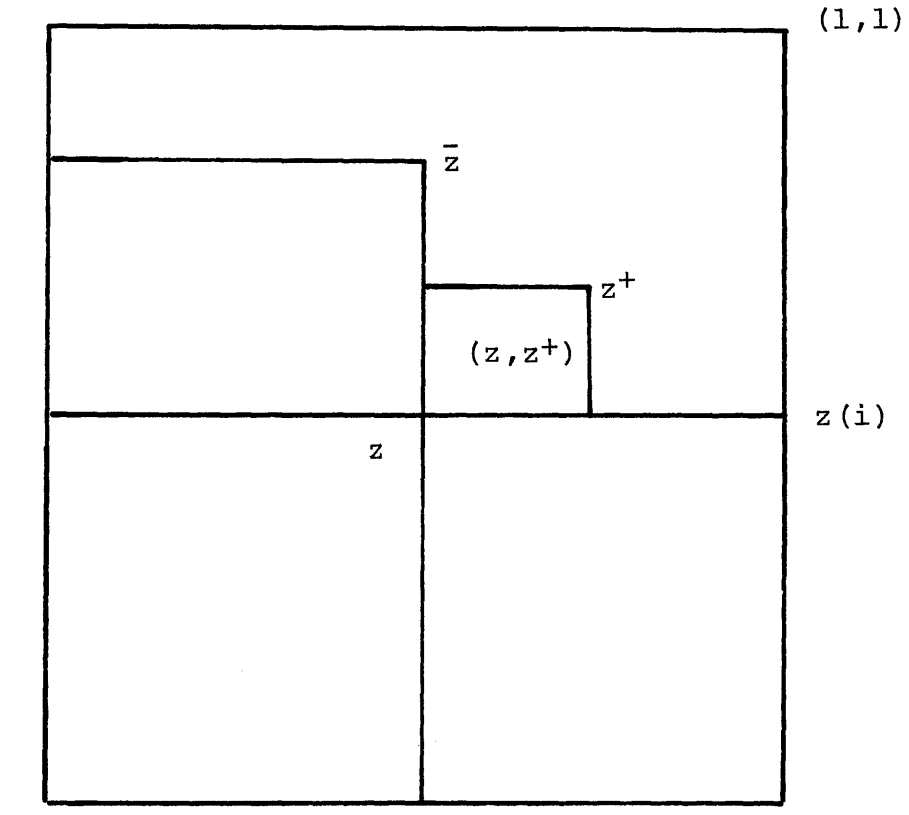


Figure 5.1

(5.4) $$\Phi_\Lambda(A) = \int \chi_A(\xi)\partial_\Lambda M_\xi .$$

The stochastic measure Φ_Λ is a Gaussian random measure in only case, namely in the case

$$\partial_{\Lambda_0} M_\xi = \partial_1\partial_2\cdots\partial_n M_\xi .$$

In the other cases Φ_Λ does not have the independent increment character of a Gaussian random measure. That is, for the Gaussian random measure Φ_{Λ_0}, we have that $\Phi_{\Lambda_0}((z,z^+])$ is actually independent of $F_{\overline{z}}$ in (5.2), but for the other Φ_Λ we do not have independence (although (5.2) is still true).

Our approach to stochastic integration will be to construct a stochastic measure first and then construct a stochastic integral from that. Thus, in chapter 6 we first define stochastic measure Φ_Λ and then define $\int \phi\partial_\Lambda M$ as the integral of Φ_Λ. In preparation for this development, we now prove that each stochastic measure defines a unique integral.

5.3 Stochastic Integrals of Stochastic Measures

Theorem 1

Suppose Φ is a stochastic measure on I as defined

in Definition 5.1. Then for each ϕ in $L_a^2(\Omega\times I,\ell)$ there is a random variable $\int\phi_\xi\Phi(d\xi)$ of $L^2(\Omega)$ such that the following are true.

(1) The mapping $\phi \to \int\phi_\xi\Phi(d\xi)$ is a continuous linear mapping from L_a^2 to $L^2(\Omega)$, and we have

$$E\{(\int\phi_\xi\Phi(d\xi))^2\} \leq \int E\{\phi_\xi^2\}\ell(d\xi). \tag{5.5}$$

(2) If z and z^+ are in I and ϕ_0 is F_z-measurable random variable, then

$$\int\phi_0\chi_{(z,z^+]}(\xi)\Phi(d\xi) = \phi_0\Phi((z,z^+]).$$

(3) For each Borel set A I, define m(A) by

$$m(A) = E\{(\int\chi_A(\xi)\Phi(d\xi))^2\}.$$

Then m is a finite, nonnegative Borel measure on I. Furthermore, we have for all $\phi\varepsilon L_a^2$ that

$$E\{(\int\phi_\xi\Phi(d\xi))^2\} = \int E\{\phi_\xi^2\}m(d\xi). \tag{5.6}$$

(4) For each ϕ in L^2, we have

$$E\{\int \phi_\xi \Phi(d\xi)\} = 0.$$

For each z in I, we have

$$(5.7) \qquad E\{\int \phi_\xi \Phi(d\xi) \mid F_z\} = \int \chi_{[0,z]}(\xi)\phi_\xi \Phi(d\xi),$$

and the process M, defined by

$$(5.8) \qquad M_z = \int \chi_{[0,z]}(\xi)\phi_\xi \Phi(d\xi),$$

is a martingale on I.

Note that equation (5.6) is the general form of (4.2) and (4) is the martingale property of (4.3)

Proof of Theorem 1.

The idea of the proof is simple. Prove everything for ϕ in L^2_{pp} and extend to L^2_a by taking limits.

For a partition process ϕ with respect to the partition P of I, define

$$(5.9) \qquad \int \phi_\xi \Phi(d\xi) = \sum_{z \in P} \phi_z \Phi(\delta z).$$

The right side of (5.9) depends only on ϕ and not on the particular partition P. Thus, $\phi \to \int \phi_\xi \Phi(d\xi)$ is a well-defined linear mapping.

For any partition on I, if $z_1 \neq z_2$, and z_1 and z_2 are in P, then either $[0,z_2^+]\ \delta z_1 = \phi$ or $[0,z_1^+]\ \delta z_2 = \phi$ (see figure 5.2). Since ϕ_z is F_z-measurable for each z in P, (5.2) implies that

$$E\{\phi_{z_1}\Phi(\delta z_1)\phi_{z_2}\Phi(\delta z_2)\} = 0$$

unless $z_1 = z_2$. Hence we find that

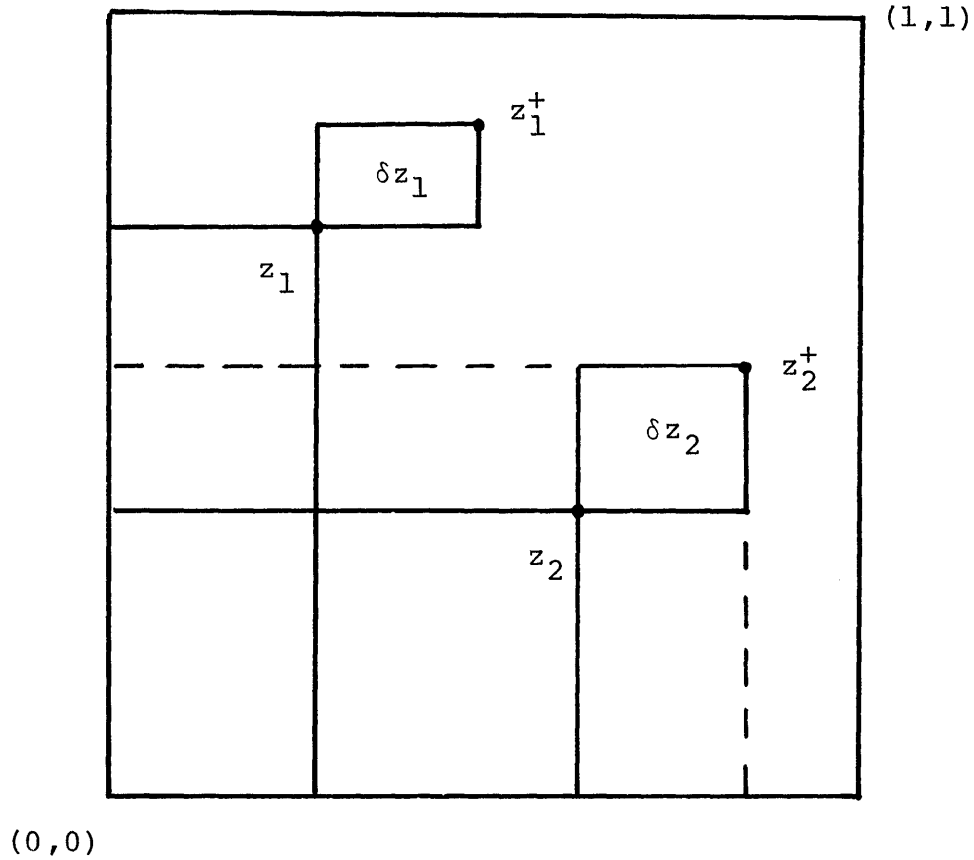
$$(5.10)\quad E\{(\int \phi_\xi \Phi(d\xi))^2\} = \sum_{z\in P} E\{\phi_z^2\}E\{\Phi(\delta z)^2\}.$$

The inequality (5.5) follows immediately from (5.10) and the inequality (5.1) of Definition 5.1. We then extend the integral to L_a^2 in the usual way, preserving (5.5). Property (2) is obvious from the definition of the stochastic integral.

To show m defined in (3) is additive we need to show

$$(5.11)\qquad E\{(\int \chi_A(\xi)\Phi(d\xi))(\int \chi_B(\xi')\Phi(d\xi'))\} = 0$$

whenever A and B are disjoint Borel sets in I. This is certainly true if A and B are in $\mathcal{A}$, and the general case follows by taking limits. Using (5.11) and the continuity



$$[0, z_2^+] \cap \delta z_1 = \emptyset$$

Figure 5.2

and additivity of the stochastic integral, we can easily show

$$m\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} m(A_n)$$

for disjoint Borel sets A_n in I. Furthermore, from (5.10) we see that (5.6) is true for ϕ in L^2_{pp}, and it remains true for limits of partition processes. Thus, (5.6) is true for all of L^2_a, and (3) is proved.

Note that (4) is true for a partition process of the form,

$$\phi_\xi = \phi_{z_1} \chi_{(z_1, z_1^+]}(\xi)$$

where ϕ_{z_1} is F_{z_1}-measurable and $z_1 < z_1^+$. That is, we first have

$$(5.12) \qquad E\{\phi_{z_1} \Phi((z_1, z_1^+]) \mid F_z\} = \phi_{z_1} \Phi((z_1, z_1^+] \cap [0,z]).$$

This follows from the fact that we can write the set $(z_1, z_1^+] - ((z_1, z_1^+] \cap [0,z])$ as a finite disjoint union, $\bigcup_{i=1}^{N} A_i$, of sets A_i where each A_i is a rectangle of the form $(\xi, \xi^+]$ for ξ and ξ^+ in I and $\xi < \xi^+$.

(see figure 5.3). Thus, additivity of Φ implies

$$\Phi((z_1,z_1^+]) = \Phi((z_1,z_1^+] \cap [0,z]) + \sum_{i=1}^{N} \Phi(A_i).$$

From condition (3) of Definition 5.1 we deduce that

$$E\{\phi_{z_1} \Phi(A_i) | F_z\} = 0$$

for each $i = 1,\ldots,N$, and

$$E\{\phi_{z_1} \Phi((z_1,z_1^+] \cap [0,z]) | F_z) = \phi_{z_1} \Phi((z_1,z_1^+] \cap [0,z])$$

This proves (5.12). The right side of (5.12) is just $\int \phi_\xi \chi_{[0,z]}(\xi) \Phi(d\xi)$ so that (5.7) is true for processes of this form. The extension to all L^2_{pp} and then to L^2_a is clear. This concludes the proof of (4) and the theorem.///

Although we have treated only stochastic measures over the unit cube in R^n, it is straightforward to extend the definition and theorem to stochastic measures over all of R^n. We can also relax our assumption that ℓ is a finite measure and assume only that it is a nonnegative Borel measure.

We digress at this point in order to show that one

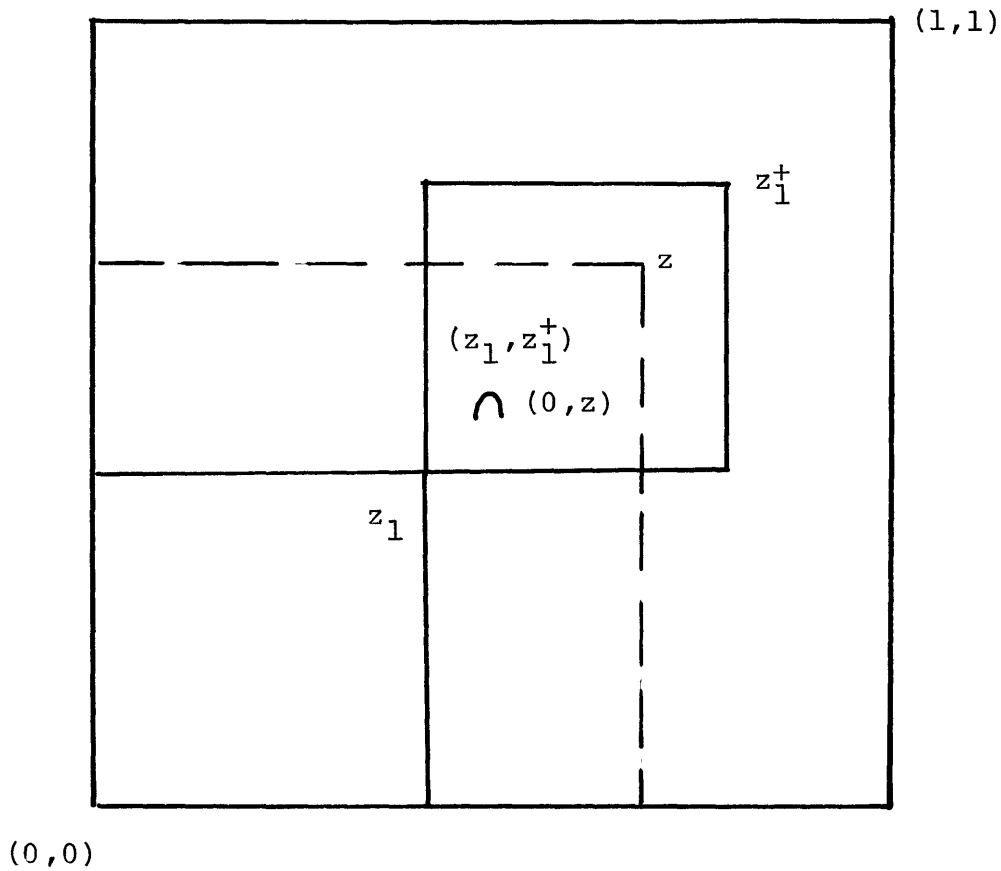


Figure 5.3

can construct a stochastic measure from any mapping $\phi \to J(\phi)$ satisfying the general properties of a stochastic integral. Thus, we see in what sense the definition of stochastic measure given in 5.2 is necessary.

Suppose J is a continuous linear mapping from $L_a^2(\Omega\times I,\ell)$ to $L^2(\Omega)$ and suppose J also has the following properties.

(i) For all ϕ in $L_a^2(\Omega\times I,\ell)$

$$E\{J(\phi)^2\} \leq \int E\{\phi_\xi^2\}\ell(d\xi).$$

(ii) If ϕ_0 is an F_z-measurable random variable and $z < z^+$, then

$$J(\phi_0\chi_{(z,z^+]}) = \phi_0 J(\chi_{(z,z^+]}).$$

(iii) The process M defined by

$$M_z = J(\chi_{[0,z]}\phi)$$

is a martingale on I for each choice of ϕ in L_a^2. Thus, J has the same properties (1), (2) and (4) as the stochastic integral in Theorem 5.1. If we define Φ by

$$\Phi(A) = J(\chi_A)$$

then Φ has all but one property of a stochastic measure. Conditions (1) and (2) of Definition 5.1 follow from (i) and the linearity of J. Property (iii) implies (5.2) in condition (3) of Definition 5.1. However, the other part of (3) is not necessarily true. It is true that $J(\chi_{(z,z^+]})$ is $\mathcal{F}_{z^+}$ -measurable (this follows from (iii)), but it is not necessarily true that $J(\chi_{(z,z^+]})$ is independent of $\mathcal{F}_z$. We need this independence condition, however, only to prove

$$E\{\phi_0^2\Phi((z,z^+])^2\} = E(\phi_0^2)E\{\Phi((z,z^+])^2\}$$

and thus, prove (5.6) in Theorem 5.1. Since J satisfies (i) above, we know

$$E\{\phi_0^2\Phi((z,z^+])^2\} \leq E\{\phi_0^2\}\ell((z,z^+]).$$

This inequality is sufficient, together with the other properties above, to define a stochastic integral, $\int\phi_\xi\Phi(d\xi)$, from Φ. Thus, we can define a stochastic integral, without the independence assumption, but then we do not have (5.6).

Condition (ii) guarantees that the integral J induces a measure Φ which in turn gives back J. That is,

$$J(\phi) = \int \phi_\xi \Phi(d\xi).$$

5.4 Orthogonal Stochastic Measures and Integrals

The following result generalizes the concept of orthogonal stochastic integrals as discussed in sections 3.4 and 4.3.

Definition 2

Two stochastic measures, Φ and Ψ, are orthogonal if for each z and z^+ in I with $z < z^+$ we have

$$E\{\Phi((z,z^+])\Psi((z,z^+])\} = 0.$$

Likewise if J is a stochastic integral on $L_a^2(\Omega\times I,\ell_1)$ and K is a stochastic integral on $L_a^2(\Omega\times I,\ell_2)$, then J and K are orthogonal if

$$E\{J(\phi)K(\Psi)\} = 0$$

for all $\phi\varepsilon L_a^2(\Omega\times I,\ell_1)$ and $\Psi\varepsilon L_a^2(\Omega\times I,\ell_2)$. With this definition we have the following easy theorem.

Theorem 2

If two stochastic measures are orthogonal, then their corresponding stochastic integrals are orthogonal.

CHAPTER 6

PARAMETERIZED STOCHASTIC MEASURES AND THEIR PRODUCTS

6.1 Introduction

The purpose of this chapter is to define the stochastic measures Φ_Λ in (5.4) and thus, to define properly the stochastic integrals introduced in chapter 4. In the process of defining Φ_Λ we give a rigorous meaning to products of the form

$$\partial_\Lambda M = \partial_{S_1} M \partial_{S_2} M \ldots \partial_{S_q} M$$

introduced in section 4.2. This will be important in chapter 7 when we study derivatives of such products and derive the differentiation product rule. Before proceeding with the definition of parameterized stochastic measures, let us agree to denote the stochastic measure Φ by the expression $\Phi(dz)$ as well as Φ. This differential notation corresponds to the notation for the stochastic integral, namely $\int \phi_\xi \Phi(d\xi)$, but it also makes clear on what parameter space Φ is defined. A similar notation will be useful sometimes in denoting parameterized measures and their products.

In 6.2 we define the parameterized stochastic measure

and discuss some of its properties. In particular, we construct a mixed stochastic-deterministic integral from it. We also discuss examples of parameterized stochastic measures relevant to the purpose of justifying chapter 4. Section 6.3 defines the very important concept of the product of two or more parameterized stochastic measures. After proving some necessary general theorems about such products we turn in section 6.4 to constructing all the stochastic and mixed integrals of chapter 4. Finally, we verify the statements of section 4.3.

6.2 Definition of Parameterized Stochastic Measures and Mixed Integrals

Definition 1

Let I_1 and I_2 be two unit hypercubes. Then $\Phi(dz_1,z_2)$ is a stochastic measure on I_1, parameterized on I_2, if the following conditions are true.

(1) Variance condition. Let A denote the set of all finite unions of rectangles of the form $(z_1,z_1^+]$ in I_1. Then for A in A, $\Phi(A,z_2)$ is a real-valued random variable of (P,Ω,F). Furthermore, $\Phi(A,z_2)$ has zero mean and a finite variance satisfying

$$E\{\Phi(A,z_2)^2\} \leq \ell(A) \tag{6.1}$$

for all z_2 in I_2. Here ℓ is a finite, nonnegative Borel measure on I_1.

(2) Additivity condition. If $A_1, A_2, \ldots, A_N$ are mutually disjoint elements of A, then

$$\Phi\left(\bigcup_{i=1}^{N} A_i, z_2\right) = \sum_{i=1}^{N} \Phi(A_i, z_2)$$

for all z_2 in I_2.

(3) Measurability condition. Let $\{F_z\}_{z \in I_1 \times I_2}$ be an increasing family of σ-fields. We will write $F_z = F_{z_1, z_2}$ where $z = (z_1, z_2)$ and z_1 is in I_1 and z_2 is in I_2. For all z_1 and z_1^+ in I_1 such that $z_1 < z_1^+$, $\Phi((z_1, z_1^+], z_2)$ is $F_{z_1^+, z_2}$-measurable and independent of $F_{z_1, \bar{z}_2}$ for all $\bar{z}_2$ in I_2. Furthermore, if $\bar{z}_1 \not> z_1$, then

$$E\{\Phi((z_1, z_2^+], z_2) \mid F_{\bar{z}_1, \bar{z}_2}\} = 0 \tag{6.2}$$

for all $\bar{z}_2$ in I_2.

(4) Continuity condition. Let $0(\varepsilon)$ be a nonnegative, nondecreasing function of $\varepsilon \geq 0$ such that $0(\varepsilon) \to 0$ as $\varepsilon \to 0$. Define the norm $||z_2||$ on I_2 by

$$||z_2|| = \max_i |z_{2,i}|$$

where $z_{2,i}$ is the i^{th} coordinate of z_2 in I_2. For any z_1 and z_1^+ in I_1 with $z_1 < z_1^+$ and for any z_2 and z_2^+ in I_2 with $z_2 < z_2^+$, we assume that the following inequality is true.

$$(6.3)\quad E\{[\Phi((z_1,z_1^+],z_2^+) - \Phi((z_1,z_1^+],z_2)]^2\} \leq \ell((z_1,z_1^+])0(||z_2^+ - z_2||)$$

For each z_2 in I_2, $\Phi(dz_1,z_2)$ is a stochastic measure on I_1 with respect to the increasing family $\{F_{z_1,z_2}\}_{z_1\varepsilon I_1}$ of σ-fields. Thus, if ϕ is an adapted and measurable process on I_1 I_2 with respect to $\{F_z\}_{z\varepsilon I_1\times I_2}$, and if

$$(6.4)\quad \int_{I_1} E\{\phi^2_{\xi_1,z_2}\}\ell(d\xi_1) < \infty,$$

then Theorem 5.1 defines the stochastic integral

$$(6.5)\quad \int \phi_{\xi_1,z_2}\, \Phi(d\xi_1,z_2).$$

At this point it is very easy to define a mixed stochastic-deterministic integral. Suppose ϕ is in $L_a^2(\Omega\times I_1\times I_2,\ell\otimes m)$

where m is the Lebesgue measure on I_2. Then (6.5) is almost surely integrable with respect to m and we define the mixed integral by

$$(6.6) \qquad \int \phi_{\xi} \Phi(d\xi_1, \xi_2) d\xi_2 = \int_{I_2} [\int_{I_1} \phi_{\xi_1, \xi_2} \Phi(d\xi_1, \xi_2)] d\xi_2.$$

To see this note that except for an m-negligible set in I_2,

$$\int E\{\phi^2_{\xi_1, z_2}\} \ell(d\xi_1) < \infty$$

and hence, we know that the stochastic integral,

$$\int \phi_{\xi_1, z_2} \Phi(d\xi_1, z_2)$$

exists for almost all (with respect to m) z_2. Thus, the integral

$$(6.7) \qquad \int_{I_2} \left| \int_{I_1} \phi_{\xi_1, \xi_2} \Phi(d\xi_1, \xi_2) \right| d\xi_2$$

is well-defined, although it may be infinite. But in fact, it is finite as the following argument shows. From Schwarz's inequality in $L^2(\Omega)$ and from Theorem 5.1 it follows that

$$E\left| \int \phi_{\xi_1,\xi_2} \Phi(d\xi_1,\xi_2) \right| < \int E\{\phi^2_{\xi_1,\xi_2}\} \ell(d\xi_1)$$

for almost all ξ_2 (with respect to m). Hence we have

$$\int E\{|\phi_{\xi_1,\xi_2} \Phi(d\xi_1,\xi_2)|\} d\xi_2 < \infty$$

and from Fubini's theorem it follows that (6.7) is almost surely finite. Thus, the right hand side of (6.6) is a well-defined element of $L^2(\Omega)$. From the fact that $m(I_2) = 1$ we deduce that

$$\left(\int [\int \phi_{\xi_1,\xi_2} \Phi(d\xi_1,\xi_2)] d\xi_2\right)^2 \leq \int [\int \phi_{\xi_1,\xi_2} \Phi(d\xi_1,\xi_2)]^2 d\xi_2.$$

Taking expectations of both sides and using (5.5) of Theorem 5.1 shows that (6.6) defines an element of $L^2(\Omega)$ satisfying the following mean square inequality.

$$(6.8) \qquad E\{(\int \phi_\xi \Phi(d\xi_1,\xi_2) d\xi_2)^2\} \leq \int E\{\phi^2_\xi\} \ell(d\xi_1) d\xi_2.$$

For the purposes of chapters 3 and 4 we are concerned with parameterized measures that come from stochastic measures in the following way. Suppose $\Phi(dz_1,z_2)$ is a stochastic measure on I_1, parameterized on I_2, and let $\Phi(dz_1 \times dz_2)$ be a stochastic measure on $I_1 \times I_2$. For

each z_1 and z_1^+ in I_1 with $z_1 < z_1^+$ and for each z_2 in I_2, we assume that

$$\Phi((z_1,z_1^+],z_2) = \Phi((z_1,z_1^+] \times [0,z_2])$$

In figure 6.1 we draw the situation for $I_1 = [0,1]^1 = I_2$. The parameterized measure $\Phi((z_1,z_1^+],z_2)$ is the Ψ-measure of the shaded region. Note that the independence condition in (3) of definition 6.1 has the following geometrical interpretation in terms of figure 6.1. The rectangles $[0,z_1] \times [0,\bar{z}_2]$ and $(z_1,z_1^+] \times [0,z_2]$ are disjoint, and $\Phi((z_1,z_1^+] \times [0,z_2])$ is independent of $\mathcal{F}_{z_1,\bar{z}_2}$. For example, this happens when Ψ is a Gaussian random measure.

Define the stochastic measure, Φ_{Λ_0}, on $[0,1]^n$ by

$$\Phi_{\Lambda_0}(A) = \int \chi_A(\xi) p(\xi) \Phi_0(d\xi)$$

where Φ_0 is the standard Gaussian random measure in $[0,1]^n$, and p is in $L^\infty([0,1]^n)$ (see 2.5 and 3.1). Then Φ_{Λ_0} is also a Guassian random measure with variance

$$E\{\Phi_{\Lambda_0}(A)^2\} = \int \chi_A(\xi) p(\xi)^2 d\xi.$$

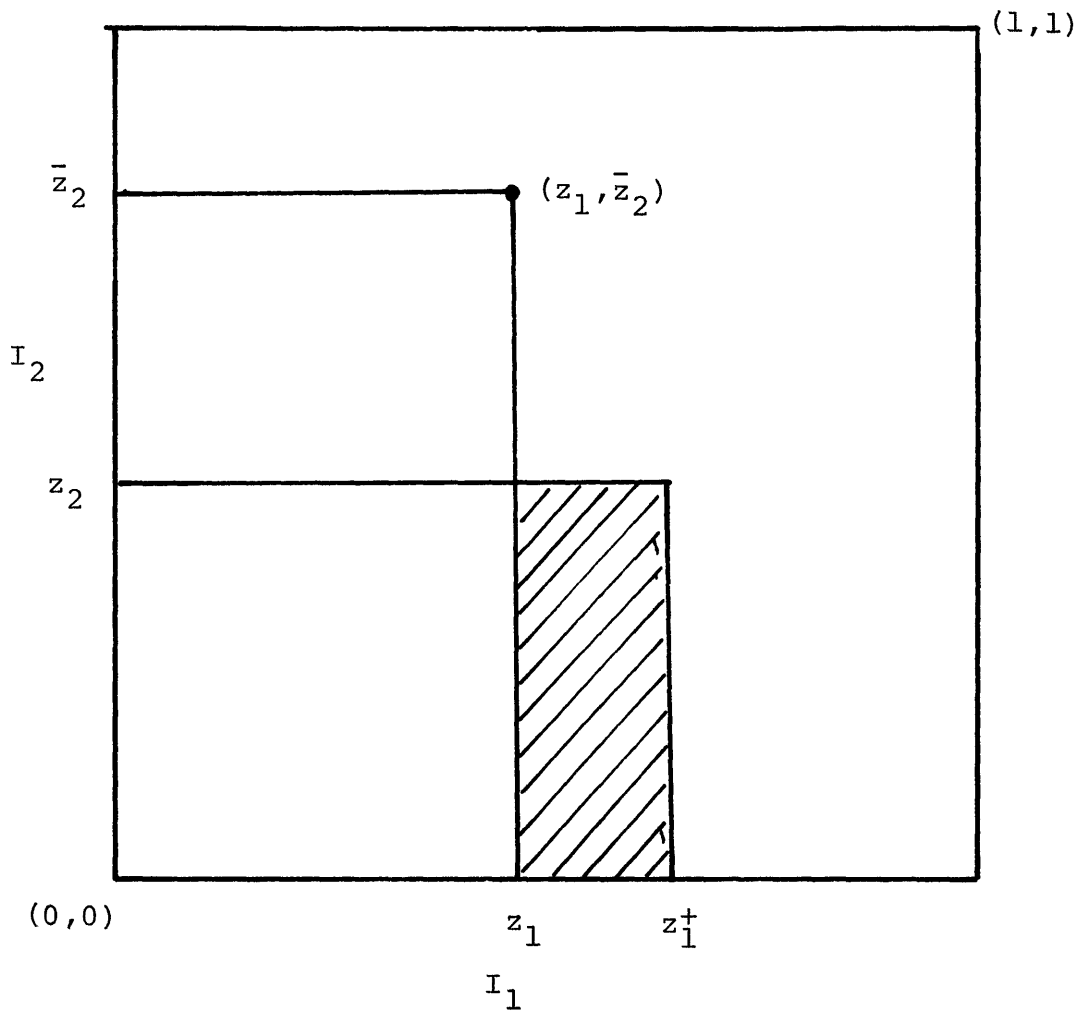


Figure 6.1

Note that Φ_{Λ_0} corresponds to $\partial_{\Lambda_0} M_z$ given by

$$\partial_{\Lambda_0} M_z = \partial_1 \partial_2 \cdots \partial_n M_z.$$

Suppose I_1 is the space of parameters corresponding to $S = \{i_1, i_2, \ldots, i_k\}$ and I_2 is the space of remaining parameters in $[0,1]^n$. Then define Φ_S by

$$\Phi_S((z_1, z_1^+], z_2) = \Phi_{\Lambda_0}((z_1, z_1^+] \times [0, z_2])$$

where z_1 and z_1^+ are on I_1, $z_1 < z_1^+$ and z_2 is in I_2. Here we identify $I_1 \times I_2$ with $[0,1]^n$ in the obvious way, and for $\xi \varepsilon [0,1]^n$ we write $\xi = (\xi_1, \xi_2)$ with ξ_1 in I_1 and ξ_2 in I_2. Thus, Φ_S corresponds to the differential

$$\partial_S M_z = \partial_{i_1} \partial_{i_2} \cdots \partial_{i_k} M_z$$

and this allows us to interpret the differential $\partial_S M$ rigorously in terms of the parameterized stochastic measure Φ_S.

To show this defines a parameterized stochastic measure, we must verify the four conditions of Definition 6.1. The additivity condition (2) is obvious. The measurability conditions (3) follow easily from the independent increment nature of the Gaussian random measure Φ_{Λ_0}. That is, for all

$z \in [0,1]^n$ and $A \subset [0,1]^n$ such that $[0,z) \cap A = \phi$, we know that $\Phi_{\Lambda_0}(A)$ and F_z are independent. This fact gives us all of condition (3). The only conditions left to verify are (1) and (4). In (1) define ℓ by

$$\ell = ||p||_\infty^2 m_1$$

where m_1 is the Lebesgue measure on I_1 and $||p||_\infty$ is the L^∞ sup norm of p on $[0,1]^n$. With this choice of ℓ, (6.1) is clearly satisfied. Define $0(\varepsilon) = (n-k)\varepsilon$ and verify that

$$E\{[\Phi_2((z_1,z_1^+],z_2^+) - \Phi_S((z_1,z_1^+],z_2)]^2\}$$

$$= \int_{(z_1,z_1^+]\times([0,z_2^+]-[0,z_2])} p(\xi)^2 d\xi$$

and

$$\int_{(z_1,z_1^+]\times[0,z_2^+]-[0,z_2])} p(\xi)^2 d\xi$$

$$< ||p||_\infty^2 m_1((z_1,z_1^+])m_2([0,z_2^+]-[0,z_2])$$

where m_2 is the Lebesgue measure on I_2. It is easy to check that

$$m_2([0,z_2^+]-[0,z_2]) \leq \sum_i |z_{2,i}^+ - z_{2,i}| \leq (n-k)||z_2^+ - z_2||.$$

Thus, we have (6.3) of definition 6.3.

Finally note that

$$E\{(\Phi_s((z_1,z_1^+],z_2))^2\} = \int_{(z_1,z_1^+]\times[0,z_2]} p(\xi)^2 d\xi$$

$$= \Delta_s V(z_1,z_2) = \int_{(z_1,z_1^+]} \partial_s V(\xi_1,z_2),^{\dagger}$$

where V is defined as in (3.4). Thus, it follows that

$$(6.9) \quad E\{(\int \Phi_{\xi_1,\xi_2} \Phi_S(d\xi_1,\xi_2))^2\} = \int_{I_1} E(\phi^2_{\xi_1,\xi_2}) \partial_S V(\xi_1,\xi_2).$$

6.3 Products of Parameterized Measures

Now that we have defined parameterized measures Φ_S corresponding to the stochastic partial differentials $\partial_S M$, we can define the other measures Φ_Λ, corresponding to $\partial_\Lambda M_1$ as products of parameterized measures Φ_S. We first construct products of two parameterized measures and then extend these results to products of more than two measures.

---

†The operators Δ_S and ∂_S are defined in 4.2 and 4.3. Of course, $\partial_S V$ is a well-defined deterministic differential.

Definition 2

Suppose $P = P_1 \times P_2$ is a partition of $I_1 \times I_2$ such that P_1 is a partition of I_1 and P_2 is a partition of I_2. Let ϕ be a partition process defined on P. Then define

$$\int_P \phi_\xi \Phi_1(d\xi_1, \xi_2) \Phi_2(d\xi_2, \xi_1)$$

or, more compactly,

$$\int_P \phi_\xi (\Phi_1 * \Phi_2)(d\xi)$$

to be the sum

$$\sum_{z_1 \in P_1, z_2 \in P_2} \phi_{z_1, z_2} \Phi_1(\delta z_1, z_2) \Phi_2(\delta z_2, z_1).$$

Theorem 1

Let ϕ be a partition process defined on a partition P_0 of $I_1 \times I_2$ and hence, on any partition $P \supset P_0$. Let $\Phi_1(dz_1, z_2)$ be a stochastic measure on I_1, parameterized on I_2, and let $\Phi_2(dz_2, z_1)$ be a stochastic measure on I_2, parameterized on I_1. Assume Φ_1 and Φ_2 are defined with respect to the Borel measures ℓ_1 on I_1 and ℓ_2 on I_2 respectively. Then, if P and P' are partitions of

$I_1 \times I_2$ with $P \supset P_0$ and $P' \supset P_0$, the following inequalities are true.

$$(6.10) \qquad E\{(\int_P \phi_\xi \Phi_1 * \Phi_2(d\xi))^2\} \leq \int E\{\phi_\xi^2\} \ell_1(d\xi_1) \ell_2(d\xi_2)$$

There is a function $0(\varepsilon)$ of $\varepsilon \geq 0$ such that $0(\varepsilon) \to 0$ as $\varepsilon \to 0$ and

$$(6.11) \quad E\{[\int_{P'} \varepsilon_\xi \Phi_1 * \Phi_2(d\xi) - \int_{P'} \phi_\xi \Phi_1 * \Phi_2(d\xi)]^2\}$$

$$\leq [0(|P|) + 0(|P'|)] \int E\{\phi_\xi^2\} \ell_1(d\xi_1) \ell_2(d\xi_2).$$

The proof of the theorem is a straightforward but lengthy application of the definition of parameterized measure and the details are left to appendix A.

If $|P^n| \to 0$ and $P^n \supset P^0$, then from Theorem 1 we know that for the partition process ϕ defined on P^0,

$$\int_{P^n} \phi_\xi \Phi_1 * \Phi_2(d\xi), \quad n = 1,2,\ldots,$$

is a Cauchy sequence in $L^2(\Omega)$ and hence, converges to a unique element of $L^2(\Omega)$, which we denote as

$$(6.12) \qquad \int \phi_\xi \, \Phi_1 * \Phi_2(d\xi).$$

The random variable (6.12) is independent of the partition P^0 on which ϕ is defined, it is a linear mapping of ϕ, and by taking limits in (6.10) we find that

$$(6.13) \qquad E\{(\int \phi_\xi \Phi_1 * \Phi_2(d\xi))^2\} \leq \int E\{\phi_\xi^2\} \ell_1 \otimes \ell_2(d\xi).$$

Thus, we can extend (6.12) to a continuous linear mapping of $L_a^2(\Omega \times I_1 \times I_2, \ell_1 \otimes \ell_2)$ satisfying (6.13). In fact, this mapping is the stochastic integral of a stochastic measure as the following theorem shows.

Theorem 2

Let $\Phi_1 * \Phi_2$ be defined by

$$(6.14) \qquad \Phi_1 * \Phi_2(A) = \int \chi_A(\xi) \Phi_1 * \Phi_2(d\xi)$$

for all Borel sets $A \subset I_1 \times I_2$ where the integral in (6.14) is the mean-square extension of (6.12). Then $\Phi_1 * \Phi_2$ is a stochastic measure on $I_1 \times I_2$ with respect to the Borel measure $\ell_1 \otimes \ell_2$, and the stochastic integral of $\Phi_1 * \Phi_2$ agrees with (6.12).

For example, in Theorem 2 let $\Phi_1 = \Phi_{S_1}$ and $\Phi_2 = \Phi_{S_2}$ where $S_1 \cap S_2 = \phi$ and $S_1 \cup S_2 = \{1,2,\ldots,n\}$. Then $\Phi_\Lambda = \Phi_1 * \Phi_2$ for $\Lambda = \{S_1, S_2\}$. It is clear from Definition 2

that $\Phi_1*\Phi_2 = \Phi_2*\Phi_1$ so Φ_Λ is independent of the order of S_1 and S_2 in Λ. It is not difficult to show that

$$(6.15) \qquad E\{\Phi_\Lambda(dz)^2\} = \partial_\Lambda V(z) = \partial_{S_1} V(z)\partial_{S_2} V(z).$$

Proof of Theorem 2 and (6.15)

For $A = (z_1,z_1^+]\times(z_2,z_2^+]$ it is easy to see that for each partition P on which χ_A is defined, we have that $\int_P \chi_A(\xi)\Phi_1*\Phi_2(d\xi)$ is $F_{z_1^+,z_2^+}$-measurable, independent of F_{z_1,z_2} and uncorrelated with $F_{\bar{z}_1,\bar{z}_2}$ if $[0,\bar{z}_1]\times[0,\bar{z}_2]$ and $(z_1,z_1^+]\times(z_2,z_2^+]$ are disjoint, and hence, if $(\bar{z}_1,\bar{z}_2) \not> (z_1,z_2)$. Thus, condition (3) of Definition 5.1 is satisfied. This condition remains true in the limit $|P| \to 0$. Thus, we see immediately that $\int\chi_A(\xi)\Phi_1*\Phi_2(d\xi)$ defines a stochastic measure. If ϕ_0 is an F_{z_1,z_2}-measurable random variable, then

$$\int\phi_0\chi_A(\xi)\Phi_1*\Phi_2(d\xi) = \phi_0\int\chi_A(\xi)\Phi_1*\Phi_2(d\xi).$$

Thus, as remarked at the end of 5.3, the integral in (6.14) agrees with the stochastic integral defined by the stochastic measure $\Phi_1*\Phi_2(A)$.

As in (6.9), we have that

$$E\{\Phi_{S_2}((z_2,z_2^+],z_1)^2 = \int_{(z_2,z_2^+]} \partial_{S_2} V(z_1,\xi_2)$$

and

$$E\{\Phi_{S_2}((z_1,z_1^+],z_2)^2\} = \int_{(z_1,z_1^+]} \partial_{S_1} V(\xi_1,z_2).$$

As shown in the proof of Theorem 1 in appendix A,

$$E\{(\int_P \phi_\xi \Phi_1 * \Phi_2(d\xi))^2\}$$

$$= \sum_{z_1\in P_1, z_2\in P_2} E\{\phi^2_{z_1,z_2}\}E\{\Phi_1(\delta z_1,z_2)^2\}E\{\Phi_2(\delta z_2,z_1)^2\}.$$

Using these last three expressions and passing to the limit $|P| \to 0$ gives the result (6.15). ///

To construct the other measures Φ_Λ we need to define products $\Phi_{S_1} * \Phi_{S_2}$ when $S_1 \cap S_2 = \phi$ but $S_1 \cup S_2 \subsetneq \{1,2,\ldots,n\}$. That is, when $\Phi_{S_1} * \Phi_{S_2}$ still has some free parameters. The following theorem gives the necessary result.

Theorem 3

Let $\Phi_1(dz_1,z_2,z_3)$ be a stochastic measure on I_1,

parameterized on $I_2 \times I_3$, and let $\Phi_2(dz_2, z_1, z_3)$ be a stochastic measure on I_2, parameterized on $I_1 \times I_3$. Then $\Phi_1(dz_1, z_2, z_3)\Phi_2(dz_2, z_1, z_3)$ is a stochastic measure on $I_1 \times I_2$, parameterized on I_3.

Proof of Theorem 3

Note that for fixed z_3 we can use Theorem 2 to show that $\Phi_1(dz_1, z_2, z_3)\Phi_2(dz_2, z_1, z_3)$ is a stochastic measure on $I_1 \times I_2$ with respect to the increasing family $\{\mathcal{F}_{z_1, z_2, z_3} \mid z_1 \in I_1, z_2 \in I_2\}$. Thus, we only need to confirm the measurability condition (3) and the continuity condition (4) in Definition 1.

From the construction of Definition 2, Theorem 1, and Theorem 2 it is clear that $\int \chi_Z(\xi_1, \xi_2)\Phi_1(d\xi_1, \xi_2, z_3)\Phi_2(d\xi_2, \xi_1, z_3)$ satisfies the measuribility conditions in (3) because $\int_P \chi_A \Phi_1(d\xi_1, \xi_2, z_3)\Phi_2(d\xi_2, \xi_1, z_3)$ does so for each partition P and these conditions hold true in the mean-square limit as $|P| \to 0$. The continuity condition follows from the equation

$$\Phi_1(\delta z_1, z_2, z_3^+)\Phi_2(\delta z_2, z_1, z_3^+) - \Phi_1(\delta z_1, z_2, z_3)\Phi_2(\delta z_2, z_1, z_3)$$
$$= [\Phi_1(\delta z_1, z_2, z_3^+) - \Phi_1(\delta z_1, z_2, z_3)]\Phi_2(\delta z_2, z_1, z_3^+)$$
$$+ \Phi_1(\delta z_1, z_2, z_3)[\Phi_2(\delta z_2, z_1, z_3^+) - \Phi_2(\delta z_2, z_1, z_3)].$$

The mean-square of each of the two terms on the right-hand-side is bounded by

$$\ell_1(\delta z_1)\ell_2(\delta z_2)0(||z_3^+ - z_3||).$$

Applying the triangle inequality shows that the mean-square of the left-hand-side has the same bound. It is easy to extend this bound to a bound for

$$\int_P \chi_A \Phi_1 * \Phi_2(d\xi_1 \times d\xi_2, z_3^+) - \int_P \chi_A \Phi_1 * \Phi_2(d\xi_1 \times d\xi_2, z_3)$$

and then to extend this bound to a bound for the limit as $|P| \to 0$, namely

$$\Phi_1 * \Phi_2(A, z_3^+) - \Phi_1 * \Phi_2(A, z_3).$$

This completes the proof. ///

Theorem 3 allows us to define arbitrary products of parameterized measures, but there are several ways to construct a product of three or more measures. The associative law of products in the next theorem shows that all these constructions lead to the same result.

Theorem 4

Let Φ_1, Φ_2 and Φ_3 be parameterized stochastic measures

on $I_1 \times I_2 \times I_3$. Then the following associative law holds true.

$$(6.16) \quad (\Phi_1(dz_1, z_2, z_3)\Phi_2(dz_2, z_1, z_3))\Phi_3(dz_3, z_1, z_2)$$

$$= \Phi_1(dz_1, z_2, z_3)(\Phi_2(dz_2, z_1, z_3)\Phi_3(dz_3, z_1, z_2))$$

Proof of Theorem 4

The idea of the proof is to show both sides of (6.16) are approximated by the same expression and that, in the limit, the approximation is exact and (6.16) is true. Define

$$\int_P \phi \; \Phi_1 * \Phi_2 * \Phi_3 =$$

$$\sum_{(z_1, z_2, z_3) \in P} \phi_{z_1, z_2, z_3} \Phi_1(\delta z_1 \cdot z_2, z_3)\Phi_2(\delta z_2, z_1, z_3)\Phi_3(\delta z_3, z_1, z_2)$$

where ϕ is a partition process defined on the partition P of $I_1 \times I_2 \times I_3$. Let us also write

$$\int_P \phi(\Phi_1 * \Phi_2) * \Phi_3 =$$

$$\sum_{(z_1, z_2, z_3) \in P} \phi_{z_1, z_2, z_3} \Phi_1 * \Phi_2(\delta z_1 \times \delta z_2, z_3)\Phi_3(\delta z_3, z_1, z_2).$$

From Theorem 1 we know that

$$E\{(\int\phi(\Phi_1*\Phi_2)*\Phi_3 - \int_P \phi(\Phi_1*\Phi_2)*\Phi_3)^2\} < ||\phi||^2 0(|P|)$$

where

$$||\phi||^2 = \int E\{\phi_\xi^2\}\ell_1 \otimes \ell_2 \otimes \ell_3(d\xi).$$

A simple calculation using the measurability condition (3) of Definition 1 for Φ_1, Φ_2 and Φ_3 shows that

$$\{\int_P \phi(\Phi_1*\Phi_2)*\Phi_3 - \int_P \phi\ \Phi_1*\Phi_2*\Phi_3)^2\} < ||\phi||^2 0(|P|).$$

Thus, as $|P| \to 0$ we have the result

$$\lim_{|P|\to 0} \int_P \phi\ \Phi_1*\Phi_2*\Phi_3 = \int\phi\ \Phi_1*\Phi_2)*\Phi_3$$

where convergence is in $L^2(\Omega)$. Since $\int_P \phi\ \Phi_1*\Phi_2*\Phi_3$ is symmetric in Φ_1, Φ_2, Φ_3, it must also converge to $\int\phi\ \Phi_1*(\Phi_2*\Phi_3)$ and thus, (6.16) must be true. ///

In fact, it is easy to see that for every multiple product, $\Phi_1*\Phi_2*\ldots*\Phi_q$, of parameterized stochastic measures, Φ_i, we have

$$\lim_{|P|\to 0} \int_P \phi\ \Phi_1*\Phi_2*\ldots*\Phi_q = \int\phi\ \Phi_1*\Phi_2*\ldots*\Phi_q$$

where ϕ is a partition process and P is always a refinement of the partition on which ϕ is defined. The convergence occurs in $L^2(\Omega)$ and the mean-square difference for any P is at most $||\phi||^2 0(|P|)$.

Note that if $\Phi_i = \Phi_{S_i}$, then

$$\int_P \phi\ \Phi_1 * \Phi_2 * \ldots * \Phi_q = \int_P \phi \partial_\Lambda M,$$

where the right-hand-side is defined as in 4.3.

6.4 Discussion of Measures and Integrals for 4.3

Theorem 4 allows us to define all the measures, Φ_Λ, for any Λ where $\cup\Lambda \subset \{1,2,\ldots,n\}$ as in 4.2. If $\cup\Lambda = \{1,2,\ldots,n\}$, then Φ_Λ is a stochastic measure on $[0,1]^n$ and it is easy to demonstrate

$$(6.17)\quad E\{(\int \phi_\xi \Phi_\Lambda(d))^2\} = \int E\{\phi_\xi^2\} \partial_\Lambda V(\xi).$$

(see (6.15).) If $\cup\Lambda \subsetneq \{1,2,\ldots,n\}$, we may use the earlier results of the section to define mixed stochastic-deterministic integrals

$$\int \phi_\xi \Phi_\Lambda(d\xi_1, \xi_2) d\xi_2,$$

which correspond to the differentials $\partial_\Lambda M_z d_S z$ where

$S \cap \cup\Lambda = \phi$. (See 4.2). From the inequality (6.8) we obtain

$$(6.18)\quad E\{(\int \phi_\xi \Phi_\Lambda(d\xi_1, \xi_2)d\xi_2)^2\} < \int E\{\phi_\xi^2\}\partial_\Lambda V(\xi)d\xi_2.$$

Note that we can also define parameterized mixed integrals

$\int \phi_{\xi_1, \xi_2, z_3} \Phi_\Lambda(d\xi_1, \xi_2, z_3)d\xi_2$ with no problem.

If ϕ is a partition process with respect to P and if as in 4.3 we have

$$\int_P \phi \partial_\Lambda M = \sum_{z \in P} \phi_z \Delta_\Lambda M_z,$$

then as $|P| \to 0$,

$$\int_P \phi \partial_\Lambda M \to \int \phi_\xi \Phi_\Lambda(d\xi)$$

in mean square. This allows us to prove very simply that Φ_{Λ_1} and Φ_{Λ_2} are orthogonal if Λ_1 and Λ_2 are not equal, but $\cup\Lambda_1 = \cup\Lambda_2$. Suppose that P is a partition of $[0,1]^n$. Then it is easy to show that

$$E\{\phi_z \psi_z, \Delta_{\Lambda_1} M_z \Delta_{\Lambda_2} M_z,\} = 0$$

for all z, z' in P and all partition processes ϕ and ψ defined on P. The argument uses the measurability conditions (3) of Definition 1 and is straightforward. It follows that

$$E\{(\int_P \phi \partial_{\Lambda_1} M)(\int_P \psi \partial_{\Lambda_2} M)\} = 0$$

and thus, by letting $|P| \to 0$, we see that

$$E\{(\int \phi \partial_{\Lambda_1} M)(\int \psi \partial_{\Lambda_2} M)\} = 0$$

for all partition processes. The result extends immediately to all ϕ and ψ in $L_a^2(\Omega \times I, m)$. That is, Φ_{Λ_1} and Φ_{Λ_2} are orthogonal measures.

The results of this section allow us to interpret rigorously the sotchastic differential equations in terms of their corresponding integral equations as in 4.4. Each stochastic differential $\partial_\Lambda M$ corresponds to a stochastic measure or a parameterized stochastic measure depending on whether $\cup\Lambda = \{1,2,\ldots,n\}$ or $\cup\Lambda \subsetneq \{1,2,\ldots,\}$. Likewise, each differential $\partial_\Lambda M d_s z$ corresponds to a mixed integral or a parameterized integral (one with free variables in it) depending on whether $\cup\Lambda \cup S = \{1,2,\ldots,n\}$ or $\cup\Lambda \cup S \subsetneq \{1,2,\ldots,n\}$. In the next section we prove Theorem 1 and 2 of 4.4.

CHAPTER 7

DIFFERENTIATION OF PRODUCTS

This section completes the technical results necessary to prove the product differentiation rules in theorems 4.1 and 4.2. Theorems 7.1 and 7.2 in this section are the precise versions of theorems 4.1 and 4.2 in terms of parameterized stochastic measures and integrals. As such, these two theorems are the main results of this paper; with them (and with Ito's one-parameter stochastic differentiation formula) one can derive stochastic differentiation formulas for any number of parameters.

Suppose $\phi(dz_1,t,z_2)$ is a stochastic measure on I_1, parameterized on $[0,1] \times I_2$ and suppose $\Psi(dz_1 \times dt,z_2)$ is a stochastic measure on $I_1 \times [0,t]$, parameterized on I_2. Assume Φ and Ψ are related by the equation

$$\Phi(A,t,z_2) = \Psi(A \times [0,t],z_2) \tag{7.1}$$

for Borel sets A in I_1, t in $[0,1]$ and z_2 in I_2. We write

$$\partial_t \Phi(dz_1,t,z_2) = \Psi(dz_1 \times dt,z_2) \tag{7.2}$$

whenever (7.1) is true. Note that if $\partial_s \partial_t \Phi$ exists, then so does $\partial_t \partial_s \Phi$ and the two measures are equal. Also note

that we never take two derivatives with respect to the same variable. Only mixed derivatives are defined. The following theorem, corresponding to Theorem 4.1 of 4.5, proves that the derivative of a product of stochastic measures satisfies the ordinary calculus product differentiation rule.

Theorem 1

Suppose $\Phi_1(dz_1,t,z_2,z_3)$ is a stochastic measure on I_1, parameterized on $[0,1] \times I_2 \times I_3$, and $\Phi_2(dz_2,t,z_1,z_3)$ is a stochastic measure on I_2, parameterized on $[0,1] \times I_1 \times I_3$. As in (7.1) suppose that

$$\Phi_1(A_1,t,z_2,z_3) = \Psi_1(A_1 \times [0,t],z_2,z_3)$$

and

$$\Phi_2(A_2,t,z_1,z_3) = \Psi_2(A_2 \times [0,t],z_1,z_3).$$

Then we have

$$(7.3) \quad \Phi_1 * \Phi_2(A_1 \times A_2,t,z_3) = \Psi_1 * \Phi_2(A_1 \times A_2 \times [0,t],z_3) + \Phi_1 * \Psi_2(A_1 \times A_2 \times [0,t],z_3).$$

In terms of the convention of (7.2) we can write (7.3) in a more suggestive form, namely,

$$(7.4) \qquad \partial_t(\Phi_1 * \Phi_2) = \partial_t\Phi_1 * \Phi_2 + \Phi_1 * \partial_t\Phi_2.$$

Proof of Theorem 1.

In this proof we omit the z_3 parameter since that more general case offers no new difficulty in the proof and only complicates the notation. Suppose ϕ is a partition process on $I_1 \times I_2$ and let $P = P_1 \times P_2$ be a partition of $I_1 \times I_2$ on which ϕ is defined. Let P_0 be a partition of $[0,1]$. Then we can write

$$\phi_1(\delta z_1, z_2, t)\Phi_2(\delta z_2, z_1, t) - \Phi_1(\delta z_1, z_2, t)\Phi_2(\delta z_2, z_1, t)$$

$$= \Psi_1(\delta z_1 \times \delta t, z_2)\Phi_2(\delta z_2, z_1, t) + \Phi_1(\delta z_1, z_2, t)\Psi_2(\delta z_2 \times \delta t, z_1)$$

$$+ \Psi_1(\delta z_1 \times \delta t, z_2)\Psi_2(\delta z_2 \times \delta t, z_1).$$

Thus, we have

$$\int_P \phi_{\xi_1,\xi_2} \Phi_1(d\xi_1,\xi_2,t)\Phi_2(d\xi_2,\xi_1,t)$$

$$= \int_{P\times P_0} \phi_{\xi_1,\xi_2} \chi_{[0,t]}(\tau)\Psi_1(d\xi_1\times d\tau,\xi_2)\Phi_2(d\xi_2,\xi_1,\tau)$$

$$+ \int_{P\times P_0} \phi_{\xi_1,\xi_2} \chi_{[0,t]}(\tau)\Phi_1(d\xi_1,\xi_2,\tau)\Psi_2(d\xi_2\times d\tau,\xi_1)$$

$$+ \sum_{\substack{z_1\in P_1, z_2\in P_2 \\ \tau\in P_0}} \phi_{z_1,z_2} \chi_{[0,t]}(\tau)\Psi_1(\delta z_1\times\delta\tau,z_2)\Psi_2(\delta z_2\times\delta\tau,z_1).$$

Using the condition (3) of Definition 6.1, we can show that the last term on the right has a mean square bounded by

$$\int_{I_1\times I_2} E\{\phi^2_{\xi_1,\xi_2}\}\ell_1(d\xi_1)\ell_2(d\xi_2). \quad 0(|P_0|)$$

and hence, this term vanishes as $|P\times P_0| \to 0$. The other terms converge to the desired result,

$$\int \phi_\xi \Phi_1{}^*\Phi_2(d\xi,t) = \int \phi_\xi \Psi_1{}^*\Phi_2(d\xi \times [0,t])$$

$$+ \int \phi_\xi \Phi_1{}^*\Psi_2(d\xi \times [0,t]).$$

This completes the proof. ///

Theorem 7.1 here gives us Theorem 4.1. If $S \subset \{1,2,\ldots,n\}$ and $i\in\{1,2,\ldots,n\}$ but $i\notin S$, then it is

clear that

$$\partial_i \Phi_S = \Phi_{S \cup \{i\}}$$

A simple induction argument applied to Theorem 7.1 gives (4.9), which now has a rigorous interpretation in terms of stochastic measures.

The next theorem justifies the statements in Theorem 4.2 by proving a differentiation rule for products of processes and measures. We introduce a useful convention for denoting the partial differential of a process. Note that this convention agrees with the interpretation of stochastic differential equations in 4.4.

Let $f_{t,z}$ be an adapted, measurable process on $[0,1]\times I$ such that

$$(7.5) \qquad f_{t,z} = \int_{[0,t]} \phi_{\tau,z} \Phi(d\tau, z) + \int_{[0,z]} \psi_{\tau,z} d\tau$$

where we assume that Φ is a stochastic measure on $[0,1]$, parameterized on I, and that the integrals in (7.5) are well-defined according to §6. Then we will write

$$(7.6) \qquad \partial_t f_{t,z} = \phi_{t,z} \Phi(dt, z) + \psi_{t,z} dt.$$

We can now present the main result of this paper, the product differentiation rule.

Theorem 2

Suppose $\psi(dz_1,z_2,t,z_3)$ is a parameterized stochastic measure and suppose that

$$\partial_t \Psi(dz_1,z_2,t,z_3) = \Theta(dz_1 \times dt, z_2, z_3).$$

Suppose f_{z_1,z_2,t,z_3} is an adapted, measurable process and suppose that

$$\partial_t f_{z_1,z_2,t,z_3} = \phi_{z_1,z_2,t,z_3}(dt,z_1,z_2,z_3) + \psi_{z_1,z_2,t,z_3} dt.$$

We assume ϕ and ψ are in appropriate L_a^2 spaces so that all integrals are well-defined. Then we have the following equation.

$$(7.7)\quad \int_{[0,z_1]\times[0,z_2]} f_{\xi_1,\xi_2,t,z_3} \Psi(d\xi_i,\xi_2,t,z_3) d\xi_2$$

$$= \int_{[0,z_1]\times[0,z_2]\times[0,t]} \phi_{\xi_1,\xi_2,\tau,z_3} \Phi(d\tau,\xi_1,\xi_2,z_3) \Psi(d\xi_1,\xi_2,\tau,z_3) d\xi_2$$

$$+ \int_{[0,z_1]\times[0,z_2]\times[0,t]} \phi_{\xi_1,\xi_2,\tau,z_3} \Psi(d\xi_1,\xi_2,\tau,z_3) d\xi_2 d\tau$$

$$+ \int_{[0,z_1]\times[0,z_2]\times[0,t]} f_{\xi_1,\xi_2,\Theta,z_3} \Theta(d\xi_1,x d\tau,\xi_2,z_3) d\xi_2$$

In terms of the convention of (7.2) and (7.6) we can write (7.7) in compact differential form as follows.

$$(7.8)\qquad \partial_t(f_{z_1,z_2,t,z_3}\Psi(dz_1,z_2,t,z_3)dz_2)$$

$$= \partial_t(f_{z_1,z_2,t,z_3})\Psi(dz_1,z_2,t,z_3)dz_2$$

$$+ f_{z_1,z_2,t,z_3}\partial_t(\Psi(dz_1,z_2,t,z_3))dz_3$$

Proof of Theorem 2.

We prove Theorem 2 in the special case $t=1,[0,z_1]=I$ and where there is no z_2 or z_3 dependence. There is no loss in generality and we gain an advantage in notational clarity. The mixed stochastic - deterministic integral was defined in such a way that we can prove (7.7) without the integration over $d\xi_2$ and then integrate both sides with respect to $d\xi_2$ to obtain the generalized equation (7.7). Thus, we must prove

$$(7.9)\qquad \int_I f_{\xi,1}\Psi(d\xi,1)$$

$$= \int_{I\times[0,1]} \phi_{\xi,\tau}\Phi(d\tau,\xi)\Psi(d\xi,\tau) + \int_{I\times[0,1]} \psi_{\xi,\tau}\Psi(d\xi,\tau)d\tau$$

$$+ \int_{I\times[0,1]} f_{\xi,\tau}\Theta(d\xi\times d\tau)$$

where we assume

$$(7.10)\quad f_{\xi,t} = \int_0^t \phi_{\xi,\tau}\Phi(d\tau,\xi) + \int_0^t \psi_{\xi,\tau}d\tau$$

and

$$(7.11)\quad \Psi(d\xi,t) = \Theta(d\xi\times[0,t]$$

Let P_0 be a partition of $[0,1]$. Then using (7.10) and (7.11) we can rewrite the left-side of (7.9) as follows.

$$(7.12)\quad \int_I f_{\xi,1}\Psi(d\xi,1) =$$

$$\sum_{t\in P_0} \int \{\int \delta t^{\phi}_{\xi,\tau}{}^{\Phi(d\tau,\xi)}\}\Psi(d\xi,t) + \sum_{t\in P_0} \int_I \{\int \delta t^{\psi}_{\xi,\tau}{}^{d\tau}\}\Psi(d\xi,t)$$

$$+ \sum_{t\in P_0} \int_I f_{\xi,t}\Theta(d\xi\times\delta t) + \sum_{t\in P_0} \int_I [f_{\xi,t^+} - f_{\xi,t}]\Theta(d\xi\times\delta t)$$

We show that as $|P_0| \to 0$ the first three terms on the right-side of (7.12) converge to the three terms on the right-hand-side of (7.9), and the last term in (7.12) converges to zero.

Consider $d\xi\to\Theta(d\xi\times\delta t)$ as a stochastic measure over I. Then it is clear that $\Theta(d\xi\times\delta t)$ and $\Theta(d\xi\times\delta t')$ are orthogonal for $t\neq t'$ in P_0. It follows immediately that

$$(7.13)\quad E\{(\sum_{t\in P_0}\int_I [f_{\xi,t^+}-f_{\xi,t}]\Theta(d\xi\times\delta t))^2\}$$

$$\leq \sum_{t\in P_0}\int_I E\{(f_{\xi,t^+}-f_{\xi,t})^2\}\ell(d\xi)m(\delta t),$$

where ℓ is a Borel measure on I and M is Lebesgue measure on [0,1]. From (7.10) we have

$$E\{(f_{\xi,t^+}-f_{\xi,t})^2\} \leq \int_{\delta t} E\{\phi^2_{\xi,\tau}\}\ell_0(d\tau) + \int_{\delta t} E\{\psi^2_{\xi,\tau}\}d\tau.$$

Thus, substituting this bound into (7.13), we obtain

$$(7.14)\quad E\{(\sum_{t\in P_0}\int_I [f_{\xi,t^+}-f_{\xi,t}]\Theta(d\xi\times\xi t))^2\}$$

$$\leq [\int_{I\times[0,1]} E\{\phi^2_{\xi,\tau}\}\ell(d\xi)\ell_0(d\tau) + \int E\{\psi^2_{\xi,\tau}\}\ell(d\xi)d\tau]0(|P_0|)$$

We assume that ψ is in $L^2_a(\Omega\times I\times[0,1],\ell\otimes\ell_0)$ and that ψ is in $L^2_a(\Omega\times I\times[0,1],\ell\otimes m)$ so that (7.14) converges to zero as $|P_0|\to 0$. The Borel measure ℓ_0 on [0,1] corresponds to the parameterized stochastic measure Φ.

Now let us show that

$$(7.15)\quad \lim_{|P_0|\to 0}\sum_{t\in P_0}\int_I f_{\xi,t}\Theta(d\xi\times\delta t) = \int_{I\times[0,1]} f_{\xi,\tau}\Theta(d\xi\times d\tau).$$

If we let

$$f^{P_0} = \sum_{t \in P_0} f_{\xi,t} \chi_{\delta t},$$

then (7.15) is equivalent to

$$\lim_{|P_0| \to 0} \int_{I \times [0,1]} f^{P_0}_{\xi,\tau} \Theta(d\xi \times d\tau) = \int_{I \times [0,1]} f_{\xi,\tau} \Theta(d\xi \times d\tau).$$

Hence, it suffices to show

$$(7.16) \qquad \int_{I \times [0,1]} E\{(f - f^{P_0})^2\} \ell(d\xi) m(d\tau) \to 0$$

as $|P_0| \to 0$. From (7.10) we have

$$E\{(f_{\xi,t'} - f^{P_0}_{\xi,t})^2\} \leq \int_{\delta t} E\{\phi^2_{\xi,\tau}\} \ell_0(d\tau) + \int_{\delta t} E\{\psi^2_{\xi,\tau}\} m(d\tau)$$

for t' in δt. Thus, (7.16) is bounded above by

$$[\int E\{\phi^2_{\xi,\tau}\} \ell(d\xi) \ell_0(d\tau) + \int E\{\psi^2_{\xi,\tau}\} \ell(d\xi) m(d\tau)] 0(|P_0|).$$

From Theorem 6.1 it is easy to deduce that

$$(7.17)\qquad E\{(\int \phi\Phi^*\Psi - \int_{P\times P_0} h\Phi^*\Psi)^2\}$$

$$\leq (||\phi-h|| + ||h||0(|P\times P_0|))^2$$

where h is a partition process on $P\times P_0$, and where

$$||h||^2 = \int_{I\times[0,1]} E\{h^2\}\ell(d\xi)\ell_0(d\tau).$$

If P_0 is a partition of $[0,1]$ and P is a partition of I, then

$$\sum_{t\in P_0}\int_I \{\int_{\delta t} \phi_{\xi,\tau}\Phi(d\tau,\xi)\}\Psi(d\xi,t) - \int_{P\times P_0} h\Phi^*\Psi$$

$$= \sum_{z\in P, t\in P_0}\int_{\delta z}\{\int_{\delta t}(\phi_{\xi,\tau}-h_{z,t})\Phi(d\tau,\xi)\}\Psi(d\xi,t)$$

$$+ \sum_{z\in P, t\in P_0}\int_{\delta z} h_{z,t}(\Phi(\delta t,\xi) - \Phi(\delta t,z))\Psi(d\xi,t).$$

It is a simple matter to check that

$$E\{(\sum_{z\in P, t\in P_0}\int_{\delta z}\{\int_{\delta t}(\phi_{\xi,\tau}-h_{z,t})\Phi(d\tau,\xi)\Psi(d\xi,t))^2\} < ||\phi-h||^2$$

and, from the continuity condition (6.3) for Φ,

$$E\{(\sum_{z\in P, t\in P_0} \int_{\delta z} h_{z,t}(\Phi(\delta t,\xi)-\Phi(\delta t,z))\Psi(d\xi,t))^2\} < ||h||^2 0(|P|)$$

Combining these results gives us

$$E\{(\int\phi\Phi^*\Psi - \sum_{t\in P_0} \int_I \{\int_{\delta t} \phi_{\xi,\tau}\Phi(d\tau,\xi)\}\Psi(d\xi,t))^2\}$$

$$\leq (2||\phi-h|| + 2||h||0(|P\times P_0|))^2.$$

By letting $h \to \phi$ and $|P\times P_0| \to 0$ we obtain

$$\lim_{|P_0|\to 0} \sum_{t\in P_0} \int_I \{\int_{\delta t} \phi_{\xi,\tau}\Phi(d\tau,\xi)\}\Psi(d\xi,t) = \int\phi\Phi^*\Psi.$$

Finally, we need to show

$$(7.18) \quad \lim_{|P_0|\to 0} \sum_{t\in P_0} \int_I \{\int_{\delta t} \psi_{\xi,\tau}d\tau\}\Psi(d\xi,t) = \int_{I\times[0,1]} \psi\ \Psi(d\xi,\tau)d\xi.$$

As above, let h be a partition process defined on $P\times P_0$.

$$\sum_{t\in P_0} \int_I \{\int_{\delta t} \psi_{\xi,\tau}d\tau\}\Psi(d\xi,t) - \int_{I\times[0,1]} h\ \Psi(d\xi,\tau)d\tau$$

$$= \sum_{t\in P_0} \int_I \{\int_{\delta t} \psi_{\xi,\tau}-h_{\xi,\tau}\ d\tau\}\Psi(d\xi,t)$$

$$+ \sum_{t\in P_0, z\in P} h_{z,t} \int_{\delta t} (\Psi(\delta z,\tau)-\Psi(\delta z,t))d\tau.$$

Noting that $\sum_{t\in P_0} m(\delta t) = m([0,1]) = 1$ we obtain the following results from the convexity of $x \to x^2$.

$$(7.19)\quad \Big(\sum_{t\in P_0} \int_I \Big\{\int_{\delta t} \psi_{\xi,\tau}\, d\tau\Big\}\Psi(d\xi,t)\Big)^2$$

$$\leq \sum_{t\in P_0} \frac{1}{m(\delta t)}\Big(\int_I \Big\{\int_{\delta t} \psi_{\xi,\tau}\, d\tau\Big\}\Psi(d\xi,t)\Big)^2$$

$$(7.20)\quad \Big\{\int_{\delta t} \psi_{\xi,\tau}\, d\tau\Big\}^2 \leq \int_{\delta t} \psi^2_{\xi,\tau}\, d\tau \cdot m(\delta t)$$

Taking expectations in (7.19) and (7.20) gives us

$$E\Big\{\Big(\sum_{t\in P_0} \int_I \Big\{\int_{\delta t} \psi_{\xi,\tau}\, d\tau\Big\}\Psi(d\xi,t)\Big)^2\Big\} \leq ||\psi||^2$$

where $||\psi||^2 = \int_{I\times[0,1]} E\{\psi_{\xi,\tau}{}^2\}\ell(d\xi)m(d\tau)$.

In particular, we have that

$$(7.21)\quad E\Big\{\Big(\sum_{t\in P_0} \int_I \Big\{\int_{\delta t} \psi_{\xi,\tau} - h_{\xi,\tau}\, d\tau\Big\}\Psi(d\xi,t)\Big)^2\Big\} \leq ||\psi - h||^2.$$

Note that

$$\Psi(\delta z,\tau) - \Psi(\delta z,t) = \Theta(\delta z x(t,\tau]).$$

Using this relation and the measurability conditions on Θ

from Definition 5.1, we find that

$$(7.22)\quad E\{(\sum_{z\in P,t\in P_0} h_{z,t}\int_{\delta t}(\Psi(\delta z,\tau)-\Psi(\delta z,t))d\tau)^2\}$$

$$\leq \sum_{z\in P,t\in P_0} \mathbb{E}\, h_{z,t}^2\}\int_{\delta t}(\tau-t)d\tau\ \ell(\delta z)$$

$$\leq ||h||^2 0(|P_0|)$$

We have already shown in (6.8) of 6.2 that

$$(7.23)\quad E\{(\int\psi\Psi(d\xi,\tau)d\tau-\int h\Psi(d\xi,\tau)d\tau)^2\} < ||\psi-h||^2.$$

From (7.21), (7.22), (7.23) we conclude that

$$E\{(\int_{I\times[0,1]}\psi\Psi(d\xi,\tau)d\tau-\sum_{t\in P_0}\int_I\{\int_\delta \psi_{\xi,\tau}d\tau\}\Psi(d\xi,t))^2\}$$

$$\leq (||\phi-h||+||h||0(|P_0|))^2$$

from which (7.18) follows immediately. Thus, the proof is complete. ///

To prove Theorem 4.2 take $\Psi=\Phi_\Lambda$ in (7.7), where Φ_Λ is defined as in 5.2, 6.2 and especially 6.4. Then take $dz_2 = d_S z$, and take $\Phi(d\tau,\xi_1,\xi_2,z_3)$ in (7.7) to be the parameterized measure $\Phi_{\{i\}}(dz_i,z_1,\ldots,z_{i-1},z_{i+1},\ldots z_n)$

generated from the Gaussian random measure as in 6.2. The integral equation corresponding to the differential equation (4.10) is precisely (7.7) with the above choices of Ψ, Φ, and so on.

CHPATER 8

REPRESENTATION OF MULTIPARAMETER SQUARE INTEGRABLE MARTINGALES BY MULTIPARAMETER STOCHASTIC INTEGRALS

8.1 Introduction

Suppose that W is the p-parameter Wiener process on $[0,1]^p$ and that $\{F_z : z \in [0,1]^p\}$ is the increasing family of σ-fields generated by W. Let $F = \bigvee_z F_z$. Then we can apply the multiparameter stochastic differentiation formula developed in previous chapters to show that every F-measurable random variable X with finite variance has a unique (a.s.) representation

$$X = c + \sum_{\Lambda} \int \phi_\Lambda \partial_\Lambda W \tag{1.1}$$

where the sum ranges over the collection of stochastic differentials $\partial_\Lambda W$ of W defined in chapter 4. The ϕ_Λ are square integrable multiparameter processes adapted to $\{F_z\}$ and c is a constant, namely $c = E\{X\}$.

In particular, if M is a square integrable martingale with respect to $\{F_z\}$, then from (1.1) we can show that M has a unique (a.s.) representation in terms of the orthogonal stochastic integrals generated by the Wiener process, namely,

$$(1.2) \qquad M_z = c + \sum_{\Lambda} \int_{[0,z]} \phi_\Lambda \partial_\Lambda W.$$

To derive these results we must first extend the multiparameter stochastic differentiation rule to cover the case of vector valued martingales. This extension is very simple and we carry it out in section 2. Using this result we apply an argument of Wong and Zakai (1974) to prove the representation (1.1) in section 3. The martingale representation follows very easily as we show at the end of section 3.

8.2 Multi-parameter, multidimensional stochastic differentiation formula

If $z \to M_z$ is a p-parameter, n-dimensional martingale generated by n Gaussian random measures, then the stochastic differentiation formula for $z \to f(M_z, z)$ follows from the one-parameter differentiation rule and the rules for differentiating products of stochastic differentials given in previous chapters.

If X maps elements of $\mathbb{R}^p$ into random variables with values in $\mathbb{R}^n$, then say that X is a p-parameter, n-dimensional stochastic process. Thus, a multidimensional, multiparameter process is a p-parameter, n-dimensional process for which p and n are both at least 2. In this section (and those following) we strive to maintain

this distinction between parameter and dimension.

Suppose that $\Phi_1, \Phi_2, \ldots, \Phi_n$, are Gaussian random measures defined on the same family of sigma-fields, $\{F_z\}$, and on the same parameter space, $[0,1]^p$. Define the n-dimensional, p-parameter martingale generated by these measures as follows: Let M be the vector $(M_1, M_2, \ldots, M_n)$ where each coordinate M_μ is given by $M_\mu(z) = \Phi_\mu((0,z])$. If we fix our attention on one parameter, say the i-th, and keep all the other parameters constant, then as functions of this one parameter, z_i, the M_μ are one-parameter martingales with respect to the obvious subfamily of $\{F_z\}$. These one-parameter martingales are square integrable and continuous so that the results of Kunita and Watanabe [2] apply. Thus, denote $\langle M_\mu, M_\nu \rangle = E\{\Phi_\mu((0,z])\Phi_\nu((0,z])\}$ by $V_{\mu\nu}$ for each $\mu, \nu = 1,2,\ldots,n$. Suppose that $f : R^n \times R^p \to R$ is a function with continuous partial derivatives up to second order in the R^n variables and to first order in the R^p variables, then

$$(2.1)\quad \Delta_i f(M_z, z) = \sum_{\mu=1}^{n} \int_0^{z_i} \frac{\partial f}{\partial u_\mu} \partial_i M_\mu + \frac{1}{2} \sum_{\mu,\nu=1}^{n} \int_0^{z_i} \frac{\partial^2 f}{\partial u_\mu \partial u_\nu} \partial_i V_{\mu\nu} + \int_0^{z_i} \frac{\partial f}{\partial \xi_i} d\zeta_i$$

In this formula all the parameters except the i-th are held constant. The forward difference operator acts only on the i-th parameter and takes the difference between 0 and z_i. The integrals in (2.1) are integrating only over the i-th parameter. The stochastic integrals are defined in the obvious way.

To obtain the multi-parameter differentiation formula, apply (2.1) repeatedly for $i = 1,2,\ldots,p$ and use the rule for differentiating stochastic products presented in chapter 4.7. For example, the differential version of (2.1) for $i = 1$ is

$$(2.2)\quad \partial_1 f = \sum_{\mu=1}^{n} f_\mu \partial_1 M_\mu + \frac{1}{2} \sum_{\mu,\nu=1}^{n} f_{\mu,\nu} \partial_1 V_{\mu\nu} + f_1 dz_1$$

where we now abbreviate partial derivatives with respect to u_μ by the Greek subscript μ, that is $f_\mu = \frac{\partial f}{\partial u_\mu}$. Note that u_μ is one of the R^n variables of f; similarly z_i denotes one of the R^p variables of f. Derivatives with respect to z_i are denoted by the Roman subscript, i, that is $f_i = \frac{\partial f}{\partial z_i}$.

The two-parameter formula is found by taking the ∂_2 differential of (3.1), giving

$$\partial_2\partial_1 f = \sum_{\mu=1}^{n} f_\mu \partial_2\partial_1 M_\mu + \sum_{\mu,\nu=1}^{n} f_{\mu,\nu}\partial_1 M_\mu \partial_2 M_\nu$$

$$+ \sum_{\lambda=1}^{n} (f_{\lambda,1}dz_1 + \frac{1}{2}\sum_{\mu,\nu=1}^{n} f_{\mu,\nu,\lambda}\partial_1 V_{\mu\nu})\partial_2 M_\lambda$$

$$+ \sum_{\lambda=1}^{n} (f_{\lambda,2}dz_2 + \frac{1}{2}\sum_{\mu,\nu=1}^{n} f_{\mu,\nu,\lambda}\partial_2 V_{\mu\nu})\partial_1 M_\lambda$$

$$(2.3) \quad + f_{1,2}dz_1 dz_2 + \frac{1}{2}\sum_{\mu,\nu=1}^{n} f_{\mu,\nu}\partial_2\partial_1 V_{\mu,\nu}$$

$$+ \frac{1}{2}\sum_{\mu,\nu=1}^{n} f_{\mu,\nu,2}\partial_1 V_{\mu\nu}dz_2$$

$$+ \frac{1}{2}\sum_{\mu,\nu=1}^{n} f_{\mu,\nu,1}dz_1\partial_2 V_{\mu\nu}$$

$$+ \frac{1}{4}\sum_{\mu,\nu,\rho,\lambda=1}^{n} f_{\mu,\nu,\rho,\lambda}\partial_1 V_{\mu\nu}\partial_2 V_{\lambda\rho}$$

The general multi-parameter, multi-dimensional differentiation formula is obtained in the same way. The product differentials denoted $\partial_1 M_\mu \partial_2 M_\nu$ are defined as stochastic product measures in chapter 4,7. The differential formula (2.3) represents a stochastic integral formula which is derived from (2.1) by means of the product differentiation rule proved in chapter 7.

8.3 Representation of Multiparameter martingales

Suppose that W is the p=parameter Wiener process on $[0,1]^p$ and that $\{\mathcal{F}_z\}$ is the family of σ-fields generated by W. Given the multi-dimensional, multi-parameter stochastic differentiation formula, it is easy to show that every $\mathcal{F}_1 = V\mathcal{F}_z$ measurable random variable X with finite variance has a unique representation,

$$X = c + \sum_{\Lambda} \int \phi_\Lambda \, \partial_\Lambda W \tag{3.1}$$

where $\{\partial_\Lambda W\}$ is the (finite) collection of stochastic differentials of W, ϕ_Λ are square integrable processes adapted to $\{\mathcal{F}_z\}$, and c is a constant, namely $c = E\{X\}$.

In particular, if M is a square integrable martingale with respect to $\{\mathcal{F}_z\}$ then M has a unique representation in terms of the orthogonal martingales generated by the Wiener process, namely,

$$M_z = c + \sum_{\Lambda} \int_{[0,z]} \phi_\Lambda \partial_\Lambda W. \tag{3.2}$$

The proof of the representation (3.1) parallels Wong and Zakai's (1974) proof in the two-parameter case. We sketch the proof in this section.

Let $M_\mu(z) = \int_{[0,z]} \chi_{[0,a_\mu]}(\xi)W(d\xi)$ for $\mu = 1,2,\ldots,n$ and let $V_{\mu\nu} = \langle M_\mu, M_\nu \rangle$. Suppose $f(u_1,\ldots,u_n,z)$ satisfies

$$(3.3) \qquad \frac{1}{2} \sum_{\mu,\nu=1}^{n} \frac{\partial}{\partial u_\mu \partial u_\nu} \frac{\partial V_{\mu\nu}}{\partial z_k} \frac{\partial f}{\partial z_k} = 0$$

for each $k = 1,2,\ldots,p$. Then the stochastic differentials of $f(M_z,z)$ have no mixed terms and

$$(3.4) \qquad \partial_1\partial_2\ldots\partial_p f = \sum_k \psi_k(z)\Phi_k(dz)$$

where $\Phi_k(dz)$ are the stochastic differentials generated by $M_1, M_2,\ldots,M_n$ and ψ_k are L_a^2-processes. It is easy to see that

$$\partial_{i_1}\partial_{i_2}\ldots\partial_{i_r} M_\mu = \chi_{[0,a_\mu]}\partial_{i_1}\ldots\partial_{i_r} W,$$

and thus, each differential Φ_k can be written

$$\Phi_k(dz) = \chi_{[0,b_k]}(z)\partial_{\Lambda_k} W$$

for some b_k in $[0,1]^p$ and some differential, $\partial_{\Lambda_k} W$, of the Wiener process W. For example, it is true that

$$\partial_1 M_\mu \partial_2 M_\nu = \chi_{[0,a_\mu\ a_\nu]} \partial_1 W \partial_2 W$$

Thus, if f satisfies (3.3), then it has the representation

$$(3.5) \qquad f(M_z,z) = c + \sum_\Lambda \int_{[0,z]} \phi_\Lambda \partial_\Lambda W$$

where $\phi_\Lambda = \sum_{\Lambda_k=\Lambda} \chi_{[0,b_k]}\psi_k$. If we take $z = \underline{1}$ in $[0,1]^p$, then (3.5) becomes

$$(3.6) \qquad f(W_{a_1},\ldots,W_{a_n},\underline{1}) = c + \sum_\Lambda \int \phi_\Lambda \partial_\Lambda W$$

Given a polynomial of $W_{a_1},W_{a_2},\ldots,W_{a_n}$ we can choose f so that $f(W_{a_1},\ldots,W_{a_n},\underline{1})$ is that given polynomial. As in Wong-Zakai (1974), take all f of the form

$$f(u_1,u_2,\ldots,u_n,z) = \exp\{i \sum_{\mu=1}^n u_\mu\alpha_\mu + \frac{1}{2} \sum_{\mu,\nu=1}^n \alpha_\mu\alpha_\nu V_{\mu\nu}(z)\}$$

together with all the partial derivatives of f with respect to α_μ, and all linear combinations of the partial derivatives. These f all satisfy (3.3), and they include all polynomials in $u_1,u_2,\ldots,u_n$ when $z = \underline{1}$ and $\alpha = \underline{0}$.

The polynomials of $W_{a_1},W_{a_2},\ldots,W_{a_n}$ for $n = 1,2,\ldots,$

and $a_1, a_2, \ldots, a_n$ in $[0,1]^p$ are dense in the space of random variables which have finite variance and are $\underline{F}_1$ measurable. One way to show this is to approximate all the elementary functions $\prod_{\mu=1}^{n} \chi_{(s_\mu, t_\mu]}(W_{a_\mu})$ by Hermite polynomials in much the same way as Cameron and Martin (1947) in their representation of one-parameter Wiener functionals.

Since all polynomials are represented as stochastic integrals, the set of all stochastic integrals is dense in $L^2(\Omega, \underline{F}_1)$. In fact these two spaces coincide. For suppose that for X in $L^2(\Omega, \underline{F}_1)$ it is true that

$$\lim_{m\to\infty} [c^m + \sum_{\Lambda} \int \phi_\Lambda \partial_\Lambda W] = X. \tag{3.7}$$

Then we can show that $c^m \to c$ and $\phi_\Lambda^m \to \phi_\Lambda$ as $m \to \infty$. The stochastic integrals (and the constant c) are mutually orthogonal as shown in chapters 4,5 and

$$||\{\phi_\Lambda, c\}||^2 \triangleq c^2 + \sum_{\Lambda} \int E\{\phi_\Lambda^2\} \partial_\Lambda V = E\{(c + \sum_{\Lambda} \int \phi_\Lambda \partial_\Lambda W)^2\} \tag{3.8}$$

where $V(z) = z_1 z_2 \ldots z_p$, defines an L^2 norm for the space of finite vectors $\{\phi_\Lambda, c\}$ where each ϕ_Λ is a square integrable process adapted to $\{F_z\}$. This space is

complete and (3.7), (3.8) imply that the sequence $\{\phi_\Lambda^m, c^m\}$ is a Cauchy sequence, and therefore, that there are $\{\phi_\Lambda, c\}$ in the same space such that

$$\lim_{m\to\infty} \phi_\Lambda^m = \phi_\Lambda \ ,$$

$$\lim_{m\to\infty} c^m = c \ ,$$

$$c + \sum_\Lambda \int \phi_\Lambda \partial_\Lambda W = X.$$

Since $||\{\phi_\Lambda, c\}|| = 0$ implies that $\phi_\Lambda = 0$ for each Λ and all (z,ω) expect for a set of measure 0, the representation is unique.

CHAPTER 9

CONCLUSIONS

Part II derives a stochastic differentiation formula for n-parameter processes of the form $z \to f(M_z, z)$ where $z \to M_z$ is an n-parameter process given by

$$M_z = \int_{[0,z]} p(\xi)\Phi_0(d\xi)$$

for p in $L^\infty([0,1]^n)$ and for Φ_0, the standard Gaussian random measure on $[0,1]^n$. The differentiation formula results from taking mixed stochastic partial derivatives of the process $z \to f(M_z, z)$. Although the rigorous definition and proofs are lengthy (5,6,7), the differentiation rules are simply those stated in section 3.3. When differentiating the product of two differentials or the product of a process and a differential, one uses the ordinary product differentiation rules of deterministic calculus. The ordinary product differentiation rule doesn't hold for the product of two processes, but it does hold if one factor is a differential.

When differentiating a process, one uses the one-parameter Ito differentiation formula. Thus, the multi-parameter stochastic differentiation formula is the same as the deterministic formula, except for one-parameter

stochastic corrections. For example, the general two-dimensional stochastic differentiation formula is

$$\partial_1\partial_2 f(M_z,z) = f_u(M_z,z)\partial_1\partial_2 M_z + f_{uu}(M_z,z)\partial_1 M_z \partial_2 M_z$$

$$+ (f_{u,1}(M_z,z) + \tfrac{1}{2} f_{uuu}(M_z,z)V_1(z))dz_1\partial_2 M_z$$

$$+ (f_{u,2}(M_z,z) + \tfrac{1}{2} f_{uuu}(M_z,z)V_2(z))\partial_1 M_z dz_2$$

$$+ (f_{1,2}(M_z,z) + \tfrac{1}{2} f_{uu}(M_z,z)V_{1,2}(z) + \tfrac{1}{2} f_{uu,2}(M_z,z)V_1(z)$$

$$+ \tfrac{1}{2} f_{uu,1}(M_z,z)V_2(z) + \tfrac{1}{4} f_{uuuu}(M_z,z)V_1(z)V_2(z))dz_1dz_2.$$

The corresponding two-dimensional deterministic differentiation formula results from setting V identically equal to zero, giving

$$\partial_1\partial_2 f(M_z,z) = f_u(M_z,z)\partial_1\partial_2 M_z + f_{uu}(M_z,z)\partial_1 M_z \partial_2 M_z$$

$$+ f_{u,1}(M_z,z)dz_1\partial_2 M_z + f_{u,2}(M_z,z)\partial_1 M_z dz_2 + f_{1,2}(M_z,z)dz_1dz_2.$$

We interpret the stochastic differential equations in terms of their corresponding integral equations. The stochastic integrals necessary for this interpretation are the natural forward-difference integrals corresponding

to the differentials $\partial_1 {}_2M$, $\partial_1 M \partial_2 M$, and so on. These integrals have nice mean-square, orthogonality and martingale properties. In addition, the multiparameter stochastic integrals provide an integral representation of multiparameter square integrable martingales in chapter 8 just as one-parameter square integrable martingales can be represented by one-parameter Ito stochastic integrals. Note, however, that in the multiparameter case, more than one type of stochastic integral is necessary.

Thus, the results of part II show that the multiparameter stochastic differentiation formula is a natural extension of the one-parameter stochastic calculus. However, there remains much to do. Certain immediate extensions of the present result are necessary in order to give multi-dimensional stochastic calculus the full power of the one-dimensional calculus. Basically, one must permit a wider class of processes, $z \to M_z$, than those defined by

$$M_z = \int_{[0,z]} p(\xi) \Phi_0(d\xi)$$

for $p \in L^\infty([0,1]^n)$. Since the stochastic calculus is essentially a mean-square calculus, one should be able to allow for $p \in L^2([0,1]^n)$. The present difficulty

is defining integrals $\int \phi \partial_1 M \partial_2 M$ in this case. In a more ambitious extension, one should allow p to be an L_a^2-stochastic process. Along the same lines, allow M to have the form

$$M_z = \sum_{i=1}^{N} \int_{[0,z]} \phi_{i,\xi} \Phi_i(d\xi)$$

where the Φ_i are orthogonal stochastic measures on $[0,1]^n$ (such as the Φ_Λ measures of chapter 6 and the ϕ_i are L_a^2-processes.

Cairoli-Walsh (1975) and Wong-Zakai (1974,1975) study more general classes of multi-parameter martingales*. Hopefully, for some braod class of continuous multi-parameter martingales M, one could define integrals $\int \phi \partial_1 \partial_2 M$, $\int \phi \partial_1 M \partial_2 M$, and so on, and extend the present results in much the same way that Kunita and Watanabe (1967) extend Ito's basic result. An extension of this paper to cases for which M is a jump process is more difficult since the definition of product measure in chapter 6 depends crucially on the mean-square continuity condition of Definition 6.1. Preliminary investigations indicate that the results of part II can be extended to

---

*i.e., strong martingales, weak martingales, and i-martingales (i=1,2).

multiparameter martingales which are strong martingales in the sense of Cairoli and Walsh (1975). A martingale M is a strong martingale if for every $z, z^+ \in [0,1]^p$ such that $z < z^+$ we have

$$E\{\Delta M_z | \mathcal{F}_z^s\} = 0$$

where Δ denotes the difference operator

$$\Delta = \Delta_{\{1,2,\ldots,p\}}$$
$$= \Delta_1\Delta_2\cdots\Delta_p$$

and where $\mathcal{F}_z^s$ denotes the σ-field

$$\mathcal{F}_z^s = \bigvee_{z \not< \zeta} \mathcal{F}_\zeta .$$

In the two-parameter case, $\mathcal{F}_z^s$ is the σ-field generated by $\mathcal{F}_\zeta$ for all ζ in the set illustrated in figure 9.1.

Cairoli and Walsh define a weak martingale M so that

$$E(\Delta M_z | \mathcal{F}_z) = 0$$

where Δ is the difference operator defined above for

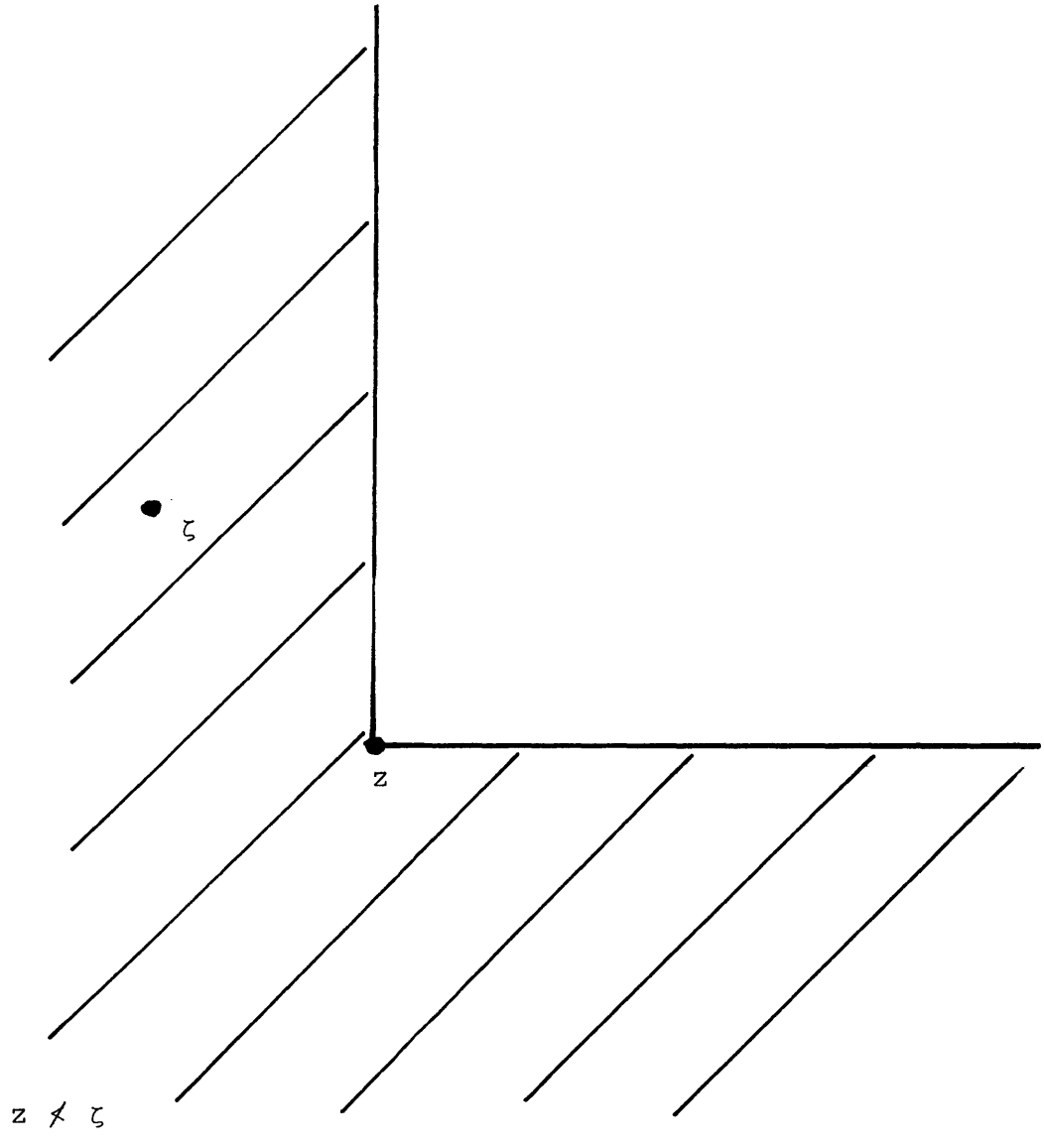


Figure 9.1

$z^+ > z$ and $\{\mathcal{F}_z\}$ is the increasing family of σ-fields for the problem. Note that $\mathcal{F}_z \neq \mathcal{F}_z^s$. Generally speaking, all strong martingales are martingales (defined as in chapter 2), and all martingales are weak martingales. The converse relationships, however, are generally not true: in general a weak martingale is not a martingale and in general a martingale is not a strong martingale. The extension of our differentiation formula to the case of weak martingales is more difficult than the extension to the case of strong martingales and indeed, an extension may not be possible. Nevertheless, if an extension is impossible, it should be possible to prove that for some weak martingales M and for some differentiable function f one cannot represent the multiparameter process $z \to f(M_z)$ by stochastic integrals over the parameter space $[0,1]^p$. That is, one cannot represent $f(M_z)$ as

$$f(M_z) = \sum_{i=1}^{N} \int_{[0,z]} \phi_{i,\xi} \Phi_i(d\xi)$$

where ϕ_i are L_a^2-processes and Φ_i are stochastic measures. Note that Wong and Zakai (1975) have proved that $f(M_z)$ can be represented by stochastic integrals over $[0,1]^2$ and $[0,1]^4$ in the case $p = 2$. In the general case, this result would extend to a representation of $f(M_z)$ using integrals over $[0,1]^{kp}$ for

$k = 1,2,...p$. Although it is more convenient to use stochastic integrals defined over $[0,1]^p$, it may be that integrals over $[0,1]^p$ alone are not able to represent all point functions of p-parameter processes such as $f(M_z)$ above.

The extensions listed so far are technical generalizations of the mathematics, but other generalizations are needed to understand the significance of the theory for physical models. One can define increasing families of σ-fields with respect to many different partial order relations. For a wide class of such partial orderings it is possible to define stochastic measure and integrals as in §5 and §6. If a smooth curvilinear coordinate system induces the partial ordering, it is not hard to derive a stochastic differentiation formula with respect to this coordinate system. For the purposes of modelling physical random fields it is important to understand the relationships between these different partial orderings and coordinate systems. Since there is no natural ordering of multi-dimensional space as there is for one-dimensional time, one must select a partial ordering for each random field model.[†] A general theory of multi-dimensional

[†]Equivalently, there is no natural notion of causality for multi-dimensional fields as there is for one-dimensional processes.

stochastic processes that indicates the significance of the partial ordering for the physical model would appreciably aid understanding of random field models, and this should be the goal of future work on multi-dimensional stochastic calculus.

APPENDIX A: Proof of Theorem 6.1

Suppose $P = P_1 \times P_2$ is a partition of $I_1 \times I_2$ and ϕ is a partition process defined on P. Consider two points of P, (z_1, z_2) and (z_1', z_2'). If $z_1 \neq z_1'$, then either $z_1 + \not> z_1'$ or $z_1' + \not> z_1$[†]. Suppose the former is true and let $a \vee b$ be the last upper bound of a and b with respect to the order relation $<$, Then it is easy to see $z_1^+ \vee z_1' \not> z_1'$. Let $\bar{z}_1 = z_1^+ \vee z_1'$, then from the measurability condition of Definition 6.1 we know that $\Phi_1(\delta z_1', z_2')$ is uncorrelated with $F_{\bar{z}_1, \bar{z}_2}$ for any $\bar{z}_2$ in I_2*. In particular, choose $\bar{z}_2 = z_2^+ \vee z_2'^+$ so that ϕ_{z_1, z_2}, ϕ_{z_1, z_2}, , $\Phi_1(\delta z_1, z_2)$, $\Phi_2(\delta z_2, z_1)$, and $\Phi_2(\delta z_2', z_1')$ are all $F_{\bar{z}_1, \bar{z}_2}$-measurable and so that, thus, we have

$$(A.1)\quad E\{\phi_{z_1, z_2} \phi_{z_1', z_2'} \Phi_1(\delta z_1, z_2) \Phi_2(\delta z_2, z_1) \Phi_1(\delta z_1', z_2') \Phi_2(\delta z_2', z_1')\} = 0.$$

The other cases are the same, and (A.1) is true whenever

[†] See the proof of Theorem 5.1 and the accompanying figure (5.2).

* See remarks on correlation and condition (3), (5.2) and (5.3) of Definition 5.1. The same remarks apply to parameterized stochastic measures. Also see figure (A.1).

$z_1 \neq z_1'$ or $z_2 \neq z_2'$. Using the independence condition in (3) of definition 6.1, we obtain the following expression for the left-hand-side of (A.1).

$$(A.2)\quad \delta_{z_1,z_1'}\delta_{z_2,z_2'}\, E\{\phi^2_{z_1,z_2}\}E\{\Phi_1(\delta z_1,z_2)^2\}E\{\Phi_2(\delta z_2,z_1)^2\}.$$

Here we have denoted the Kronecker delta-function by $\delta_{z,z'}$. The inequality (6.10) follows immediately from (A.2) and bounds of the type (6.1) for Φ_1 and Φ_2.

To prove (6.11), first let $P' \supset P$ where P and P' are partitions of $I_1 \times I_2$ and ϕ is defined on both of these partitions. Also, let $P = P_1 \times P_2$ and $P' = P_1' \times P_2'$. The difference

$$\int_P \phi_\xi \Phi_1 * \Phi_2(d\xi) - \int_{P'} \phi_\xi \Phi_1 * \Phi_2(d\xi)$$

can be rewritten as the sum

$$(A.3)\quad \sum_{\substack{z_1 \in P_1 \\ z_2 \in P_2}} \{\phi_{z_1,z_2} \Phi_1(\delta z_1, z_2)\Phi_2(\delta z_2, z_1)$$

$$- \sum_{\substack{z_1' \in \delta z_1 \cap P_1' \\ z_2' \in \delta z_2 \cap P_2'}} \Phi_1(\delta z_1', z_2')\Phi_2(\delta z_2', z_1')\}.$$

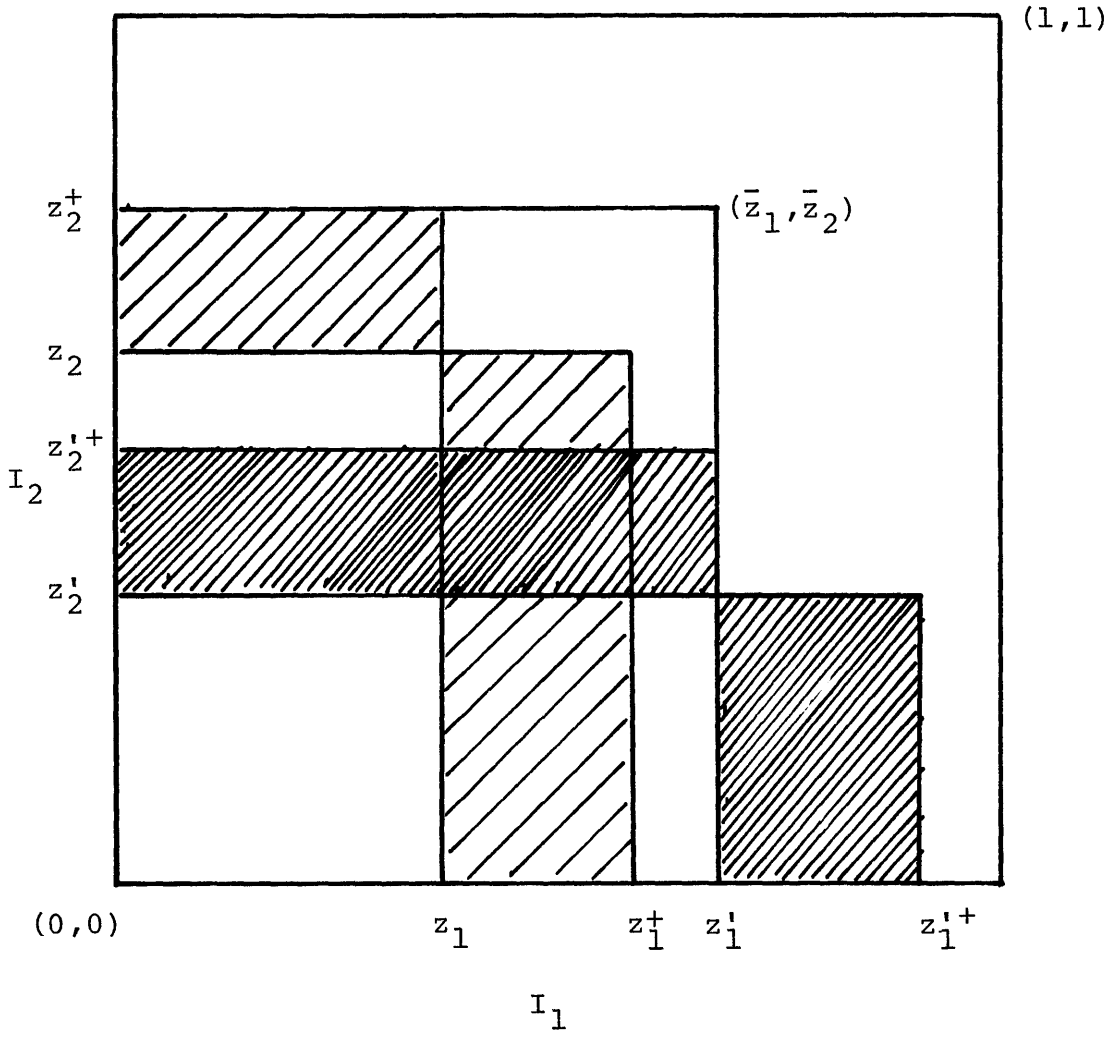


Figure A.1

Note that $\phi_{z_1,z_2} = \phi_{z_1',z_2'}$ for $z_1' \varepsilon \delta z$, and $z_2' \varepsilon \delta z_2$ since P' P.

Since P' is a refinement of P, we have the relation,

$$\delta z_1 = \bigcup_{z_1' \varepsilon \delta z_1 \cap P_1'} \delta z_1'$$

for each z_1 in P_1, and a similar relation for each z_2 in P_2. Using the additivity of Φ_1 and Φ_2 as stochastic measures, we can rewrite (A.3) as

$$(A.4) \quad \sum_{\substack{z_1 \varepsilon P_1 \\ z_2 \varepsilon P_2}} \phi_{z_1,z_2} \sum_{\substack{z_1' \varepsilon \delta z_1 \cap P_1' \\ z_2' \varepsilon \delta z_2 \cap P_2'}} \{\Phi_1(\delta z_1', z_2)\Phi_2(\delta z_2', z_1) - \Phi_1(\delta z_1', z_2')\Phi_2(\delta z_2', z_1')\}.$$

We again use the measurability conditions as for (A.1) to show that distinct terms below are uncorrelated.

$$\phi z_1,z_2\{\Phi_1(\delta z_i, z_2)\Phi_2(\delta z_2', z_1) - \Phi_1(\delta z_1', z_2')\Phi_2(\delta z_2', z_1')\}$$

Using the independence condition, we can show that the mean square of (A.3) is

$$(A.5) \quad \sum_{(z_1,z_2) \varepsilon P} E\{\phi_{z_1,z_2}^2\} \sum_{\substack{z_1' \varepsilon \delta z_1, z_2' \varepsilon \delta z_2 \\ (z_1',z_2') \varepsilon P'}} E\{\Psi^2_{z_1',z_2'}\}$$

where

$$\Psi_{z_1',z_2'} = \{\Phi_1(\delta z_1', z_2)\Phi_2(\delta z_2', z_1) - \Phi_1(\delta z_1', z_2')\Phi_2(\delta z_2', z_1')\}.$$

Note that there is only one z_1 in P_1 such that $z_1' \varepsilon \delta z_1 \cap P_1'$ and only one z_2 in P_2 such that $z_2' \varepsilon \delta z_2 \cap P_2'$. Fix z_1' and z_2' and define the random variables, A, B, C, D, as follows.†

$$A = \Phi_1(\delta z_1', z_2)$$

$$B = \Phi_1(\delta z_1', z_2') - \Phi_1(\delta z_1', z_2)$$

$$C = \Phi_2(\delta z_2', z_2') - \Phi_2(\delta z_2', z_1)$$

$$D = \Phi_2(\delta z_2', z_1)$$

Thus, we have

$$\psi_{z_1',z_2'} = -AC - BC - BD. \tag{A.6}$$

The independence condition of (3) in Definition 6.1 implies that A is independent of D and C, and B is independent

†Figure A.2 may help visualize the argument at this point in the proof.

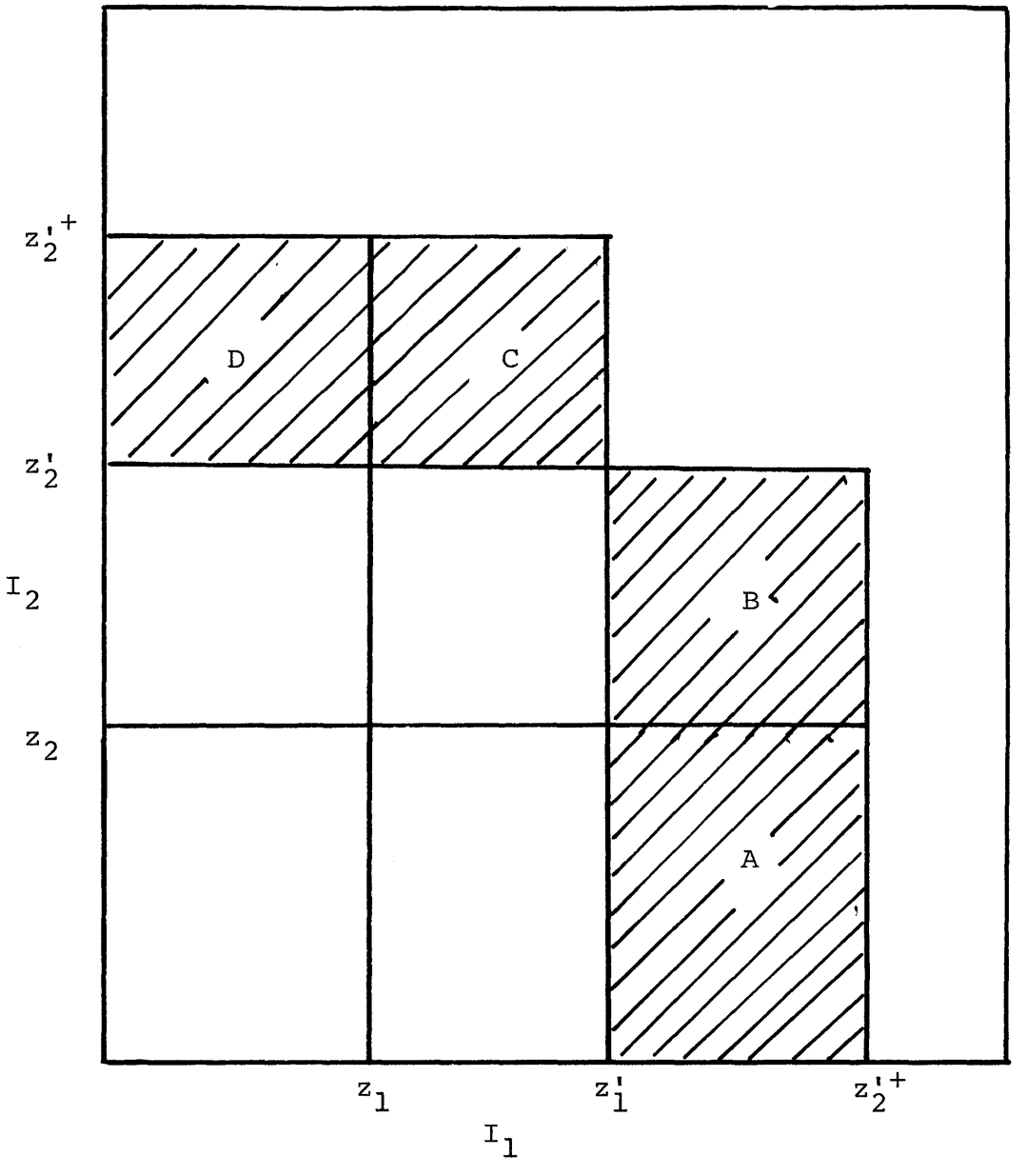


Figure A.2

of D and C. Thus, we have from (A.6) that $E\{\psi^2_{z_1',z_2'}\}$ is given by

$$(A.7)\quad E\{A^2\}E\{C^2\} + E\{B^2\}E\{C^2\} + E\{B^2\}E\{D^2\}$$
$$+ 2E\{AB\}E\{C^2\} + 2E\{B^2\}E\{CD\} + 2E\{AB\}E\{CD\}.$$

Use the continuity condition (4) of Definition 6.1 to estimate (A.7) as follows. From the variance condition (6.1) we have

$$E\{A^2\} \leq \ell_1(\delta z_1')$$

and

$$E\{D^2\} \leq \ell_2(\delta z_2')$$

From the continuity condition (6.3), we have

$$E\{B^2\} \leq \ell_1(\delta z_1')0(||z_2'-z_2||_{I_2})$$

and

$$E\{C^2\} \leq \ell_2(\delta z_2')0(||z_1'-z_1||_{I_1}).$$

Note that $||z_2'-z_2||_{I_2} \leq |P|$ and $||z_1'-z_1||_{I_1} \leq |P|$.

Using Schwarz's inequality on the terms $E\{AB\}$ and $E\{CD\}$ in (A.7), we derive the following upper bound for (A.7) and hence, for $E\{\psi^2_{z_1',z_2'}\}$

$$(A.8) \qquad E\left\{\psi^2_{z_1',z_2'}\right\} \leq \ell_1(\delta z_1')\ell_2(\delta z_2')O(|P|)$$

The inequality (6.11) follows from (A.5) and (A.8) in the case $P' \supset P$. For general P' and P we take P'' so that $P'' \supset P'$ and $P'' \supset P$. Apply (6.11) for the pairs P, P'' and P', P'' and use the triangle inequality to obtain (6.11) for P, P'. ///

APPENDIX B: Stochastic Integrals and Differentiation Formulas of Wong-Zakai and Cairoli-Walsh

Wong-Zakai

Wong and Zakai (1975) define three types of stochastic integrals in the plane in addition to the Ito-type integral described in 2.5. We now show these integrals correspond to the ones defined in this paper.

Suppose ψ is an $\mathcal{F}_{z_1 \vee z_2}$-measurable random variable with finite variance. Let δz_1 and δz_2 be rectangles in $[0,1]^2$ such that

$$\chi_{\delta z_1}(\xi)\chi_{\delta z_2}(\xi') = 0$$

unless $\xi \wedge \xi'$. The relation $\xi \wedge \xi'$ is true if $\xi_1 \leq \xi'$ and $\xi_2 \geq \xi_2'$. Similarly, $I(\xi \wedge \xi')$ is the function of ξ and ξ' which 1 or 0 depending on whether or not $\xi \wedge \xi'$. Define a process $\psi_{\xi,\xi'}$ on $[0,1]^2 \times [0,1]^2$ by

$$\psi_{\xi,\xi'} = \psi\chi_{\delta z_1}(\xi)\chi_{\delta z_2}(\xi')$$

for all ξ and ξ' in $[0,1]^2$. Then Wong and Zakai (1975) define integrals of $\psi_{\xi,\xi'}$ over $[0,1]^2 \times [0,1]^2$ as follows.

(B.1) $\int \psi_{\xi,\xi'}\, dW_{\xi} dW_{\xi'} = \psi W(\delta z_1) W(\delta z_2)$

(B.2) $\int \psi_{\xi,\xi'}\, d\xi\, dW_{\xi'} = \psi m(\delta z_1) W(\delta z_2)$

(B.3) $\int \phi_{\xi,\xi'}\, dW_{\xi}\, d\xi' = \psi W(\delta z_1) m(\delta z_2)$

In equations (B.1), (B.2), and (B.3), $W(\cdot)$ denotes the standard Gaussian random measure on $[0,1]^2$, and $m(\cdot)$ denotes the Legesque measure on $[0,1]^2$. Using linearity and mean-square continuity, one extends the definitions to define integrals of processes $\psi_{\xi,\xi'}$ such that

(B.4) $\psi_{\xi,\xi'}$ is $F_{\xi,\xi'}$-measurable,

(B.5) $\int \{\psi_{\xi,\xi'},2\} d\xi d\xi' < \infty$,

and

(B.6) $\psi_{\xi,\xi'} = 0$ unless $\xi \Lambda \xi'$.

For processes $\psi_{\xi,\xi'}$ that satisfy (B.4) and (B.5) one can define the integral of $\psi_{\xi,\xi'}$ to be the same as the integral of $I(\xi \Lambda \xi')\psi_{\xi,\xi'}$. Note that each integral is a continuous linear mapping with respect to the norm defined by (B.5).

We must evaluate these integrals for processes $\psi_{\xi,\xi'}$ which are given by

$$\psi_{\xi,\xi'} = \phi_{\xi v \xi'} ,$$

where ϕ is in $L_a^2(\Omega \times [0,1]^2, m)$. First we prove the following lemma.

LEMMA:

Suppose ϕ is the partition process defined by

$$\phi_\xi = \sum_{z \in P} \phi_z \chi_{\delta z}(\xi)$$

where P is a partition of $[0,1]^2$. Then define $\phi(\xi,\xi',P)$ as

$$\phi(\xi,\xi',P) = \sum_{z \in P} \phi_z \chi_{[0,x] \times \delta y}^{(\xi)} \chi_{\delta x \times [0,y]}^{(\xi')}$$

where we always understand that $z = (x,y)$ and $\delta z = \delta x \times \delta y$ for all z in P. Then the following inequality is true.

$$\text{(B.7)} \quad \int E\{(\phi_{\xi v \xi'} - \phi(\xi,\xi',P))^2 \, d\xi d\xi' \leq 0(|P|) \int E\{\phi_\xi^2\} d\xi$$

Proof of lemma

Note that

$$\chi_{\delta z}(\xi v \xi') = \chi^{(\xi)}_{[0,x]\times\delta y}\ \chi_{\delta x \times [0,y]}^{(\xi')} + \chi^{(\xi)}_{[0,x]\times\delta y}\ \chi_{\delta x \times \delta y}^{(\xi')}$$

$$\chi_{\delta x \times \delta y}^{(\xi)}\ \chi_{\delta x \times [0,y]}^{(\xi')} + \chi_{\delta x \times \delta y}^{(\xi)}\ \chi_{\delta x \times \delta y}^{(\xi')} .$$

It is easy to calculate that since P is a partition, we must have that

$$\Big(\sum_{z\in P} \phi_z [\chi_{\delta z}(\xi v \xi') - \chi^{(\xi)}_{[0,x]\times\delta y}\ \chi^{(\xi')}_{\delta x\times[0,y]} \Big)^2$$

$$= \sum_{z\in P} \phi_z^{\,2} [\chi^{(\xi)}_{[0,x]\times\delta y}\ \chi^{(\xi')}_{\delta x\times\delta y} + \chi^{(\xi)}_{\delta x\times\delta y}\ \chi^{(\xi')}_{\delta x\times[0,y]}$$

$$+ \chi^{(\xi)}_{\delta x\times\delta y}\ \chi^{(\xi')}_{\delta x\times\delta y}] .$$

The inequality (B.7) follows immediately from taking expectation and integral of this equation. ///

From (B.1) we see that

$$\int \phi(\xi,\xi',P)\, dW_\xi dW_{\xi'} = \sum_{z\in P} \phi_z W([0,x]\times\delta y) W(\delta x\times[0,y])$$

and thus, we see that

$$\text{(B.8)} \qquad \int \phi(\xi,\xi',P)\,dW_\xi dW_{\xi'} = \int_P \phi_\xi \partial_1 W_\xi \partial_2 W_\xi .$$

By using (B.7) and taking $|P| \to 0$ in (B.8) we find that

$$\text{(B.9)} \qquad \int \phi_{\xi v \xi'}\,dW_\xi dW_{\xi'} = \int \phi_\xi \partial_1 W_\xi$$

for all partition processes ϕ. It is a simple matter to extend the result (B.9) to all processes ϕ in $L_a^2(\Omega\times[0,1]^2,m)$.

$$\text{(b.10)} \qquad \int \phi(\xi,\xi',P)\,d\xi dW_{\xi'} = \sum_{z\in P} \phi_z \times \delta y W(\delta x\ X[0,y])$$

where we have abused our notation slightly by allowing δy to denote the length as well as the set $\delta y = (y,y^+]$. Let $\phi(\xi,P)$ denote the partition process defined by

$$\phi(\xi,P) = \sum_{z\in P} \phi_z x\ \chi_{\delta z}(\xi).$$

It is simple to check that

$$\text{(B.11)} \qquad \int E\{(\phi_\xi \xi_1 - \phi(\xi,P))^2\}d\xi \le 0(|P|). \quad \int E\{\phi_\xi^2\}d\xi.$$

Using the inequality (7.22) of §7 for $h = \phi$ and $\Psi(dx,y) = W(dx\ X[0,y])$, we obtain

$$(B.12)\quad E\{\{(\int\phi(\xi,\xi',P)d\xi dW_{\xi'} - \int\phi(\xi,P)\partial_1 W_\xi d\xi_2)^2\}$$

$$\leq \int E\{\phi(\xi,P)^2\}d\xi \quad 0(|P|).$$

Taking the limit $|P| \to 0$ gives

$$\lim_{|P|\to 0} \int\phi(\xi,P)\partial_1 W_\xi d\xi_1 = \int\phi_{\xi v\xi'}\, d\xi dW_{\xi'}.$$

Let $v(\xi) = \xi_1\xi_2$, then from (B.11) we have that

$$\lim_{|P|\to 0} \int\phi(\xi,P)\partial_1 W_\xi d\xi_1 = \int\phi_\xi \xi_2 \partial_1 W_\xi d\xi_1$$

and thus, we have that

$$(B.13)\qquad \int\phi_{\xi v\xi'}\, d\xi dW_{\xi'} = \int\phi_\xi \xi_2 \partial_1 W_\xi d\xi_1.$$

Similarly, from (B.3) one can deduce that

$$(B.14)\qquad \int\phi_{\xi v\xi'}\, dW_\xi d\xi' = \int\phi_\xi \xi_1 \partial_2 W_\xi d\xi_1.$$

Finally, it is a simple calculation to show that

$$(B.15)\qquad \int I(\xi\wedge\xi')\phi_{\xi v\xi'}\, d\xi d\xi' = \int\phi_\xi \xi_1\xi_2 d\xi$$

where both integrals are Lebesgue integrals.

Thus, we see that the integrals defined by Wong and Zakai in [3] agree with the integrals defined in this paper. Therefore, the Wong-Zakai differentiation formula given below is equivalent to (3.11). (Wong-Zakai differentiation formula for Wiener process W)

$$f(W_z) = f(0) + \int_{[0,z]} f'(W_\xi)dW_\xi + \int_{[0,z]} f''(W_\xi)d\xi$$

$$+ \int_{[0,z]\times[0,z]} f'''(W_{\xi v_{\xi'}})dW_\xi dW_{\xi'} + \frac{1}{2}\int_{[0,z]\times[0,z]} f'''(W_{\xi v\xi'})d_\xi dW_{\xi'}$$

$$+ \frac{1}{2}\int_{[0,z]\times[0,z]} f'''(W_{\xi v\xi'})dW_\xi d\xi'$$

$$+ \frac{1}{4}\int_{[0,z]\times[0,z]} f''''(W_{\xi v\xi'})I(\xi\wedge\xi')d\xi d\xi'$$

By a similar argument one can show that (3.11) is equivalent to the Wong-Zakai formula in the more general case of the process M is defined as in (3.3).

Cairoli-Walsh

Cairoli-Walsh (1975) prove a stochastic Green's formula

and then apply it to prove a two dimensional version of Ito's formula, namely*

$$(B.16)\quad f(W_z) = f(0) + \int_{[0,z]} f'(W)dW + \int_{[0,z]} f''(W)dJ$$

$$- \frac{1}{2}\int_{[0,z]} [F''(W) + \frac{uv}{2}f''''(W)]dudv - \frac{1}{2}\int_{[0,z]} f''(W)(udv-vdu).$$

The first integral corresponds to our $\partial_1\partial_2 W$ integral, that is, the Ito-type stochastic integral. Cairoli and Walsh derive the integrals in terms of the multiple Wiener integral as in Wong-Zakai (1974, 1975, 1976). Thus, it corresponds to our $\partial_1 W\partial_2 W$ integral. In fact, Cairoli and Walsh also show that dJ is a measure corresponding to the product, $\partial_1 W\partial_2 W$, which has the same meaning as in this work**. The remaining integrals in (B.16) are (almost surely) Lebesgue integrals, the last one being a line integral over the boundary of the rectangle [0,z].

Although (B.16) does not resemble (3.11), we can show (3.11) is equivalent to another equation of [4] which is equivalent to (B.16). Thus, in a preceding step of their argument Cairoli and Walsh present the following equation equivalent to (B.16).

---

*This is (6.22), p. 155 of Cairoli and Walsh (1975).

**See page 147, §6 of Cairoli-Walsh (1975).

$$(B.17)^* \qquad f(W_z) = f(0) + \int_{[0,z]} f'(W)\,dW + \int_{[0,z]} f''(W)\,dJ + \int_0^{z_2} \left[\int_0^{z_1} \frac{u}{2} f'''(W_{uv})\,d_u W_{uv}\right] dv + \frac{1}{2}\int_0^{z_2} t f''(W_{ut})\,du$$

The mixed integral is the same as our $\partial_1 W_{uv}\,dv$ integral, and by applying Ito's formula to the last integral one obtains

$$\frac{1}{2}\int_0^{z_2} t f''(W_{ut})\,du = \frac{1}{2}\left\{\int_0^{z_1} v f'''(W_{uv})\,\partial_2 W_{uv}\right\} du + \int_0^{z_2}\int_0^{z_1} \left[\frac{1}{2} f''(W_{uv}) + \frac{1}{4} f''''(W_{uv})\,uv\right] du\,dv.$$

Substituting this expression back into (B.17), one sees that (B.17) is equivalent to (3.11) and hence, (3.11) is equivalent to (B.16).

---

* (B.17) is equation (6.20), p. 154 in Cairoli-Walsh (1975).

REFERENCES

1. Bartlett, M. S., The statistical analysis of spatial pattern, Chapman and Hall, London, 1975.
2. Batchelor, G. K., The theory of homogeneous turbulence, University Press, Cambridge, England, 1953.
3. Bauer, H., Probability Theory and Elements of Measure Theory, Holt, Rinehart and Winston, N. Y., 1971.
4. Bellman, R., Dynamic Programming, Princeton Univ. Press, Princeton, 1957.
5. Bertele, U. and Brioschi, F., Nonserial dynamic programming, Acad. Press, N. Y., 1972.
6. Besieris, I. M. and Kohler, W. E., Underwater Sound Wave Propagation in the Presence of a Randomly Perturbed Sound Speed Profile. Part I, SIAM J. Appl. Math. 34(3):423-436, 1978.
7. Bochner, S., Partial ordering in the theory of martingales, Ann. Math. 62(1), 1955 b .
8. Bose, N. K., Problems and Progress in Multidimensional Systems, Proc. IEEE 65(6):824-840, 1977.
9. Cairoli, R., Une inegalité pour martingales à indices multiples et ses applications, Séminaire de probabilitésIV, Univ. Strasbourg, Springer-Verlag, Berlin, 1-27, 1970.

10. Cairoli, R., Décomposition de processus à indices doubles, Séminaire de probabilités V, Université de Strasbourg, Springer, Berlin, 1971.

11. Cairoli, R. and Walsh, J. B., Stochastic integrals in the plane, Acta Mathematica, 134, pp. 111-183, 1975.

12. Cairoli, R. and Walsh, J. B., Martingale representations and holomorphic processes, Ann. Prob. 5(4), 511-521, 1977.

13. Cameron, M. A. and Hannan, E. J., Measuring properties of plane waves, Math. Geology, 10(1):1-22, 1978.

14. Cameron, R. H. and Martin, W. T., "The orthogonal development of nonlinear functionals in a series of Fourier-Hermite functions." Ann. of Math. 48, 385-392, 1947.

15. Chan, D. S. K., Computation Issues in the Implementation of 2-D Filters, 1976 Arden House Workshop on Digital Signal Processing, Harriman, N. Y., 1976.

16. Chow, Y. S., "Martingales in a σ-finite measure space indexed by directed sets," T.AMS 97, 254-285, 1960.

17. Csanady, G. T., Turbulent diffusion in the environment, D. Reidel Publ., Dordrecht, Holland, 1973.

18. Dellacherie, Claude, Capacités et processus stochastiques, Springer, Berlin, 1972.

19. Delfiner, P. and Delhomme, J. P., Optimum interpolation by kriging, in Display and analysis of spatial data, J. C. Davis and M. J. McCullagh (eds), Wiley, pp. 96-114, 1975.

20. Dence, D. and Spence, J. E., Wave propagation in random anisotropic media, in A. T. Bharucha-Reid, Probabilistic Methods in Applied Mathematics, v. 3, Acad. Press, N. Y., 1973.

21. Doob, J. L., Stochastic Processes, John Wiley and Sons, N. Y., 1953.

22. Epstein, E. S., Stochastic dynamic preduction, Tellus 21(6):739-759, 1969 b .

23. Fel'dbaum, A. A., "Dual Control Theory. I." Automation and Remote Control, 21, 9, pp. 874-880, Engl. trans. from Russian Avtomatika i Telemekhanika, Vol. 21, No. 9, 1240-1249, 1960.

24. Fraedrich, K. and Böttger, H., A wavenumber frequency analysis of the 500 mb geopotential at 50°N., J. Atm. Sci. 35:745-750, 1978.

25. Frisch, U., Wave propagation in random media, pp. 76-198 in A. T. Bharucha-Reid, Probabilistic Methods in Applied Mathematics, V. 1, Acad. Press, N. Y., 1968.

26. Gelfand, I. M. and Vilenkin, N. Ya, Applications of Harmonic Analysis, Generalized Functions, Vol. 4, Acad. Press, N. Y., 1964.

27. Haggstrom, G. W., "Optional stopping and experimental design." Ann. Math. Stat. 37: 7-29, 1966.

28. Hannan, E. J., Group representations and applied probability, J. App. Prob. 2, 1-68, 1965.

29. Hannan, E. J., The concept of a filter, Proc. Camb. Phil. Soc. 63, 221-227, 1967.

30. Hannan, E. J., Fourier methods and random processes, Bull. Int. Statist. Inst. 42, 475-496, 1969.

31. Harbaugh, J. W., Doveton, J. J., David, J. C., "Proability Methods in Oil Exploration," John Wiley, N. Y., 1977.

32. Ho, Y. C. and Chu, K. C., Team Decision Theory and Information Structures in Optimal Control Problems: Part I., IEEE Trans. Aut. Contr., AC-17, 1972.

33. Ito, K., Multiple Wiener integral, J. Math. Soc. of Japan, 3(1), 157-169, 1951.

34, Kallianpur, G. and Mandrekar, V., The Markov property for generalized Gaussian random fields, Ann. Inst. Fourier, Grenoble 24(2):143-167, 1974.

35. Kemeny, J. G., Snell, J. L. and Knapp, A. W., Denumerable Markov Chains, Springer-Verlag, N. Y., 1976.

36. Kunita, J. and Watanabe, S., "On square integrable martingales", Nagoya Math. J., 30, 209-245, 1967.

37. Krickeberg, K., Convergence of martingales with a directed index set, Trans. Amer. Math. Soc., 83, 313-337, 1956.

38. Kurtz, T. G., The optional sampling theorem for martingales indexed by directed sets, Séries de mathématiques pures et appliquées, Univ. Louis Pasteur, Strasbourg, 1977.

39. Kurtz, T. G., Representations of Markov processes as multiparameter time change, Séries de mathématiques pures et appliquées, Institut de recherche mathematique avancée, laboratoire associe au C.N.R.S. N^{O} 1, dept. de math. Univ. Louis Pasteur, Strasbourg, 1978.

40. Larrimore, W. E., Statistical inference on stationary random fields, Proc. IEEE, 65(6), pp. 961-976, 1977.

41. Lévy, P., A special problem of Brownian motion, and a general theory of Gaussian random functions, in Proc. 3rd Berkeley Symposium on Math. Stat. and Prob., 2:133-175, U. C. Press, Berkeley, 1956.

42. Longshaw, T. C. and Gilbertson, B., Multispectral aerial photography as an exploration tool - III: two applications in the North-Western Cape Province South Africa, Remote Sensing of Environment, 5(1), 1976.

43. Lumley, J. L., Stochastic tools in Turbulence, Acad. Press, N. Y., 1970.

44. MacLane, S. and Birkhoff, G., Algebra, MacMillan, N. Y., 1967.

45. Marschak, J. and Radner, R., Economic Theory of Teams, Yale Univ. Press, New Haven, 1972.

46. Matheron, G., Les variables régionalisées et leurs estimation (une application de la theorie des fonctions aléatoires aux sciences de la nature), Mason and Cie, Paris, 305 pp., 1965.

47. McKean, H. P., Brownian motion with a several dimensional time, Theory Prob. and Appl. 3(4):335-354, 1963.

48. Merriam, D. F., Geostatistics: a colloquium, Plenum Press, N. Y., 1970.

49. Meyer, P. A., Probability and Potentials, Blaisdell, N. Y., 1966.

50. Monin, A. S. and Yaglom, A. M., Statistical Fluid Mechanics: Mechanics of Turbulence, V. 1, MIT Press, Cambridge, Mass., 1971.

51. Monin, A. S. and Yaglom, A. M., Statistical Fluid Mechanics: Mechanics of Turbulence, V. 2, MIT Press, Cambridge, Mass., 1975.

52. Nash, R. A. Jr. and Jordan, S. K., Statistical geodesy - an engineering perspective. Proc. IEEE, Vol. 66(5), 532-550, 1978.

53. Park. W. J., A multiparameter Gaussian process, Ann. Math. Stat. 41:1582-1595, 1970.

54. Preston, C. J., Gibbs states on countable sets, Cambridge U. Press, Cambridge, 1974.

55. Radner, R., Team decision problems, Ann. Math. Stat. 33:857-881, 1962.

56. Rhenals-Figueredo, A. E., Rodriguez-Iturbe, and Schaake, J. C. Jr., Bidimensional spectral analysis of rainfall events, MIT Ralph M Parsons lab. report 193, Cambridge, Mass. 1974.

57. Rosenfeld, A. and Kak, A. C., Digital picture processing, Acad. Press, N. Y., 1976.

58. Sandell, N. R. and Athans, M., Relativistic Control Theory and the Dynamic Control of Communication Networks, Proc. of the 1975 Conference on Decision and Control, 451-455, 1975.

59. Snell, J. L., "Applications of martingale system theorems," Trans. Am. Math. Soc. 73:293-312, 1952.

60. Spitzer, F., Random fields and interacting particle systems, Math. Assoc. Amer., 1971.

61. Tatarskaya, M. S., Spatial statistical structure of geopotential fields, Izv. Atmos. and Oceanic Phys. 1(7):437-439, 1965.

62. Taylor, G. I., The spectrum of turbulence, Proc. Roy. Soc. London, A164, 476-490, 1938.

63. Von Neumann, J. E. and Morgenstern, O., Theory of Games and Economic Behavior, Wiley, N. Y., 1944.

64. Whittle, P., On stationary processes in the plane, Biometrika, 41:434-449, 1954.

65. Willsky, A. S., Digital Signal Processing and Control and Estimation Theory - Points of Tangency, Area of Intersection, and Parallel Directions, Electronics Systems Lab, Dept. of Elec. Eng. and Comp. Sci., M.I.T., to be published, MIT Press, 1976.

66. Witsenhausen, H., A counterexample in stochastic optimum control, SIAM J. Contr. 6:131-147, 1968.

67. Witsenhausen, Hans S., "On information structures, feedback and causality," SIAM J. Control, 9, 2, 149-160, 1971.

68. Woods, J. W., Two-dimensional Markov spectral estimation, IEEE Trans. on Inf. Th. IT-22(5): 552-559, 1976.

69. Woods, J. W. and Radewan, C. J., Kalman filtering in two dimensions, IEEE Trans. on Inf. Th. IT-23(4):473-482, 1977.

70. Wong, E., Two-dimensional random fields and representation of images, SIAM J. Appl. Math. 16(4):756-770, 1968.

71. Wong, E., Homogeneous Gauss-Markov random fields, Ann. Math. Stat. 40:1625-1634, 1969.

72. Wong, E., Sotachastic processes in information and dynamical systems, McGraw-Hill, N. Y., 1971.

73. Wong, E., Detection and filtering for two-dimensional random fields, Proc. IEEE Conference on Decision and Control, 1976.

74. Wong, E., Recursive causal linear filtering for two-dimensional random fields, IEEE Trans. on Inf. Th. IT-24(1): 50-59, 1978.

75. Wong, E. and Zakai, M., Martingales and stochastic integrals for processes with a multidimensional parameter, Z. Wahrscheinlichkeitstheorie und Verw. Gebeite 29, 1974.

76. Wong, E. and Zakai, M., Differentiation formulas for stochastic integrals in the plane, Mem. ERL-M540, Electronics Research Lab., Berkeley, 1975 a .

77. Wong, E. and Zakai, M., Weak martingales and stochastic integrals in the plane, Mem. ERL-M496, Electronics Research Lab, Berkeley, 1975 b .

78. Wong, E. and Zakai, M., "Weak martingales and stochastic integrals in the plane", Ann. Prob. 4(4), 570-586, 1976.

79. Yaglom, A. M., "Second Order Homogeneous random fields". Proc. 4th Berkeley Symposium on Math. Stat. and Prob., Berkeley, L. A. U. of Ca. Press, V. 2, 593-622, 1961.

80. Yaglom, A. M., An introduction to the theory of stationary random functions, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

81. Yor, Marc, Représentation des martingales de carré intégrable relative aux processus de Wiener et de Poisson à n-paramètres. C. R. Acad. Sci. Paris, t. 281, (21 juillet 1975) série A-111, no. 2-3, 1975.

BIOGRAPHICAL NOTE

Robert B. Washburn, Jr. was born on January 30, 1951, in Bryn Mawr, Pennsylvaina. He attended Conestoga High School in Berwyn, Pennsylvania and graduated from Yale University in 1973 with the Bachelor of Science degree in mathematics.

The author started graduate work at the Massachusetts Institute of Technology in the fall of 1973. In 1973-1974 and 1974-1975 he held a M.I.T. Applied Mathematics Fellowship. In 1974 through 1977 he was a teaching assistant in the Department of Mathematics. Since 1978 he has been employed at Scientific Systems, Inc. in Cambridge, Massachusetts.