Stochastic Optimization in Continuous Time

Fwu-Ranq Chang

CAMBRIDGE

Stochastic Optimization in Continuous Time

FWU-RANQ CHANG
Indiana University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2004

Printed in the United States of America

Typeface Times New Roman PS 10.5/13 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Chang, Fwu-Ranq, 1947–
Stochastic optimization in continuous time / Fwu-Ranq Chang.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-83406-6
1. Economics–Mathematical models. 2. Stochastic control theory. I. Title.
HB135.C444 2004
330′.01′51923–dc22 *2003061745*

ISBN 0 521 83406 6 hardback

Preface

This is an introduction to stochastic control theory with applications to economics. There are many texts on this mathematical subject; however, most of them are written for students in mathematics or in finance. For those who are interested in the relevance and applications of this powerful mathematical machinery to economics, there must be a thorough and concise resource for learning. This book is designed for that purpose. The mathematical methods are discussed intuitively whenever possible and illustrated with many economic examples. More importantly, the mathematical concepts are introduced in language and terminology familiar to first-year graduate students in economics.

The book is, therefore, at a second-year graduate level. The first part covers the basic elements of stochastic calculus. Chapter 1 is a brief review of probability theory focusing on the mathematical structure of the information set at time t, and the concept of conditional expectations. Many theorems related to conditional expectations are explained intuitively without formal proofs.

Chapter 2 is devoted to the Wiener process with emphasis on its irregularities. The Wiener process is an essential component of modeling shocks in continuous time. We introduce this important concept via three different approaches: as a limit of random walks, as a Markov process with a specific transition probability, and as a formal mathematical definition which enables us to derive and verify variants of the Wiener process. The best way to understand the irregularities of the Wiener process is to examine its sample paths closely. We devote substantial time to the zero sets of the Wiener process and the concept and examples of stopping times. It is the belief of the author that one cannot have a good grasp of the Wiener process without a decent understanding of the zero sets.

In Chapter 3 we define the stochastic integrals, discuss stochastic differential equation, and examine the celebrated Ito lemma. The Ito integral is defined as the limit of Riemann sums evaluated only at the left endpoint of each subinterval and hence is not a Riemann integral. However, this formulation fits economic reasoning very well, because under uncertainty future events are indeed nonanticipating. It is similar to the discrete-time formulation in which an economic agent makes a decision at the beginning of a time period and then subjects herself to the consequences after the state of nature is revealed. These mathematical tools enable us to study the Black–Scholes option pricing formula and issues related to irreversible investment. To make the presentation self-contained, we include a brief discussion of the heat equation and Euler's homogeneous equation. More importantly, we caution the reader throughout the chapter that some of results and intuitions cherished by economists may no longer be true in stochastic calculus.

The second part of the book is on the stochastic optimization methods and applications. In Chapter 4 we study the Bellman equation of stochastic control problems; a set of sufficient conditions, among them the transversality condition, for verifying the optimal control; and the conditions for the existence and differentiability of the value function. We guide the reader through a step-by-step argument that leads to the Bellman equation. We apply this solution method to many well-known examples, such as Merton's consumption and portfolio rules, demand for index bonds, exhaustible resources, the adjustment-cost theory of investment, and the demand for life insurance. We also derive the Bellman equation for a certain class of recursive utility functions to broaden the scope of applications to models with variable discount rates. Most of all, we wish to show that setting up the Bellman equation for a stochastic optimization problem in continuous time is as easy as setting up the Lagrange function in a static, constrained optimization problem. We hope applied economists will find this powerful tool readily accessible.

In Chapter 5 we discuss various methods of finding a closed-form representation for the value function of a stochastic control problem. In many economic problems, the functional form of the value function is absolutely essential to ascertain the optimal policy functions. We present the commonly employed methods systematically, from brute force to educated guess. Some of the problems are solved using more than one method so that the reader can compare the method's strengths and weaknesses. We also introduce the *inverse optimum* methodology that enables us to ascertain the underlying economic structure from the observed policy functions. The chapter title, "How to Solve It," which is borrowed from Pólya's book, summarizes the spirit of the presentation. We hope researchers will find this chapter useful.

In Chapter 6 we investigate two classes of economic problems related to the boundaries of a controlled diffusion process. The first class of problems relates to the nonnegativity constraint, which is not addressed in the mathematical literature. Specifically, the mathematical solution to a controlled diffusion process assumes values on the whole real line, while the economic variables such as consumption and capital–labor ratios cannot be negative. We introduce several approaches to address this issue. As an example, we employ a reflection method to show that the capital–labor ratio in the stochastic Solow equation can never become negative. The second class of problems uses the optimal stopping time technique. We show the reader how to formulate and solve this type of problem through two well-known economic models: precautionary and transactions demand for money, and the tree-cutting problem. We also show that, even though the optimal policy function is implicitly defined, comparative dynamics can still be performed if we do the mathematics right.

The book includes many exercises, which follow immediately after each topic so that the reader can check and practice their understanding of the subject matter. Many of them are provided with useful hints; those, however, should be used only after an honest attempt has been made. Notes and suggested readings are provided at the end of each chapter for more relevant references and possible extensions. The "Miscellaneous Applications and Exercises" in the Appendix provide the reader with more applications to economics and can also be used as review exercises on stochastic optimization methods.

Acknowledgement

This book has grown out of a graduate course in mathematical economics I gave at Indiana University, Bloomington. A draft of this book was presented at the Center for Economic Studies of the University of Munich in 1999. Feedback from many experts at these two universities has greatly improved the presentation of the book. In particular, I would like to thank Jinwon Ahn, Manoj Atolia, Bob Becker, John Boyd, Hess Chung, Franz Gehrels, Slava Govoroi, Sumit Joshi, Peter Pedroni, and Pravin Trivedi.

Most of all, I would like to thank Buz Brock for introducing me to this interesting subject and for his intellectual inspiration, and Tassos Malliaris for his many suggestions, encouragement, and generosity, which played a very important role in getting this book off the ground. I would also like to thank Indiana University for financial support in the preparation of this book. Finally, I would like to thank my family for their support through thick and thin.

Contents

List of Figures

1

Probability Theory

1.1 Introduction

In this chapter we introduce probability theory using a measure-theoretic approach. There are two main subjects that are closely related to economics. First, the concept of σ-algebra is closely related to the notion of information set used widely in economics. We shall formalize it. Second, the concepts of conditional probability and conditional expectation are defined in terms of the underlying σ-algebra. These are background materials for understanding Wiener processes and stochastic dynamic programming.

We keep proofs to the bare minimum. In their place, we emphasize the intuition so that the reader can gain some insights into the subject matter. In fact, we shall go over many commonly employed theorems on conditional expectation with intuitive explanations.

1.2 Stochastic Processes

1.2.1 Information Sets and σ-Algebras

Let Ω be a point set. In probability theory, it is the set of elementary events. The power set of Ω, denoted by 2^{Ω}, is the set of all subsets of Ω. For example, if the experiment is tossing a coin twice, then the set Ω is $\{HH, HT, TH, TT\}$. It is easy to write down all $2^4 = 16$ elements in the power set. Specifically,

$$\begin{aligned} 2^{\Omega} = \{&\varnothing, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \{HH, HT\}, \{HH, TH\}, \\ &\{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}, \{HH, HT, TH\}, \\ &\{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\}, \Omega\}. \end{aligned}$$

In general, the cardinality of the power set is $2^{|\Omega|}$, where $|\Omega|$ is the cardinality of the set Ω. Power sets are very large. To convince yourself, let the experiment be rolling a die twice, a rather simple experiment. In this simple experiment, $|\Omega| = 36$ and the cardinality of the power set is $2^{36} = 6.87 \times 10^{10}$. It would be impractical to write down all elements in this power set. What we are interested in is subsets of the power set with certain structure.

Definition 1.1 *A class $\mathcal{F}$ of subsets of Ω, i.e., $\mathcal{F} \subset 2^{\Omega}$, is an algebra (or a field) if:*

(i) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$, where A^c is the complement of A in Ω.
(ii) $A, B \in \mathcal{F}$ imply that $A \cup B \in \mathcal{F}$.
(iii) $\Omega \in \mathcal{F}$ (equivalently, $\varnothing \in \mathcal{F}$).

Conditions (i) and (ii) imply $A \cap B \in \mathcal{F}$, because $A \cap B = (A^c \cup B^c)^c$.

Definition 1.2 *A class $\mathcal{F}$ of subsets of Ω is a σ-algebra if it is an algebra satisfying*

(iv) if $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The Greek letter "σ" simply indicates that the number of sets forming the union is *countable* (including finite numbers).

Any $A \in \mathcal{F}$ is called a *measurable* set, or simply, an $\mathcal{F}$-set. We use $\mathcal{F}$ to represent the *information set*, because it captures our economic intuition. Conditions (i) through (iv) provide a mathematical structure for an information set.

Intuitively, we can treat a measurable set as an *observable set*. An object under study ($\omega \in \Omega$) is observable if we can detect that it has certain characteristics. For example, let Ω be the set of flying objects and let A be the set of flying objects that are green. Then A^c represents the set of all flying objects that are not green. Condition (i) simply says that if, in our information set, we can observe that a flying object is green (i.e., A is observable), then we should be able to observe that other flying objects are not green. That means A^c is also observable. Another example is this: if we were able to observe when the general price level is rising, then we should be able to observe when the general price level is not rising. Formally, if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

Condition (ii) says that, if we can observe the things or objects described by characteristics A and those described by characteristics B, then we should be able to observe the objects characterized by the properties of A or B. That is, $A, B \in \mathcal{F}$ imply $A \cup B \in \mathcal{F}$. For example, if we are able to observe when the price level is rising, and if we are able to observe the unemployment level is rising, then we should be able to observe the rising of price level *or* rising unemployment. The same argument applies to countably many observable sets, which is condition (iv). These mathematical structures make σ-algebras very suitable for representing information.

It is clear that the power set 2^{Ω} is itself a σ-algebra. But there are lots of σ-algebras that are smaller than the power set. For example, in the experiment of tossing a coin twice, $\mathcal{F}_1 = \{\Omega, \varnothing, \{HH\}, \{HT, TH, TT\}\}$ and $\mathcal{F}_2 = \{\Omega, \varnothing, \{HH, TT\}, \{HT, TH\}\}$ are both algebras. The information content of $\mathcal{F}_1$ is this: we can tell whether tossing a coin twice ends up with both heads or otherwise. The information content of $\mathcal{F}_2$ is this: we can tell whether both tosses have the same outcome or not. The reader should try to find other algebras in this setup. An obvious one is to "combine" $\mathcal{F}_1$ and $\mathcal{F}_2$. See the exercise below. We will return to these two examples in Example 1.12.

Exercise 1.2.1

(1) Verify that $\mathcal{F}_1$ and $\mathcal{F}_2$ are algebras.

(2) Show that $\mathcal{F}_1 \cup \mathcal{F}_2$, while containing $\mathcal{F}_1$ and $\mathcal{F}_2$, is not an algebra.

(3) Find the smallest algebra $\mathcal{G}$ that contains $\mathcal{F}_1$ and $\mathcal{F}_2$ in the sense that for any algebra $\mathcal{H}$ which contains $\mathcal{F}_1$ and $\mathcal{F}_2$, then $\mathcal{G} \subset \mathcal{H}$.

Definition 1.3 *A set function $P : \mathcal{F} \to \mathbb{R}$ is a probability measure if P satisfies*

(i) $0 \le P(A) \le 1$ for all $A \in \mathcal{F}$;

(ii) $P(\varnothing) = 0$ and $P(\Omega) = 1$;

(iii) if $A_i \in \mathcal{F}$ and the A_i's are mutually disjoint, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Property (iii) is called *countable additivity*. The triplet $(\Omega, \mathcal{F}, P)$ is used to denote a probability space.

Example 1.4 *(Borel Sets and Lebesgue Measure) When $\Omega = \mathbb{R}$ (the whole real line) or $\Omega = [0, 1]$ (the unit interval), and the σ-algebra* *is the one generated by the open sets in $\mathbb{R}$ (or in $[0, 1]$), we call this σ-field the Borel field. It is usually denoted by $\mathcal{B}$. An element in the Borel field is a Borel set.*

Examples of Borel sets are open sets, closed sets, semi-open, semi-closed sets, F_σ sets (countable unions of closed sets), and G_δ sets (countable intersections of open sets). When $\Omega = [0, 1]$, $\mathcal{B}$ is the σ-algebra, and $P(A)$ is the "length" (measure) of $A \in \mathcal{F}$, we can verify that P is a probability measure on $\mathcal{B}$. Such a measure is called the Lebesgue measure on $[0, 1]$.

However, not all subsets of $\mathbb{R}$ are Borel sets, i.e., not all subsets of $\mathbb{R}$ are observable. For example, the Vitali set is not a Borel set. See, for example, Reed and Simon (1972, p. 33). For curious souls, the Vitali set V is constructed as follows. Call two numbers $x, y \in [0, 1)$ equivalent if $x - y$ is rational. Let V be the set consists of exactly one number from each equivalent class. Then V is not Lebesgue measurable.

A single point and, therefore, any set composed of countably many points are of Lebesgue measure zero. The question then is this: Are sets with uncountably many points necessarily of positive Lebesgue measure? The answer is negative, and the best-known example is the Cantor set.

1.2.2 The Cantor Set

Since the Cantor set contains many important properties that are essential to understanding the nature of a Wiener process, we shall elaborate on this celebrated set. The construction of the Cantor set proceeds as follows. Evenly divide the unit interval $[0, 1]$ into three subintervals. Remove the middle *open* interval, $(1/3, 2/3)$, from $[0, 1]$. The remaining two closed intervals are $[0, 1/3]$ and $[2/3, 1]$. Then remove the two middle open intervals, $(1/9, 2/9)$ and $(7/9, 8/9)$, from $[0, 1/3]$ and $[2/3, 1]$ respectively. Continue to remove the four middle open intervals from the remaining four closed intervals, $[0, 1/9]$, $[2/9, 1/3]$, $[2/3, 7/9]$, and $[8/9, 1]$, and so on indefinitely. The set of points that are not removed is called the *Cantor set*, $\mathfrak{C}$.

Any point in the Cantor set can be represented by

$$\sum_{n=1}^{\infty} \frac{i_n}{3^n}, \qquad \text{where } i_n = 0 \text{ or } 2.$$

For example,

$$\frac{7}{9}=\frac{2}{3}+\frac{0}{9}+\frac{2}{27}+\frac{2}{81}+\frac{2}{243}+\cdots,$$

i.e., $7/9=(2,0,2,2,2,\dots)$. Similarly, $8/9=(2,2,0,0,\dots)$, $0=(0,0,0,\dots)$, $8/27=(0,2,2,0,0,0,\dots)$, and $1=(2,2,2,\dots)$. Therefore, the cardinality of the Cantor set is that of the continuum. Since the Lebesgue measure of the intervals removed through this process is

$$\frac{1}{3}+\frac{1}{9}\cdot 2+\frac{1}{27}\cdot 4+\cdots=\sum_{n=1}^{\infty}\frac{2^{n-1}}{3^n}=1,$$

the Cantor set must be of Lebesgue measure zero.

The main properties that are of interest to us are three. First, here is a set with uncountably many elements that has a zero Lebesgue measure. Second, every point in the Cantor set can be approached by a sequence of subintervals that were removed. In other words, every point in the Cantor set is a limit point. Such a set is called a *perfect set*. Third, for any interval $I\subset[0,1]$, it must contain some subinterval that was eventually removed, i.e., we can find a subinterval $J\subset I$ such that J and the Cantor set $\mathfrak{C}$ are disjoint: $J\cap\mathfrak{C}=\varnothing$. That is, $\mathfrak{C}$ is *nowhere dense* in $[0,1]$. These three properties are the basic features of the zero set of a Wiener process, as we shall see later.

1.2.3 Borel–Cantelli Lemmas

Definition 1.5 *The limit superior and the limit inferior of a sequence of sets* $\{A_n\}$ *are*

$$\limsup_{n\to\infty} A_n=\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k,$$

$$\liminf_{n\to\infty} A_n=\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}A_k.$$

Simply put, $x\in\limsup_{n\to\infty}A_n$ means x belongs to infinitely many A_k. In contrast, $x\in\liminf_{n\to\infty}A_n$ means x belongs to virtually all A_k, in the sense that there exists N such that $x\in A_k$ for $k\geq N$. Since $\mathcal{F}$ is a