Morfismos, Vol. 7, No. 2, 2003, pp. 47–73

Application of modularity to optimal resource allocation with risk sensitivity

Guadalupe Avila-Godoy [1]

Abstract

An optimal allocation problem with a risk-sensitive controller is modelled by a controlled Markov chain with exponential total cost criterion. Some general results recently obtained are applied to show that the particular model studied here has a monotone optimal policy and monotone optimal value function. Moreover, it is shown that under certain conditions, the allocation problem with both risk-neutral and risk-sensitive performance criteria has an optimal policy of the threshold type.

2000 Mathematics Subject Classification: 90C40, 60J05.
Keywords and phrases: Controlled Markov Chains, Optimal Resource Allocation, Exponential Total Cost Criterion, Risk Sensitivity, Modularity.

1 Introduction

In this paper we study a finite horizon controlled Markov chain (CMC) modeling an optimal allocation problem with exponential total cost (ETC) as performance risk-sensitive criterion. The CMC considered has finite state space, compact action space and bounded cost function. Models of dynamic systems that incorporate risk-sensitivity by means of an exponential utility function have recently received considerable attention in the literature, see for example [2, 3, 6, 7, 8, 9] and references therein. However, in contrast with the risk-neutral literature (see

[1]This paper is part of the author's doctoral research under the direction of Dr. Emmanuel Fernández Gaucherand at the Department of Mathematics of the University of Arizona.

[11, 13, 14, 15, 16, 17, 18, 19] and references therein), only a few (and recent) contributions dealing with structural properties of the CMC have been made in the risk-sensitive case; see [5, 9]. We are particularly concerned with the contributions made by Avila-Godoy [4], which extend some results in [13, 14] from the risk-neutral to the risk-sensitive case. It is proved in [4] that appropiate structural conditions (like modularity and/or monotonicity) of certain functions defined in terms of the cost and the transition kernel, imply the monotonicity of the optimal exponential value function and the existence of monotone optimal policies. Herein, we apply some results in [4] to show that the CMC modeling the optimal allocation problem has monotone optimal policies and monotone optimal value function. See, e.g., [13, 14] for an analysis of this problem with risk-neutral total cost.

The paper is organized as follows. Section 2 includes the description of the model and collects the results in [4] needed for our study. In Sections 3 and 4, the CMC model for the finite horizon optimal allocation problem is given and the main results of the paper are proved. First, it is shown that the optimal value function for this model, $J_t(x)$, is increasing in the state x and decreasing in t (Lemma 3.1.6), and then the existence of a monotone optimal policy is established (Proposition 3.1.12), that is, we show that the decision function of the optimal policy at the t-th stage is increasing (as a function of the state), for $t = 0, 1, \ldots$, and increasing in t, for each $x \in \mathbf{X}$. Moreover, under additional conditions, we prove that the allocation problem can be reduced to a problem with two actions and that the optimal policy is of the threshold type (Proposition 4.1.13). Finally, we apply those results to a particular example with a linear final cost. For the purpose of comparing the results obtained in Sections 3 and 4, we include an appendix on risk-neutral resource allocation problems. The proof of the result relative to the reduction of the risk-neutral allocation problem to a problem with two actions and that the optimal policy is of the threshold type is also a contribution of this paper's author (Section 5.2).

2 Description of the model and basic results

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

- $\mathbf{X} = \{1, 2, \ldots\}$ is the state space, a countable set.
- $\mathbf{A}$, the action (or control) set, is a compact subset of $\mathbb{R}$. The set

$\mathbf{K} := \big\{(x, a) : x \in \mathbf{X},\ a \in \mathbf{A}\big\}$ is called the set of state-action pairs.

• $\mathbf{P}$, the transition kernel, is a family of transition probabilities on $\mathbf{X}$ given $\mathbf{K}$:

$$\mathbf{P} = \{P(\cdot \mid x, a) : (x, a) \in \mathbf{K}\}.$$

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

• $\mathbf{C} : \mathbf{K} \longrightarrow \mathbb{R}$ is the one-stage cost function. We will assume that $\mathbf{C}$ is nonnegative and bounded: $0 \leqslant \mathbf{C}(x, a) \leqslant K < \infty$ for every $(x, a) \in \mathbf{K}$, and $c : \mathbf{X} \longrightarrow \mathbb{R}$ is the final penalty cost.

The above defined CMC represents a stochastic dynamical system observed at times $t = 0, 1, \cdots, n$, whose evolution is as follows. Let X_t and A_t respectively denote the state of the system and the action chosen at time t. If $X_0 = x \in \mathbf{X}$, and $A_0 = a \in \mathbf{A}$, then (i) a cost $\mathbf{C}(x, a)$ is incurred, and (ii) the system moves to a new state X_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 n times; see [1, 10, 13].

The strategy followed to choose the actions at each stage is called a policy. The most general set $\mathbf{\Pi}$ of policies considered in the literature includes the admissible, history dependent, randomized policies; see [1, 10, 13]. Herein, we will be concerned only with the subset of $\mathbf{\Pi}$ consisting of the Markov deterministic policies, denoted by $\mathbf{\Pi}_{MD}$. For a policy $\boldsymbol{\pi} \in \mathbf{\Pi}$ and initial state $x \in \mathbf{X}$, E_x^π will denote the expectation operator with respect to the probability measure induced by π and x in the space of trayectories of the chain.

Risk-sensitivity of the controller is modelled by grading the total cost with the exponential (disutility) function $\mathcal{U}_\gamma(x) = (\operatorname{sgn} \gamma)e^{\gamma x}$, $\gamma \neq 0$, where the parameter γ turns out to be the (constant) risk-sensitivity coefficient associated to $\mathcal{U}_\gamma$, see [12, 20]. In this work, only the case $\gamma > 0$, the risk-averse case, will be considered. Thus, the performance criterion for a policy π when the initial state is x and we proceed for n stages, is given by

$$J_n^\pi(x, \gamma) := E_x^\pi \left[e^{\gamma(\sum_{t=0}^{n-1} \mathbf{C}(X_t, A_t) + c(X_n))} \right]. \tag{1}$$

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

$$J_n^{\pi^*}(x, \gamma) = \inf_\pi \{J_n^\pi(x, \gamma)\} =: J_n(x, \gamma). \tag{2}$$

The optimal policy π^* is called ETC-optimal and $J_n(x,\gamma)$ is the optimal ETC. We can interpret $J_n(x,\gamma)$ as the minimal ETC that can be obtained starting at state x with risk-sensitivity coefficient γ and proceeding for n stages.

For ease of reference, we end the section by stating (without proofs) some general results that will be needed in the next section; see [4] for their proofs. The following known assumption will be made in the rest of this section (see [10]).

Assumption 2.1

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

First, we recall a typical forward dynamic programming recursion.

Theorem 2.1.1 (Dynamic Programming Algorithm) *The optimal ETC, $J_n(x,\gamma)$, satisfies the following recursion:*

$$\begin{aligned} J_0(x,\gamma) &= e^{\gamma c(x)}, \\ &\vdots \\ J_{s+1}(x,\gamma) &= \inf_{a\in\mathbf{A}} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_s(y,\gamma) \right\}, \end{aligned} \tag{3}$$

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

$$\begin{aligned} & e^{\gamma \mathbf{C}\left(x, f_s(x)\right)} \sum_y p_{xy}\left(f_s(x)\right) J_{n-s-1}(y,\gamma) = \\ & \qquad \inf_{a\in\mathbf{A}} \left\{ e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_{n-s-1}(y,\gamma) \right\}. \end{aligned} \tag{4}$$

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

Next, a lemma that provides sufficient conditions for monotonicity of the optimal value function is stated.

Lemma 2.1.2 *Suppose that*

i) $\mathbf{C}(x,a)$ *is increasing (decreasing) in* x *for each* a*, and* $c(x)$ *is increasing (decreasing).*

ii) $\sum_{y=z}^{\infty} p_{xy}(a)$ *is increasing in* x *for all* $z \in \mathbf{X}$ *and* $a \in \mathbf{A}$.
Then, the optimal value function $J_s(x,\gamma)$ *is increasing (decreasing) in* x*, for* $s = 0, 1, \cdots n$.

Finally, some standard definitions and notation, and two key theorems about structural properties of CMC's are stated (see [4].)

Let $(S, \preccurlyeq_S)$ be a lattice, i.e., a partially ordered 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 *superadditive (or supermodular)* on S if $-G(\cdot)$ is subadditive on S.

We will assume the state and action spaces to be subsets of $\mathbb{R}$ with the usual order and 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 $z \leqslant z'$.

A Markov deterministic policy $\pi = (f_0, f_1, \ldots, f_{n-1})$ is said to be monotone (with respect to x) if all the decision rules f_t are monotone functions of the state x. In the particular case that the action space contains only two actions, say a_1 and a_2, a monotone policy is called a threshold policy. That is, a threshold policy is a deterministic Markov policy $\pi = (f_0, f_1, \ldots, f_{n-1})$ such that, for $t = 0, 1, \ldots, n-1$, the decision rule f_t is given by

$$f_t(x) = \begin{cases} a_1 & \text{if } x \geqslant x_t^* \\ a_2 & \text{if } x < x_t^*, \end{cases} \tag{5}$$

where x_t^* is the control limit or threshold.

It is clearly useful to know in advance when a monotone optimal policy exists, because the search for an optimal policy can then be restricted from the class of Markov deterministic policies to the much smaller subclass of monotone policies [11, 13]. The following theorem provides sufficient conditions for the existence of optimal monotone (with respect to x) policies for CMC's with ETC criterion.

Notation. For $t = 0, 1, 2, \dots, n-1$, let

$$H_t(x, a, \gamma) := e^{\gamma \mathbf{C}(x,a)} \sum_y p_{xy}(a) J_{n-t-1}(y, \gamma), \tag{6}$$

i.e., H_t denotes the function within brackets in (4).

Theorem 2.1.3 *For $t = 0, 1, 2, \cdots n-1$, set*

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

and $f_t(x) := \min A_t^*(x)$ (respectively $f_t(x) := \max A_t^*(x)$). Suppose that $\log H_t(\cdot, \cdot, \gamma)$ is subadditive (respectively superadditive) on $(\mathbf{X} \times \mathbf{A}, \preccurlyeq)$, for fixed γ.

Then, $(f_0, f_1, \dots f_{n-1})$ is an optimal policy with $f_t(x)$ increasing (respectively decreasing) in x for each t.

Additionally, due to the fact that the optimal policy $\pi = (f_0, f_1, \dots f_{n-1})$ is in general non-stationary, it is natural to ask how the optimal action $f_t(x)$ varies with respect to t for each fixed x. Thus, a Markov deterministic policy $\pi = (f_0, f_1, \dots, f_{n-1})$ is said to be monotone (with respect to t) if for each fixed x, the sequence of actions $f_t(x)$ is monotone in t. The following theorem provides sufficient conditions for the existence of optimal monotone (with respect to t) policies for CMC's with ETC criterion.

Theorem 2.1.4 *Let $A_t^*(x)$ and $f_t(x)$ be as in Theorem 2.1.3. Assume that for each $x \in \mathbf{X}$, the function $\log H_{(\cdot)}(x, \cdot, \gamma)$ is superadditive (respectively subadditive) on the lattice $(\mathbf{A} \times \{0, 1, 2, \cdots n-1\}, \preccurlyeq)$. Then, $(f_0, f_1, \dots f_{n-1})$ is an optimal policy such that the sequence of actions $f_t(x)$ is decreasing (respectively increasing) in t for each x.*

3 An Optimal Allocation Problem.

In this section we follow Ross [14] to model an optimal allocation problem by means of a finite horizon CMC. However, unlike the mentioned reference, we introduce a risk-sensitive performance criterion. The general results stated in Section 2 are applied to show that, under standard conditions, the optimal value function is increasing in the state and decreasing in t (Lemma 3.1.6) and that the optimal policy is increasing in x and increasing in t (Proposition 3.1.12). Moreover, under additional conditions, we prove that the allocation problem can be reduced to a problem with two actions and that the optimal policy is of the threshold-type (Proposition 4.1.13). Finally, we apply those results to a particular example of a linear final cost; see [14] for an analysis of this problem with risk-neutral total cost.

The optimal allocation problem can be described as follows. Suppose we have N stages to construct sequentially I successful components . 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 strictly increasing 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 expected ETC. A CMC $(\mathbf{X}, \mathbf{A}, P, \mathbf{C})$ which models the described allocation problem can be defined by taking the state space $\mathbf{X} = \{0, 1, \dots I\}$, 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

$$p_{xy}(a) = \begin{cases} P(a) & \text{if } y \;=\; x-1 \\ 1 - P(a) & \text{if } y \;= x \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

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.

We recall that $J_t(x, \gamma)$ denotes the minimal cost starting at state x with t stages to go, $x \in \mathbf{X}$ and $t = 0, 1, \dots, N$.

Remark 3.1.5 This model satisfies Assumption 2.1 since $\mathbf{C}(x, a)$ and

$$\sum_y p_{xy}(a) J_t(y, \gamma) = P(a) J_t(x-1, \gamma) + (1 - P(a)) J_t(x, \gamma)$$

are continuous functions in a, for each $x \in \mathbf{X}$.

According to (3), $J_0(x,\gamma) = e^{\gamma c(x)}$ and for $t = 0, 1, \ldots N-1$,

$$
\begin{aligned}
& J_{t+1}(x,\gamma) \\
(8) \quad &= \inf_{a \in [0,M]} \left\{ e^{\gamma a}[P(a)J_t(x-1,\gamma) + (1-P(a))J_t(x,\gamma)] \right\} \\
(9) \quad &= \inf_{a \in [0,M]} \{ e^{\gamma a}[J_t(x,\gamma) - P(a)(J_t(x,\gamma) - J_t(x-1,\gamma)] \} \\
(10) \quad & \\
&= \inf_{a \in [0,M]} \{ e^{\gamma a}[J_t(x-1,\gamma) + (1-P(a))(J_t(x,\gamma) - J_t(x-1,\gamma)] \} .
\end{aligned}
$$

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

Lemma 3.1.6 *The optimal value function $J_t(x,\gamma)$ is increasing in x and decreasing in t.*

Proof: We will apply Lemma 2.1.2 to prove that $J_t(x,\gamma)$ is increasing in x. First, we see that this model satisfies (i) of the mentioned lemma since $\mathbf{C}(x,a)$ is constant in x, and the terminal cost $c(x)$ is increasing. Finally, it follows from (7) that

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

and hence, (ii) of Lemma 2.1.2 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 (8) that

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

and hence,

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

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

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

$$
(12) \qquad \log \{ e^{\gamma a}[P(a)J_{N-t-1}(x-1,\gamma) + (1-P(a))J_{N-t-1}(x,\gamma)] \}
$$

is subadditive on $\mathbf{X} \times \mathbf{A}$ and subadditive on $\mathbf{A} \times \{0, 1, 2, \dots N-1\}$, so that the mentioned monotonicity properties will follow from Theorems 2.1.3 and 2.1.4 since in this model the function H_t defined in (6) is the function within brackets in (12), i.e.,

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

Set $g_t(x, a, \gamma) = P(a)J_t(x-1, \gamma) + (1 - P(a))J_t(x, \gamma)$ and

$$G_t(x, a, \gamma) := e^{\gamma a} g_t(x, a, \gamma), \tag{13}$$

so that

$$H_t(x, a, \gamma) = G_{N-t-1}(x, a, \gamma). \tag{14}$$

First, it follows from (14) that each of the structural properties of $\log H_t(x, a, \gamma)$ we need is equivalent to a structural property of $\log G_t(x, a, \gamma)$.

Lemma 3.1.7 *a)* $\log H_t(x, a, \gamma)$ *is subadditive on* $\mathbf{X} \times \mathbf{A}$ *iff* $\log G_t(x, a, \gamma)$ *is subadditive on* $\mathbf{X} \times \mathbf{A}$. *b)* $\log H_t(x, a, \gamma)$ *is subadditive on* $\mathbf{A} \times \{0, 1, \dots N-1\}$ *iff* $\log G_t(x, a, \gamma)$ *is superadditive on* $\mathbf{A} \times \{0, 1, \dots N-1\}$.

Next, we will see that each of the structural properties of $\log G_t(x, a, \gamma)$ we need is equivalent to a structural property of $\log J_t(x, \gamma)$.

Lemma 3.1.8 *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* $\mathbf{A} \times \{0, 1, \dots N-1\}$ *iff* $\log J_t(x, \gamma)$ *is subadditive on* $\mathbf{X} \times \{0, 1, \dots N-1\}$.

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

Then

$$
\begin{aligned}
&\log J_t(x,\gamma) \text{ is convex in } x\\
&\iff \log J_t(x+1,\gamma) - \log J_t(x,\gamma) \geqslant \log J_t(x,\gamma) - \log J_t(x-1,\gamma)\\
&\iff J_t(x+1,\gamma)J_t(x-1,\gamma) \geqslant J_t^2(x,\gamma)\\
&\iff J_t(x,\gamma)D_t(x) \geqslant J_t(x+1,\gamma)D_t(x-1)\\
&\iff (P(a')-P(a))J_t(x,\gamma)D_t(x) \geqslant (P(a')-P(a))J_t(x+1,\gamma)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad D_t(x-1)\\
&\iff -P(a)J_t(x,\gamma)D_t(x) - P(a')J_t(x+1,\gamma)D_t(x-1) \geqslant\\
&\qquad\qquad - P(a')J_t(x,\gamma)D_t(x) - P(a)J_t(x+1,\gamma)D_t(x-1)\\
&\iff [J_t(x+1,\gamma) - P(a)D_t(x)][J_t(x,\gamma) - P(a')D_t(x-1)] \geqslant\\
&\qquad\qquad [J_t(x+1,\gamma) - P(a')D_t(x)][J_t(x,\gamma) - P(a)D_t(x-1)]\\
&\iff \frac{g_t(x+1,a)}{g_t(x,a)} \geqslant \frac{g_t(x+1,a')}{g_t(x,a')}\\
&\iff \log g_t(x,a,\gamma) \text{ is subadditive on } \mathbf{X}\times\mathbf{A}\\
&\iff \log G_t(x,a,\gamma) \text{ is subadditive on } \mathbf{X}\times\mathbf{A}.
\end{aligned}
$$

Note that the last step follows from the equality

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

b) Let $a' > a$. Then

$$
\begin{aligned}
&\log J_t(x,\gamma) \text{ is subadditive on } \mathbf{X}\times\{0,1,\dots N-1\}\\
&\iff \log J_{t+1}(x-1) - \log J_{t+1}(x) \geqslant \log J_t(x-1) - \log J_t(x)\\
&\iff J_{t+1}(x-1)J_t(x) \geqslant J_{t+1}(x)J_t(x-1)\\
&\iff J_t(x)D_{t+1}(x-1) \leqslant J_{t+1}(x)D_t(x-1)\\
&\iff (P(a')-P(a))J_t(x)D_{t+1}(x-1) \leqslant (P(a')-P(a))J_{t+1}(x)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad D_t(x-1)\\
&\iff -P(a)J_t(x)D_{t+1}(x-1) - P(a')J_{t+1}(x)D_t(x-1) \leqslant\\
&\qquad\qquad - P(a')J_t(x)D_{t+1}(x-1) - P(a)J_{t+1}(x)D_t(x-1)\\
&\iff [J_{t+1}(x) - P(a)D_{t+1}(x-1)][J_t(x) - P(a')D_t(x-1)] \leqslant\\
&\qquad\qquad [J_{t+1}(x) - P(a')D_{t+1}(x-1)][J_t(x) - P(a)D_t(x-1)]\\
&\iff \frac{g_{t+1}(x,a,\gamma)}{g_t(x,a,\gamma)} \leqslant \frac{g_{t+1}(x,a',\gamma)}{g_t(x,a',\gamma)}\\
&\iff \log g_t(x,a,\gamma) \text{ is superadditive on } \mathbf{A}\times\{0,1,\cdots N-1\}\\
&\iff \log G_t(x,a,\gamma) \text{ is superadditive on } \mathbf{A}\times\{0,1,\cdots N-1\}.\ \Box
\end{aligned}
$$

Now, we show that $\log J_t(x,\gamma)$ is indeed convex in x for each t, and subadditive on $\mathbf{X} \times \{0,1,\dots N-1\}$. Throughout the rest of this paper we will assume the following condition, which is reasonable for some situations.

Assumption 3.1. The terminal cost $c(x)$ is convex.

Lemma 3.1.9 *Under Assumption 3.1, 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* $(\mathbf{X} \times \{0,1,\dots N-1\}, \preccurlyeq)$.

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

$$A_{x,t}: \quad \frac{J_t(x+2,\gamma)}{J_t(x+1,\gamma)} \geqslant \frac{J_t(x+1,\gamma)}{J_t(x,\gamma)} \tag{15}$$

$$B_{x,t}: \quad \frac{J_{t+2}(x,\gamma)}{J_{t+1}(x,\gamma)} \geqslant \frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)} \tag{16}$$

$$C_{x,t}: \quad \frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)} \geqslant \frac{J_{t+1}(x+1,\gamma)}{J_t(x+1,\gamma)} \tag{17}$$

respectively. We will show that those inequalities hold for $t = 0,1,\dots$ $N-2$ and $x = 0,1,\dots I-2$. 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 3.1.6). $B_{0,0}$ is an obvious equality, and $A_{0,0}$ follows from Assumption 3.1. 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 (9) that for some a, say $\bar{a}$,

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

and hence

$$\frac{J_{t+1}(x,\gamma)}{J_t(x,\gamma)} = e^{\gamma\bar{a}}\left[1 - P(\bar{a})\frac{J_t(x,\gamma) - J_t(x-1,\gamma)}{J_t(x,\gamma)}\right]. \tag{18}$$

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 (18) we obtain

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

and $C_{x,t}$ follows. In a similar way, to prove $B_{x,t}$ we have that it follows from (9) that for some a, say a',

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

and hence

$$\frac{J_{t+2}(x,\gamma)}{J_{t+1}(x,\gamma)} = e^{\gamma a'}\left[1-P(a')\frac{J_{t+1}(x,\gamma)-J_{t+1}(x-1,\gamma)}{J_{t+1}(x,\gamma)}\right]. \tag{19}$$

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 (19) we obtain

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

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 (15) we have to show that

$$J_t(x+2,\gamma)J_t(x,\gamma) \geqslant J_{t+1}(x+1,\gamma)J_{t-1}(x+1,\gamma). \tag{20}$$

It follows from (10) that for some a, say $\tilde{a}$,

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

and hence

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

$$\left.\frac{J_{t-1}(x+2,\gamma) - J_{t-1}(x+1,\gamma)}{J_{t-1}(x+1,\gamma)}\right]. \tag{21}$$

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 (21) we obtain

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

and (20) follows. Thus, the proof is complete. □

Corollary 3.1.10 *Under Assumption 3.1, $J_t(x,\gamma)$ is convex in x for each t.*

Proof: Since $J_t(x,\gamma) = \exp(\log J_t(x,\gamma))$, the claim follows from Lemma 3.1.9 (a). □

Lemma 3.1.11 *Under Assumption 3.1,*

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

b) $\log[G_t(x,a,\gamma)]$ *is superadditive on* $\mathbf{A} \times \{0,1,\ldots N-1\}$.

Proof: The results in (a) and (b) follow from Lemmas 3.1.8 and 3.1.9. □

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 [14]. In the following proposition we show an analogous result for the risk-sensitive case.

Proposition 3.1.12 *Under Assumption 3.1, there exists an optimal policy* $\pi = (f_0^*, \dots f_{N-1}^*)$ *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 Lemmas 3.1.7 and 3.1.11 that for $t = 0, 1, \dots N-1$, $\log H_t(x, a, \gamma)$ is subadditive on $\mathbf{X} \times \mathbf{A}$, and subadditive on $\mathbf{A} \times \{0, 1, \dots N-1\}$. The result follows from Theorems 2.1.3 and 2.1.4. □

In the following section we analyze the allocation control problem with ETC criterion for the case in which the probability function $P(a)$ is convex and the final cost $c(x)$ is strictly increasing. We show that under the mentioned conditions, the optimal policy obtained in Proposition 3.1.12 has further structural properties. Moreover, we compare those structured optimal policies with those corresponding to the risk-neutral allocation problem (which are obtained in the appendix). Finally, we apply the obtained results to the particular case of a linear terminal cost function, and again we compare the conclusions with those corresponding to the risk-neutral problem.

4 Allocation Problem with P(a) Convex and c(x) Strictly Increasing.

Throughout this section, $\pi^* = (f_0^*, \dots, f_{t-1}^*)$ will denote the monotone optimal policy obtained in Proposition 3.1.12.

Proposition 4.1.13 *Assume that* $P(a)$ *is convex and twice differentiable and* $c(x)$ *strictly increasing. Then, under Assumption 3.1, the optimal allocation problem can be reduced to a problem with the action set* $\{0, M\}$*. Moreover, the optimal policy* $\pi^* = (f_0^*, f_1^*, \dots f_{N-1}^*)$ *is of the threshold type, that is, there exist states* $x_0^*, x_1^*, \dots, x_{N-1}^*$ *such that*

$$f_t^*(x) = \begin{cases} 0 & \text{if } x < x_t^* \\ M & \text{if } x \geqslant x_t^*, \end{cases} \tag{22}$$

$t = 0, 1, \dots N-1$*. Furthermore, the sequence of thresholds is decreasing.*

Proof: First, we will show by induction on t, that $J_t(x, \gamma)$ is strictly increasing in x. Since $J_0(x, \gamma) = e^{\gamma c(x)}$, the result holds for $t = 0$. Now assume that $J_t(x, \gamma)$ is strictly increasing in x for some $t \geq 0$. Then, from Corollary 3.1.10 and by using the induction hypothesis we have that

$$\begin{aligned} e^{\gamma a}[J_t(x,\gamma) + &(1 - P(a))(J_t(x+1,\gamma) - J_t(x,\gamma))] \\ &> e^{\gamma a}[J_t(x-1,\gamma) + (1 - P(a))(J_t(x,\gamma) - J_t(x-1,\gamma))] \end{aligned}$$

and since $e^{\gamma a}[J_t(x,\gamma) + (1 - P(a))(J_t(x+1,\gamma) - J_t(x,\gamma))]$ is continuous in a,

$$\begin{aligned} &\inf_{a \in [0,M]} \{e^{\gamma a}[J_t(x,\gamma) + (1 - P(a))(J_t(x+1,\gamma) - J_t(x,\gamma))]\} \\ &> \inf_{a \in [0,M]} \{e^{\gamma a}[J_t(x-1,\gamma) + (1 - P(a))(J_t(x,\gamma) - J_t(x-1,\gamma))]\}. \end{aligned}$$

Thus, from (10), $J_{t+1}(x+1,\gamma) > J_{t+1}(x,\gamma)$. Next, 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 (13) that

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

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

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

and

$$\begin{aligned} \frac{\partial^2 G_t}{\partial^2 a}(x, a, \gamma) &= -e^{\gamma a}[J_t(x,\gamma) - J_t(x-1,\gamma)]P''(a) - \\ &\quad \gamma e^{\gamma a}[J_t(x,\gamma) - J_t(x-1,\gamma)]P'(a) + \gamma \frac{\partial G_t}{\partial a}(x, a, \gamma) \\ &= \gamma \frac{\partial G_t}{\partial a}(x, a, \gamma) + e^{\gamma a}[J_t(x-1,\gamma) - J_t(x,\gamma)] \\ &\qquad [\gamma P'(a) + P''(a)] \end{aligned}$$

If $a_x \in (0, M)$ is such that $\frac{\partial G_t}{\partial a}(x, a_x, \gamma) = 0$, then $\frac{\partial^2 G_t}{\partial^2 a}(x, a, \gamma) < 0$ since $J_t(x-1,\gamma) - J_t(x,\gamma) < 0$ and $\gamma P'(a) + P''(a) > 0$.

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

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

$t = 0, 1, \dots N-1$, then (22) 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. □

Now, to gain further insight of the consequences of Proposition 4.1.13, we apply this proposition to compute the optimal policy in a particular example with linear final cost.

Example 4.1.14 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 4.1.13, 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,

$$G_0(x, 0, \gamma) = e^{2\gamma x} \tag{23}$$

and

$$G_0(x, 1, \gamma) = e^{2\gamma x}[P(1) e^{-\gamma} + (1 - P(1)) e^{\gamma}]. \tag{24}$$

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

$$\begin{aligned} 1 \leq P(1) e^{-\gamma} + (1 - P(1)) e^{\gamma} &\iff e^{\gamma} \leq P(1) + e^{2\gamma}(1 - P(1)) \\ &\iff (1 - P(1)) \left[e^{2\gamma} - \frac{1}{1 - P(1)} e^{\gamma} + \frac{P(1)}{1 - P(1)} \right] \geq 0 \\ &\iff e^{2\gamma} - \frac{1}{1 - P(1)} e^{\gamma} + \frac{P(1)}{1 - P(1)} \geq 0 \\ &\iff \left(e^{\gamma} - \frac{P(1)}{1 - P(1)} \right) \left(e^{\gamma} - 1 \right) \geq 0 \\ &\iff \gamma \geq \log \frac{P(1)}{1 - P(1)}. \end{aligned} \tag{25}$$

Thus, it follows from (23), (24) and (25) 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

(26)
$$f^*_{N-1}(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geqslant 1, \end{cases}$$

and
(27)
$$J_1(x,\gamma) = \begin{cases} 1 & \text{if } x = 0 \\ e^{\gamma}[P(1)J_0(x-1,\gamma) + (1-P(1))J_0(x,\gamma)] & \text{if } x \geq 1, \end{cases}$$

and for (b) and (c) by

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

and

(28)
$$J_1(x,\gamma) = e^{2\gamma x}, \quad x \geq 0.$$

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 $t = 1, \ldots N-1$,

$$J_t(x,\gamma) = \begin{cases} 1 & \text{if } x = 0 \\ e^{\gamma}[P(1)J_{t-1}(x-1,\gamma) + (1-P(1))J_t(x,\gamma)] & \text{if } x \geq 1, \end{cases}$$

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 $t = 1, \ldots N-1$,

$$J_t(x,\gamma) = \begin{cases} 1 & \text{if } x = 0 \\ e^{\gamma}[P(1)J_{t-1}(x-1,\gamma) + (1-P(1))J_t(x,\gamma)] & \text{if } x \geq 1, \end{cases}$$

First, let's prove (I). The validity of assertion (I) for $t = 1$ follows from (27). Next, by (8),

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

where

$$G_t(x, a, \gamma) = e^{\gamma a}[P(a)J_t(x-1, \gamma) + (1-P(a))J_t(x, \gamma)], \quad x \geq 1. \tag{29}$$

Thus,

$$\begin{aligned} J_{t+1}(x,\gamma) &= \min\{J_t(x,\gamma), e^{\gamma}[P(1)J_t(x-1,\gamma) + (1-P(1))J_t(x,\gamma)]\} \\ &= \min\{e^{\gamma}[P(1)J_{t-1}(x-1,\gamma) + (1-P(1))J_{t-1}(x,\gamma)], \\ &\qquad\qquad e^{\gamma}[P(1)J_t(x-1,\gamma) + (1-P(1))J_t(x,\gamma)]\} \qquad (30) \\ &= e^{\gamma}[P(1)J_t(x-1,\gamma) + (1-P(1))J_t(x,\gamma)], \qquad (31) \end{aligned}$$

where (30) and (31) follow from the induction hypothesis and Lemma 3.1.6 respectively. Thus, the proof of (I) is complete.

Now, let's prove (II). First, (28) implies that (II) holds for $t = 1$. Next, similarly as above,

$$\begin{aligned} J_{t+1}(I,\gamma) &= \min\{G_t(I,0,\gamma),\, G_t(I,1,\gamma)\} \\ &= \min\{J_t(I,\gamma), e^{\gamma}[P(1)J_t(I-1,\gamma) + (1-P(1))J_t(I,\gamma)]\} \qquad (32) \\ &= \min\{J_0(I,\gamma), e^{\gamma}[P(1)J_0(I-1,\gamma) + (1-P(1))J_0(I,\gamma)]\} \qquad (33) \\ &= \min\{J_0(I), J_0(I)[e^{-\gamma}P(1) + e^{\gamma}(1-P(1))], \\ &= J_0(I) \qquad (34) \end{aligned}$$

where (32), (33) and (34) follow from (29), the induction hypothesis and (25) 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 (34) follows from (25) 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

$$f^*_t(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geqslant 1 \end{cases} \tag{35}$$

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 $P(1) \leq \frac{1}{2}$ and $\gamma > 0$.

Remark 4.1.15 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 the appendix;

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 the appendix.

5 Appendix

For the purpose of comparing the results obtained in Sections 3 and 4 about structured optimal policies for an optimal allocation problem (with ETC criterion), in this appendix we study the corresponding risk-neutral allocation problem. Section 5.1 summarizes some results about monotonicity and convexity properties of the optimal value function and monotonicity properties of the policies. For the proof of those results we refer the reader to Ross [14]. In Section 5.2 we derive further structural properties of the optimal policies under the assumptions that the probability function $P(a)$ is convex and the final cost $c(x)$ is strictly increasing.

5.1 Monotone Optimal Policies

For $t = 0, 1, \ldots, N-1$, denote

$$F_t(x,a) := a + P(a)J_t(x-1) + (1-P(a))J_t(x), \quad x \geq 1, \tag{36}$$

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

$$J_0(x) = c(x) \tag{37}$$

$$\vdots \qquad \vdots$$

$$J_{t+1}(x) = \inf_{a \in A(x)} \big\{ \mathbf{C}(x,a) + \sum_y p_{xy}(a) J_t(y) \big\}. \tag{38}$$

Let

$$\bar{A}_t(x) := \{a : F_t(x,a) = \inf_{a'} \{F_t(x,a')\}\}$$

and

$$\bar{f}_t(x) := \min \bar{A}_t(x).$$

Lemma 5.1.1 *The optimal value function $J_t(x)$ is increasing in x and decreasing in t. Moreover, under Assumption 3.1, $J_t(x)$ is convex in x.*

Proposition 5.1.2 *Under Assumption 3.1, $\bar{\pi} = (\bar{f}_0, \ldots, \bar{f}_{N-1})$ 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.*

5.2 Risk-neutral Allocation Problem with P(a) Strictly Convex and c(x) Strictly Increasing.

Throughout this appendix, the policy $\bar{\pi} = (\bar{f}_0, \ldots, \bar{f}_{N-1})$ will denote the monotone optimal policy obtained in Proposition 5.1.2. In the following proposition we will show that when the probability function $P(a)$ is strictly convex and the final cost $c(x)$ is strictly increasing, 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 5.2.1 *Assume that $P(a)$ is strictly convex and twice differentiable and $c(x)$ is strictly increasing. Then, under Assumption 3.1, 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*

$$\bar{f}_t(x) = \begin{cases} 0 & \text{if } x < \bar{x}_t \\ M & \text{if } x \geqslant \bar{x}_t, \end{cases} \tag{39}$$

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

Proof: It follows from (36) that

$$F_t(x, a) = a + [J_t(x) - J_t(x-1)](1 - P(a)) + J_t(x-1). \tag{40}$$

First, we will show by induction on t, that $J_t(x)$ is strictly increasing in x. Since $J_0(x) = c(x)$, the result holds for $t = 0$. Now assume that $J_t(x) < J_t(x+1)$. Then, from Lemma 5.1.1 and by using the induction hypothesis we have that

$$\begin{aligned} a + J_t(x) + &(1 - P(a))[J_t(x+1) - J_t(x)] \\ &> a + J_t(x-1) + (1 - P(a))[J_t(x) - J_t(x-1)] \end{aligned}$$

and since $a + J_t(x) + (1 - P(a))[J_t(x+1) - J_t(x)]$ is continuous in a,

$$\begin{aligned} J_{t+1}(x+1) &= \inf_{a \in [0,M]} \{a + J_t(x) + (1 - P(a))[J_t(x+1) - J_t(x)]\} \\ &> \inf_{a \in [0,M]} \{a + J_t(x-1) + (1 - P(a))[J_t(x) - J_t(x-1)]\} = J_{t+1}(x). \end{aligned}$$

It follows from (40) that

$$\frac{\partial F_t}{\partial a}(x,a) = 1 - P'(a)[J_t(x) - J_t(x-1)]$$

and

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

Thus, since $P''(a) > 0$ and $J_t(x)$ is strictly increasing in x 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 (39) 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. □

Now, we will apply Proposition 5.2.1 to compute the optimal policy for the example considered in Section 4.

Example 5.2.2 (revisited.) Take the example considered in Section 4 with $P(a)$ strictly convex. First we compute $\bar{f}_{N-1}(x)$. To do that, by Proposition 5.2.1, 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 (36) 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

$$\bar{f}_{N-1}(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geqslant 1, \end{cases} \tag{41}$$

and

$$J_1(x) = \begin{cases} 0 & \text{if } x = 0 \\ 2x + (1 - 2P(1)) & \text{if } x \geq 1; \end{cases} \tag{42}$$

for the case (b) by

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

and

$$J_1(x) = 2x, \quad x \geq 0; \tag{43}$$

and for the case (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 $P(1) \leq \frac{1}{2}$ 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 (42). Next, by the dynamic programming algorithm

$$J_{t+1}(1) = \min\{F_t(1,0),\, F_t(1,1)\}.$$

Thus,

$$\begin{aligned} J_{t+1}(1) &= \min\{J_t(1), 1 + (1 - P(1))J_t(1)\} \\ &= \min\{1 + (1 - P(1))J_{t-1}(1), 1 + (1 - P(1))J_t(1)\} && (44) \\ &= 1 + (1 - P(1))J_t(1), && (45) \end{aligned}$$

where (44) and (45) follow from the induction hypothesis and Lemma 5.1.1 respectively. Thus the proof of (I) is complete.

Now, let's prove (II). First, (43) implies that (II) holds for $t = 1$. Next, similarly as above

$$\begin{aligned} J_{t+1}(I) &= \min\{F_t(I,0),\ F_t(I,1)\} \\ &= \min\{J_t(I), 1 + P(1)J_t(I-1) + (1-P(1))J_t(I)\} \quad (46) \\ &= \min\{2I, 1 + P(1)2(I-1) + (1-P(1))2I\} \quad (47) \\ &= \min\{2I, 2I + (1-2P(1))\} \\ &= 2I \quad (48) \end{aligned}$$

where (46), (47) and (48) follow from (36), 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:

$$f_t^y(x) = \begin{cases} 0 & \text{if } x \leq y \\ 1 & \text{if } x > y, \end{cases}$$

$y = 0, 1, \ldots I$.

Acknowledgement

This paper is part of the author's doctoral research under the direction of Dr. Emmanuel Fernández Gaucherand at the Department of Mathematics of the University of Arizona. I want to thank him for his guidance during the preparation of this work. I am specially grateful to Dr. Onésimo Hernández Lerma for his helpful comments and suggestions in helping me see this article to its completion.

Guadalupe Avila Godoy
Departamento de Matemáticas,
Universidad de Sonora,
Rosales y Boulevard Luis Encinas,
Hermosillo, Sonora, 83000
gavila@gauss.mat.uson.mx

References

[1] Arapostathis A.; Borkar V. S.; Fernández-Gaucherand E.; Ghosh M. K.; Marcus S. I., *Discrete-time Controlled Markov Processes with Average Cost Criterion: a Survey*, SIAM J. Control Optim., **31** (1993), 282-344.

[2] Avila-Godoy G.; Brau A.; Fernández-Gaucherand E., *Controlled Markov Chains with Discounted Risk-sensitive Criteria: Applications to Machine Replacement*, In Proc. 36th IEEE Conf. Decision Control, (1997), 1115-1120.

[3] Avila-Godoy G.; Fernández-Gaucherand E., *Exponential Risk-sensitive Optimal Scheduling*, In Proc. 36th IEEE Conf. Decision Control, (1997), 3958-3963.

[4] Avila-Godoy G., *Controlled Markov Chains with Exponential Risk Sensitive Criteria: Modularity, Structured Policies and Applications*, PhD thesis, Department of Mathematics, University of Arizona, (1998).

[5] Bouakiz M.; Sobel M. J., *Inventory Control with an Exponential Utility Criterion*, Math. Oper. Res., **40** (1992), 603-608.

[6] Brau A.; Fernández-Gaucherand E., *Controlled Markov Chains with Risk Sensitive Exponential Average Cost Criterion*, In Proc. 36th IEEE Conf. Decision Control, (1997), 2260-2264.

[7] Brau A.; Fernández-Gaucherand E., *Controlled Markov Chains with Risk-Sensitive Average Cost Criterion: The Non-Irreducible Case*, In Proc. 40th IEEE Conf. Decision Control, (2001), 2108-2109.

[8] Brau A.; Fernández-Gaucherand E., *On Weak Conditions and Optimality Inequality Solutions in Risk-Sensitive Controlled Markov Processes with Average Criterion*, In Proc. 41th IEEE Conf. Decision Control, (2002), 1375-1379.

[9] Fernández-Gaucherand E.; Marcus S. I., *Risk-sensitive Optimal Control of Hidden Markov Models: Structural Results*, IEEE Trans. Autom. Control, **42** (1997), 1418-1442.

[10] Hernández-Lerma O.; Lasserre J., Discrete Time Markov Control Processes, Springer, New York, (1996).

[11] Hinderer K. F., *On the Structure of Solutions of Stochastic Dynamic Programs*, In Proc. 7th Conf. Probability Theory, (1982), 173-182.

[12] Pratt J. W., *Risk Aversion in the Small and in the Large*, Econometrica, **32** (1964), 122-136.

[13] Puterman M. L., Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley, New York, (1994).

[14] Ross S. M., Introduction to Stochastic Dynamic Programming, Academic Press, New York, (1983).

[15] Serfozo R. F., *Monotone Optimal Policies for Markov Decision Processes*, Math. Prog. Study, **6** (1976), 202-215.

[16] Serfozo R. F., *Optimal Control of Random Walks, Birth and Death Processes, and Queues*, Adv. Appl. Prob., **13** (1981), 61-83.

[17] Stidham S., Jr.; Weber R. R., *Monotonic and Insensitive Optimal Policies for Control of Queues with Undiscounted Costs*, Oper. Res., **87** (1989), 611-625.

[18] Vega-Amaya O.; Montes-de-Oca R., *Application of Average Dynamic Programming to Inventory Systems*, Math. Met. Oper. Res., **47** (1998), 451-471.

[19] Weber R. R.; Stidham S. Jr., *Optimal Control of Service Rates in Networks of Queues*, Adv. Appl. Prob., **19** (1987), 202-218.

[20] Whittle P., Risk-sensitive Optimal Control, John Wiley & Sons, Chichester, (1990).