CONTINUOUS-TIME ZERO-SUM GAMES FOR MARKOV DECISION PROCESSES WITH RISK-SENSITIVE FINITE-HORIZON COST CRITERION ON A GENERAL STATE SPACE

Subrata Golui

(M.Sc. in Mathematics, Indian Institute of Engineering Science and Technology Shibpur). Senior Research Scholar, Department of Mathematics, Indian Institute of Technology Guwahati. Guwahati (India).

golui@iitg.ac.in:

https://orcid.org/0000-0001-7232-562X:

Chandan Pal

(PhD in Mathematics, Indian Institute of Technology Bombay). Assistant Professor, Department of Mathematics, Indian Institute of Technology Guwahati. Guwahati (India).

cpal@iitg.ac.in:

https://orcid.org/0000-0002-4684-0481:

Reception: 27/08/2022 **Acceptance:** 11/09/2022 **Publication:** 29/12/2022

Suggested citation:

Subrata Golui and Chanda Pal (2022). Continuous-time zero-sum games for Markov decision processes with risk-sensitive finite-horizon cost criterio on a general state space. *3C Empresa. Investigación y pensamiento crítico*, *11*(2), 76-92. https://doi.org/10.17993/3cemp.2022.110250.76-92

ABSTRACT

In this manuscript, we study continuous-time risk-sensitive finite-horizon time-homogeneous zero-sum dynamic games for controlled Markov decision processes (MDP) on a Borel space. Here, the transition and payoff functions are extended real-valued functions. We prove the existence of the game's value and the uniqueness of the solution of Shapley equation under some reasonable assumptions. Moreover, all possible saddle-point equilibria are completely characterized in the class of all admissible feedback multi-strategies. We also provide an example to support our assumptions.

KEYWORDS

Zero-sum stochastic game, Borel state space, risk-sensitive utility, finite-horizon cost criterion, optimality equation, saddle-point

1 INTRODUCTION

In the literature of game theory, there are two types of game models: a zero-sum model and a nonzero-sum game model. We know that in the zero-sum two-person game, one player tries to maximize his/her payoff and another tries to minimize his/her payoff, whereas in the nonzero-sum game, both players try to minimize their payoff. We can study game theory either in discrete-time or in continuous-time. In continuous time, the players observe the state space continuously, whereas in discrete-time, they observe the state space in discrete-time. Also, there are two types of game models with respect to the risk measure; one is a risk-neutral game, and another is a risk-sensitive game. Risk-sensitive, or "exponential of integral utility" cost criterion is popular, particularly in finance (see, e.g., Bielecki and Pliska (1999)) since it has the property to capture the effects of more than first order (expectation) moments of the cost.

There are large number of literatures for the risk-neutral utility cost criterion for continuous-time controlled Markov decision processes (CTCMDPs) with different setup, see Guo (2007), Guo and Hernandez-Lerma (2009), Guo et al. (2015), Guo et al. (2012), Guo and Piunovskiy (2011), Huang (2018), Piunovskiy and Zhang (2011), Piunovskiy and Zhang (2014) for single controller model and Guo and Hernandez-Lerma (2003), Guo and Hernandez-Lerma (2005), Guo and Hernandez-Lerma (2007), wei and Chen (2016), Zhang and Guo (2012) for game models. Players want to ignore risk in risk-neutral stochastic games because of the additive feature of this criterion. If the variance is high, the risk-neutral criterion is not useful since there can be issues with optimal control. Regarding risk preferences, different controllers may exhibit various perspectives. So, risk preferences are considered by the decision-makers to be the performance criterion. Bell (1995) gave a model containing the interpretation of risk-sensitive utility. This paper considers finite-horizon risk-sensitive two-person zero-sum dynamic games for controlled CTMDPs with unbounded rates (transition and payoff rates) under admissible feedback strategies. State and action spaces are considered to be Borel spaces. The main target of this manuscript is to find the solution of the optimality equation (6) (Shapley equation), to provide the proof of the existence of game's value, and to give a proof of complete characterization of saddle-point equilibrium.

The finite-horizon optimality criterion generally comes up in real-life scenarios. where the cost criterion may not be risk-neutral. For finite-horizon risk-neutral CTMDPs, see Guo et al. (2015), Huang (2018) while for the corresponding game, see Wei and Chen (2016) and its references. In this context, for risk-sensitive finite-horizon controlled CTMDP, one can see Ghosh and Saha (2014), Guo et al. (2019), Wei (2016), while the research for infinite-horizon risk-sensitive CTMDP are available in, Ghosh and Saha (2014), Golui and Pal (2022), Guo and Zhang (2018), Kumar and Pal (2013), Kumar and Pal (2015), Zhang (2017) and the references therein. At the same time the corresponding finite/infinite-horizon dynamic games are studied in Ghosh et at. (2022), Golui and Pal (2021a), Golui and Pal (2021b), Golui et al. (2022), wei (2019). Study on CTMDPs for risk-sensitive control on a denumerable state space are available greatly, see Ghosh and saha (2014), Guo and Liao (2019), Guo et al. (2019) but some times we see the countable state space dose not help to study some models specially in chemical reactions problem, water reservoir management problem, inventory problem, cash-flow problem, insurance problem etc. We see that the literature in controlled CTMDPs considering on general state space is very narrow. Some exceptions for single controller are Golui and Pal (2022), Guo et al. (2012), Guo and Zhang (2019), Pal and Pradhan (2019), Piunovskiy and Zhang (2014), Piunovskiy and Zhang (2020) and for corresponding stochastic games are Bauerle and Rieder (2017), Golui and Pal (2021b), Guo and Hernandez-Lerma (2007), Wei (2017). So, it is very interesting and very important to consider the game problem in some general state space. In Guo and Zhang (2019), the authors studied the same as in Guo et al. (2019) but on general state space, whereas in Wei (2017), the finite-horizon risk-sensitive zero-sum game for a controlled Markov jump process with bounded costs and unbounded transition rates was studied. Where in Ghosh et al. (2016), the authors studied dynamic games on the infinite-horizon for controlled CTMDP by considering bounded transition and payoff rates. However this boundedness condition is a restrictive conditon for many real life scenarios. Someone may note queuing and population processes for the requirement of unboundedness in transition and payoff functions. In Golui and Pal (2021a), finite-horizon continuous-time risk-sensitive zero-sum games for

https://doi.org/10.17993/3cemp.2022.110250.76-92

unbounded transition and payoff function on countable state space is considered. But the extension of the same results to a general Borel state space were unknown to us. We solve this problem in this paper. Here we are dealing with finite-horizon risk-sensitive dynamic games employing the unbounded payoff and transition rates in the class of all admissible feedback strategies on some general Borel state space, whose results were unknown until now. In this paper, we try to find the solution to the risk-sensitive finite-horizon optimality equation and, at the same time, try to obtain the existence of an optimal equilibrium point for this jump process. We take homogeneous game model. In Theorem 4, we prove our final results, i.e., we show that if the cost rates are real-valued functions, then the Shapley equation (6), has a solution. The existence of optimal-point equilibria is proved by using the measurable selection theorem in Nowak (1985). The claim of uniqueness of the solution is due to the well known Feynman-Kac formula. The value of the game has also been established.

The remaining portions of this work are presented. Section 2 describes the model of our stochastic game, some definitions, and the finite-horizon cost criterion. In Section 3, preliminary results, conditions, and the extension of the Feynman-Kac formula are provided. Also, we establish the probabilistic representation of the solution of the finite horizon optimality equation (6) there. The uniqueness of this optimal solution as well as the game's value are proved in section 4. We also completely characterize the Nash equilibrium among the class of admissible Markov strategies for this game model here. In Section 5, we verify our results with an example.

2 THE ZERO-SUM DYNAMIC GAME MODEL

First, we introduce a time-homogeneous continuous-time zero-sum dynamic game model in this section, which contains the following:

$$\mathcal{G} := \{\boldsymbol{X}, U, V, (U(x) \subset U, x \in \boldsymbol{X}), (V(x) \subset V, x \in \boldsymbol{X}), q(\cdot|x,u,v), c(x,u,v), g(x)\}. \tag{1}$$

Here $\boldsymbol{X}$ is our state space which is a Borel space and the corresponding Borel σ-algebra is $\mathcal{B}(\boldsymbol{X})$. The action spaces are U and V for first and second players, respectively, and are considered to be Borel spaces. Their corresponding Borel σ-algebras are, respectively, $\mathcal{B}(U)$ and $\mathcal{B}(V)$. For each $x \in \boldsymbol{X}$, the admissible action spaces are denoted by $U(x) \in \mathcal{B}(U)$ and $V(x) \in \mathcal{B}(V)$, respectively and these spaces are assumed to be compact. Now let us define a Borel subset of $\boldsymbol{X} \times U \times V$ denoted by $\mathcal{K} := \{(x,u,v) | x \in \boldsymbol{X}, u \in U(x), v \in V(x)\}$.

Next, for any $(x,u,v) \in \mathcal{K}$, we know that the transition rate of the CTMDPs denoted by $q(\cdot|x,u,v)$ is a signed kernel on $\boldsymbol{X}$ such that $q(D|x,u,v) \geq 0$ where $(x,u,v) \in \mathcal{K}$ and $x \notin D$. Also, $q(\cdot|x,u,v)$ is assumed to be conservative i.e., $q(\boldsymbol{X}|x,u,v) \equiv 0$, as well as stable i.e.,

$$q^*(x) := \sup_{u \in U(x), v \in V(x)} [q_x(u,v)] < \infty \quad \forall x \in \boldsymbol{X}, \tag{2}$$

$q_x(u,v) := -q(\{x\}|x,u,v) \geq 0$ for all $(x,u,v) \in \mathcal{K}$. Our running cost is c, assumed to be measurable on $\mathcal{K}$ and the terminal cost is g, assumed to be measurable on $\boldsymbol{X}$. These costs are taken to be real-valued.

The dynamic game is played as following. The players take actions continuously. At time moment $t \geq 0$, if the system's state is $x \in S$, the players take their own actions $u_t \in U(x)$ and $v_t \in V(x)$ independently as their corresponding strategies. As a results the following events occurs:

- the first player gets an reward at rate $c(x, u_t, v_t)$ immediately and second player gives a cost at a rate $c(x, u_t, v_t)$; and
- staying for a random time in state x, the system leaves the state x at a rate given by the quantity $q_x(u_t, v_t)$, and it jumps to a set D, $(x \notin D)$ with some probability determined by $\dfrac{q(D|x, u_t, v_t)}{q_x(u_t, v_t)}$ (for details, see Proposition B.8 in Guo and Hernandez-Lerma (2009), p. 205 for details).

Now suppose the system is at a new state y. Then the above operation is replicated till the fixed time $\hat{T} > 0$. Moreover, at time $\hat{T}$ if the system occupies a state $y_{\hat{T}}$, second player pays a terminal cost $g(y_{\hat{T}})$ to first player.

 https://doi.org/10.17993/3cemp.2022.110250.76-92

Consequently, first player always tries to maximize his/her payoff, whereas second player wants to minimize his/her payoff according to some cost measurement criterion $\mathcal{H}^{\cdot,\cdot}(\cdot,\cdot)$, that is presented below by equation (4). Next the construction of the CTMDPs will be presented under possibly pair of admissible feedback strategies. For construction of the corresponding CTMDPs (as in Kitaev (1986), Piunovskiy and Zhang (2011)), we imopse some usefull notations: define $\boldsymbol{X}(\Delta) := \boldsymbol{X} \cup \{\Delta\}$ (for some $\Delta \notin \boldsymbol{X}$),
$\Omega^0 := (\boldsymbol{X} \times (0,\infty))^\infty$, $\Omega := \Omega^0 \cup \{(x_0,\theta_1,x_1,\cdots,\theta_{\hat{k}},x_{\hat{k}},\infty,\Delta,\infty,\Delta,\cdots)|x_0 \in \boldsymbol{X}, x_l \in \boldsymbol{X}, \theta_l \in (0,\infty)$, for each $1 \leq l \leq \hat{k}, \hat{k} \geq 1\}$, and suppose $\mathcal{F}$ be the corresponding Borel σ-algebra on Ω. Then we get a Borel measurable space $(\Omega,\mathcal{F})$. For each $\hat{k} \geq 0$, $\omega := (x_0,\theta_1,x_1,\cdots,\theta_{\hat{k}},x_{\hat{k}},\cdots) \in \Omega$, let us define $T_0(\omega) := 0$, $T_{\hat{k}}(\omega) - T_{\hat{k}-1}(\omega) := \theta_{\hat{k}}$, $T_\infty(\omega) := \lim_{\hat{k}\to\infty} T_{\hat{k}}(\omega)$. Now in view of the definition of $\{T_{\hat{k}}\}$, we define the state process $\{\xi_t\}_{t\geq 0}$ defined by

$$\xi_t(\omega) := \sum_{\hat{k}\geq 0} I_{\{T_{\hat{k}} \leq t < T_{\hat{k}+1}\}} x_{\hat{k}} + I_{\{t \geq T_\infty\}}\Delta,\ t \geq 0. \tag{3}$$

Here I_E is the standard notation for indicator function corresponding to a set E, and we use $0 + z =: z$ and $0z =: 0$ for any $z \in \boldsymbol{X}(\Delta)$ as convention. The process after the time T_∞ is treated for absorbtion in the state Δ. Hence, let us define $q(\cdot|\Delta,u_\Delta,v_\Delta) :\equiv 0$, $U(\Delta) := U \cup U_\Delta$, $V(\Delta) := V \cup V_\Delta$, $U_\Delta := \{u_\Delta\}$, $V_\Delta := \{v_\Delta\}$, $c(\Delta,u,v) :\equiv 0$ for all $(u,v) \in U(\Delta) \times V(\Delta)$, u_Δ, v_Δ are treated as isolated points. Furthermore, define $\mathcal{F}_t := \sigma(\{T_{\hat{k}} \leq s, \xi_{T_{\hat{k}}} \in D\} : D \in \mathcal{B}(\boldsymbol{X}), 0 \leq s \leq t, \hat{k} \geq 0)\ \forall t \in \mathbb{R}_+$, and $\mathcal{F}_{s-} =: \bigvee_{0\leq t<s} \mathcal{F}_t$. Lastly the σ-algebra of predictable sets on $\Omega \times [0,\infty)$ corresponding to $\{\mathcal{F}_t\}_{t\geq 0}$ is denoted by $\mathcal{P} := \sigma(\{U \times \{0\}, U \in \mathcal{F}_0\} \cup \{V \times (s,\infty), V \in \mathcal{F}_{s-}\})$. Now we intoduce strategies of players to define the risk sensitive cost criterion:

Definition 1. *An admissible feedback strategy for player 1, denoted by $\zeta^1 = \{\zeta^1_t\}_{t\geq 0}$, is defined to be a transition probability $\zeta^1(du|\omega,t)$ from $(\Omega \times [0,\infty), \mathcal{P})$ onto $(U_\Delta, \mathcal{B}(U_\Delta))$, for which $\zeta^1(U(\xi_{t-}(\omega))|\omega,t) = 1$.*

For more informations, one can see [Guo and Song (2011), Definition 2.1, Remark 2.2], Piunovskiy and Zhang (2011), Zhang (2017).

Let Π^1_{Ad} denote the set of all admissible feedback strategies for player 1. A strategy $\zeta^1 \in \Pi^1_{Ad}$ for player 1, is said to be Markov if for every $\omega \in \Omega$ and $t \geq 0$ the relation $\zeta^1(du|\omega,t) = \zeta^1(du|\xi_{t-}(\omega),t)$ holds, $\lim_{s\uparrow t} \xi_s(\omega) := \xi_{t-}(w)$. We call a Markov strategy $\{\zeta^1_t\}$ as a stationary Markov for player 1, if it not explicitly dependent on time t. The family of all Markov strategies and all stationary strategies are denoted by Π^1_M and Π^1_{SM}, respectively, for first player. The sets Π^2_{Ad}, Π^2_M, Π^2_{SM} stand for all admissible feedback strategies, all Markov strategies, and all stationary strategies, respectively, for second player are defined similarly. In view of Assumption 1, below, for any initial distribution γ on $\boldsymbol{X}$ and any multi-strategy $(\zeta^1,\zeta^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, in view of Theorem 4.27 in Kitaev and Rykov (1985) a unique probability measure exists and denoted by $P^{\zeta^1,\zeta^2}_\gamma$ (depending on γ and (ζ^1,ζ^2)) on $(\Omega,\mathcal{F})$ for which $P^{\zeta^1,\zeta^2}_\gamma(\xi_0 = x) = 1$. Let us define the corresponding expectation operator as $E^{\zeta^1,\zeta^2}_\gamma$. Particularly, when γ represents the Dirac measure at a state $x \in \boldsymbol{X}$, $P^{\zeta^1,\zeta^2}_\gamma$ and $E^{\zeta^1,\zeta^2}_\gamma$ will be written as P^{ζ^1,ζ^2}_x and E^{ζ^1,ζ^2}_x, respectively. For any compact metric space Y, the space of probability measures on Y is denoted by $\mathcal{P}(Y)$ with Prohorov topology. As $U(x)$ and $V(x)$ are compact sets for each $x \in \boldsymbol{X}$, $\mathcal{P}(U(x))$ and $\mathcal{P}(V(x))$ are also compact and convex metric spaces. Now for each fixed $x \in \boldsymbol{X}$, $\vartheta \in \mathcal{P}(U(x))$ and $\eta \in \mathcal{P}(V(x))$, the corresponding transition and payoff rates are defined, as below:

$$q(D|x,\vartheta,\eta) := \int_{V(x)} \int_{U(x)} q(D|x,u,v)\vartheta(du)\eta(dv),\ D \subseteq \boldsymbol{X}.$$

$$c(x,\vartheta,\eta) := \int_{V(x)} \int_{U(x)} c(x,u,v)\vartheta(du)\eta(dv),$$

Note that $\zeta^1 \in \Pi^1_{SM}$ can be identified by a mapping $\zeta^1 : \boldsymbol{X} \to \mathcal{P}(U)$ for which $\zeta^1(\cdot|x) \in \mathcal{P}(U(x))$ for each $x \in \boldsymbol{X}$. So, we can write $\Pi^1_{SM} = \Pi_{x\in S}\mathcal{P}(U(x))$ and $\Pi^2_{SM} = \Pi_{x\in\boldsymbol{X}}\mathcal{P}(V(x))$. So, the sets Π^1_{SM} and Π^2_{SM} are compact metric spaces by using Tychonoff theorem.

 https://doi.org/10.17993/3cemp.2022.110250.76-92

Next take $\lambda \in (0,1]$ as a fixed risk-sensitivity coefficient and fix a finite time horizon $\hat{T} > 0$. Then for each $x \in \boldsymbol{X}$, $t \in [0,\hat{T}]$ and $(\zeta^1,\zeta^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, define the risk-sensitive finite-horizon ($\hat{T}$-horizon) cost criterion as

$$\mathcal{H}^{\zeta^1,\zeta^2}(0,x) := E_x^{\zeta^1,\zeta^2}\left[e^{\lambda \int_0^{\hat{T}} \int_V \int_U c(\xi_t,u,v)\zeta^1(da|\omega,t)\zeta^2(dv|\omega,t)dt + \lambda g(\xi_{\hat{T}})}\right], \tag{4}$$

whence it is given that the integral is well defined. For each $(\zeta^1,\zeta^2) \in \Pi^1_M \times \Pi^2_M$, we know that $\{\xi_t, \geq 0\}$ is a controlled Markov Process on $(\Omega, \mathcal{F}, P_\gamma^{\zeta^1,\zeta^2})$, and hence for any γ (initial distribution on $\boldsymbol{X}$), for each $x \in \boldsymbol{X}$, $t \in [0,\hat{T}]$,

$$\mathcal{H}^{\zeta^1,\zeta^2}(t,x) := E_\gamma^{\zeta^1,\zeta^2}\left[e^{\lambda \int_t^{\hat{T}} \int_V \int_U c(\xi_t,u,v)\zeta^1(du|\xi_t,t)\zeta^2(dv|\xi_t,t)dt + \lambda g(\xi_{\hat{T}})}|\xi_t = x\right], \tag{5}$$

is well defined.
We define the lower value of the game on $\boldsymbol{X}$ as $\mathcal{L}(x) := \sup_{\zeta^2 \in \Pi^2_{Ad}} \inf_{\zeta^1 \in \Pi^1_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(0,x)$.
Similarly, define the upper value of the game on $\boldsymbol{X}$ as $\mathcal{U}(x) := \inf_{\zeta^1 \in \Pi^1_{Ad}} \sup_{\zeta^2 \in \Pi^2_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(0,x)$.
It is easy to see that

$$\mathcal{L}(x) \leq \mathcal{U}(x) \text{ for each } x \in \boldsymbol{X}.$$

If $\mathcal{L}(x) = \mathcal{U}(x)$, $\forall x \in \boldsymbol{X}$, define $\mathcal{L}(\cdot) \equiv \mathcal{U}(\cdot) :\equiv \mathcal{H}^*(\cdot)$, and then the function $\mathcal{H}^*(x)$ is called the value of the game. Also, if $\sup_{\zeta^2 \in \Pi^2_M} \inf_{\zeta^1 \in \Pi^1_M} \mathcal{H}^{\zeta^1,\zeta^2}(t,x) = \inf_{\zeta^1 \in \Pi^1_M} \sup_{\zeta^2 \in \Pi^2_M} \mathcal{H}^{\zeta^1,\zeta^2}(t,x)$, $\forall (t,x) \in [0,\hat{T}] \times \boldsymbol{X}$, the common function is denoted by $\mathcal{H}^*(\cdot,\cdot)$.
A strategy $\zeta^{*1} \in \Pi^1_{Ad}$ is called optimal for first player if

$$\mathcal{H}^{\zeta^{*1},\zeta^2}(x,c) \leq \sup_{\zeta^2 \in \Pi^2_{Ad}} \inf_{\zeta^1 \in \Pi^1_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(x) = \mathcal{L}(x) \ \forall \ x \in \boldsymbol{X}, \ \forall \zeta^2 \in \Pi^2_{Ad}.$$

Similarly, for second player, the strategy $\zeta^{*2} \in \Pi^2_{Ad}$ is optimal if

$$\mathcal{H}^{\zeta^1,\zeta^{*2}}(x,c) \geq \inf_{\zeta^1 \in \Pi^1_{Ad}} \sup_{\pi^2 \in \Pi^2_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(x) = \mathcal{U}(x) \ \forall \ x \in \boldsymbol{X}, \ \forall \zeta^1 \in \Pi^1_{Ad}.$$

If for k^{th} player, (k=1,2), $\zeta^{*k} \in \Pi^k_{Ad}$ is optimal, then (ζ^{*1},ζ^{*2}) is said to be a pair of optimal strategies. Now for the pair of strategies (ζ^{*1},ζ^{*2}) if

$$\mathcal{H}^{\zeta^{*1},\zeta^2}(x,c) \leq \mathcal{H}^{\zeta^{*1},\zeta^{*2}}(x,c) \leq \mathcal{H}^{\zeta^1,\zeta^{*2}}(x,c), \ \ \forall \zeta^1 \in \Pi^1_{Ad}, \ \forall \zeta^2 \in \Pi^2_{Ad},$$

then (ζ^{*1},ζ^{*2}) is said to a saddle-point equilibrium, and then the strategies ζ^{*1} and ζ^{*2} are optimal strategies corresponding to first player and second player, respectively.

3 PRELIMS

For proving the existence of an optimal pair of strategies, we recall some statndard results for the risk-sensitive finite time horizon CTMDPs. Due to the unboundedness of the rates $q(dy|x,u,v)$ and $c(x,u,v)$, we impose some conditions to make the processes $\{\xi_t, t \geq 0\}$ nonexplosive, and to make $\mathcal{H}^{\pi^1,\pi^2}(0,x)$ finite, which were used greatly in CTMDPs; see, Golui and Pal (2021a), Guo and Liao (2019), Guo et al. (2019), Guo and Zhang (2019) and references therein. For bounded rates, following Assumption 1 (ii)-(iii) are not required, see Ghosh and Saha (2014), Kumar and Pal (2015).

Assumption 1. *There exists a function $\mathcal{W} : \boldsymbol{X} \to [1,\infty)$ for which the followings hold:*

(i) The relation $\int_S \mathcal{W}(y)q(dy|x,u,v) \leq \rho_1 \mathcal{W}(x) + b_1$ holds, for each $(x,u,v) \in \mathcal{K}$, for some constants $\rho_1 > 0$, $b_1 \geq 0$;

 https://doi.org/10.17993/3cemp.2022.110250.76-92

(ii) $q^*(x) \le M_1 \mathcal{W}(x)$, $\forall x \in \boldsymbol{X}$, *for some nonnegative constant* $M_1 \ge 1$, *where* $q^*(x)$ *is as in (2.2);*

(iii) $e^{2(\hat{T}+1)\lambda|c(x,u,v)|} \le M_2 \mathcal{W}(x)$ *for any* $(x,u,v) \in \mathcal{K}$, *and* $e^{2(\hat{T}+1)\lambda|g(x)|} \le M_2 \mathcal{W}(x)$ *for each* $x \in \boldsymbol{X}$, *for some constant* $M_2 \ge 1$.

The non-exlposion of the state process $\{\xi_t, t \ge 0\}$ and the finiteness of $\mathcal{H}^{\zeta^1,\zeta^2}(0,x)$ is shown in following Lemma. Here we see that the function $\mathcal{H}^{\zeta^1,\zeta^2}(0,x)$ has upper and lower bound in terms of the function $\mathcal{W}$.

Lemma 1. *We grant Assumption 1. Then for each* $(\zeta^1,\zeta^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, *we obtain the following results.*

(a) $P_x^{\zeta^1,\zeta^2}(T_\infty = \infty) = 1$, $P_x^{\zeta^1,\zeta^2}(\xi_t \in \boldsymbol{X}) = 1$, *and* $P_x^{\zeta^1,\zeta^2}(\xi_0 = x) = 1$ *for each* $t \ge 0$ *and* $x \in \boldsymbol{X}$.

(b) (b_1) $e^{-L_1\mathcal{W}(x)} \le \mathcal{H}^{\zeta^1,\zeta^2}(0,x) \le L_1 \mathcal{W}(x)$ *for* $x \in \boldsymbol{X}$ *and* $(\zeta^1,\zeta^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, *where* $L_1 := M_2 e^{\rho_1 \hat{T}}\left[1 + \frac{b_1}{\rho_1}\right]$.

(b_2) $e^{-L_1\mathcal{W}(x)} \le \mathcal{H}^{\zeta^1,\zeta^2}(t,x) \le L_1 \mathcal{W}(x)$ *for* $(t,x) \in [0,\hat{T}] \times \boldsymbol{X}$ *and* $(\zeta^1,\zeta^2) \in \Pi^1_M \times \Pi^2_M$.

Proof. These results can be proved by using Guo et al. (2019), Lemma 3.1 and Guo and Zhang (2019), Lemma 3.1.

In order to apply the extended Feynman-Kac formula, we impose the following assumption for unbounded functions. If the rates are bounded, the following Assumption is not required, see Ghosh and Saha (2014). Since we are dealing with unbounded rates, we require the following condition.

Assumption 2. *There exists* $[1,\infty)$*-valued function* $\mathcal{W}_1$ *on* $\boldsymbol{X}$ *such that*

(i) $\int_{\boldsymbol{X}} \mathcal{W}_1^2(y) q(dy|x,u,v) \le \rho_2 \mathcal{W}_1^2(x) + b_2$, *for each* $(x,u,v) \in \mathcal{K}$ *for some constants* $\rho_2 > 0$ *and* $b_2 > 0$;

(ii) $\mathcal{W}^2(x) \le M_3 \mathcal{W}_1(x)$, $\forall x \in \boldsymbol{X}$, *for some constant* $M_3 \ge 1$, *where the function* $\mathcal{W}$ *is as in Assumption 1.*

In addition of Assumptions 1, 2, we impose the following conditions to guarantee the existence of a pair of optimal strategies.

Assumption 3. *(i) The cost and transition rate functions,* $c(x,u,v)$ *and* $q(\cdot|x,u,v)$ *are continuous on* $U(x) \times V(x)$, *for each* $x \in \boldsymbol{X}$.

(ii) The integral functions $\displaystyle\int_{\boldsymbol{X}} f(y) q(dy|x,u,v)$ *and* $\displaystyle\int_{\boldsymbol{X}} \mathcal{W}(y) q(dy|x,u,v)$ *are continuous on* $U(x) \times V(x)$, *for each* $x \in \boldsymbol{X}$, *for all bounded measurable functions* f *on* $\boldsymbol{X}$ *and* $\mathcal{W}$ *as previous in Assumption 1.*

We next introduce some useful notations. Let $\mathcal{A}_c(\Omega \times [0,\hat{T}] \times \boldsymbol{X})$ denote the space of all real-valued, $\mathcal{P} \times \mathcal{B}(\boldsymbol{X})$-measurable functions $\varphi(\omega,t,x)$ which are differentiable in $t \in [0,\hat{T}]$ a.e. i.e., $\mathcal{A}_c(\Omega \times [0,\hat{T}] \times \boldsymbol{X})$ contains the said measurable functions φ with the following facets: Given any $x \in \boldsymbol{X}$, $(\zeta^1,\zeta^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$, and a.s. $\omega \in \Omega$, there exists a $\mathcal{E}_{(\varphi,\omega,x,\zeta^1,\zeta^2)} \subseteq [0,\hat{T}]$ (a Borel subset of $[0,\hat{T}]$ that depends on $\varphi,\omega,x,\zeta^1,\zeta^2)$ such that $\frac{\partial\varphi}{\partial t}$ (the partial derivative with respect to time $t \in [0,\hat{T}]$) exists for every $t \in \mathcal{E}_{(\varphi,\omega,x,\zeta^1,\zeta^2)}$ and $m_L(\mathcal{E}^c_{(\varphi,\omega,x,\zeta^1,\zeta^2)}) = 0$, where m_L is the Lebesgue measure on $\mathbb{R}$. Now if for some $(\omega,t,x) \in \Omega \times [0,\hat{T}] \times \boldsymbol{X}$, $\frac{\partial\varphi}{\partial t}(\omega,t,x)$ does not exists, we take this as any real number, and so $\frac{\partial\varphi}{\partial t}(\cdot,\cdot,\cdot)$ can be made definable on $\Omega \times [0,\hat{T}] \times \boldsymbol{X}$. For any given function $W \ge 1$ on $\boldsymbol{X}$, a function f (real-valued) on $\Omega \times [0,\hat{T}] \times \boldsymbol{X}$ is said to be a W-bounded if $\|f\|^\infty_W := \sup_{(\omega,t,x)\in\Omega\times[0,\hat{T}]\times\boldsymbol{X}} \frac{|f(\omega,t,x)|}{W(x)} < \infty$. The W-bounded Banach space is denoted by $\mathcal{B}_W(\Omega \times [0,\hat{T}] \times \boldsymbol{X})$. Note that if $W \equiv 1$, $\mathcal{B}_1(\Omega \times [0,\hat{T}] \times \boldsymbol{X})$ is the space of all bounded functions on $\Omega \times [0,\hat{T}] \times \boldsymbol{X}$.

Now define $\mathcal{C}^1_{W_0,W_1}(\Omega \times [0,\hat{T}] \times \boldsymbol{X}) := \{\psi \in \mathcal{B}_{W_0}(\Omega \times [0,\hat{T}] \times \boldsymbol{X}) \cap \mathcal{A}_c(\Omega \times [0,\hat{T}] \times \boldsymbol{X}) : \frac{\partial \psi}{\partial t} \in \mathcal{B}_{W_1}(\Omega \times [0,\hat{T}] \times \boldsymbol{X})\}$. If any function $\psi(\omega,t,x) \in \mathcal{C}^1_{W_0,W_1}(\Omega \times [0,\hat{T}] \times \boldsymbol{X})$ does not depend on ω, we write it as $\psi(t,x)$ and the corresponding space is $\mathcal{C}^1_{W_0,W_1}([0,\hat{T}] \times \boldsymbol{X})$.
In the the next theorem, we state the extended Feynman-Kac formula, which is very useful for us.

Theorem 1. *We grant Assumptions 1 and 2.*

(a) Then, for each $x \in \boldsymbol{X}$, $(\zeta^1,\zeta^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ *and* $\psi \in \mathcal{C}^1_{\mathcal{W},\mathcal{W}_1}(\Omega \times [0,\hat{T}] \times \boldsymbol{X})$,

$$E_x^{\zeta^1,\zeta^2}\left[\int_0^{\hat{T}}\left(\frac{\partial \psi}{\partial t}(\omega,t,\xi_t) + \int_{\boldsymbol{X}} \psi(\omega,t,y)\int_V\int_U q(dy|\xi_t,u,v)\zeta^1(du|\omega,t)\zeta^2(dv|\omega,t)\right)dt\right]$$
$$= E_x^{\zeta^1,\zeta^2}[\psi(\omega,\hat{T},\xi_{\hat{T}})] - E_x^{\zeta^1,\zeta^2}\psi(\omega,0,x).$$

Note that since $(\zeta^1,\zeta^2) \in \Pi^1_{Ad} \times \Pi^2_{Ad}$ *may be dependent on histories,* $\{\xi_t, t \geq 0\}$ *may be not Markovian.*

(b) For each $x \in \boldsymbol{X}$, $(\zeta^1,\zeta^2) \in \Pi^1_M \times \Pi^2_M$ *and* $\psi \in \mathcal{C}^1_{\mathcal{W},\mathcal{W}_1}([0,\hat{T}] \times \boldsymbol{X})$,

$$E_\gamma^{\zeta^1,\zeta^2}\left[\int_s^{\hat{T}}\left(\left(\frac{\partial \psi}{\partial t}(t,\xi_t) + \lambda c(\xi_t,\zeta^1_t,\zeta^2_t)\right)e^{\int_s^t \lambda c(\xi_\beta,\zeta^1_\beta,\zeta^2_\beta)d\beta}\psi(t,\xi_t)\right.\right.$$
$$\left.\left. + \int_{\boldsymbol{X}} e^{\int_s^t \lambda c(\xi_\beta,\zeta^1_\beta,\zeta^2_\beta)d\beta}\psi(t,y)q(dy|\xi_t,\zeta^1_t,\zeta^2_t)\right)dt|\xi_s = x\right]$$
$$= E_\gamma^{\zeta^1,\zeta^2}\left[e^{\int_s^{\hat{T}} \lambda c(\xi_\beta,\zeta^1_\beta,\zeta^2_\beta)d\beta}\psi(\hat{T},\xi_{\hat{T}})|\xi_s = x\right] - \psi(s,x).$$

Proof. See Guo and Zhang (2019), Theorem 3.1.

Next, we present a theorem which shows that the solutions of the optimality equations (Shapley equations) have unique probabilistic representations. In section 4, we also illustrate how this verification theorem can be used to determine the game's value.

Theorem 2. *Assume that Assumptions 1 and 2 are true. If there exist a function* $\psi \in \mathcal{C}^1_{\mathcal{W},\mathcal{W}_1}([0,\hat{T}] \times \boldsymbol{X})$ *and a pair of stationary strategies* $(\zeta^{*1},\zeta^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *for which*

$$\psi(s,x) - e^{\lambda g(x)}$$
$$= E_1 = \int_s^{\hat{T}} \sup_{\vartheta \in \mathcal{P}(U(x))} \inf_{\eta \in \mathcal{P}(V(x))}\left[\lambda c(x,\vartheta,\eta)\psi(t,x) + \int_{\boldsymbol{X}} \psi(t,y)q(dy|x,\vartheta,\eta)\right]dt$$
$$= E_2 = \int_s^{\hat{T}} \inf_{\eta \in \mathcal{P}(V(x))} \sup_{\vartheta \in \mathcal{P}(U(x))}\left[\lambda c(x,\vartheta,\eta)\psi(t,x) + \int_{\boldsymbol{X}} \psi(t,y)q(dy|x,\vartheta,\eta)\right]dt$$
$$= \int_s^{\hat{T}} \inf_{\eta \in \mathcal{P}(V(x))}\left[\lambda c(x,\zeta^{*1}(\cdot|x,t),\eta)\psi(t,x) + \int_{\boldsymbol{X}} \psi(t,y)q(dy|x,\zeta^{*1}(\cdot|x,t),\eta)\right]dt$$
$$= \int_s^{\hat{T}} \sup_{\vartheta \in \mathcal{P}(U(x))}\left[\lambda c(x,\vartheta,\zeta^{*2}(\cdot|x,t))\psi(t,x) + \int_{\boldsymbol{X}} \psi(t,y)q(dy|x,\vartheta,\zeta^{*2}(\cdot|x,t))\right]dt$$
$$s \in [0,\hat{T}],\ x \in \boldsymbol{X}, \tag{6}$$

then

(a)

$$\psi(0,x) = \sup_{\zeta^1 \in \Pi^1_{Ad}} \inf_{\zeta^2 \in \Pi^2_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(0,x) = \inf_{\zeta^2 \in \Pi^2_{Ad}} \sup_{\zeta^1 \in \Pi^1_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(0,x)$$
$$= \inf_{\zeta^2 \in \Pi^2_{Ad}} \mathcal{H}^{\zeta^{*1},\zeta^2}(0,x) = \sup_{\zeta^1 \in \Pi^1_{Ad}} \mathcal{H}^{\zeta^1,\zeta^{*2}}(0,x),\ x \in \boldsymbol{X} \tag{7}$$

and

https://doi.org/10.17993/3cemp.2022.110250.76-92

(b)

$$\psi(t,x) = \sup_{\zeta^1\in\Pi^1_M}\inf_{\zeta^2\in\Pi^2_M}\mathcal{H}^{\zeta^1,\zeta^2}(t,x) = \inf_{\zeta^2\in\Pi^2_M}\sup_{\zeta^1\in\Pi^1_M}\mathcal{H}^{\zeta^1,\pi^2}(t,x)$$
$$= \inf_{\zeta^2\in\Pi^2_M}\mathcal{H}^{\zeta^{*1},\zeta^2}(t,x) = \sup_{\zeta^1\in\Pi^1_M}\mathcal{H}^{\zeta^1,\zeta^{*2}}(t,x) = \mathcal{H}^*(t,x),\ t\in[0,\hat{T}],\ x\in\boldsymbol{X}. \tag{8}$$

Proof.

(a) See Golui and Pal (2021a), Corollary 3.1.

(b) This proof follows from part (a).

4 THE EXISTENCE OF OPTIMAL SOLUTION AND SADDLE POINT EQUILIBRIUM

This section provides the proof that optimality equation (6) has a solution in the space $\mathcal{C}^1_{\mathcal{W},\mathcal{W}_1}([0,\hat{T}]\times\boldsymbol{X})$. Furthermore, we use the optimality equation (6) to prove the existence of saddle point equilibrium. The next Proposition proves the optimality equation (6) has a solution when the rates are bounded.

Proposition 1. *Suppose Assumption 3 holds. Also, assume that* $\|q\|<\infty$, $\|c\|<\infty$, $\|g\|<\infty$, $c(x,u,v)\geq 0$ *and* $g(x)\geq 0$, *for all* $(x,u,v)\in\mathcal{K}$. *Then the following results are true.*

(a) There exists a bounded function $\psi\in\mathcal{B}_1([0,\hat{T}]\times\boldsymbol{X})$ *satisfying first two equations (*E_1 *and* E_2*) of (6).*

(b) There exists a pair of strategies $(\zeta^{*1},\zeta^{*2})\in\Pi^1_{SM}\times\Pi^2_{SM}$ *satisfying the equations (6), (7) and (8) and hence this forms a saddle-point equilibrium.*

(c) $\mathcal{H}^*(t,x)$ *(and so* $\psi(t,x)$*) is non-increasing in* t *for fixed* $x\in\boldsymbol{X}$, *where* $t\in[0,\hat{T}]$.

Proof. (a) From Wei (2017), Theorem 4.1, there exists $\psi\in\mathcal{B}_1([0,\hat{T}]\times\boldsymbol{X})$ satisfying first two equations (E_1 and E_2) of (6).

(b) In view of measurable selection theorem as in Nowak (1985), we get the existence of $(\zeta^{*1},\zeta^{*2})\in\Pi^1_{SM}\times\Pi^2_{SM}$ for which (6) holds. So, by Theorem 2, we get

$$\sup_{\zeta^1\in\Pi^1_{Ad}}\inf_{\zeta^2\in\Pi^2_{Ad}}\mathcal{H}^{\zeta^1,\zeta^2}(0,x) = \inf_{\zeta^2\in\Pi^2_{Ad}}\sup_{\zeta^1\in\Pi^1_{Ad}}\mathcal{H}^{\zeta^1,\zeta^2}(0,x) = \sup_{\zeta^1\in\Pi^1_{Ad}}\mathcal{H}^{\zeta^1,\zeta^{*2}}(0,x)$$
$$= \inf_{\zeta^2\in\Pi^2_{Ad}}\mathcal{H}^{\zeta^{*1},\zeta^2}(0,x) = \psi(0,x) \tag{9}$$

and

$$\sup_{\zeta^1\in\Pi^1_M}\inf_{\zeta^2\in\Pi^2_M}\mathcal{H}^{\zeta^1,\zeta^2}(t,x) = \inf_{\zeta^2\in\Pi^2_M}\sup_{\zeta^1\in\Pi^1_M}\mathcal{H}^{\zeta^1,\zeta^2}(t,x) = \sup_{\zeta^1\in\Pi^1_M}\mathcal{H}^{\zeta^1,\zeta^{*2}}(t,x)$$
$$= \inf_{\zeta^2\in\Pi^2_M}\mathcal{H}^{\zeta^{*1},\zeta^2}(t,x) = \mathcal{H}^*(t,x) = \psi(t,x). \tag{10}$$

Thus the game's value exists and $(\zeta^{*1},\zeta^{*2})\in\Pi^1_{SM}\times\Pi^2_{SM}$ forms a saddle-point equilibrium.

(c) First we fix any $s,t\in[0,\hat{T}]$ where $s<t$. Also fix any $(\zeta^1,\zeta^2)\in\Pi^1_M\times\Pi^2_M$. Now for each $x\in\boldsymbol{X}$, define a Markov strategy corresponding to $\zeta^1\in\Pi^1_M$ as

$$\zeta^1_{s,t}(du|x,\beta) = \begin{cases} \zeta^1(du|x,\beta+t-s) \text{ if } \beta\geq s \\ \zeta^1(du|x,\beta) \text{ otherwise.}\end{cases} \tag{11}$$

https://doi.org/10.17993/3cemp.2022.110250.76-92

Similarly, for each $\zeta^2 \in \Pi_M^2$, we define $\zeta_{s,t}^2$.
Then, for each $\beta \in [s, s+\hat{T}-t]$ and $x \in \boldsymbol{X}$, $q(dy|x, \zeta_{s,t}^1(du|x,\beta), \zeta_{s,t}^2(dv|x,\beta)) = q(dy|x, \zeta^1(du|x, \beta + t - s), \zeta^2(dv|x, \beta+t-s))$,
$c(x, \zeta_{s,t}^1(du|x,\beta), \zeta_{s,t}^2(dv|x,\beta)) = c(x, \zeta^1(du|x, \beta+t-s), \zeta^2(dv|x,\beta+t-s))$. Next define

$$\mathcal{H}^{\zeta^1,\zeta^2}(s \rightsquigarrow t, x) := E_\gamma^{\zeta^1,\zeta^2}\left[e^{\lambda \int_s^t c(\xi_\beta, \zeta^1(du|\xi_\beta,\beta), \zeta^2(dv|\xi_\beta,\beta))d\beta + \lambda g(\xi_t)} | \xi_s = x\right], \tag{12}$$

$$\mathcal{H}^*(s \rightsquigarrow t, x) := \inf_{\zeta^2 \in \Pi_M^2} \sup_{\zeta^1 \in \Pi_M^1} \mathcal{H}^{\zeta^1,\zeta^2}(s \rightsquigarrow t, x). \tag{13}$$

Now in view of the Markov property of $\{\xi_t, t \geq 0\}$ under any $(\zeta^1, \zeta^2) \in \Pi_M^1 \times \Pi_M^2$ and (11)-(13), we have $\mathcal{H}^{\zeta^1,\zeta^2}(t \rightsquigarrow \hat{T}, x) = \mathcal{H}^{\zeta_{s,t}^1,\zeta_{s,t}^2}(s \rightsquigarrow \hat{T}+s-t, x)$.
It can be easily shown that $\sup_{\zeta_{s,t}^1 \in \Pi_M^1} \mathcal{H}^{\zeta_{s,t}^1,\zeta_{s,t}^2}(s \rightsquigarrow \hat{T}+s-t, x) \leq \sup_{\zeta^1 \in \Pi_M^1} \mathcal{H}^{\zeta^1,\zeta^2}(t \rightsquigarrow T, x)$ and $\sup_{\zeta^1 \in \Pi_M^1} \mathcal{H}^{\zeta^1,\zeta^2}(t \rightsquigarrow \hat{T}, x) \leq \sup_{\zeta_{s,t}^1 \in \Pi_M^1} \mathcal{H}^{\zeta_{s,t}^1,\zeta_{s,t}^2}(s \rightsquigarrow \hat{T}+s-t, x)$ for all $\zeta^2 \in \Pi_2^M$. Hence, $\sup_{\zeta^1 \in \Pi_M^1} \mathcal{H}^{\zeta^1,\zeta^2}(t \rightsquigarrow \hat{T}, x) = \sup_{\zeta_{s,t}^1 \in \Pi_M^1} \mathcal{H}^{\zeta_{s,t}^1,\zeta_{s,t}^2}(s \rightsquigarrow \hat{T}+s-t, x)$ for all $\zeta^2 \in \Pi_2^M$. Similarly, we can show that $\mathcal{H}^*(t \rightsquigarrow \hat{T}, x) = \mathcal{H}^*(s \rightsquigarrow \hat{T}+s-t, x)$. Now since $c(x,u,v) \geq 0$ on $\mathcal{K}$, by (13) and $t > s$, we have $\mathcal{H}^*(t \rightsquigarrow \hat{T}, x) = \mathcal{H}^*(s \rightsquigarrow \hat{T}+s-t, x) \leq \mathcal{H}^*(s \rightsquigarrow \hat{T}, x)$. But by (10), (12) and (13), we have $\mathcal{H}^*(t \rightsquigarrow \hat{T}, x) = \mathcal{H}^*(t,x)$. Hence, we obtain $\mathcal{H}^*(s,x) \geq \mathcal{H}^*(t,x)$ i.e. $\mathcal{H}^*(t,x)$ is decreasing in t. Now from part (b), we have $\mathcal{H}^*(t,x) = \psi(t,x)$. Hence, $\psi(t,x)$ is also decreasing in t.

Theorem 3. *Suppose Assumptions 1, 2 and 3 hold. Also, in addition suppose $c(x,u,v) \geq 0$ and $g(x) \geq 0$ for all $(x,u,v) \in \mathcal{K}$. Then there exist a unique $\psi \in \mathcal{C}^1_{\mathcal{W},\mathcal{W}_1}([0,\hat{T}] \times \boldsymbol{X})$ and some pair of strategies $(\zeta^{*1}, \zeta^{*2}) \in \Pi_{SM}^1 \times \Pi_{SM}^2$ satisfying the equations (6), (7) and (8) and hence this is a saddle-point equilibrium.*
Proof. *First observe that $1 \leq e^{2(\hat{T}+1)\lambda c(x,u,v)} \leq M_2 \mathcal{W}(x)$ and $1 \leq e^{2(\hat{T}+1)\lambda g(x)} \leq M_2 \mathcal{W}(x)$. For each integer $n \geq 1$, $x \in \boldsymbol{X}$, define $\boldsymbol{X}_n := \{x \in \boldsymbol{X} | \mathcal{W}(x) \leq n\}$, $U_n(x) := U(x)$ and $V_n(x) := V(x)$. Also for each $(x,u,v) \in \mathcal{K}_n := \{(x,u,v) : x \in \boldsymbol{X}, u \in U_n(x), v \in V_n(x)\}$, define*

$$q_n(dy|x,u,v) := \begin{cases} q(dy|x,u,v) & \text{if } x \in \boldsymbol{X}_n, \\ 0 \quad \text{if} \quad x \notin \boldsymbol{X}_n, \end{cases} \tag{14}$$

$$c_n^+(x,u,v) := \begin{cases} c(x,u,v) \wedge \ min\left\{n, \frac{1}{\lambda(\hat{T}+1)} \ln \sqrt{M_2 \mathcal{W}(x)}\right\} & \text{if } x \in \boldsymbol{X}_n, \\ 0 \quad \text{if} \quad x \notin \boldsymbol{X}_n. \end{cases} \tag{15}$$

and

$$g_n^+(x) := \begin{cases} g(x) \wedge \ min\left\{n, \frac{1}{\lambda(\hat{T}+1)} \ln \sqrt{M_2 \mathcal{W}(x)}\right\} & \text{if } x \in \boldsymbol{X}_n, \\ 0 \quad \text{if} \quad x \notin \boldsymbol{X}_n. \end{cases} \tag{16}$$

By (14), obviously $q_n(dy|x,u,v)$ is transition rates on $\boldsymbol{X}$ satisfying conservative and stable conditions. Now consider the sequence of CTMDPs models with bounded rates $\mathcal{G}_n^+ := \{\boldsymbol{X}, U, V, (U_n(x), V_n(x), x \in \boldsymbol{X}), c_n^+, g_n^+, q_n\}$. Fix a n. Corresponding to a pair of Markov strategies $(\zeta^1, \zeta^2) \in \Pi_M^1 \times \Pi_M^2$, suppose for this model the risk-sensitive cost criterion is $\mathcal{H}_n^{\zeta^1,\zeta^2}(t,x)$ and the value function is

$$\mathcal{H}_n(t,x) := \sup_{\zeta^1 \in \Pi_M^1} \inf_{\zeta^2 \in \Pi_M^2} \mathcal{H}_n^{\zeta^1,\zeta^2}(t,x).$$

Then by Proposition 1, for each $n \geq 1$, we get a unique ψ_n in $\mathcal{C}^1_{1,1}([0,\hat{T}]) \times S$ and $(\zeta_n^{*1}, \zeta_n^{*2}) \in \Pi_{SM}^1 \times \Pi_{SM}^2$ satisfying

$$\psi_n(s,x) - e^{\lambda g_n^+(x)}$$

 https://doi.org/10.17993/3cemp.2022.110250.76-92

$$= \int_s^{\hat{T}} \inf_{\eta \in \mathcal{P}(V(x))} \left[\lambda c_n^+(x, \zeta_n^{*1}(\cdot|x,t), \eta)\psi_n(t,x) + \int_{\boldsymbol{X}} \psi_n(t,y) q_n(dy|x, \zeta_n^{*1}(\cdot|x,t), \eta) \right] dt$$
$$= \int_s^{\hat{T}} \sup_{\vartheta \in \mathcal{P}(U(x))} \left[\lambda c_n^+(x, \vartheta, \zeta_n^{*2}(\cdot|x,t))\psi_n(t,x) + \int_{\boldsymbol{X}} \psi_n(t,y) q_n(dy|x, \vartheta, \zeta_n^{*2}(\cdot|x,t)) \right] dt$$
$$s \in [0, \hat{T}], \ x \in \boldsymbol{X}. \tag{17}$$

Now, $e^{2\lambda(\hat{T}+1)c_n^+(x,u,v)} \leq M_2 \mathcal{W}(x)$, $e^{2\lambda(\hat{T}+1)g_n^+(x)} \leq M_2 \mathcal{W}(x)$ *and* $\psi_n(\hat{T}, x) = e^{\lambda g_n^+(x)}$*. Hence by Lemma 1, Theorem 2 and (17), we have*

$$e^{-L_1 \mathcal{W}(x)} \leq \psi_n(t,x) = \sup_{\zeta^1 \in \Pi_1^M} \mathcal{H}_n^{\zeta^1, \zeta_n^{*2}}(t,x) \leq L_1 \mathcal{W}(x) \ \forall \ n \geq 1. \tag{18}$$

Moreover, since $\psi_n(t,x) \geq 0$, $c_{n-1}^+(x,u,v) \leq c_n^+(x,u,v)$*, and* $g_{n-1}^+(t,x) \leq g_n^+(x)$ $\forall (x,u,v) \in \mathcal{K}$*, using (14), (15), (17) and Proposition 1,* $\forall x \in \boldsymbol{X}$ *and a.e.* t*, we obtain,*

$$\begin{cases} \frac{\partial \psi_n}{\partial t}(t,x) + \left[\lambda c_{n-1}^+(x, \vartheta, \zeta_n^{*2}(\cdot|x,t))\psi_n(t,x) + \int_{\boldsymbol{X}} \psi_n(t,y) q_{n-1}(dy|x, \vartheta, \zeta_n^{*2}(\cdot|x,t)) \right] \\ \leq 0 \qquad \text{if } x \in \boldsymbol{X}_{n-1} \end{cases} \tag{19}$$

and

$$\begin{cases} \frac{\partial \psi_n}{\partial t}(t,x) + \left[\lambda c_{n-1}^+(x, \vartheta, \zeta_n^{*2}(\cdot|x,t))\psi_n(t,x) + \int_{\boldsymbol{X}} \psi_n(t,y) q_{n-1}(dy|x, \vartheta, \zeta_n^{*2}(\cdot|x,t)) \right] \\ = \frac{\partial \psi_n}{\partial t}(t,x) \leq 0 \qquad \text{if } x \notin \boldsymbol{X}_{n-1}, \end{cases} \tag{20}$$

(for details see, Golui and Pal (2021b), Theorem 4.1, p. 24). So, for any $\zeta^1 \in \Pi_M^1$*, by Feynman-Kac formula (similar proof as in Theorem 2), we get*

$$\mathcal{H}_{n-1}^{\zeta^1, \zeta_n^{*2}}(t,x) \leq \psi_n(t,x).$$

Since $\zeta^1 \in \Pi_M^1$ *is arbitrary*

$$\inf_{\zeta^2 \in \Pi_M^2} \sup_{\zeta^1 \in \Pi_M^1} \mathcal{H}_{n-1}^{\zeta^1, \zeta^2}(t,x) \leq \sup_{\zeta^1 \in \Pi_M^1} \mathcal{H}_{n-1}^{\zeta^1, \zeta_n^{*2}}(t,x) \leq \psi_n(t,x). \tag{21}$$

Also using (17) and Feynman-Kac formula (similar proof as in Theorem 2), we have

$$\sup_{\zeta^1 \in \Pi_M^1} \inf_{\zeta^2 \in \Pi_M^2} \mathcal{H}_{n-1}^{\zeta^1, \zeta^2}(t,x) = \inf_{\zeta^2 \in \Pi_M^2} \sup_{\zeta^1 \in \Pi_M^1} \mathcal{H}_{n-1}^{\zeta^1, \zeta^2}(t,x) = \psi_{n-1}(t,x). \tag{22}$$

From (21) and (22), we obtain $\psi_{n-1}(t,x) \leq \psi_n(t,x)$*. Also, since* ψ_n *has an upper bound,* $\lim_{n\to\infty} \psi_n$ *exists. Let*

$$\lim_{n\to\infty} \psi_n(t,x) := \psi(t,x) \ \forall t \in [0, \hat{T}], \ \forall x \in \boldsymbol{X}. \tag{23}$$

Next by Lemma 1, we get

$$|\psi(t,x)| \leq L_1 \mathcal{W}(x) \quad \forall t \in [0, \hat{T}]. \tag{24}$$

Let

$$I_n(t,x) := \sup_{\vartheta \in \mathcal{P}(U_n(x))} \inf_{\eta \in \mathcal{P}(V_n(x))} \left[\lambda c_n^+(x, \vartheta, \eta)\psi_n(t,x) + \int_{\boldsymbol{X}} \psi_n(t,y) q_n(dy|x, \vartheta, \eta) \right],$$
$$\forall t \in [0, \hat{T}], \ \forall x \in \boldsymbol{X}.$$

Then, applying Fan's minimax theorem, Fan, (1953), we obtain

$$I_n(t,x) := \inf_{\eta \in \mathcal{P}(V_n(x))} \sup_{\vartheta \in \mathcal{P}(U_n(x))} \left[\lambda c_n^+(x, \vartheta, \eta)\psi_n(t,x) + \int_{\boldsymbol{X}} \psi_n(t,y) q_n(dy|x, \vartheta, \eta) \right],$$

 https://doi.org/10.17993/3cemp.2022.110250.76-92

$\forall t \in [0,\hat{T}],\ \forall x \in \boldsymbol{X}.$

Then, by Assumptions 1 and 2 and the fact that $\lambda \leq 1$, we get the following result

$$\begin{aligned}|I_n(t,x)| &\leq L_1\Big(M_2\mathcal{W}^2(x) + (b_1+\rho_1)\mathcal{W}^2(x) + 2M_1\mathcal{W}^2(x)\Big)\\ &\leq L_1M_3\mathcal{W}_1(x)(M_2+b_1+\rho_1+2M_1) =: \mathfrak{R}(x),\ (t,x)\in[0,\hat{T}]\times\boldsymbol{X}.\end{aligned} \tag{25}$$

Let

$$I(t,x) := \sup_{\vartheta\in\mathcal{P}(U(x))}\inf_{\eta\in\mathcal{P}(V(x))}\Big[\lambda c(x,\vartheta,\eta)\psi(t,x) + \int_{\boldsymbol{X}}\psi(t,y)q(dy|x,\vartheta,\eta)\Big],$$
$$\forall t\in[0,\hat{T}],\ \forall x\in\boldsymbol{X}.$$

Hence in view of Fan's minimax theorem, Fan, (1953), we obtain

$$I(t,x) := \inf_{\eta\in\mathcal{P}(V(x))}\sup_{\vartheta\in\mathcal{P}(U(x))}\Big[\lambda c(x,\vartheta,\eta)\psi(t,x) + \int_{\boldsymbol{X}}\psi(t,y)q(dy|x,\vartheta,\eta)\Big],$$
$$\forall t\in[0,\hat{T}],\ \forall x\in\boldsymbol{X}.$$

We next prove that for each fixed $x \in \boldsymbol{X}$ and $t \in [0,\hat{T}]$, along some suitable subsequence of $\{n\}$ (if necessary), $\lim_{n\to\infty} I_n(t,x) = I(t,x)$. Now, using Assumption 3, the functions $c(x,\vartheta,\eta)$ and $\int_{\boldsymbol{X}} q(dy|x,\vartheta,\eta)\psi_n(t,y)$ are continuous on $\mathcal{P}(U(x))\times\mathcal{P}(V(x))$ for each $x\in\boldsymbol{X}$. So, we find a sequence of pair of measurable functions $(\vartheta_n^*,\eta_n^*)\in\mathcal{P}(U(x))\times\mathcal{P}(V(x))$ such that

$$\begin{aligned}I_n(t,x) :&= \inf_{\eta\in\mathcal{P}(V(x))}\Big[\lambda c_n^+(x,\vartheta_n^*,\eta)\psi_n(t,x) + \int_{\boldsymbol{X}}\psi_n(t,y)q_n(dy|x,\vartheta_n^*,\eta)\Big]\\ &= \sup_{\vartheta\in\mathcal{P}(U(x))}\Big[\lambda c_n^+(x,\vartheta,\eta_n^*)\psi_n(t,x) + \int_{\boldsymbol{X}}\psi_n(t,y)q_n(dy|x,\vartheta,\eta_n^*)\Big].\end{aligned} \tag{26}$$

Now, $\mathcal{P}(U(x))$ and $\mathcal{P}(V(x))$ are compact. So, there exists a subsequences (here, we take the same sequence for simplicity) that $\vartheta_n^*\to\vartheta^*$ and $\eta_n^*\to\eta^*$ as $n\to\infty$ for some $(\vartheta^*,\eta^*)\in\mathcal{P}(U(x))\times\mathcal{P}(V(x))$.

Taking $n\to\infty$ in (26), by the generalized version of Fatou's lemma Feinberge et al. (2014), Hernandez-Lerma and Lasserre (1999), Lemma 8.3.7, for arbitrarily fixed $\vartheta\in\mathcal{P}(U(x))$, we have

$$\liminf_{n\to\infty} I_n(t,x) \geq \Big[\lambda c(x,\vartheta,\eta^*)\psi(t,x) + \int_{\boldsymbol{X}}\psi(t,y)q(dy|x,\vartheta,\eta^*)\Big].$$

Since $\vartheta\in\mathcal{P}(U(x))$ is arbitrary,

$$\begin{aligned}\liminf_{n\to\infty} I_n(t,x) &\geq \sup_{\vartheta\in\mathcal{P}(U(x))}\Big[\lambda c(x,\vartheta,\eta^*)\psi(t,x) + \int_{\boldsymbol{X}}\psi(t,y)q(dy|x,\vartheta,\eta^*)\Big]\\ &\geq \inf_{\eta\in\mathcal{P}(V(x))}\sup_{\vartheta\in\mathcal{P}(U(x))}\Big[\lambda c(x,\vartheta,\eta)\psi(t,x) + \int_{\boldsymbol{X}}\psi(t,y)q(dy|x,\vartheta,\eta)\Big].\end{aligned} \tag{27}$$

Using analogous arguments from (26), by the generalized version of Fatou's Lemma, Feinberge et al. (2014), Hernandez-Lerma and lasserre (1999), Lemma 8.3.7, we have

$$\limsup_{n\to\infty} I_n(t,x) \leq \sup_{\vartheta\in\mathcal{P}(U(x))}\inf_{\eta\in\mathcal{P}(V(x))}\Big[\lambda c(x,\vartheta,\eta)\psi(t,x) + \int_{\boldsymbol{X}}\psi(t,y)q(dy|x,\vartheta,\eta)\Big]. \tag{28}$$

So, by (27) and (28), we get

$$\lim_{n\to\infty} I_n(t,x) = I(t,x)\ \forall t\in[0,\hat{T}],\ \forall x\in\boldsymbol{X}. \tag{29}$$

 https://doi.org/10.17993/3cemp.2022.110250.76-92

Since $\lim_{n\to\infty}\psi_n(t,x)=\psi(t,x)$ *and* $\forall t\in[0,\hat{T}]$, $\forall x\in\boldsymbol{X}$, *in view of (29) and the dominated convergent theorem (since* $|I_n(t,x)|\le\mathcal{R}(x)$*), taking limit* $n\to\infty$ *in (17), we say that* ψ *satisfies first two equations (*E_1 *and* E_2*) of (6) and hence* $\psi(\cdot,x)$ *is differentiable almost everywhere on* $[0,\hat{T}]$*, see Athreya (2006), Theorem 4.4.1. Again, by the analogous arguments as in (25), we obtain*

$$\left|\frac{\partial\psi(t,x)}{\partial t}\right|=|I(t,x)|\le R(x),\quad\forall t\in[0,\hat{T}],\ \forall x\in\boldsymbol{X}.$$

Therefore, we see that $\psi\in\mathcal{C}^1_{\mathcal{W},\mathcal{W}_1}([0,\hat{T}]\times\boldsymbol{X})$*. Furthermore, using analogous arguments as in Proposition 1 (b),* ψ *is the unique solution of (6) satisfying (7) and (8) and hence saddle-point equilibrium exists.*

Next we state the main optimal results that provide the proof of the existence of saddle point equilibrium and game's value when payoff rates are extended real valued functions.

Theorem 4. *We grant Assumptions 1, 2 and 3. Then, the following claims are true.*

(a) There exists a unique function $\psi\in\mathcal{C}^1_{\mathcal{W},\mathcal{W}_1}([0,\hat{T}]\times\boldsymbol{X})$ *that satisfies first two equations (*E_1 *and* E_2*) of (6).*

(b) There exists a pair of strategies $(\zeta^{*1},\zeta^{*2})\in\Pi^1_{SM}\times\Pi^2_{SM}$ *that satisfies the equations (6), (7) and (8) and hence this pair of strategies becomes a saddle-point equilibrium.*

Proof. *We only need prove part (a) since part (b) follows from Proposition 1 (b). Now, for each* $n\ge1$*, define* c_n *and* g_n *on* $\mathcal{K}$ *as:*

$$c_n(x,u,v):=\max\{-n,c(x,u,v)\},\ \ g_n(x):=\max\{-n,g(x)\}$$

for each $(x,u,v)\in\mathcal{K}$*. Then* $\lim_{n\to\infty}c_n(x,u,v)=c(x,u,v)$ *and* $\lim_{n\to\infty}g_n(x)=g(x)$*. Define* $r_n(x,u,v):=c_n(x,u,v)+n$ *and* $\tilde{g}_n(x):=g_n(x)+n$*. So,* $r_n(x,u,v)\ge0$ *and* $\tilde{g}_n(x)\ge0$ *for each* $n\ge1$ *and* $(x,u,v)\in\mathcal{K}$*. Now by Assumption 1, we have*

$$-\frac{\ln\sqrt{M_2\mathcal{W}(x)}}{\lambda(\hat{T}+1)}\le\max\left\{-n,-\frac{\ln\sqrt{M_2\mathcal{W}(x)}}{\lambda(\hat{T}+1)}\right\}\le c_n(x,u,v)\le\frac{\ln\sqrt{M_2\mathcal{W}(x)}}{\lambda(\hat{T}+1)}\tag{30}$$

and

$$-\frac{\ln\sqrt{M_2\mathcal{W}(x)}}{\lambda(\hat{T}+1)}\le\max\left\{-n,-\frac{\ln\sqrt{M_2\mathcal{W}(x)}}{\lambda(\hat{T}+1)}\right\}\le g_n(x)\le\frac{\ln\sqrt{M_2\mathcal{W}(x)}}{\lambda(\hat{T}+1)}.\tag{31}$$

So, we have $e^{2\lambda(\hat{T}+1)r_n(x,u,v)}\le e^{2\lambda(\hat{T}+1)n}M_2\mathcal{W}(x)$ *and* $e^{2\lambda(\hat{T}+1)\tilde{g}_n(x)}\le e^{2\lambda(\hat{T}+1)n}M_2\mathcal{W}(x)$, $\forall n\ge1$ *and* $(x,u,v)\in\mathcal{K}$*. Define a new model* $\mathcal{R}_n:=\{\boldsymbol{X},U,V,(U_n(x),V_n(x),x\in\boldsymbol{X}),r_n,\tilde{g}_n,q\}$*. Now for any real-valued measurable functions* $\tilde{\psi}$ *and* ϕ *defined on* $\mathcal{K}$ *and* $[0,\hat{T}]\times\boldsymbol{X}$*, respectively, define*

$$\mathcal{H}(s,x,\tilde{\psi},\phi):=\sup_{\zeta^1\in\Pi^1_M}\inf_{\zeta^2\in\Pi^2_M}E^{\zeta^1,\zeta^2}_{\gamma}\left[\exp\left(\lambda\int_s^{\hat{T}}\tilde{\psi}(\xi_t,\pi^1_t,\pi^2_t)dt+\lambda\phi(\xi_{\hat{T}})\right)\Big|\xi_s=x\right]\tag{32}$$

assuming that the integral exists. Now since $r_n\ge0$, $\tilde{g}_n\ge0$ *and all Assumptions hold for the model* $\mathcal{R}_n$*, by Theorem 3, we have*

$$\begin{aligned}&-\frac{\partial\mathcal{H}(s,x,r_n,\tilde{g}_n)}{\partial s}\\&=\sup_{\vartheta\in\mathcal{P}(U(x))}\inf_{\eta\in\mathcal{P}(V(x))}\left[\lambda r_n(x,\vartheta,\nu)\mathcal{H}(s,x,r_n,\tilde{g}_n)+\int_{\boldsymbol{X}}\mathcal{H}(s,y,r_n,\tilde{g}_n)q(dy|x,\vartheta,\eta)\right]\\&=\inf_{\eta\in\mathcal{P}(V(x))}\sup_{\vartheta\in\mathcal{P}(U(x))}\left[\lambda r_n(x,\vartheta,\nu)\mathcal{H}(s,x,r_n,\tilde{g}_n)+\int_{\boldsymbol{X}}\mathcal{H}(s,y,r_n,\tilde{g}_n)q(dy|x,\vartheta,\eta)\right]\end{aligned}\tag{33}$$

https://doi.org/10.17993/3cemp.2022.110250.76-92

for almost all $s \in [0, \hat{T}]$. *Now*

$$\mathcal{H}(s,x,r_n,\tilde{g}_n) = \mathcal{H}(s,x,c_n+n,g_n+n) = \mathcal{H}(s,x,c_n,g_n)e^{\lambda(\hat{T}-s+1)n}.$$

So, by (33), we can write for a.e. $s \in [0, \hat{T}]$,

$$\begin{aligned} -\frac{\partial \mathcal{H}(s,x,c_n,g_n)}{\partial s} &= \sup_{\vartheta\in\mathcal{P}(U(x))}\inf_{\eta\in\mathcal{P}(V(x))}\Big[\lambda c_n(x,\vartheta,\eta)\mathcal{H}(s,x,c_n,g_n) + \int_{\boldsymbol{X}} \mathcal{H}(s,y,c_n,g_n)q(dy|x,\vartheta,\eta)\Big] \\ &= \inf_{\eta\in\mathcal{P}(V(x))}\sup_{\vartheta\in\mathcal{P}(U(x))}\Big[\lambda c_n(x,\vartheta,\eta)\mathcal{H}(s,x,c_n,g_n) + \int_{\boldsymbol{X}} \mathcal{H}(s,y,c_n,g_n)q(dy|x,\vartheta,\eta)\Big]. \end{aligned}$$

Hence

$$\begin{aligned} &\mathcal{H}(s,x,c_n,g_n) - e^{\lambda g_n(x)} \\ &= \int_s^{\hat{T}} \sup_{\vartheta\in\mathcal{P}(U(x))}\inf_{\eta\in\mathcal{P}(V(x))}\Big[\lambda c_n(x,\vartheta,\eta)\mathcal{H}(t,x,c_n,g_n) + \int_{\boldsymbol{X}} \mathcal{H}(t,y,c_n,g_n)q(dy|x,\vartheta,\eta)\Big]dt \\ &= \int_s^{\hat{T}} \inf_{\eta\in\mathcal{P}(V(x))}\sup_{\vartheta\in\mathcal{P}(U(x))}\Big[\lambda c_n(x,\vartheta,\eta)\mathcal{H}(t,x,c_n,g_n) + \int_{\boldsymbol{X}} \mathcal{H}(t,y,c_n,g_n)q(dy|x,\vartheta,\eta)\Big]dt. \end{aligned} \tag{34}$$

Now by (34) and Lemma 1, we obtain

$$|\mathcal{H}(t,x,c_n,g_n)| \le L_1\mathcal{W}(x) \quad n \ge 1. \tag{35}$$

Now since $c_n(x,u,v)$ *and* $g_n(x)$ *are non-increasing in* $n \ge 1$*, hence its corresponding value function* $\mathcal{H}(t,x,c_n,g_n)$ *is also non-increasing in* n*. Also by Lemma 1, we know that* $\mathcal{H}(\cdot,\cdot,c_n,g_n)$ *has a lower bound. So,* $\lim_{n\to\infty}\mathcal{H}(t,x,c_n,g_n)$ *exists. Let* $\lim_{n\to\infty}\mathcal{H}(t,x,c_n,g_n) =: \psi(t,x)$, $(t,x) \in [0,\hat{T}] \times \boldsymbol{X}$*. Then using analogous arguments as Theorem 4.1, and using the function* $\mathcal{H}(t,x,c_n,g_n)$ *in the place of the function* $\psi_n(t,x)$ *here, by (34), (35), Assumptions 1, and 2, we see that (a) is true.*

The converse of Theorem 4 is given below.

Theorem 5. *Under Assumptions 1, 2 and 3, suppose* $(\hat{\zeta}^{*1}, \hat{\zeta}^{*2}) \in \Pi^1_{SM} \times \Pi^2_{SM}$ *is a saddle-point equilibria. Then* $(\hat{\zeta}^{*1}, \hat{\zeta}^{*2})$ *is a mini-max selector of eq. (6).*

Proof. Using the definition of saddle-point equilibrium, we have

$$\begin{aligned} \mathcal{H}^{\hat{\zeta}^{*1},\hat{\zeta}^{*2}}(0,x) &= \sup_{\zeta^2\in\Pi^2_{Ad}}\inf_{\zeta^1\in\Pi^1_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(0,x) \\ &= \inf_{\zeta^1\in\Pi^1_{Ad}}\sup_{\zeta^2\in\Pi^2_{Ad}} \mathcal{H}^{\zeta^1,\zeta^2}(0,x) = \sup_{\zeta^2\in\Pi^2_{Ad}} \mathcal{H}^{\hat{\zeta}^{*1},\zeta^2}(0,x) = \inf_{\zeta^1\in\Pi^1_{Ad}} \mathcal{H}^{\zeta^1,\hat{\zeta}^{*2}}(0,x). \end{aligned} \tag{36}$$

Now arguing as in Theorem 4, it follows that for $\hat{\zeta}^{*1} \in \Pi^1_{SM}$ there exists a function $\tilde{\psi} \in C^1_{\mathcal{W},\mathcal{W}_1}([0,\hat{T}]\times\boldsymbol{X})$ such that

$$\begin{aligned} &\tilde{\psi}(s,x) - e^{\lambda g(x)} \\ &= \int_s^{\hat{T}} \inf_{\eta\in\mathcal{P}(V(x))}\Big[\lambda c(x,\hat{\zeta}^{*1}(\cdot|x,t),\eta)\tilde{\psi}(t,x) + \int_{\boldsymbol{X}} \tilde{\psi}(t,y)q(dy|x,\hat{\zeta}^{*1}(\cdot|x,t),\eta)\Big]dt \\ &s \in [0,\hat{T}],\ x \in \boldsymbol{X}, \end{aligned} \tag{37}$$

satisfying

$$\tilde{\psi}(0,x) = \inf_{\zeta^2\in\Pi^2_{Ad}} \mathcal{H}^{\hat{\zeta}^{*1},\zeta^2}(0,x) \tag{38}$$

and

$$\tilde{\psi}(t,x) = \inf_{\zeta^2\in\Pi^2_{m}} \mathcal{H}^{\hat{\zeta}^{*1},\zeta^2}(t,x). \tag{39}$$

Then by (6), (36), (37), (38), (39), Theorem 2, Theorem 4, we say that $\hat{\zeta}^{*1}$ is outer maximizing selector of (6). By analogous arguments, $\hat{\zeta}^{*2}$ is outer minimizing selector of (6).

 https://doi.org/10.17993/3cemp.2022.110250.76-92

5 EXAMPLE

This section is dedicated for an example to validate assumptions in this paper, where transition and cost functions are not bounded.

Example 1. *Consider a model of a zero-sum game as*

$$\mathcal{G} := \{\boldsymbol{X}, (U, U(x), x \in \boldsymbol{X}), (V, V(x), x \in \boldsymbol{X}), c(x,u,v), q(dy|x,u,v)\}.$$

Suppose our state space is $\boldsymbol{X} = (-\infty, \infty)$ *and transition rate is given by*

$$q(\hat{D}|x,u,v) = \hat{\lambda}(x,a,b)\left[\int_{y\in\hat{D}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}} dy - \delta_x(\hat{D})\right], \ x \in \boldsymbol{X}, \hat{D} \in \mathcal{B}(\boldsymbol{X}), (u,v) \in U(x) \times V(x). \tag{40}$$

We take the following requirements to see if our model has a saddle-point equilibrium.

(I) $U(x)$ *and* $V(x)$ *are compact subsets of the Borel spaces* U *and* V*, respectively, for each fixed* $x \in \boldsymbol{X}$*,.*

(II) The payoff function $c(x,u,v)$ *and the rate function* $\hat{\lambda}(x,u,v)$ *are continuous on* $U(x) \times V(x)$*, for each* $x \in S$*. Also, assume that* $e^{2\lambda(\hat{T}+1)|c(x,u,v)|} \leq M_2\mathcal{W}(x)$, $e^{2\lambda(\hat{T}+1)|g(x)|} \leq M_2\mathcal{W}(x)$ *and* $0 < \sup_{(u,v)\in U(x)\times V(x)} \hat{\lambda}(x,u,v) \leq M_0(x^2+1)$ *for each* $(x,u,v) \in \mathcal{K}$*.*

Proposition 2. *In view of conditions (I)-(II), Assumptions 1, 2, and 3 are satisfied by above controlled system. Therefore, the existence of a saddle point equilibrium is proved by Theorem 4.*

Proof. See Guo and Zhang (2019), Proposition 5.1.

6 CONCLUSIONS

A finite-time horizon dynamic zero-sum game with risk-sensitive cost criteria on a Borel state space is studied. Here for each state x, the admissible action spaces ($U(x)$ and $V(x)$) are compact metric spaces and costs and transition rate functions are unbounded. Under certain assumptions, we have solved the Shapley equation and have established a saddle point equilibrium.

Risk-sensitive non-zero-sum game with unbounded rates (costs and transition rates) over countable state space was investigated in Wei (2019). It would be a challenging problem to study the same problem on the Borel state space.

ACKNOWLEDGMENT

The second named author's research is supported partially by SERB, India, grant MTR/2021/000307.

REFERENCES

[1] **Athreya, K.B.**, & **Lahiri, S.N.** (2006). *Measure Theory and Probability Theory*, Springer, New York.

[2] **Bauerle, E.,** & **Rieder, U.** (2017). Zero-sum risk-sensitive stochastic games, *Stoch. Process. Appl.*, 127, 622-642.

[3] **Bell, D.E.** (1995). Risk, return and utility, *Manage Sci.*, 41, 23-30.

[4] **Bielecki, T.R.,** & **Pliska, S.R.** (1999). Risk-sensitive dynamic optimization, *Appl. Math. Optim.*, 39, 337-360.

[5] **Fan, K.** (1953). Minimax Theorems, *Proc. Natl. Acad. Sci., USA, 39*, 42-47.

[6] **Feinberge, E.A. Kasyanov, P.O.** & **Zadoianchuk, N.V.,** (2014). Fatou's lemma for weakly convergening probabilities, *Theory Probab. Appl.,* 58, 683-689.

[7] **Ghosh, M.K., Kumar, K.S.,** & **Pal, C.** (2016). Zero-sum risk-sensitive stochastic games for continuous-time Markov chains, *Stoch. Anal. Appl.,* 34, 835-851.

[8] **Ghosh, M.K., Golui, S., Pal. C.** & **Pradhan, S.** (2022). Nonzero-Sum Risk-Sensitive Continuous-Time Stochastic Games with Ergodic Costs, *Appl. Math. Optim.*, https://doi.org/10.1007/s00245-022-09878-9.

[9] **Ghosh, M.K.,** & **Saha, S.** (2014). Risk-sensitive control continuous-time Markov chains, *Stochastics,* 86, 655-675.

[10] **Golui, S.,** & **Pal, C.** (2021a). Continuous-time zero-sum games for Markov chains with risk-sensitive finite-horizon cost criterion, *Stoch. Anal. Appl., 40*, 78-95.

[11] **Golui, S., Pal, C.** (2021b). Continuous-time zero-sum games for Markov decision processes with discounted risk-sensitive cost criterion on a general state space, *Stoch. Anal. Appl.*,https://doi.org/10.1080/07362994.2021.2013889.

[12] **Golui, S.,** & **Pal, C.** (2022). Risk-sensitive discounted cost criterion for continuous-time Markov decision processes on a general state space, *Math. Meth. Oper. Res., 95*, 219-247.

[13] **Golui, S., Pal, C.,** & **Saha, S.** (2022). Continuous-Time Zero-Sum Games for Markov Decision Processes with Discounted Risk-Sensitive Cost Criterion, *Dyn. Games Appl., 12*, 485-512.

[14] **Guo, X.P.,** & **Hernandez-Lerma, O.** (2003). Zero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates, *J. Appl. Probab.*, 40, 327-345.

[15] **Guo, X.P.,** & **Hernandez-Lerma, O.** (2005). Nonzero-sum games for continuous-time Markov chains with unbounded discounted payoffs, *J. Appl. Probab.*, 303-320.

[16] **Guo, X.P.,** & **Hernandez-Lerma, O.** (2007). Zero-sum games for continuous-time jump Markov processes in Polish spaces: discounted payoffs, *Adv. Appl. Probab.,* 645-668.

[17] **Guo, X.P.** (2007). Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces, *Math. Oper. Res.*, 32, 73-87.

[18] **Guo, X.P.,** & **Hernandez-Lerma, O.** (2009). *Continuous-time Markov decision processes: Theory and Applications*, Springer, Berlin.

[19] **Guo, X.P., Huang X.,** & **Huang, Y.** (2015). Finite-horizon optimality for continuous-time Markov decision processes with unbounded transition rates, *Adv. Appl. Probab.,* 47, 1064-1087.

[20] **Guo, X.P, Huang, Y.,** & **Song, X.** (2012). Linear programming and constrained average optimality for general continuous-time Markov decision processes in history-dependent polices, *Siam J. Control Optim.,* 50, 23-47.

[21] **Guo, X.P.** & **Liao, Z.W.** (2019). Risk-sensitive discounted continuous-time Markov decision processes with unbounded rates, *SIAM J. Control Optim.,* 57, 3857-3883.

[22] **Guo, X., Liu, Q.,** & **Zhang, Y.** (2019). Finite-horizon risk-sensitive continuous-time Markov decision processes with unbounded transition and cost rates, *4OR,* 427-442.

[23] **Guo, X.** & **Piunovskiy, A.** (2011). Discounted continuous-time Markov decision processes with constraints: Unbounded transition and loss rates, *Math. Oper. Res.,* 36, 105-132.

[24] **Guo, X.P.,** & **Song, X.** (2011). Discounted continuous-time constrained Markov decision processes in polish spaces, *Ann. Appl. Probab.,* 21, 2016-2049.

[25] **Guo, X.P.,** & **Zhang, J.** (2018). On risk-sensitive piecewise deterministic Markov decision processes, *Appl. Math. Optim.* 81, 685-710.

[26] **Guo, X.P.,** & **Zhang, J.** (2019). Risk-sensitive continuous-time Markov decision processes with unbounded rates and Borel spaces, *Discrete Event Dyn. Syst.,* 29, 445-471.

[27] **Hernandez-Lerma, O.,** & **Lasserre, J.,** (1999). *Further topics on discrete-time Markov control processes*, Springer, New York.

[28] **Huang, Y.** (2018). Finite horizon continuous-time Markov decision processes with mean and variance criteria, *Discrete Event Dyn. Sys.,* 28, 539–564.

[29] **Kitaev, M.Y.** (1986). Semi-Markov and jump Markov controlled models: Average cost criterion, T*heory Probab. Appl.*, 30, 272-288.

[30] **Kitaev, M.Y.,** & **Rykov, V.V.** (1995). *Controlled Queueing Systems*, CRC Press, Boca Raton.

[31] **Kumar, K.S.,** & **Pal, C.** (2013). Risk-sensitive control of jump process on denumerable state space with near monotone cost, *Appl. Math. Optim.,* 68, 311-331.

[32] **Kumar, K.S.,** & **Pal, C.** (2015). Risk-sensitive control of continuous-time Markov processes with denumerable state space, *Stoch. Anal. Appl.,* 33, 863-881.

[33] **Nowak, A.S.** (1985). Measurable selection theorems for minimax stochastic optimization problems, *SIAM J. Control Optim.,* 23, 466-476.

[34] **Pal, C.,** & **Pradhan, S.** (2019). Risk-sensitive control of pure jump processes on a general state space, *Stochastics,* 91, 155-174.

[35] **Piunovsiy, A.,** & **Zhang, Z.** (2011). Discounted continuous-time Markov decision processes with unbounded rates: The convex analytic approach, *SIAM J. Control Optim.,* 49, 2032-2061.

[36] **Piunovsiy, A.,** & **Zhang, Y.** (2014). Discounted continuous-time Markov decision processes with unbounded rates and randomized history-dependent policies: the dynamic programming approach, *4OR-Q. J. Oper. Res.,* 12, 49–75.

[37] **Piunovsiy, A.,** & **Zhang, Y.** (2020). Continuous-time Markov decision processes. In: Probability Theory and Stochastic Modelling, *Springer*, https://doi.org/10.1007/978-3-030-54987-9.

[38] **Wei, Q.** (2016). Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion, *Math. Meth. Oper. Res.*, 84, 461-487.

[39] **Wei, Q.** (2017). Zero-sum games for continuous-time Markov jump processes with risk-sensitive finite-horizon cost criterion, *Oper. Res. Lett.*, 46, 69-75.

[40] **Wei, Q.** (2019). Nonzero-sum risk-sensitive finite-horizon continuous-time stochastic games, *Stat. Probab. Lett.*, 147, 96-104.

[41] **Wei, Q.,** & **Chen, X.** (2016). Stochastic games for continuous-time jump processes under finite-horizon payoff criterion, *Appl. Math. Optim.*, 74, 273-301.

[42] **Zhang, W.Z.,** & **Guo, X.P.** (2012). Nonzero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates, *Sci. China Math.*, 55, 2405-2416.

[43] **Zhang, Y.** (2017). Continuous-time Markov decision processes with exponential utility, *SIAM J. Control Optim.*, 55, 2636-2660.