INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

CONTROLLED MARKOV CHAINS WITH EXPONENTIAL RISK-SENSITIVE CRITERIA: MODULARITY, STRUCTURED POLICIES AND APPLICATIONS

by

Micaela Guadalupe Avila Godoy

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1999

UMI Number: 9957946

UMI Microform 9957946

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Micaela Guadalupe Avila de Brau entitled CONTROLLED MARKOV PROCESSES WITH RISK-SENSITIVE TOTAL AND DISCOUNTED COST CRITERION

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

W.M. Greenlee	10/08/99
W.M. Greenlee	Date
Carl DeVito	10/8/99
Carl DeVito	Date
Robert Maier	10/8/99
Robert Maier	Date
Emmanuel Fernandez	8 Oct '99
Emmanuel Fernandez	Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

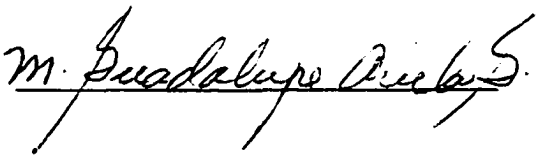
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

W. M. Greenlee	10/08/99
Dissertation Director W.M. Greenlee	Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: 

ACKNOWLEDGMENTS

I want to specially thank my advisor, Professor Emmanuel Fernández Gaucherand, for his invaluable guidance and effort in helping me see this work to its completion. Professor Fernández exemplifies the meaning of advisor, teacher and friend.

Also, I want to thank Professors Robert Maier, Moshe Shaked and Carl DeVito for devoting their time to serve in my dissertation committee.

I am specially grateful to the other member of my dissertation committee, Professor W.M. Greenlee, for his kind disposition, advice and support.

Perhaps most importantly of all, I want to thank my sons, Agustín, Ernesto and Julio, for having relinquished many hours of my time that rightfully belonged to them. Without their love and comprehension, the completion of this project would not have been possible.

Finally, I want to express my deepest thanks to Agustín, my husband, for the long hours of discussion we shared during my doctoral program, for his helpful comments and suggestions, and most of all for his love. I will be forever grateful for the gift of his presence in my life.

DEDICATION

A mi madre, Inés Godoy de Avila, con mucho cariño.

A mis hijos, Agustín, Ernesto y Julio, con todo mi amor.

Al compañero de mi vida, Agustín, por todo lo que ha significado para mi. en mi formación profesional.

A Jesus y Lupita, por la solidaridad que me brindaron en momentos difíciles durante mi programa de doctorado.

A todos mis hermanos.

Table of Contents

TABLE OF CONTENTS—*Continued*

ABSTRACT

Controlled Markov chains (CMC's) are mathematical models for the control of sequential decision stochastic systems. Starting in the early 1950's with the work of R. Bellman, many basic contributions to CMC's have been made, and numerous applications to engineering, operation research, and economics, among other areas, have been developed.

The optimal control problem for CMC's with a countable state space, and with a general action space, is studied for (exponential) total and discounted risk-sensitive cost criteria. General (dynamic programming) results for the finite and the infinite horizon cases are obtained. A set of general conditions is presented to obtain structural properties of the optimal value function and policies. In particular, monotonicity properties of value functions and optimal policies are established. The approach followed is to show the (sub)modularity of certain functions (related to the optimality equations). Four applications studies are used to illustrate the general results obtained is this dissertation: equipment replacement, optimal resource allocation, scheduling of uncertain jobs, and inventory control.

Chapter 1

INTRODUCTION

1.1 Motivation

It is often necessary to study sequential decision processes for stochastic systems, i.e., systems with the following characteristics. The system is dynamic: it evolves in time and its state is observed at each of a sequence of points in time; the decision making is sequential: a decision is made (an action is applied) each time the state of the system is observed; and the behavior of the system is stochastic: the observed state and the applied action at a decision epoch do not completely determine the state at the next decision epoch, but only the probability distribution of that state. Controlled Markov chains (CMC's), the type of stochastic models studied in this dissertation, capture the essential characteristics of sequential decision stochastic systems.

In CMC's, a cost is incurred at each decision epoch, which depends on the current state and the chosen action. To evaluate the overall performance of the system, the Decision Maker (DM) selects a functional of the aggregated costs per stage (a cost criterion). A policy or strategy is a rule or prescription that gives the actions to be applied at each decision epoch under any eventuality, i.e., for each possible state. The optimal control problem is then formulated as the problem of finding a policy that minimizes the selected cost criterion. Most of the literature on the subject consider risk neutral cost criteria, that is, criteria that do not take into account the DM's attitude toward risk with respect to the randomness of the system, e.g., the variance of the cost process. However, in many areas of application, the DM may wish to incorporate that attitude to the cost criterion when choosing a policy. Various approaches to the formulation of risk sensitive criteria for CMC's have been proposed in the literature. Herein, we consider the formulation of the risk-sensitive criteria

based on von Neumann- Morgenstern expected utility theory. For that approach, it is assumed that the DM's risk preferences are represented by a utility function $\mathcal{U} : \mathbb{R} \longrightarrow \mathbb{R}$. We focus in particular on the class of exponential utility functions, which, as we will see, have appealing decision-theoretic properties.

In this dissertation, a study of CMC's with *exponential total cost* (ETC) and *exponential discounted cost* (EDC) as risk-sensitive performance criteria is presented. Our study is closely related to that of Cavazos-Cadena and Fernández-Gaucherand [13, 14, 15, 16, 17], who have recently made important contributions to CMC's with an *exponential average cost criterion* and its relation with the EDC criterion; see also [2, 3, 4, 10, 11, 22, 26] for other recent results. We consider the model with a countable state space, a Borel action space and bounded costs per stage. General (dynamic programming) optimality results, such as optimality equations and value iteration schemes, are obtained.

It is well known that "explicit" solutions arising from the optimality equations are obtained only in a few cases, even in risk-neutral models. This fact has motivated the study of "structural" properties of the value function and the optimal policies: it is clearly useful, for analysis and computations, to know when a "structured" optimal policy exists, because then the search for an optimal policy can thus be restricted to the much smaller subclass of such "structured" policies. While there is a vast literature dealing with structured models in the risk-neutral case (see [1, 30, 40, 41, 42, 43, 44, 46] and references therein), only a few (and recent) contributions have been made on that subject in the risk-sensitive context; see [2, 9, 22, 26].

In this dissertation, we study CMC's structured models, and general conditions inducing structural properties of the exponential optimal value function are presented. In particular, monotonicity properties of the value function and optimal policies are established. We extend an approach frequently used in the risk-neutral context, to obtain monotone optimal policies to the risk-sensitive context. The key point in this approach is to show the (sub)modularity of some functions related to the optimality

equation. In order to follow this approach, basic results are developed concerning the product of (sub)modular functions.

The advantageous use of the specific structure of different classes of problems leads to specific structural results, e.g., the existence of threshold optimal policies. In this dissertation, we develop structural properties of problems in the following areas: equipment replacement, optimal resource allocation, and inventory control.

The inventory control problem, unlike the other two mentioned applications, is modeled as a CMC with cost per stage function depending not only on the present state-action pair, but also on a "random disturbance" (i.e., the conditional distribution of the one-stage cost given the present state-action pair is non-degenerate). When dealing with risk-neutral criteria, the mentioned dependence can be eliminated by replacing the random costs per stage by their expected values. However, for risk-sensitive exponential criteria, the described procedure for solving within-period randomness does not yield an equivalent control problem. Furthermore, unlike models studied more frequently (see e.g., Bertsekas [6]), ours considers random disturbances whose distribution is given by a stochastic kernel that depends explicitly on the prior disturbance. To further illustrate this model we include an application in the area of jobs scheduling.

1.2 Summary of Results

The organization is as follows: in Chapter 2, the formal definition of the stochastic sequential decision problems studied in this dissertation is introduced. The basic framework for CMC's and the classification of admissible policies are respectively presented in the first and second sections. That material is contained in many books and journals on the subject. Particularly important for my study of those topics were the book by Bertsekas [6], Hernández-Lasserre [29], Puterman [40], Ross [41] and the survey by Arapostathis et al [1]. In Section 2.3, a summary of the basic notions

of expected utility theory used in the subsequent analysis is included. Finally, in Sections 2.4 and 2.5, the risk -sensitive versions of the standard criteria, and the corresponding optimal control problems, are defined.

The ETC and the EDC were first studied in, e.g., [18, 20, 31, 32, 33] for the finite state space model. Results in [18, 20, 31, 33], characterizing the (exponential) optimal value function and policies are extended in Chapter 3, to infinite state space models. Moreover, optimization with respect to the set of randomized, history dependent policies is considered here, whereas only Markovian deterministic policies were considered in [18, 20, 31, 32, 33]; see also [7] for related results. In Section 3.1, we present the exponential version of the policy evaluation algorithm (over a finite horizon), which plays a key role in the rest of the chapter. Then, in the same section, we show that, adequately translated to a (exponential) multiplicative scheme, the (finite horizon) dynamic programming algorithm for the (additive) standard discounted and total cost criteria also holds for exponential criteria. In Section 3.2 (3.3), it is shown that the EDC (ETC) is given by a solution of an *exponential discounted (total) cost optimality equation*, and that deterministic Markovian optimal policies can be obtained from that equation. Finally, it is also shown that the value iteration scheme holds for the EDC (ETC).

Chapter 4 is devoted to CMC's structured models. In Section 4.1, we provide a set of sufficient conditions on the cost and the probabilistic structure of a CMC under which the optimal EDC (or ETC) function is shown to be monotone (as a function of the initial state). In Section 4.2, it is shown that, under certain assumptions, the decision function of the optimal policy at the t-th stage is monotone (as a function of the state), for $t = 0, 1, \ldots$. In Section 4.3, we show that the modularity on the set of admissible state-actions pairs of certain functions (related to the exponential optimality equation) guarantees the existence of monotone optimal policies. Moreover, besides considering modularity with respect to the set of admissible state-actions pairs, as in the risk-neutral case, modularity with respect to $\mathbf{A} \times \{\gamma\beta^t : t = 0, 1, \ldots\}$ is also

considered, leading to structural properties of the optimal policies with respect to t, where $\mathbf{A}$ is the set of actions, γ is the risk-sensitivity coefficient, and β is the discount factor. . Finally, some results about modularity of the product of two functions are established in Section 4.3.

In Chapter 5, we apply results previously obtained in the dissertation to infinite horizon equipment replacement problems with EDC criterion. In Section 5.1, the formulation of the problem as a CMC is presented. In Sections 5.2 and 5.3, respectively, it is shown that under standard conditions, the optimal policy is of the threshold-type and ultimately stationary. Moreover, under mild additional conditions, it is also proved in Section 5.3 that ultimately, it is optimal to follow the risk-neutral stationary optimal policy. Finally, in Section 5.4, we prove the monotonic character of each optimal decision rule with respect to t, the time in which the action is applied.

Chapter 6 includes an application of finite horizon CMC's with ETC criterion to an optimal resource allocation problem. In Section 6.1, the formulation of the problem as a CMC is given. In Section 6.2, we prove structural properties of the optimal value function. In the same section, we show the existence of an optimal policy such that the optimal decision rules are increasing functions of both the state x and the time t. In Section 6.3, we analyze the allocation problem under additional convexity conditions on the transition law, obtaining further structural properties for the optimal policy. Moreover, this structured optimal policy is compared with that corresponding to the risk-neutral allocation problem (which is obtained in Appendix A). In addition, we apply the results developed to the particular case of a problem with linear terminal cost function, and compare the conclusions with those corresponding to the risk-neutral problem, as well.

In Section 7.1, we present the basic framework for CMC's with cost per stage function depending explicity on a random disturbance. In Section 7.2, a DP algorithm is developed. It is shown that the risk-sensitive optimal value function at each stage depends on the prior disturbance. Moreover, we prove that the risk-sensitive optimal

policies yielded by this algorithm are not Markovian, because the optimal decision function at each stage depends on the prior disturbance in general. In Section 7.3, we collect some optimality results for the infinite horizon optimal control problem associated to our general model.

In Chapter 8, a CMC model for scheduling uncertain jobs is presented. To facilitate comparisons with the results we derive there, in Section 8.2 we include the analysis of the stochastic optimal control problem corresponding to the risk-null performance criterion given by an expected total weighted completion time (see [6, 38]). Then, in Section 8.3 we introduce risk- sensitivity by considering the minimization of the *expected exponential utility* of the total weighted completion time. We use a particular example to show how, similarly to the risk-neutral case, the optimal policies obtained from the DP algorithm are Markovian (i.e., the decision functions at each stage do not depend on the prior disturbance as in the general case). It is interesting to note, as we shall show, that for the risk-sensitive criterion a simple interchange argument is not applicable, and thus the only general computational and analytical tool for this situation is the DP algorithm. By means of a simple example, in Section 8.4 we illustrate how the optimal schedule depends on the risk sensitivity coefficient γ.

Finally, in Chapter 9, a formulation of an inventory control problem as a CMC is given. To that end, we consider a simplified model in which the random disturbance space is countable, and the transition law is independent of the prior disturbance. In Section 9.1 it is shown that, similarly to the risk-neutral case, the optimal value function depends only on the present state. This is due to the fact that the transition law is independent on the prior perturbation. In Section 9.2 (9.3), it is shown that the finite (infinite) horizon inventory control has a base-stock optimal policy.

Chapter 2

PRELIMINARIES

In Section 2.1, CMC'models are presented in the form needed for subsequent development. First, the basic structure is defined, and then we mention its intuitive interpretation. In Section 2.2, the set of admissible policies are defined and classified. Some basic notions of the expected utility approach to decision making under uncertainty are briefly summarized in Section 2.3. We conclude this chapter with the formulation, in Sections 2.4 and 2.5, of the risk-sensitive criteria that we are going to analyze later in this dissertation and the optimal control problem respectively.

Terminology and notation. Given a Borel space Y (i.e., a Borel subset of a complete separable metric space), its Borel σ-algebra is denoted by $\mathcal{B}(Y)$. When a measurable function (set) is written, it is understood as Borel measurable function (set). If Y and Z are Borel spaces, a stochastic kernel on Y given Z is a function $Q(\cdot \mid \cdot)$ such that $Q(\cdot \mid z)$ is a probability measure on Y for each $z \in Z$, and $Q(B \mid \cdot)$ is a measurable function on Z for each $B \in \mathcal{B}(Y)$.

2.1 The Model

Let us consider a CMC specified by the four-tuple $(\mathbf{X}, \mathbf{A}, \mathbf{P}, \mathbf{C})$, where:

- $\mathbf{X} = \{1, 2, \dots\}$ is the countable state space;

- $\mathbf{A}$, the action (or control) set, is a Borel space. To each $x \in \mathbf{X}$ we associate a nonempty measurable subset $A(x)$ of $\mathbf{A}$. $A(x)$ represents the set of admissible actions when the system is in state x. The set $\mathbf{K} := \{(x, a) : x \in \mathbf{X},\ a \in A(x)\}$, is called the set of admissible state-action pairs. In addition,

- $P(\cdot \mid \cdot)$, the transition law, is a stochastic kernel on $\mathbf{X}$ given $\mathbf{K}$.

We will also denote $p_{xx'}(a) := P(x' \mid x, a)$. Finally,

- $\mathbf{C} : \mathbf{K} \longrightarrow \mathbb{R}$, the one-stage cost function, is a measurable function.

The CMC ($\mathbf{X}$, $\mathbf{A}$, P, $\mathbf{C}$) represents a stochastic dynamical system observed at times $t = 0, 1, 2, \ldots$. The evolution of the system is as follows. Let X_t denote the state at time $t \in \mathbb{N}$, and A_t the action chosen at that time. If the system is in state $X_t = x \in \mathbf{X}$, and the control $A_t = a \in A(x)$ is chosen then (i) a cost $\mathbf{C}(x, a)$ is incurred, and (ii) the system moves to a new state X_{t+1} according to the probability distribution $P(\cdot \mid x, a)$. Once the transition into the new state has occurred, a new action is chosen, and the process is repeated. For some generalizations of this model see [1, 29, 40]; for more details see also [7].

The total period of time over which the system is to be observed is called the control horizon, and it is denoted by T. It can be finite, $T = \{0, 1, \ldots n-1\}$, or it can be infinite, $T = \mathbb{N}$.

The admissible history spaces are defined by

$$\mathbf{H}_0 := \mathbf{X}, \quad \mathbf{H}_t := \mathbf{K}^t \times \mathbf{X} \quad , \; t \geq 1,$$

and the canonical sample space is defined as

$$\Omega = (\mathbf{X} \times \mathbf{A})^\infty.$$

A generic element $\omega \in \Omega$ is of the form $\omega = (x_0, a_0, x_1, a_1, \ldots)$, where $x_t \in \mathbf{X}$ and $a_t \in \mathbf{A}$, for $t \geq 0$. An element of $\mathbf{H}_t$ is a vector of the form

$$h_t = (x_0, a_0, \ldots x_{t-1}, a_{t-1}, x_t),$$

where $(x_k, a_k) \in \mathbf{K}$ denotes the state and action of the system at decision epoch k, $k = 0, 1, \ldots, t-1$, and $x_t \in \mathbf{X}$ denotes the state at time t. The history $h_t \in \mathbf{H}_t$ follows the recursion $h_0 = x_0$, $h_t = (h_{t-1}, a_{t-1}, x_t)$. Note that $\mathbf{K}^\infty \subset (\mathbf{X} \times \mathbf{A})^\infty$.

The state, action, and history processes, denoted respectively by $\{X_t\}_{t \in T}$, $\{A_t\}_{t \in T}$, and $\{H_t\}_{t \in T}$ are defined on the measurable space $(\Omega, \mathcal{B}(\Omega))$ by the

projections

$$X_t(\omega) = x_t, \quad A_t(\omega) = a_t, \quad H_t(\omega) = (x_0, \ldots, a_{t-1}, x_t),$$

where $\mathcal{B}(\Omega)$ is the corresponding product σ-algebra. This means that when the observed path of states and actions is ω, the random variable X_t denotes the state at time t, A_t the chosen action at time t, and H_t the history up to time t.

2.2 Policies

A central feature of a CMC is the DM's possibility of influencing the evolution of the system by choosing actions. The way this influence is exerted is determined at the beginning of the process. This is done not by specifying at that point the actions to be taken at each future stage; but by determining a policy or sequence of decision rules. Those decision rules prescribe the actions to be taken at each stage. They range in generality from deterministic Markovian to randomized history dependent, depending on how do they incorporate past information and how do they select actions.

An admissible randomized control policy is a sequence $\pi = (q_t)_{t \in T}$ of stochastic kernels on $\mathbf{A}$ given $\mathbf{H}_t$ satisfying the constraint

$$q_t(A(x_t) \mid h_t) = 1, \quad h_t = (h_{t-1}, a_{t-1}, x_t) \in \mathbf{H}_t, \quad t \in T,$$

$q_t(a \mid h_t)$ denotes the probability of choosing action a at stage $t+1$ given that h_t is the history up to time t. The set of all admissible policies is denoted by $\mathbf{\Pi}$.

An admissible deterministic control policy is a sequence $\pi = (f_t)_{t \in T}$ of measurable functions (or decision functions)

$$f_t : \mathbf{H}_t \longrightarrow \mathbf{A}$$

such that

$$f_t(h_t) \in A(x_t), \quad h_t = (h_{t-1}, a_{t-1}, x_t) \in \mathbf{H}_t, \quad t \in T.$$

$f_t(h_t)$ is the action chosen at stage $t+1$ given that h_t is the history up to time t. We will denote the class of deterministic policies by $\mathbf{\Pi}_D$. Notice that a deterministic policy $(f_t)_{t\in T}$ can be seen as a randomized policy $(q_t)_{t\in T}$: $q_t\left(f_t(h_t) \mid h_t\right) = 1$.

The policies (randomized or deterministic) defined above are said to be history dependent because the chosen action at each stage may depend on stage and action at previous stages. The following two classes of policies given below are of special interest.

A randomized Markov policy is a sequence $\pi = (q_t)_{t\in T}$ of stochastic kernels on $\mathbf{A}$ given $\mathbf{X}$ such that

$$q_t\left(A(x_t) \mid x_t\right) = 1, \quad x_t \in \mathbf{X}, \quad t \in T.$$

A deterministic Markov policy is a sequence $\pi = (f_t)_{t\in T}$ of decision functions

$$f_t : \mathbf{X} \longrightarrow \mathbf{A}$$

such that $f_t(x_t) \in A(x_t), \quad x_t \in \mathbf{X}, \quad t \in T.$

For Markov policies, the chosen action at each stage t depends only on the current state x_t, and not on the whole history h_t. We will denote the set of Markov policies by $\mathbf{\Pi}_M$ and by $\mathbf{\Pi}_{MD}$ the set of deterministic Markov policies. Clearly $\mathbf{\Pi}_{MD} \subset \mathbf{\Pi}_M \subset \mathbf{\Pi}$; see [1].

Furthermore, a policy $\pi = (f_0, f_1, \ldots) \in \mathbf{\Pi}_{MD}$ is called a stationary deterministic policy if there exists a decision rule f such that $f_t = f$, for all t. In this case we use the notation $f^\infty := (f, f, f, \ldots)$, and in some cases just f. The set of these policies is denoted by $\mathbf{\Pi}_{SD}$.

Throughout this dissertation, we will assume that the CMC ($\mathbf{X}$, $\mathbf{A}$, $\mathbf{P}$, $\mathbf{C}$) satisfies the following condition.

Assumption 2.1. There exists $K \in \mathbb{R}$, such that $0 \leq \mathbf{C}(x,a) \leq K < \infty$, for all pairs $(x,a) \in \mathbf{K}$.

Let $\pi = (q_0, q_1, \ldots)$ be an arbitrary control policy, and x an initial state. Then using a theorem of C. Ionescu Tulcea (see e.g., [7, 37]) there exists a unique probability measure P_x^π on $\big(\Omega, \mathcal{B}(\Omega)\big)$ such that $P_x^\pi(\mathbf{K} \times \mathbf{K} \times \ldots) = 1$, and, moreover, for all $y \in \mathbf{X}$, $B \in \mathcal{B}(\mathbf{A})$, and $h_t \in \mathbf{H}_t$,

$$P_x^\pi(X_0 = x) = 1;$$

$$P_x^\pi(A_t \in B \mid H_t = h_t) = q_t(B \mid h_t);$$

and

$$P_x^\pi\left(X_{t+1} = y \mid H_t = (h_{t-1}, a_{t-1}, x_t), A_t = a\right) = p_{x_t y}(a). \tag{2.1}$$

This probability measure is such that its restriction to a finite horizon, $P_x^\pi \mid_{\mathbf{H}_n}$, satisfies

$$P_x^\pi(\{x\} \times B_0 \times \{x_1\} \times B_1 \times \ldots \times \{x_{n-1}\} \times B_{n-1} \times \{x_n\}) = \int_{B_0} q_0(da_0 \mid x) p_{xx_1}(a_0) \ldots \int_{B_{n-1}} q_{n-1}(da_{n-1} \mid x_{n-1}) p_{x_{n-1}x_n}(a_{n-1}),$$

for $x, x_1, \ldots, x_n \in \mathbf{X}$, and $B_0, B_1, \ldots, B_{n-1} \in \mathcal{B}(\mathbf{A})$.

We will often abuse notation by denoting $P_x^\pi \mid_{\mathbf{H}_n}$ by P_x^π when no risk of confusion exists. Also, the expectation operator with respect to P_x^π (or $P_x^\pi \mid_{\mathbf{H}_n}$) will be denoted by E_x^π.

2.3 Von Neumann-Morgenstern Utility Theory

In this section we briefly summarize some basic notions of the expected utility approach to decision making under uncertainty. For an extended discussion about expected utility theory and its applications, we refer to [21, 27, 34, 39].

Let us consider the situation wherein a DM has to choose among several policies or alternatives $\pi \in \mathbf{\Pi}$, each of which randomly generates an outcome $\omega \in \Omega$ according

to a known probability distribution P^π. A basic assumption is that decisions (choices) are made according to a DM's well defined preference pattern for a set $\mathcal{P}$ of probability measures on Ω such that $\mathcal{P} \supset \{P^\pi : \pi \in \mathbf{\Pi}\}$. That preferences pattern is described by an order relation $\preccurlyeq$ in $\mathcal{P}$. We say that $\preccurlyeq$ satisfies the *expected utility hypothesis* when there exists a measurable function $\mathcal{U} : \Omega \longrightarrow \mathbb{R}$ such that

$$P \preccurlyeq Q \iff E^P[\mathcal{U}] \leq E^Q[\mathcal{U}].$$

The function $\mathcal{U}$ is said to be the *utility function* of the order relation $\preccurlyeq$ (or of the corresponding DM.) The fundamental result of Utility Theory, von Neumann and Morgenstern (vNM) Theorem, establishes the necessary and sufficient conditions, the so called 'rationality' axioms, under which an order relation $\preccurlyeq$ in $\mathcal{P}$ satisfies the *expected utility axioms.* The reasonability of the 'rationality' axioms gives intuitive support to utility based preference patterns which are furthermore very convenient for their mathematical tractability. Nevertheless, it should be mentioned that those axioms are subject to serious criticism concerning their validity and meaning in practical situations (see [27, 34]).

For the case in which Ω is finite, the 'rationality' axioms consist of the fundamental premises of transitivity and completeness of $\preccurlyeq$ and some "continuity" conditions of $\preccurlyeq$ with respect to the *mixture* of probability measures on Ω, defined by:

$$(\alpha P_1 + (1-\alpha)P_2)(\{\omega\}) = \alpha P_1(\{\omega\}) + (1-\alpha)P_2(\{\omega\}),$$

for $\omega \in \Omega$, and $P_1, P_2 \in \mathcal{P}$. For an elementary proof of vNM Theorem (in the finite case), see e.g., [5, 34]. When Ω is infinite, some technical measure theoretic conditions must be added to the 'rationality' axioms for the vNM Theorem to hold (see e.g. [19]). These additional conditions will guarantee, for example, the measurability of the utility function $\mathcal{U}$.

Now, following Pratt [39], we summarize some important concepts and results concerning the particular case of utility based order relations on the space $\mathcal{P}_B$ of

probability measures on $\mathcal{B}$ with bounded support, where $\mathcal{B}$ is the Borel σ–algebra in $\mathbb{R}$. Here, P is of bounded support if $P(B) = 1$, for some bounded set $B \in \mathcal{B}$.

As it is customary in the literature we will rephrase all the discussion above as given in terms of random variables, via the correspondence $X \longrightarrow P_X$, where P_X is, as usual, the probability induced by a random variable X on the image space. Hence, $X \preccurlyeq Y$, just means $E^{P_X}[\mathcal{U}] \leqslant E^{P_Y}[\mathcal{U}]$, or equivalently, $E[\mathcal{U}(X)] \leqslant E[\mathcal{U}(Y)]$. We will refer to real random variables as *lotteries*, and to the number $E[\mathcal{U}(X)]$ as the *utility of the lottery* X. Note that if X is a degenerate lottery, that is, $P([X = x]) = 1$ for some $x \in \mathbb{R}$, then the expected utility of X is $\mathcal{U}(x)$.

Remark 1. *Although not explicitly indicated, all the lotteries can [8, 12] and will be considered to be defined in the same probability space* $(\Omega, \mathcal{A}, P)$*, and the expectation operator taken with respect to that* P*.*

In the sequel, whenever we mention a DM and a function $\mathcal{U}$, we are assuming that the later is the utility function of the former.

A DM is called *risk neutral* if

$$E[\mathcal{U}(X)] = \mathcal{U}(E[X]),$$

for every lottery X: a risk neutral DM is one who is always indifferent between taking the risks associated to a lottery X, and receiving for certain the expected value of this lottery.

A DM is called (strictly) *risk averse* if

$$E[\mathcal{U}(X)] < \mathcal{U}(E[X]) \tag{2.2}$$

for every non-degenerate lottery X: a risk averse DM always prefers the expected value of an uncertain lottery to the lottery itself. On the other hand, a DM is called (strictly) *risk seeking* if

$$E[\mathcal{U}(X)] > \mathcal{U}(E[X]) \tag{2.3}$$

for every non-degenerate lottery X.

The following proposition [39] shows that conditions (2.2) and (2.3) have strong implications on the shape of $\mathcal{U}$:

Proposition 1. *A DM is strictly risk averse (seeking) if and only if her utility function is strictly concave (convex).*

Since we will be dealing with utilities of positive amounts representing costs, we may and will assume in the sequel that the DM's preferences are described by a *strictly decreasing utility function.* In particular, the definitions and results in the rest of this section are for that kind of utilities. Of course, similar definitions and results can be given for strictly increasing utilities. Note that, under the assumption that $\mathcal{U}$ is strictly decreasing , if $\mathcal{U}$ is a twice differentiable function then Proposition 1 implies that the DM is risk averse, or risk seeking according to whether $\mathcal{U}''$ is negative or positive respectively.

The *certainty equivalent* of a lottery X is defined as the number $\varsigma(X)$ such that

$$\mathcal{U}(\varsigma(X)) = E[\mathcal{U}(X)].$$

That is, the DM is indifferent between the lottery X and the amount $\varsigma(X)$ for certain.

The *risk premium* of a lottery X is defined as the number $\Psi(X)$ such that

$$\mathcal{U}(\Psi(X) + E[X]) = E[\mathcal{U}(X)]. \tag{2.4}$$

That is, $\Psi(X)$ is the (rational) amount the DM would be willing to pay in addition to the expected value of the lottery in order to avoid that lottery. Note that

$$\begin{aligned} \varsigma(X) > E[X] \quad \forall \text{non-degenerate}\, X &\iff \Psi(X) > 0 \quad \forall \text{non-degenerate}\, X \\ &\iff \text{the DM is risk averse,} \end{aligned}$$

$$\begin{aligned} \varsigma(X) < E[X] \quad \forall \text{non-degenerate}\, X &\iff \Psi(X) < 0 \quad \forall \text{non-degenerate}\, X \\ &\iff \text{the DM is risk seeking.} \end{aligned}$$

For strictly decreasing and twice differentiable utility functions, Pratt [39] defined a notion of local measure of risk sensitivity in the following way. By taking the Taylor expansion of $\mathcal{U}$ around $E[X]$ in both sides of (2.4) we obtain

$$\mathcal{U}(\Psi(X) + E[X]) = \mathcal{U}(E[X]) + \Psi(X)\mathcal{U}'(E[X]) + o(\Psi(X)), \tag{2.5}$$

and

$$\begin{aligned} E[\mathcal{U}(X)] &= E\{\mathcal{U}(E[X]) + (X - E[X])\mathcal{U}'(E[X]) \\ &\quad + \frac{1}{2}(X - E[X])^2\mathcal{U}''(E[X]) + \dots\} \\ &= \mathcal{U}(E[X]) + \frac{1}{2}\sigma_X^2\mathcal{U}''(E[X]) + o(\sigma_X^2), \end{aligned} \tag{2.6}$$

where σ_X^2 denotes the variance of X. We are assuming in (2.6) that $E[(X - E[X])^k]$ is $o(\sigma_X^2)$ for $k > 2$. By equating (2.5) and (2.6), and neglecting terms of higher order in both equations we obtain

$$\Psi(X)\mathcal{U}'(E[X]) \simeq \frac{1}{2}\sigma^2\mathcal{U}''(E[X]). \tag{2.7}$$

From (2.7) we see that if the lottery X has small variance then its risk premium is proportional, up to first order, to the variance of X, and the proportionality factor is $\frac{1}{2}\frac{\mathcal{U}''(E[X])}{\mathcal{U}'(E[X])}$. In that sense, the function (introduced by Pratt) called the *risk sensitivity coefficient*,

$$r(x) := \frac{\mathcal{U}''(x)}{\mathcal{U}'(x)},$$

'measures' the degree of risk aversion in the neighborhood of x. Notice that if $r(x) > 0$ ($r(x) < 0$) then the DM is risk averse (risk seeking), and if $r(x) = 0$, the DM is risk neutral.

Constant Risk Sensitivity Coefficient. In this dissertation, we will be concerned only with the case of constant risk sensitivity coefficient. It is not hard to see that the utility functions with constant risk sensitivity coefficient are characterized by the

so called "$\Delta-$property" [31]: if X is increased by a constant $\Delta \in \mathbb{R}$, then the certain equivalent is increased by the same constant, that is, $\zeta(X+\Delta) = \Delta+\zeta(X)$. Moreover, from the definition of $r(x)$ we have that, up to a positive affine transformation,

$$r(x) \equiv \gamma \in \mathbb{R}, \quad \forall x \quad \text{if and only if}$$

$$\mathcal{U}(x) = \begin{cases} -x & \text{for } \gamma = 0 \text{ (risk neutral)} \\ -e^{\gamma x} & \text{for } \gamma > 0 \text{ (risk averse)} \\ e^{\gamma x} & \text{for } \gamma < 0 \text{ (risk seeking)} \end{cases}$$

For $\gamma \neq 0$, we will denote

$$\mathcal{U}_\gamma(x) = (sgn\gamma)e^{\gamma x}. \tag{2.8}$$

This function, being the negative of a utility function, is often called a (dis)utility function.

For brevity, we will refer to risk sensitivity corresponding to the (dis)utility function (2.8) as *γ-exponential risk sensitivity*, or simply as *exponential risk sensitivity.*

2.4 Risk Sensitive Criteria

Risk neutral cost criteria suffer from several shortcomings; see [23, 25, 45] and references therein. A notable limitation of the standard criteria is that they do not consider the sensitivity to risk of the DM who employs them. This fact has motivated the definition and development of risk sensitive criteria, see [24, 28, 31, 33]. In this dissertation, we are concerned (mainly) with the risk sensitive versions of the total and discounted cost standard criteria obtained by using the utility theoretic approach studied in the previous section: *exponential total cost and exponential discounted cost criteria.* These will be defined in this section and analyzed with more detail in Chapter 3.

Exponential Total Cost Criterion. Under the standard total cost criterion, the DM's ranking of policies

$$\pi \preccurlyeq \pi' \ \ (\pi' \text{ is prefered to } \ \pi) \ \text{ iff } \ \ E_x^{\pi}[-C] \leq E_x^{\pi'}[-C],$$

is induced by the order relation $\preccurlyeq_C$ in $\mathcal{P}_\Omega$ (the space of probabilities measures on the sample space) given by

$$P \preccurlyeq_C Q \quad \text{iff} \quad E^P[-C] \leq E^Q[-C],$$

where $C = \sum_{t \in T} \mathbf{C}(X_t, A_t)$, $P, Q \in \mathcal{P}_\Omega$. We can think of that ranking of policies as based also on the order relation $\preccurlyeq$ in $\mathcal{P}_B$ (the set of Borel probability measures of bounded support on $\mathbb{R}$) determined by $\mathcal{U}(x) = -x$ as utility function, since obviously for $P, Q \in \mathcal{P}_\Omega$

$$E^P[-C] \leq E^Q[-C] \Longleftrightarrow E^{P_C}[\mathcal{U}] \leq E^{Q_C}[\mathcal{U}], \tag{2.9}$$

From that point of view, according to the discussion of the previous section, we can see the total cost (or the DM who employs it) as a risk-neutral criterion. Consequently, a natural way of incorporating γ-risk sensitivity to the preference relation $\preccurlyeq_C$ is by considering the utility function $-\mathcal{U}_\gamma$ instead of the utility function in (2.9). In that way, the policy ranking for γ-exponential risk sensitivity will be given by

$$\pi_1 \preccurlyeq \pi_2 \quad \Longleftrightarrow \quad E_x^{\pi_1}[-\mathcal{U}_\gamma(C)] \leqslant E_x^{\pi_2}[-\mathcal{U}_\gamma(C)], \tag{2.10}$$

for $x \in \mathbf{X}$. In order to have a minimization problem as in the standard case, in the following definition we will consider the *expected disutility* instead of the expected utility that we have in (2.10).

The *γ-exponential total cost* incurred by a policy $\pi \in \mathbf{\Pi}$, and an initial state x is defined by

$$J_T^\pi(x,\gamma) := E_x^\pi\left[\mathcal{U}_\gamma(\mathcal{C})\right] = E_x^\pi\left[(sgn\gamma)e^{\gamma\sum_{t\in T}\mathbf{C}(X_t,A_t)}\right]. \tag{2.11}$$

When the horizon is finite (infinite), i.e., $T = \{0,1,\ldots n-1\}$, $n \in \mathbb{N}$, $(T = \mathbb{N})$, we denote the above more explicitly as $J_n^\pi(x,\gamma)$ $(J^\pi(x,\gamma))$.

Exponential Discounted Cost Criterion. Similarly as in the previous case, the DM's ranking of policies under the discounted cost criterion is based on the order relation $\preccurlyeq_{\mathcal{D}}$ in $\mathcal{P}_\Omega$ given by

$$P \preccurlyeq_{\mathcal{D}} Q \iff E^P[-\mathcal{D}] \leq E^Q[-\mathcal{D}] \iff E^{P_{\mathcal{D}}}[\mathcal{U}] \leq E^{Q_{\mathcal{D}}}[\mathcal{U}],$$

where $\mathcal{D} = \sum_{t\in T}\beta^t\mathbf{C}(X_t,A_t)$, and $\mathcal{U}(x) = -x$. Thus the same arguments lead us to the following definition of the exponential risk sensitive version of the standard discounted cost.

The γ-*exponential discounted cost* incurred by the policy $\pi \in \mathbf{\Pi}$, and the initial state x is defined by

$$J_{\beta,T}^\pi(x,\gamma) := E_x^\pi\left[\mathcal{U}_\gamma(\mathcal{D})\right] = E_x^\pi\left[(sgn\gamma)e^{\gamma\sum_{t\in T}\beta^t\mathbf{C}(X_t,A_t)}\right]. \tag{2.12}$$

When the horizon is finite (infinite), i.e., $T = \{0,1,\ldots,n-1\}$, $n \in \mathbb{N}$, $(T = \mathbb{N})$, we denote the above as $J_{\beta,n}^\pi(x,\gamma)$ $(J_\beta^\pi(x,\gamma))$.

2.5 The Optimal Stochastic Control Problem

The *optimal stochastic control problem* is that of selecting an admissible policy, such that a given performance criterion is minimized, over all admissible policies. For example, if a policy π^* is such that

$$J_T^{\pi^*}(x,\gamma) \leq J_T^\pi(x,\gamma), \quad \forall\pi \in \mathbf{\Pi}, \quad \forall x \in \mathbf{X},$$

then π^* is said to be (ETC)-utility optimal. The *optimal value function* is given by

$$J_T(x,\gamma) := \inf_{\pi\in\mathbf{\Pi}}\{J_T^{\pi}(x,\gamma)\}. \tag{2.13}$$

Similar definitions are made for the EDC criterion, and $J_{\beta,T}(x,\gamma)$ will denote the optimal value function.

Remark 2. *Note that even when the infima in (2.13) are attained, an optimal policy may not exist because for each $x \in \mathbf{X}$, a different π may attain the infimum.*

Chapter 3

Exponential Total and Discounted Cost Criteria

In this chapter we consider CMC's with ETC and EDC criteria. First, in Section 3.1 we prove the existence of a recursive scheme for the evaluation of the ETC and the EDC corresponding to a policy π, in a finite number of stages. Then, in the same section we derive the exponential version of the dynamic programming algorithm (DPA). In Section 3.2 (3.3) we derive a discounted (total) optimality equation that plays the same role as in the risk null model, that is, the optimal exponential cost and a utility optimal policy can be obtain from that equation. Finally, we prove that the optimal EDC (ETC) over a finite horizon converges to the optimal (EDC) ETC over an infinite horizon, when the number of stages goes to ∞. All the results discussed in Sections 3.2-3.3 were obtained in, e.g., [18, 20, 31, 33] for a finite state space, and restricting the optimization to the set of deterministic Markovian policies ($\mathbf{\Pi}_{MD}$). Here, we show that those results are valid for an infinite countable state space and we consider optimization over the set of all admissible policies ($\mathbf{\Pi}$).

3.1 Basic Results

Let $\mathcal{U}_\gamma(\cdot)$ be the exponential disutility function with constant risk-sensitivity coefficient $\gamma \neq 0$, that is, $\mathcal{U}_\gamma(x) = (sgn\gamma)e^{\gamma x}$; see Section 2.3. Then, for $h_s \in \mathbf{H}_s$, $s = 0, 1, \ldots n-1$, the EDC to go from time s to time n due to a policy $\pi \in \mathbf{\Pi}$ is defined as

$$u^\pi_{\beta,s}(h_s, \gamma) := E^\pi_{h_s}\left[\mathcal{U}_\gamma(\mathcal{D}^s)\right],$$

where $\mathcal{D}^s := \sum_{m=s}^{n-1} \beta^m \mathbf{C}(X_m, A_m)$, $0 < \beta < 1$, and for $h_n \in \mathbf{H}_n$, $u^\pi_{\beta,n}(h_n, \gamma) := sgn\gamma$. For this finite horizon case, taking $\beta = 1$ in the previous definitions yields

the corresponding ETC quantities. The corresponding optimal EDC (or ETC, with $\beta = 1$) to go is denoted by

$$u^*_{\beta,s}(h_s, \gamma) := \inf_{\pi}\{u^{\pi}_{\beta,s}(h_s, \gamma)\}.$$

In this chapter we provide proofs under the following condition.

Assumption 3.1. The action space is finite.

Additional conditions are required when the action set is a Borel space; see Remark 6.

The following two results for the finite horizon decision problem constitute the essence of the dynamic programming technique.

Lemma 1. (Finite Horizon Policy Evaluation Algorithm)
For arbitrary $\pi = \{q_0, q_1, ..., q_{n-1}\} \in \mathbf{\Pi})$, *and* $h_s = (h_{s-1}, a_{s-1}, x) \in \mathbf{H}_s$, *the functions* $u^{\pi}_{\beta,s}(h_s, \gamma)$ *satisfy*

$$u^{\pi}_{\beta,s}(h_s, \gamma) = \sum_{a \in A(x)} q_s(a \mid h_s) e^{\gamma \beta^s \mathbf{C}(x,a)} \sum_y p_{xy}(a) u^{\pi}_{\beta,s+1}((h_s, a, y), \gamma), \tag{3.1}$$

for $s = 0, 1, \ldots, n-2, n-1$.

Proof. It follows from the definition of the operators $E^{\pi}_{(\cdot)}$ that

$$\begin{aligned} u^{\pi}_{\beta,s}(h_s, \gamma) &= E^{\pi}_{h_s}\left[\mathcal{U}_\gamma(\mathcal{D}^s)\right] \\ &= \sum_{a \in A(x)} q_s(a \mid h_s) \sum_y p_{xy}(a) e^{\gamma \beta^s \mathbf{C}(x,a)} E^{\pi}_{((h_s,a,y))} \left[\mathcal{U}_\gamma\left(\sum_{t=s+1}^{n-1} \beta^t \mathbf{C}(X_t, A_t)\right)\right] \\ &= \sum_{a \in A(x)} q_s(a \mid h_s) e^{\gamma \beta^s \mathbf{C}(x,a)} \sum_y p_{xy}(a) u^{\pi}_{\beta,s+1}((h_s, a, y), \gamma) \end{aligned}$$

□

The following result, which provides an algorithm for finding both the optimal value function and an optimal policy, shows also that the optimal utility $u^*_{\beta,s}(h_s, \gamma)$ does not depend on the whole history h_s but only on the state at time s.

Theorem 1. (Dynamic Programming Algorithm) *Let $u_{\beta,s}$; $s = 0, 1, \cdots, n$ be the functions defined on* $\mathbf{X}$ *by*

$$u_{\beta,n}(x_n, \gamma) = (sgn\gamma), \tag{3.2}$$

$$\vdots \qquad\qquad \vdots$$

$$u_{\beta,s}(x_s, \gamma) = \min_{a \in A(x_s)} \left\{ e^{\gamma \beta^s \mathbf{C}(x_s, a)} p_{x_s y}(a) u_{\beta,s+1}(y, \gamma) \right\}. \tag{3.3}$$

For $s = 0, 1, 2, \ldots n-1$, let $f_s : \mathbf{X} \longrightarrow \mathbf{A}$ be a decision rule defined by

$$e^{\gamma \beta^s \mathbf{C}\left(x, f_s(x)\right)} \sum_y p_{xy}\left(f_s(x)\right) u_{\beta,s+1}(y, \gamma) = \min_{a \in A(x)} \left\{ e^{\gamma \beta^s \mathbf{C}(x,a)} \sum_y p_{xy}(a) u_{\beta,s+1}(y, \gamma) \right\}.$$

Then the Markov deterministic policy $\pi^* = (f_0, f_1, f_2, \ldots f_{n-1})$ is EDC-utility optimal, and

$$u^*_{\beta,s}(h_s, \gamma) = u_{\beta,s}(x_s, \gamma), \quad \forall\ h_s = (x_0, a_0, x_1, \ldots a_{s-1}, x_s), \quad s = 0, 1, \cdots, n.$$

Proof. Let $\pi = (q_0, \ldots, q_{n-1})$ be an arbitrary policy and $h_s = (h_{s-1}, a_{s-1}, x)$ an arbitrary history up to time s. To prove the theorem, we will prove that for $s = 0, 1, \ldots n-1$,

$$u^{\pi}_{\beta,s}(h_s, \gamma) \geqslant u_{\beta,s}(x, \gamma), \tag{3.4}$$

with equality if $\pi = \pi^*$, i.e.,

$$u^{\pi^*}_{\beta,s}(h_s, \gamma) = u_{\beta,s}(x, \gamma). \tag{3.5}$$

The proof of (3.4) and (3.5) is by backward induction. The inductive step holds for $t = n$ since

$$u^{\pi}_{\beta,n}(h_n, \gamma) = sgn\gamma = u_{\beta,n}(x_n, \gamma).$$

Now, assume that (3.4) and (3.5) hold for $s+1, \ldots n$. Then

$$u^{\pi}_{\beta,s}(h_s,\gamma) = E^{\pi}_{h_s}\left[\mathcal{U}_{\gamma}(\mathcal{D}_s)\right] \tag{3.6}$$

$$= \sum_{a\in A(x)} q_s(a \mid h_s) e^{\gamma\beta^s \mathbf{C}(x,a)} \sum_y p_{xy}(a) u^{\pi}_{\beta,s+1}((h_s,a,y),\gamma) \tag{3.7}$$

$$\geq \sum_{a\in A(x)} q_s(a \mid h_s) e^{\gamma\beta^s \mathbf{C}(x,a)} \sum_y p_{xy}(a) u_{\beta,s+1}(y,\gamma) \tag{3.8}$$

$$\geq \min_{a\in A(x)} \left\{ e^{\gamma\beta^s \mathbf{C}(x,a)} \sum_y p_{xy}(a) u_{\beta,s+1}(y,\gamma) \right\} \tag{3.9}$$

$$= u_{\beta,s}(x,\gamma), \tag{3.10}$$

where (3.7) follows from Lemma 3.1, and (3.8) from the induction hypothesis. This proves (3.4). Now, if $\pi = \pi^*$ then equality holds throughout the previous calculations and (3.5) follows. □

Remark 3. *a) The above theorem shows that the optimal value functions $u^*_{\beta,s}(x,\gamma)$ satisfy the recursion $u^*_{\beta,n}(x,\gamma) = sgn\gamma$, and for $s = n-1,\ldots,1,0$,*

$$u^*_{\beta,s}(x,\gamma) = \min_{a\in A(x)} \left\{ e^{\gamma\beta^s \mathbf{C}(x,a)} \sum_y p_{xy}(a) u^*_{\beta,s+1}(y,\gamma) \right\}. \tag{3.11}$$

b) When $\beta = 1$, (3.11) becomes

$$u^*_s(x,\gamma) = \min_{a\in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) u^*_{s+1}(y,\gamma) \right\}, \tag{3.12}$$

where $u^*_s(x,\gamma) := u^*_{1,s}(x,\gamma)$.

3.2 Exponential Discounted Optimality Equation

In this section, it is shown that the value of the infinite horizon optimal EDC satisfies the discounted optimality equation, that this equation characterizes Markov deterministic (EDC)-utility optimal policies, and also the convergence of the DPA.

Recall that, in Section 2.4, we defined the infinite horizon optimal EDC as

$$J_{\beta}(x,\gamma) = \inf_{\pi\in\mathbf{\Pi}} \{J^{\pi}_{\beta}(x,\gamma)\},$$

where

$$J_\beta^\pi(x,\gamma) = E_x^\pi\left[(sgn\gamma)e^{\gamma\sum_{t=0}^{\infty}\beta^t\mathbf{C}(X_t,A_t)}\right]. \tag{3.13}$$

Theorem 2. (Exponential Discounted Optimality Equations) *For $k = 0, 1, \ldots,$ the optimal exponential discounted costs, $J_\beta(x, \gamma\beta^k)$, satisfy the exponential discounted optimality equations (EDOE's)*

$$J_\beta(x,\gamma\beta^k) = \min_{a\in A(x)}\left\{e^{\gamma\beta^k\mathbf{C}(x,a)}\sum_y p_{xy}(a)J_\beta(y,\gamma\beta^{k+1})\right\}. \tag{3.14}$$

Remark 4. *Contrary to Bellman's equations in the risk-null case, see [1, 6, 29, 40, 41], 3.14 is non-stationary.*

Proof. (of Theorem 2.) Consider an arbitrary policy $\pi = (q_0, q_1, q_2, \ldots) \in \mathbf{\Pi}$. For each $(x,a) \in \mathbf{K}$ define a policy $\pi^{xa} = \{q_0^{xa}, q_1^{xa}, \ldots\}$ by $q_t^{xa}(a' \mid h_t) = q_{t+1}(a' \mid x, a, h_t)$, $t = 0, 1 \ldots$. Then

$$\begin{aligned}
&J_\beta^\pi(x,\gamma\beta^k) = E_x^\pi\left[\mathcal{U}_{\gamma\beta^k}(\mathcal{D})\right]\\
&= E_x^\pi\left[e^{\gamma\beta^k\mathbf{C}(x,A_0)}\mathcal{U}_{\gamma\beta^k}\left(\sum_{t=1}^{\infty}\beta^t\mathbf{C}(X_t,A_t)\right)\right]\\
&= \sum_{a\in A(x)} q_0(a|x)e^{\gamma\beta^k\mathbf{C}(x,a)}\sum_y p_{xy}(a)J_\beta^{\pi^{xa}}(y,\gamma\beta^{k+1})\\
&\geqslant \sum_{a\in A(x)} q_0(a|x)e^{\gamma\beta^k\mathbf{C}(x,a)}\sum_y p_{xy}(a)J_\beta(y,\gamma\beta^{k+1})\\
&\geq \min_{a\in A(x)}\left\{e^{\gamma\beta^k\mathbf{C}(x,a)}\sum_y p_{xy}(a)J_\beta(y,\gamma\beta^{k+1})\right\}.
\end{aligned}$$

Since π was arbitrary, it follows that

$$J_\beta(x,\gamma\beta^k) \geq \min_{a\in A(x)}\left\{e^{\gamma\beta^k\mathbf{C}(x,a)}\sum_y p_{xy}(a)J_\beta(y,\gamma\beta^{k+1})\right\}.$$

To obtain the reverse inequality, consider an arbitrary $\epsilon > 0$ and any initial state x. First, let $f_k : \mathbf{X} \longrightarrow \mathbf{A}$ be a decision rule defined by

$$\min_{a \in A(x)} \left\{ e^{\gamma\beta^k \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_\beta(y, \gamma\beta^{k+1}) \right\} = e^{\gamma\beta^k \mathbf{C}(x, f(x))} \sum_y p_{xy}(f_k(x)) J_\beta(y, \gamma\beta^{k+1}),$$

and one such $f_k(\cdot)$ exists by Assumption 3.1. Next, for each $y \in \mathbf{X}$ choose a policy $\pi_y = (q_0^y, q_1^y, \dots)$ such that

$$J_\beta^{\pi_y}(y, \gamma\beta^{k+1}) \leq J_\beta(y, \gamma\beta^{k+1}) + \epsilon.$$

Finally, define the policy $\bar{\pi} = (\bar{q}_0, \bar{q}_1, \dots)$ by taking $\bar{q}_0$ given by f_k, and

$$\bar{q}_t(a \mid x, a_0, x_1, \dots, x_{t-1}, a_{t-1}, x_t) = q_{t-1}^{x_1}(a \mid x_1, \dots, x_{t-1}, a_{t-1}, x_t),$$

for $t \geqslant 1$. Then

$$\begin{aligned} J_\beta(x, \gamma\beta^k) &\leq J_\beta^{\bar{\pi}}(x, \gamma\beta^k) \\ &= e^{\gamma\beta^k \mathbf{C}(x, f_k(x))} \sum_y p_{xy}(f_k(x)) J_\beta^{\pi_y}(y, \gamma\beta^{k+1}) \\ &\leq e^{\gamma\beta^k \mathbf{C}(x, f_k(x))} \sum_y p_{xy}(f_k(x)) [J_\beta(y, \gamma\beta^{k+1}) + \epsilon] \\ &= \min_{a \in A(x)} \left\{ e^{\gamma\beta^k \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_\beta(y, \gamma\beta^{k+1}) \right\} \\ &\qquad + \epsilon e^{\gamma\beta^k \mathbf{C}(x,a)}. \end{aligned}$$

Since ϵ was arbitrary and $\mathbf{C}(\cdot, \cdot)$ is bounded, we obtain

$$J_\beta(x, \gamma\beta^k) \leq \min_{a \in A(x)} \left\{ e^{\gamma\beta^k \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_\beta(y, \gamma\beta^{k+1}) \right\},$$

and the proof of the theorem is complete. □

Theorem 3. *For $k = 0, 1, 2, \dots$, let $f_k : \mathbf{X} \longrightarrow \mathbf{A}$ be a decision rule defined by*

$$e^{\gamma\beta^k \mathbf{C}\left(x, f_k(x)\right)} \sum_y p_{xy}\left(f_k(x)\right) J_\beta(y, \gamma\beta^{k+1}) = \min_{a \in A(x)} \left\{ e^{\gamma\beta^k \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_\beta(y, \gamma\beta^{k+1}) \right\}. \tag{3.15}$$

Then the policy $\pi^* = (f_0, f_1, f_2, \dots)$ *is (EDC)-utility optimal. That is,*

$$J_\beta^{\pi^*}(x,\gamma) = J_\beta(x,\gamma), \quad \textit{for all } x. \tag{3.16}$$

Remark 5. *Note that, like the analogous result in the risk-null case, see [1, 6, 29, 40, 41], the optimal policy provided by the above theorem is Markovian and deterministic but unlike the risk-null case in general the optimal policy is non-stationary.*

Proof. (of Theorem 3.) Denote by $\pi_k = (f_k, f_{k+1}, \dots)$, so that $\pi_0 = \pi^*$. To show (3.16) we need only to prove that $J_\beta^{\pi^*}(x,\gamma) \leqslant J_\beta(x,\gamma)$. First, we are going to prove by induction on n, that for each $k = 0, 1, 2, \dots$, and for all $x \in \mathbf{X}$,

$$J_{\beta,n}^{\pi_k}(x, \gamma\beta^k) \leqslant J_\beta(x, \gamma\beta^k), \ n = 1, 2, \dots . \tag{3.17}$$

Recall that, in Section 2.4, we defined

$$J_{\beta,n}^{\pi}(x,\gamma) := E_x^{\pi}[(sgn\gamma)e^{\gamma\sum_{t=0}^{n-1}\beta^t C(X_t,A_t)}].$$

We have that $J_{\beta,1}^{\pi_k}(x,\gamma\beta^k) = (sgn\gamma)e^{\gamma\beta^k \mathbf{C}(x,f_k(x))}$. On the other hand, we have from the EDOE and (3.15) that

$$\begin{aligned} J_\beta(x,\gamma\beta^k) &= \min_{a\in A(x)}\left\{e^{\gamma\beta^k \mathbf{C}(x,a)}\sum_y p_{xy}(a)J_\beta(y,\gamma\beta^{k+1})\right\} \\ &= e^{\gamma\beta^k \mathbf{C}\left(x,f_k(x)\right)}\sum_y p_{xy}\left(f_k(x)\right)J_\beta(y,\gamma\beta^{k+1}). \end{aligned}$$

From (3.13), note that $J_\beta(x,\ \gamma\beta^{k+1}) \geqslant sgn\gamma$, for each k, and for all x. Therefore, we obtain from (3.2) that

$$\begin{aligned} J_\beta(x,\ \gamma\beta^k) &\geqslant (sgn\gamma)e^{\gamma\beta^k \mathbf{C}\left(x,f_k(x)\right)} \\ &= J_{\beta,1}^{\pi_k}(x,\gamma\beta^k). \end{aligned} \tag{3.18}$$

Thus, from (3.18) we obtain that (3.17) holds for $n = 1$ for each k. Suppose now that, for each k,

$$J^{\pi_k}_{\beta,n-1}(x,\gamma\beta^k) \leqslant J_\beta(x,\gamma\beta^k), \quad \text{for all } x.$$

Then

$$\begin{aligned}
&J^{\pi_k}_{\beta,n}(x,\gamma\beta^k)\\
&= E^{\pi_k}_x[(sgn\gamma)e^{\gamma\beta^k\sum_{t=0}^{n-1}\beta^t C(X_t,A_t)}]\\
&= e^{\gamma\beta^k \mathbf{C}\left(x,f_k(x)\right)}\sum_y p_{xy}\left(f_k(x)\right) J^{\pi_{k+1}}_{\beta,n-1}(y,\gamma\beta^{k+1})\\
&\leqslant e^{\gamma\beta^k \mathbf{C}\left(x,f_k(x)\right)}\sum_y p_{xy}\left(f_k(x)\right) J_\beta(y,\gamma\beta^{k+1})\\
&= J_\beta(x,\gamma\beta^k)
\end{aligned}$$

Now, by the Monotone Convergence Theorem we have that

$$J^{\pi_k}_{\beta,n}(x,\gamma\beta^k) \to J^{\pi_k}_\beta(x,\gamma\beta^k).$$

Therefore, by letting $n \to \infty$ in (3.17), we obtain

$$J^{\pi_k}_\beta(x,\gamma\beta^k) \leqslant J_\beta(x,\gamma\beta^k), \tag{3.19}$$

$\forall x \in \mathbf{X}$, $\forall k \geqslant 0$. In particular, (3.16) is obtained from (3.19) by taking $k = 0$. □

Consider the n-stage problem obtained from the infinite horizon EDC by truncation. The optimal EDC for this problem is given by the last step of the algorithm (3.2)-(3.3). Note that

$$J_{\beta,n-s}(x,\gamma) = u_{\beta,s}(x,\gamma\beta^{-s}), \quad s = 0,1,\ldots n.$$

Then, an algorithm (forward in time) equivalent to (3.2)-(3.3) is given in the following lemma; see [6] for an analogous result in the risk-null case.

Lemma 2. *The functions $J_{\beta,s}(x,\gamma)$, $s = 0, 1, \cdots, n$, satisfy the recursion*

$$\begin{aligned} J_{\beta,0}(x,\gamma) &= sgn\gamma, \\ &\vdots \qquad \vdots \\ J_{\beta,s+1}(x,\gamma) &= \min_{a\in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_{\beta,s}(y,\gamma\beta) \right\}. \end{aligned} \tag{3.20}$$

The next proposition shows that the DP algorithm in (3.20) may be used to successively approximate $J_\beta(x,\gamma)$.

Theorem 4. (Value Iteration): *For $n = 0, 1, 2, \ldots$, we have that*

$$J_{\beta,n}(x,\gamma) \leqslant J_{\beta,n+1}(x,\gamma) \leqslant J_\beta(x,\gamma), \quad \forall x \in \mathbf{X}.$$

Furthermore,

$$\lim_{n\to\infty} J_{\beta,n}(x,\gamma) = J_\beta(x,\gamma). \tag{3.21}$$

Proof. Take an arbitrary $\pi \in \mathbf{\Pi}$. Then, since $\mathcal{U}_\gamma(\cdot)$ is increasing and, by Assumption 2.1, $0 \leqslant \mathbf{C}(x,a)$, we have that

$$\begin{aligned} J^\pi_{\beta,n+1}(x,\gamma) &= E^\pi_x \left[\mathcal{U}_\gamma(\mathcal{D}_{n+1})\right] \geqslant E^\pi_x \left[\mathcal{U}_\gamma(\mathcal{D}_n)\right] \\ &= J^\pi_{\beta,n}(x,\gamma) \geqslant J_{\beta,n}(x,\gamma), \end{aligned}$$

where $\mathcal{D}_n = \sum_{t=0}^n \beta^t \mathbf{C}(X_t, A_t)$. Since π was arbitrary, it follows that

$$J_{\beta,n+1}(x,\gamma) \geqslant J_{\beta,n}(x,\gamma).$$

Similarly, we can show that

$$J_\beta(x,\gamma) \geqslant J_{\beta,n}(x,\gamma),$$

for all n. Therefore, to obtain (3.21) it only remains to show that for arbitrary $\epsilon > 0$, there exists n such that

$$J_\beta(x,\gamma) \leqslant J_{\beta,n}(x,\gamma) + \epsilon. \tag{3.22}$$

We have that

$$\begin{aligned} J_\beta^\pi(x,\gamma) &= E_x^\pi\left[\mathcal{U}_\gamma(\mathcal{D}_n)e^{\gamma\sum_{t=n}^\infty \beta^t \mathbf{C}(X_t,A_t)}\right] \\ &\leqslant e^{\frac{\gamma K\beta^n}{1-\beta}} E_x^\pi\left[\mathcal{U}_\gamma(\mathcal{D}_n)\right] = e^{\frac{\gamma K\beta^n}{1-\beta}} J_{\beta,n}^\pi(x,\gamma) \\ &= J_{\beta,n}^\pi(x,\gamma)\left[1 + (sgn\gamma)v_n\right] \\ &= J_{\beta,n}^\pi(x,\gamma) + v_n|J_{\beta,n}^\pi(x,\gamma)|, \end{aligned}$$

where K is the bound in Assumption 2.1, and $v_n \geqslant 0$, $v_n \to 0$. If $\gamma < 0$, then $|J_{\beta,n}^\pi(x,\gamma)| \leqslant 1$, and if $\gamma > 0$, then $|J_{\beta,n}^\pi(x,\gamma)| \leqslant e^{\frac{\gamma K}{1-\beta}}$. Therefore

$$J_\beta^\pi(x,\gamma) \leqslant \begin{cases} J_{\beta,n}^\pi(x,\gamma) + v_n & \text{if } \gamma < 0 \\ J_{\beta,n}^\pi(x,\gamma) + v_n e^{\frac{\gamma K}{1-\beta}} & \text{if } \gamma > 0. \end{cases}$$

Thus, in both cases we can find n large enough so that

$$J_\beta^\pi(x,\gamma) - \epsilon \leqslant J_{\beta,n}^\pi(x,\gamma) \leqslant J_\beta^\pi(x,\gamma).$$

Since π was arbitrary, we obtain (3.22). □

3.3 Exponential Total Optimality Equation

In this Section, we prove for the ETC the validity of results similar to Theorems 2-4 in Section 3.2.

Recall that, in Section 2.4, we defined the infinite horizon optimal ETC as

$$J(x,\gamma) = \inf_{\pi\in\mathbf{\Pi}}\{J^\pi(x,\gamma)\},$$

where

$$J^\pi(x,\gamma) := E_x^\pi\left[(sgn\gamma)e^{\gamma\sum_{t=0}^\infty \mathbf{C}(X_t,A_t)}\right].$$

Theorem 5. (Exponential Total Optimality Equation) *The optimal ETC, $J(x,\gamma)$, satisfies the exponential total optimality equation (ETOE)*

$$J(x,\gamma) = \min_{a\in A(x)}\left\{e^{\gamma\mathbf{C}(x,a)}\sum_y p_{xy}(a)J(y,\gamma)\right\}. \tag{3.23}$$

Proof. The proof of Theorem 5 is very similar to the proof of Theorem 2, and thus is omitted here. □

Theorem 6. *Let f be a decision rule defined by*

$$e^{\gamma \mathbf{C}(x,f(x))} \sum_y p_{xy}\big(f(x)\big) J(y,\gamma) = \min_{a \in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J(y,\gamma) \right\}.$$

Then the stationary policy $\pi^* = (f, f, \dots)$ is (ETC)-utility optimal, that is,

$$J^f(x,\gamma) = J(x,\gamma), \quad \forall x \in \mathbf{X}.$$

Proof. The optimality of π^* will follow from the inequality

$$J(x,\gamma) \geqslant J^f(x,\gamma), \quad \forall x \in \mathbf{X}. \tag{3.24}$$

First, we prove by induction on n that

$$J(x,\gamma) \geqslant J_n^f(x,\gamma), \forall x \in \mathbf{X}, \quad \forall n. \tag{3.25}$$

It follows from the ETOE that

$$J(x,\gamma) = e^{\gamma \mathbf{C}(x,f(x))} \sum_y p_{xy}\big(f(x)\big) J(y,\gamma). \tag{3.26}$$

If $\gamma > 0$ then $J(y,\gamma) \geqslant 1$, for all y, and hence from (3.26) it follows that $J(x,\gamma) \geqslant e^{\gamma \mathbf{C}(x,f(x))}$, and thus

$$J(x,\gamma) \geqslant (sgn\gamma)\, e^{\gamma \mathbf{C}(x,f(x))} = J_1^f(x,\gamma). \tag{3.27}$$

On the other hand, if $\gamma < 0$ then $J(y,\gamma) \geqslant -1$, for all y, and hence $J(x,\gamma) \geqslant -e^{\gamma \mathbf{C}(x,f(x))}$. Therefore, again it follows from (3.26) that

$$J(x,\gamma) \geqslant (sgn\gamma)\, e^{\gamma \mathbf{C}(x,f(x))} = J_1^f(x,\gamma). \tag{3.28}$$

Thus, by (3.27) and (3.28) we obtain that (3.25) holds for $n = 1$. Now, we assume that for $n > 1$, $J_n^f(x,\gamma) \leq J(x,\gamma)$, for all $x \in \mathbf{X}$. Then

$$\begin{aligned} J^f_{n+1}(x,\gamma) &= e^{\gamma \mathbf{C}(x,f(x))} \sum_y p_{xy}\big(f(x)\big) J^f_n(y,\gamma) \\ &\leqslant e^{\gamma \mathbf{C}(x,f(x))} \sum_y p_{xy}\big(f(x)\big) J(y,\gamma) \\ &= J(x,\gamma), \end{aligned}$$

hence proving (3.25). Next, since, as $n \to \infty$,

$$\mathcal{U}_\gamma(\mathcal{C}_n) \uparrow \mathcal{U}_\gamma(\mathcal{C}),$$

then it follows from the Monotone Convergence Theorem that

$$E^f_x\left[\mathcal{U}_\gamma(\mathcal{C}_n)\right] \uparrow E^f_x\left[\mathcal{U}_\gamma(\mathcal{C})\right],$$

i.e.,

$$\lim_{n\to\infty} J^f_n(x,\gamma) = J^f(x,\gamma).$$

Finally, (3.24) is obtained by letting $n \to \infty$ in (3.25). □

Similarly as in (3.20), an algorithm equivalent to (3.2)-(3.3), with $\beta = 1$, is given by

$$\begin{aligned} J_0(x,\gamma) &= sgn\gamma, \\ &\vdots \qquad\quad \vdots \\ J_{s+1}(x,\gamma) &= \min_{a\in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_s(y,\gamma) \right\}, \end{aligned} \tag{3.29}$$

$s = 0, 1, 2, \ldots$.

The last result we will prove in this section is the ETC analogue of Theorem 4 in Section 3.2. The following three lemmas will be used in the proof of it.

Lemma 3. *For $n = 0, 1, \ldots,$*

$$J_n(x,\gamma) \leqslant J_{n+1}(x,\gamma) \leqslant J(x,\gamma), \quad \forall x \in \mathbf{X}.$$

Proof. We will prove only the first inequality, and the second one can be proved similarly. The proof will be done by induction on n. The induction step for $n = 0$ in the first inequality follows from the following trivial inequality

$$J_0(x,\gamma) = sgn\gamma \leqslant \min_a\{(sgn\gamma)e^{\gamma \mathbf{C}(x,a)}\} = J_1(x,\gamma).$$

Now, let's assume that $J_n(x,\gamma) \leqslant J_{n+1}(x,\gamma)$. Then, for $x \in \mathbf{X}$,

$$\begin{aligned} J_{n+1}(x,\gamma) &= \min_a \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_n(y,\gamma) \right\} \\ &\leqslant \min_a \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_{n+1}(y,\gamma) \right\} \\ &= J_{n+2}(x,\gamma). \end{aligned}$$

□

Lemma 4. $\lim_{n\to\infty} J_n(x,\gamma) =: J_\infty(x,\gamma)$ *satisfies the ETOE.*

Proof. We have that

$$\begin{aligned} J_\infty(x,\gamma) &= \lim_{n\to\infty} J_n(x,\gamma) \\ &= \lim_{n\to\infty} \min_{a\in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_{n-1}(y,\gamma) \right\} \\ &= \min_{a\in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_\infty(y,\gamma) \right\}, \end{aligned}$$

where the second and last equality follow respectively from (3.29) and Monotone Convergence Theorem. □

In the following lemma, it is shown that $J(x,\gamma)$ is the smallest bounded solution, not smaller than $(sgn\gamma)$, of the ETOE.

Lemma 5. *Let* $v : \mathbf{X} \longrightarrow \mathbf{R}$ *be a bounded function such that* $v \geqslant (sgn\gamma)$ *and*

$$v(x) = \min_{a\in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) v(y) \right\}. \tag{3.30}$$

Then, for fixed γ,

$$J(x,\gamma) \leqslant v(x), \quad \textit{for all } x. \tag{3.31}$$

Proof. Let g be a function defined by

$$e^{\gamma \mathbf{C}(x,g(x))} p_{xy}(g(x)) v(y) = \min_a \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) v(y) \right\}.$$

First, we will prove that for each $x \in \mathbf{X}$,

$$J_n^g(x,\gamma) \leqslant v(x), \quad \text{for } n = 1,2,3,\ldots. \tag{3.32}$$

It follows from (3.30) and the hypothesis $v \geqslant (sgn\gamma)$ that

$$\begin{aligned} v(x) &= e^{\gamma \mathbf{C}(x,g(x))} \sum_y p_{xy}\left(g(x)\right) v(y) \\ &\geqslant (sgn\gamma) e^{\gamma \mathbf{C}(x,g(x))} \\ &= J_1^g(x,\gamma). \end{aligned}$$

Thus, (3.32) holds for $n = 1$. Next, we assume that $J_n^g(x,\gamma) \leqslant v(x)$ for all x. Then

$$\begin{aligned} J_{n+1}^g(x,\gamma) &= e^{\gamma \mathbf{C}(x,g(x))} \sum_y p_{xy}\left(g(x)\right) J_n^g(y,\gamma) \\ &\leqslant e^{\gamma \mathbf{C}(x,g(x))} \sum_y p_{xy}\left(g(x)\right) v(y) \\ &= v(x). \end{aligned}$$

Finally, we obtain (3.31), by letting $n \to \infty$ in (3.32). □

Theorem 7. (Value Iteration) $J(x,\gamma) = \lim_{n\to\infty} J_n(x,\gamma)$.

Proof. From Lemma 3 it follows that

$$\lim_{n\to\infty} J_n(x,\gamma) \leqslant J(x,\gamma), \quad \text{for all } x.$$

The reverse inequality follows from Lemmas 4 and 5 since $\lim_{n\to\infty} J_n(x,\gamma) \geqslant (sgn\gamma)$, and the proof is complete. □

Remark 6. *The results obtained in this chapter continue to hold for more general CMC models with $A(x)$ not necessarily finite (see Section 2.1), under the following assumption:*

Assumption 3.1.

1) $A(x)$ is compact for each $x \in \mathbf{X}$;

2) $\mathbf{C}(x, \cdot)$ is continuous for each $x \in \mathbf{X}$; and

3) If $v : \mathbf{X} \longrightarrow \mathbb{R}$ is bounded then the function

$$a \longmapsto \sum_y p_{xy}(a)v(y) \quad \text{is continuous,}$$

for each $x \in \mathbf{X}$.

The proofs are straightforward modifications of the previous ones. Under Assumption 3.1 there exists a decision function $f : \mathbf{X} \longrightarrow \mathbf{A}$ (see [1, 29]) such that

$$\min_{a \in A(x)} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a)v(y) \right\} = e^{\gamma \mathbf{C}(x,f(x))} \sum_y p_{xy}(f(x))v(y),$$

for $v : \mathbf{X} \longrightarrow \mathbb{R}$ bounded.

Chapter 4

MODULARITY AND STRUCTURED POLICIES

In the sequel, the state and the action spaces are considered as subsets of the real numbers with the usual order: $(\mathbf{X}, \leqslant)$, $(\mathbf{A}, \leqslant)$. It is well known that "explicit" solutions arising from the optimality equations are obtained only in a few cases, even in risk-neutral models [1, 6, 9, 22, 26, 30, 40, 47]. This fact has motivated the study of structural properties of the value function and the optimal policies, both from the viewpoint of analysis and computations. By a "structural property" it is understood herein any special dependence or parametrization of, for example, $J_{\beta,t}(x, \gamma)$ and $f_t(x)$ on t, the initial state x, or γ. In this chapter, general conditions to obtain structural properties of the optimal exponential value function and policies are presented. In Section 4.1, we provide sufficient conditions on the basic components of a CMC for the optimal value function to be monotone, as a function of the initial state. In Section 4.2, the existence of monotone policies for CMC's with risk-sensitive criteria is established. Finally, in Section 4.3, it is shown that the modularity of certain functions (related to the exponential optimality equation) guarantees the existence of monotone policies.

4.1 Structural Properties of the Value Function.

Since the essential components of a CMC are the set $\mathbf{K}$ of admissible state-action pairs, the stochastic kernel $P(\cdot \mid \cdot)$ and the one stage cost $\mathbf{C}(\cdot, \cdot)$, properties of these components will induce properties of the value function.

First, a technical lemma is presented, which will be needed to prove other results in the sequel; see [40] for a proof.

Lemma 6. *Let $\{z_j\}$, and $\{z'_j\}$ be real-valued non-negatives sequences satisfying*

$$\sum_{j=k}^{\infty} z_j \geq \sum_{j=k}^{\infty} z'_j, \quad \text{for all } k,$$

with equality holding for $k = 0$. Suppose that $v_{j+1} \geq v_j$ for $j = 0, 1, \ldots$. Then

$$\sum_{j=0}^{\infty} v_j z_j \geqslant \sum_{j=0}^{\infty} v_j z'_j,$$

for all sequences $\{v_j\}$ for which both series converge.

Throughout this chapter, we will assume that the CMC ($\mathbf{X}$, $\mathbf{A}$, P, $\mathbf{C}$) satisfies Assumption 3.1.

The following lemma establishes structural results for the optimal value function of a CMC with risk-sensitive performance criterion; see [30, 40] for analogous results in the risk-neutral case.

Lemma 7. *Suppose that:*

i) $\mathbf{C}(x, a)$ is increasing (decreasing) in x, for each a;

ii) $\sum_{y=z}^{\infty} p_{xy}(a)$ is increasing in x for all $z \in \mathbf{X}$ and $a \in \mathbf{A}$; and

iii) $x \longmapsto A(x)$ is decreasing, i.e., $x' \geq x \Longrightarrow A(x') \subset A(x)$.

Then, the optimal EDC, $J_\beta(x, \gamma\beta^t)$, is increasing (decreasing) in x, for each t, and the optimal exponential total cost, $J(x, \gamma)$, is increasing (decreasing) in x.

Proof. We only prove the lemma for the exponential discounted case, and the exponential total case is similar. Also, we only prove the lemma for $\mathbf{C}(\cdot, a)$ increasing, the $\mathbf{C}(\cdot, a)$ decreasing case following similarly. By Theorem 4, we only need to prove that for each t, $J_{\beta,n}(x, \gamma\beta^t)$ is increasing in x, for all n. We will do it by induction on n. First, since $J_{\beta,0}(x, \gamma\beta^t) = sgn\gamma$, the result holds for $n = 0$. Now assume that

$J_{\beta,n}(x,\gamma\beta^t)$ is increasing in x, $\forall t$, and let $x' \geq x$. Then, using (3.20),

$$\begin{aligned} J_{\beta,n+1}(x,\gamma\beta^t) &\leqslant e^{\gamma\beta^t \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_{\beta,n}(y,\gamma\beta^{t+1}), \qquad \forall a \in A(x) \\ &\leqslant e^{\gamma\beta^t \mathbf{C}(x',a)} \sum_y p_{x'y}(a) J_{\beta,n}(y,\gamma\beta^{t+1}), \qquad \forall a \in A(x) \\ &\leqslant e^{\gamma\beta^t \mathbf{C}(x',a)} \sum_y p_{x'y}(a) J_{\beta,n}(y,\gamma\beta^{t+1}), \qquad \forall a \in A(x'), \end{aligned}$$

where the 1^{st} inequality follows from Condition (i), the second inequality follows from Lemma 6 by using the induction hypothesis and Condition (ii), and the last inequality follows from Condition (iii). Thus

$$\begin{aligned} J_{\beta,n+1}(x,\gamma\beta^t) &\leqslant \min_{a \in A(x')} \{ e^{\gamma\beta^t \mathbf{C}(x',a)} \sum_y p_{x'y}(a) J_{\beta,n}(y,\gamma\beta^{t+1}) \} \\ &= J_{\beta,n+1}(x',\gamma\beta^t), \end{aligned}$$

and the proof is complete. □

4.2 Monotone Policies.

In this section, the existence of optimal policies with special structure, e.g., threshold policies, is established. A threshold policy is a Markov deterministic policy $\pi = (f_0, f_1, \ldots)$ such that, for $t = 0, 1, \ldots$, the decision rule f_t is given by, e.g.,

$$f_t(x) = \begin{cases} a_1 & \text{if } x < x_t^* \\ a_2 & \text{if } x \geqslant x_t^*, \end{cases} \tag{4.1}$$

where $x_0^*, x_1^*, \ldots$ are the control limits or thresholds. More general structured policies are the monotone policies. A monotone policy is a Markov deterministic policy $\pi = (f_0, f_1, \ldots)$ such that the decision rules f_t, $t = 0, 1, \ldots$, are monotone functions of the state x. It is clearly benefitial to know when such an optimal policy exists, because the search for an optimal policy within the class of the Markov deterministic policies can be then restricted to the much smaller subclass of monotone policies [30, 40]. Due to the fact that the optimal policy $\pi = (f_0, f_1, \ldots)$ is in general non-stationary, a natural question that arises here is: how does the optimal action $f_t(x)$ varies with respect to

t, for each fixed x? The objective in this section is to find sufficient conditions that guarantee monotonicity with respect to x and/or with respect to t for the optimal decision rules. Thus, in the sequel, we refer to a monotone policy as a Markov deterministic policy $\pi = (f_0, f_1, \dots)$ such that the decision rules f_t, $t = 0, 1, \dots$, are monotone functions of the state x, and/or the stage t. For ease of presentation and clarity, we treat the finite and the infinite horizon models separately. We will consider a risk sensitivity coefficient $\gamma > 0$ throughout the rest of the dissertation.

4.2.1 Infinite Horizon CMC with EDC Criterion

For $\gamma > 0$ fixed, let $\Gamma := \{\gamma\beta^t : t = 0, 1, 2, \dots\}$, and $\mathcal{M} := \{v : \mathbf{X} \times \Gamma \longrightarrow \mathbb{R},\ v \text{ bounded}\}$. Similarly as in the risk-neutral case [6, 7, 29, 40], define the DP operator $T : \mathcal{M} \longrightarrow \mathcal{M}$ by

$$T(v)(x, \gamma\beta^t) = \min_{a \in A(x)} \{e^{\gamma\beta^t \mathbf{C}(x,a)} \sum_y p_{xy}(a) v(y, \gamma\beta^{t+1})\}$$

Now, set $\mathcal{N} := \{u : \mathbf{K} \times \Gamma \longrightarrow \mathbb{R},\ u \text{ bounded}\}$, and define the linear operator $L : \mathcal{M} \longrightarrow \mathcal{N}$, by

$$L(v)(x, a, \gamma\beta^t) = e^{\gamma\beta^t C(x,a)} \sum_y p_{xy}(a) v(y, \gamma\beta^{t+1}). \tag{4.2}$$

Remark 7. *Note that*

a) $J_\beta^\pi \in \mathcal{M}$, *for all* $\pi \in \mathbf{\Pi}$:

$$\mid J_\beta^\pi(x, \gamma\beta^t) \mid \leqslant e^{\frac{\gamma K}{1-\beta}}, \quad \forall x \in \mathbf{X},\ t = 0, 1, \dots;$$

b) $L(J_\beta)(x, a, \gamma\beta^t)$ *is the function within brackets in the EDOE's (3.14).*

In the following two theorems, sufficient conditions are presented for an infinite horizon CMC with EDC criterion to have an optimal policy $(f_0, f_1, \dots)$ such that the decision rules $f_t(x)$, $t = 0, 1, \dots$, are monotone functions in x, and monotone functions in t, respectively.

Theorem 8. *Set* $A_t^*(x) = \{a \in A(x) : L(J_\beta)(x, a, \gamma\beta^t) = T(J_\beta)(x, \gamma\beta^t)\}$, *and* $f_t(x) := \inf A_t^*(x)$. *Suppose that:*

i) $x \longmapsto A(x)$ *is a decreasing function, i.e.,* $x \leqslant x'$ *implies* $A(x') \subset A(x)$;

ii) for each $x \in \mathbf{X}$, *the set* $A(x)$ *is such that* $a \in A(x)$ *and* $a' \geq a$ *imply* $a' \in A(x)$; *and*

iii) If $x \leqslant x'$, $L(J_\beta)(x', \cdot, \gamma\beta^t) - L(J_\beta)(x, \cdot, \gamma\beta^t)$ *is decreasing on* $A(x')$, *for each fixed* t.

Then $(f_0, f_1, \dots)$ *is an optimal policy such that* $f_t(x)$ *is increasing in* x *for each* t.

Proof. First, it follows from Assumption 3.1 that the function $L(J_\beta)(x, \cdot, \gamma\beta^t)$ is continuous, and hence $A_t^*(x)$ is closed. In fact, from Assumption 3.1(1), we obtain that $A_t^*(x)$ is compact and therefore $f_t(x) \in A_t^*(x)$. Thus, the optimality of $(f_0, f_1, \dots)$ follows from Theorem 3. To see that $f_t(\cdot)$ is increasing, we suppose that $f_t(x') < f_t(x)$ for some $x' \geq x$. Then, it follows from i), ii) and iii) respectively that $f_t(x') \in A(x)$, $f_t(x) \in A(x')$ and

$$L(J_\beta)(x', f_t(x'), \gamma\beta^t) - L(J_\beta)(x, f_t(x'), \gamma\beta^t) \geq L(J_\beta)(x', f_t(x), \gamma\beta^t) - L(J_\beta)(x, f_t(x), \gamma\beta^t).$$

Consequently, we obtain

$$\begin{aligned} 0 &\geqslant L(J_\beta)(x', f_t(x'), \gamma\beta^t) - L(J_\beta)(x', f_t(x), \gamma\beta^t) \\ &\geqslant L(J_\beta)(x, f_t(x'), \gamma\beta^t) - L(J_\beta)(x, f_t(x), \gamma\beta^t) \geqslant 0, \end{aligned}$$

where the outer inequality follows from optimality of $f_t(x)$ and $f_t(x')$, and the inner inequality follows from the previous equation. But $f_t(x') \in A(x)$ and $L(J_\beta)(x, f_t(x'), \gamma\beta^t) = L(J_\beta)(x, f_t(x), \gamma\beta^t)$ contradict the definition of f_t since it was supposed that $f_t(x') < f_t(x)$ (strictly). Thus, $f_t(x') \geqslant f_t(x)\ \forall x' \geqslant x$. □

Theorem 9. *Set* $A_t^*(x) = \{a \in A(x) : L(J_\beta)(x, a, \gamma\beta^t) = T(J_\beta)(x, \gamma\beta^t)\}$, *and* $f_t(x) := \inf A_t^*(x)$. *Suppose that* $\gamma\beta^u \leqslant \gamma\beta^t$ *implies*

$$L(J_\beta)(x, \cdot, \gamma\beta^u) - L(J_\beta)(x, \cdot, \gamma\beta^t) \text{ is decreasing on } A(x), \text{ for each fixed } x. \quad (4.3)$$

Then $(f_0, f_1, \dots)$ *is an optimal policy such that* $f_t(x)$ *is increasing in* t*, for each fixed* x*.*

Proof. Similarly to the above theorem, the optimality of $(f_0, f_1, \dots)$ follows from Theorem 3. Now, we assume that $f_t(x) > f_u(x)$ for some $t \leqslant u$. Then, by (4.3) we obtain

$$L(J_\beta)(x, f_t(x), \gamma\beta^u) - L(J_\beta)(x, f_t(x), \gamma\beta^t) \\ \leqslant L(J_\beta)(x, f_u(x), \gamma\beta^u) - L(J_\beta)(x, f_u(x), \gamma\beta^t),$$

and hence

$$0 \leqslant L(J_\beta)(x, f_t(x), \gamma\beta^u) - L(J_\beta)(x, f_u(x), \gamma\beta^u) \\ \leqslant L(J_\beta)(x, f_t(x), \gamma\beta^t) - L(J_\beta)(x, f_u(x), \gamma\beta^t) \leqslant 0.$$

But $f_u(x) < f_t(x)$ and $L(J_\beta)(x, f_u(x), \gamma\beta^t) = L(J_\beta)(x, f_t(x), \gamma\beta^t)$ contradict the definition of f_t. Thus, $f_t(x) \leqslant f_u(x)$ $\forall t \leqslant u$. □

4.2.2 Finite Horizon CMC with ETC Criterion.

Consider a n-stages CMC as in Section 3.1. For $t = 1, 2, \dots n$, let

$$H_t(x, a, \gamma) := e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) u_t^*(y, \gamma), \tag{4.4}$$

and

$$\mathcal{E}_t(x, a, \gamma) := \frac{1}{\gamma} \log[H_t(x, a, \gamma)]$$

In the following two theorems similar results to the ones obtained in Theorems 8 and 9 are presented. Those results are stated in terms of certain equivalent instead of utilities and will be used in Chapter 6 to prove optimality of monotone policies, for a finite horizon optimal allocation model.

Theorem 10. *For* $t = 1, 2, \dots n$*, set*

$$A_t^*(x) = \left\{ a \in A(x) : \mathcal{E}_t(x, a, \gamma) = \min_{a' \in A(x)} \{\mathcal{E}_t(x, a', \gamma)\} \right\};$$

and $f_{t-1}(x) := \inf A_t^*(x)$. *Assume that*

i) $x \longmapsto A(x)$ *is a decreasing function;*

ii) for each $x \in \mathbf{X}$, *the set* $A(x)$ *is such that* $a \in A(x)$ *and* $a' \geq a$ *imply* $a' \in A(x)$; *and*

iii) If $x \leqslant x'$, $\mathcal{E}_t(x', \cdot, \gamma) - \mathcal{E}_t(x, \cdot, \gamma)$ *is decreasing on* $A(x')$, *for each fixed* t.

Then $(f_0, f_1, \ldots f_{n-1})$ *is an optimal policy such that* $f_t(x)$ *is increasing in* x *for each* t.

Proof. First, note that if $a \in A_t^*(x)$ then a satisfies that

$$H_t(x, a, \gamma) = \min_{a' \in A(x)} H_t(x, a', \gamma).$$

Thus, the optimality of $(f_0, f_1, \ldots f_{n-1})$ will follow from Theorem 1, if we show that $f_t(x) \in A_t^*(x)$. Indeed, it follows from Assumption 3.1 that the function $H_t(x, \cdot, \gamma)$ is continuous, and therefore $\log H_t(x, \cdot, \gamma)$ is continuous. Hence $A_t^*(x)$ is closed. Now, from Assumption 3.1(1), we obtain that $A_t^*(x)$ is compact and therefore $f_t(x) \in A_t^*(x)$. The rest of the proof is similar to the proof of Theorem 8. □

Theorem 11. *For* $t = 1, 2, \ldots n$, *set*

$$A_t^*(x) = \left\{a \in A(x) : \mathcal{E}_t(x, a, \gamma) = \min_{a' \in A(x)} \{\mathcal{E}_t(x, a', \gamma)\}\right\},$$

and $f_{t-1}(x) := \inf A_t^*(x)$. *Assume that*

$$\mathcal{E}_{t+1}(x, \cdot, \gamma) - \mathcal{E}_t(x, \cdot, \gamma) \quad \textit{is decreasing on } A(x), \textit{ for each fixed } x.$$

Then $(f_0, f_1, \ldots f_{n-1})$ *is an optimal policy such that* $f_t(x)$ *is increasing in* t *for each* x.

Proof. The proof is similar to the proof of Theorem 9. □

4.3 Modular Functions.

The results in Theorems 8-11 were obtained under the strong assumptions of monotonicity of the DP operators, e.g., (iii) in Theorem 8 and (4.3). In this section, we give corresponding modularity conditions that imply the above mentioned assumptions.

Let $(S, \preccurlyeq_S)$ be a lattice, i.e., a partially order set such that if $s, r \in S$ then $s \vee r \in S$ and $s \wedge r \in S$, and let $G : S \longrightarrow \mathbb{R}$. We say that

a) $G(\cdot)$ is *subadditive (or submodular)* on S if

$$G(s \vee r) + G(s \wedge r) \leqslant G(s) + G(r)$$

for every $s, r \in S$;

b) $G(\cdot)$ is *strictly subadditive* on S if

$$G(s \vee r) + G(s \wedge r) < G(s) + G(r)$$

for every non-comparable $s, r \in S$; and

c) $G(\cdot)$ is *superadditive (or supermodular)* on S if $-G(\cdot)$ is subadditive on S.

Unless otherwise stated, we will consider the product order $\preccurlyeq$ on $\mathbb{R}^2$, that is, $\preccurlyeq$ is defined by $(y, z) \preccurlyeq (y', z')$ if $y \leqslant y'$ and $z \leqslant z'$.

Submodularity on $\mathbf{K}$ of the function within brackets in the (risk-neutral) discounted cost optimality equation, i.e.,

$$J_\beta(x) = \min_a \{C(x, a) + \beta \sum_y p_{xy}(a) J_\beta(y)\}, \tag{4.5}$$

is frequently used to obtain monotone optimal policies for the corresponding stochastic control problem; see, e.g., [40]. In this section, it is shown that the subadditivity on $\mathbf{K}$ of certain functions, related to the exponential optimality equation, implies the existence of optimal policies whose decision rules are monotone functions in the state x. Moreover, besides submodularity with respect to $\mathbf{K}$ as in the risk-neutral case, submodularity with respect to $\mathbf{A} \times \Gamma$ ($\mathbf{A} \times \{0, 1, \ldots n\}$) is also considered, leading

to structural properties of the optimal policies with respect to γ (t) (recall that $\Gamma := \{\gamma\beta^t : t = 0, 1, \cdot\}$).

The next two theorems state results for the infinite horizon discounted case, while in Theorems 14-15 the finite horizon total cost is treated.

Theorem 12. *Assume that*

i) $x \longmapsto A(x)$ *is a decreasing function;*

ii) for each $x \in \mathbf{X}$, *the set* $A(x)$ *is such that* $a \in A(x)$ *and* $a' \geq a$ *imply* $a' \in A(x)$*).*

Then $(\mathbf{K}, \preccurlyeq)$ *is a lattice. Moreover, assume that*

iii) $L(J_\beta)(\cdot, \cdot, \gamma\beta^t)$ *is subadditive on* $(\mathbf{K}, \preccurlyeq)$.

Then there exists an optimal policy $(f_0, f_1, \ldots)$ *such that* $f_t(x)$ *is increasing in* x *for each* t.

Proof. We will show that result by applying Theorem 8. That is, we will show that if $x < x'$ then

$$L(J_\beta)(x', \cdot, \gamma\beta^t) - L(J_\beta)(x, \cdot, \gamma\beta^t)$$

is decreasing on $A(x')$. Let $a,\ a' \in A(x'),\ a < a'$. Then

$$(x, a') \vee (x', a) = (x', a') \in \mathbf{K}, \quad \text{and} \quad (x, a') \wedge (x', a) = (x, a) \in \mathbf{K},$$

where $(x, a) \in \mathbf{K}$ follows from (i) since $a \in A(x') \subset A(x)$. Hence, under (i)-(ii), $(\mathbf{K}, \preccurlyeq)$ is a lattice. Moreover, by the subadditivity of $L(J_\beta)(\cdot, \cdot, \gamma\beta^t)$, we obtain

$$L(J_\beta)(x', a', \gamma\beta^t) + L(J_\beta)(x, a, \gamma\beta^t) \leqslant$$
$$L(J_\beta)(x', a, \gamma\beta^t) + L(J_\beta)(x, a', \gamma\beta^t),$$

and hence

$$L(J_\beta)(x', a', \gamma\beta^t) - L(J_\beta)(x, a', \gamma\beta^t) \leqslant$$
$$L(J_\beta)(x', a, \gamma\beta^t) - L(J_\beta)(x, a, \gamma\beta^t),$$

i.e., *(iii)* in Theorem 8 holds, and the proof is complete. □

Theorem 13. *Consider the partial order $\preccurlyeq_{inv}$ on $\mathbb{R}^2$ defined by $(y,z) \preccurlyeq_{inv} (y',z')$ if $y \leqslant y'$ and $z \geqslant z'$. Then: (i) for each $x \in \mathbf{X}$, $(A(x) \times \Gamma, \preccurlyeq_{inv})$ is a lattice. Assume that, for each $x \in \mathbf{X}$, the function $L(J_\beta)(x,\cdot,\cdot)$ is subadditive on $(A(x) \times \Gamma, \preccurlyeq_{inv})$. Then: (ii) there exists an optimal policy $(f_0, f_1, \dots)$ such that $f_t(x)$ is increasing in t for each x.*

Proof. We will apply Theorem 9, that is, we will prove that $\gamma\beta^u \leqslant \gamma\beta^t$ implies

$$L(J_\beta)(x,\cdot,\gamma\beta^u) - L(J_\beta)(x,\cdot,\gamma\beta^t) \text{ is decreasing on } A(x), \text{ for each fixed } x. \quad (4.6)$$

Let $\gamma\beta^u \leqslant \gamma\beta^t$, and $a \leqslant a'$. Then

$$(a,\gamma\beta^u) \vee (a',\gamma\beta^t) = (a',\gamma\beta^u), \text{ and } (a,\gamma\beta^u) \wedge (a',\gamma\beta^t) = (a,\gamma\beta^t).$$

Hence, $(A(x) \times \Gamma, \preccurlyeq_{inv})$ is clearly a lattice. Thus, by the subadditivity we have that

$$L(J_\beta)(x,a',\gamma\beta^u) + L(J_\beta)(x,a,\gamma\beta^t) \leqslant L(J_\beta)(x,a',\gamma\beta^t) + L(J_\beta)(x,a,\gamma\beta^u),$$

from which (4.6) is obtained. □

Theorem 14. *Assume that*

i) $x \longmapsto A(x)$ is decreasing;

ii) for each $x \in \mathbf{X}$, the set $A(x)$ is such that $a \in A(x)$ and $a' \geq a$ imply $a' \in A(x)$); and

iii) $\mathcal{E}_t(\cdot,\cdot,\gamma)$ is subadditive on $(\mathbf{K}, \preccurlyeq)$.

Then there exists an optimal policy $(f_0, f_1, \dots f_{n-1})$ such that $f_t(x)$ is increasing in x for each t.

Proof. The proof of this result is similar to that of the above theorem, but in this case we apply Theorem 10. □

Theorem 15. *Assume that for each $x \in \mathbf{X}$, the function $\mathcal{E}_{(\cdot)}(x,\cdot,\gamma)$ is subadditive on the lattice $(A(x) \times \{1,\dots n\}, \preccurlyeq)$. Then there exists an optimal policy $(f_0, f_1, \dots f_{n-1})$ such that $f_t(x)$ is increasing in t for each x.*

Proof. We will apply Theorem 11. That is, we need to prove that for $t = 0, 1, \ldots, n-1$

$$\mathcal{E}_{t+1}(x, \cdot, \gamma) - \mathcal{E}_t(x, \cdot, \gamma) \text{ is decreasing on } A(x). \tag{4.7}$$

Let $a, a' \in A(x)$, $a \leqslant a'$. Then, $(A(x) \times \{1, \ldots n\}, \preccurlyeq)$ is clearly a lattice:

$$(a, t+1) \vee (a', t) = (a', t+1), \text{ and } (a, t+1) \wedge (a', t) = (a, t).$$

Thus, by the subadditivity we have that

$$\mathcal{E}_{t+1}(x, a', \gamma) + \mathcal{E}_t(x, a, \gamma) \leqslant \mathcal{E}_{t+1}(x, a, \gamma) + \mathcal{E}_t(x, a', \gamma)$$

from which (4.7) follows. □

In the risk-neutral model, one is interested in establishing verifiable conditions under which the function within brackets in (4.5) is submodular. This task is simplified by the fact that submodularity is preserved under addition, because it is easier to find conditions for each of the functions $C(x, a)$ and $\sum_y p_{xy}(a) J_\beta(y)$ to be submodular separately.

Therefore, motivated by Theorems 12 and 13 and work in the risk-neutral literature, in the sequel we pursue the following: obtain conditions which guarantee submodularity on $\mathbf{X} \times \mathbf{A}$ and/or on $\mathbf{A} \times \Gamma$ of the function

$$L(J_\beta)(x, a, \gamma\beta^t) = g(x, a, \gamma\beta^t) h(x, a, \gamma\beta^t), \tag{4.8}$$

where

$$g(x, a, \gamma\beta^t) := e^{\gamma\beta^t C(x,a)}, \qquad h(x, a, \gamma\beta^t) := \sum_y p_{xy}(a) J_\beta(y, \gamma\beta^{t+1}). \tag{4.9}$$

However, an additional challenge is faced here in that now additional conditions must be established for the *product* of two submodular functions to be a submodular function. In Subsection 4.3.1, seemingly novel results about modularity of the product

of two functions are established, and then conditions to ensure that the submodularity of $L(J_\beta)$ follows from that of g and h are given.

A second and novel approach used in this dissertation to decompose the problem as in the risk-neutral case is motivated by Theorems 14 and 15; the challenge here is to find conditions that guarantee submodularity on $\mathbf{X} \times \mathbf{A}$ and/or on $\mathbf{A} \times \{1, \ldots n\}$ of the function

$$\mathcal{E}_t(x, a, \gamma) = \mathbf{C}(x, a) + \frac{1}{\gamma} \log[h_t(x, a, \gamma)],$$

where

$$h_t(x, a, \gamma) = \sum_y p_{xy}(a) u_t^*(y, \gamma).$$

In that case, similarly to the risk-neutral case, the submodularity of the two functions $\log[h_t(x, a, \gamma)]$ and $\mathbf{C}(x, a)$ would be needed separately. In Chapter 6, for a particular resource allocation problem, we will show that those two functions are subadditive (and therefore so is the function $\mathcal{E}_t(x, a, \gamma)$). It is important to mention that submodularity of the function $\mathcal{E}_t(x, a, \gamma)$ ($H_t(x, a, \gamma)$) does not necessarily imply submodularity of the function $H_t(x, a, \gamma)$ ($\mathcal{E}_t(x, a, \gamma)$).

4.3.1 Product of Submodular Functions.

Theorem 16. *Let $(S, \preccurlyeq_S)$ be a lattice, and let $g, h : S \longrightarrow \mathbb{R}$ be non-negative and subadditive functions on S. If $g(\cdot)$ is increasing (decreasing) and $h(\cdot)$ is decreasing (increasing) then the product $g(\cdot)h(\cdot)$ is subadditive on $(S, \preccurlyeq_S)$.*

Proof.

$$\begin{aligned}
&g(s \vee r)h(s \vee r) - g(s)h(s) \\
&\qquad \leqslant [g(s) + g(r) - g(s \wedge r)]\, h(s \vee r) - g(s)h(s) \\
&\qquad = g(r)h(s \vee r) + [g(s) - g(s \wedge r)]\, h(s \vee r) - g(s)h(s) \\
&\qquad \leqslant g(r)h(s \vee r) + [g(s) - g(s \wedge r)]\, h(s) - g(s)h(s) \\
&\qquad = g(r)h(s \vee r) - g(s \wedge r)h(s) \\
&\qquad \leqslant g(r)\,[h(s) + h(r) - h(s \wedge r)] - g(s \wedge r)h(s) \\
&\qquad \leqslant g(r)h(r) + g(s \wedge r)\,[h(s) - h(s \wedge r)] - g(s \wedge r)h(s) \\
&\qquad \leqslant g(r)h(r) - g(s \wedge r)h(s \wedge r).
\end{aligned}$$

□

In the following two theorems, the monotonicity property of the functions $g(\cdot)$ and $h(\cdot)$ is substituted for some alternative conditions that also lead to results as in Theorem 16. Moreover, these two theorems will be used in Chapter 5, to show submodularity (with respect to $\mathbf{A} \times \Gamma$ and $\mathbf{X} \times \mathbf{A}$ respectively) of the product $g(\cdot)h(\cdot)$ in the optimality equations arising in a machine replacement problem.

Theorem 17. *Let $(S, \preccurlyeq_S)$ be a lattice, and let $g, h : S \longrightarrow \mathbb{R}$ be non-negative, and subadditive functions on S. Assume that $g(s) = g(s \wedge r)$ and $h(s) \leqslant h(s \wedge r)$ for every non-comparable $s, r \in S$ such that $g(s) \leqslant g(r)$. Then the product $g(\cdot)h(\cdot)$ is subadditive on S.*

Proof. Take arbitrary non-comparable $s, r \in S$ such that $g(s) \leqslant g(r)$. We deduce at once from the hypotheses that $g(r) \geq g(s \vee r)$ and $g(r) \geq g(s \wedge r)$. Thus,

$$\begin{aligned}
&g(s \vee r)h(s \vee r) - g(s)h(s) \\
&\qquad \leqslant g(r)\,[h(s) + h(r) - h(s \wedge r)] - g(s \wedge r)h(s) \\
&\qquad = g(r)h(r) + g(s \wedge r)\,[h(s) - h(s \wedge r)] - g(s \wedge r)h(s) \\
&\qquad \leqslant g(r)h(r) - g(s \wedge r)h(s \wedge r),
\end{aligned}$$

i.e., $g(\cdot)h(\cdot)$ is subadditive. □

Theorem 18. *Let $(S, \preccurlyeq_S)$ be a lattice, and let $g, h : S \longrightarrow \mathbb{R}$ be non-negative, and subadditive functions on S. Assume that for every non-comparable $s, r \in S$ such that $g(s) \leqslant g(r)$, the following two conditions hold: a) $g(s) = g(s \wedge r)$ and b) $h(s \vee r) = h(r)$. Then the product $g(\cdot)h(\cdot)$ is subadditive on S.*

Proof. From the hypotheses we have that $g(s \vee r) \leqslant g(r)$ and $h(s \wedge r) \leqslant h(s)$, which yields

$$g(s \vee r)h(s \vee r) + g(s \wedge r)h(s \wedge r) \leqslant g(s)h(s) + g(r)h(r).$$

□

Next, we provide conditions on the costs and transition probabilities under which the functions $g(\cdot, \cdot, \gamma\beta^t)$ and $h(\cdot, \cdot, \gamma\beta^t)$ in (4.9) satisfy the modularity property. For ease of presentation, $A(x) = \mathbf{A}$, for all $x \in \mathbf{X}$, will be assumed.

Assumption 4.1 $\mathbf{C}(x, a)$ is separable, i.e., $\mathbf{C}(x, a) = u(x) + r(a)$ for some functions $u : \mathbf{X} \longrightarrow \mathbb{R}$ and $r : \mathbf{A} \longrightarrow \mathbb{R}$.

Assumption 4.2 $\mathbf{C}(x, a)$ is increasing in (x, a).

Assumption 4.3 $\sum_{y=z}^{\infty} p_{xy}(a)$ is increasing in x, for all $z \in \mathbf{X}$ and $a \in \mathbf{A}$.

Assumption 4.4 $\sum_{y=z}^{\infty} p_{xy}(a)$ is subadditive in $\mathbf{X} \times \mathbf{A}$, for all $z \in \mathbf{X}$.

Note that if the cost function satisfies Assumption 4.1 then the discounted optimality equation is given as follows:

$$J_\beta(x, \gamma\beta^t) = e^{\gamma\beta^t u(x)} \min_a \{e^{\gamma\beta^t r(a)} \sum_y p_{xy}(a) J_\beta(y, \gamma\beta^{t+1})\}.$$

Thus, the function $L(J_\beta)(x, a, \gamma\beta^t)$ in (4.8), is in this case given by

$$L(J_\beta)(x, a, \gamma\beta^t) = e^{\gamma\beta^t r(a)} \sum_y p_{xy}(a) J_\beta(y, \gamma\beta^{t+1}).$$

In the rest of this dissertation we denote $\bar{g}(x, a, \gamma\beta^t) := e^{\gamma\beta^t r(a)}$.

Lemma 8. *Under Assumption 4.1, $\bar{g}(\cdot,\cdot,\gamma\beta^t)$ is subadditive and superadditive on* $\mathbf{X}\times\mathbf{A}$.

Proof. The result follows since $\bar{g}(\cdot,a,\gamma\beta^t)$ is constant. □

Lemma 9. *Under Assumption 4.2, 4.3, and 4.4, $h(\cdot,\cdot,\gamma\beta^t)$ is subadditive in* $\mathbf{X}\times\mathbf{A}$.

Proof. By Lemma 7, Assumptions 4.2 and 4.3 imply that $J_\beta(x,\gamma\beta^t)$ is increasing in x. Now, by Assumption 4.4, we have that if $x'\geq x$ and $a'\geq a$ then

$$\sum_{y=z}^{\infty}[p_{x'y}(a')+p_{xy}(a)]\leqslant\sum_{y=z}^{\infty}[p_{x'y}(a)+p_{xy}(a')]$$

and by applying Lemma 6,

$$\sum_{y=z}^{\infty}[p_{x'y}(a')+p_{xy}(a)]J_\beta(y,\gamma\beta^t)\leqslant\sum_{y=z}^{\infty}[p_{x'y}(a)+p_{xy}(a')]J_\beta(y,\gamma\beta^t)$$

is obtained, i.e., $h(\cdot,\cdot,\gamma\beta^t)$ is subadditive in $\mathbf{X}\times\mathbf{A}$. □

Chapter 5

Application 1: An Equipment Replacement Model

In this chapter, an infinite horizon CMC model of an equipment replacement problem is studied. In Section 5.1, the formulation of the model is presented. In Sections 5.2 and 5.3, respectively, it is shown that under standard conditions, the optimal policy is of the threshold-type and ultimately stationary. Moreover, under mild additional conditions, it is also shown in Section 5.3 that ultimately it is optimal to follow the risk-neutral stationary optimal policy. Finally, in Section 5.4, it is proved that the optimal decision functions $f_t(x)$, $t = 0, 1, \ldots$, are decreasing functions in t for each x; see, e.g., [6, 40, 41] for an analysis of this problem with risk-neutral discounted cost, and [26] for a finite horizon, finite states/actions model with a risk-sensitive criterion and under partial state observations.

5.1 Formulation of the Equipment Replacement Problem as a CMC

The state of an equipment unit used in a, e.g., manufacturing process deteriorates over the time. The state $x \in \mathbf{X} = \{0, 1, 2, \ldots\}$ represents the level of deterioration. At each decision epoch, the decision maker can choose one of two actions from the action space $\mathbf{A} = \{1, 2\}$. Action 1 corresponds to *keep the unit* (do not replace it) while action 2 corresponds to *replace the unit.* If the decision is *to replace*, then a cost $R > 0$ is immediately incurred and the state at the beginning of the next time period is 0, the "as new" state. If the present state is x and the action is *to keep*, then the state at the beginning of the next time will be y with probability p_{xy}. In addition, an operating cost $0 \leqslant c(x) \leqslant K$ is incurred when the present state of the

unit is x. Thus, the cost function per stage is given by

$$\mathbf{C}(x,1) = c(x), \qquad \mathbf{C}(x,2) = c(x) + R.$$

The following is a natural assumption to make.

Assumption 5.1. $c(x)$ is increasing in x.

Since the component tends to turn worse gradually with use, it is natural to model the probabilistic structure such that the conditional probability of the component being at a state greater or equal than k at the end of the period, given that it was a state x at the beginning of the period is increasing in x, i.e.,

$$\sum_{y=z}^{\infty} p_{xy}(1) \quad \text{is increasing in } x \tag{5.1}$$

5.2 Threshold Optimal Policies

First, it will be shown that the optimal value function for the model above is increasing on the state, and then the existence of a threshold optimal policy will be established.

Lemma 10. *The optimal value function $J_\beta(x,\gamma)$ is increasing in x.*

Proof. . We will apply Lemma 7 to prove that $J_\beta(x,\gamma)$ is increasing in x. It follows from Assumption 5.1 and (5.1) respectively that Conditions 1 and 2 hold in this model, and Condition 3 is obviously verified since $A(x) = \mathbf{A}$, $\forall x$. Therefore the result follows. □

Proposition 2. *There exists an increasing optimal policy $\pi^* = (f_0^*, f_1^*, \ldots)$ for the equipment replacement problem with risk-sensitive discounted cost criterion, that is, there exists a sequence $\{x_0^*, x_1^*, \ldots\}$ such that f_t^* is given by*

$$f_t^*(x) = \begin{cases} 1 \text{ (no-replace)} & \text{if } x < x_t^* \\ 2 \text{ (replace)} & \text{if } x \geqslant x_t^* \end{cases} \tag{5.2}$$

Proof. First, Theorem 18 will be applied to show that the function

$$L(J_\beta)(\cdot,\cdot,\gamma\beta^t) = g(\cdot,\cdot,\gamma\beta^t)h(\cdot,\cdot,\gamma\beta^t)$$

is subadditive on $\mathbf{X} \times \mathbf{A}$. From the conditions of the model, we have that the cost function is separable:

$$\mathbf{C}(x,a) = c(x) + r(a),$$

where

$$r(1) = 0 \quad \text{and} \quad r(2) = R. \tag{5.3}$$

Thus, by Lemma 8, $g(\cdot,\cdot,\gamma\beta^t)$ is subadditive on $\mathbf{X} \times \mathbf{A}$. Now, it follows from Assumption 5.1, (5.3) and (5.1) respectively that $\mathbf{C}(x,a)$ is increasing in x for each $a \in \mathbf{A}$, increasing in a for each $x \in \mathbf{X}$ and $\sum_{y=z}^{\infty} p_{xy}(a)$ is increasing in x for each $z \in \mathbf{X}$ and $a \in \mathbf{A}$. Thus, this model satisfies Assumptions 4.2 and 4.3. To verify that Assumption 4.4 holds in this model, we note that if $x' > x$ then

$$\sum_{y=z}^{\infty}[p_{x'y}(2) - p_{xy}(2)] = 0,$$

and by (5.1)

$$\sum_{y=z}^{\infty}[p_{x'y}(1) - p_{xy}(1)] \geqslant 0.$$

Therefore, $\sum_{y=z}^{\infty} p_{xy}(a)$ is subadditive on $\mathbf{X} \times \mathbf{A}$. Thus, by Lemma 9 we obtain that h is subadditive on $\mathbf{X} \times \mathbf{A}$. Finally,

$$g(x,a,\gamma\beta^t) = g(x',a,\gamma\beta^t), \quad \forall x,x' \in \mathbf{X},$$

and

$$h(x,2,\gamma\beta^t) = J_\beta(0,\gamma\beta^{t+1}) = h(x',2,\gamma\beta^t), \quad \forall x,x' \in \mathbf{X}$$

imply respectively the conditions (a) and (b) of Theorem 18. Therefore $g \cdot h$ is subadditive on $\mathbf{X} \times \mathbf{A}$. The result follows from Theorem 12. □

For the rest of this chapter, when $\{x_k^*\}_{k=0}^{\infty}$ or π^* is written, it is to be understood as the sequence of thresholds or, respectively, the optimal policy obtained in Proposition 2.

5.3 Ultimately Stationary Optimal Policies

It is well known that for the risk-neutral replacement problem, there exists a threshold stationary β-optimal policy; see [6, 40, 41]. The case of interest is that for which R is small enough so that the threshold x^* of that optimal policy is finite. Moreover, intuitively, it is to be expected that when $\mid \gamma \mid$ is small the risk-sensitive model should be close to the risk-neutral model. Thus, it is reasonable to expect that the following assumption holds for most situations of interest.

Assumption 5.2. $\{x_k^*\}_{k=0}^{\infty}$ does not diverge to ∞.

Jaquette [32, 33] showed that optimal policies are ultimately stationary for every CMC with exponential utility criterion and *finite* state space. Here, it shall be proved in Proposition 3 that, under Assumption 5.2, the optimal policy π^* is ultimately stationary for the equipment replacement model with infinite state space.

Lemma 11. *For arbitrary $\pi_1, \pi_2 \in \Pi$, and $x \in \mathbf{X}$, there exists $\gamma(x, \pi_1, \pi_2) > 0$, such that one (and only one) of the following statements is true*

$$J_\beta^{\pi_1}(x,\cdot) - J_\beta^{\pi_2}(x,\cdot) = 0, \quad on \quad (0, \gamma(x,\pi_1,\pi_2))$$
$$J_\beta^{\pi_1}(x,\cdot) - J_\beta^{\pi_2}(x,\cdot) > 0 \quad on \quad (0, \gamma(x,\pi_1,\pi_2))$$
$$J_\beta^{\pi_1}(x,\cdot) - J_\beta^{\pi_2}(x,\cdot) < 0 \quad on \quad (0, \gamma(x,\pi_1,\pi_2)).$$

Proof. For every $\pi \in \mathbf{\Pi}$ and $x \in \mathbf{X}$, the function $J_\beta^\pi(x,\cdot) : (0,\infty) \longrightarrow \mathbb{R}$ is given by

$$\begin{aligned} J_\beta^\pi(x,\gamma) &= E_x^\pi[e^{\gamma \mathcal{D}}] \\ &= \sum_{n=0}^{\infty} \left(\frac{E_x^\pi[\mathcal{D}^n]}{n!} \right) \gamma^n < \infty. \end{aligned} \tag{5.4}$$

It follows from (5.4) that $J_\beta^\pi(x,\cdot)$ is analytic in $(0,\infty)$. Therefore, the function

$$J_\beta^{\pi_1}(x,\cdot) - J_\beta^{\pi_2}(x,\cdot)$$

cannot have an infinite number of zeros in any finite interval without being identically zero, and thus the existence of a number $\gamma(x,\pi_1,\pi_2)$ as above follows. □

Lemma 12. *Set $\pi_1 = (f, f_1^*, f_2^*, \ldots)$ and $\pi_2 = (g, f_1^*, f_2^*, \ldots)$, where $f(x) = 2$ and $g(x) = 1\ \forall x$. Let $\bar{x}$ be an accumulation point of the sequence $\{x_k^*\}$. Then for $x = 0, \ldots \bar{x}$ there exists a number $\bar{\gamma} > 0$ such that on $(0,\bar{\gamma})$ one (and only one) of the following statements is true:*

$$\begin{aligned} &e^{\gamma c(x)} e^{\gamma R} J_\beta(0,\gamma\beta) - e^{\gamma c(x)} \sum_y p_{xy} J_\beta(y,\gamma\beta) = 0 \\ &e^{\gamma c(x)} e^{\gamma R} J_\beta(0,\gamma\beta) - e^{\gamma c(x)} \sum_y p_{xy} J_\beta(y,\gamma\beta) > 0 \\ &e^{\gamma c(x)} e^{\gamma R} J_\beta(0,\gamma\beta) - e^{\gamma c(x)} \sum_y p_{xy} J_\beta(y,\gamma\beta) < 0 \end{aligned} \tag{5.5}$$

Proof. It follows from the definition of the operator E_x^π that

$$J_\beta^{\pi_1}(x,\gamma) = e^{\gamma c(x)} e^{\gamma R} J_\beta(0,\gamma\beta), \text{ and}$$

$$J_\beta^{\pi_2}(x,\gamma) = e^{\gamma c(x)} \sum_y p_{xy} J_\beta(y,\gamma\beta).$$

Then, by Lemma 11, for each $0 \leqslant x \leqslant \bar{x}$, there exists $\gamma(x) > 0$ such that one and only one of the above statements (5.5) is true. Thus, the result follows by taking $\bar{\gamma} = \min\{\gamma(0), \gamma(1), \ldots \gamma(\bar{x})\}$. □

Now, we are ready to show that the policy π^* is ultimately stationary.

Proposition 3. *Under Assumptions 5.1 and 5.2, the policy $\pi^* = (f_0^*, f_1^*, \ldots)$ (see (5.2) is ultimately stationary, that is, there exists N such that $x_k^* = x_N^*$, $\forall k \geqslant N$.*

Proof. Let $\bar{x}$ be an accumulation point of the sequence $\{x_k^*\}$, and for $k = 0, 1, \ldots,$ let

$$h_k = e^{\gamma\beta^k R} J_\beta(0, \gamma\beta^{k+1}), \quad \text{and} \quad w_k(x) = \sum_y p_{xy} J_\beta(y, \gamma\beta^{k+1}).$$

Take N (large enough) so that $\gamma\beta^N < \bar{\gamma}$ and $x_N^* = \bar{x}$. Then since x_N^* is the threshold corresponding to f_N^*, we deduce that

$$h_N - w_N(x_N^* - 1) > 0 \text{ and } \quad h_N - w_N(x_N^*) \leqslant 0.$$

Thus, by Lemma 11 we obtain that $\forall k \geqslant N$ (i.e. $\gamma\beta^k < \gamma\beta^N$),

$$h_k - w_k(x_N^* - 1) > 0 \text{ and } h_k - w_k(x_N^*) \leqslant 0.$$

Therefore $x_k^* = x_N^* = \bar{x}$, $\forall k \geqslant N$. □

Thus, we have proved that the policy π^* has the form

$$\pi^* = (f_0^*, f_1^*, \ldots f_{N-1}^*, f^*, f^*, f^* \ldots), \tag{5.6}$$

where f^* is defined by $f^*(x) = 1$ if $x < \bar{x}$, and $f^*(x) = 2$ if $x \geqslant \bar{x}$. It is natural to expect the policy $(f^{*\infty})$ to be optimal for the risk-neutral replacement problem, given that for large t the (effective) risk-factor $\gamma\beta^t$ will be close to zero, i.e., close to the risk-null case. The following lemma will be used to show, in Proposition 4, that this is the case.

Lemma 13. *Let* $\mathcal{E}^\pi(x, \gamma) := \frac{1}{\gamma} \log E_x^\pi[e^{\gamma \mathcal{D}}]$ *be the certain equivalent of the random cost* $\mathcal{D}$. *Then*

$$\lim_{\gamma \to 0} \mathcal{E}^\pi(x, \gamma) = E_x^\pi[\mathcal{D}].$$

Proof. Let $u(\gamma) = E_x^\pi[e^{\gamma \mathcal{D}}]$. Thus

$$\begin{aligned} \lim_{\gamma \to 0} \mathcal{E}^\pi(x, \gamma) &= \lim_{\gamma \to 0} \frac{1}{\gamma} \log u(\gamma) = \lim_{\gamma \to 0} \frac{u'(\gamma)}{u(\gamma)} \\ &= u'(0) = E_x^\pi[\mathcal{D}] \end{aligned}$$

□

Proposition 4. *Let $\bar{\pi} = (f^*, f^*, \dots)$ be the tail of the policy π^* given in (5.6). Then under Assumptions 5.1 and 5.2, the stationary policy $\bar{\pi}$ is optimal for the risk-neutral replacement problem.*

Proof. It follows from (5.6) that for $k = N, N+1, \dots$, the policy $\bar{\pi}$ is optimal for the $\gamma\beta^k$-risk sensitive problem. Thus, for any stationary deterministic policy $\pi = (g, g, \dots)$,

$$J_\beta^{\bar{\pi}}(x, \gamma\beta^k) \leqslant J_\beta^{\pi}(x, \gamma\beta^k), \ \forall x, \ \forall k \geqslant N. \tag{5.7}$$

Since $\mathcal{U}_\gamma^{-1}(\cdot)$ is increasing, it follows from (5.7) that

$$\mathcal{E}^{\bar{\pi}}(x, \gamma\beta^k) \leqslant \mathcal{E}^{\pi}(x, \gamma\beta^k), \ \forall x, \ \forall k \geqslant N.$$

Thus, by taking limit when $k \to \infty$ one obtains from the previous lemma that

$$E_x^{\bar{\pi}}[\mathcal{D}] \leqslant E_x^{\pi}[\mathcal{D}],$$

and therefore $\bar{\pi}$ is risk-neutral optimal. □

5.4 Convergence of the Control Limits

In Section 4.2, the question was posed about how each decision rule $f_t(x)$, $t = 0, 1, \dots$, of the optimal policy varies with respect to t. In particular, in the machine replacement problem, it will be shown in Proposition 5 that under reasonable conditions those decision rules are monotone functions of t. More specifically, under certain conditions the function

$$L(J_\beta)(x, \cdot, \cdot) = g(x, \cdot, \cdot)h(x, \cdot, \cdot)$$

is superadditive on $\mathbf{A} \times \Gamma$, for fixed x.

The following general result gives a way of constructing a superadditive function by taking the composition of two other functions with certain properties.

Lemma 14. *Let $(S, \preccurlyeq_S)$ be a lattice, $G : S \longrightarrow \mathbb{R}$ an increasing (decreasing) and superadditive function on S, and $H : \mathbb{R} \longrightarrow \mathbb{R}$ an increasing and convex function. Then the composition of H and G is an increasing (decreasing) and superadditive function on S.*

Proof. The proof of this result follows immediately from the following identity:

$$\begin{aligned} H(G(s)) + H(G(r)) - H(G(s \vee r)) - H(G(s \wedge r)) = \\ [H(G(s)) - H(G(s \vee r) + G(s \wedge r) - G(r))] \\ + [H(G(s \vee r) + G(s \wedge r) - G(r)) - H(G(s \vee r))] \\ - H(G(s \wedge r)) + H(G(r)) \end{aligned}$$

□

Lemma 15. *The function $g(x, a, \gamma\beta^t) = e^{\gamma\beta^t r(a)}$ is increasing and superadditive on $\mathbf{A} \times \Gamma$.*

Proof. The result follows applying Lemma 14 with

$$H(z) = e^z \quad \text{and} \quad G(a, \gamma\beta^t) = \gamma\beta^t r(a).$$

□

Now, the following assumption will be made in the rest of this chapter.

Assumption 5.3 If $u \geqslant t$ then

$$J_\beta(x, \gamma\beta^t) - J_\beta(x, \gamma\beta^u) \leqslant J_\beta(x+1, \gamma\beta^t) - J_\beta(x+1, \gamma\beta^u).$$

Lemma 16. *Under Assumptions 5.1, 5.2 and 5.3, $h(x, a, \gamma\beta^t)$ is superadditive on $(\mathbf{A} \times \Gamma, \preccurlyeq)$ for each x.*

Proof. First, from the conditions of the model it is easy to see that

$$\sum_{y=z}^{\infty} p_{xy}(1) \geqslant \sum_{y=z}^{\infty} p_{xy}(2), \quad \text{for all } x, z \in \mathbf{X}.$$

Thus, by Lemma 6 and Assumption 5.3 we have that if $u \geqslant t$ then

$$\sum_{y=0}^{\infty} p_{xy}(1)[J_\beta(y, \gamma\beta^t) - J_\beta(y, \gamma\beta^u)] \geqslant \sum_{y=0}^{\infty} p_{xy}(2)[J_\beta(y, \gamma\beta^t) - J_\beta(y, \gamma\beta^u)],$$

and hence $h(x, 1, \gamma\beta^t) - h(x, 1, \gamma\beta^u) \geqslant h(x, 2, \gamma\beta^t) - h(x, 2, \gamma\beta^u)$, i.e., h is superadditive on the lattice $(\mathbf{A} \times \Gamma, \preccurlyeq)$. □

The following lemma proves that the optimal EDC decrease to 1 when the sensitivity coefficient decreases to 0. This result will be needed in the proof of Lemma 18.

Lemma 17. $\lim_{t \to \infty} J_\beta(x, \gamma\beta^t) = 1.$

Proof. By Dominated Convergence Theorem, $J_\beta^\pi(x, \gamma) \downarrow 1$ as $\gamma \downarrow 0$, $\forall \pi$. The claim follows from the inequality $1 \leqslant J_\beta(x, \gamma) \leqslant J_\beta^\pi(x, \gamma)$. □

Lemma 18. *Let* $S := (\mathbf{A} \times \Gamma, \preccurlyeq_{inv})$ *be the partial ordered set with* $(a, \gamma\beta^t) \preccurlyeq_{inv} (a', \gamma\beta^u)$ *if* $a \leqslant a'$ *and* $t \leqslant u$. *Then, under Assumptions 5.1, 5.2 and 5.3, the function* $g(x, \cdot, \cdot)h(x, \cdot, \cdot)$ *is subadditive on* S.

Proof. The result is proved by applying Theorem 17. Lemmas 15 and 16 imply respectively that g and h are subadditive functions on S. Now, take $s, r \in S$, noncomparable, i.e., $s = (1, \gamma\beta^u)$ and $r = (2, \gamma\beta^t)$ with $t \leqslant u$. Then since $g(x, 1, \gamma\beta^u) \leqslant g(x, 2, \gamma\beta^t)$ and $s \wedge r = (1, \gamma\beta^t)$, we need only to prove that

$$g(x, 1, \gamma\beta^t) = g(x, 1, \gamma\beta^u) \quad \text{and} \tag{5.8}$$

$$h(x, 1, \gamma\beta^u) \leqslant h(x, 1, \gamma\beta^t). \tag{5.9}$$

Then, (5.8) follows from the equality $g(x, 1, \gamma\beta^t) = 1$, $\forall t$, and (5.9) follows from Lemma 17. Thus, the proof is complete. □

Proposition 5. *Under Assumptions 5.1, 5.2 and 5.3, the decision rules* $f_t^*(x)$, $t = 0, 1, \ldots N-1$, *and* $f^*(x)$ *are increasing functions in* t, *for each* $x \in \mathbf{X}$. *Consequently, the sequence of thresholds* $\{x_0^*, x_1^*, \ldots\}$ *converges decreasingly to* $\bar{x}$.

Proof. By Lemma 18 we have that $L(J_\beta)(x, \cdot, \cdot)$ is subadditive on $(\mathbf{A} \times \Gamma, \preccurlyeq_{inv})$. Thus, by applying Theorem 13 we obtain that if $t \leqslant u$ then $f_t(x) \leqslant f_u(x)$, for each x, and consequently $x_0^* \geqslant x_1^* \geqslant \ldots$. The convergence of $\{x_k^*\}$ to $\bar{x}$ follows from Proposition 3. □

Chapter 6

Application 2: Optimal Resource Allocation

In this chapter we study a finite horizon CMC modeling an optimal allocation problem. In Section 6.1, the description of the problem is given. In Section 6.2, we apply results previously obtained in this work to prove structural properties of the optimal value function corresponding to the exponential total cost. It is also shown that there exists an optimal policy $(f_0, f_1, \ldots f_{N-1})$ such that the decision rules $f_t(x)$, $t = 0, \ldots N-1$, are increasing functions in x and increasing in t. Moreover, under additional conditions, we prove in Section 6.3 that the allocation problem can be reduced to a problem with two actions and that the optimal policy is of the threshold-type. Finally, in the same section, we apply those results to a particular example of a linear final cost; see [41] for an analysis of this problem with risk-neutral total cost. The optimal allocation problem can be described as follows.

6.1 Formulation of the Allocation Problem as a CMC

Suppose we have N stages to construct I successful components sequentially. At each stage we allocate a certain amount of money for the construction of a component. If a is the amount allocated, then the component constructed will be a success with probability $P(a)$, where P is a continuous nondecreasing function such that $P(0) = 0$. After each component is constructed, we are informed whether or not it is successful. If at the end of N stages, we are x components short, then a final penalty cost $C(x)$ is incurred, where $C(x)$ is increasing. The problem is to determine how much money to allocate at each stage to minimize the total expected utility. A CMC $(\mathbf{X}, \mathbf{A}, P, \mathbf{C})$ which models the described allocation problem can be defined by taking the state space $\mathbf{X} = \{0, 1, \ldots I\}$, the action space $\mathbf{A} = [0, M]$, where M is a positive real

number, the cost function $\mathbf{C}(x,a) = a$, and the transition probabilities

$$p_{xy}(a) = \begin{cases} P(a) & \text{if } y = x-1 \\ 1-P(a) & \text{if } y = x \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

The state X_t is the number of successful components still needed at time t and the action A_t is the amount of money allocated at time t.

For notational convenience in this section, $J_t(x,\gamma)$ will denote the minimal cost at state x with t stages to go, $x \in \mathbf{X}$ and $t = 0, 1, \ldots, N$. (Note that in Section 3.3, $J_t(x,\gamma)$ denoted the minimal cost at state x with $N-t$ stages to go.)

According to (3.2) and (3.3), $J_0(x,\gamma) = e^{\gamma C(x)}$ and for $t = 1, \ldots N$,

$$J_{t+1}(x,\gamma) = \inf_a \{e^{\gamma a}[P(a)J_t(x-1) + (1-P(a))J_t(x)]\} \tag{6.2}$$

$$= \inf_a \{e^{\gamma a}[J_t(x) - P(a)(J_t(x) - J_t(x-1)]\} \tag{6.3}$$

$$= \inf_a \{e^{\gamma a}[J_t(x-1) + (1-P(a))(J_t(x) - J_t(x-1)]\}. \tag{6.4}$$

6.2 Structural Properties of the Optimal Value Function and Policies.

First, we will show that the optimal value function $J_t(x,\gamma)$ is increasing in the state x and decreasing in the number of stages to go t.

Lemma 19. *The optimal value function $J_t(x,\gamma)$ is increasing in x and decreasing in t.*

Proof. We will apply Lemma 7 to prove that $J_t(x,\gamma)$ is increasing in x. First, we see that this model satisfies Conditions 1 and 3 of the mentioned lemma since $\mathbf{C}(x,a)$ is constant in x, the terminal cost $C(x)$ is increasing and $A(x) = \mathbf{A}$, $\forall x$. Finally, it follows from (6.1) that

$$\sum_{y=k}^{I} p_{xy}(a) = \begin{cases} 1 & \text{if } k \leqslant x-1 \\ 1-P(a) & \text{if } k = x \\ 0 & \text{if } k > x, \end{cases} \tag{6.5}$$

and hence, Condition 2 of Lemma 7 is valid for this model. Therefore, $J_t(x,\gamma)$ is increasing in x. Now, since $a = 0$ is an admissible action, it follows from (6.2) that

$$J_t(x,\gamma) \leqslant e^{\gamma \cdot 0}[P(0)J_{t-1}(x-1,\gamma) + (1-P(0))J_{t-1}(x,\gamma)],$$

and hence,

$$J_t(x,\gamma) \leqslant J_{t-1}(x,\gamma).$$

Thus, $J_t(x,\gamma)$ is decreasing in t for each x. □

The next goal is to show that the allocation problem has optimal policies that are increasing in x and decreasing in t. To this end, we will prove that

$$\log\left(e^{\gamma a}[P(a)J_t(x-1) + (1-P(a))J_t(x)]\right)$$

is subadditive on $\mathbf{X} \times \mathbf{A}$ and superadditive on $\mathbf{A} \times \{1, 2, \ldots N\}$, so that the mentioned monotonicity properties will follow from Theorems 14 and 15.

Set

$$G_t(x,a,\gamma) := e^{\gamma a} h_t(x,a,\gamma), \tag{6.6}$$

where $h_t(x,a,\gamma) = P(a)J_t(x-1) + (1-P(a))J_t(x)$. First, it will be shown that each of the structural properties of $\log G_t(x,a,\gamma)$ we need is equivalent to a structural property of $\log J_t(x,\gamma)$.

Proposition 6. *a)* $\log G_t(x,a,\gamma)$ *is subadditive on* $\mathbf{X} \times \mathbf{A}$ *iff* $\log J_t(x,\gamma)$ *is convex in* x. *b)* $\log G_t(x,a,\gamma)$ *is superadditive on* $\mathbf{A} \times \{1, \ldots N\}$ *iff* $\log J_t(x,\gamma)$ *is subadditive on* $\mathbf{A} \times \{0, 1, \ldots N\}$.

Proof. a) Let $a' > a$ and denote by $D_t(x) := J_t(x+1,\gamma) - J_t(x,\gamma)$. Then

$\log J_t(x,\gamma)$ is convex in x

$$\iff \log J_t(x+1) - \log J_t(x) \geqslant \log J_t(x) - \log J_t(x-1)$$

$$\iff J_t(x+1)J_t(x-1) \geqslant J_t^2(x)$$

$$\iff J_t(x)D_t(x) \geqslant J_t(x+1)D_t(x-1)$$

$$\iff (P(a') - P(a))J_t(x)D_t(x) \geqslant (P(a') - P(a))J_t(x+1)D_t(x-1)$$

$$\iff -P(a)J_t(x)D_t(x) - P(a')J_t(x+1)D_t(x-1) \geqslant$$
$$- P(a')J_t(x)D_t(x) - P(a)J_t(x+1)D_t(x-1)$$

$$\iff [J_t(x+1) - P(a)D_t(x)][J_t(x) - P(a')D_t(x-1)] \geqslant$$
$$[J_t(x+1) - P(a')D_t(x)][J_t(x) - P(a)D_t(x-1)]$$

$$\iff \frac{h_t(x+1,a)}{h_t(x,a)} \geqslant \frac{h_t(x+1,a')}{h_t(x,a')}$$

$$\iff \log h_t(x,a,\gamma) \text{ is subadditive on } \mathbf{X} \times \mathbf{A}$$

$$\iff \log G_t(x,a,\gamma) \text{ is subadditive on } \mathbf{X} \times \mathbf{A}.$$

Note the last step follows from the equality

$$\log G_t(x,a,\gamma) = \gamma a + \log h_t(x,a,\gamma).$$

b) Let $a' > a$. Then

$\log J_t(x,\gamma)$ is subadditive on $\mathbf{X} \times \{1,\dots N\}$

$$\iff \log J_{t+1}(x-1) - \log J_{t+1}(x) \geqslant \log J_t(x-1) - \log J_t(x)$$

$$\iff J_{t+1}(x-1)J_t(x) \geqslant J_{t+1}(x)J_t(x-1)$$

$$\iff J_t(x)D_{t+1}(x-1) \leqslant J_{t+1}(x)D_t(x-1)$$

$$\iff (P(a') - P(a))J_t(x)D_{t+1}(x-1) \leqslant (P(a') - P(a))J_{t+1}(x)D_t(x-1)$$

$$\iff -P(a)J_t(x)D_{t+1}(x-1) - P(a')J_{t+1}(x)D_t(x-1) \leqslant$$
$$- P(a')J_t(x)D_{t+1}(x-1) - P(a)J_{t+1}(x)D_t(x-1)$$

$$\iff [J_{t+1}(x) - P(a)D_{t+1}(x-1)][J_t(x) - P(a')D_t(x-1)] \leqslant$$
$$[J_{t+1}(x) - P(a')D_{t+1}(x-1)][J_t(x) - P(a)D_t(x-1)]$$

$$\iff \frac{h_{t+1}(x,a)}{h_t(x,a)} \leqslant \frac{h_{t+1}(x,a')}{h_t(x,a')}$$

$\iff \log h_t(x,a,\gamma)$ is superadditive on $\mathbf{X} \times \mathbf{A}$

$\iff \log G_t(x,a,\gamma)$ is superadditive on $\mathbf{X} \times \mathbf{A}$.

□

Now, we show that $\log J_t(x)$ is indeed convex in x for each t, and subadditive on $\mathbf{X} \times \{0,1,\dots N\}$. For the rest of this chapter, it will be assumed the following condition, which is reasonable for most situations of interest.

Assumption 6.1. The terminal cost $C(x)$ is convex.

Proposition 7. *Under Assumption 6.1, the following three statements hold:*

a) $\log J_t(x,\gamma)$ *is convex in* x *for each* t.

b) $\log J_t(x,\gamma)$ *is convex in* t *for each* x.

c) $\log J_t(x,\gamma)$ *is subadditive on* $(\mathbf{X} \times \{0,1,\dots N\}, \preccurlyeq)$.

Proof. First, note that (a), (b) and (c) are equivalent to

$$A_{x,t}: \quad \frac{J_t(x+2,\gamma)}{J_t(x+1,\gamma)} \geqslant \frac{J_t(x+1,\gamma)}{J_t(x,\gamma)}, \tag{6.7}$$

$$B_{x,t}: \quad \frac{J_{t+2}(x,\gamma)}{J_{t+1}(x,\gamma)} \geqslant \frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)}, \text{ and} \tag{6.8}$$

$$C_{x,t}: \quad \frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)} \geqslant \frac{J_{t+1}(x+1,\gamma)}{J_t(x+1,\gamma)} \tag{6.9}$$

respectively. We will show that those inequalities hold for $t = 0, 1, \ldots N-2$ and $x = 0, 1, \ldots I-2$. The proof will be by induction on $k = t + x$. We have that $C_{0,0}$ is true since J_t is decreasing in t (Lemma 19). $B_{0,0}$ is an obvious equality, and $A_{0,0}$ follows from the hypothesis about $C(x)$. Thus the inequalities are true for $k = 0$. We assume that they are true whenever $t + x < k$ and let $k = t + x$. Let's prove $C_{x,t}$. It follows from (6.3) that for some a, say $\bar{a}$,

$$J_{t+1}(x,\gamma) = e^{\gamma\bar{a}}[J_t(x,\gamma) - P(\bar{a})(J_t(x,\gamma) - J_t(x-1))],$$

and hence

$$\frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)} = e^{\gamma\bar{a}}\left[1 - P(\bar{a})\frac{J_t(x,\gamma) - J_t(x-1,\gamma)}{J_t(x)}\right]. \tag{6.10}$$

On the other hand, it follows from $A_{x-1,t}$ that

$$\frac{J_t(x,\gamma) - J_t(x-1,\gamma)}{J_t(x,\gamma)} \leqslant \frac{J_t(x+1,\gamma) - J_t(x,\gamma)}{J_t(x+1,\gamma)}.$$

Therefore, from (6.10) we obtain

$$\begin{aligned}\frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)} &\geq e^{\gamma\bar{a}}\left[1 - P(\bar{a})\frac{J_t(x+1,\gamma) - J_t(x,\gamma)}{J_t(x+1)}\right] \\ &\geq \frac{J_{t+1}(x+1,\gamma)}{J_t(x+1,\gamma)},\end{aligned}$$

and $C_{x,t}$ follows. In a similar way, to prove $B_{x,t}$ we have that it follows from (6.3) that for some a, say a',

$$J_{t+2}(x,\gamma) = e^{\gamma a'}[J_{t+1}(x,\gamma) - P(a')(J_{t+1}(x,\gamma) - J_{t+1}(x-1))],$$

and hence

$$\frac{J_{t+2}(x,\gamma)}{J_{t+1}(x,\gamma)} = e^{\gamma a'}\left[1 - P(a')\frac{J_{t+1}(x,\gamma) - J_{t+1}(x-1,\gamma)}{J_{t+1}(x)}\right]. \tag{6.11}$$

On the other hand, it follows from $C_{x-1,t}$ that

$$\frac{J_{t+1}(x,\gamma) - J_{t+1}(x-1,\gamma)}{J_{t+1}(x,\gamma)} \leqslant \frac{J_t(x,\gamma) - J_t(x-1,\gamma)}{J_t(x,\gamma)}.$$

Therefore, from (6.11) we obtain

$$\begin{aligned}\frac{J_{t+2}(x,\gamma)}{J_{t+1}(x,\gamma)} &\geqslant e^{\gamma a'}\left[1 - P(a')\frac{J_t(x,\gamma) - J_t(x-1,\gamma)}{J_t(x)}\right] \\ &\geqslant \frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)},\end{aligned}$$

and $B_{x,t}$ follows.

Finally, to prove $A_{x,t}$, note that $B_{x+1,t-1}$ is just

$$\frac{J_{t+1}(x+1,\gamma)}{J_t(x+1,\gamma)} \geqslant \frac{J_t(x+1,\gamma)}{J_{t-1}(x+1,\gamma)},$$

or equivalently,

$$J_{t+1}(x+1,\gamma)J_{t-1}(x+1,\gamma) \geqslant J_t^2(x+1,\gamma).$$

Thus to complete the proof of (6.7) we have to show that

$$J_t(x+2,\gamma)J_t(x,\gamma) \geqslant J_{t+1}(x+1,\gamma)J_{t-1}(x+1,\gamma). \tag{6.12}$$

It follows from (6.4) that for some a, say $\tilde{a}$,

$$J_t(x+2,\gamma) = e^{\gamma\tilde{a}}\left[J_{t-1}(x+1,\gamma) + (1 - P(\tilde{a}))(J_{t-1}(x+2,\gamma) - J_{t-1}(x+1,\gamma))\right],$$

and hence

$$\frac{J_t(x+2,\gamma)}{J_{t-1}(x+1,\gamma)} = e^{\gamma\tilde{a}}\left[1 + (1 - P(\tilde{a}))\frac{J_{t-1}(x+2,\gamma) - J_{t-1}(x+1,\gamma)}{J_{t-1}(x+1,\gamma)}\right]. \tag{6.13}$$

On the other hand, it follows from $A_{x,t-1}$ and $C_{x,t-1}$ that

$$\frac{J_{t-1}(x+2,\gamma)}{J_{t-1}(x+1,\gamma)} \geqslant \frac{J_t(x+1,\gamma)}{J_t(x,\gamma)},$$

and hence

$$\frac{J_{t-1}(x+2,\gamma)-J_{t-1}(x+1,\gamma)}{J_{t-1}(x+1,\gamma)} \geqslant \frac{J_t(x+1,\gamma)-J_t(x,\gamma)}{J_t(x,\gamma)}.$$

Thus, from (6.13) we obtain

$$\begin{aligned}\frac{J_t(x+2,\gamma)}{J_{t-1}(x+1,\gamma)} &\geqslant \frac{e^{\gamma\tilde{a}}}{J_t(x,\gamma)}[J_t(x,\gamma)+(1-P(\tilde{a}))(J_t(x+1)-J_t(x))] \\ &\geq \frac{J_{t+1}(x+1)}{J_t(x)},\end{aligned}$$

and (6.12) follows. Thus, the proof is complete. □

Corollary 1. *Under Assumption 6.1, $J_t(x)$ is convex in x for each t.*

Proof. Since $J_t(x) = \exp(\log J_t(x))$, the claim follows from Proposition 7 (a). □

We know that for the risk-neutral allocation problem there exists an optimal policy $\pi = (f_0, \ldots f_{N-1})$ such that $f_t(x)$ is increasing in x for each t, and increasing in t for each x; see [41]. In the following proposition we show a result analogous for the risk-sensitive case.

Proposition 8. *Under Assumption 6.1, there exists an optimal policy $\pi = (f_0^*, \ldots f_{N-1}^*)$ for the allocation problem with exponential total cost criterion such that $f_t^*(x)$ is increasing in x, for each t, and increasing in t, for each x.*

Proof. It follows from Propositions 6 and 7 that $\log G_t(x,a,\gamma)$ is subadditive on $\mathbf{X}\times\mathbf{A}$ and superadditive on $\mathbf{A}\times\{0,\ldots N-1\}$. Now, for $t = 0,1,\ldots N$, define $H_t(x,a,\gamma) = G_{N-t}(x,a,\gamma)$. Then $\log H_t(x,a,\gamma)$ is subadditive on $\mathbf{X}\times\mathbf{A}$ and subadditive on $\mathbf{A}\times\{0,\ldots N-1\}$. Finally, set $A_t^*(x) = \{a\in\mathbf{A} : \log H_t(x,a,\gamma) = \min_{a'}\log H_t(x,a',\gamma)$ and define $f_{t-1}^*(x) = \min A_t^*$. The result follows Theorems 14 and 15. □

6.3 Allocation Problem with P(a) Convex.

In this section, we analyze the allocation control problem with ETC criterion for the case in which the probability function $P(a)$ is convex. It is shown that, under the mentioned convexity condition, the optimal policy obtained in Proposition 8 has further structural properties. Moreover, those structured optimal policies are compared with those corresponding to the risk-neutral allocation problem (which are obtained in Appendix A). Finally, we apply the obtained results to the particular case of a linear terminal cost function, and again the conclusions are compared with those corresponding to the risk-neutral problem.

Throughout this section, $\pi^* = (f_0^*, \ldots, f_{t-1}^*)$ will denote the monotone optimal policy obtained in Proposition 8.

Proposition 9. *Assume that $P(a)$ is convex and twice differentiable. Then, under Assumption 6.1, the optimal allocation problem can be reduced to a problem with the action set $\{0, M\}$. Moreover, the optimal policy $\pi^* = (f_0^*, f_1^*, \ldots f_{N-1}^*)$ is of the threshold-type, that is, there exist states $x_0^*, x_1^*, \ldots, x_{N-1}^*$ such that*

$$f_t^*(x) = \begin{cases} 0 & \text{if } x < x_t^* \\ M & \text{if } x \geqslant x_t^*, \end{cases} \tag{6.14}$$

$t = 0, 1, \ldots N-1$. Moreover, the sequence of thresholds is decreasing.

Proof. First, we will show that for $a_x \in (0, M)$,

$$\frac{\partial G_t}{\partial a}(x, a_x, \gamma) = 0 \Longrightarrow \frac{\partial^2 G_t}{\partial^2 a}(x, a_x, \gamma) < 0;$$

that is, that there are no minimal points in $(0, M)$. Indeed, it follows from (6.6) that

$$G_t(x, a, \gamma) = e^{\gamma a}[J_t(x, \gamma) - J_t(x-1, \gamma)](1 - P(a)) + e^{\gamma a} J_t(x-1, \gamma),$$

which yields by differentiating both sides two times with respect to a:

$$\frac{\partial G_t}{\partial a}(x,a,\gamma) = -e^{\gamma a}[J_t(x,\gamma) - J_t(x-1,\gamma)][\gamma P(a) + P'(a)] + \gamma e^{\gamma a} J_t(x,\gamma), \quad (6.15)$$

and

$$\frac{\partial^2 G_t}{\partial^2 a}(x,a,\gamma) = -[J_t(x,\gamma) - J_t(x-1,\gamma)][\gamma^2 P(a) + 2\gamma P'(a) + P''(a)] + \gamma^2 e^{\gamma a} J_t(x,\gamma).$$

Hence,

$$\frac{\partial G_t}{\partial a}(x,a,\gamma) = 0 \Longleftrightarrow \frac{J_t(x,\gamma)}{J_t(x,\gamma) - J_t(x-1,\gamma)} = P(a) + \frac{1}{\gamma}P'(a), \quad (6.16)$$

and

$$\frac{\partial^2 G_t}{\partial^2 a}(x,a,\gamma) < 0 \Longleftrightarrow \frac{J_t(x,\gamma)}{J_t(x,\gamma) - J_t(x-1,\gamma)} < \frac{1}{\gamma^2}P''(a) + \frac{2}{\gamma}P'(a) + P(a). \quad (6.17)$$

Now, if $a_x \in (0,M)$ is such that $\frac{\partial G_t}{\partial a}(x,a_x,\gamma) = 0$, then after adding $\frac{1}{\gamma^2}P''(a_x) + \frac{1}{\gamma}P'(a_x)$ to the right side of 6.16 we obtain

$$\frac{1}{\gamma^2}P''(a_x) + \frac{2}{\gamma}P'(a_x) + P(a_x) > \frac{J_t(x,\gamma)}{J_t(x,\gamma) - J_t(x-1,\gamma)},$$

because $P''(a)$ and $P'(a)$ are positive. Thus, $\frac{\partial^2 G_t}{\partial^2 a}(x,a_x,\gamma) < 0$ follows from (6.17).

Since there are no minimal points in $(0,M)$, then we must have $f_t^*(x) \in \{0,M\}$ $\forall t$, $\forall x$. Moreover, if we define

$$x_t^* := \min\{x : f_t^*(x) = M\},$$

$t = 0,1,\ldots N-1$, then (6.14) follows from the fact that $f_t^*(x)$ is increasing in x. Finally, the sequence $\{x_t^*\}$ is decreasing since $f_t^*(x)$ is increasing in t. □

6.3.1 Allocation Problem with Linear Terminal Cost

Now, to gain further insight of the consequences of Proposition 9, we apply this proposition to compute the optimal policy in a particular example with linear final cost.

Example 1. Take $C(x) = 2x$, $\mathbf{A} = [0,1]$, and $P(a)$ convex. We start by computing $f^*_{N-1}(x)$. To do that, by Proposition 9, we need only to compare the values of the function $G_0(x,a,\gamma)$ at the extreme actions $a = 0$ and $a = 1$. We have that

$$\begin{aligned} G_0(x,a,\gamma) &= e^{\gamma a}[P(a)J_0(x-1,\gamma) + (1-P(a))J_0(x,\gamma)], \quad x \geq 1 \\ &= e^{\gamma a}[P(a)e^{\gamma(2x-2)} + (1-P(a))e^{2\gamma x}] \quad x \geq 1. \end{aligned}$$

Thus,

$$G_0(x,0,\gamma) = e^{2\gamma x} \tag{6.18}$$

and

$$G_0(x,1,\gamma) = e^{2\gamma x}[P(1)e^{-\gamma} + (1-P(1))e^{\gamma}]. \tag{6.19}$$

On the other hand, assuming that $P(1) \neq 1$, we obtain that

$$\begin{aligned} 1 \leq P(1)e^{-\gamma} + (1-P(1))e^{\gamma} &\iff e^{\gamma} \leq P(1) + e^{2\gamma}(1-P(1)) \\ &\iff (1-P(1))\left[e^{2\gamma} - \frac{1}{1-P(1)}e^{\gamma} + \frac{P(1)}{1-P(1)}\right] \geq 0 \\ &\iff e^{2\gamma} - \frac{1}{1-P(1)}e^{\gamma} + \frac{P(1)}{1-P(1)} \geq 0 \\ &\iff \left(e^{\gamma} - \frac{P(1)}{1-P(1)}\right)\left(e^{\gamma} - 1\right) \geq 0 \\ &\iff \gamma \geq \log\frac{P(1)}{1-P(1)}. \end{aligned} \tag{6.20}$$

Thus, it follows from (6.18), (6.19) and (6.20) that

a) if $\frac{1}{2} < P(1) < 1$ and $0 < \gamma \leq \log(\frac{P(1)}{1-P(1)})$ then

$$G_0(x,1,\gamma) \leq G_0(x,0,\gamma);$$

b) if $\frac{1}{2} < P(1) < 1$ and $\gamma \geq \log(\frac{P(1)}{1-P(1)})$ then

$$G_0(x,0,\gamma) \leq G_0(x,1,\gamma);$$

c) if $P(1) \leq \frac{1}{2}$ and $\gamma > 0$ then

$$G_0(x, 0, \gamma) < G_0(x, 1, \gamma);$$

d) if $P(1) = 1$ and $\gamma > 0$ then

$$G_0(x, 1, \gamma) < G_0(x, 0, \gamma).$$

Therefore the optimal decision rule f^*_{N-1} and the optimal value function J_1 for the cases (a) and (d) are given by

$$f^*_{N-1}(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geqslant 1, \end{cases} \tag{6.21}$$

and

$$J_1(x, \gamma) = \begin{cases} 1 & \text{if } x = 0 \\ e^{\gamma}[P(1)J_0(x-1) + (1 - P(1))J_0(x, \gamma)] & \text{if } x \geq 1, \end{cases} \tag{6.22}$$

and for (b) and (c) by

$$f^*_{N-1}(x) = 0, \quad \forall x$$

and

$$J_1(x, \gamma) = e^{2\gamma x}, \quad x \geq 0. \tag{6.23}$$

Now, to compute the optimal decision rules f^*_t, $t = 0, \ldots, N-2$, we will first prove each one of the following statements by induction on t:

I) if $\frac{1}{2} < P(1) < 1$ and $0 < \gamma \leq \log(\frac{P(1)}{1-P(1)})$ then, for $t = 1, \ldots N-1$,

$$J_t(1, \gamma) = e^{\gamma}[P(1) + (1 - P(1))J_{t-1}(1, \gamma)];$$

II) if $\frac{1}{2} < P(1) < 1$ and $\gamma \geq \log(\frac{P(1)}{1-P(1)})$ then, for $t = 1, \ldots N-1$,

$$J_t(x, \gamma) = J_0(x, \gamma), \quad x \in \mathbf{X};$$

III) if $P(1) \leq \frac{1}{2}$ and $\gamma > 0$ then for $t = 1, \ldots N-1$

$$J_t(x,\gamma) = J_0(x,\gamma), \quad x \in \mathbf{X};$$

IV) if $P(1) = 1$ and $\gamma > 0$ then for $t = 1, \ldots N-1$

$$J_t(1,\gamma) = e^{\gamma}[P(1) + (1-P(1))J_{t-1}(1,\gamma)].$$

First, let's prove (I). The validity of assertion (I) for $t = 1$ follows from (6.22). Next, by (6.2),

$$J_{t+1}(1,\gamma) = \min\{G_t(1,0,\gamma),\, G_t(1,1,\gamma)\},$$

where

$$G_t(x,a,\gamma) = e^{\gamma a}[P(a)J_t(x-1,\gamma) + (1-P(a))J_t(x,\gamma)], \quad x \geq 1. \tag{6.24}$$

Thus,

$$\begin{aligned} J_{t+1}(1,\gamma) &= \min\{J_t(1,\gamma), e^{\gamma}[P(1) + (1-P(1))J_t(1,\gamma)]\} \\ &= \min\{e^{\gamma}[P(1) + (1-P(1))J_{t-1}(1,\gamma)], \\ &\qquad e^{\gamma}[P(1) + (1-P(1))J_t(1,\gamma)]\} \qquad (6.25) \\ &= e^{\gamma}[P(1) + (1-P(1))J_t(1,\gamma)], \qquad (6.26) \end{aligned}$$

where (6.25) and (6.26) follow from the induction hypothesis and Lemma 19 respectively. Thus the proof of (I) is complete.

Now, let's prove (II). First, (6.23) implies that (II) holds for $t = 1$. Next, similarly as above,

$$\begin{aligned} J_{t+1}(I,\gamma) &= \min\{G_t(I,0,\gamma),\, G_t(I,1,\gamma)\} \\ &= \min\{J_t(I,\gamma), e^{\gamma}[P(1)J_t(I-1,\gamma) + (1-P(1))J_t(I,\gamma)]\} \qquad (6.27) \\ &= \min\{J_0(I,\gamma), e^{\gamma}[P(1)J_0(I-1,\gamma) + (1-P(1))J_0(I,\gamma)]\} \qquad (6.28) \\ &= \min\{J_0(I), J_0(I)[e^{-\gamma}P(1) + e^{\gamma}(1-P(1))], \\ &= J_0(I) \qquad (6.29) \end{aligned}$$

where (6.27), (6.28) and (6.29) follow from (6.24), the induction hypothesis and (6.20) respectively. Thus $f^*_{N-t-1}(I) = 0$ and since $f^*_{N-t-1}(x)$ is increasing in x, we obtain that $f^*_{N-t-1}(x) = 0$, for all x. Therefore

$$\begin{aligned} J_{t+1}(x,\gamma) &= \min\{G_t(x,0,\gamma),\, G_t(x,1,\gamma)\} \\ &= G_t(x,0,\gamma) \\ &= J_t(x,\gamma) \\ &= J_0(x,\gamma), \quad \forall x \in \mathbf{X}, \end{aligned}$$

and the proof of (II) is complete.

The proof of (III) is similar to the proof of (II) but in this case (6.29) follows from (6.20) since $P(1) \leq \frac{1}{2} \Longrightarrow \log \frac{P(1)}{1-P(1)} \leq 0$. The proof of (IV) is similar to the proof of (I).

Finally, it follows from (I), (II), (III) and (IV) that $f^*_t(x)$, $t = 0, 1, \ldots, N-2, N-1$, are given by

$$f^*_t(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geqslant 1 \end{cases} \tag{6.30}$$

if $\frac{1}{2} < P(1) < 1$ and $0 < \gamma \leq \log\left(\frac{P(1)}{1-P(1)}\right)$, or if $P(1) = 1$ and $\gamma > 0$; and

$$f^*_t(x) = 0, \quad \forall x$$

if $\frac{1}{2} < P(1) < 1$ and $\gamma \geq \log\left(\frac{P(1)}{1-P(1)}\right)$, or if $P(1) < \frac{1}{2}$ and $\gamma > 0$.

Remark 8. *Note that*

a) if $\frac{1}{2} < P(1) < 1$ *and* $\gamma \geq \log(\frac{P(1)}{1-P(1)})$ *then the preferences of the* γ*-decision maker differ from those of the risk-neutral decision maker: the* γ*-decision maker prefers the action* $a = 0$*, whereas the risk-neutral decision maker prefers the action* $a = 1$*; see Appendix A;*

b) if $P(1) = \frac{1}{2}$ *then the* γ*-decision maker prefers the action* $a = 0$*, whereas the risk-neutral decision maker is indifferent between the actions* $a = 0$ *and* $a = 1$*; see Appendix A.*

Chapter 7

CMC's WITH NON-DEGENERATE COSTS

In this chapter, we consider CMC's with cost per stage function depending on a disturbance. Unlike the model studied more frequently (see e.g., Bertsekas [6]), ours considers random disturbances whose distribution is given by a stochastic kernel that depends explicitly on the prior disturbance. We show that, due to this fact and to the introduction of risk-sensitivity, the risk-sensitive optimal value functions at each stage depend on the prior disturbance. Moreover, the risk-sensitive optimal policies yielded by the corresponding dynamic programming (DP) algorithm are not Markovian, because the optimal decision function at each stage depends on the prior disturbance in general. In Section 7.1, the model is described and the general notation is given. In Section 7.2, a DP algorithm within our framework is proved in full detail. Finally, in Section 7.3, we collect some optimality results for the infinite horizon optimal control problem associated to our general model.

7.1 The Model.

Let us consider a CMC specified by the tuple $(\mathbf{X}, \mathbf{A}, \mathbf{D}, Q_0, Q, F, \mathbf{C})$, where:

- $\mathbf{X}$, the state space, is a countable set.

- $\mathbf{A}$, the action (or control) set, is a countable set. To each $x \in \mathbf{X}$ we associate a finite non-empty subset $A(x)$ of $\mathbf{A}$. $A(x)$ represents the set of admissible actions when the system is in state x. The set $\mathbf{K} := \{(x, a) : x \in \mathbf{X},\ a \in A(x)\}$, is called the set of admissible state-action pairs.

- $\mathbf{D}$, the disturbance space , is a Borel subset of $\mathbb{R}$;

- Q_0, the initial transition law, is a stochastic kernel on $\mathbf{D}$ given $\mathbf{K}$;

• Q, the transition law, is a stochastic kernel on $\mathbf{D}$ given $\mathbf{D} \times \mathbf{K}$; $Q(B \mid \nu, x, a)$ is the probability that the system undergoes a disturbance from the set B, given that the last disturbance was ν, the present state is x and the chosen action is a. In addition,

• F, the transition function, is a measurable map from $\mathbf{K} \times \mathbf{D}$ into $\mathbf{X}$; $F(x, a, \nu)$ is the state at time $t+1$ if at time t, a state x and an action a occured, and a disturbance $\nu \in \mathbf{D}$ was selected. Finally,

• $\mathbf{C} : \mathbf{K} \times \mathbf{D} \longrightarrow \mathbb{R}$, the one-stage cost function, is a measurable function. We will assume that $\mathbf{C}$ is non-negative and bounded: $0 \leqslant \mathbf{C}(x, a, \nu) \leqslant K < \infty$ for every $(x, a) \in \mathbf{K}$, $\nu \in \mathbf{D}$.

Note that the CMC we are considering is not completely stationary since the kernel giving the transition probability from the first stage to the second one is different from that governing the next transitions.

The CMC $(\mathbf{X}, \mathbf{A}, \mathbf{D}, Q_0, Q, F, \mathbf{C})$ represents a stochastic dynamical system observed at times $t = 0, 1, \ldots$. The evolution of the system is as follows. Let X_t, A_t, and D_t denote the state, the action chosen and the disturbance at time $t \in \mathbb{N}$, respectively. If the system is in state $X_t = x \in \mathbf{X}$, and the control $A_t = a \in A(x)$ is chosen then a disturbance D_t occurs according to the transition probability $Q(\cdot \mid D_{t-1}, x, a)$ (or $Q_0(\cdot \mid x, a)$ if $t = 0$). If the occurred perturbation is ν then (i) a cost $\mathbf{C}(x, a, \nu)$ is incurred, and (ii) the system moves to a new state $X_{t+1} = F(x, a, \nu)$. Once the transition into the new state has occurred, a new action is chosen, and the process is repeated; see [1, 29, 40].

The admissible history spaces are defined by

$$\mathbf{H}_0 := \mathbf{X}, \quad \mathbf{H}_t := (\mathbf{K} \times \mathbf{D})^t \times \mathbf{X} \quad , \ t \geq 1,$$

and the canonical sample space is defined as

$$\Omega = (\mathbf{X} \times \mathbf{A} \times \mathbf{D})^\infty.$$

The state, action, disturbance, and history processes, denoted respectively by $\{X_t\}_{t\in\mathbf{T}}$, $\{A_t\}_{t\in\mathbf{T}}$, $\{D_t\}_{t\in\mathbf{T}}$ and $\{H_t\}_{t\in\mathbf{T}}$ are defined on the measurable space $(\Omega, \mathcal{B}(\Omega))$ by the projections

$$X_t(\omega) = x_t, \quad A_t(\omega) = a_t, \quad D_t(\omega) = \nu_t, \quad H_t(\omega) = (x_0, a_0, \nu_0, x_1, \ldots, \nu_{t-1}, x_t),$$

where $\mathcal{B}(\Omega)$ is the corresponding product sigma-algebra. This means that when the observed path of states, actions and disturbances is ω, the random variable X_t denotes the state at time t, A_t the chosen action at time t, D_t the disturbance occurred at time t and H_t the history up to time t. For $\pi \in \mathbf{\Pi}$ and $h_s \in H_s$, $P^{\pi}_{h_s}$ and $E^{\pi}_{h_s}$ denote the appropriate operators; see [29, 37].

7.2 Dynamic Programming.

For $h_s \in \mathbf{H}_s$, $s = 0, 1, \ldots, n-1$, the EDC (ETC) to go from time s to time n due to a policy $\pi \in \mathbf{\Pi}$ is defined as

$$u^{\pi}_{\beta,s}(h_s, \gamma) := E^{\pi}_{h_s}\left[\mathcal{U}_{\gamma}(\mathcal{D}_s)\right],$$

where $\mathcal{D}_s := \sum_{m=s}^{n-1} \beta^m \mathbf{C}(X_m, A_m, D_m)$, $0 < \beta < 1$ ($\beta = 1$), and for $h_n \in \mathbf{H}_n$, $u^{\pi}_{\beta,n}(h_n, \gamma) := sgn\gamma$. The stochastic optimization problem is to find a policy π^* such that

$$u^{\pi^*}_{\beta,0}(x, \gamma) = \inf_{\pi}\{u^{\pi}_{\beta,0}(x, \gamma)\}, \quad \forall x \in \mathbf{X}.$$

We refer to such a policy as a γ-optimal policy.

The following result, which provides an algorithm for finding an optimal policy, also shows that for an arbitrary history $h_s = (x_0, a_0, \nu_0, \ldots \nu_{s-1}, x_s)$, the optimal utility-to-go

$$\inf_{\pi}\{u^{\pi}_{\beta,s}(h_s, \gamma)\}$$

does not depend on the whole history h_s but only on the disturbance ν_{s-1} and the state x_s.

Theorem 19. (Dynamic Programming Algorithm) *For $s = n, \ldots 1 \quad (s = 0)$, and $h_s = (x_0, a_0, \nu_0, \ldots, \nu_{s-1}, x_s) \in \mathbf{H}_s$, let $u^*_{\beta,s}$ be the function defined on $\mathbf{D} \times \mathbf{X} \quad (\mathbf{X})$ by*

$$u^*_{\beta,n}(\nu_{n-1}, x_n, \gamma) = (sgn\gamma), \tag{7.1}$$

$$\vdots \qquad\qquad \vdots$$

$$u^*_{\beta,s}(\nu_{s-1}, x_s, \gamma) = \min_{a_s \in A(x_s)} \left\{ \int_{\mathbf{D}} Q(d\nu_s \mid \nu_{s-1}, x_s, a_s) \right.$$
$$\left. e^{\gamma \beta^s \mathbf{C}(x_s, a_s, \nu_s)} u^*_{\beta,s+1}(\nu_s, x_{s+1}, \gamma) \right\} \tag{7.2}$$

$$\vdots \qquad\qquad \vdots$$

$$u^*_{\beta,1}(\nu_0, x_1, \gamma) = \min_{a_1 \in A(x_1)} \left\{ \int_{\mathbf{D}} Q(d\nu_1 \mid \nu_0, x_1, a_1) e^{\gamma \beta \mathbf{C}(x_1, a_1, \nu_1)} u^*_{\beta,2}(\nu_1, x_2, \gamma) \right\}$$

$$u^*_{\beta,0}(x_0, \gamma) = \min_{a_0 \in A(x_0)} \left\{ \int_{\mathbf{D}} Q_0(d\nu_0 \mid x_0, a_0) e^{\gamma \mathbf{C}(x_0, a_0, \nu_0)} u^*_{\beta,1}(\nu_0, x_1, \gamma) \right\} \tag{7.3}$$

where $x_{s+1} = F(x_s, a_s, \nu_s)$. Let $f_0 : \mathbf{X} \longrightarrow \mathbf{A}$ be a decision rule defined by

$$\int_{\mathbf{D}} Q_0(d\nu_0 \mid x_0, f_0(x_0)) e^{\gamma \mathbf{C}(x_0, f_0(x_0), \nu_0)} u^*_{\beta,1}(\nu_0, F(x_0, f_0(x_0), \nu_0), \gamma)$$
$$= \min_{a \in A(x_0)} \left\{ \int_{\mathbf{D}} Q_0(d\nu_s \mid x_0, a) e^{\gamma \mathbf{C}(x_0, a, \nu_0)} u^*_{\beta,1}(\nu_0, x_1, \gamma) \right\},$$

and for $s = 1, 2, \ldots n-1$, let $f_s : \mathbf{D} \times \mathbf{X} \longrightarrow \mathbf{A}$ be a decision rule defined by

$$\int_{\mathbf{D}} Q(d\nu_s \mid \nu_{s-1}, x_s, f_s(\nu_{s-1}, x_s))$$
$$e^{\gamma \beta^s \mathbf{C}(x_s, f_s(\nu_{s-1}, x_s), \nu_s)} u^*_{\beta,s+1}(\nu_s, F(x_s, f_s(\nu_{s-1}, x_s), \nu_s), \gamma)$$
$$= \min_{a_s \in A(x_s)} \left\{ \int_{\mathbf{D}} Q(d\nu_s \mid \nu_{s-1}, x_s, a_s) e^{\gamma \beta^s \mathbf{C}(x_s, a_s, \nu_s)} u^*_{\beta,s+1}(\nu_s, x_{s+1}, \gamma) \right\},$$

where $x_{s+1} = F(x_s, f_s(x_s), \nu_s)$. Then the deterministic policy $\pi^* = (f_0, f_1, \ldots, f_{n-1})$ is γ-optimal, and for $s = 0, 1, \ldots n-1$,

$$u^*_{\beta,s}(\nu_{s-1}, x_s, \gamma) = \inf_{\pi \in \mathbf{\Pi}} \{u^\pi_{\beta,s}(h_s, \gamma)\}, \quad \forall\ h_s = (x_0, a_0, \nu_0, \ldots \nu_{s-1}, x_s) \in \mathbf{H}_s.$$

The functions $u^*_{\beta,s}$ are called the optimal value functions since they give the optimal utility-to-go for each history h_s.

Proof. Let $\pi = (q_0, \ldots, q_{n-1})$ be an arbitrary policy and $h_s = (h_{s-1}, a_{s-1}, \nu_{s-1}, x)$ an arbitrary history up to time s. To prove the theorem, we will prove that for $s = 0, 1, \ldots n-1$,

$$u^{\pi}_{\beta,s}(h_s, \gamma) \geqslant u^{*}_{\beta,s}(\nu_{s-1}, x, \gamma), \tag{7.4}$$

with equality if $\pi = \pi^*$, i.e.,

$$u^{\pi^*}_{\beta,s}(h_s, \gamma) = u^{*}_{\beta,s}(\nu_{s-1}, x, \gamma). \tag{7.5}$$

The proof of (7.4) and (7.5) is by backward induction. The inductive step holds for $t = n$ since

$$u^{\pi}_{\beta,n}(h_n, \gamma) = sgn\gamma = u^{*}_{\beta,n}(\nu_{n-1}, x_n, \gamma).$$

Now, assume that (7.4) and (7.5) hold for $s+1, \ldots n$. Then

$$u^{\pi}_{\beta,s}(h_s, \gamma) = E^{\pi}_{h_s}\left[\mathcal{U}_{\gamma}(\mathcal{D}_s)\right] \tag{7.6}$$

$$= \sum_{a \in A(x)} q_s(a \mid h_s) \int_{\mathbf{D}} Q(d\nu \mid \nu_{s-1}, x, a) e^{\gamma \beta^s \mathbf{C}(x,a,\nu)} u^{\pi}_{\beta,s+1}(h_s, a, \nu, F(x,a,\nu), \gamma) \tag{7.7}$$

$$\geq \sum_{a \in A(x)} q_s(a \mid h_s) \int_{\mathbf{D}} Q(d\nu \mid \nu_{s-1}, x, a) e^{\gamma \beta^s \mathbf{C}(x,a,\nu)} u^{*}_{\beta,s+1}(\nu, F(x,a,\nu), \gamma) \tag{7.8}$$

$$\geq \min_{a \in A(x)} \left\{ \int_{\mathbf{D}} Q(d\nu \mid \nu_{s-1}, x, a) e^{\gamma \beta^s \mathbf{C}(x,a,\nu)} u^{*}_{\beta,s+1}(\nu, F(x,a,\nu), \gamma) \right\} \tag{7.9}$$

$$= u^{*}_{\beta,s}(\nu_{s-1}, x, \gamma), \tag{7.10}$$

where (7.7) follows from the definition of the operators $E^{\pi}_{(\cdot)}$, (7.8) from the induction hypothesis, and (7.10) from (7.2). This proves (7.4). Now, if $\pi = \pi^*$ then equality holds throughout the previous calculations and (7.5) follows. □

Remark 9. *Note that, unlike the analogous result in the risk-null case, the optimal policies provided by the above theorem are deterministic but in general non-Markovian.*

Remark 10. *When $\beta = 1$, (7.2) becomes*

$$u^{*}_{s}(\nu_{s-1}, x_s, \gamma) = \min_{a \in A(x)} \left\{ \int_{\mathbf{D}} Q(d\nu_s \mid \nu_{s-1}, x_s, a) e^{\gamma \mathbf{C}(x_s, a, \nu_s)} u^{*}_{s+1}(\nu_s, x_{s+1}, \gamma) \right\} \tag{7.11}$$

where $u^{*}_{s}(\nu_{s-1}, x_s, \gamma) := u^{*}_{1,s}(\nu_{s-1}, x_s, \gamma)$.

7.3 Infinite Horizon Optimality Results.

In this section we collect some general results about the infinite horizon control problem associated to the CMC $(\mathbf{X}, \mathbf{A}, \mathbf{D}, Q, F, \mathbf{C})$, which will be necessary to study the inventory problem for an infinite horizon in Chapter 8. To formulate this inventory control problem as a CMC, it is sufficient to consider, and so we do for the rest of this chapter, a simplified model in which the disturbance space $\mathbf{D}$ is countable, and the transition law Q is a stochastic kernel on $\mathbf{D}$ given $\mathbf{K}$, i.e., Q is independent of the prior disturbance . The proofs of the mentioned results can be obtained by minor modifications of similar results in Chapter 3.

Denote $\mathcal{D} := \sum_{t=o}^{\infty} \beta^t \mathbf{C}(X_t, A_t, D_t)$. The infinite horizon EDC due to a policy π is defined as

$$\begin{aligned} J_\beta^\pi(x,\gamma) &:= E_x^\pi \left[(sgn\gamma) e^{\gamma \sum_{t=0}^{\infty} \beta^t \mathbf{C}(X_t, A_t, D_t)} \right] \\ &= E_x^\pi \left[\mathcal{U}_\gamma(\mathcal{D}) \right], \end{aligned} \tag{7.12}$$

$0 < \beta < 1$. The stochastic optimal control problem is to find a policy π^* within the class $\mathbf{\Pi}$ such that (7.12) is minimized, that is, such that

$$J_\beta^{\pi^*}(x,\gamma) = \inf_\pi \{ J_\beta^\pi(x,\gamma) \} =: J_\beta(x,\gamma). \tag{7.13}$$

We refer to the γ-optimal policy π^* as a EDC-optimal policy, and $J_\beta(x,\gamma)$ is the optimal EDC. The following two theorems establish, respectively, that the value of the optimal EDC satisfies a functional equation, and that this equation characterizes Markov deterministic EDC-optimal policies.

Theorem 20. (Discounted Optimality Equation) *For $t = 0, 1, \ldots$, the optimal value function satisfies the equation*

$$J_\beta(x, \gamma\beta^t) = \min_{a \in A(x)} \{ \sum_\nu Q(\nu \mid x, a) e^{\gamma \beta^t C(x,a,\nu)} J_\beta(F(x,a,\nu), \gamma\beta^{t+1}) \}$$

Theorem 21. *Let $\pi = (f_0, f_1, \dots,)$ be the policy such that $f_t(x)$ is defined by*

$$\sum_{\nu} Q(\nu \mid x, a) e^{\gamma \beta^t C(x, f_t(x), \nu)} J_\beta(F(x, f_t(x), \nu), \gamma \beta^{t+1}) = \min_{a \in A(x)} \sum_{d} Q(\nu \mid x, a) e^{\gamma \beta^t C(x, a, \nu)} J_\beta(F(x, a, \nu), \gamma \beta^{t+1}).$$

Then π is (EDC)-optimal for the infinite horizon problem.

Denoting

$$J_{\beta, n-s}(x, \gamma) := u_{\beta, s}(x, \gamma \beta^{-s}),$$

we obtain the following algorithm equivalent to (7.1)-(7.2):

$$J_{\beta,0}(x, \gamma) = sgn\gamma,$$

$$\vdots \qquad \vdots$$

$$J_{\beta, s+1}(x, \gamma) = \min_{a \in A(x)} \Big\{ \sum_{\nu} Q(\nu \mid x, a) e^{\gamma \mathbf{C}(x, a, nu} J_{\beta, s}(F(x, a, \nu), \gamma \beta) \Big\}.$$

We can interpret $J_{\beta,s}(x, \gamma)$ as the minimal EDC that can be obtained starting at state x, with risk-sensitivity coefficient γ, and proceeding for s stages.

The next theorem establish that the DP algorithm may be used to successively approximate $J_\beta(x, \gamma)$.

Theorem 22. (Value iteration):

$$\lim_{n \to \infty} J_{\beta, n}(x, \gamma) = J_\beta(x, \gamma).$$

Chapter 8

APPLICATION 3: SCHEDULING JOBS.

In this chapter, a CMC model of a jobs scheduling problem is studied. To facilitate comparisons with the results we derive in this chapter, we include the analysis of the stochastic optimal control problem corresponding to the risk-null performance criterion given by an expected total weighted completion time (see [6] and [38]). Then, we introduce risk- sensitivity by considering the minimization of the *expected exponential utility* of the total weighted completion time. We use a particular example to show how, similarly to the risk-neutral case, the optimal policies obtained from the DP algorithm developed in the previous chapter are Markovian (i.e., the decision functions at each stage do not depend on the prior disturbance as in the general case). It is interesting to note, as we shall show, that for the risk-sensitive criterion a simple interchange argument is not applicable, and thus the only general computational and analytical tool for this situation is the mentioned DP algorithm. Finally, by means of a simple example, we illustrate how the optimal schedule depends on the risk sensitivity coefficient γ.

8.1 Formulation of the Scheduling Jobs Problem as a CMC.

The key elements of a job scheduling problem are as follows; see [6, 38].

N: number of jobs to process.

T_i: processing time of job i. We assume that the processing times are independent random variables with known distribution.

w_i: weight (usually interpreted as a holding cost) of job i.

Z_i: random completion time of job i. The distribution of the completion times depend on the chosen schedule.

The problem is faced as follows. We have N jobs with independent processing times that are to be processed non preemptively (that is, determined before any processing begins) on a single machine. For simplicity we assume there are no arrivals, breakdowns, setup or switch over costs, or precedences. If job j is completed at time t, the cost incurred is $w_j t$, where $w_j \geqslant 0$. The problem is to find a job schedule such that the expected total weighted completion time $E\left[\sum_j w_j Z_j\right]$ is minimized, where the expectation is understood to be taken with respect to the joint distribution of the $\{T_i\}$.

A formulation of the previous problem within the context of CMC's is as follows. Define $x_0 := \phi$ (the empty set), $x_t :=$\{the collection of jobs completed up to time t\}. The (finite) state space $\mathbf{X}$ is defined as the collection of subsets of $\{1, 2, ..., N\}$, i.e., $\mathbf{X} = 2^{\{1,2,\ldots,N\}}$, and $\mathbf{A} := \{1, 2, ..., N\}$ denotes the action space. The set of admissible state-action pairs is given by $\mathbf{K} = \{(x, a) : x \in \mathbf{X}, a \in x^c\}$. The model is *event driven*, i.e., time is incremented when a job is completed. Let a_t denote the action taken at time t, i.e., the $t - th$ scheduled job, and compute

$$D_0 = T_{a_0}, \quad D_t = D_{t-1} + T_{a_t}, \quad t = 1, 2, \ldots, N-1. \tag{8.1}$$

Let the one stage cost function be given as

$$\mathbf{C}(x_t, a_t, D_t) = w_{a_t} D_t = w_{a_t}\Big(\sum_{i \in x_t} T_i + T_{a_t}\Big), \tag{8.2}$$

and the transition function F is given as

$$F(x, a, \nu) = x \cup \{a\}.$$

The random disturbance D_t given by (8.1) takes values $\nu_t \in \mathbb{R}^+$ and has a distribution

$$P(D_0 \in B) = P_{T_{a_0}}(B),$$

$$P(D_t \in B \mid \nu_{t-1}, a_t) = P_{\nu_{t-1}+T_{a_t}}(B), \quad t = 1, \ldots, N-1,$$

where B is a Borel subset of $\mathbb{R}^+$, and P_X denotes the distribution of the random variable X. The history spaces are given by $\mathbf{H}_0 = \{(\phi)\}$, and for $t = 1, 2, \ldots N$, $\mathbf{H}_t = (\mathbf{K} \times \mathbb{R}^+)^{t-1} \times \mathbf{X}$. The initial stochastic kernel $Q_0(\cdot \mid \cdot)$ on $\mathbb{R}^+$ given $\mathbf{A}$ is defined by

$$Q_0(\cdot \mid a) := P_{T_a}(\cdot),$$

and the stochastic kernel $Q(\cdot \mid \cdot)$ on $\mathbb{R}^+$ given $\mathbb{R}^+ \times \mathbf{A}$ is defined by

$$Q(\cdot \mid \nu, a) := P_{\nu+T_a}(\cdot).$$

An admissible Markovian deterministic policy $\pi = \{f_0, f_1, ..., f_{n-1}\} \in \mathbf{\Pi_{MD}}$ satisfies $f_t(x) \in x^c$. To the deterministic policy $\pi = \{f_0, ..., f_{n-1}\}$ corresponds the sequential order of the jobs

$$(f_0(\phi), f_1(\{f_0(\phi)\}), f_2(\{f_0(\phi), f_1(\{f_0(\phi)\})\}), ...).$$

Thus, a policy $\pi \in \mathbf{\Pi_{MD}}$ is associated with an open-loop schedule $\{a_0, \ldots, a_{N-1}\}$.

8.2 Risk Neutral Case.

Let $\pi = (a_0, a_1, \ldots, a_{N-1}) \in \mathbf{\Pi}_{MD}$, and ϕ the initial state. The total cost incurred by π is given by

$$u_0^\pi(\phi) := \tilde{\mathbf{C}}(\phi, a_0) + \ldots + \tilde{\mathbf{C}}(\{a_0, \ldots, a_{N-2}\}, a_{N-1}), \tag{8.3}$$

where

$$\tilde{\mathbf{C}}(x_t, a_t) = w_{a_t}\Big(\sum_{i \in x_t} ET_i + ET_{a_t}\Big). \tag{8.4}$$

Remark 11. *Note that given the linear dependence of risk neutral criteria on the one stage costs, all that one needs in this case is the mean processing times given by (8.4). However, as we shall see later, for risk-sensitive criteria the dependence of the one stage cost on the "disturbance" T_{a_t} will need to be made explicit, and hence the mean processing times will not suffice in that situation. Thus, the explicit use of (8.2) will be needed.*

The optimal cost, $u_0^*(\phi) := \inf_{\pi \in \Pi_{MD}} \{u_0^\pi(\phi)\}$, satisfies the last step of the standard Dynamic Programming recursion:

$$\begin{aligned} u_N^*(x) &= 0, \\ &\vdots \qquad \vdots \\ u_t^*(x) &= \min_{a \in x^c} \left\{ w_a \left[\sum_{k \in x \cup \{a\}} E[T_k] \right] + u_{t+1}^*(x \cup \{a\}) \right\}. \end{aligned}$$

Of course, in this case the optimal schedule could also be obtained by total enumeration. As an illustration of this algorithm, we obtain next, the optimal schedule, and the optimal cost for a simple problem with 3 jobs.

Example 1. We suppose that there are 3 jobs to be processed in sequential order. Let f_1, f_2, f_3 be the distributions of T_1, T_2, T_3 respectively, where f_1, f_3 are concentrated in 1, and f_2 is given by

$$f_2 = \begin{cases} 0 & \text{with probability } \frac{9}{20} \\ 1 & \text{with probability } \frac{1}{2} \\ 10 & \text{with probability } \frac{1}{20} \end{cases}$$

The weights and expected processing times are given by

jobs	1	2	3
w_j	6	10	8
$E(T_j)$	1	1	1

We have that $u_3^*(\{1,2,3\}) = 0$,

$$\begin{aligned} u_2^*(\{1,2\}) &= w_3 E[T_1 + T_2 + T_3] = \mathbf{24}, \\ u_2^*(\{1,3\}) &= w_2 E[T_1 + T_2 + T_3] = \mathbf{30}, \\ u_2^*(\{2,3\}) &= w_1 E[T_1 + T_2 + T_3] = \mathbf{18}. \end{aligned}$$

Substituting these values in u_1^* we obtain

$$\begin{aligned} u_1^*(\{1\}) &= \min\left\{w_2 E[T_1 + T_2] + u_2^*(\{1,2\}),\ w_3 E[T_1 + T_3] + u_2^*(\{1,3\})\right\} = \mathbf{44}, \\ u_1^*(\{2\}) &= \min\left\{w_1 E[T_1 + T_2] + u_2^*(\{1,2\}),\ w_3 E[T_2 + T_3] + u_2^*(\{2,3\})\right\} = \mathbf{34}, \\ u_1^*(\{3\}) &= \min\left\{w_1 E[T_1 + T_3] + u_2^*(\{1,3\}),\ w_2 E[T_2 + T_3] + u_2^*(\{2,3\})\right\} = \mathbf{38}. \end{aligned}$$

Finally, we calculate the optimal cost:

$$\begin{aligned} u_0^*(\phi) &= \min\left\{w_1 E(T_1) + u_1^*(\{1\}),\ w_2 E(T_2) + u_1^*(\{2\}),\ w_3 E(T_3) + u_1^*(\{3\})\right\} \\ &= \min\left\{6 + 44, 10 + 34, 8 + 38\right\} = \mathbf{44}, \end{aligned}$$

and the optimal schedule is $\mathbf{S} = (\mathbf{2\,3\,1})$.

Interchange Argument. In problems for which there exists an optimal open loop policy, an *interchange argument* may be used to obtain (necessary) optimality conditions [6, 38]. Let $\pi^* = (i_0^*, i_1^*, \ldots, i_{r-1}^*, \mathbf{i}, \mathbf{j}, i_{r+2}^*, \ldots, i_{N-1}^*)$ be an optimal schedule, and consider the schedule $\pi' = (i_0^*, \ldots, i_{r-1}^*, \mathbf{j}, \mathbf{i}, i_{r+2}^*, \ldots i_{N-1}^*)$ obtained from π^* by interchanging the jobs $\mathbf{i}$ and $\mathbf{j}$. We have that

$$\begin{aligned} u_0^{\pi^*}(\phi) &= c_1 + \tilde{\mathbf{C}}(\{i_0^*, \ldots, i_{r-1}^*\}, \mathbf{i}) + \tilde{\mathbf{C}}(\{i_0^*, \ldots, i_{r-1}^*, \mathbf{i}\}, \mathbf{j}) + c_2, \\ u_0^{\pi'}(\phi) &= c_1 + \tilde{\mathbf{C}}(\{i_0^*, \ldots, i_{r-1}^*\}, \mathbf{j}) + \tilde{\mathbf{C}}(\{i_0^*, \ldots, i_{r-1}^*, \mathbf{j}\}, \mathbf{i}) + c_2, \end{aligned}$$

where

$$c_1 = \tilde{\mathbf{C}}(\phi, i_0^*) + \tilde{\mathbf{C}}(\{i_0^*\}, i_1^*) + \ldots + \tilde{\mathbf{C}}(\{i_0^*, \ldots i_{r-2}^*\}, i_{r-1}^*),$$

and

$$c_2 = \tilde{\mathbf{C}}(\{i_0^*, \ldots \mathbf{i}, \mathbf{j}\}, i_{r+2}^*) + \ldots + \tilde{\mathbf{C}}(\{i_0^*, \ldots i_{N-2}^*\}, i_{N-1}^*).$$

Since π^* is supposed to be optimal, we obtain

$$w_{\mathbf{i}} E(t_r + T_{\mathbf{i}}) + w_{\mathbf{j}} E(t_r + T_{\mathbf{i}} + T_{\mathbf{j}}) \leqslant$$
$$w_{\mathbf{j}} E(t_r + T_{\mathbf{j}}) + w_{\mathbf{i}} E(t_r + T_{\mathbf{j}} + T_{\mathbf{i}}),$$

where $t_r = \sum_{k=0}^{r-1} T_{i_k^*}$. Canceling equal terms in the last inequality we obtain

$$\frac{w_{\mathbf{j}}}{E(T_{\mathbf{j}})} \leqslant \frac{w_{\mathbf{i}}}{E(T_{\mathbf{i}})}. \tag{8.5}$$

Proposition 10. *The rule that processes the jobs in decreasing order of* $\frac{w_k}{E(T_k)}$ *is optimal.*

Proof. Suppose that the schedule $\pi^* = (i_0, i_1, \ldots, i_{r-1}, \mathbf{i}, \mathbf{j}, i_{r+2}, \ldots i_{N-1})$ is optimal and the jobs $\mathbf{i}$ and $\mathbf{j}$ do not satisfy the rule, that is, $\frac{w_{\mathbf{i}}}{E(T_{\mathbf{i}})} < \frac{w_j}{E(T_j)}$. Then by taking $\pi' = (i_0, i_1, \ldots, i_{r-1}, \mathbf{j}, \mathbf{i}, i_{r+2}, \ldots i_{N-1})$ we obtain from (8.5) that

$$u_0^{\pi^*}(\phi) \geqslant u_0^{\pi'}(\phi).$$

This contradicts the optimality of π^*. □

This result allows us to give a simple algorithm to find the optimal schedule:

Scheduling Algorithm (Risk-null case.)

Step 1: Compute $E(T_i)$ for all i, and write

$$R_i = \frac{w_i}{E(T_i)}.$$

Step 2: Sort $\{R_i\}$ in decreasing order; break ties in any arbitrary way. Denote by

$$\{R_{i_1^*}, R_{i_2^*}, \ldots R_{i_N^*}\}$$

the sorted array, i.e.,

$$R_{i_k^*} \geqslant R_{i_{k+1}^*}.$$

Step 3: Obtain the optimal schedule

$$\pi^* = \{i_1^*, i_2^*, \ldots i_N^*\}.$$

As an illustration of the simplicity of this algorithm, we obtain the optimal schedule in Example 1. We have that

$$R_1 = 6, \quad R_2 = 10, \quad R_3 = 8.$$

Therefore, the optimal schedule is $\mathbf{S} = (\mathbf{231})$, since

$$R_2 > R_3 > R_1.$$

Remark 12. *In the optimal schedule we have that job 2, the first job scheduled, is the one with processing time having largest variance. Note that the variability of job 2 impacts greatly the variability of the total cost. For example if T_2 takes the value 10, then the total cost will be $10(10) + 8(10+1) + 6(10+1+1) = 260$, while if T_2 takes the value 0, then the total cost will be 20. Thus a risk averse Decision Maker (DM) may prefer to schedule later the job 2; that is, the schedule $S = (231)$ may not to be his/her optimal schedule.*

8.3 Risk Sensitive Case.

Suppose now that the Decision Maker (DM) has constant risk sensitivity coefficient $\gamma \neq 0$; see [35, 39, 47]. The case $\gamma > 0$ ($\gamma < 0$) corresponds to a risk-averse (risk-seeking) DM, and if $\gamma = 0$ we recover the standard risk-neutral situation. Thus the exponential total cost due to a policy $\pi = (a_0, a_1, \ldots, a_{N-1}) \in \mathbf{\Pi}_{MD}$ would be

computed as

$$\begin{aligned} u_0^{\pi}(\phi,\gamma) &= E^{\pi}\left[\mathcal{U}_{\gamma}\Big(\sum_{t=0}^{N-1}\mathbf{C}(x_t,a_t,D_t)\Big)\right] \\ &= E^{\pi}[(sgn\gamma)e^{\sum_{t=0}^{N-1}\mathbf{C}(x_t,a_t,D_t)}], \end{aligned} \tag{8.6}$$

and thus one sees that (8.4) is not enough to compute (8.6), but instead (8.2) is needed.

For $\pi = (a_0, \ldots, a_{N-1}) \in \mathbf{\Pi}_{MD}$, and $h_t \in \mathbf{H}_t$, the utility-to-go functions u_t^{π} are given by $u_N^{\pi}(h_N) = sgn\gamma$, and for $t < N$ by

$$u_t^{\pi}(h_t,\gamma) = E_{h_t}^{\pi}\left[\mathcal{U}_{\gamma}\left(\sum_{l=t}^{N-1} w_{a_l} D_l\right)\right],$$

where $D_0 = T_{a_0}$, $D_l = D_{l-1} + T_{a_l}$, $l = 1, \ldots N-1$. According to Theorem 19, the optimal utility-to-go functions $u_t^*(\nu_{t-1}, x_t, \gamma)$ satisfy the recursion $u_N^*(\nu_{N-1}, x, \gamma) = sgn\gamma$, for $t = N-1, \ldots, 1$,

$$u_t^*(\nu_{t-1}, x, \gamma) = \min_{a \in x^c}\left\{\int_{\mathbb{R}^+} Q(d\nu \mid \nu_{t-1}, a)\left[e^{\gamma\beta^t w_a \nu} u_{t+1}^*(\nu, x \cup \{a\}, \gamma)\right]\right\},$$

where $Q(\cdot \mid \nu_{t-1}, a) = P_{\nu_{t-1}+T_a}(\cdot)$, and

$$u_0^*(\phi,\gamma) = \min_{a \in \mathbf{A}}\left\{\int_{\mathbb{R}^+} Q_0(d\nu \mid a)\left[e^{\gamma w_a \nu} u_1^*(\nu, \{a\}, \gamma)\right]\right\},$$

where $Q_0(\cdot \mid a) = P_{T_a}(\cdot)$.

Next, we apply the DP algorithm to Example 1. This particular example will illustrate how, for the jobs scheduling problem, the optimal decision function at each stage obtained from the DP algorithm does not depend on the prior disturbance (as it does in the general case).

Example 1 (revisited.) For simplicity we take $\gamma = 1$ (risk averse case.) Then we have that

$$u_3^*(\nu_2, \{1,2,3\}, \gamma) = 1.$$

Thus,

$$\begin{aligned} u_2^*(\nu_1,\{2,3\},\gamma) &= \int_{\mathbb{R}^+} Q(d\nu_2 \mid \nu_1,1)e^{w_1\nu_2} = e^{w_1\nu_1}E[e^{w_1T_1}] = e^{6\nu_1}e^6; \\ u_2^*(\nu_1,\{1,3\},\gamma) &= \int_{\mathbb{R}^+} Q(d\nu_2 \mid \nu_1,2)e^{w_2\nu_2} = e^{w_2\nu_1}E[e^{w_2T_2}] = e^{10\nu_1}E[e^{10T_2}]; \\ u_2^*(\nu_1,\{1,2\},\gamma) &= \int_{\mathbb{R}^+} Q(d\nu_2 \mid \nu_1,3)e^{w_3\nu_2} = e^{w_3\nu_1}E[e^{w_3T_3}] = e^{8\nu_1}e^8. \end{aligned} \tag{8.7}$$

Substituting these values in u_1^* we obtain

$$\begin{aligned} u_1^*(\nu_0,\{1\},\gamma) &= \min\left\{\int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,2)e^{w_2\nu_1}u_2^*(\nu_1,\{1,2\},\gamma)\ , \right. \\ &\qquad \left. \int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,3)e^{w_3\nu_1}u_2^*(\nu_1,\{1,3\},\gamma)\right\} \\ &= \min\left\{e^{(w_2+w_3)\nu_0}E[e^{w_3T_3}]E[e^{(w_2+w_3)T_2}], e^{(w_2+w_3)\nu_0}E[e^{w_2T_2}]E[e^{(w_2+w_3)T_3}]\right\} \\ &= e^{(w_2+w_3)\nu_0}\min\left\{E[e^{w_3T_3}]E[e^{(w_2+w_3)T_2}], \quad E[e^{w_2T_2}]E[e^{(w_2+w_3)T_3}]\right\} \\ &= e^{18\nu_0}\min\left\{e^8E[e^{18T_2}], \quad e^{18}E[e^{10T_2}]\right\} \\ &= \mathbf{e^{18\nu_0}e^{18}E[e^{10T_2}]}; \end{aligned}$$

Similarly,

$$\begin{aligned} u_1^*(\nu_0,\{2\},\gamma) &= \min\left\{\int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,1)e^{6\nu_1}u_2^*(\nu_1,\{1,2\},\gamma)\ , \right. \\ &\qquad \left. \int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,3)e^{8\nu_1}u_2^*(\nu_1,\{2,3\},\gamma)\right\} \\ &= \int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,3)e^{8\nu_1}u_2^*(\nu_1,\{2,3\},\gamma) = \mathbf{e^{14\nu_0}e^{20}}; \end{aligned}$$

and

$$\begin{aligned} u_1^*(\nu_0,\{3\},\gamma) &= \min\left\{\int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,1)e^{6\nu_1}u_2^*(\nu_1,\{1,3\},\gamma)\ , \right. \\ &\qquad \left. \int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,2)e^{10\nu_1}u_2^*(\nu_1,\{2,3\},\gamma)\right\} \\ &= \int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,1)e^{6\nu_1}u_2^*(\nu_1,\{1,3\},\gamma) = \mathbf{e^{16\nu_0}e^{16}E[e^{10T_2}]}. \end{aligned} \tag{8.8}$$

Finally, let's calculate

$$u_0^*(\phi,\gamma) = \min\left\{\int_{\mathbb{R}^+} Q_0(d\nu_0 \mid 1)e^{6\nu_0}u_1^*(\nu_0,\{1\}), \quad \int_{\mathbb{R}^+} Q_0(d\nu_0 \mid 2)e^{10\nu_0}u_1^*(\nu_0,\{2\}), \right.$$
$$\left. \int_{\mathbb{R}^+} Q_0(d\nu_0 \mid 3)e^{8\nu_0}u_1^*(\nu_0,\{3\})\right\}$$
$$= \int_{\mathbb{R}^+} Q(d\nu_0 \mid 3)e^{8\nu_0}u_1^*(\nu_0,\{3\})\} = \mathbf{e^{40}E[e^{10T_2}]}. \tag{8.9}$$

Thus, from (8.7), (8.8) and (8.9), we obtain

$$\begin{aligned} u_0^*(\phi,\gamma) &= \int_{\mathbb{R}^+} Q_0(d\nu_0 \mid 3)e^{w_3\nu_0}\int_{\mathbb{R}^+} Q(d\nu_1 \mid \nu_0,1)e^{w_1\nu_1}\int_{\mathbb{R}^+} Q(d\nu_2 \mid \nu_1,2)e^{w_2\nu_2} \\ &= E[e^{(w_1+w_2+w_3)T_3}]E[^{(w_1+w_2)T_1}]E[^{w_2T_2}]. \end{aligned} \tag{8.10}$$

Thus, it follows from (8.10) that the optimal schedule is $\mathbf{S} = (\mathbf{3\,1\,2})$, which, as we expected (see Remark 12), is different to the optimal schedule $\mathbf{S} = (\mathbf{2\,3\,1})$ obtained in the risk neutral case, and one that schedules the most uncertain job for last. Now, since the optimal schedule is an open-loop policy and motivated by the risk neutral case, we can pose the following question.

Question: Applying an interchange argument, can we derive some simple necessary and sufficient conditions for optimality?

We proceed to investigate this question. Suppose that the schedule $\pi^* = (i_0, \ldots, i_{r-1}, \mathbf{i}, \mathbf{j}, i_{r+2}, \ldots, i_{N-1})$ is optimal. Let $M_l = \sum_{k=l}^{r-1} w_{i_k}$, for $l = 0, 1, \ldots r-1$, and let $S_{r+l} = \sum_{k=l}^{N-r-1} w_{i_{r+k}}$, for $l = 2, 3, \ldots N-r-1$. We have that

$$u_0^{\pi^*}(\phi,\gamma) = sgn\gamma K_1 E[e^{\gamma(w_i+w_j+S_{r+2})T_i}]E[e^{\gamma(w_j+S_{r+2})T_j}]K_2,$$

where $K_2 = E[e^{\gamma S_{r+2}T_{i_{r+2}}}]\ldots E[e^{\gamma S_{N-1}T_{i_{N-1}}}]$, and

$$K_1 = E[e^{\gamma(M_0+w_i+w_j+S_{r+2})T_{i_0}}]\ldots E[e^{\gamma(M_{r-1}+w_i+w_j+S_{r+2})T_{i_{r-1}}}].$$

Let $\pi' = (i_0, \ldots, i_{r-1}, \mathbf{j}, \mathbf{i}, i_{r+2} \ldots i_{N-1})$ be the schedule which interchanges the jobs **i** and **j**. Then

$$u_0^{\pi'}(\phi, \gamma) = sgn\gamma K_1 E[e^{\gamma(w_i + w_j + S_{r+2})T_j}] E[e^{\gamma(w_i + S_{r+2})T_i}] K_2.$$

Since the policy π^* is supposed to be optimal, we have that, if $\gamma > 0$ then

$$E[e^{\gamma(w_i+w_j+S_{r+2})T_i}]E[e^{\gamma(w_j+S_{r+2})T_j}] \leqslant E[e^{\gamma(w_i+w_j+S_{r+2})T_j}]E[e^{\gamma(w_i+S_{r+2})T_i}], \quad (8.11)$$

and if $\gamma < 0$ then the reverse inequality holds. Hence, we arrive at the following answer.

Answer: As opposite to the risk-neutral case, we notice that by applying an interchange argument, we do not derive any simple general condition for optimality. Observe that (8.11) involves the jobs scheduled from the job i onward. Thus, in this case, the DP algorithm is the only way we have to compute the optimal schedule. Certainty, a possibility is to explore particular distributions for $\{T_i\}$, e.g., see [36], and perhaps (8.11) may then yield simpler optimality conditions.

8.4 Optimal Schedule Dependence on γ

In this section, we will illustrate, via a simple example with 2 jobs, how the optimal schedule depends on γ. We will show that there exists $\gamma^* > 0$ such that $\mathbf{S} = (\mathbf{1}\ \mathbf{2})$ will be optimal for a DM with risk sensitivity coefficient less or equal than γ^*, while $\mathbf{S} = (\mathbf{2}\ \mathbf{1})$ will be optimal for a DM with risk coefficient great or equal than γ^*.

Example 2. Suppose that we want to process 2 jobs whose weights and processing times are given by

jobs	1	2
w_j	7	6

T_1 takes values 2 and 4 with probability $\frac{1}{2}$ each one, and T_2 is concentrate at 3. Note that $E(T_1) = E(T_2) = 3$. Thus since

$$\frac{w_1}{E(T_1)} > \frac{w_2}{E(T_2)},$$

it follows from Proposition 1 that

$$\mathbf{S = (1\,2) \quad is\ optimal\ if\ the\ DM\ is\ risk\text{-}neutral.} \tag{8.12}$$

We consider now a risk-sensitive DM. Note that in any example with 2 jobs, the inequalities (8.11) turn out to be very simple to manipulate. We have that

$$E[e^{\gamma(w_1+w_2)T_1}]E[e^{\gamma w_2 T_2}] = \frac{e^{26\gamma}+e^{52\gamma}}{2}e^{18\gamma};$$

and

$$E[e^{\gamma(w_1+w_2)T_2}]E[e^{w_1 T_1}] = e^{39\gamma}\frac{e^{14\gamma}+e^{28\gamma}}{2}.$$

Thus,

$$\begin{aligned} 2\left(E\right.&\left.[e^{\gamma(w_1+w_2)T_1}]E[e^{\gamma w_2 T_2}] - E[e^{\gamma(w_1+w_2)T_2}]E[e^{w_1 T_1}]\right) \\ &= \left(e^{26\gamma}+e^{52\gamma}\right)e^{18\gamma} - e^{39\gamma}\left(e^{14\gamma}+e^{28\gamma}\right) \\ &= e^{44\gamma}g(\gamma), \end{aligned}$$

where $g(\gamma) = e^{26\gamma} - e^{23\gamma} - e^{9\gamma} + 1$. Thus, we obtain that

$$E[e^{\gamma(w_1+w_2)T_1}]E[e^{\gamma w_2 T_2}] > E[e^{\gamma(w_1+w_2)T_2}]E[e^{w_1 T_1}] \quad \text{if } g(\gamma) > 0; \tag{8.13}$$

and

$$E[e^{\gamma(w_1+w_2)T_1}]E[e^{\gamma w_2 T_2}] < E[e^{\gamma(w_1+w_2)T_2}]E[e^{w_1 T_1}] \quad \text{if } g(\gamma) < 0. \tag{8.14}$$

Lemma 20. *Let $g(\gamma) := e^{26\gamma} - e^{23\gamma} - e^{9\gamma} + 1$. Then, the following statements hold:*

i) $g(\gamma)$ has only two real zeros:

$$\gamma_1 = 0, \quad \text{and } \gamma^* \in (\frac{1}{3}\ln\frac{3}{2}, 1);$$

ii) $g(\gamma) < 0$ if $0 < \gamma < \gamma^*$;

iii) $g(\gamma) > 0$ if $\gamma < 0$, or $\gamma > \gamma^*$.

Proof. We have that

$$\begin{aligned} g'(\gamma) &= 26e^{26\gamma} - 23e^{23\gamma} - 9e^{9\gamma} \\ &= e^{9\gamma}h(\gamma), \end{aligned} \tag{8.15}$$

where

$$h(\gamma) = 26e^{17\gamma} - 23e^{14\gamma} - 9.$$

It is easy to see that $h'(\gamma) = 0$ if and only if $\gamma = \frac{1}{3}\ln\frac{161}{221}$. Therefore $h(\gamma)$ has at most two real zeros and hence, from (8.15), g' has at most two real zeros. Moreover, g' has a zero in $(0,1)$ since $g'(0) = 26 - 23 - 9 < 0$, and $g'(1) > 0$. Let $\tilde{\gamma} > 0$ be the smallest zero (if there exist two) of g'. Then since $g'(0) = -6$, we have that $g'(\gamma) < 0$, for all $\gamma \in (0, \tilde{\gamma})$, and therefore g is decreasing in $(0, \tilde{\gamma})$. Since $g(0) = 0$ we have that

$$g(\gamma) < 0, \quad \text{for all } \gamma \in (0, \tilde{\gamma}). \tag{8.16}$$

Now, $g(1) = e^9(e^{17} - e^{14} - 1) > 0$. Therefore there exists $\gamma^* \in (\tilde{\gamma}, 1)$ such that $g(\gamma^*) = 0$, and

$$g(\gamma) < 0, \quad \text{for all } \gamma \in (\tilde{\gamma}, \gamma^*) \tag{8.17}$$

Thus (8.16) and (8.17) imply ii).

Now, let $\gamma < 0$, and $b = -\gamma$. Then

$$\begin{aligned} g'(\gamma) = g'(-b) &= e^{-9b}\left(26e^{-17b} - 23e^{-14b} - 9\right) \\ &\quad e^{-9\gamma}\left(\frac{26 - 23e^{3b} - 9e^{17b}}{e^{17b}}\right) < 0. \end{aligned}$$

Then g is decreasing in $(-\infty, 0)$, and since $g(0) = 0$, we have that

$$g(\gamma) > 0, \quad \text{for all } \gamma \in (-\infty, 0). \tag{8.18}$$

Finally,

$$g(\gamma) > 0, \quad \text{if } \gamma > 1. \tag{8.19}$$

Thus, (8.18) and (8.19) imply iii).

Moreover,

$$g'(\gamma) = 0 \Longleftrightarrow \gamma = \frac{1}{3}\ln\frac{3}{2}. \tag{8.20}$$

Therefore $\tilde{\gamma} = \frac{1}{3}\ln\frac{3}{2}$, and g has only two real zeros: $\gamma_1 = 0$, and $\gamma^* > \frac{1}{3}\ln\frac{3}{2}$. Thus, the proof is complete. □

Proposition 11. *There exists* $\frac{1}{3}\ln\frac{3}{2} < \gamma^* < 1$ *such that:*

$$\mathbf{S} = (\mathbf{1}\ \mathbf{2}) \quad \textit{is optimal if } \gamma \leqslant \gamma^*,$$
$$\mathbf{S} = (\mathbf{2}\ \mathbf{1}) \quad \textit{is optimal if } \gamma \geqslant \gamma^*.$$

Proof. The proof follows from (8.12), (8.13), (8.14), and Lemma 20. □

Chapter 9

APPLICATION 4: INVENTORY CONTROL

In this chapter, a formulation of an inventory control problem as a CMC is given. To that end, we consider a simplified model in which the disturbance space **D** is countable, and the transition law Q is a stochastic kernel on **D** given **K**, i.e., Q is independent of the prior disturbance. In Section 9.2 (9.3), it is shown that the finite (infinite) horizon inventory control has a base-stock optimal policy. See, e.g., [6, 40] for an analysis of the inventory control problem with risk-neutral discounted cost.

9.1 Formulation of an Inventory Control Problem as a CMC

. At the beginning of each period, the manager of a warehouse determines current inventory of a product. He decides whether or not to order additional stock and therefore what will be the inventory level for that period. In doing so, he is faced with a tradeoff between the costs associated with keeping inventory and the penalties associated with being unable to satisfy customer demand. Let us denote

X_t: stock available at the beginning of the t-th period.

A_t: number of units ordered at time t.

D_t: demand at time t.

We assume that the demand has a known probability distribution $p_\nu = P[D_t = \nu]$, $\nu = 0, 1, \ldots, L$, $L \in \mathbb{N}$, $t = 0, 1, \ldots$, and that excess demand is back logged and filled as soon as additional inventory becomes available. Thus the inventory at time $t+1$ is related to the state, the action, and the demand at time t, through the transition

function

$$F(x, a, \nu) = x + a - \nu. \tag{9.1}$$

The state and action spaces are given by

$$\mathbf{X} = \{\dots, -1, 0, 1, \dots, M\}, \quad \text{and} \quad \mathbf{A} = \{0, 1, 2, \dots\},$$

where M is the capacity of the warehouse. For $x \in \mathbf{X}$, the set of admissible actions is the finite set

$$A(x) = \{0, 1, \dots M - x\}. \tag{9.2}$$

Note that $x < 0$ corresponds to back logged demand. The transition law Q is given by

$$Q(\nu \mid x, a) = p_\nu \tag{9.3}$$

and it is independent on (x, a). The cost incurred at each period t consists of two components:

(1) the purchasing cost, $c \cdot (a_t)$, where c is the cost per unit ordered; and

(2) the inventory-related cost, $H(x_t, a_t, \nu_t) = h \cdot (x_t + a_t - \nu_t)^+ + p \cdot (\nu_t - x_t - a_t)^+$, where h and p are the holding and penalty costs per unit respectively.

In finite-horizon problems, each remaining unit at the last period time has a salvage value of c. The discounted cost up to time N is denoted by

$$\begin{aligned}
\tilde{\mathcal{D}}_N &:= \sum_{t=0}^{N-1} \beta^t [c \cdot (a_t) + H(x_t, a_t, \nu_t)] - \beta^N c \cdot x_N \\
&= \sum_{t=0}^{N-1} \beta^t [c \cdot (x_t + a_t) + H(x_t, a_t, \nu_t) - \beta c \cdot (x_t + a_t - \nu_t)] \\
&\quad - \sum_{t=0}^{N-1} \beta^t c \cdot (x_t) + \sum_{t=0}^{N-1} \beta^{t+1} c \cdot (x_{t+1}) \\
&= \sum_{t=0}^{N-1} \beta^t [c \cdot (x_t + a_t) + H(x_t, a_t, \nu_t) - \beta c \cdot (x_t + a_t - \nu_t)] - c \cdot x_0.
\end{aligned}$$

Denote $C(x,a,\nu) := c\cdot(x+a) + H(x,a,\nu) - \beta c\cdot(x+a-\nu)$. Then

$$\tilde{\mathcal{D}}_N = \sum_{t=0}^{N-1} \beta^t C(x_t, a_t, d_t) - cx_0.$$

Let $\mathcal{D}_N := \sum_{t=0}^{N-1} \beta^t C(X_t, A_t, D_t)$. For a deterministic policy π, let

$$u^{\pi}_{\beta,0}(x,\gamma) := E^{\pi}_x[\mathcal{U}_\gamma(\mathcal{D}_N)].$$

Note that the policy that optimizes

$$\inf_{\pi}\{u^{\pi}_{\beta,0}(x,\gamma)\}$$

also optimizes

$$\inf_{\pi}\{E^{\pi}_x[\mathcal{U}_\gamma(\tilde{\mathcal{D}}_N)]\}.$$

In this section, we consider the optimal control problem with discounted cost $\mathcal{D}_N$. According to Theorem 19, the DP algorithm for this model is as follow:
$u_{\beta,N}(x,\gamma) = sgn\gamma$; and for $t = N-1,\ldots,0$,

$$u^*_{\beta,t}(x,\gamma) = \min_{a\in A(x)}\left\{\sum_{\nu} p_\nu e^{\gamma\beta^t C(x,a,\nu)} u^*_{\beta,t+1}(x+a-\nu,\gamma)\right\},$$

where $C(x,a,\nu) = c\cdot(x+a) + h\cdot(x+a-\nu)^+ + p\cdot(\nu-x-a)^+ - \beta c\cdot(x+a-\nu)$.

Remark 13. *Note that, while the optimal value function for the scheduling jobs problem studied in Chapter 8, $u^*_{\beta,s}(\nu_{s-1}, x_s, \gamma)$, depends on the disturbance-state pair (ν_{s-1}, x_s), the optimal value function for the inventory problem, $u^*_{\beta,s}(x_s,\gamma)$, depends only on the present state x_s. This is due to the fact that the transition law, Q, is independent on the prior perturbation.*

9.2 Base-Stock Optimal Policies.

In this section, it is shown that the inventory control model has a base-stock optimal policy. A base-stock policy is a decreasing policy $\pi = (f_0, f_1, \ldots)$ such that, for

$t = 0, 1, \ldots$, the decision rule f_t is given by

$$f_t(x) = \begin{cases} x_t^* - x & \text{if } x \leqslant x_t^* \\ 0 & \text{if } x > x_t^*, \end{cases} \tag{9.4}$$

where $x_0^*, x_1^*, \ldots$ are the base-stock levels.

The following theorem, which will be used to show that the inventory problem has a base-stock optimal policy, gives sufficient conditions under which the finite horizon CMC, has a decreasing optimal policy. For $(x, a) \in \mathbf{K}$, let

$$H_t(x, a, \gamma) := \sum_{\nu} Q(\nu \mid x, a) e^{\gamma \beta^t C(x,a,\nu)} u^*_{\beta,t+1}(F(x, a, \nu), \gamma). \tag{9.5}$$

Theorem 23. *For $t = 0, 1, \ldots n-1$, let*

$$A_t^*(x) = \{a \in A(x) : H_t(x, a, \gamma) = \min_{a' \in A(x)} H_t(x, a', \gamma)\};$$

and $f_t(x) := \max A_t^*(x)$. Suppose that

i) $x \longmapsto A(x)$ is a decreasing function, i.e., $x \leqslant x'$ implies $A(x') \subset A(x)$;

ii) for each $x \in \mathbf{X}$, the set $A(x)$ is such that $a \in A(x)$ and $a' \leq a$ imply $a' \in A(x)$; and

iii) For $t = 0, 1, \ldots n-1$, if $x \leqslant x'$, $H_t(x', \cdot, \gamma) - H_t(x, \cdot, \gamma)$ is increasing on $A(x')$. Then $(f_0, f_1, \ldots f_{n-1})$ is an optimal policy such that $f_t(x)$ is decreasing in x for each t.

Proof. The optimality of $(f_0, f_1, \ldots f_{n-1})$ follows from Theorem 19. Now, assume that $f_t(x') > f_t(x)$ for some $x' \geqslant x$. Then, it follows respectively from i), ii) and iii) that $f_t(x') \in A(x)$, $f_t(x) \in A(x')$ and

$$H_t(x', f_t(x'), \gamma) - H_t(x, f_t(x'), \gamma) \geqslant H_t(x', f_t(x), \gamma) - H_t(x, f_t(x), \gamma).$$

Consequently, we obtain

$$\begin{aligned} 0 &\geqslant H_t(x', f_t(x'), \gamma) - H_t(x', f_t(x), \gamma) \\ &\geqslant H_t(x, f_t(x'), \gamma) - H_t(x, f_t(x), \gamma) \\ &\geqslant 0. \end{aligned}$$

But $f_t(x') \in A(x)$, and $H_t(x, f_t(x'), \gamma) = H_t(x, f_t(x), \gamma)$ contradict the definition of f_t. □

Define $\bar{C} : \mathbf{X} \times \mathbf{D} \longrightarrow \mathbb{R}$ by

$$\bar{C}(y, \nu) = c \cdot y + h \cdot (y - \nu)^+ + p \cdot (\nu - y)^+ - \beta c \cdot (y - \nu).$$

Then, we obtain that $C(x, a, \nu) = \bar{C}(x + a, \nu)$, for all $(x, a) \in \mathbf{K}$. Bouakiz and Sobel [9] showed that if $p > c(1 - \beta)$ then, for each $\gamma' > 0$, and $t \in \{0, 1 \dots\}$, the function $V_{t,\gamma'} : \mathbf{X} \longrightarrow \mathbb{R}$ defined by

$$V_{t,\gamma'}(y) := \sum_{\nu} p_\nu e^{\gamma' \bar{C}(y,\nu)} J_{\beta,t}(y - \nu, \gamma'\beta) \tag{9.6}$$

is convex in y, where $J_{\beta,t}(x, \gamma)$ denotes the optimal EDC that can be obtained starting at state x, with risk-sensitivity coefficient γ, and proceeding for t stages, i.e.,

$$J_{\beta,t}(x, \gamma) = u^*_{\beta,N-t}(x, \gamma\beta^{N-t}). \tag{9.7}$$

This convexity property and the following lemma will be used in the proof of Proposition 12.

Lemma 21. *Let Y and Z be subsets of $\mathbb{R}$. Let G be a real-valued increasing function on $Y \times Z$, and H a real valued, convex function on the real line. Assume that*

$$G(y, z) + G(y', z') = G(y', z) + G(y, z'), \quad \forall y' \geq y, \ z' \geq z.$$

Then the composition of the functions H and G, $H(G)$, satisfies

$$H(G(y', z)) + H(G(y, z')) \leq H(G(y', z')) + H(G(y, z)), \quad \forall y' \geq y, \ z' \geq z.$$

Proof. Let $y' \geq y$, and $z' \geq z$. The proof of this result follows from the following identity:

$$\begin{aligned} H(G(y',z)) + H(G(y,z')) - H(G(y',z')) - H(G(y,z)) = \\ [H(G(y',z)) - H(G(y',z') + G(y,z) - G(y,z'))] \\ + [H(G(y',z') + G(y,z) - G(y,z')) - H(G(y',z'))] \\ - H(G(y,z)) + H(G(y,z')) \end{aligned}$$

□

Proposition 12. *If $p > c(1-\beta)$, then there exists a decreasing optimal policy $\pi = (f_0, f_1, \ldots f_{N-1})$ for the inventory control with an exponential discounted cost criterion. Furthermore, for $t = 0, 1, \ldots N-1$,*

$$f_t(x) = (f_t(0) - x)^+, \quad \forall x \in \mathbf{X},$$

where $f_t(0)$ is the base-stock level for the period t.

Proof. For $(x,a) \in \mathbf{K}$, let H_t be the function defined in (9.5), i.e.,

$$H_t(x,a,\gamma) = \sum_{\nu} p_\nu e^{\gamma\beta^t C(x,a,\nu)} u^*_{\beta,t+1}(x+a-\nu,\gamma).$$

Let $f_t(x) := \max A^*_t(x)$, where

$$A^*_t(x) = \{a \in A(x) : H_t(x,a,\gamma) = \min_{a' \in A(x)} \{H_t(x,a',\gamma)\}\}.$$

First, it follows from Theorem 19 that $(f_0, f_1, \ldots f_{N-1})$ is optimal. Next, we will apply Theorem 23 to prove that, for $t = 0, 1, \ldots N-1$, $f_t(x)$ is decreasing in x. Note that the set $\mathbf{K}$ for the inventory control problem satisfies the conditions (i) and (ii) of the mentioned theorem. On the other hand, it follows from (9.5), (9.6), and (9.7) that

$$\begin{aligned} H_t(x,a,\gamma) &= \sum_d p_d e^{\gamma\beta^t C(x,a,d)} u^*_{\beta,t+1}(x+a-d,\gamma) \\ &= \sum_d p_d e^{\gamma\beta^t \tilde{C}(x+a,d)} J_{\beta,N-t-1}(x+a-d,\gamma\beta^{t+1}) \qquad (9.8) \\ &= V_{N-t-1,\gamma\beta^t}(x+a). \end{aligned}$$

Thus, since $V_{N-t-1,\gamma\beta^t}(\cdot)$ is convex, and $\phi(x,a)+\phi(x',a')=\phi(x',a)+\phi(x,a')$, where $\phi(x,a)=x+a$, then by Lemma 21 we obtain that

$$H_t(x,a,\gamma\beta^t)+H_t(x',a',\gamma\beta^t)\geqslant H_t(x',a,\gamma\beta^t)+H_t(x,a',\gamma\beta^t).$$

Hence, if $x'\geqslant x$, then

$$H_t(x',\cdot,\gamma)-H_t(x,\cdot,\gamma)$$

is increasing on $A(x')$, i.e., H_t satisfies Condition (iii) of Theorem 23, and therefore, for $t=0,1,\ldots N-1$, $f_t(x)$ is decreasing. Now, it follows from (9.8) and the definition of f_t that

$$\begin{aligned} H_t(x+1,f_t(x)-1,\gamma) &= H_t(x,f_t(x),\gamma) \\ &\leqslant H_t(x,a,\gamma), \quad \forall a\leqslant M-x \\ &= H_t(x+1,a-1,\gamma), \quad \forall a\leqslant M-x. \end{aligned} \tag{9.9}$$

Therefore, $f_t(x+1)=f_t(x)-1$. Hence,

$$f_t(x)=\begin{cases} f_t(0)-x & \text{if} \quad x\leqslant f_t(0), \\ 0 & \text{if} \quad x>f_t(0). \end{cases} \tag{9.10}$$

The optimal policy obtained is a base-stock policy and $f_t(0)$ is the base-stock level for the period t. □

9.3 Infinite Horizon Inventory Control Model.

In this section, we will show that the infinite horizon inventory problem has an ultimately stationary base-stock optimal policy. The following theorem, which will be used to show that the infinite horizon inventory problem has a base-stock optimal policy, gives sufficient conditions under which the infinite horizon CMC, has a decreasing optimal policy. For $(x,a)\in\mathbf{K}$, let

$$G_t(x,a):=\sum_{\nu}Q(\nu\mid x,a)e^{\gamma\beta^t C(x,a,\nu)}J_\beta(F(x,a,\nu),\gamma\beta^{t+1}). \tag{9.11}$$

Theorem 24. *For $t = 0, 1, \ldots n-1$, let*

$$A_t^*(x) = \{a \in A(x) : G_t(x,a) = \min_{a' \in A(x)} G_t(x,a')\};$$

and $f_t(x) := \max A_t^*(x)$. Suppose that

i) $x \longmapsto A(x)$ is a decreasing function, i.e., $x \leqslant x'$ implies $A(x') \subset A(x)$;

ii) for each $x \in \mathbf{X}$, the set $A(x)$ is such that $a \in A(x)$ and $a' \leq a$ imply $a' \in A(x)$; and

iii) For $t = 0, 1, \ldots$, if $x \leqslant x'$, $G_t(x', \cdot) - G_t(x, \cdot)$ is increasing on $A(x')$.

Then $(f_0, f_1, \ldots)$ is an optimal policy such that $f_t(x)$ is decreasing in x for each t.

Proof. The optimality of $(f_0, f_1, \ldots)$ follows from Theorem 21. The rest of the proof is similar to that of Theorem 23. □

Let

$$\mathcal{D} := \sum_{t=0}^{\infty} \beta^t [c \cdot (a_t) + H(x_t, a_t, d_t)]$$

$$= \sum_{t=0}^{\infty} \beta^t [c \cdot (x_t + a_t) + H(x_t, a_t, d_t) - \beta c \cdot (x_t + a_t - d_t)] - \sum_{t=0}^{\infty} \beta^t c \cdot (x_t) + \sum_{t=0}^{\infty} \beta^{t+1} c \cdot (x_{t+1})$$

$$= \sum_{t=0}^{\infty} \beta^t [c \cdot (x_t + a_t) + H(x_t, a_t, d_t) - \beta c \cdot (x_t + a_t - d_t)] - c \cdot x_0.$$

Thus

$$\mathcal{D} = \sum_{t=0}^{\infty} \beta^t C(x_t, a_t, d_t) - c x_0.$$

For a deterministic policy $\pi = (f_0, f_1, \ldots)$, let

$$J_\beta^\pi(x, \gamma) = E_x^\pi [e^{\gamma \sum_{t=0}^{\infty} \beta^t C(X_t, A_t, D_t)}]$$

and

$$J_\beta(x, \gamma) = \inf_\pi \{J_\beta^\pi(x, \gamma)\} \tag{9.12}$$

Note that the policy which optimizes (9.12) also optimizes the problem with discounted cost $\mathcal{D}$.

The discounted optimality equations for this model are given by

$$J_\beta(x,\gamma\beta^t) = \min_{a\in A(x)}\{\sum_\nu p_\nu e^{\gamma\beta^t C(x,a,\nu)} J_\beta(x+a-\nu,\gamma\beta^{t+1})\}.$$

Proposition 13. *If $p > c(1-\beta)$ then there exists a base-stock and ultimately stationary optimal policy for the infinite horizon inventory control with exponential discounted cost criterion.*

Proof. It follows from Theorem 22 that

$$\lim_{n\to\infty} V_{n,\gamma'}(y) = \sum_d p_d e^{\gamma'\bar{C}(y,d)} J_\beta(y-d,\gamma'\beta)] =: V_{\gamma'}(y).$$

Thus, since the limit of convex functions is convex, we obtain that $V_{\gamma'}(y)$ is convex in y for each $\gamma' > 0$. On the other hand, we have that

$$\begin{aligned} G_t(x,a) &= \sum_\nu p_\nu e^{\gamma\beta^t C(x,a,\nu)} J_\beta(x+a-\nu,\gamma\beta^{t+1}) \\ &= V_{\gamma\beta^t}(x+a). \end{aligned}$$

Hence, by Lemma 21 we obtain that

$$G_t(x,a) + G_t(x',a') \geqslant G_t(x',a) + G_t(x,a'), \quad x' \geqslant x, \quad a' \geqslant a,$$

or equivalently,

$$\text{if } x' \geqslant x, \quad \text{then } G_t(x',\cdot) - G_t(x,\cdot) \quad \text{is increasing.}$$

Thus, analogously to the proof of Proposition 12, the existence of a base-stock optimal policy follows from Theorem 23. Moreover, since the state space is in fact finite, i.e., $\mathbf{X} = \{-L,\ldots,-1,0,\ldots M\}$, there exists N such that $f_t = f$ for $t \geqslant N$; see [32, 33]. Therefore there is an infinite horizon optimal policy which is base-stock and ultimately stationary. □

Appendix A

Risk-Neutral Optimal Resource Allocation

For the purpose of comparing the results obtained in Section 6.3 about structured optimal policies for an optimal allocation problem (with ETC criterion), in this appendix we study the corresponding risk-neutral allocation problem. Section 1 summarizes some results about monotonicity properties of the optimal value function and policies. For the proofs of those results we refer the reader to Ross [41]. In Sections 2 and 3 we derive further structural properties of the optimal policies under the assumption that the probability function P(a) is convex.

A.1 Monotone Optimal Policies

For $t = 0, 1, \ldots, N-1$, denote

$$F_t(x,a) := a + P(a)J_t(x-1) + (1-P(a))J_t(x), \quad x \geq 1, \tag{A.1}$$

where $J_t(x)$ is the risk-neutral optimal total cost when t stages remain to go and the state at time $N-t$ is x. Note that $F_t(x,a)$ is the function within brackets in the (risk-neutral) dynamic programming algorithm

$$J_0(x) = 0 \tag{A.2}$$

$$\vdots \qquad\qquad \vdots$$

$$J_{t+1}(x) = \min_{a \in A(x)} \{\mathbf{C}(x,a) + \sum_y p_{xy}(a)J_t(y)\}. \tag{A.3}$$

Let

$$\bar{A}_t(x) := \{a : F_t(x,a) = \inf_{a'}\{F_t(x,a')\}\}$$

and

$$\bar{f}_t(x) := \min \bar{A}_t(x).$$

For a proof of the following two results we refer to [41].

Lemma 22. *The optimal value function $J_t(x)$ is increasing in x and decreasing in t.*

Proposition 14. *Assume that $C(x)$ is convex. Then $\bar{\pi} = (\bar{f}_0, \ldots, \bar{f}_{N-1})$ is an optimal policy for the risk-neutral allocation problem such that for $t = 0, \ldots N-1$, $f_t(x)$ is increasing in x; and for fixed x, $f_t(x)$ is increasing in t.*

A.2 Risk-neutral Allocation Problem with P(a) Convex

Throughout this appendix, the policy $\bar{\pi} = (\bar{f}_0, \ldots, \bar{f}_{N-1})$ will denote the monotone optimal policy obtained in Proposition 14. In the following proposition we will show that when the probability function $P(a)$ is convex, the allocation model is reduced to a problem with two actions: the extreme points of the interval $[0, M]$. Consequently, there exists an optimal threshold policy.

Proposition 15. *Assume that $C(x)$ is convex and $P(a)$ is convex and twice differentiable. Then the allocation optimal control problem (with total cost criterion) can be reduced to a problem with two actions: the extreme points of the interval $[0, M]$. Moreover, the optimal policy $\bar{\pi} = (\bar{f}_0, \bar{f}_1, \ldots \bar{f}_{N-1})$ is of the threshold-type, that is, there exist states $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_{N-1}$ such that*

$$\bar{f}_t(x) = \begin{cases} 0 & \text{if } x < \bar{x}_t \\ M & \text{if } x \geqslant \bar{x}_t, \end{cases} \tag{A.4}$$

$t = 0, 1, \ldots N-1$. Moreover, the sequence of thresholds is decreasing.

Proof. It follows from (A.1) that

$$F_t(x,a) = a + [J_t(x) - J_t(x-1)](1 - P(a)) + J_t(x-1),$$

and hence

$$\frac{\partial F_t}{\partial a}(x,a) = 1 - P'(a)[J_t(x) - J_t(x-1)]$$

and

$$\frac{\partial^2 F_t}{\partial^2 a}(x,a) = -P''(a)[J_t(x) - J_t(x-1)].$$

Thus, since $P''(a) \geq 0$ and $J_t(x)$ is increasing in x (by Lemma 22) we obtain that $\frac{\partial^2 F_t}{\partial^2 a}(x,a) < 0$, and therefore $F_t(x,a)$ is concave in a. Consequently,

$$\bar{A}_t(x) = \{0, M\},$$

and hence, $\bar{f}_t(x) \in \{0, M\}$. Moreover, if we define

$$\bar{x}_t := \min\{x : \bar{f}_t(x) = M\},$$

then (A.4) follows from the fact that $\bar{f}_t(x)$ is increasing in x. Finally, the sequence $\{\bar{x}_t\}$ is decreasing since $\bar{f}_t(x)$ is increasing in t. □

A.2.1 Risk-neutral Allocation Problem with Linear Terminal Cost

Now, we will apply Proposition 15 to compute the optimal policy for the Example 1 considered in Section 7.

Example 1 (revisited) First we compute $\bar{f}_{N-1}(x)$. To do that, by Proposition 15, we need only to compare the values of the function $F_0(x,a)$ at the extreme actions $a = 0$ and $a = 1$. It follows from (A.1) that

$$\begin{aligned} F_0(x,a) &= a + P(a)J_0(x-1) + (1 - P(a))J_0(x), \quad x \geq 1 \\ &= a + P(a)(2x-2) + (1 - P(a))2x, \quad x \geq 1. \end{aligned}$$

Thus,

$$F_0(x,0) = 2x, \quad x \geq 1, \quad \text{and}$$

$$F_0(x,1) = 2x + (1 - 2P(1)), \quad x \geq 1.$$

Thus, we obtain

a) if $P(1) > \frac{1}{2}$ then

$$F_0(x,1) < F_0(x,0), \quad x \geq 1$$

b) if $P(1) < \frac{1}{2}$ then

$$F_0(x,0) < F_0(x,1), \quad x \geq 1, \quad \text{and}$$

c) if $P(1) = \frac{1}{2}$ then

$$F_0(x,1) = F_0(x,0), \quad x \geq 1.$$

Therefore the optimal decision rule $\bar{f}_{N-1}$ and the optimal value function J_1 for the case (a) are given by

$$\bar{f}_{N-1}(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geqslant 1, \end{cases} \tag{A.5}$$

and

$$J_1(x) = \begin{cases} 0 & \text{if } x = 0 \\ 2x + (1 - 2P(1)) & \text{if } x \geq 1; \end{cases} \tag{A.6}$$

for (b) by

$$\bar{f}_{N-1}(x) = 0, \quad \forall x,$$

and

$$J_1(x) = 2x, \quad x \geq 0; \tag{A.7}$$

and for (c) we obtain that both actions $a = 0$ and $a = 1$ are optimal.

Now, to compute the optimal decision rules $\bar{f}_t$, $t = 0, \ldots, N-2$, we will first prove each one of the following statements by induction on t :

I) If $P(1) > \frac{1}{2}$ then for $t = 1, \ldots N-1$,

$$J_t(1) = 1 + (1 - P(1))J_{t-1}(1);$$

II) If $P(1) > \frac{1}{2}$ then for $t = 1, \ldots N-1$,

$$J_t(x) = J_0(x), \quad x \in \mathbf{X}.$$

First, let's prove (I). The validity of assertion (I) for $t = 1$ follows from (A.6). Next, by the dynamic programming algorithm

$$J_{t+1}(1) = \min\{F_t(1,0),\ F_t(1,1)\}.$$

Thus,

$$\begin{aligned} J_{t+1}(1) &= \min\{J_t(1), 1 + (1 - P(1))J_t(1)\} \\ &= \min\{1 + (1 - P(1))J_{t-1}(1), 1 + (1 - P(1))J_t(1)\} \qquad \text{(A.8)} \\ &= 1 + (1 - P(1))J_t(1), \qquad \text{(A.9)} \end{aligned}$$

where (A.8) and (A.9) follow from the induction hypothesis and Lemma 22 respectively. Thus the proof of (I) is complete.

Now, let's prove (II). First, (A.7) implies that (II) holds for $t = 1$. Next, similarly as above

$$\begin{aligned} J_{t+1}(I) &= \min\{F_t(I,0),\ F_t(I,1)\} \\ &= \min\{J_t(I), 1 + P(1)J_t(I-1) + (1 - P(1))J_t(I)\} \qquad \text{(A.10)} \\ &= \min\{2I, 1 + P(1)2(I-1) + (1 - P(1))2I\} \qquad \text{(A.11)} \\ &= \min\{2I, 2I + (1 - 2P(1))\} \\ &= 2I \qquad \text{(A.12)} \end{aligned}$$

where (A.10), (A.11) and (A.12) follow from (A.1), the induction hypothesis and the hypothesis $P(1) < \frac{1}{2}$ respectively. Thus $\bar{f}_{N-t-1}(I) = 0$ and since $\bar{f}_{N-t-1}(x)$ is increasing in x we obtain that $\bar{f}_{N-t-1}(x) = 0$, for all x. Therefore

$$\begin{aligned} J_{t+1}(x) &= \min\{F_t(x,0),\ F_t(x,1)\} \\ &= F_t(x,0) \\ &= J_t(x) \\ &= J_0(x), \quad \forall x \in \mathbf{X}, \end{aligned}$$

and the proof of (II) is complete.

Finally, it follows from (I), (II) and (c) that $\bar{f}_t(x)$, $t = 0, 1, \ldots, N-1$, are given by

$$\bar{f}_t(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geqslant 1 \end{cases}$$

if $P(1) > \frac{1}{2}$;

$$\bar{f}_t(x) = 0, \quad \forall x$$

if $P(1) < \frac{1}{2}$; and if $P(1) = \frac{1}{2}$ then there are $I+1$ threshold optimal policies:

$$f_t^y(x) = \begin{cases} 0 & \text{if } x \leq y \\ 1 & \text{if } x > y, \end{cases}$$

$y = 0, 1, \ldots I$.

REFERENCES

[1] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J. Control and Optimization*, 31(2):282–344, March 1993.

[2] G. Avila-Godoy, A. Brau, and E. Fernández-Gaucherand. Controlled Markov chains with discounted risk-sensitive criteria: applications to machine replacement. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 1115–1120, San Diego, CA, 1997.

[3] G. Avila-Godoy and E. Fernández-Gaucherand. Controlled Markov chains with exponential risk-sensitive criteria: Modularity, structured policies and applications. (Preprint), 1998.

[4] G. Avila-Godoy and E. Fernández-Gaucherand. Exponential risk-sensitive optimal scheduling. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 3958–3963, San Diego, CA, 97.

[5] D. P. Bertsekas. *Dynamic Programming and Stochastic Control.* Academic Press, New York, 1976.

[6] D. P. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models.* Prentice Hall, Englewood Cliffs, N.J., 1987.

[7] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete Time Case.* Academic Press, New York, 1978.

[8] P. Billingsley. *Probability and Measure.* John Wiley & Sons, New York, 1979.

[9] M. Bouakiz and M. J. Sobel. Inventory control with an exponential utility criterion. *Mathematics of Operations Research*, 40(3):603–608, May–June 1992.

[10] A. Brau and E. Fernández-Gaucherand. Controlled Markov chains with risk-sensitive exponential average cost criterion. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 2260–2264, San Diego, CA, 1997.

[11] A. Brau and E. Fernández-Gaucherand. Controlled Markov chains with risk-sensitive criteria: Some (counter) examples. (Preprint), 1998.

[12] L. Breiman. *Probability.* Addison-Wesley Publishing Company, California, 1968.

[13] R. Cavazos-Cadena and E. Fernández-Gaucherand. Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions. To appear, Mathematical Methods in Operations Research, March 1998.

[14] R. Cavazos-Cadena and E. Fernández-Gaucherand. Controlled Markov chains with risk-sensitive average cost criterion: A counter-example and necessary conditions for optimal solutions under strong recurrence assumptions. Submitted for pubication, 1998.

[15] R. Cavazos-Cadena and E. Fernández-Gaucherand. Markov decision processes with risk-sensitive average cost criterion: The discounted stochastic games approach. Submitted for publication, 1998.

[16] R. Cavazos-Cadena and E. Fernández-Gaucherand. Risk-sensitive optimal control in communicating average Markov decision chains. Submitted for pubication, 1998.

[17] R. Cavazos-Cadena and E. Fernández-Gaucherand. The vanishing discount approach in Markov chains with risk-sensitive criteria. Submitted for publication, 1998.

[18] K.-J. Chung and M. J. Sobel. Discounted MDP's: distribution functions and exponential utility maximization. *SIAM J. Control and Optimization*, 25(1):49–61, January 1987.

[19] M. DeGroot. *Optimal Statistical decisions.* McGraw-Hill, New York, 1970.

[20] E. V. Denardo and U. G. Rothblum. Optimal stopping, exponential utility, and linear programming. *Mathematical Programming*, 16:228–244, 1979.

[21] E. Fernández-Gaucherand. Utility theory. Unpublished manuscript, 1994.

[22] E. Fernández-Gaucherand. Non-standard optimality criteria for controlled Markov processes. *Zeitschrift für Angewandte Mathematik und Mechanik*, pages 423–424, 1996. Special issue on Applied Stochastic and Optimization.

[23] E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus. Controlled Markov processes on the infinite planning horizon: Weighted and overtaking cost criteria. *Methods and Models of Operations Research*, 39:131–155, 1994.

[24] E. Fernández-Gaucherand and S. I. Marcus. Risk-sensitive optimal control of hidden Markov models: a case study. In *Proceedings of the 33rd IEEE Conference on Decision and Control*, pages 1657–1662, 1994.

[25] E. Fernández-Gaucherand and S. I. Marcus. Non-standard optimality criteria for stochastic control problems. In *Proceedings of the 34th IEEE Conference on Decision and Control*, pages 585–589, New Orleans, LA, 1995.

[26] E. Fernández-Gaucherand and S. I. Marcus. Risk-sensitive optimal control of hidden Markov models: structural results. *IEEE Transactions on Automatic Control*, 42:1418–1442, 1997.

[27] P. C. Fishburn. *Utility Theory for Decision Making.* John Wiley & Sons, Inc., New York, 1970.

[28] W. H. Fleming and D. Hernández-Hernández. Risk sensitive control of finite state machines on an infinite horizon I. *SIAM Journal on Control and Optimization*, 35(5):1970–1810, September 1997.

[29] O. Hernández-Lerma and J. Lasserre. *Discrete Time Markov Control Processes.* Springer, New York, 1996.

[30] K. F. Hinderer. On the structure of solutions of stochastic dynamic programs. In *Proceedings of the Seventh Conference on Probability Theory*, pages 173–182, Brasov, Romania, 1982.

[31] R. A. Howard and J. E. Matheson. Risk-sensitive Markov decision processes. *Management Science*, 18(7):356–369, March 1972.

[32] S. C. Jaquette. Markov decision processes with a new optimality criterion: discrete time. *The Annals of Statistics*, 1:496–505, September 1973.

[33] S. C. Jaquette. A utility criterion for Markov decision processes. *Management Science*, 23(1):43–49, September 1976.

[34] R. D. Luce and H. Raiffa. *Games and Decisions: Introduction and Critical Survey.* Dover Publications, Inc., New York, 1989.

[35] S. I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernández, S. Coraluppi, and P. Fard. Risk sensitive Markov decision processes. In C. Byrnes, B. Data, D. Gilliam, and C. Martin, editors, *Systems and Control in the Twenty-First Century*, Progress in Systems and Control, pages 263–279. Birkhauser, 1997.

[36] R. A. Milito. Single machine scheduling with non-linear costs: Necessary conditions for optimality in the transform domain. In *Proceedings 34th CDC*, pages 3650–3651, New Orleans, LA, 1995.

[37] J. Neveu. *Mathematical Foundations of the Calculus of Probability.* Holden-Day,Inc., San Francisco, Cal., 1965.

[38] M. Pinedo. *Scheduling: Theory, Algorithms, and Systems.* Prentice Hall, Englewood Cliffs, N.J., 1995.

[39] J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1):122–136, January-April 1964.

[40] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, 1994.

[41] S. M. Ross. *Introduction to Stochastic Dynamic Programming*. Academic Press, New York, 1983.

[42] R. F. Serfozo. Monotone optimal policies for Markov decision processes. *Mathematical Programming Study*, 6:202–215, 1976.

[43] R. F. Serfozo. Optimal control of random walks, birth and death processes, and queues. *Advances in Applied Probability*, 13:61–83, 1981.

[44] J. Shaler Stidham and R. R. Weber. Monotonic and insensitive optimal policies for control of queues with undiscounted costs. *Operations Research*, 87(4):611–625, August 1989.

[45] M. W. Sobel. An overview of nonstandard criteria in Markov decision processes. In *Proceedings Airlie House NSF Conference, Airlie, Virginia*, Washington, 1988. NSF.

[46] R. R. Weber and J. Shaler Stidham. Optimal control of service rates in networks of queues. *Advances in Applied Probability*, 19:202–218, 1987.

[47] P. Whittle. *Risk-sensitive Optimal Control*. John Wiley & Sons, Chichester, 1990.