



Zero-sum infinite-horizon discounted piecewise deterministic Markov games

Yonghui Huang¹ · Zhaotong Lian² · Xianping Guo¹

Received: 11 September 2021 / Revised: 22 September 2022 / Accepted: 22 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

This paper is devoted to zero-sum piecewise deterministic Markov games with Borel state and action spaces, where the expected infinite-horizon discounted payoff criterion is considered. Both the transition rate and payoff rate are allowed to be unbounded. The policies of the two players are history-dependent, and the controls continuously act on the transition rate and the payoff rate. Under suitable conditions, Dynkin's formula and the comparison theorem are developed in our setup, via which the game is shown to have the value function as the unique solution to the associated Shapley equation. By the Shapley equation in the form of a differential equation, we establish the existence of a saddle point with a very simple form, which only depends on the current state and can be applied at any time. A potential algorithm for computing saddle points is proposed.

Keywords Piecewise deterministic Markov games · Zero sum · Unbounded transition rate · Shapley equation · Saddle point

Mathematics Subject Classification 91A15 · 91A25

✉ Yonghui Huang
hyongh5@mail.sysu.edu.cn
Zhaotong Lian
lianzt@um.edu.mo
Xianping Guo
mcsqxp@mail.sysu.edu.cn

¹ School of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China

² Faculty of Business Administration, University of Macau, Macau SAR, China

1 Introduction

As is well known, game theory is a mathematical framework to analyze social situations among competing players and produce satisfactory decision-making for competing players. Game theory has a wide range of applications such as psychology, evolutionary biology, politics, social sciences, economics and business. Since John Nash received the Nobel Prize in 1994, twenty game theorists have been awarded the Nobel Prize in Economic Sciences. Nowadays, there have been many advances along with plenty of branches in game theory; see, for instance, the monographs (Barron 2013; Haurie et al. 2012).

Markov games are one type of stochastic dynamic games, where the state dynamics of the games are driven by Markov processes. To date, Markov games have received increasing attentions and have been widely investigated; see, for instance, Minjárez-Sosa (2020) for discrete-time Markov games (DTMGs) with discounted criterion; Gensbittel and Renault (2015), Minjárez-Sosa (2020) for DTMGs with long-run average criterion; Guo and Zhang (2017) for continuous-time (pure jump) Markov games (CTMGs) with finite-horizon payoff criterion; Guo and Hernández-Lerma (2007), Prieto-Rumeau and Lorenzo (2015) for CTMGs with discounted payoff criterion; Guo and Hernández-Lerma (2003), Lorenzo et al. (2015) for CTMGs with average payoff criterion; Jaśkiewicz (2009), Mondal (2017) for semi-Markov games (SMGs) with average criterion; Ghosh and Goswami (2006) for SMGs with discounted criterion; Costa and Dufour (2018) for piecewise deterministic Markov games (PDMGs) with discounted payoff criterion.

This paper is concerned with zero-sum PDMGs, one type of Markov games whose state dynamics are driven by piecewise deterministic Markov processes (PDMPs). PDMPs evolve through random jumps at random time points while the motion between jumps follows a flow. In particular, if the flow remains unchanged over time, PDMPs reduce to continuous-time (pure jump) Markov processes (CTMPs). These features of PDMPs make them have wide applications in many areas such as management science, operations research, and engineering. There has been a vast literature on piecewise deterministic Markov decision processes (PDMDPs) where only one decision maker is considered (Bäuerle and Rieder 2011; Costa et al. 2016; Huang and Guo 2019). However, as far as we can tell, there is only one paper (Costa and Dufour 2018) devoted to PDMGs. In fact, Costa and Dufour (2018) deal with zero-sum PDMGs with the infinite horizon total expected discounted reward criterion. The transition rate and reward functions are assumed to be unbounded. The authors use the special features of the PDMPs to reformulate the problem as a discrete-stage zero-sum game problem. They derive conditions for the existence of min-max strategies.

In this paper, our problem and the assumptions on model data are similar to those in Costa and Dufour (2018). We also consider the expected infinite-horizon discounted payoff criterion. The state space is assumed to be a Borel space. Both the transition rate and payoff function are allowed to be unbounded. The policies of the two players are history-dependent, and the controls continuously act on the transition rate and the payoff rate. However, there are some differences between the work in Costa and Dufour (2018) and ours. First, the decisions for the two players in Costa and Dufour (2018) are taken only after a jump time, while the transition rate, reward rate and boundary reward

are affected until the next jump occurs through a predefined mapping $l(x, a, b, t)$, where x is the state at the beginning of the current jump, (a, b) is the action pair chosen by players, and t is the time elapsed since this jump occurred. In our work, the decisions for the two players are taken continuously over time, and these actions will be continuously acting on the transition rate and the payoff rate. Second, our approach to dealing with the problem is different from the reduction technique in Costa and Dufour (2018). In fact, we use a so-called infinitesimal approach, which characterizes the value function as a solution to some differential equation; see Costa et al. (2016) for infinite-horizon discounted PDMDPs with bounded jump rates and Huang and Guo (2019) for finite-horizon PDMDPs with unbounded jump rates. In more detail, we develop Dynkin's formula and the comparison theorem in our setup, and then show that the game has the value function as the unique solution to the Shapley equation that is in the form of a differential equation. As far as we know, this paper is the first attempt to apply the infinitesimal approach to studying PDMGs. Third, because our approach is different from the one in Costa and Dufour (2018), some assumptions in this paper are different from those in Costa and Dufour (2018); see Remarks 2 and 3 for details. We also provide a simple example to verify our assumptions. Finally, by the Shapley equation derived in this paper, we establish the existence of saddle points, and propose a potential algorithm for computing a saddle point. We mention that, since the Shapley equation in this paper is in the form of a differential equation, the saddle point we obtain is in a very simple form, which only depends on the current state and can be applied at any time.

The rest of the paper is organized as follows. Section 2 describes the model of PDMGs and the problem formulation. Section 3 provides some preliminaries such as Dynkin's formula and the comparison theorem. The main results on the Shapley equation and the existence of saddle points are given in Sect. 4. Section 5 provides an example that verifies all the assumptions in this paper. An appendix about the proof of Proposition 1 is included in Sect. 6.

2 Problem formulation

Notation. Let $R = (-\infty, \infty)$, and $R_+ = [0, \infty)$. For a Borel space X , we denote by D^c the complement of a set $D \subseteq X$, by $\delta_{\{x\}}(\cdot)$ the Dirac measure concentrated on $x \in X$, by $\mathbb{1}_D(x)$ the indicator function on a set $D \subseteq X$, by $\mathcal{P}(X)$ the family of all non-empty subsets of X , by $\mathcal{B}(X)$ the Borel σ -algebra on X , and by $P(X)$ the space of all probability measures on $(X, \mathcal{B}(X))$.

The model of a zero-sum PDMG is a tuple as below:

$$\{E, A, B, \{A(x), B(x), x \in E\}, q(\cdot|x, a, b), \phi(x, t), r(x, a, b)\}. \quad (1)$$

Here, E is the state space of the PDMP, while A and B are the action space for player 1 and player 2, respectively. These spaces are all assumed to be Borel spaces endowed with Borel σ -algebras. $A(\cdot)$ and $B(\cdot)$ are *measurable compact-valued* multi-functions from E to $\mathcal{P}(A)$ and $\mathcal{P}(B)$, respectively. For each $x \in E$, $A(x)$ and $B(x)$ denote the sets of available actions to player 1 and player 2, respectively, when the system is at

the state $x \in E$. For convenience, let's introduce the set

$$K := \{(x, a, b) : x \in E, a \in A(x), b \in B(x)\}.$$

That is, K is the graph of the multi-function $A(\cdot) \times B(\cdot)$. Since $A(\cdot)$ and $B(\cdot)$ are measurable compact-valued multi-functions, by Lemma 1.7 in Nowak (1984), K is a measurable subset of $E \times A \times B$. Moreover, $q(\cdot|x, a, b)$ is referred to a transition rate, which is a measurable signed kernel on E given K such that, for all $(x, a, b) \in K$, (i) $0 \leq q(D|x, a, b) < +\infty$ for $x \notin D \in \mathcal{B}(E)$, (ii) $q(E|x, a, b) \equiv 0$, and (iii) $q^*(x) := \sup_{a \in A(x), b \in B(x)} q(x, a, b) < \infty$, where $q(x, a, b) := -q(\{x\}|x, a, b) \geq 0$. The motion between jumps of the PDMP is determined by $\phi(x, t)$ in (1), called a flow, which is a measurable function from $E \times R$ to E . We assume that $\phi(x, s + t) = \phi(\phi(x, s), t)$ for all $x \in E$ and $(s, t) \in R^2$. Finally, the measurable function $r(x, a, b)$ on K denotes the payoff rate for player 1. In a zero-sum game, one player's gain is equivalent to another's loss, so that $r(x, a, b)$ is the loss rate for player 2.

Remark 1 Unlike the models of PDMPs in Costa et al. (2016), Costa and Dufour (2018), we do not consider jumps when hitting the boundary and the related impulsive control in our model. In some cases, there are actually no boundary jumps since the boundary will be never reached, as indicated in the example in Sect. 5 below.

To construct the PDMP based on the data above, let $E_\Delta := E \cup \{x_\infty\}$, where x_∞ is an isolated artificial point corresponding to the case when no jump occurs in the future. For $n \geq 0$, we put $\Omega_n = E \times ((0, \infty) \times E)^n \times (\{\infty\} \times \{x_\infty\})^\infty$. The sample space is $\Omega = \cup_{n=0}^\infty \Omega_n \cup (E \times ((0, \infty) \times E)^\infty)$, and let \mathcal{F} be the Borel σ -algebra on Ω . Then we obtain a measurable space (Ω, \mathcal{F}) . For a trajectory $\omega = (x_0, \theta_1, x_1, \dots, \theta_n, x_n, \dots) \in \Omega$, x_0 denotes the initial state of the process, and for $n \geq 1$, $\theta_n > 0$ and x_n correspond to the time interval between two consecutive jumps and the state of the process immediately after the jump. In case $\theta_n < \infty$ and $\theta_{n+1} = \infty$, the trajectory has only n jumps, and we put $\theta_m = \infty$ and $x_m = x_\infty$ for all $m \geq n + 1$. On the measurable space (Ω, \mathcal{F}) , we define a sequence of random variables $\{\Theta_n, X_n, n \geq 0\}$ by $\Theta_0(\omega) := 0$, $\Theta_{n+1}(\omega) := \theta_{n+1}$, $X_n(\omega) := x_n$, for each $n \geq 0$ and any trajectory $\omega = (x_0, \theta_1, x_1, \dots, \theta_k, x_k, \dots) \in \Omega$. Further, we define $T_0(\omega) := 0$, $T_n(\omega) := \sum_{i=1}^n \Theta_i(\omega)$ for every $n \geq 1$, and $T_\infty(\omega) := \lim_{n \rightarrow \infty} T_n(\omega)$. Then, the PDMP $\{\xi_t, t \in R_+\}$ is defined by

$$\xi_t(\omega) := \begin{cases} \phi(X_n(\omega), t - T_n(\omega)), & \text{if } T_n(\omega) \leq t < T_{n+1}(\omega), \\ x_\infty, & \text{if } t \geq T_\infty(\omega), \end{cases}$$

for each $t \in R_+$. To be complete, when the PDMP occupies the state x_∞ , we introduce artificial actions a_∞ and b_∞ for player 1 and player 2, respectively. Moreover, let $q(\cdot|x_\infty, a_\infty, b_\infty) \equiv 0$, $A(x_\infty) = \{a_\infty\}$, $B(x_\infty) = \{b_\infty\}$, and $A_\infty = A \cup \{a_\infty\}$, $B_\infty = B \cup \{b_\infty\}$.

The probabilistic properties of the PDMP $\{\xi_t, t \in R_+\}$ are determined by the transition rate and policies of players. First of all, we need to define policies. To

this end, we introduce the random measure μ associated with $\{\Theta_n, X_n, n \geq 0\}$ on $(R_+ \times E, \mathcal{B}(R_+ \times E))$ by

$$\mu(\omega; dt, dx) := \sum_{n \geq 1} \mathbb{1}_{\{T_n(\omega) < \infty\}} \delta_{\{T_n(\omega), X_n(\omega)\}}(dt, dx).$$

Moreover, we take the right-continuous family of σ -algebras $\{\mathcal{F}_t\}_{t \geq 0}$ with $\mathcal{F}_t := \sigma(\mu([0, s] \times D) : 0 \leq s \leq t, D \in \mathcal{B}(E))$, and let $\mathcal{P} := \sigma(\{\Gamma \times \{0\}, \Gamma \in \mathcal{F}_0\} \cup \{\Gamma \times (s, \infty), \Gamma \in \mathcal{F}_{s-}, s > 0\})$ be the σ -algebra of predictable sets on $\Omega \times R_+$ related to $\{\mathcal{F}_t\}_{t \geq 0}$, where $\mathcal{F}_{s-} := \vee_{t < s} \mathcal{F}_t$.

Definition 1 A randomized history-dependent policy, or simply, a policy for player 1 is a transition probability $\pi^1(da|\omega, t)$ from $(\Omega \times R_+, \mathcal{P})$ onto $(A_\infty, \mathcal{B}(A_\infty))$ such that $\pi^1(A(\xi_{t-}(\omega))|\omega, t) = 1$. In particular, a policy $\pi^1(da|\omega, t)$ is called randomized Markov for player 1 if the action selection depends on the history only through the current state, i.e., the policy has the form $\pi^1(da|\xi_{t-}(\omega))$. Policies for player 2 can be defined similarly.

For each $i = 1, 2$, we denote by Π^i and Π_{RS}^i the families of all policies and that of all randomized Markov policies for player i , respectively.

Now consider a state $x \in E$ and a pair of policies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$. By Theorem 3.6 in Jacod (1975), there exists a probability measure $\mathbb{P}_x^{\pi^1, \pi^2}$ on (Ω, \mathcal{F}) such that the restriction of $\mathbb{P}_x^{\pi^1, \pi^2}$ on (Ω, \mathcal{F}_0) is given by

$$\mathbb{P}_x^{\pi^1, \pi^2}(X_0 = x) = 1,$$

and the random measure ν^{π^1, π^2} defined on $(0, \infty) \times E$ by

$$\begin{aligned} \nu^{\pi^1, \pi^2}(\omega; dt, dx) &:= \int_B \int_A q(dx \setminus \{\xi_{t-}(\omega)\} | \xi_{t-}(\omega), a, b) \pi^1(da|\omega, t) \pi^2(db|\omega, t) dt. \end{aligned}$$

is the predictable projection of μ with respect to $\mathbb{P}_x^{\pi^1, \pi^2}$; see Costa et al. (2016) for further details.

Let $\mathbb{E}_x^{\pi^1, \pi^2}$ be the corresponding expectation operator with respect to $\mathbb{P}_x^{\pi^1, \pi^2}$. For each pair of policies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, we define the expected infinite-horizon discounted payoff criterion by

$$\begin{aligned} V^{\pi^1, \pi^2}(x) &:= \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^\infty e^{-\alpha t} \int_B \int_A r(\xi_t, a, b) \pi^1(da|\omega, t) \pi^2(db|\omega, t) dt \right] \forall x \in E, \end{aligned}$$

provided that the integral is well defined. Here, α is a discount factor.

As is well known, the functions $\underline{V}^*(x)$ and $\overline{V}^*(x)$ on E defined by

$$\underline{V}^*(x) := \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} V^{\pi^1, \pi^2}(x),$$

and $\overline{V}^*(x) := \inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} V^{\pi^1, \pi^2}(x), x \in E,$

are called the *lower value* and the *upper value* of the game, respectively. It is clear that $\underline{V}^*(x) \leq \overline{V}^*(x)$ for all $x \in E$.

Definition 2 If $\underline{V}^*(x) = \overline{V}^*(x)$ for all $x \in E$, the common function is called the value function of the game and denoted by $V^*(x)$.

Definition 3 A pair of policies $(\hat{\pi}^1, \hat{\pi}^2)$ is called a saddle-point if

$$V^{\pi^1, \hat{\pi}^2}(x) \leq V^{\hat{\pi}^1, \hat{\pi}^2}(x) \leq V^{\hat{\pi}^1, \pi^2}(x) \quad \forall \pi^1 \in \Pi^1, \pi^2 \in \Pi^2, x \in E.$$

At a saddle point, no player can improve (reduce) his/her payoff (loss) by deviating unilaterally from the saddle-point policy. Note that if a saddle point $(\hat{\pi}^1, \hat{\pi}^2)$ exists, we must have

$$\overline{V}^*(x) \leq V^{\hat{\pi}^1, \hat{\pi}^2}(x) \leq \underline{V}^*(x) \quad \forall x \in E,$$

which implies that $V^{\hat{\pi}^1, \hat{\pi}^2}(x) = V^*(x)$ for all $x \in E$, meaning that the game has a value. In the following sections, we show when a saddle point exists and how to find a saddle point for the game.

3 Preliminaries

Since the transition rate is unbounded, we shall first avoid the explosion of the PDMP $\{\xi_t\}$, and so we propose the following assumption.

Assumption 1 (a) There exist a measurable function $w_0 \geq 1$ on E , and constants $c_0 > 0, d_0 \geq 0$ such that $\int_E w_0(y)q(dy|x, a, b) \leq c_0 w_0(x) + d_0$, for all $(x, a, b) \in K$; (b) There exists a sequence $\{E_m, m \geq 1\}$ of Borel subsets of E such that $E_m \uparrow E$, $\sup_{x \in E_m} q^*(x) < \infty$, and $\lim_{m \rightarrow \infty} \inf_{x \notin E_m} w_0(x) = \infty$; (c) $w_0(\phi(x, t)) \leq w_0(x)$ for all $(x, t) \in E \times R_+$.

Remark 2 (a) Assumption 1 is inspired by Assumption A in Guo and Song (2011) for continuous-time Markov decision processes (CTMDPs) and Assumption 3.1 in Guo and Hernández-Lerma (2007) for CTMGs. However, due to the presence of the flow ϕ in PDMPs, we add Assumption 1(c) additionally to ensure the non-explosion of PDMPs. Moreover, Assumption 1 here is similar to Assumption 3.1 in Huang and Guo (2019) for finite-horizon PDMDPs, but the two assumptions slightly differ.

(b) The parallel assumption in Costa and Dufour (2018) for PDMGs to Assumption 1 here is Assumption B therein, but it is different from ours owing to different techniques applied, mentioned in the introduction above.

Proposition 1 *Let Assumption 1 be fulfilled. For each $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $x \in E$, and $t \in R_+$, the following assertions hold.*

- (a) $\mathbb{P}_x^{\pi^1, \pi^2}(T_\infty = +\infty) = 1.$
- (b) $\mathbb{E}_x^{\pi^1, \pi^2}[w_0(\xi_t)] \leq e^{c_0 t} w_0(x) + \frac{d_0}{c_0}(e^{c_0 t} - 1).$

Proof The proof is postponed to Sect. 6. □

To ensure the finiteness of V^{π^1, π^2} , we also propose some growth conditions on the payoff rate and the discount factor.

Assumption 2 (a) The constant c_0 in Assumption 1 satisfies that $c_0 < \alpha$.
 (b) There exists a constant $M_0 > 0$ such that

$$|r(x, a, b)| \leq M_0 w_0(x) \quad \forall (x, a, b) \in K.$$

We can now show that V^{π^1, π^2} is finite for each pair of policies (π^1, π^2) .

Lemma 1 *Under Assumptions 1 and 2, we have*

$$|V^{\pi^1, \pi^2}(x)| \leq \frac{M_0}{\alpha - c_0} w_0(x) + \frac{d_0 M_0}{\alpha(\alpha - c_0)} \quad \forall (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2, x \in E.$$

Proof Under Assumptions 1 and 2, it follows from Proposition 1(b) that

$$|V^{\pi^1, \pi^2}(x)| \leq M_0 \int_0^\infty e^{-\alpha t} \mathbb{E}_x^{\pi^1, \pi^2}[w_0(\xi_t)] dt \leq \frac{M_0}{\alpha - c_0} w_0(x) + \frac{d_0 M_0}{\alpha(\alpha - c_0)}.$$

The proof is complete. □

We are going to derive Dynkin’s formula for discounted PDMGs with unbounded transition rates. To this end, we introduce a framework. Let $w \geq 1$ be a real-valued measurable function on E , called a weight function. For every measurable function φ on E , we introduce its w -norm $\|\varphi\|_w$ by

$$\|\varphi\|_w := \sup_{x \in E} |\varphi(x)|/w(x).$$

Let $\mathbb{B}_w(E)$ be the Banach space of all measurable functions φ on E such that $\|\varphi\|_w < \infty$, while let $\mathbb{B}_w^{ac}(E)$ be the collection of all measurable functions φ in $\mathbb{B}_w(E)$ such that $\varphi(\phi(x, t))$ is absolutely continuous in $t \in R_+$ for all $x \in E$. For a function $\varphi \in \mathbb{B}_w^{ac}(E)$, by Lemma A.1 in Piunovskiy and Zhang (2021), there is some measurable function $L^\phi \varphi$ on E satisfying

$$\varphi(\phi(x, t)) - \varphi(x) = \int_0^t L^\phi \varphi(\phi(x, v)) dv \quad \forall t \in R_+, x \in E.$$

Here, for $x \in E$, the function $L^\phi \varphi(\phi(x, v))$ on R_+ coincides with the partial derivative, $\partial \varphi(\phi(x, v))/\partial v$, of the function $\varphi(\phi(x, v))$ in $v \in R_+$ apart from on a null set $Z_\varphi(x) \subset R_+$ with respect to the Lebesgue measure. For such a function φ , let

$$\mathcal{D}^\varphi := \{\phi(x, t) \in E : t \in Z_\varphi^c(x), x \in E\}.$$

Then, the function $L^\phi \varphi$ on E can be defined as below:

$$L^\phi \varphi(x) := \begin{cases} \lim_{\Delta s \rightarrow 0} \frac{\varphi(\phi(x, \Delta s)) - \varphi(x)}{\Delta s}, & x \in \mathcal{D}^\varphi, \\ \text{arbitrary,} & \text{otherwise.} \end{cases}$$

In particular, if $\phi(x, t) \equiv x$ in which case a PDMP becomes a CTMP, $L^\phi \varphi(x) = 0$ for all $x \in E$. Moreover, for $\varphi \in \mathbb{B}_w^{ac}(E)$ and a weight function \bar{w} on E , we let

$$\|L^\phi \varphi\|_{\bar{w}}^{es} := \sup_{x \in \mathcal{D}^\varphi} |L^\phi \varphi(x)|/\bar{w}(x),$$

and $\mathbb{B}_{w, \bar{w}}^{ac}(E) := \{\varphi \in \mathbb{B}_w^{ac}(E) : \|L^\phi \varphi\|_{\bar{w}}^{es} < \infty\}$.

Some more assumptions are required to ensure Dynkin’s formula works.

Assumption 3 There exist a measurable function $w_1 \geq 1$ on E , and constants $M_1 > 0$, $c_1 < \alpha$ and $d_1 \geq 0$ such that

- (a) $(1 + q(x, a, b))w_0(x) \leq M_1 w_1(x)$, for all $(x, a, b) \in K$;
- (b) $\int_E w_1(y)q(dy|x, a, b) \leq c_1 w_1(x) + d_1$ for all $(x, a, b) \in K$;
- (c) $w_1(\phi(x, t)) \leq w_1(x)$ for all $(x, t) \in E \times R_+$.

Remark 3 The parallel assumption in Costa and Dufour (2018) for PDMGs to Assumption 3 here is Assumption D therein, but it is different from ours. Assumption D together with Assumption B and C therein are used to ensure the convergence of the expected infinite horizon discounted payoffs, while Assumption 3 together with Assumption 1 here are used to justify Dynkin’s formula in our setup.

When Assumptions 3(b) and 3(c) are further imposed, Proposition 1(b) holds with w_0 being replaced by w_1 so that we have

$$\mathbb{E}_x^{\pi^1, \pi^2} [w_1(\xi_t)] \leq e^{c_1 t} w_1(x) + \frac{d_1}{c_1} (e^{c_1 t} - 1). \tag{2}$$

We are ready to state Dynkin’s formula for discounted PDMGs.

Theorem 2 (Dynkin’s formula) Suppose Assumptions 1, 2(a) and 3 are satisfied. For each $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $\varphi \in \mathbb{B}_{w_0, w_1}^{ac}(E)$, $T \in R_+$ and $x \in E$,

$$\mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T (L^\phi [e^{-\alpha t} \varphi(\xi_t)]) \right]$$

$$\begin{aligned}
 & + e^{-\alpha t} \int_B \int_A \int_E \varphi(y)q(dy|\xi_t, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) dt \Big] \\
 & = \mathbb{E}_x^{\pi^1, \pi^2} [e^{-\alpha T} \varphi(\xi_T)] - \varphi(x).
 \end{aligned}$$

Proof It follows from Proposition 1(a) that, for almost all (a.a.) $\omega \in \Omega$ with respect to $\mathbb{P}_x^{\pi^1, \pi^2}$, there are only finite number of jumps of the PDMP $\{\xi_t(\omega)\}$ up to any time T . Therefore, by the construction of ξ_t and the definition of \mathcal{D}^φ , we conclude that for a.a. $t \in R_+$, $\xi_t(\omega) \in \mathcal{D}^\varphi$. Hence, by the definition of the operator L^ϕ , we have

$$L^\phi [e^{-\alpha t} \varphi(\xi_t(\omega))] = \frac{\partial [e^{-\alpha t} \varphi(\xi_t(\omega))]}{\partial t}, \quad \text{a.a. } \omega \in \Omega, t \in R_+.$$

Thus, using the equality (8) in Avrachenkov et al. (2015) yields that

$$\begin{aligned}
 & e^{-\alpha T} \varphi(\xi_T) - \varphi(\xi_0) \\
 & = \int_0^T L^\phi [e^{-\alpha t} \varphi(\xi_t)] dt \\
 & + \int_{(0, T] \times E} [e^{-\alpha t} \varphi(y) - e^{-\alpha t} \varphi(\xi_{t-})] \mu(\omega; dt, dy), \quad \text{a.a. } \omega - \mathbb{P}_x^{\pi^1, \pi^2}. \tag{3}
 \end{aligned}$$

We are going to take expectations in both sides of (3) to derive Dynkin’s formula, but, we should first discuss if the expectations are well defined, as shown below.

Since $\varphi \in \mathbb{B}_{w_0, w_1}^{\alpha c}(E)$, it is clear that $|\varphi(x)| \leq \|\varphi\|_{w_0} w_0(x)$ for all $x \in E$, and $|L^\phi \varphi(x)| \leq \|L^\phi \varphi\|_{w_1}^{es} w_1(x)$ for any $x \in \mathcal{D}^\varphi$. On the one hand, observe that $(\xi_t(\omega)) \in \mathcal{D}^\varphi$ for a.a. $\omega \in \Omega$ and $t \in R_+$, and so we have

$$L^\phi [e^{-\alpha t} \varphi(\xi_t(\omega))] = -\alpha e^{-\alpha t} \varphi(\xi_t(\omega)) + e^{-\alpha t} L^\phi [\varphi(\xi_t(\omega))], \quad \text{a.a. } \omega \in \Omega, t \in R_+,$$

which together with Proposition 1(b) and (2) give

$$\begin{aligned}
 & \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T |L^\phi [e^{-\alpha t} \varphi(\xi_t)]| dt \right] \\
 & \leq \int_0^T \mathbb{E}_x^{\pi^1, \pi^2} [\alpha e^{-\alpha t} \|\varphi\|_{w_0} w_0(\xi_t) + e^{-\alpha t} \|L^\phi \varphi\|_{w_1}^{es} w_1(\xi_t)] dt \\
 & \leq \frac{\alpha \|\varphi\|_{w_0}}{\alpha - c_0} (w_0(x) + \frac{d_0}{c_0}) + \frac{\|L^\phi \varphi\|_{w_1}^{es}}{\alpha - c_1} (w_1(x) + \frac{d_1}{c_1}) \\
 & < \infty. \tag{4}
 \end{aligned}$$

On the other hand, under Assumptions 1 and 3, we have

$$\begin{aligned}
 & \left| \int_B \int_A \int_E \varphi(y)q(dy|\xi_t, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) \right| \\
 & \leq \|\varphi\|_{w_0} \int_B \int_A \left[\int_E w_0(y)q(dy|\xi_t, a, , b) + 2q(\xi_t, a, b)w_0(\xi_t) \right]
 \end{aligned}$$

$$\begin{aligned} & \pi^1(da|\omega, t)\pi^2(db|\omega, t) \\ & \leq \|\varphi\|_{w_0} \left[c_0 w_0(\xi_t) + d_0 + 2M_1 w_1(\xi_t) \right]. \end{aligned} \tag{5}$$

Thus, by (5), Proposition 1(b) and (2), we obtain

$$\begin{aligned} & \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T e^{-\alpha t} \left| \int_B \int_A \int_E \varphi(y)q(dy|\xi_t, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) \right| dt \right] \\ & \leq \|\varphi\|_{w_0} \int_0^T e^{-\alpha t} \mathbb{E}_x^{\pi^1, \pi^2} \left[c_0 w_0(\xi_t) + d_0 + 2M_1 w_1(\xi_t) \right] dt \\ & \leq \|\varphi\|_{w_0} \left[\frac{c_0}{\alpha - c_0} (w_0(x) + \frac{d_0}{c_0}) + \frac{d_0}{\alpha} + \frac{2M_1}{\alpha - c_1} (w_1(x) + \frac{d_1}{c_1}) \right] \\ & < \infty. \end{aligned} \tag{6}$$

Now, taking expectation in both sides of (3), using (4) and (6) yields that

$$\begin{aligned} & \mathbb{E}_x^{\pi^1, \pi^2} \left[e^{-\alpha T} \varphi(\xi_T(\omega)) \right] - \varphi(x) \\ & = \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T L^\phi \left[e^{-\alpha t} \varphi(\xi_t) \right] dt \right] \\ & \quad + \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_E \int_{(0, T]} \left[e^{-\alpha t} \varphi(y) - e^{-\alpha t} \varphi(\xi_{t-}) \right] \mu(dt, dy) \right] \\ & = \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T L^\phi \left[e^{-\alpha t} \varphi(\xi_t) \right] dt \right] \\ & \quad + \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_E \int_{(0, T]} e^{-\alpha t} (\varphi(y) - \varphi(\xi_{t-})) v^{\pi^1, \pi^2}(dt, dy) \right] \\ & = \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T L^\phi \left[e^{-\alpha t} \varphi(\xi_t) \right] dt \right] \\ & \quad + \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T e^{-\alpha t} \int_B \int_A \int_E \varphi(y)q(dy|\xi_{t-}, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) dt \right], \end{aligned}$$

where the second equality follows from the fact that the random measure v^{π^1, π^2} is a dual predictable projection of the random measure μ under $\mathbb{P}_x^{\pi^1, \pi^2}$. Note that, for every $\omega \in \Omega$, $\xi_{t-}(\omega) = \xi_t(\omega)$ on $(0, T]$ except countable many time points. Hence,

$$\begin{aligned} & \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T e^{-\alpha t} \int_B \int_A \int_E \varphi(y)q(dy|\xi_{t-}, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) dt \right] \\ & = \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T e^{-\alpha t} \int_B \int_A \int_E \varphi(y)q(dy|\xi_t, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) dt \right]. \end{aligned}$$

The formula then follows. □

Remark 4 Costa et al. (2016) derive Dynkin’s formula for infinite-horizon discounted PDMDPs, where bounded transition rate and boundary jumps are considered.

Now, using Dynkin’s formula, we can compare V^{π^1, π^2} with the solution to some differential inequality or equation. To begin with, we introduce the following notations: for each $x \in E$, $(\eta, \gamma) \in P(A(x)) \times P(B(x))$ and $\varphi \in \mathbb{B}_{w_0}(E)$,

$$r(x, \eta, \gamma) := \int_{B(x)} \int_{A(x)} r(x, a, b)\eta(da)\gamma(db), \tag{7}$$

$$\int_E \varphi(y)q(dy|x, \eta, \gamma) := \int_{B(x)} \int_{A(x)} \int_E \varphi(y)q(dy|x, a, b)\eta(da)\gamma(db). \tag{8}$$

Theorem 3 (Comparison Theorem) *Suppose that Assumptions 1–3 hold.*

(a) *If there exists a function $\varphi \in \mathbb{B}_{w_0, w_1}^{ac}(E)$ such that*

$$L^\phi \varphi(x) - \alpha\varphi(x) + r(x, \eta, \gamma) + \int_E \varphi(y)q(dy|x, \eta, \gamma) \leq 0 \quad \forall (\eta, \gamma) \in P(A(x)) \times P(B(x)), x \in \mathcal{D}^\phi,$$

we have $V^{\pi^1, \pi^2}(x) \leq \varphi(x)$, for all $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and $x \in E$.

(b) *If there exists a function $\varphi \in \mathbb{B}_{w_0, w_1}^{ac}(E)$ such that*

$$L^\phi \varphi(x) - \alpha\varphi(x) + r(x, \eta, \gamma) + \int_E \varphi(y)q(dy|x, \eta, \gamma) \geq 0 \quad \forall (\eta, \gamma) \in P(A(x)) \times P(B(x)), x \in \mathcal{D}^\phi,$$

we have $V^{\pi^1, \pi^2}(x) \geq \varphi(x)$, for all $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and $x \in E$.

Proof (a) As in the proof of Theorem 2, we conclude that, for a.a. $\omega \in \Omega$ with respect to $\mathbb{P}_x^{\pi^1, \pi^2}$ and a.a. $t \in R_+$ with respect to Lebesgue measure, $\xi_t(\omega) \in \mathcal{D}^\phi$, which together with the condition in (a) implies that, for a.a. $\omega \in \Omega$ and $t \in R_+$,

$$L^\phi \varphi(\xi_t(\omega)) - \alpha\varphi(\xi_t(\omega)) + \int_B \int_A r(\xi_t(\omega), a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) + \int_B \int_A \int_E \varphi(y)q(dy|\xi_t, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) \leq 0,$$

which yields that

$$L^\phi \left[e^{-\alpha t} \varphi(\xi_t(\omega)) \right] + e^{-\alpha t} \int_B \int_A r(\xi_t(\omega), a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) + e^{-\alpha t} \int_B \int_A \int_E \varphi(y)q(dy|\xi_t, a, b)\pi^1(da|\omega, t)\pi^2(db|\omega, t) \leq 0.$$

Thus, by Theorem 2, for all $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and $x \in E$, we have

$$\begin{aligned} & \mathbb{E}_x^{\pi^1, \pi^2} [e^{-\alpha T} \varphi(\xi_T)] - \varphi(x) \\ &= \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T \left(L^\phi [e^{\alpha t} \varphi(\xi_t)] \right. \right. \\ & \quad \left. \left. + e^{\alpha t} \int_B \int_A \int_E \varphi(y) q(dy | \xi_t, a, b) \pi^1(da | \omega, t) \pi^2(db | \omega, t) \right) dt \right] \\ & \leq -\mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T e^{-\alpha t} \int_B \int_A r(\xi_t, a, b) \pi^1(da | \omega, t) \pi^2(db | \omega, t) dt \right], \end{aligned}$$

which indicates that

$$\begin{aligned} & \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^T e^{-\alpha t} \int_B \int_A r(\xi_t, a, b) \pi^1(da | \omega, t) \pi^2(db | \omega, t) dt \right] \\ & \leq \varphi(x) - \mathbb{E}_x^{\pi^1, \pi^2} [e^{-\alpha T} \varphi(\xi_T)]. \end{aligned}$$

Letting $T \rightarrow \infty$ on both sides of the above inequality, the dominated convergence theorem yields that $V^{\pi^1, \pi^2}(x) \leq \varphi(x)$.

(b) The proof is similar to that of part (a). □

Theorem 4 *Suppose Assumptions 1–3 hold. For every $(\pi^1, \pi^2) \in \Pi_{RS}^1 \times \Pi_{RS}^2$, V^{π^1, π^2} is the unique solution solution in $\mathbb{B}_{w_0, w_1}^{ac}(E)$ to the differential equation*

$$\begin{aligned} & L^\phi \varphi(x) - \alpha \varphi(x) + r(x, \pi^1(\cdot | x), \pi^2(\cdot | x)) \\ & \quad + \int_E \varphi(y) q(dy | x, \pi^1(\cdot | x), \pi^2(\cdot | x)) = 0 \quad \forall x \in \mathcal{D}^\phi. \end{aligned} \tag{9}$$

Proof From Lemma 1, we see that $V^{\pi^1, \pi^2} \in \mathbb{B}_{w_0}(E)$. Now, for each $x \in E$, conditioning on the first-jump time and the post-jump state, we have

$$\begin{aligned} & V^{\pi^1, \pi^2}(x) \\ &= \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_0^{T_1} e^{-\alpha t} \int_B \int_A r(\xi_t, a, b) \pi^1(da | \xi_{t-}) \pi^2(db | \xi_{t-}) dt \right] \\ & \quad + \mathbb{E}_x^{\pi^1, \pi^2} \left[e^{-\alpha T_1} \int_{T_1}^\infty e^{-\alpha(t-T_1)} \int_B \int_A r(\xi_t, a, b) \pi^1(da | \xi_{t-}) \pi^2(db | \xi_{t-}) dt \right] \\ &= \int_0^\infty e^{-\alpha t} \mathbb{P}_x^{\pi^1, \pi^2}(T_1 > t) r(\phi(x, t), \pi^1(\cdot | \phi(x, t)), \pi^2(\cdot | \phi(x, t))) dt \\ & \quad + \mathbb{E}_x^{\pi^1, \pi^2} \left[e^{-\alpha T_1} \mathbb{E}_x^{\pi^1, \pi^2} \left[\int_{T_1}^\infty e^{-\alpha(t-T_1)} \right. \right. \\ & \quad \left. \left. \times \left[\int_B \int_A r(\xi_t, a, b) \pi^1(da | \xi_{t-}) \pi^2(db | \xi_{t-}) \right] dt \middle| T_1, X_1 \right] \right] \end{aligned}$$

$$\begin{aligned}
 &= \int_0^\infty e^{-\alpha t - \int_0^t q(\phi(x, v), \pi^1(\cdot|\phi(x, v)), \pi^2(\cdot|\phi(x, v)))dv} \\
 &\quad \times \left[r(\phi(x, t), \pi^1(\cdot|\phi(x, t)), \pi^2(\cdot|\phi(x, t))) \right. \\
 &\quad \left. + \int_{E \setminus \{\phi(x, t)\}} V^{\pi^1, \pi^2}(y) q(dy|\phi(x, t), \pi^1(\cdot|\phi(x, t)), \pi^2(\cdot|\phi(x, t))) \right] dt. \tag{10}
 \end{aligned}$$

For all $s \in R_+$, replacing x with $\phi(x, s)$ in (10), we obtain

$$\begin{aligned}
 &V^{\pi^1, \pi^2}(\phi(x, s)) \\
 &= \int_0^\infty e^{-\alpha t} e^{-\int_0^t q(\phi(x, v+s), \pi^1(\cdot|\phi(x, t+v)), \pi^2(\cdot|\phi(x, t+v)))dv} \\
 &\quad \times \left[r(\phi(x, t+s), \pi^1(\cdot|\phi(x, t+s)), \pi^2(\cdot|\phi(x, t+s))) \right. \\
 &\quad \left. + \int_{E \setminus \{\phi(x, t+s)\}} V^{\pi^1, \pi^2}(y) \right. \\
 &\quad \left. \cdot q(dy|\phi(x, t+s), \pi^1(\cdot|\phi(x, t+s)), \pi^2(\cdot|\phi(x, t+s))) \right] dt. \tag{11}
 \end{aligned}$$

Now, changing the integral variable t to u with $u = t + s$ in the R.H.S. of (11), and then multiplying by $e^{-\int_0^s (\alpha + q(\phi(x, v), \pi^1(\cdot|\phi(x, v)), \pi^2(\cdot|\phi(x, v))))dv}$ in both sides of (11), we find that

$$\begin{aligned}
 &e^{-\int_0^s (\alpha + q(\phi(x, v), \pi^1(\cdot|\phi(x, v)), \pi^2(\cdot|\phi(x, v))))dv} V^{\pi^1, \pi^2}(\phi(x, s)) \\
 &= \int_s^\infty e^{-\int_0^u (\alpha + q(\phi(x, v), \pi^1(\cdot|\phi(x, v)), \pi^2(\cdot|\phi(x, v))))dv} \\
 &\quad \times \left[r(\phi(x, u), \pi^1(\cdot|\phi(x, u)), \pi^2(\cdot|\phi(x, u))) \right. \\
 &\quad \left. + \int_{E \setminus \{\phi(x, u)\}} V^{\pi^1, \pi^2}(y) q(dy|\phi(x, u), \pi^1(\cdot|\phi(x, u)), \pi^2(\cdot|\phi(x, u))) \right] du,
 \end{aligned}$$

which shows that $V^{\pi^1, \pi^2}(\phi(x, s))$ is absolutely continuous in $s \in R_+$, and thus, is differentiable almost everywhere on R_+ . Therefore, differentiating both sides of the above equality with respect to s , and then dividing by $e^{-\int_0^s (\alpha + q(\phi(x, v), \pi^1(\cdot|\phi(x, v)), \pi^2(\cdot|\phi(x, v))))dv}$ both sides of the resulting equality yields

$$\begin{aligned}
 &L^\phi V^{\pi^1, \pi^2}(\phi(x, s)) - \alpha V^{\pi^1, \pi^2}(\phi(x, s)) + r(\phi(x, s), \pi^1(\cdot|\phi(x, s)), \pi^2(\cdot|\phi(x, s))) \\
 &\quad + \int_E V^{\pi^1, \pi^2}(y) q(dy|\phi(x, s), \pi^1(\cdot|\phi(x, s)), \\
 &\quad \pi^2(\cdot|\phi(x, s))) = 0 \quad \forall s \in Z_{V^{\pi^1, \pi^2}}^\phi(x).
 \end{aligned}$$

This implies that

$$L^\phi V^{\pi^1, \pi^2}(x) - \alpha V^{\pi^1, \pi^2}(x) + r(x, \pi^1(\cdot|x), \pi^2(\cdot|x))$$

$$+ \int_E V^{\pi^1, \pi^2}(y)q(dy|x, \pi^1(\cdot|x), \pi^2(\cdot|x)) = 0,$$

for all $x \in \mathcal{D}^{V^{\pi^1, \pi^2}}$. That is, V^{π^1, π^2} satisfies (9).

To show that V^{π^1, π^2} is in $\mathbb{B}_{w_0, w_1}^{ac}(E)$, it remains to verify that $\|L^\phi V^{\pi^1, \pi^2}\|_{w_1}^{es} < \infty$. Indeed, by (9) and Assumptions 1-3, a simple calculation gives that

$$|L^\phi V^{\pi^1, \pi^2}(x)| \leq \left[\frac{M_0}{\alpha - c_0} \left(1 + \frac{d_0}{\alpha}\right) (\alpha + c_0 + d_0 + 2M_1) + M_0 \right] w_1(x) \forall x \in \mathcal{D}^{V^{\pi^1, \pi^2}},$$

which indicates that $\|L^\phi V^{\pi^1, \pi^2}\|_{w_1}^{es} < \infty$. Now, if φ is another solution in $\mathbb{B}_{w_0, w_1}^{ac}(E)$ to equation (9), by Theorem 3, we must have $\varphi = V^{\pi^1, \pi^2}$. The proof is achieved. \square

4 Main results

4.1 Shapley equation and saddle points

In this subsection, we prove that the game has the value function V^* satisfying the associated Shapley equation and a saddle point exists.

To proceed, we introduce a set of continuity assumptions.

Assumption 4 Let $x \in E$ be arbitrary.

- (a) The payoff rate $r(x, a, b)$ is continuous in $(a, b) \in A(x) \times B(x)$;
- (b) The function $q(D|x, a, b)$ is continuous in $(a, b) \in A(x) \times B(x)$ for every $D \in \mathcal{B}(E)$;
- (c) The function $\int_E w_0(y)q(dy|x, a, b)$ is continuous in $(a, b) \in A(x) \times B(x)$.

Moreover, let $m(x)$ be a measurable function on E such that $m(x) \geq q^*(x) + 1$, and

$$Q(D|x, a, b) := \frac{q(D|x, a, b)}{m(x)} + \delta_{\{x\}}(D) \quad \forall (x, a, b) \in K \text{ and } D \in \mathcal{B}(E).$$

Clearly, $Q(\cdot|x, a, b)$ is a stochastic kernel on E given K .

Lemma 2 (a) $P(A(\cdot))$ and $P(B(\cdot))$ are measurable compact-valued multi-functions from E to $\mathcal{P}(P(A))$ and $\mathcal{P}(P(B))$, respectively;

- (b) Under Assumptions 2(b) and 4(a), $r(x, \eta, \gamma)$ defined by (7) is continuous in $(\eta, \gamma) \in P(A(x)) \times P(B(x))$ with respect to the weak topology for each $x \in E$;
- (c) Suppose Assumptions 1(a), 3(a), 4(b) and 4(c) hold. For every function φ in $\mathbb{B}_{w_0}(E)$ and $x \in E$, $\int_E \varphi(y)Q(dy|x, \eta, \gamma)$ and $\int_E \varphi(y)q(dy|x, \eta, \gamma)$ defined by (8) are continuous in $(\eta, \gamma) \in P(A(x)) \times P(B(x))$ with respect to the weak topology.

Proof (a) When the model of PDMG is introduced, we have assumed that $A(\cdot)$ and $B(\cdot)$ are measurable compact-valued multi-functions from E to $\mathcal{P}(A)$ and $\mathcal{P}(B)$, respectively. Then, the result follows from Lemma 1.11 in Nowak (1984).

- (b) For each fixed $x \in E$, Assumptions 2(b) and 4(a) implies that $r(x, a, b)$ is a bounded continuous function on $A(x) \times B(x)$. Hence, for any sequence (η_n, γ_n) weakly converging to (η, γ) in $P(A(x)) \times P(B(x))$, we have $r(x, \eta_n, \gamma_n)$ converge to $r(x, \eta, \gamma)$, which means that $r(x, \eta, \gamma)$ is continuous in $(\eta, \gamma) \in P(A(x)) \times P(B(x))$.
- (c) For each $x \in E$, Assumption 4(c) implies that $\int_E \varphi(y)Q(dy|x, a, b)$ is continuous in $(a, b) \in A(x) \times B(x)$ for every bounded measurable function φ on E , while Assumption 4(c) implies that $\int_E w_0(y)Q(dy|x, a, b)$ is continuous in $(a, b) \in A(x) \times B(x)$. Using this fact and a similar argument as in the proof of Lemma 8.3.7 in Hernández-Lerma and Lasserre (1999), we can prove that $\int_E \varphi(y)Q(dy|x, a, b)$ is continuous in $(a, b) \in A(x) \times B(x)$ for every φ in $\mathbb{B}_{w_0}(E)$, and so is $\int_E \varphi(y)q(dy|x, a, b)$.

For each fixed $x \in E$, Assumptions 1(a) and 3(a) as well as what we have proved above imply that $\int_E \varphi(y)Q(dy|x, a, b)$ and $\int_E \varphi(y)q(dy|x, a, b)$ are bounded continuous function on $A(x) \times B(x)$. Hence, for any sequence (η_n, γ_n) weakly converging to (η, γ) in $P(A(x)) \times P(B(x))$, we have $\int_E \varphi(y)Q(dy|x, \eta_n, \gamma_n) \rightarrow \int_E \varphi(y)Q(dy|x, \eta, \gamma)$ and $\int_E \varphi(y)q(dy|x, \eta_n, \gamma_n) \rightarrow \int_E \varphi(y)q(dy|x, \eta, \gamma)$, completing the proof. \square

For a function φ in $\mathbb{B}_{w_0}(E)$, we define a dynamic programming operator H as below:

$$H\varphi(x) := \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x, v))dv} \sup_{\eta \in P(A(\phi(x, s)))} \inf_{\gamma \in P(B(\phi(x, s)))} \left[r(\phi(x, s), \eta, \gamma) + m(\phi(x, s)) \int_E \varphi(y)Q(dy|\phi(x, s), \eta, \gamma) \right] ds, \quad x \in E.$$

In general, the function

$$x \mapsto \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \varphi(y)Q(dy|x, \eta, \gamma) \right]$$

need not be measurable. However, under those conditions in Lemma 2, by Fan’s minimax theorem in Fan (1953) and Lemma 4.1 in Nowak (1984), it is the case and thus the operator H is well defined.

Now, for each $n = 0, 1, \dots$ and $x \in E$, we recursively define a sequence of functions as follows: $\psi_{n+1}(x) := H\psi_n(x)$, where the initial function ψ_0 is defined by

$$\psi_0(x) := -\frac{M_0}{\alpha - c_0} w_0(x) - \frac{d_0 M_0}{\alpha(\alpha - c_0)}.$$

The choice of ψ_0 is inspired by Lemma 1, i.e., ψ_0 is the lower bound of V^{π^1, π^2} . Note that H need not be a contraction operator due to the unbounded transition rate, and so the convergence of $\{\psi_n\}$ may be sensitive to the choice of ψ_0 .

Theorem 5 *Suppose that Assumptions 1, 2, 3(a) and 4 are satisfied.*

- (a) The sequence $\{\psi_n\}$ is increasing in n , and the limit $\psi_\infty := \lim_{n \rightarrow \infty} \psi_n$ is in $\mathbb{B}_{w_0}(E)$.
- (b) The function ψ_∞ in (a) satisfies the equation $\psi_\infty = H\psi_\infty$.
- (c) The function ψ_∞ in (a) is in $\mathbb{B}_{w_0, w_1}^{ac}(E)$ and verifies the following Shapley equation:

$$L^\phi \varphi(x) - \alpha \varphi(x) + \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E \varphi(y) q(dy|x, \eta, \gamma) \right] = 0 \quad \forall x \in \mathcal{D} \tag{12}$$

Proof (a) To prove the monotonicity of the sequence $\{\psi_n\}$, we first show that $\psi_1 \geq \psi_0$. Indeed, under Assumptions 1, and 2, for every $x \in E$, a direct calculation gives

$$\begin{aligned} &\psi_1(x) \\ &\geq - \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v))dv} \sup_{\eta \in P(A(\phi(x,s)))} \inf_{\gamma \in P(B(\phi(x,s)))} \left[M_0 w_0(\phi(x, s)) \right. \\ &\quad \left. + m(\phi(x, s)) \int_E \left(\frac{M_0}{\alpha - c_0} w_0(y) + \frac{d_0 M_0}{\alpha(\alpha - c_0)} \right) Q(dy|\phi(x, s), \eta, \gamma) \right] ds \\ &\geq - \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v))dv} \left[M_0 w_0(\phi(x, s)) \right. \\ &\quad \left. + \frac{M_0}{\alpha - c_0} \left(c_0 w_0(\phi(x, s)) + d_0 + m(\phi(x, s)) w_0(\phi(x, s)) \right) \right. \\ &\quad \left. + m(\phi(x, s)) \frac{d_0 M_0}{\alpha(\alpha - c_0)} \right] ds \\ &\geq - \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v))dv} \left[M_0 w_0(x) \right. \\ &\quad \left. + \frac{M_0}{\alpha - c_0} \left(c_0 w_0(x) + d_0 + m(\phi(x, s)) w_0(x) \right) + m(\phi(x, s)) \frac{d_0 M_0}{\alpha(\alpha - c_0)} \right] ds \\ &= - \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v))dv} \left[\frac{\alpha M_0 w_0(x)}{\alpha - c_0} + \frac{M_0 d_0}{\alpha - c_0} \right. \\ &\quad \left. + \left(\frac{M_0 w_0(x)}{\alpha - c_0} + \frac{d_0 M_0}{\alpha(\alpha - c_0)} \right) m(\phi(x, s)) \right] ds \\ &= - \left(\frac{M_0 w_0(x)}{\alpha - c_0} + \frac{d_0 M_0}{\alpha(\alpha - c_0)} \right) \int_0^\infty e^{-\int_0^s (\alpha + m(\phi(x,v)))dv} (\alpha + m(\phi(x, s))) ds \\ &= - \left(\frac{M_0 w_0(x)}{\alpha - c_0} + \frac{d_0 M_0}{\alpha(\alpha - c_0)} \right) \\ &= \psi_0(x). \end{aligned}$$

Thus, the monotonicity of the operator H yields

$$\psi_{n+1} = H^n \psi_1 \geq H^n \psi_0 = \psi_n \quad \forall n \geq 1,$$

which implies the monotonicity of the sequence $\{\psi_n\}$, and thus the existence of the point-wise limit ψ_∞ . Moreover, by a similar calculation as in the proof of

$\psi_1 \geq \psi_0$ and an induction argument, one can show that

$$|\psi_n(x)| \leq \frac{M_0}{\alpha - c_0} w_0(x) + \frac{d_0 M_0}{\alpha(\alpha - c_0)} \forall x \in E, n \geq 0,$$

and so is ψ_∞ , which indicates that $\psi_\infty \in \mathbb{B}_{w_0}(E)$.

- (b) On the one hand, by the monotonicity of H , we have that $\psi_{n+1} = H\psi_n \leq H\psi_\infty$ for all $n \geq 0$. Hence, $\psi_\infty \leq H\psi_\infty$. On the other hand, for every fixed $x \in E$ and any $\eta \in P(A(x))$, it is clear that

$$\begin{aligned} & \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_n(y) Q(dy|x, \eta, \gamma) \right] \\ & \geq \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_n(y) Q(dy|x, \eta, \gamma) \right]. \end{aligned} \tag{13}$$

By Lemma 2, for each $n \geq 0$, there exists $\gamma_n \in P(B(x))$ such that

$$\begin{aligned} & \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_n(y) Q(dy|x, \eta, \gamma) \right] \\ & = r(x, \eta, \gamma_n) + m(x) \int_E \psi_n(y) Q(dy|x, \eta, \gamma_n). \end{aligned} \tag{14}$$

Note that $P(B(x))$ is compact with respect to the weak topology, there is a subsequence $\{\gamma_{n_l}, l \geq 0\}$ of $\{\gamma_n, n \geq 0\}$ such that γ_{n_l} weakly converges to some $\gamma^* \in P(B(x))$. Hence, it follows from Lemma 2(b) that

$$\lim_{l \rightarrow \infty} r(x, \eta, \gamma_{n_l}) = r(x, \eta, \gamma^*). \tag{15}$$

Furthermore,

$$\begin{aligned} & \left| \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma_{n_l}) - \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma^*) \right| \\ & \leq \left| \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma_{n_l}) - \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma_{n_l}) \right| \\ & \quad + \left| \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma_{n_l}) - \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma^*) \right| \\ & \leq \int_E (\psi_\infty(y) - \psi_{n_l}(y)) Q(dy|x, \eta, \gamma_{n_l}) \\ & \quad + \left| \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma_{n_l}) - \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma^*) \right|. \end{aligned} \tag{16}$$

Note that $(\psi_\infty - \psi_{n_l})$ are non-increasing in l . Under Assumption 4, using Theorem A. 1. 5 in Bäuerle and Rieder (2011) and the dominated convergence theorem, we

conclude that

$$\begin{aligned}
 & \lim_{l \rightarrow \infty} \int_E (\psi_\infty(y) - \psi_{n_l}(y)) Q(dy|x, \eta, \gamma_{n_l}) \\
 & \leq \lim_{l \rightarrow \infty} \sup_{b \in B(x)} \left[\int_E (\psi_\infty(y) - \psi_{n_l}(y)) Q(dy|x, \eta, b) \right] \\
 & = \sup_{b \in B(x)} \lim_{l \rightarrow \infty} \left[\int_E (\psi_\infty(y) - \psi_{n_l}(y)) Q(dy|x, \eta, b) \right] \\
 & = \sup_{b \in B(x)} \left[\int_E \lim_{l \rightarrow \infty} (\psi_\infty(y) - \psi_{n_l}(y)) Q(dy|x, \eta, b) \right] \\
 & = 0.
 \end{aligned} \tag{17}$$

Moreover, since ψ_∞ is in $\mathbb{B}_{w_0}(E)$, it follows from Lemma 2(c) that

$$\lim_{l \rightarrow \infty} \left| \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma_{n_l}) - \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma^*) \right| = 0. \tag{18}$$

Now, using (16)-(18), we have

$$\lim_{l \rightarrow \infty} \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma_{n_l}) = \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma^*). \tag{19}$$

Hence, it follows from (14), (15) and (19) that

$$\begin{aligned}
 & \lim_{l \rightarrow \infty} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma) \right] \\
 & = \lim_{l \rightarrow \infty} \left[r(x, \eta, \gamma_{n_l}) + m(x) \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma_{n_l}) \right] \\
 & = r(x, \eta, \gamma^*) + m(x) \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma^*),
 \end{aligned} \tag{20}$$

which together with (13) gives that

$$\begin{aligned}
 & \lim_{l \rightarrow \infty} \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma) \right] \\
 & \geq \lim_{l \rightarrow \infty} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma) \right] \\
 & = r(x, \eta, \gamma^*) + m(x) \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma^*) \\
 & \geq \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma) \right].
 \end{aligned} \tag{21}$$

Hence, using the arbitrariness of η and (21), we obtain

$$\begin{aligned} & \lim_{l \rightarrow \infty} \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_{n_l}(y) Q(dy|x, \eta, \gamma) \right] \\ & \geq \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_\infty(y) Q(dy|x, \eta, \gamma) \right] \quad \forall x \in (\mathcal{X}) \end{aligned}$$

Thus, using the dominated convergence theorem and (22), we have

$$\begin{aligned} & \psi_\infty(x) \\ & = \lim_{l \rightarrow \infty} H\psi_{n_l}(x) \\ & = \lim_{l \rightarrow \infty} \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v))dv} \sup_{\eta \in P(A(\phi(x,s)))} \inf_{\gamma \in P(B(\phi(x,s)))} \left[r(\phi(x,s), \eta, \gamma) \right. \\ & \quad \left. + m(\phi(x,s)) \int_E \psi_{n_l}(y) Q(dy|\phi(x,s), \eta, \gamma) \right] ds \\ & = \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v))dv} \lim_{l \rightarrow \infty} \sup_{\eta \in P(A(\phi(x,s)))} \inf_{\gamma \in P(B(\phi(x,s)))} \left[r(\phi(x,s), \eta, \gamma) \right. \\ & \quad \left. + m(\phi(x,s)) \int_E \psi_{n_l}(y) Q(dy|\phi(x,s), \eta, \gamma) \right] ds \\ & \geq \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v))dv} \sup_{\eta \in P(A(\phi(x,s)))} \inf_{\gamma \in P(B(\phi(x,s)))} \left[r(\phi(x,s), \eta, \gamma) \right. \\ & \quad \left. + m(\phi(x,s)) \int_E \psi_\infty(y) Q(dy|\phi(x,s), \eta, \gamma) \right] ds \quad \forall x \in E, \end{aligned}$$

which gives the reverse inequality $\psi_\infty \geq H\psi_\infty$. This leads to that $\psi_\infty = H\psi_\infty$.

(c) Clearly, for every $x \in E$ and $t \in R_+$, we see that

$$\begin{aligned} & \psi_\infty(\phi(x,t)) \\ & = \int_0^\infty e^{-\alpha s} e^{-\int_0^s m(\phi(x,v+t))dv} \\ & \quad \times \sup_{\eta \in P(A(\phi(x,s+t)))} \inf_{\gamma \in P(B(\phi(x,s+t)))} \left[r(\phi(x,s+t), \eta, \gamma) \right. \\ & \quad \left. + m(\phi(x,s+t)) \int_E \psi_\infty(y) Q(dy|\phi(x,s+t), \eta, \gamma) \right] ds, \quad x \in E, \end{aligned}$$

which is equivalent to that

$$\begin{aligned} & e^{-\int_0^t (\alpha + m(\phi(x,v)))dv} \psi_\infty(\phi(x,t)) \\ & = \int_t^\infty e^{-\int_0^s (\alpha + m(\phi(x,v)))dv} \sup_{\eta \in P(A(\phi(x,s)))} \inf_{\gamma \in P(B(\phi(x,s)))} \left[r(\phi(x,s), \eta, \gamma) \right. \\ & \quad \left. + m(\phi(x,s)) \int_E \psi_\infty(y) Q(dy|\phi(x,s), \eta, \gamma) \right] ds, \quad x \in E. \end{aligned} \quad (23)$$

This equality shows that $\psi_\infty(\phi(x, t))$ is absolutely continuous in $t \in R_+$, and thus, is differentiable almost everywhere on R_+ , which indicates that ψ_∞ is in $\mathbb{B}_{w_0}^{ac}(E)$.

For $x \in E$ and $t \in R_+$, differentiating both sides of (23) with respect to t eventually leads to

$$L^\phi \psi_\infty(x) - \alpha \psi_\infty(x) + \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E \psi_\infty(y)q(dy|x, \eta, \gamma) \right] = 0 \quad \forall x \in \mathcal{D}^{\psi_\infty},$$

which in turn implies that $\|L^\phi \psi_\infty\|_{w_1}^{es} < \infty$. The proof is complete. □

We are now ready to state our main results.

Theorem 6 *Under Assumptions 1–3 and 4, the following assertions hold.*

(a) *The game has the value function $V^*(x)$ as the unique solution in $\mathbb{B}_{w_0, w_1}^{ac}(E)$ to the Shapley equation*

$$L^\phi V^*(x) - \alpha V^*(x) + \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E V^*(y)q(dy|x, \eta, \gamma) \right] = 0 \quad \forall x \in \mathcal{D}^{V^*}.$$

(b) *There exists a pair of policies $(\hat{\pi}^1, \hat{\pi}^2) \in \Pi_{RS}^1 \times \Pi_{RS}^2$ such that*

$$\begin{aligned} & \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E V^*(y)q(dy|x, \eta, \gamma) \right] \\ &= r(x, \hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x)) + \int_E V^*(y)q(dy|x, \hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x)) \\ &= \inf_{\gamma \in P(B(x))} \left[r(x, \hat{\pi}^1(\cdot|x), \gamma) + \int_E V^*(y)q(dy|x, \hat{\pi}^1(\cdot|x), \gamma) \right] \\ &= \sup_{\eta \in P(A(x))} \left[r(x, \eta, \hat{\pi}^2(\cdot|x)) + \int_E V^*(y)q(dy|x, \eta, \hat{\pi}^2(\cdot|x)) \right] \end{aligned}$$

for all $x \in E$, and such a pair of policies $(\hat{\pi}^1, \hat{\pi}^2) \in \Pi_{RS}^1 \times \Pi_{RS}^2$ is a saddle point.

Proof For each $x \in E$, by Lemma 2, $P(A(x))$ and $P(B(x))$ is compact, and the function

$$(\eta, \gamma) \mapsto r(x, \eta, \gamma) + \int_E \psi_\infty(y)q(dy|x, \eta, \gamma) \tag{24}$$

is continuous on $P(A(x)) \times P(B(x))$ with respect to the weak topology. Moreover, (24) is linear in $(\eta, \gamma) \in P(A(x)) \times P(B(x))$. Thus, using Fan’s minimax theorem

in Fan (1953), we obtain

$$\begin{aligned} & \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E \psi_\infty(y)q(dy|x, \eta, \gamma) \right] \\ &= \inf_{\gamma \in P(B(x))} \sup_{\eta \in P(A(x))} \left[r(x, \eta, \gamma) + \int_E \psi_\infty(y)q(dy|x, \eta, \gamma) \right] \forall x \in E. \end{aligned}$$

Now, by Lemma 2 above and Lemma 4.1 in Nowak (1984), there exists a pair of policies $(\hat{\pi}^1, \hat{\pi}^2) \in \Pi_{RS}^1 \times \Pi_{RS}^2$ such that

$$\begin{aligned} & \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E \psi_\infty(y)q(dy|x, \eta, \gamma) \right] \\ &= \inf_{\gamma \in P(B(x))} \left[r(x, \hat{\pi}^1(\cdot|x), \gamma) + \int_E \psi_\infty(y)q(dy|x, \hat{\pi}^1(\cdot|x), \gamma) \right] \forall x \in E, \end{aligned}$$

and

$$\begin{aligned} & \inf_{\gamma \in P(B(x))} \sup_{\eta \in P(A(x))} \left[r(x, \eta, \gamma) + \int_E \psi_\infty(y)q(dy|x, \eta, \gamma) \right] \\ &= \sup_{\eta \in P(A(x))} \left[r(x, \eta, \hat{\pi}^2(\cdot|x)) + \int_E \psi_\infty(y)q(dy|x, \eta, \hat{\pi}^2(\cdot|x)) \right] \forall x \in E, \end{aligned}$$

which implies that

$$\begin{aligned} & \sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E \psi_\infty(y)q(dy|x, \eta, \gamma) \right] \\ &= r(x, \hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x)) + \int_E \psi_\infty(y)q(dy|x, \hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x)) \forall x \in E. \end{aligned}$$

Therefore, since the function ψ_∞ satisfies Shapley equation (12) by Theorem 5, we have

$$\begin{aligned} & L^\phi \psi_\infty(x) - \alpha \psi_\infty(x) + r(x, \hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x)) \\ &+ \int_E \psi_\infty(y)q(dy|x, \hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x)) = 0, \end{aligned} \tag{25}$$

$$\begin{aligned} & L^\phi \psi_\infty(x) - \alpha \psi_\infty(x) + \inf_{\gamma \in P(B(x))} \left[r(x, \hat{\pi}^1(\cdot|x), \gamma) \right. \\ &+ \left. \int_E \psi_\infty(y)q(dy|x, \hat{\pi}^1(\cdot|x), \gamma) \right] = 0, \end{aligned} \tag{26}$$

$$\begin{aligned} & L^\phi \psi_\infty(x) - \alpha \psi_\infty(x) + \sup_{\eta \in P(A(x))} \left[r(x, \eta, \hat{\pi}^2(\cdot|x)) \right. \\ &+ \left. \int_E \psi_\infty(y)q(dy|x, \eta, \hat{\pi}^2(\cdot|x)) \right] = 0, \end{aligned} \tag{27}$$

for every $x \in \mathcal{D}^{\psi_\infty}$. Using Theorem 4 and (25), we have $\psi_\infty = V^{\hat{\pi}^1, \hat{\pi}^2}$. Using Theorem 3(b) and (26), we have $\psi_\infty \leq V^{\hat{\pi}^1, \pi^2}$ for every $\pi^2 \in \Pi^2$. Using Theorem 3(a) and (27), we have $V^{\pi^1, \hat{\pi}^2} \leq \psi_\infty$ for every $\pi^1 \in \Pi^1$. These three facts show that

$$V^{\pi^1, \hat{\pi}^2}(x) \leq V^{\hat{\pi}^1, \hat{\pi}^2}(x) \leq V^{\hat{\pi}^1, \pi^2}(x) \quad \forall \pi^1 \in \Pi^1, \pi^2 \in \Pi^2, x \in E, \tag{28}$$

which implies that the pair of policies $(\hat{\pi}^1, \hat{\pi}^2) \in \Pi^1_{RS} \times \Pi^2_{RS}$ is a saddle point, and thus $V^* = \psi_\infty$ is the value of the game. Hence, part (a) and part (b) follow. \square

Remark 5 From Theorem 6, one can see that the max-min points of the mapping

$$(\eta, \gamma) \mapsto r(x, \eta, \gamma) + \int_E V^*(y)q(dy|x, \eta, \gamma)$$

over $P(A(x)) \times P(B(x))$ for all $x \in E$, denoted by $(\hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x))$, constitute a saddle point. It has a very simple form, which depends only on the current state and can be applied at any time.

4.2 How to compute a saddle point

Although Theorem 6 shows the existence of a saddle point, how to compute a saddle point in practice is still a difficult task. There are two problems on this issue. The first one is how to compute the value function $V^*(x)$. The second one is how to solve the static game

$$\sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E V^*(y)q(dy|x, \eta, \gamma) \right].$$

Fortunately, Theorem 5 proposes a value iteration algorithm to compute the value function $V^*(x)$, and Sect. 2.4.2 in Barron (2013) provides a linear program formulation to solve the static game. Below, we propose a potential algorithm to compute a saddle point.

Algorithm for saddle point:

Step 1. Specify an accuracy $\epsilon > 0$, and set $n = 0$. Let

$$\psi_0(x) := -\frac{M_0}{\alpha - c_0}w_0(x) - \frac{d_0M_0}{\alpha(\alpha - c_0)} \quad \forall x \in E.$$

Step 2. For each $x \in E$, carry out the linear program:

$$\begin{aligned} \text{LP1 : } & \max_{\eta, V(x)} V(x), \\ & \text{s.t. } r(x, \eta, b) + m(x) \int_E \psi_n(y)Q(dy|x, \eta, b) \geq V(x) \quad \forall b \in B(x), \\ & \eta \in P(A(x)). \end{aligned}$$

One of the outputs of LP is the value $\hat{V}(x)$ of the static game

$$\sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + m(x) \int_E \psi_n(y) Q(dy|x, \eta, \gamma) \right].$$

Step 3. For each $x \in E$, compute

$$\psi_{n+1} := \int_0^\infty e^{-\int_0^s (\alpha + m(\phi(x, v))) dv} \hat{V}(\phi(x, s)) ds.$$

Step 4. If $|\psi_{n+1}(x) - \psi_n(x)| < \epsilon$ for every $x \in E$, go to Step 5. Otherwise, increment n by 1 and return to Step 2.

Step 5. For each $x \in E$, carry out the two linear programs:

$$\begin{aligned} \mathbf{LP2} : \quad & \max_{\eta, V(x)} V(x), \\ & \text{s.t.} \quad r(x, \eta, b) + \int_E \psi_{n+1}(y) q(dy|x, \eta, b) \geq V(x) \quad \forall b \in B(x), \\ & \quad \eta \in P(A(x)). \end{aligned}$$

and

$$\begin{aligned} \mathbf{LP3} : \quad & \min_{\gamma, W(x)} W(x), \\ & \text{s.t.} \quad r(x, a, \gamma) + \int_E \psi_{n+1}(y) q(dy|x, a, \gamma) \leq W(x) \quad \forall a \in A(x), \\ & \quad \gamma \in P(B(x)). \end{aligned}$$

Denote by $(\hat{V}(x), \hat{\pi}^1(\cdot|x))$ the solution of LP2 and by $(\hat{W}(x), \hat{\pi}^2(\cdot|x))$ the solution of LP3. Then, $\hat{V}(x) = \hat{W}(x)$ and $(\hat{\pi}^1(\cdot|x), \hat{\pi}^2(\cdot|x))$ are the value and a saddle point of the static game

$$\sup_{\eta \in P(A(x))} \inf_{\gamma \in P(B(x))} \left[r(x, \eta, \gamma) + \int_E \psi_{n+1}(y) q(dy|x, \eta, \gamma) \right], x \in E,$$

respectively.

Since $\psi_{n+1} \approx V^*$, by Theorem 6, we can expect that $(\hat{\pi}^1, \hat{\pi}^2)$ obtained in Step 5 above is a near saddle point for the original PDMGs.

It should be noted that it is very difficult to implement the algorithm since the state space and the action spaces are all uncountable.

5 Example

In this section, we provide an example wherein all the assumptions in this paper are fulfilled.

Example 1 Let $E = (0, \infty)$, $A(x) = [-x - 1, -x]$, and $B(x) = [0, \frac{\sqrt{2}}{2}x]$ for each $x \in E$. Suppose that the flow $\phi(x, t) = xe^{-t}$, the payoff rate $r(x, a, b) = \frac{x}{2} + a + b$ for each $(x, a, b) \in K$ and $t \in R_+$. The transition rates are given by

$$q(D|x, a, b) = (x + a + b + 1) \left[\int_{D \setminus \{x\}} \frac{1}{x + a + b + 1} e^{-\frac{y}{x+a+b+1}} dy - \delta_{\{x\}}(D) \right],$$

$$\forall (x, a, b) \in K, D \in \mathcal{B}(E).$$

Note that, starting from any state $x \in E$, the flow never touches the boundary of E , say, $\{0\}$. Moreover, it is obvious that the multi-functions $A(\cdot)$ and $B(\cdot)$ are compact-valued. Also, we can verify that the inverses $A^{-1}(O) := \{x \in E : A(x) \cap O = \emptyset\} \in \mathcal{B}(E)$ and $B^{-1}(O) \in \mathcal{B}(E)$ for every open interval O in R , which together with the fact that any open set in R can be represented by countable union of open intervals, $A^{-1}(O) \in \mathcal{B}(E)$ and $B^{-1}(O) \in \mathcal{B}(E)$ for every open set O in R . This means that the multi-functions $A(\cdot)$ and $B(\cdot)$ are measurable.

Now, we verify that all the assumptions in this paper are satisfied. To do so, take $w_0(x) = x + 1$, $w_1(x) = x^2 + 1$, and $E_m = [0, m]$ for all $x \in E$ and $m \geq 1$. Then, it is clear that $w_0(\phi(x, t)) \leq w_0(x)$, $w_1(\phi(x, t)) \leq w_1(x)$, $E_m \uparrow E$, $\sup_{x \in E_m} q^*(x) = \frac{\sqrt{2}}{2}m + 1 < \infty$, $\lim_{m \rightarrow \infty} \inf_{x \notin E_m} w_0(x) = \infty$, $|r(x, a, b)| \leq \frac{x}{2} + 1 \leq w_0(x)$, and

$$(1 + q(x, a, b))w_0(x) \leq \left(\frac{\sqrt{2}}{2}x + 2\right)(x + 1) \leq 3w_1(x) \quad \forall (x, a, b) \in K.$$

This indicates that Assumptions 1(b), 1(c), 2(b), 3(a) and 3(c) are fulfilled.

Regarding Assumption 1(a), for all $(x, a, b) \in K$,

$$\int_E w_0(y)q(dy|x, a, b) \leq (x + a + b + 1) \left[\left(\frac{\sqrt{2}}{2} - 1\right)x + 1 \right].$$

If $x \leq 2 + \sqrt{2}$, then $(\frac{\sqrt{2}}{2} - 1)x + 1 \geq 0$, and thus, $(x + a + b + 1) \left[\left(\frac{\sqrt{2}}{2} - 1\right)x + 1 \right] \leq (\frac{\sqrt{2}}{2}x + 1) \left[\left(\frac{\sqrt{2}}{2} - 1\right)x + 1 \right] \leq w_0(x)$. If $x > 2 + \sqrt{2}$, then $(\frac{\sqrt{2}}{2} - 1)x + 1 < 0$, and thus, $(x + a + b + 1) \left[\left(\frac{\sqrt{2}}{2} - 1\right)x + 1 \right] \leq 0 < w_0(x)$. This means that, in either case, Assumption 1(a) is fulfilled with $c_0 = 1$ and $d_0 = 0$.

Regarding Assumption 3(b), for all $(x, a, b) \in K$, we have

$$\begin{aligned} & \int_E w_1(y)q(dy|x, a, b) \\ &= (x + a + b + 1) \left[2(x + a + b + 1)^2 - x^2 \right] \\ &\leq \left(\frac{\sqrt{2}}{2}x + 1\right) \left[2\left(\frac{\sqrt{2}}{2}x + 1\right)^2 - x^2 \right] \\ &\leq \left(\frac{\sqrt{2}}{2}x + 1\right) (2\sqrt{2}x + 2) \end{aligned}$$

$$\leq 4w_1(x) + 2,$$

which implies that Assumption 3(b) is satisfied with $c_1 = 4$ and $d_1 = 2$. Hence, if we take $\alpha > 4$, the requirements on the discount factor in Assumptions 2(a) and 3 are both fulfilled.

Finally, Assumption 4 is obviously true for the data in this example.

Therefore, by Theorem 6, there is a saddle point for this game model.

Acknowledgements This research was supported in part by the National Natural Science Foundation of China (11931018), the National Key Research and Development Program of China (2022YFA1004600), the University of Macau (MYRG2019-00031-FBA), and Guangdong Basic and Applied Basic Research Foundation. The authors are grateful to the two reviewers and the associate editor for their careful reading and many constructive suggestions that have improved this paper.

6 Appendix: Proof of Proposition 1

In fact, we extend the previous results (Guo and Song 2011, Theorem 3.1) for CTMDPs and (Huang and Guo 2019, Proposition 3.1) for PDMDPs to the case of PDMGs. Thus, the proof of Proposition 1 proceeds in a similar way. Since Assumption 1 in this paper slightly differs from Assumption 3.1 in Huang and Guo (2019), the proofs are accordingly adjusted. First, we provide two lemmas.

Lemma 3 *Let some function $w \geq 0$ on E , and some signed kernel $\bar{q}(dy|x, u)$ on E given $E \times R_+$ satisfy the following conditions: (i) $\bar{q}(E|x, u) = 0$; $\bar{q}(D|x, u) \geq 0$ if $\phi(x, u) \notin D$; $\bar{q}(x, u) := \bar{q}(E \setminus \{\phi(x, u)\}|x, u) < \infty$; (ii) $\int_E \bar{q}(dy|x, u)w(y) \leq cw(\phi(x, u)) + d$, where $u \geq 0$, $c > 0$ and $d \geq 0$ are some constants; (iii) $w(\phi(x, t)) \leq w(x)$ for all $x \in E$ and $t \in R_+$. Then the function*

$$\Psi(s, x) := e^{cs}w(x) + \frac{d}{c}(e^{cs} - 1), \quad s \in R_+, x \in E, \quad (29)$$

satisfies the following inequality

$$\begin{aligned} & \int_0^s \int_{E \setminus \{\phi(x, u)\}} e^{-\int_0^u \bar{q}(x, v)dv} \bar{q}(dy|x, u) \Psi(s - u, y) du + e^{-\int_0^s \bar{q}(x, v)dv} w(\phi(x, s)) \\ & \leq \Psi(s, x) \end{aligned}$$

for all $s \in R_+$ and $x \in E$.

Proof Using the conditions in this lemma, it is easy to verify that

$$\begin{aligned} & \int_0^s \int_{E \setminus \{\phi(x, u)\}} e^{-\int_0^u \bar{q}(x, v)dv} \bar{q}(dy|x, u) \left[e^{c(s-u)}w(y) + \frac{d}{c}(e^{c(s-u)} - 1) \right] du \\ & \quad + e^{-\int_0^s \bar{q}(x, v)dv} w(\phi(x, s)) \\ & \leq \int_0^s e^{-\int_0^u \bar{q}(x, v)dv} \left[e^{c(s-u)}(cw(\phi(x, u)) + d + \bar{q}(x, u)w(\phi(x, u))) \right] \end{aligned}$$

$$\begin{aligned}
 & + \frac{b}{c} (e^{c(s-u)} - 1) \bar{q}(x, u) \Big] du + e^{-\int_0^s \bar{q}(x,v) dv} w(\phi(x, s)) \\
 & \leq \int_0^s e^{-\int_0^u \bar{q}(x,v) dv} \Big[e^{c(s-u)} (cw(x) + d + \bar{q}(x, u)w(x)) \\
 & + \frac{d}{c} (e^{c(s-u)} - 1) \bar{q}(x, u) \Big] du \\
 & \quad + e^{-\int_0^s \bar{q}(x,v) dv} w(x) \\
 & = \left(w(x) + \frac{d}{c} \right) \int_0^s e^{-\int_0^u \bar{q}(x,v) dv} d \left(-e^{c(s-u)} \right) \\
 & \quad + w(x) \int_0^s e^{c(s-u)} d \left(-e^{-\int_0^u \bar{q}(x,v) dv} \right) \\
 & \quad + \frac{d}{c} \int_0^s \left(e^{c(s-u)} - 1 \right) d \left(-e^{-\int_0^u \bar{q}(x,v) dv} \right) + e^{-\int_0^s \bar{q}(x,v) dv} w(x) \\
 & = \left(w(x) + \frac{d}{c} \right) \left[e^{cs} - e^{-\int_0^s \bar{q}(x,v) dv} - \int_0^s e^{c(s-u)} d \left(-e^{-\int_0^u \bar{q}(x,v) dv} \right) \right] \\
 & \quad + \left(w(x) + \frac{d}{c} \right) \int_0^s e^{c(s-u)} d \left(-e^{-\int_0^u \bar{q}(x,v) dv} \right) + \frac{d}{c} \left(e^{-\int_0^s \bar{q}(x,v) dv} - 1 \right) \\
 & \quad + e^{-\int_0^s \bar{q}(x,v) dv} w(x) \\
 & = e^{cs} w(x) + \frac{d}{c} (e^{cs} - 1) \\
 & = \Psi(s, x),
 \end{aligned}$$

where the first and second inequalities follow from conditions (ii) and (iii), respectively. □

Lemma 4 *Let Assumption 1(a) and Assumption 1(c) be fulfilled. Then, for each $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $x \in E$, and $m = 0, 1, 2, \dots$,*

$$\mathbb{E}_x^{\pi^1, \pi^2} [w_0(\xi_t) \mathbb{1}_{\{t < T_{m+1}\}}] \leq e^{c_0 t} w_0(x) + \frac{d_0}{c_0} (e^{c_0 t} - 1) \quad \forall t \in R_+.$$

Proof The proof is similar to that of (Huang and Guo 2019, Lemma A.2) for PDMDPs. We omit it to save space. □

Proof of Proposition 1 Based on Lemma 4 and Assumption 1, the proof of Proposition 1 can be proceeded in a similar way to that of (Guo and Song 2011, Theorem 3.1) for CTMDPs and that of (Huang and Guo 2019, Proposition 3.1) for PDMDPs. However, we omit the proof to save space. □

References

Avrachenkov K, Habachi O, Piunovskiy A, Zhang Y (2015) Infinite horizon impulsive optimal control with applications to Internet congestion control. *Int J Control* 88:703–716

- Barron EN (2013) *Game Theory: An Introduction*. Wiley series in operations research and management science, 2nd edn. Wiley series in operations research and management science Wiley series in operations research and management science Wiley series in operations research and management science. John Wiley and Sons Inc, Hoboken, NJ
- Bäuerle N, Rieder U (2011) Markov Decision Processes with Applications to Finance. Chapter 8: piecewise deterministic Markov decision processes. Universitext, Springer, Heidelberg
- Costa OLV, Dufour F, Piunovskiy AB (2016) Constrained and unconstrained optimal discounted control of piecewise deterministic Markov processes. *SIAM J Control Optim* 54:1444–1474
- Costa OLV, Dufour F (2018) Zero-sum discounted reward criterion games for piecewise deterministic Markov processes. *Appl Math Optim* 78:587–611
- Fan K (1953) Minimax theorems. *Proc Natl Acad Sci USA* 39:42–47
- Gensbittel F, Renault J (2015) The value of Markov chain games with incomplete information on both sides. *Math Oper Res* 40:820–841
- Ghosh MK, Goswami A (2006) Partially observable semi-Markov games with discounted payoff. *Stoch Anal Appl* 24:1035–1059
- Guo X, Zhang Y (2017) Zero-sum continuous-time Markov pure jump game over a fixed duration. *J Math Anal Appl* 452:1194–1208
- Guo XP, Hernández-Lerma O (2003) Zero-sum games for continuous-time Markov chains with unbounded transition and average payoff rates. *J Appl Probab* 40:327–345
- Guo XP, Hernández-Lerma O (2007) Zero-sum games for continuous-time jump Markov processes in Polish spaces: discounted payoffs. *Adv Appl Probab* 39:645–668
- Guo XP, Song XY (2011) Discounted continuous-time constrained Markov decision processes in Polish spaces. *Ann Appl Probab* 21:2016–2049
- Haurie A, Krawczyk JB, Zaccour G (2012) *Games and Dynamic Games*. World Scientific Now Publishers Series in Business, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ
- Hernández-Lerma O, Lasserre JB (1999) *Further topics on discrete-time Markov control processes*. Springer-Verlag, New York
- Huang YH, Guo XP (2019) Finite-horizon piecewise deterministic Markov decision processes with unbounded transition rates. *Stochastics* 91:67–95
- Jacod J (1975) Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und Verw Gebiete* 31:235–253
- Jaśkiewicz A (2009) Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities. *J Optim Theory Appl* 141:321–347
- Lorenzo JM, Hernández-Noriega I, Prieto-Rumeau T (2015) Approximation of two-person zero-sum continuous-time Markov games with average payoff criterion. *Oper Res Lett* 43:110–116
- Minjárez-Sosa JA (2020) Zero-sum discrete-time Markov games with unknown disturbance distribution. Discounted and average criteria. *SpringerBriefs in Probability and Mathematical Statistics*. Springer, Cham
- Mondal P (2017) On zero-sum two-person undiscounted semi-Markov games with a multichain structure. *Adv Appl Probab* 49:826–849
- Nowak AS (1984) On zero-sum stochastic games with general state space. I. *Probab Math Stat* 4:13–32
- Piunovskiy A, Zhang Y (2021) Aggregated occupation measures and linear programming approach to constrained impulse control problems. *J Math Anal Appl* 499(2):125070
- Prieto-Rumeau T, Lorenzo JM (2015) Approximation of zero-sum continuous-time Markov games under the discounted payoff criterion. *TOP* 23:799–836

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.