Research Article

Risk-Sensitive Multiagent Decision-Theoretic Planning Based on MDP and One-Switch Utility Functions

Wei Zeng, Hongtao Zhou, and Mingshan You

School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China

Correspondence should be addressed to Wei Zeng; zengwei@mail.hust.edu.cn

Received 28 April 2014; Accepted 16 June 2014; Published 7 July 2014

Academic Editor: Wei Zhang

Copyright © 2014 Wei Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In high stakes situations decision-makers are often risk-averse and decision-making processes often take place in group settings. This paper studies multiagent decision-theoretic planning under Markov decision processes (MDPs) framework with considering the change of agent's risk attitude as his wealth level varies. Based on one-switch utility function that describes agent's risk attitude change with his wealth level, we give the additive and multiplicative aggregation models of group utility and adopt maximizing expected group utility as planning objective. When the wealth level approaches infinity, the characteristics of optimal policy are analyzed for the additive and multiplicative aggregation model, respectively. Then a backward-induction method is proposed to divide the wealth level interval from negative infinity to initial wealth level into subintervals and determine the optimal policy in states and subintervals. The proposed method is illustrated by numerical examples and the influences of agent's risk aversion parameters and weights on group decision-making are also analyzed.

1. Introduction

Decision-theoretic planning is to compute optimal policy that is formed by courses of action to maximize expected reward with considering actions that have uncertain outcomes [1]. In high stakes situations with the possibility of high wins and losses, such as emergency and crisis response, business and investment decision, military battle, and lottery, decision-makers are often risk-averse. In the risk-sensitive decisions, exponential utility function is one of the typical utility functions to model decision-maker's risk aversion and maximizing expected utility is the most commonly used rule instead of maximizing expected reward. However, the risk attitude of a decision-maker modeled by exponential utility function is independent of his wealth level and does not change as his wealth level varies, while in reality personal risk attitude often changes with his wealth level [2–4]. Bell proposes a kind of utility function, named one-switch utility function, to model an agent who is always risk-averse but becomes risk neutral as his wealth increases [2]. Bell and Fishburn take further studies on the characteristics of a form

of one-switch utility function that is a combination of linear utility function and exponential utility function [5]. Liu and Koenig give a form of one-switch utility function with considering agent's risk aversion parameter γ ($0 < \gamma < 1$); that is, $u(w) = w - D\gamma^w$, where w denotes the wealth level and D is a parameter to adjust tradeoff between risk neutrality and risk aversion. This form of one-switch utility function not only describes the change of agent's risk attitude, but also presents the degree of agent's risk aversion by the quantitative risk aversion parameter γ [6–8].

For decision-theoretic planning problems, Markov decision processes (MDPs) framework is adopted broadly as an underlying model. Howard and Matheson in their seminal paper introduce risk-sensitive MDPs based on maximizing the expected exponential utility [9]. In the follow-up related studies structural properties of optimal solution and algorithms to compute optimal policy are investigated based on exponential utility function [10–12]. If an agent is risk-sensitive, it is necessary to consider the possible change of agent's risk attitude when his wealth level varies and further influence on the decisions in the next stage. Liu and Koenig

study Markov decision processes with considering agent's risk-sensitive attitude modeled by one-switch utility function and propose an exact backward-induction algorithm to compute optimal policy [8].

In reality decision-making processes often take place in group settings due to a single decision-maker's limited decision-making ability. For the group decision-making problem, group utility is usually got by aggregating personal utilities and then group decisions are made based on the group utility. The aggregation methods include additive value model and multiplicative value rule. Other methods such as multiobjective linear programming [13], fuzzy sets method [14, 15], and interactive approach [16, 17] are used to aggregate individual decision information including attribute weights and attribute values into group decisions. Besides, some researches on group decision-making problem take time into consideration. Xu investigates multistage multiattribute group decision-making problems in which the weight information on a collection of attributes and the decision information on a finite set of alternatives with respect to the attributes are collected at different stages [18].

This paper focuses on decision-theoretic planning problem in which sequential decisions are made by a group of risk-sensitive members. Considering agent's risk-sensitive attitude and wealth level, this paper studies the risk-sensitive multiagent decision-theoretic planning problem based on one-switch utility function and MDP framework. Two group utility functions based, respectively, on additive value model and multiplicative value model of one-switch utility functions are given. Backward-induction algorithms for these two kinds of group utility functions to compute optimal policy of risk-sensitive group decision-making under MDP framework are proposed.

The rest of this paper is organized as follows. One-switch utility function and risk-sensitive MDP model augmented with wealth level are introduced in Section 2. In Section 3, additive and multiplicative aggregation model of one-switch utility functions are given. We analyze the characteristics of optimal policy when the wealth level approaches negative infinity for additive and multiplicative aggregation model, respectively, in Section 3. In Section 4, detailed backward-induction algorithms are proposed to solve the multiagent decision-theoretic planning problem based, respectively, on additive and multiplicative aggregation model. Numerical examples are used to illustrate the proposed method and analyze the influences of agent's risk aversion parameters and weights on group decision-making in Section 5. Finally, a conclusion of this paper and suggested topics for future research are presented in Section 6.

2. Risk-Sensitive MDP Model Augmented with Wealth Level

2.1. One-Switch Utility Function. One-switch utility function is a kind of utility function to describe the change of agent's risk attitude as his wealth level varies. In detail, there exists a wealth level w ; when the agent's wealth level is below w , the agent is risk-averse, but when his wealth level increases and

becomes higher than w , the agent becomes risk neutral. For agent i , one-switch utility function given by Liu and Koenig is shown as follows [6–8]:

$$u_i(w) = w - D_i \gamma_i^w = u_l + D \cdot u_{i,e}(w), \quad (1)$$

where w is wealth level, γ_i is agent i 's risk aversion parameter, and $0 < \gamma_i < 1$. D_i is a constant that provides an adjustable tradeoff between risk neutrality (linear term) and risk aversion (exponential term). $u_l = w$ is a linear utility function. $u_{i,e}(w) = -\gamma_i^w$ is agent i 's exponential utility function.

2.2. Risk-Sensitive MDP Model Augmented with Wealth Level. In the paper goal directed Markov decision problem (GDMDP) is adopted as underlying model of decision-theoretic planning problem [8]. GDMDP is a kind of MDP with a finite set of goal states. When an agent reaches a goal state, he stops acting and receives no more rewards thereafter. One-switch utility function is used to describe the agent's risk-sensitive attitude and maximizing expected utility is adopted as planning objective instead of maximizing expected reward. As wealth level is included in the one-switch utility function, it is necessary to consider the wealth level as a component of the system state of GDMDP.

Formally, a GDMDP consists of a finite set of states S with wealth levels W , so the augmented state set of GDMDP is denoted by (S, W) . Goal state set is (G, W) , where $G \subseteq W$. Nongoal state set is (S', W) , where $S' = S \setminus G$.

The agent's action set is A_s . The agent chooses an action $a \in A_s$ to execute in its current state $s \in S$.

The agent's execution of action a in state s results in a finite reward $r(s, a, s')$ and a transition to successor state $s' \in S$ with probability $P(s' | s, a)$. In the paper only cost is considered and assumed reward is strictly negative, $r(s, a, s') < 0$.

We also use s_t and a_t to denote the state and action at time step t ($t = 0, 1, 2, \dots$). $r_t = r(s_t, a_t, s_{t+1})$ is used to denote the reward for executing action a_t . After the agent reaches a goal state, $r_t = 0$.

$w_t = w_0 + \sum_{i=0}^{t-1} r_i$ is the agent's wealth level at time step t , where the initial wealth level is denoted by w_0 .

For the MDP model augmented with wealth level, the optimal policy maps every combination of a state $s \in S'$ and wealth level w to an action $a \in A_s$ that an agent in state s with wealth level w should execute to maximize expected utility.

For agent i 's exponential utility function $u_{i,e}(w)$ and all policies π , we define the value $v_{i,e}^\pi(s, w) = \lim_{t \rightarrow \infty} E_{s,w}^\pi[u_{i,e}(w_t)]$ as the expected exponential utility of agent i with initial state s and initial wealth level w that follows policy π .

The optimal value $v_{i,e}^*(s, w) = \max_\pi v_{i,e}^\pi(s, w)$ is defined as the highest possible expected exponential utility of agent i with initial state s and initial wealth level w . Assume $v_{i,e}^*(s, w)$ is finite for all state $s \in S$ and wealth levels w .

An optimal policy $\pi_{i,e}^*(s, w)$ is defined as a policy with $v_{i,e}^{\pi_{i,e}^*}(s, w) = v_{i,e}^*(s, w)$ for all state $s \in S$ and wealth levels w .

Similarly, for linear utility function $u_l(w)$, we use $v_l^\pi(s, w)$, $v_l^*(s, w)$, and $\pi_l^*(s, w)$ to denote expected utility, optimal value, and optimal policy, respectively.

It is worth noting that differently from Liu and Koenig [8], the paper focuses on the decision-making in group setting. So the value function in MDP will be replaced by group utility function which is the aggregation of personal one-switch utilities and the planning objective is to maximize the expected group utility.

3. Utility Aggregation Model of One-Switch Utility Functions

Group utility is the aggregation of personal utilities. The common methods include additive value model and multiplicative value model. In the following sections we will discuss additive and multiplicative value model for the aggregation of personal one-switch utility functions, respectively.

3.1. Additive Aggregation Model of One-Switch Utility Functions. In general, additive aggregation model of group utility is defined as follows:

$$u(w) = \sum_{i=1}^n k_i u_i(w), \quad (2)$$

where k_i is the weight of agent i 's utility, $0 < k_i < 1$, and $\sum_{i=1}^n k_i = 1$.

Thus the additive aggregation model of one-switch utility functions is defined as follows:

$$u(w) = \sum_{i=1}^n k_i u_i(w) \quad (3)$$

$$= w - k_1 D_1 \gamma_1^w - k_2 D_2 \gamma_2^w - \dots - k_n D_n \gamma_n^w.$$

According to MDP, the expected group utility for all policies π and state (s, w) is presented as follows:

$$\begin{aligned} v^\pi(s, w) &= u(w) \\ &= w - k_1 D_1 \gamma_1^w \\ &\quad - k_2 D_2 \gamma_2^w - \dots - k_n D_n \gamma_n^w \quad \forall s \in G, \quad \forall w \\ v^\pi(s, w) &= \sum_{s' \in S} P(s' | s, \pi(s, w)) \\ &\quad \cdot v^\pi(s', w + r(s, \pi(s, w), s')) \quad \forall s \in S', \quad \forall w. \end{aligned} \quad (4)$$

Then, the optimal value $v^*(s, w)$ is presented as follows:

$$\begin{aligned} v^*(s, w) &= u(w) \\ &= w - k_1 D_1 \gamma_1^w \\ &\quad - k_2 D_2 \gamma_2^w - \dots - k_n D_n \gamma_n^w \quad \forall s \in G, \quad \forall w \\ v^*(s, w) &= \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) \\ &\quad \cdot v^*(s', w + r(s, a, s')) \quad \forall s \in S', \quad \forall w. \end{aligned} \quad (5)$$

Next, we will derive the relationship between the expected group utility and the expected personal linear and exponential utility for the additive aggregation model.

For all policies π , the expected group utility of the additive aggregation model of one-switch utility functions is presented as follows:

$$\begin{aligned} v^\pi(s, w) &= \lim_{t \rightarrow \infty} E_{s,w}^\pi [u(w_t)] \\ &= \lim_{t \rightarrow \infty} E_{s,w}^\pi [u_l(w_t) + k_1 D_1 u_{1,e}(w_t) \\ &\quad + k_2 D_2 u_{2,e}(w_t) \\ &\quad + \dots + k_n D_n u_{n,e}(w_t)] \\ &= \lim_{t \rightarrow \infty} E_{s,w}^\pi [u_l(w_t)] \\ &\quad + \lim_{t \rightarrow \infty} E_{s,w}^\pi [k_1 D_1 u_{1,e}(w_t)] \\ &\quad + \dots + \lim_{t \rightarrow \infty} E_{s,w}^\pi [k_n D_n u_{n,e}(w_t)] \\ &= \lim_{t \rightarrow \infty} E_{s,w}^\pi \left[w + \sum_{i=0}^{t-1} r_i \right] \\ &\quad + \lim_{t \rightarrow \infty} E_{s,w}^\pi \left[k_1 D_1 \gamma_1^{w + \sum_{i=0}^{t-1} r_i} \right] \\ &\quad + \dots + \lim_{t \rightarrow \infty} E_{s,w}^\pi \left[k_n D_n \gamma_n^{w + \sum_{i=0}^{t-1} r_i} \right] \\ &= w + \lim_{t \rightarrow \infty} E_{s,w}^\pi \left[\sum_{i=0}^{t-1} r_i \right] \\ &\quad + k_1 D_1 \gamma_1^w \lim_{t \rightarrow \infty} E_{s,w}^\pi \left[\gamma_1^{\sum_{i=0}^{t-1} r_i} \right] \\ &\quad + \dots + k_n D_n \gamma_n^w \lim_{t \rightarrow \infty} E_{s,w}^\pi \left[\gamma_n^{\sum_{i=0}^{t-1} r_i} \right] \\ &= w + v_l^\pi(s) + k_1 D_1 \gamma_1^w v_{1,e}^\pi(s) \\ &\quad + \dots + k_n D_n \gamma_n^w v_{n,e}^\pi(s), \end{aligned} \quad (6)$$

where $v_l^\pi(s)$ and $v_{i,e}^\pi(s)$ satisfy the following policy-evaluation equations, respectively, [8, 19, 20]:

$$\begin{aligned} v_l^\pi(s) &= 0 \quad s \in G \\ v_l^\pi(s) &= \sum_{s' \in S} P(s' | s, \pi(s)) [r(s, \pi(s), s') + v_l^\pi(s')] \quad s \in S' \\ v_{i,e}^\pi(s) &= -1 \quad s \in G \\ v_{i,e}^\pi(s) &= \sum_{s' \in S} P(s' | s, \pi(s)) \left[\gamma_i^{r(s, \pi(s), s')} \cdot v_{i,e}^\pi(s') \right] \quad s \in S'. \end{aligned} \quad (7)$$

From the above policy-evaluation equation (6), we obtain the relationship between $v^\pi(s, w)$ and expected linear utility $v_l^\pi(s)$ and expected personal exponential utility $v_{i,e}^\pi(s)$ for the additive aggregation model, where $v_l^\pi(s)$ and $v_{i,e}^\pi(s)$ are independent of the wealth level w .

3.2. Multiplicative Aggregation Model of One-Switch Utility Functions. In the paper we adopt the following multiplicative aggregation model of group utility:

$$u(w) = \sum_{i=1}^n k_i u_i(w) + \sum_{i=1, j>i}^n k_{ij} u_i(w) u_j(w) + \cdots + k_{123 \dots n} u_1(w) u_2(w) \cdots u_n(w), \quad (8)$$

where $k_1, k_2, \dots, k_n, k_{12}, k_{23}, \dots, k_{123 \dots n}$ are constants and $0 < k_i < 1$.

For simplicity, in the paper we only consider the case $n = 2$. For $n > 2$, the derivation of multiplicative aggregation model is similar.

For $n = 2$, multiplicative aggregation model of one-switch utility functions is simplified as follows:

$$\begin{aligned} u(w) &= k_1 u_1(w) + k_2 u_2(w) + k_{12} u_1(w) u_2(w) \\ &= k_{12} w^2 - k_{12} D_1 w \gamma_1^w - k_{12} D_2 w \gamma_2^w \\ &\quad + k_{12} D_1 D_2 (\gamma_1 \gamma_2)^w + k_1 w + k_2 w \\ &\quad - k_1 D_1 \gamma_1^w - k_2 D_2 \gamma_2^w. \end{aligned} \quad (9)$$

According to MDP, the optimal value $v^*(s, w)$ is presented as follows:

$$\begin{aligned} v^*(s, w) &= u(w) \quad \forall s \in G, \quad \forall w \\ v^*(s, w) &= \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) \\ &\quad \cdot v^*(s', w + r(s, a, s')) \quad \forall s \in S', \quad \forall w. \end{aligned} \quad (10)$$

For the multiplicative aggregation model of one-switch utility functions and all policies π , the expected group utility is

$$\begin{aligned} v^\pi(s, w) &= \lim_{t \rightarrow \infty} E_{s,w}^\pi [u(w_t)] \\ &= \lim_{t \rightarrow \infty} E_{s,w}^\pi [k_1 u_1(w) + k_2 u_2(w) + k_{12} u_1(w) u_2(w)] \\ &= k_1 \lim_{t \rightarrow \infty} E_{s,w}^\pi [u_1(w)] + k_2 \lim_{t \rightarrow \infty} E_{s,w}^\pi [u_2(w)] \\ &\quad + k_{12} \lim_{t \rightarrow \infty} E_{s,w}^\pi [u_1(w)] \cdot \lim_{t \rightarrow \infty} E_{s,w}^\pi [u_2(w)]. \end{aligned} \quad (11)$$

According to the fact that $\lim_{t \rightarrow \infty} E_{s,w}^\pi [u_i(w)] = w + v_i^\pi(s) + D_1 \gamma_1^w v_{1,e}^\pi(s)$ [8], we have

$$\begin{aligned} v^\pi(s, w) &= k_1 (w + v_1^\pi(s) + D_1 \gamma_1^w v_{1,e}^\pi(s)) \\ &\quad + k_2 (w + v_2^\pi(s) + D_2 \gamma_2^w v_{2,e}^\pi(s)) \\ &\quad + k_{12} (w + v_1^\pi(s) + D_1 \gamma_1^w v_{1,e}^\pi(s)) \\ &\quad \times (w + v_2^\pi(s) + D_2 \gamma_2^w v_{2,e}^\pi(s)). \end{aligned} \quad (12)$$

From the above policy-evaluation equation (12), we obtain the relationship between expected group utility $v^\pi(s, w)$ and expected linear utility $v_i^\pi(s)$ and expected personal exponential utility $v_{i,e}^\pi(s)$ for the multiplicative aggregation model.

4. Preparation for Backward-Induction Method

To solve the optimal policy of the additive and multiplicative aggregation model of one-switch utility functions, backward-induction method is adopted. In the paper the value range of wealth level is a continuous interval $(-\infty, w_0]$; we first compute the optimal policy when wealth level $w \rightarrow -\infty$ that is represented by $\pi_{-\infty}^*$. Then increase the wealth level until $\pi_{-\infty}^*$ is no longer an optimal policy and get a wealth level threshold. Increase further the wealth level and get the next wealth level threshold similarly. The backward-induction method ends when the wealth level is larger than initial wealth level w_0 . Thus the continuous wealth level interval is divided into subintervals by the thresholds. We use $((s, (w^i, w^{i+1}]), a)$ to denote action a executed in state s and wealth level interval $(w^i, w^{i+1}]$ (w^i denotes a wealth level threshold and $i = 0, 1, 2, \dots$). In this section we will analyze the characteristics of optimal policy when the wealth level w approaches negative infinity for additive and multiplicative aggregation model, respectively.

Lemma 1. For additive aggregation model of one-switch utility functions, if agent i is the most risk-averse one, that is, $\gamma_i < \gamma_j$ for any $j \neq i, i, j = 1, 2, \dots, n$, then

$$\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_i^{-w} = k_i D_i v_{i,e}^*(s) \quad \text{for all states } s \in S. \quad (13)$$

Proof. For all optimal policies π^* , we have

$$\begin{aligned} \lim_{w \rightarrow -\infty} v^*(s, w) \gamma_i^{-w} &= \lim_{w \rightarrow -\infty} v^{\pi^*}(s, w) \gamma_i^{-w} \\ &= \lim_{w \rightarrow -\infty} \left[w \gamma_i^{-w} + v_{1,e}^{\pi^*}(s) \gamma_i^{-w} + k_1 D_1 \left(\frac{\gamma_1}{\gamma_i} \right)^w v_{1,e}^{\pi^*}(s) \right. \\ &\quad \left. + \cdots + k_i D_i v_{i,e}^{\pi^*}(s) + \cdots + k_n D_n \left(\frac{\gamma_n}{\gamma_i} \right)^w v_{n,e}^{\pi^*}(s) \right]. \end{aligned} \quad (14)$$

As $\gamma_j / \gamma_i > 1, j \neq i$, we can derive $\lim_{w \rightarrow -\infty} k_j D_j (\gamma_j / \gamma_i)^w v_{j,e}^{\pi^*}(s) = 0$.

Thus $\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_i^{-w} = k_i D_i v_{i,e}^{\pi^*}(s) \leq k_i D_i v_{i,e}^*(s)$.

On the other hand, for all optimal policies $\pi_{i,e}^*$, according to the fact that $v^*(s, w) \geq v^\pi(s, w)$ for all policies π , we have

$$\begin{aligned} \lim_{w \rightarrow -\infty} v^*(s, w) \gamma_i^{-w} &\geq \lim_{w \rightarrow -\infty} v^{\pi_{i,e}^*}(s, w) \gamma_i^{-w} \\ &= \lim_{w \rightarrow -\infty} \left[w \gamma_i^{-w} + v_{1,e}^{\pi_{i,e}^*}(s) \gamma_i^{-w} \right. \\ &\quad \left. + k_1 D_1 \left(\frac{\gamma_1}{\gamma_i} \right)^w v_{1,e}^{\pi_{i,e}^*}(s) \right] \end{aligned}$$

$$\begin{aligned}
& + \cdots + k_i D_i v_{i,e}^{\pi^*}(s) \\
& + \cdots + k_n D_n \left(\frac{\gamma_n}{\gamma_i} \right)^w v_{n,e}^{\pi^*}(s) \Big] \\
& = k_i D_i v_{i,e}^{\pi^*}(s).
\end{aligned} \tag{15}$$

Therefore, the lemma holds. $\square$

Lemma 1 implies that the optimal policy for the additive aggregation model of one-switch utility functions is the same as the most risk-averse agent's optimal policy for the exponential utility function as the wealth level $w \rightarrow -\infty$.

Lemma 2. For multiplicative aggregation model of one-switch utility functions, $\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_3^{-w} = k_{12} D_1 D_2 v_{3,e}^{\pi^*}(s)$ for all states $s \in S$ and $\gamma_3 = \gamma_1 \gamma_2$, where $v_{3,e}^{\pi^*}(s, w)$ is the highest expected exponential utility with risk aversion parameter γ_3 .

Proof. For all optimal policies π^* , we have

$$\begin{aligned}
\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_3^{-w} &= \lim_{w \rightarrow -\infty} v^{\pi^*}(s, w) \gamma_3^{-w} \\
&= \lim_{w \rightarrow -\infty} \gamma_3^{-w} \left[k_1 \left(w + v_l^{\pi^*}(s) + D_1 \gamma_1^w v_{1,e}^{\pi^*}(s) \right) \right. \\
&\quad + k_2 \left(w + v_l^{\pi^*}(s) + D_2 \gamma_2^w v_{2,e}^{\pi^*}(s) \right) \\
&\quad + k_{12} \left(w + v_l^{\pi^*}(s) + D_1 \gamma_1^w v_{1,e}^{\pi^*}(s) \right) \\
&\quad \times \left. \left(w + v_l^{\pi^*}(s) + D_2 \gamma_2^w v_{2,e}^{\pi^*}(s) \right) \right].
\end{aligned} \tag{16}$$

As $\gamma_i/\gamma_3 > 1, i = 1, 2$,

Then $\lim_{w \rightarrow -\infty} k_i D_i (\gamma_i/\gamma_3)^w v_{i,e}^{\pi^*}(s) = 0, \lim_{w \rightarrow -\infty} k_i D_i w (\gamma_i/\gamma_3)^w v_{i,e}^{\pi^*}(s) = 0$.

Therefore,

$$\begin{aligned}
\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_3^{-w} &= k_{12} D_1 D_2 v_{1,e}^{\pi^*}(s) v_{2,e}^{\pi^*}(s) \\
&= k_{12} D_1 D_2 \lim_{w \rightarrow -\infty} E_{s,w}^{\pi^*} \left[\gamma_1^{\sum_{i=0}^{t-1} r_i} \right] \cdot \lim_{w \rightarrow -\infty} E_{s,w}^{\pi^*} \left[\gamma_2^{\sum_{i=0}^{t-1} r_i} \right] \\
&= k_{12} D_1 D_2 \lim_{w \rightarrow -\infty} E_{s,w}^{\pi^*} \left[(\gamma_1 \gamma_2)^{\sum_{i=0}^{t-1} r_i} \right] \\
&\leq k_{12} D_1 D_2 v_{3,e}^{\pi^*}(s).
\end{aligned} \tag{17}$$

On the other hand, for all optimal policies $\pi_{3,e}^*$, according to the fact that $v^*(s, w) \geq v^{\pi}(s, w)$ for all policies π , we have

$$\begin{aligned}
\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_3^{-w} &\geq \lim_{w \rightarrow -\infty} v^{\pi_{3,e}^*}(s, w) \gamma_3^{-w}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{w \rightarrow -\infty} \gamma_3^{-w} \left[k_1 \left(w + v_l^{\pi_{3,e}^*}(s) + D_1 \gamma_1^w v_{1,e}^{\pi_{3,e}^*}(s) \right) \right. \\
&\quad + k_2 \left(w + v_l^{\pi_{3,e}^*}(s) + D_2 \gamma_2^w v_{2,e}^{\pi_{3,e}^*}(s) \right) \\
&\quad + k_{12} \left(w + v_l^{\pi_{3,e}^*}(s) + D_1 \gamma_1^w v_{1,e}^{\pi_{3,e}^*}(s) \right) \\
&\quad \times \left. \left(w + v_l^{\pi_{3,e}^*}(s) + D_2 \gamma_2^w v_{2,e}^{\pi_{3,e}^*}(s) \right) \right] \\
&= k_{12} D_1 D_2 v_{1,e}^{\pi_{3,e}^*}(s) v_{2,e}^{\pi_{3,e}^*}(s) = k_{12} D_1 D_2 v_{3,e}^{\pi^*}(s).
\end{aligned} \tag{18}$$

Therefore, the lemma holds. $\square$

Lemma 2 implies that the optimal policy for the multiplicative aggregation of one-switch utility functions is the same as the optimal policy for a virtual agent's exponential utility function as the wealth level $w \rightarrow -\infty$. The virtual agent's risk aversion parameter is the product of every agent's risk aversion parameter in the group.

5. Division of Wealth Level Interval and Backward-Induction Method

The above section gives the optimal policy as the wealth level approaches negative infinity for additive and multiplicative aggregation model of one-switch utility functions, respectively. The next step is to divide the wealth level interval and determine the wealth level thresholds and optimal policies in the intervals by using backward-induction method. In this section we will discuss the backward-induction method in the cases of additive and multiplicative aggregation model.

5.1. The Case of Additive Aggregation Model. For the additive aggregation model of one-switch utility functions, we first give the following theorem to prove the existence of a wealth level threshold $\underline{w}$ and then give the backward-induction algorithm.

Theorem 3. For all optimal policies π^* , there exists a wealth level threshold $\underline{w}$ such that it holds for all states $s \in S'$ and all wealth levels $w \leq \underline{w}$ that

$$\begin{aligned}
v_l^*(s, w) &= v_l^{\pi^*}(s), \quad v_{i,e}^*(s, w) = v_{i,e}^{\pi^*}(s), \\
v^*(s, w) &= v^{\pi^*}(s, w), \quad i = 1, 2, \dots, n.
\end{aligned} \tag{19}$$

Please see Appendix A for the proof of Theorem 3.

Theorem 3 shows the existence of the wealth level threshold $\underline{w}$. Next we will show how to determine the wealth level threshold $\underline{w}$.

After getting $\pi_{-\infty}^*$ when wealth level $w \rightarrow -\infty$, assume $\pi_{-\infty}^*$ is the optimal policy for the wealth level interval $(-\infty, \underline{w}]$; then for wealth level w in the interval $(\underline{w}, \min_{a \in A_s} (\underline{w} - r(s, a, s'))]$, where $r(s, a, s')$ is the reward got by executing one step action, $\pi_{-\infty}^*$ is no longer the optimal

policy, and assume π' is the optimal policy; according to (5); for all nongol states, we have

$$v^*(s, w) = \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) \cdot v^{\pi^*}_{-\infty}(s', w + r(s, a, s')). \quad (20)$$

As $w + r(s, a, s') < \underline{w}$,

$$\begin{aligned} v^*(s, w) &= \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) \cdot v^{\pi^*}_{-\infty}(s', w + r(s, a, s')) \\ &= \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) \cdot \left(w + r(s, a, s') + v^{\pi^*}_{-\infty}(s) \right. \\ &\quad \left. + k_1 D_1 \gamma_1^{w+r(s, a, s')} v^{\pi^*}_{-\infty}(s) \right. \\ &\quad \left. + \dots + k_n D_n \gamma_n^{w+r(s, a, s')} v^{\pi^*}_{-\infty}(s) \right) \\ &= \max_{a \in A_s} \left[w + v^{\pi^*}_{-\infty}(s, a) + k_1 D_1 \gamma_1^w v^{\pi^*}_{-\infty}(s, a) \right. \\ &\quad \left. + \dots + k_n D_n \gamma_n^w v^{\pi^*}_{-\infty}(s, a) \right], \end{aligned} \quad (21)$$

where

$$\begin{aligned} v_l^*(s, a) &= \sum_{s' \in S} P(s' | s, a) [r(s, a, s') + v_l^*(s')], \\ v_{i,e}^*(s, a) &= \sum_{s' \in S} P(s' | s, a) [\gamma_i^{r(s, a, s')} v_{i,e}^*(s')]. \end{aligned} \quad (22)$$

For $s \in S' \setminus G$, $a_j \in A_s \setminus \pi^*_{-\infty}(s)$, because the optimal policy is π' , not $\pi^*_{-\infty}$ now under assumption, we have

$$\begin{aligned} &\sum_{s' \in S} P(s' | s, \pi^*_{-\infty}(s)) \cdot v^{\pi^*}_{-\infty}(s', w + r(s, \pi^*_{-\infty}(s), s')) \\ &\geq \sum_{s' \in S} P(s' | s, a_j) \cdot v^{\pi^*}_{-\infty}(s', w + r(s, a_j, s')) \end{aligned} \quad (23)$$

or, equivalently,

$$\begin{aligned} &w + v^{\pi^*}_{-\infty}(s, \pi^*_{-\infty}(s)) + k_1 D_1 \gamma_1^w v^{\pi^*}_{-\infty}(s, \pi^*_{-\infty}(s)) \\ &\quad + \dots + k_n D_n \gamma_n^w v^{\pi^*}_{-\infty}(s, \pi^*_{-\infty}(s)) \\ &\geq w + v^{\pi^*}_{-\infty}(s, a_j) + k_1 D_1 \gamma_1^w v^{\pi^*}_{-\infty}(s, a_j) \\ &\quad + \dots + k_n D_n \gamma_n^w v^{\pi^*}_{-\infty}(s, a_j). \end{aligned} \quad (24)$$

We can get a wealth level threshold $\underline{w}_{\pi^*_{-\infty}, s, a_j}$ in equality case of the above weak inequality.

From the algorithm above, we can get the wealth level threshold $\underline{w}$:

$$\underline{w} = \min_{s \in S'} \min_{a_j \in A_s \setminus \pi^*_{-\infty}(s)} \min(w_0, \underline{w}_{\pi^*_{-\infty}, s, a_j}). \quad (25)$$

After getting $\underline{w}$, the next step is to divide further the wealth level interval $(\underline{w}, w_0]$ into subintervals and solve the optimal policy for each subinterval similarly to the above algorithm. The main procedure of the backward-induction algorithm for group decision-making in the case of additive aggregation model is listed as follows.

Step 1. By maximizing the expected exponential utility, get the optimal policy $\pi^*_{i,e}$ of the most risk-averse agent when $w \rightarrow -\infty$.

Step 2. According to (6), we have

$$v^*(s, w) = \max_{a \in A_{i,e}^*(s)} \left\{ w + v_l^{\pi^*_{i,e}}(s, a) + k_1 D_1 \gamma_1^w v_{1,e}^{\pi^*_{i,e}}(s, a) \right. \\ \left. + \dots + k_n D_n \gamma_n^w v_{n,e}^{\pi^*_{i,e}}(s, a) \right\}, \quad (26)$$

where $A_{i,e}^*(s) = \arg \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) [\gamma_i^{r(s, a, s')} v_{i,e}^*(s')]$; get the optimal value $v^*(s, w)$ and the values of $v_{1,e}^*(s), v_{2,e}^*(s), \dots, v_{n,e}^*(s), v_l^*(s)$ for all states $s \in S'$ when $w \rightarrow -\infty$.

Step 3. For all states $s \in S', a \in A_s \setminus \pi^*_{-\infty}(s)$, get the values of $v_l^{\pi^*_{-\infty}}(s, a), v_{1,e}^{\pi^*_{-\infty}}(s, a), \dots, v_{n,e}^{\pi^*_{-\infty}}(s, a)$ by (7); then get $\underline{w}_{\pi^*_{-\infty}, s, a_j}$ by Expression (24).

Step 4. Calculate the wealth level threshold $\underline{w}$ according to (25).

Step 5. For the wealth level interval $(\underline{w}, w_0]$, increase further the wealth level according to the reward got by executing one step action, and determine the wealth level threshold and optimal policy similarly to the above steps.

Step 6. If, for all $s \in S'$, the wealth level w is larger than w_0 , then end the algorithm.

5.2. The Case of Multiplicative Aggregation Model. For the multiplicative aggregation model of one-switch utility functions, we also have the following theorem that shows the existence of a wealth level threshold $\underline{w}$.

Theorem 4. For all optimal policies π^* , there exists a wealth threshold $\underline{w}$. For any state $s \in S'$, wealth level $w \leq \underline{w}$,

$$\begin{aligned} v_l^*(s, w) &= v_l^{\pi^*}(s), \quad v_{i,e}^*(s, w) = v_{i,e}^{\pi^*}(s), \\ v^*(s, w) &= v^{\pi^*}(s, w), \quad i = 1, 2. \end{aligned} \quad (27)$$

Please see Appendix B for the proof of Theorem 4.

Similarly to additive aggregation model, we determine the wealth level threshold $\underline{w}$ for the multiplicative aggregation model. Assume that $\pi^*_{-\infty}$ is the optimal policy for the wealth level interval $(-\infty, \underline{w}]$; then for the wealth level w in the interval $(\underline{w}, \min_{a \in A_s} (\underline{w} - r(s, a, s'))]$, $\pi^*_{-\infty}$ is no longer

the optimal policy, and assuming π^* is the optimal policy, according to (10), for all nongoal states, we have

$$v^*(s, w) = \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) \cdot v^{\pi^*}_{-\infty}(s', w + r(s, a, s')). \quad (28)$$

As $w + r(s, a, s') < \underline{w}$,

$$\begin{aligned} v^*(s, w) &= \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) \\ &\cdot \left[k_1 \left(w + r(s, a, s') + v^{\pi^*}_{-\infty}(s) \right) \right. \\ &\quad + D_1 \gamma_1^{w+r(s,a,s')} v^{\pi^*}_{1,e}(s) \\ &\quad + k_2 \left(w + r(s, a, s') + v^{\pi^*}_{-\infty}(s) \right) \\ &\quad + D_2 \gamma_2^{w+r(s,a,s')} v^{\pi^*}_{2,e}(s) \\ &\quad + k_{12} \left(w + r(s, a, s') + v^{\pi^*}_{-\infty}(s) \right) \\ &\quad \left. + D_1 \gamma_1^{w+r(s,a,s')} v^{\pi^*}_{1,e}(s) \right) \\ &\times \left(w + r(s, a, s') + v^{\pi^*}_{-\infty}(s) \right) \\ &\quad + D_2 \gamma_2^{w+r(s,a,s')} v^{\pi^*}_{2,e}(s) \Big] \\ &= \max_{a \in A_s} \left[k_1 \left(w + v^{\pi^*}_{-\infty}(s) + D_1 \gamma_1^w v^{\pi^*}_{1,e}(s) \right) \right. \\ &\quad + k_2 \left(w + v^{\pi^*}_{-\infty}(s) + D_2 \gamma_2^w v^{\pi^*}_{2,e}(s) \right) \\ &\quad + k_{12} \left(w + v^{\pi^*}_{-\infty}(s) + D_1 \gamma_1^w v^{\pi^*}_{1,e}(s) \right) \\ &\quad \left. \times \left(w + v^{\pi^*}_{-\infty}(s) + D_2 \gamma_2^w v^{\pi^*}_{2,e}(s) \right) \right]. \quad (29) \end{aligned}$$

For w , because the optimal policy is π^* , not $\pi^*_{-\infty}$ now under assumption, we have

$$\begin{aligned} &\sum_{s' \in S'} P(s' | s, \pi^*_{-\infty}(s)) \cdot v^{\pi^*}_{-\infty}(s', w + r(s, \pi^*_{-\infty}(s), s')) \\ &\leq \sum_{s' \in S'} P(s' | s, a_j) \cdot v^{\pi^*}_{-\infty}(s', w + r(s, a_j, s')) \end{aligned} \quad (30)$$

or, equivalently,

$$\begin{aligned} &k_1 \left(w + v^{\pi^*}_{-\infty}(s, \pi^*_{-\infty}(s)) + D_1 \gamma_1^w v^{\pi^*}_{1,e}(s, \pi^*_{-\infty}(s)) \right) \\ &\quad + k_2 \left(w + v^{\pi^*}_{-\infty}(s, \pi^*_{-\infty}(s)) + D_2 \gamma_2^w v^{\pi^*}_{2,e}(s, \pi^*_{-\infty}(s)) \right) \\ &\quad + k_{12} \left(w + v^{\pi^*}_{-\infty}(s, \pi^*_{-\infty}(s)) + D_1 \gamma_1^w v^{\pi^*}_{1,e}(s, \pi^*_{-\infty}(s)) \right) \\ &\quad \times \left(w + v^{\pi^*}_{-\infty}(s, \pi^*_{-\infty}(s)) + D_2 \gamma_2^w v^{\pi^*}_{2,e}(s, \pi^*_{-\infty}(s)) \right) \end{aligned}$$

$$\begin{aligned} &\leq k_1 \left(w + v^{\pi^*}_{-\infty}(s, a_j) + D_1 \gamma_1^w v^{\pi^*}_{1,e}(s, a_j) \right) \\ &\quad + k_2 \left(w + v^{\pi^*}_{-\infty}(s, a_j) + D_2 \gamma_2^w v^{\pi^*}_{2,e}(s, a_j) \right) \\ &\quad + k_{12} \left(w + v^{\pi^*}_{-\infty}(s, a_j) + D_1 \gamma_1^w v^{\pi^*}_{1,e}(s, a_j) \right) \\ &\quad \times \left(w + v^{\pi^*}_{-\infty}(s, a_j) + D_2 \gamma_2^w v^{\pi^*}_{2,e}(s, a_j) \right). \end{aligned} \quad (31)$$

We can get a wealth level point $\underline{w}_{\pi^*_{-\infty}, s, a_j}$ in equality case of the above weak inequality.

Then, we can get the wealth level threshold $\underline{w}$ according to (25).

After getting $\underline{w}$, the next step is to divide further the interval $(\underline{w}, w_0]$ into subintervals and compute the optimal policy for each subinterval. The main procedure of the backward-induction algorithm for group decision-making in the case of multiplicative aggregation model is listed as follows.

Step 1. By maximizing the expected exponential utility, get the optimal policy $\pi^*_{3,e}$ of the visual agent when $w \rightarrow -\infty$.

Step 2. According to (12), we have

$$\begin{aligned} v^*(s, w) &= \max_{a \in A^*_{3,e}(s)} \left\{ k_1 \left(w + v^{\pi^*_{3,e}}_{-\infty}(s) + D_1 \gamma_1^w v^{\pi^*_{3,e}}_{1,e}(s) \right) \right. \\ &\quad + k_2 \left(w + v^{\pi^*_{3,e}}_{-\infty}(s) + D_2 \gamma_2^w v^{\pi^*_{3,e}}_{2,e}(s) \right) \\ &\quad + k_{12} \left(w + v^{\pi^*_{3,e}}_{-\infty}(s) + D_1 \gamma_1^w v^{\pi^*_{3,e}}_{1,e}(s) \right) \\ &\quad \left. \times \left(w + v^{\pi^*_{3,e}}_{-\infty}(s) + D_2 \gamma_2^w v^{\pi^*_{3,e}}_{2,e}(s) \right) \right\}, \quad (32) \end{aligned}$$

where $A^*_{3,e}(s) = \arg \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) [\gamma_3^{r(s,a,s')} \cdot v^*_{3,e}(s')]$; get the optimal value $v^*(s, w)$ and the values of $v^*_{1,e}(s)$, $v^*_{2,e}(s)$, $v^*_l(s)$ for all states $s \in s'$ when $w \rightarrow -\infty$.

Step 3. For all states $s \in s'$, $a \in A_s \setminus \pi^*_{-\infty}(s)$, get the values of $v^{\pi^*}_{-\infty}(s, a)$, $v^{\pi^*}_{1,e}(s, a)$, $v^{\pi^*}_{2,e}(s, a)$ by (7); then get $\underline{w}_{\pi^*_{-\infty}, s, a_j}$ by Expression (31).

Step 4. Calculate the wealth level threshold $\underline{w}$ according to (25).

Step 5. For the wealth level interval $(\underline{w}, w_0]$, increase further the wealth level according to the reward got by executing one step action and determine the wealth level threshold and optimal policy similarly to the above steps.

Step 6. If, for all $s \in s'$, the wealth level w is larger than w_0 , then end the algorithm.

FIGURE 1: System state transitions.

6. Numerical Examples

Consider a simple GDMDP model. There are two agents named Agent₁ and Agent₂ with risk aversion parameters γ_1 and γ_2 , respectively. The state set of the GDMDP model includes initial state s_0 and goal state g . Agent's action set is $\{a_1, a_2\}$. In state s_0 executing action a_i ($i = 1, 2$) results in a finite reward r_i and a transition to goal state g with possibility p_i . When agent reaches the goal state g it stops acting and receives no more rewards thereafter. Figure 1 shows the transitions of system states. Agents need to make an optimal policy together to reach the goal state.

Without loss of generality, the GDMDP model's parameters are assumed as follows: $p_1 = 0.4$, $p_2 = 0.005$, $r_1 = -15$, and $r_2 = -40$; the one-switch utility functions of two agents are defined as $u_1(w) = w - 0.952^w$ and $u_2(w) = w - 0.93^w$ with $\gamma_1 = 0.952$ and $\gamma_2 = 0.93$, respectively. Initial wealth level of each agent is set 0.

First, consider the situation that each agent makes decisions alone. The agent's optimal policy and the wealth level threshold are solved by utilizing the method proposed by Liu and Koenig [8]. The results are shown as follows:

$$\begin{aligned} \text{Agent}_1: \pi^*(s_0, w) &= \begin{cases} a_1 & w \in (-106.5, 0] \\ a_2 & w \in (-\infty, -106.5] \end{cases} \\ \text{Agent}_2: \pi^*(s_0, w) &= \begin{cases} a_1 & w \in (-6.9, 0] \\ a_2 & w \in (-\infty, -6.9] \end{cases} \end{aligned} \quad (33)$$

Next, we consider the group decision-making based on additive and multiplicative aggregation model of one-switch utility functions. In the case of additive aggregation model, assume each agent has equal weight; that is, $k_1 = k_2 = 0.5$; then the optimal policy and the wealth level threshold based on the proposed method in the paper are solved as follows:

$$\pi^*(s_0, w) = \begin{cases} a_1 & w \in (-16.4, 0] \\ a_2 & w \in (-\infty, -16.4] \end{cases} \quad (34)$$

The above result shows that, in the wealth level interval $(-106.5, -16.4)$, if Agent₁ makes decisions alone, the optimal policy is taking action a_1 in state s_0 but a_2 if Agent₂ makes decisions alone. If they make decisions together, then action a_2 is taken.

Now we consider how the wealth level threshold $\underline{w}$ of group decision-making changes as the weights of agents vary. In detail, k_1 changes from 0.05 to 0.95; meanwhile $k_2 = 1 - k_1$

FIGURE 2: The change of the wealth level threshold of group decision-making with different risk aversion parameters of agents.

changes from 0.95 to 0.05. The change of the wealth level threshold of group decision-making is shown in Figure 2.

Figure 2 shows that the values of the wealth level threshold $\underline{w}$ of group decision-making are near to the wealth level threshold of Agent₂ who is more risk-averse even if k_2 is small and $k_1 = k_2 - 1$ is large. This means that the influence of weights on group decision-making is not obvious if the risk aversion parameters of agents are different, while the risk aversion parameters play an important role in this situation.

Consider the situation that two agents have similar risk attitude; that is, their risk aversion parameters are similar; for example, their one-switch utility functions are $u_1(w) = w - 0.952^w$ and $u_2(w) = w - 0.95^w$, respectively. When each agent makes decisions alone their optimal policies and the wealth level thresholds are solved as follows:

$$\begin{aligned} \text{Agent}_1: \pi^*(s_0, w) &= \begin{cases} a_1 & w \in (-106.5, 0] \\ a_2 & w \in (-\infty, -106.5] \end{cases} \\ \text{Agent}_2: \pi^*(s_0, w) &= \begin{cases} a_1 & w \in (-71.5, 0] \\ a_2 & w \in (-\infty, -71.5] \end{cases} \end{aligned} \quad (35)$$

Change the values of weights in the same way as Figure 2; the result is shown in Figure 3. Difference from Figures 2 and 3 shows that the values of the wealth level threshold $\underline{w}$ of group decision-making are near to the wealth level threshold of Agent₁ when k_2 is small and $k_1 = k_2 - 1$ is large. So if the difference between the risk aversion parameters of agents is not obvious, the weights of agents will play a critical role.

Finally, we consider group decision-making based on the multiplicative aggregation model and especially focus on the influence of product term of group utility, that is, $k_{12}u_1(w)u_2(w)$, on the group decision-making. Given the same one-switch utility functions of agents in Figure 3 and assuming two sets of k_1 and k_2 value, in detail, $k_1 = 0.85$, $k_2 = 0.1$ and $k_1 = 0.1$, $k_2 = 0.85$. If the value of k_{12} is changed

FIGURE 3: The change of the wealth level threshold of group decision-making with similar risk aversion parameters of agents.

from -0.001 to -0.05 , we get two curved lines of wealth level threshold of group decision-making in Figure 4.

The two curved lines gradually approach each other to a point when the absolute value of k_{12} increases from 0.001 to 0.05 . If compared with the result of additive aggregation model, we can find that the point is very close to the wealth level threshold of additive aggregation model with $k_1 = k_2 = 0.5$. This is because when the absolute value of k_{12} is small the absolute values of $k_1 u_1(w)$ and $k_2 u_2(w)$ in group utility are larger than the absolute value of $k_{12} u_1(w) u_2(w)$, so $k_{12} u_1(w) u_2(w)$ has little influence on group decision-making. When the absolute value of k_{12} increases enough $k_{12} u_1(w) u_2(w)$ will mainly influence the group decision-making; furthermore, $k_{12} u_1(w) u_2(w)$ has same influence on Agent₁ and Agent₂; therefore the wealth level threshold of multiplicative aggregation model will approach the threshold of additive aggregation model with $k_1 = k_2 = 0.5$. This implies that the multiplicative aggregation model avoids group decision-making being dominated by the weights of individuals completely.

7. Conclusion and Future Works

This paper has put an effort on how to extend a single agent's risk-sensitive decision-theoretic planning under the MDP framework to the multiagent problem. Based on one-switch utility function that is used to describe agent's risk-sensitive attitude, the additive and multiplicative aggregation models of group utility have been proposed in this paper. According to the characteristics of group utility, a backward-induction method has been presented to divide the wealth level interval and compute the optimal policy. The paper has also offered numerical examples and discussed how the weights and risk aversion parameters influence the group decision-making.

FIGURE 4: The change of the wealth level threshold of group decision-making based on the multiplicative aggregation model.

From numerical examples we can observe that, for the additive aggregation model, if the risk aversion parameters of agents are different, the risk aversion parameters will have an obvious influence on the group decision-making, while the weights of agents will play a critical role if the risk aversion parameters are similar. For the multiplicative aggregation model, group decision-making will not be dominated by the weights of individuals completely. The product term of group utility will also influence the group decision-making.

In the future we intend to further study multiattribute group decision-making under the MDP framework with one-switch utility function. Based on the work of Tsetlin and Winkler [21], we will further study how to extend our method to the group decision-making problem.

Appendices

A. Proof of Theorem 3

Proof. Let $A_{i,e}^*(s) = \arg \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) [\gamma_i^{r(s,a,s')} \cdot v_{i,e}^*(s')]$.

$$\varepsilon = \min_{s \in S} \left(v_{i,e}^*(s) - \max_{a \in A_s \setminus A_{i,e}^*(s)} \sum_{s' \in S} P(s' | s, a) \times [\gamma_i^{r(s,a,s')} \cdot v_{i,e}^*(s')] \right). \quad (\text{A.1})$$

As $v_{i,e}^*(s) = \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) [\gamma_i^{r(s,a,s')} \cdot v_{i,e}^*(s')]$, and

$$\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_i^{-w} = k_i D_i v_{i,e}^*(s), \quad (\text{A.2})$$

thus there exists a wealth level w_0 , for wealth level $w \leq w_0$, $s \in S$,

$$k_i D_i v_{i,e}^*(s) - v^*(s, w) \gamma_i^{-w} \leq \frac{k_i D_i}{2} \varepsilon. \quad (\text{A.3})$$

On the other hand, for all wealth levels $w \leq w_0$, $s \in S$, let $f^\pi(s, w) = w + v_l^\pi(s) + k_1 D_1 \gamma_1^w v_{1,e}^\pi(s) + \dots + k_{i-1} D_{i-1} \gamma_{i-1}^w v_{i-1,e}^\pi(s) + k_{i+1} D_{i+1} \gamma_{i+1}^w v_{i+1,e}^\pi(s) + \dots + k_n D_n \gamma_n^w v_{n,e}^\pi(s)$; then $v_{i,e}^{\pi^*}(s, w) = f^{\pi^*}(s, w) + k_i D_i v_{i,e}^*(s)$.

Assume that there exists a state $s' \in S'$; $\pi'(s)$ does not belong to $A_{i,e}^*(s)$; then

$$\begin{aligned} v_{i,e}^{\pi'}(s') &= \sum_{s'' \in S} P(s'' | s', \pi'(s')) \\ &\quad \times \left[\gamma_i^{r(s', \pi'(s'), s'')} \cdot v_{i,e}^{\pi'}(s'') \right] \\ &\leq \sum_{s'' \in S} P(s'' | s', \pi_{i,e}^*(s')) \\ &\quad \times \left[\gamma_i^{r(s', \pi_{i,e}^*(s'), s'')} \cdot v_{i,e}^*(s'') \right]. \end{aligned} \quad (\text{A.4})$$

According to the fact that w , $v_l^{\pi'}(s')$, and $v_{j,e}^{\pi'}(s')$ ($j = 1, 2, \dots, n$) are all less than 0, we have

$$\begin{aligned} v^{\pi'}(s', w) &= f^{\pi'}(s, w) + k_i D_i v_{i,e}^{\pi'}(s) \leq k_i D_i v_{i,e}^{\pi'}(s) \\ &\leq \sum_{s'' \in S} P(s'' | s', \pi_{i,e}^*(s')) \\ &\quad \times \left[\gamma_i^{r(s', \pi_{i,e}^*(s'), s'')} \cdot v_{i,e}^*(s'') \right]. \end{aligned} \quad (\text{A.5})$$

Therefore,

$$\begin{aligned} k_i D_i v_{i,e}^{\pi'}(s') - v^*(s', w) \gamma_i^{-w} &\geq k_i D_i \\ &\quad \times \left(v_{i,e}^{\pi'}(s') - \sum_{s'' \in S} P(s'' | s', \pi_{i,e}^*(s')) \right. \\ &\quad \left. \times \left[\gamma_i^{r(s', \pi_{i,e}^*(s'), s'')} \cdot v_{i,e}^*(s'') \right] \right) \geq k_i D_i \varepsilon. \end{aligned} \quad (\text{A.6})$$

This is contradictory to (A.3). Thus for $w \leq w_0$, $s \in S$, optimal action $\pi^*(s, w) \in A_{i,e}^*(s)$.

Additionally, as $v_{i,e}^{\pi^*}(s) = v_{i,e}^*(s)$, and $f^{\pi^*}(s, w) \leq f^{\pi^*}(s, w)$, then

$$\begin{aligned} v_{i,e}^{\pi^*}(s, w) &= f^{\pi^*}(s, w) + k_i D_i v_{i,e}^{\pi^*}(s) \\ &\leq f^{\pi^*}(s, w) + k_i D_i v_{i,e}^*(s) = v^{\pi^*}(s, w). \end{aligned} \quad (\text{A.7})$$

Therefore, for $w \leq w_0$, $s \in S$, $v^*(s, w) = v^{\pi^*}(s, w)$. $\square$

B. Proof of Theorem 4

Proof. Let $A_{3,e}^*(s) = \arg \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) [\gamma_3^{r(s,a,s')} \cdot v_{3,e}^*(s')]$,

$$\begin{aligned} \varepsilon &= \min_{s \in S} \left(v_{3,e}^*(s) - \max_{a \in A_s \setminus A_{3,e}^*(s)} \sum_{s' \in S} P(s' | s, a) \right. \\ &\quad \left. \times \left[\gamma_3^{r(s,a,s')} \cdot v_{3,e}^*(s') \right] \right). \end{aligned} \quad (\text{B.1})$$

As $v_{3,e}^*(s) = \max_{a \in A_s} \sum_{s' \in S} P(s' | s, a) [\gamma_3^{r(s,a,s')} \cdot v_{3,e}^*(s')]$, and

$$\lim_{w \rightarrow -\infty} v^*(s, w) \gamma_3^{-w} = k_{12} D_1 D_2 v_{3,e}^*(s), \quad (\text{B.2})$$

there exists a wealth level w_0 , for all $w \leq w_0$, $s \in S$,

$$k_{12} D_1 D_2 v_{3,e}^*(s) - v^*(s, w) \gamma_3^{-w} \leq \frac{k_{12} D_1 D_2}{2} \varepsilon. \quad (\text{B.3})$$

On the other hand, for all $w \leq w_0$, $s \in S$, let

$$\begin{aligned} f^\pi(s, w) &= (w + v_l^\pi(s)) + k_1 D_1 \gamma_1^w v_{1,e}^\pi(s) \\ &\quad + k_2 D_2 \gamma_2^w v_{2,e}^\pi(s) \\ &\quad + k_{12} (w + v_l^\pi(s))^2 + k_{12} (w + v_l^\pi(s)) \\ &\quad \times (D_1 \gamma_1^w v_{1,e}^\pi(s) + D_2 \gamma_2^w v_{2,e}^\pi(s)); \end{aligned} \quad (\text{B.4})$$

then $v_{3,e}^{\pi^*}(s, w) = f^{\pi^*}(s, w) + k_{12} D_1 D_2 v_{3,e}^*(s)$.

Assume that there exists some state $s' \in S'$; $\pi'(s)$ does not belong to $A_{3,e}^*(s)$; then

$$\begin{aligned} v_{3,e}^{\pi'}(s') &= \sum_{s'' \in S} P(s'' | s', \pi'(s')) \left[\gamma_3^{r(s', \pi'(s'), s'')} \cdot v_{3,e}^{\pi'}(s'') \right] \\ &\leq \sum_{s'' \in S} P(s'' | s', \pi_{3,e}^*(s')) \\ &\quad \times \left[\gamma_3^{r(s', \pi_{3,e}^*(s'), s'')} \cdot v_{3,e}^*(s'') \right]. \end{aligned} \quad (\text{B.5})$$

According to the fact that w , $v_l^{\pi'}(s')$, and $v_{i,e}^{\pi'}(s')$ are all less than 0, we have

$$\begin{aligned} v^{\pi'}(s', w) &= f^{\pi'}(s, w) + k_{12} D_1 D_2 v_{3,e}^{\pi'}(s) \\ &\leq k_{12} D_1 D_2 v_{3,e}^{\pi'}(s) \\ &\leq \sum_{s'' \in S} P(s'' | s', \pi_{3,e}^*(s')) \\ &\quad \times \left[\gamma_3^{r(s', \pi_{3,e}^*(s'), s'')} \cdot v_{3,e}^*(s'') \right]. \end{aligned} \quad (\text{B.6})$$

Therefore,

$$\begin{aligned}
 & k_{12}D_1D_2v_{3,e}^*(s') - v^*(s', w)\gamma_3^{-w} \\
 & \geq k_{12}D_1D_2 \left(v_{3,e}^*(s') - \sum_{s'' \in S} P(s'' | s', \pi_{3,e}^*(s')) \right. \\
 & \quad \left. \times [\gamma_3^{r(s', \pi_{3,e}^*(s'), s'')} \cdot v_{3,e}^*(s'')] \right) \\
 & \geq k_{12}D_1D_2\varepsilon.
 \end{aligned} \tag{B.7}$$

This is contradictory to (B.3). So for all $w \leq w_0$, $s \in S$, optimal policy $\pi^*(s, w) \in A_{3,e}^*(s)$.

Additionally, as $v_{3,e}^{\pi_{3,e}^*}(s) = v_{3,e}^*(s)$, and $f^{\pi_{3,e}^*}(s, w) \leq f^{\pi^*}(s, w)$, so

$$\begin{aligned}
 v^{\pi_{3,e}^*}(s, w) &= f^{\pi_{3,e}^*}(s, w) + k_{12}D_1D_2v_{3,e}^{\pi_{3,e}^*}(s) \\
 &\leq f^{\pi^*}(s, w) + k_{12}D_1D_2v_{3,e}^*(s) = v^{\pi^*}(s, w).
 \end{aligned} \tag{B.8}$$

Therefore, for all $w \leq w_0$, $s \in S$, $v^*(s, w) = v^{\pi^*}(s, w)$. $\square$

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 70971048.

References

- [1] C. Boutilier, T. Dean, and S. Hanks, "Decision-theoretic planning: structural assumptions and computational leverage," *Journal of Artificial Intelligence Research*, vol. 11, pp. 1–94, 1999.
- [2] D. E. Bell, "One-switch utility functions and a measure of risk," *Management Science*, vol. 34, no. 12, pp. 1416–1424, 1988.
- [3] G. M. Gelles and D. W. Mitchell, "Broadly decreasing risk aversion," *Management Science*, vol. 45, no. 10, pp. 1432–1439, 1999.
- [4] Y. Nakamura, "Sumex utility functions," *Mathematical Social Sciences*, vol. 31, no. 1, pp. 39–47, 1996.
- [5] D. E. Bell and P. C. Fishburn, "Strong one-switch utility," *Management Science*, vol. 47, no. 4, pp. 601–604, 2001.
- [6] Y. Liu, *Decision-theoretic planning under risk-sensitive planning objectives [Ph.D. thesis]*, College of Computing, Georgia Institute of Technology, 2005.
- [7] Y. Liu and S. Koenig, "Risk-sensitive planning with one-switch utility functions: value iteration," in *Proceedings of the 20th National Conference on Artificial Intelligence*, pp. 993–999, 2005.
- [8] Y. Liu and S. Koenig, "An exact algorithm for solving MDPs under risk-sensitive planning objectives with one-switch utility functions," in *Proceedings of the 7th International Conference on Autonomous Agents and Multi-Agent Systems*, 2008.
- [9] R. A. Howard and J. E. Matheson, "Risk-sensitive Markov decision processes," *Management Science*, vol. 18, no. 7, pp. 356–369, 1972.
- [10] S. D. Patek, "On terminating Markov decision processes with a risk-averse objective function," *Automatica*, vol. 37, no. 9, pp. 1379–1386, 2001.
- [11] D. Hernandez-Hernández and S. I. Marcus, "Risk sensitive control of Markov processes in countable state space," *Systems and Control Letters*, vol. 29, no. 3, pp. 147–155, 1996.
- [12] Y. Le and L. Tallec, *Robust, risk-sensitive, and data-driven control of Markov decision processes [Ph.D dissertation]*, Sloan School of Management, Massachusetts Institute of Technology, 2007.
- [13] P. H. Iz, "Two multiple criteria group decision support systems based on mathematical programming and ranking methods," *European Journal of Operational Research*, vol. 61, no. 1-2, pp. 245–253, 1992.
- [14] S. J. Chen and C. L. Hwang, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, vol. 375 of *Lecture Notes in Economics and Mathematical Systems*, Springer, New York, NY, USA, 1992.
- [15] R. J. Li, "Fuzzy method in group decision making," *Computers & Mathematics with Applications*, vol. 38, no. 1, pp. 91–101, 1999.
- [16] R. Benayoun, J. de Montgolfier, and J. Tergny, "Linear programming with multiple objective functions: step method (stem)," *Mathematical Programming*, vol. 1, no. 1, pp. 366–375, 1971.
- [17] A. M. Geoffrion, J. S. Dyer, and A. Feinberg, "An interactive approach for multi-criterion optimization, with an application to the operation of academic department," *Management Science*, vol. 19, no. 4, pp. 357–368, 1972.
- [18] Z. Xu, "Approaches to multi-stage multi-attribute group decision making," *International Journal of Information Technology and Decision Making*, vol. 10, no. 1, pp. 121–146, 2011.
- [19] D. P. Bertsekas and J. N. Tsitsiklis, "An analysis of stochastic shortest path problems," *Mathematics of Operations Research*, vol. 16, no. 3, pp. 580–595, 1991.
- [20] S. D. Patek, "On terminating Markov decision processes with a risk-averse objective function," *Automatica*, vol. 37, no. 9, pp. 1379–1386, 2001.
- [21] I. Tsetlin and R. L. Winkler, "Multiattribute one-switch utility," *Management Science*, vol. 58, no. 3, pp. 602–605, 2012.