

Nonzero-sum Discrete-time Stochastic Games with Risk-sensitive Ergodic Cost Criterion

Bivakar Bose* Chandan Pal† Somnath Pradhan‡ and Subhamay Saha§

Abstract

In this paper we study infinite horizon nonzero-sum stochastic games for controlled discrete-time Markov chains on a Polish state space with risk-sensitive ergodic cost criterion. Under suitable assumptions we show that the associated ergodic optimality equations admit unique solutions. Finally, the existence of Nash-equilibrium in randomized stationary strategies is established by showing that an appropriate set-valued map has a fixed point.

Key Words : Nonzero-sum game, risk-sensitive ergodic cost criterion, optimality equations, Nash equilibrium, set-valued map

Mathematics Subject Classification: 91A15, 91A50

1 Introduction

We consider infinite horizon ergodic-cost risk-sensitive two-person nonzero-sum stochastic games for discrete time Markov decision processes (MDPs) on a general state space with bounded cost function. We first establish the existence of unique solutions of the corresponding optimality equations. This is used to establish the existence of optimal randomized stationary strategies for a player fixing the strategy of the other player. We then define a suitable topology on the space of randomized stationary strategies. Then under certain separability assumptions we show that a suitably defined set-valued map has a fixed point. This gives the existence of a Nash equilibrium for the nonzero-sum game under consideration. The main step towards this end is to establish the upper semi-continuity of the defined set-valued map.

*Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati-781039, India, email: b.bivakar@iitg.ernet.in

†Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati-781039, India, email: cpal@iitg.ernet.in

‡Department of Mathematics, Indian Institute Science Education and Research, Bhopal, Bhopal-462066 India, email: somnath@iiserb.ac.in

§Department of Mathematics, Indian Institute of Technology, Guwahati, Guwahati-781039 India, email: saha.subhamay@iitg.ac.in

In a stochastic control problem, since the cost payable depends on the random evolution of the underlying stochastic process, the same is also random. So the most basic approach is to minimize the expected cost. This is the risk-neutral approach. But the obvious short coming of this approach is that it does not take care of the risk factor. In the risk-sensitive criterion, the expected value of the exponential of the cost is considered. Generally the risk associated with a random quantity is quantified by the standard deviation. Hence risk-sensitive criterion provides significantly better protection from the risk since it captures the effects of the higher order moments of the cost as well, see [24] for more details. The analysis of risk-sensitive control is technically more involved because of the exponential nature of the cost; accumulated costs are multiplicative over time as opposed to additive in the risk-neutral case. Risk-sensitive control problems also has deep connections to robust control theory, control theory literature which deals with model uncertainty, see [24] and references therein. Risk-sensitive criterion also arises naturally in portfolio optimization problems in mathematical finance, see [6] and associated references. Nonzero-sum stochastic games arise naturally in strategic multi-agent decision-making system like socioeconomic systems [16] , network security [1], routing and scheduling [19] and so on.

Following the seminal work of Howard and Matheson [15] there has been a lot of work on risk-sensitive control problems. For an up to date survey on ergodic risk-sensitive control problems with underlying stochastic process being discrete-time Markov chains, see [7]. In the multi-controller set up, discrete-time risk-sensitive zero-sum games on countable state space have been studied in [2, 11]. In [4] the authors consider discrete-time zero-sum risk-sensitive stochastic games on Borel state space for both discounted and ergodic cost criterion. Their analysis involves transforming the original risk-sensitive problem to a risk-neutral problem. Nonzero-sum risk-sensitive stochastic games with both discounted and ergodic cost criterion for countable state space and bounded cost has been studied in [3, 22]. In [23] discrete-time nonzero-sum stochastic games under the risk-sensitive ergodic cost criterion with countable state space and unbounded cost function has been investigated. To the best of our knowledge, this is the first paper which considers risk-sensitive nonzero-sum games on a general state space. The analysis of nonzero-sum games on general state space is significantly more involved as compared to the case of countable state space. One of the reasons for this is that in case of countable state space the topology on the space of randomized stationary controls is fairly simple to work with. In case of general state space this becomes a substantial challenge. In the risk-neutral setup itself the analysis becomes substantially involved and requires additional assumptions. In the risk-neutral set up discrete-time nonzero-sum ergodic cost games with general state space has been studied in [10, 17, 21]. In [10], the authors prove the existence of a Nash-equilibrium under an additive reward additive transition (ARAT) assumption.

In [17], the authors establish the existence of an ϵ -Nash equilibrium. In [21], the authors assume that the transition law is a combination of finitely many probability measures. In this work we also assume ARAT condition.

In this paper we first consider ergodic optimality equations which correspond to control problems with the strategy of one player fixed. We show the existence of unique solutions to these equations. We follow a span contraction approach. For this analysis we make certain ergodicity assumption along with a few extra assumptions on the state transition kernel. Analogous assumptions appear in [8] where the authors consider risk-sensitive control problems with ergodic cost criterion. In order to establish the existence of a Nash equilibrium we first define an appropriate set valued map. In order to show the existence of a Nash equilibrium for the nonzero-sum game we use Fan's fixed point theorem [9]. This involves showing that the defined set valued map is upper semi-continuous under the appropriate topology.

The rest of the paper is organized as follows. Section 2 introduces the game model, some preliminaries and notations. In section 3 we show the existence of a unique solution to the optimality equation using a certain span-norm contraction. Existence of the Nash-equilibrium has been shown in section 4.

2 The Game Model

In this section, we present the discrete-time nonzero-sum stochastic game model and introduce the notations utilized throughout the paper. The following elements are needed to describe the discrete-time nonzero-sum stochastic game:

$$\{X, A, B, P, c_1, c_2\},$$

where each component is described below.

- X is the state space assumed to be Polish space endowed with the Borel σ -algebra $\mathcal{X}$.
- A and B are action spaces of player 1 and 2 respectively, assumed to be compact metric spaces. Let $\mathcal{A}$ and $\mathcal{B}$ denote the Borel σ -algebras on A and B respectively.
- P is the transition kernel from $X \times A \times B \rightarrow \mathcal{P}(X)$, where for any metric space D , $\mathcal{P}(D)$ denotes the space of all probability measures on D with the topology of weak convergence.
- $c_i : X \times A \times B \rightarrow \mathbb{R}$, $i = 1, 2$, is one-stage cost function for player i , assumed to be bounded and continuous on $A \times B$. Since, the cost is bounded, without loss of generality let, $0 \leq c_i \leq \bar{c}$.

At each stage(time) the players observe the current state $x \in X$ of the system and then player 1 and 2 independently choose actions $a \in A$ and $b \in B$ respectively. As a consequence two things happen:

- (i) player $i, i = 1, 2$, pays an immediate cost $c_i(x, a, b)$.
- (ii) the system moves to a new state y with the distribution $P(\cdot|x, a, b)$.

The whole process then repeats from the new state y . The cost accumulates throughout the course of the game. The planning horizon is taken to be infinite and each player wants to minimize his/her infinite-horizon risk-sensitive cost with respect to some cost criterion, which is in our case defined by (2.1) below.

At each stage the players choose their actions independently on the basis of the available information. The information available for decision making at time $t \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$ is given by the history of the process up to that time

$$h_t := (x_0, a_0, b_0, x_1, a_1, b_1, \dots, a_{t-1}, b_{t-1}, x_t) \in H_t,$$

where $H_0 = X, H_t = H_{t-1} \times (A \times B \times X), \dots, H_\infty = (X \times A \times B)^\infty$ are the history spaces. The history spaces are endowed with the corresponding Borel σ -algebra. A strategy for player 1 is a sequence $\pi^1 = \{\pi_t^1\}_{t \in \mathbb{N}_0}$ of stochastic kernels $\pi_t^1 : H_t \rightarrow \mathcal{P}(A)$. The set of all strategies for player 1 is denoted by Π_1 . A strategy $\pi^1 \in \Pi_1$ is called a Markov strategy if

$$\pi_t^1(h_{t-1}, a, b, x)(\cdot) = \pi_t^1(h'_{t-1}, a', b', x)(\cdot)$$

for all $h_{t-1}, h'_{t-1} \in H_{t-1}, a, a' \in A, b, b' \in B, x \in X, t \in \mathbb{N}_0$. Thus a Markov strategy for player 1 can be identified with a sequence of measurable maps $\{\Phi_t^1\}, \Phi_t^1 : X \rightarrow \mathcal{P}(A)$. A Markov strategy $\{\Phi_t^1\}$ is called a stationary strategy if $\Phi_t^1 = \Phi : X \rightarrow \mathcal{P}(A)$ for all t . Let M_1 and S_1 denote the set of Markov and stationary strategies for player 1, respectively. The strategies for player 2 are defined similarly. Let Π_2, M_2 , and S_2 denote the set of arbitrary, Markov and stationary strategies for player 2, respectively.

Given an initial distribution $\tilde{\pi}_0 \in \mathcal{P}(X)$ and a pair of strategies $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$, the corresponding state and action processes $\{X_t\}, \{A_t\}, \{B_t\}$ are stochastic processes defined on the canonical space $(H_\infty, \mathfrak{B}(H_\infty), P_{\tilde{\pi}_0}^{\pi^1, \pi^2})$ (where $\mathfrak{B}(H_\infty)$ is the Borel σ -field on H_∞) via the projections $X_t(h_\infty) = x_t, A_t(h_\infty) = a_t, B_t(h_\infty) = b_t$, where $P_{\tilde{\pi}_0}^{\pi^1, \pi^2}$ is uniquely determined by π^1, π^2 and $\tilde{\pi}_0$ by

Ionescu Tulcea's theorem [5, Proposition 7.28]. When $\tilde{\pi}_0 = \delta_{x_0}$ (the dirac measure at x_0), $x_0 \in X$, we simply write this probability measure as $P_{x_0}^{\pi_1, \pi_2}$. For $h_t \in H_t, a \in A, b \in B$, we have

$$\begin{aligned} P_{x_0}^{\pi_1, \pi_2}(X_0 = x_0) &= 1, \\ P_{x_0}^{\pi_1, \pi_2}(X_{t+1} \in E | h_t, A_t = a, B_t = b) &= P(E | x_t, a, b) \quad \forall E \in \mathcal{X}, \\ P_{x_0}^{\pi_1, \pi_2}(A_t \in F, B_t \in G | h_t) &= \pi_t^1(h_t)(F) \pi_t^2(h_t)(G) \quad \forall F \in \mathcal{A}, \quad \forall G \in \mathcal{B}. \end{aligned}$$

Let the corresponding expectation operator be denoted by $E_{\tilde{\pi}_0}^{\pi_1, \pi_2}(E_{x_0}^{\pi_1, \pi_2})$ with respect to the probability measure $P_{\tilde{\pi}_0}^{\pi_1, \pi_2}(P_{x_0}^{\pi_1, \pi_2})$. Now from [13] we know that under any $(\Phi \times \Psi) \in S_1 \times S_2$, the corresponding stochastic process X_t is a Markov process.

Ergodic cost criterion: We now define the risk-sensitive ergodic cost criterion for nonzero-sum discrete-time game. Let (X_t, A_t, B_t) be the corresponding process with $X_0 = x \in X$ and $\theta > 0$ be the risk-sensitive parameter. For a pair of strategies $(\pi^1, \pi^2) \in (\Pi_1 \times \Pi_2)$, the risk-sensitive ergodic cost criterion for player $i = 1, 2$ is given by

$$J_i^{\pi^1, \pi^2}(x) = \limsup_{n \rightarrow \infty} \frac{1}{n} \ln E_x^{\pi^1, \pi^2} \left[e^{\theta \sum_{t=0}^{n-1} c_i(x_t, a_t, b_t)} \right] \quad (2.1)$$

Since the risk-sensitive parameter remains the same throughout, we assume without loss of generality that $\theta = 1$. Note that, $J_i^{\pi^1, \pi^2}$ for $i = 1, 2$ are bounded as our cost functions are bounded.

Nash equilibrium: A pair of strategies $(\pi^{*1}, \pi^{*2}) \in \Pi_1 \times \Pi_2$ is called a Nash equilibrium (for the ergodic cost criterion) if

$$J_1^{\pi^{*1}, \pi^{*2}}(x) \leq J_1^{\pi^1, \pi^{*2}}(x) \text{ for all } \pi^1 \in \Pi_1 \text{ and } x \in X$$

and

$$J_2^{\pi^{*1}, \pi^{*2}}(x) \leq J_2^{\pi^{*1}, \pi^2}(x) \text{ for all } \pi^2 \in \Pi_2 \text{ and } x \in X.$$

Our primary goal is to establish the existence of a Nash equilibrium in stationary strategies.

For $i = 1, 2$, let us define transition measures $\tilde{P}_i$ from $X \times A \times B \rightarrow \mathcal{P}(X)$ by

$$\tilde{P}_i(dy | x, a, b) = e^{c_i(x, a, b)} P(dy | x, a, b). \quad (2.2)$$

Moreover, define for $i = 1, 2$, $\varphi \in \mathcal{P}(A)$ and $\psi \in \mathcal{P}(B)$

$$P(dy | x, \varphi, \psi) := \int_B \int_A P(dy | x, a, b) \varphi(da) \psi(db)$$

and

$$\tilde{P}_i(dy | x, \varphi, \psi) := \int_B \int_A \tilde{P}_i(dy | x, a, b) \varphi(da) \psi(db).$$

Obviously $\tilde{P}_i$ for $i = 1, 2$ is in general is not a probability measure. The normalizing constant for $x \in X, \varphi \in \mathcal{P}(A)$, and $\psi \in \mathcal{P}(B)$ is given by

$$\tilde{c}_i(x, \varphi, \psi) = \int_X \tilde{P}_i(dy|x, \varphi, \psi) = \int_B \int_A e^{c_i(x, a, b)} \varphi(da) \psi(db). \quad (2.3)$$

Since $0 \leq c_i \leq \bar{c}$, the function $\tilde{c}_i$ is also bounded. More precisely $1 \leq \tilde{c}_i(x, \varphi, \psi) \leq e^{\bar{c}}$ for $i = 1, 2$ and for each $x \in X, \varphi \in \mathcal{P}(A)$ and $\psi \in \mathcal{P}(B)$. Thus for $i = 1, 2$

$$\hat{P}_i(\cdot|x, \varphi, \psi) := \frac{\tilde{P}_i(\cdot|x, \varphi, \psi)}{\tilde{c}_i(x, \varphi, \psi)} \quad (2.4)$$

defines a probability transition kernel and we also use the notation $\hat{c}_i(x, \varphi, \psi) := \ln \tilde{c}_i(x, \varphi, \psi)$ for $i = 1, 2$.

We will use the above transformations, to convert our optimality equations (3.5) and (3.9) into a well-known equation. This process is beneficial as it helps us to prove the existence of the unique solution to the optimality equations as we will see in the next section.

Define $\mathbb{B}(X)$, the space of all real valued bounded measurable functions on X endowed with the supremum norm $\|\cdot\|$. For a fixed $\varphi \in \mathcal{P}(A)$, and $v \in \mathbb{B}(X)$ define the operator:

$$Tv(x) = \inf_{\psi \in \mathcal{P}(B)} \left[\hat{c}_2(x, \varphi, \psi) + \ln \int_X e^{v(y)} \hat{P}_2(dy|x, \varphi, \psi) \right]. \quad (2.5)$$

Due to the dual representation of the exponential certainty equivalent [12, Lemma 3.3] it is possible to write (2.5) as

$$Tv(x) = \inf_{\psi} \sup_{\mu} \left[\hat{c}_2(x, \varphi, \psi) + \int_X v(y) \mu(dy) - I(\mu, \hat{P}_2(\cdot|x, \varphi, \psi)) \right], \quad (2.6)$$

where the supremum is over all probability measures $\mu \in \mathcal{P}(X)$ and $I(p, q)$ is the relative entropy of the two probability measures p, q which is defined by

$$I(p, q) := \int_X \ln \frac{dp}{dq} p(dx)$$

when $p \ll q$ and $+\infty$ otherwise. Note that the supremum in (2.6) is attained at the probability measure given by

$$\mu_{x\psi v}^{\varphi}(E) := \frac{\int_E e^{v(y)} \hat{P}_2(dy|x, \varphi, \psi)}{\int_X e^{v(y)} \hat{P}_2(dy|x, \varphi, \psi)} \quad (2.7)$$

for measurable set $E \subset X$. Obviously $\mu_{x\psi v}^{\varphi}$ can be interpreted as a transition kernel.

Let $v \in \mathbb{B}(X)$. The span semi-norm of v is defined as:

$$\|v\|_{sp} = \sup_{x \in X} v(x) - \inf_{x \in X} v(x). \quad (2.8)$$

In the last part of this section we make a couple of assumptions that will be in force throughout the rest of the paper. First we will consider the following continuity assumption which is quite standard in literature, see [4, 8, 10, 18] for instance. This will allow us to show that the map T defined in (2.5) is a contraction in the span semi-norm.

Assumption 2.1. *For a fixed $x \in X$ the transition measure P is strongly continuous in (a, b) , i.e. for all bounded and measurable $v : X \rightarrow \mathbb{R}$ we have that $(a, b) \mapsto \int_X v(y)P(dy|x, a, b)$ is continuous in (a, b) .*

It follows from (2.2) that $\tilde{P}_i(dy|x, a, b)$ is also strongly continuous in (a, b) for $i = 1, 2$ which lead us the following remark.

Remark 2.1. *Due to the fact that we consider for any metric space D , the space $\mathcal{P}(D)$ is endowed with the topology of weak convergence, from Assumption 2.1 it follows immediately that for all bounded and measurable functions $v : X \mapsto \mathbb{R}$ and a fixed $x \in X$ the map $(\varphi, \psi) \mapsto \int_X v(y)\hat{P}_i(dy|x, \varphi, \psi)$, $i = 1, 2$, is continuous in $(\varphi, \psi) \in \mathcal{P}(A) \times \mathcal{P}(B)$.*

Next we have the following ergodicity assumption.

Assumption 2.2.

(i) *There exists a real number $0 < \delta < 1$ such that*

$$\sup \|P(\cdot|x, \varphi, \psi) - P(\cdot|x', \varphi', \psi')\|_{\text{TV}} \leq 2\delta,$$

where the supremum is over all $x, x' \in X, \varphi, \varphi' \in \mathcal{P}(A), \psi, \psi' \in \mathcal{P}(B)$ and $\|\cdot\|_{\text{TV}}$ denotes the total variation norm.

(ii) *For $x \in X, a \in A, b \in B$, $P(\mathcal{O}|x, a, b) > 0$ for any open set $\mathcal{O} \subset X$.*

3 Solution to the optimality equations

In this section, we demonstrate that the operator T defined in (2.5) is a contraction. The fixed point of T corresponds to the solution of the optimality equation for player 1. In the latter part of this section, we define another operator U corresponding to player 2 and establish results analogous to those obtained for T .

Proposition 3.1. *Under Assumptions 2.1 and 2.2 the operator T maps $\mathbb{B}(X)$ to $\mathbb{B}(X)$ and for each $M > 0$, there exists a positive constant $\alpha(M) < 1$ such that for all $v_1, v_2 \in \mathbb{B}_M(X)$*

$$\|Tv_1 - Tv_2\|_{sp} \leq \alpha(M) \|v_1 - v_2\|_{sp},$$

where $\mathbb{B}_M(X) = \{v \in \mathbb{B}(X) : \|v\|_{sp} \leq M\}$.

Proof. From the definition of T it follows that T transforms $\mathbb{B}(X)$ into itself and the infimum is attained. For given functions $v_1, v_2 \in \mathbb{B}(X)$ and $x_1, x_2 \in X$, let $\Psi_1, \Psi_2 \in S_2$ be such that for $i = 1, 2$

$$Tv_i(x_i) = \sup_{\mu} [\hat{c}_2(x_i, \varphi, \Psi_i(x_i)) + \int_X v_i(y) \mu(dy) - I(\mu, \hat{P}_2(\cdot|x_i, \varphi, \Psi_i^*(x_i)))].$$

Then we obtain that

$$\begin{aligned} & (Tv_1)(x_2) - (Tv_2)(x_2) - ((Tv_1)(x_1) - (Tv_2)(x_1)) \\ & \leq \sup_{\mu} \left\{ \hat{c}_2(x_2, \varphi, \Psi_2(x_2)) + \int_X v_1(y) \mu(dy) - I\left(\mu, \hat{P}_2(\cdot|x_2, \varphi, \Psi_2(x_2))\right) \right\} \\ & \quad - \sup_{\mu} \left\{ \hat{c}_2(x_2, \varphi, \Psi_2(x_2)) + \int_X v_2(y) \mu(dy) - I\left(\mu, \hat{P}_2(\cdot|x_2, \varphi, \Psi_2(x_2))\right) \right\} \\ & \quad - \sup_{\mu} \left\{ \hat{c}_2(x_1, \varphi, \Psi_1(x_1)) + \int_X v_1(y) \mu(dy) - I\left(\mu, \hat{P}_2(\cdot|x_1, \varphi, \Psi_1(x_1))\right) \right\} \\ & \quad + \sup_{\mu} \left\{ \hat{c}_2(x_1, \varphi, \Psi_1(x_1)) + \int_X v_2(y) \mu(dy) - I\left(\mu, \hat{P}_2(\cdot|x_1, \varphi, \Psi_1(x_1))\right) \right\} \\ & \leq \hat{c}_2(x_2, \varphi, \Psi_2(x_2)) + \int_X v_1(y) \mu_{x_2\Psi_2v_1}^{\varphi}(dy) - I\left(\mu_{x_2\Psi_2v_1}^{\varphi}, \hat{P}_2(\cdot|x_2, \varphi, \Psi_2(x_2))\right) \\ & \quad - \hat{c}_2(x_2, \varphi, \Psi_2(x_2)) - \int_X v_2(y) \mu_{x_2\Psi_2v_1}^{\varphi}(dy) + I\left(\mu_{x_2\Psi_2v_1}^{\varphi}, \hat{P}_2(\cdot|x_2, \varphi, \Psi_2(x_2))\right) \\ & \quad - \hat{c}_2(x_1, \varphi, \Psi_1(x_1)) - \int_X v_1(y) \mu_{x_1\Psi_1v_2}^{\varphi}(dy) + I\left(\mu_{x_1\Psi_1v_2}^{\varphi}, \hat{P}_2(\cdot|x_1, \varphi, \Psi_1(x_1))\right) \\ & \quad + \hat{c}_2(x_1, \varphi, \Psi_1(x_1)) + \int_X v_2(y) \mu_{x_1\Psi_1v_2}^{\varphi}(dy) - I\left(\mu_{x_1\Psi_1v_2}^{\varphi}, \hat{P}_2(\cdot|x_1, \varphi, \Psi_1(x_1))\right) \\ & = \int_{\Delta} (v_1(y) - v_2(y)) \left(\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi} \right) (dy) + \int_{\Delta^c} (v_1(y) - v_2(y)) \left(\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi} \right) (dy) \\ & \leq \sup_{y \in X} (v_1(y) - v_2(y)) \left(\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi} \right) (\Delta) + \inf_{y \in X} (v_1(y) - v_2(y)) \left(\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi} \right) (\Delta^c) \\ & = \|v_1 - v_2\|_{sp} (\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi})(\Delta), \end{aligned}$$

where the set Δ comes from the Hahn-Jordan decomposition of $\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi}$ and Δ^c denotes the complement of Δ . Now, taking supremum over $x_1, x_2 \in X$ in the above set-up we have

$$\|Tv_1 - Tv_2\|_{sp} \leq \|v_1 - v_2\|_{sp} \sup_{E \in \mathcal{X}} \sup_{x_1, x_2 \in X} \sup_{\Psi_1, \Psi_2 \in S_2} (\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi})(E).$$

We claim that

$$\sup_{v_1, v_2; \|v_1\|_{sp}, \|v_2\|_{sp} \leq M} \sup_{E \in \mathcal{X}} \sup_{x_1, x_2 \in X} \sup_{\Psi_1, \Psi_2 \in S_2} (\mu_{x_2\Psi_2v_1}^{\varphi} - \mu_{x_1\Psi_1v_2}^{\varphi})(E) = \alpha(M) < 1. \quad (3.1)$$

Suppose (3.1) does not hold. Then there exists sequences $\{v_{1n}\}, \{v_{2n}\}$ with $\|v_{1n}\|_{sp} \leq M, \|v_{2n}\|_{sp} \leq M$, $\{E_n\}, E_n \in \mathcal{X}, \{x_{1n}\}, \{x_{2n}\}$ and $\{\Psi_{1n}\}, \{\Psi_{2n}\}$ such that

$$(\mu_{x_{2n}\Psi_{2n}v_{1n}}^{\varphi} - \mu_{x_{1n}\Psi_{1n}v_{2n}}^{\varphi})(E_n) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

As $\mu_{x_{2n}\Psi_{2n}v_{1n}}^\varphi$ and $\mu_{x_{1n}\Psi_{1n}v_{2n}}^\varphi$ are probability measures therefore

$$\mu_{x_{2n}\Psi_{2n}v_{1n}}^\varphi(E_n) \rightarrow 1 \quad \text{as } n \rightarrow \infty,$$

and

$$\mu_{x_{1n}\Psi_{1n}v_{2n}}^\varphi(E_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Since for each $x \in X, \psi \in \mathcal{P}(B)$ and $v \in \mathbb{B}(X)$ from (2.7) we get

$$e^{-\|v\|_{sp}} \hat{P}_2(E|x, \varphi, \psi) \leq \mu_{x\psi v}^\varphi(E),$$

we have

$$\hat{P}_2(E_n^c|x_{2n}, \varphi, \Psi_{2n}(x_{2n})) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and

$$\hat{P}_2(E_n|x_{1n}, \varphi, \Psi_{1n}(x_{1n})) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Consequently using (2.4) direct calculations imply

$$P(E_n^c|x_{2n}, \varphi, \Psi_{2n}(x_{2n})) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and

$$P(E_n|x_{1n}, \varphi, \Psi_{1n}(x_{1n})) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Hence

$$\lim_{n \rightarrow \infty} \left(P(E_n|x_{2n}, \varphi, \Psi_{2n}(x_{2n})) - P(E_n|x_{1n}, \varphi, \Psi_{1n}(x_{1n})) \right) = 1 \quad (3.2)$$

But from Assumption 2.2(i) we get, $\forall x_1, x_2 \in X, \forall \psi, \psi' \in \mathcal{P}(B), \forall E \in \mathcal{X}$

$$P(E|x, \varphi, \psi) - P(E|x', \varphi, \psi') \leq \delta,$$

which contradicts (3.2). Hence (3.1) holds and therefore the theorem also holds true. $\square$

We will now make additional assumptions to show that T is a global contraction in $\mathbb{B}_L(X)$.

Assumption 3.1. *There exists $\lambda \in \mathcal{P}(X)$ such that $P(\cdot|x, a, b) \ll \lambda$ for all $x \in X, a \in A$, and $b \in B$. Also let $h : X \times A \times B \times X \rightarrow \mathbb{R}$ be the Radon-Nikodym derivative of $P(\cdot|x, a, b)$ with respect to λ .*

Assumption 3.2.

$$\sup_{x, x' \in X} \sup_{y \in X} \sup_{a \in A} \sup_{b \in B} \frac{h(x, a, b, y)}{h(x', a, b, y)} = \kappa < \infty.$$

Lemma 3.1. *Under the Assumptions 2.1, 3.1 and 3.2 the operator T transforms $\mathbb{B}(X)$ into $\mathbb{B}_L(X)$, where $L = \ln \kappa + 3\bar{c}$. Furthermore, T is a global contraction in $\mathbb{B}_L(X)$.*

Proof. Notice that for a $v \in \mathbb{B}(X)$ we have

$$Tv(x) - Tv(x') \leq \sup_{\psi \in \mathcal{P}(B)} \left[\hat{c}_2(x, \varphi, \psi) - \hat{c}_2(x', \varphi, \psi) + \ln \frac{\int_X e^{v(y)} \hat{P}_2(dy|x, \varphi, \psi)}{\int_X e^{v(y)} \hat{P}_2(dy|x', \varphi, \psi)} \right]. \quad (3.3)$$

Now we get

$$\begin{aligned} & \ln \frac{\int_X e^{v(y)} \hat{P}_2(dy|x, \varphi, \psi)}{\int_X e^{v(y)} \hat{P}_2(dy|x', \varphi, \psi)} \\ &= \ln \frac{\int_X e^{v(y)} \int_B \int_A e^{c_2(x, a, b)} \frac{h(x, a, b, y)}{h(x', a, b, y)} h(x', a, b, y) \varphi(da) \psi(db) \lambda(dy)}{\int_X e^{v(y)} \int_B \int_A e^{c_2(x', a, b)} h(x', a, b, y) \varphi(da) \psi(db) \lambda(dy)} \cdot \frac{\tilde{c}_2(x', \varphi, \psi)}{\tilde{c}_2(x, \varphi, \psi)} \\ &\leq \ln \kappa + \ln \frac{\tilde{c}_2(x', \varphi, \psi)}{\tilde{c}_2(x, \varphi, \psi)} + \ln \frac{\int_X e^{v(y)} \int_B \int_A e^{c_2(x, a, b)} h(x', a, b, y) \varphi(da) \psi(db) \lambda(dy)}{\int_X e^{v(y)} \int_B \int_A e^{c_2(x', a, b)} h(x', a, b, y) \varphi(da) \psi(db) \lambda(dy)} \\ &\leq \ln \kappa + \ln e^{\bar{c}} + \ln \frac{e^{\bar{c}} \int_X e^{v(y)} \int_B \int_A h(x', a, b, y) \varphi(da) \psi(db) \lambda(dy)}{\int_X e^{v(y)} \int_B \int_A h(x', a, b, y) \varphi(da) \psi(db) \lambda(dy)} \\ &= \ln \kappa + 2\bar{c}. \end{aligned}$$

In the above expression the first equality follows from Assumption 3.1, first inequality follows from Assumption 3.2 and the last inequality follows from (2.3) and the fact that $0 \leq c_2 \leq \bar{c}$. So, from (3.3) we have

$$Tv(x) - Tv(x') \leq \ln \kappa + 3\bar{c}. \quad (3.4)$$

From (3.4) it follows that $\|Tv\|_{sp} \leq L$. Hence T is a global contraction in the span norm on $\mathbb{B}_L(X)$. $\square$

Before proceeding to the main theorem of this section, we briefly outline some main points. Suppose player 2 announces that he/she is going to employ a strategy $\Psi \in S_2$. In such a scenario, player 1 attempts to minimize

$$J_1^{\pi^1, \Psi}(x) = \limsup_{n \rightarrow \infty} \frac{1}{n} \ln E_x^{\pi^1, \Psi} \left[e^{\sum_{t=0}^{n-1} c_i(x_t, a_t, b_t)} \right]$$

over $\pi^1 \in \Pi_1$. Thus for player 1 it is a discrete-time Markov decision problem with risk sensitive ergodic cost. Player 2 go through equivalent situations when player 1 announces his strategy to be $\Phi \in S_1$. This leads us to the following theorem.

Theorem 3.1. *Suppose Assumptions 2.1, 2.2, 3.1 and 3.2 are satisfied. Then for $\Phi \in S_1$, there exists a unique solution pair $(\rho_2^*, v_2^*) \in \mathbb{R}_+ \times \mathbb{B}_L(X)$ with $v_2^*(x_0) = 0$, satisfying*

$$e^{v(x)+\rho} = \inf_{\psi \in \mathcal{P}(B)} \int_B \int_A e^{c_2(x, a, b)} \int_X e^{v(y)} P(dy|x, a, b) \Phi(x)(da) \psi(db). \quad (3.5)$$

In addition, a strategy $\Psi^* \in S_2$ is an optimal strategy of player 2 given player 1 chooses Φ if and only if (3.5) attains point-wise minimum at Ψ^* . Moreover,

$$\rho_2^* = \inf_{\pi^2 \in \Pi_2} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln E_x^{\Phi, \pi^2} \left[e^{\sum_{t=0}^{n-1} c_2(x_t, a_t, b_t)} \right]. \quad (3.6)$$

$$\left(:= \rho_2^{*\Phi} = \inf_{\pi^2 \in \Pi_2} J_2^{\Phi, \pi^2} \right)$$

Proof. Notice that (3.5) can be rewritten as

$$v(x) + \rho = \inf_{\psi \in \mathcal{P}(B)} \left[\hat{c}_2(x, \Phi(x), \psi) + \ln \int_X e^{v(y)} \hat{P}_2(dy|x, \Phi(x), \psi) \right]. \quad (3.7)$$

By Lemma 3.1, T is a global contraction in the span norm in $\mathbb{B}_L(X)$, so that it has a fixed point $\hat{v}_2$ in $\mathbb{B}_L(X)$ and $\hat{v}_2$ (which is unique up to an additive constant) and the constant $\rho_2^* = T\hat{v}_2 - \hat{v}_2$ are solutions to (3.7) and consequently to (3.5).

Let $v_2^*(x) = \hat{v}_2(x) - \hat{v}_2(x_0)$. Then $v_2^*(x_0) = 0$ and it can be easily seen that (ρ_2^*, v_2^*) satisfies (3.7). Since, $v_2^*(x) = \hat{v}_2(x) - \hat{v}_2(x_0)$ and $\|\hat{v}_2\|_{sp} \leq L$, so $\|v_2^*\|_{sp} \leq L$ as well.

Let (ρ', v') be another solution of (3.5) i.e., it satisfies $\rho' + v' = Tv'(x)$ with $v'(x_0) = 0$. Then clearly v' is also a span fixed point of T . Hence $v_2^*(x) - v'(x) = \text{constant}$. Since $v_2^*(x_0) - v'(x_0) = 0$, it follows that $v_2^* \equiv v'$. It then easily follows that $\rho_2^* = \rho'$.

The proof of the remaining part is analogous to the proof in [12, Theorem 2.1] which has been done for countable state space but can easily be extended to our general state space case. $\square$

For a fixed $\psi \in \mathcal{P}(B)$, and $v \in \mathbb{B}(X)$ define the operator:

$$Uv(x) = \inf_{\varphi \in \mathcal{P}(A)} \left[\hat{c}_1(x, \varphi, \psi) + \ln \int_X e^{v(y)} \hat{P}_1(dy|x, \varphi, \psi) \right]. \quad (3.8)$$

By similar arguments we can also show that U is a global contraction in the span norm in $\mathbb{B}_L(X)$ and the following theorem holds true.

Theorem 3.2. *Suppose Assumptions 2.1, 2.2, 3.1 and 3.2 are satisfied. Then for $\Psi \in S_2$, there exists a unique solution pair $(\rho_1^*, v_1^*) \in \mathbb{R}_+ \times \mathbb{B}_L(X)$ with $v_1^*(x_0) = 0$ (where x_0 is some fixed state), satisfying*

$$e^{v(x)+\rho} = \inf_{\varphi \in \mathcal{P}(A)} \int_B \int_A e^{c_1(x,a,b)} \int_X e^{v(y)} P(dy|x, a, b) \varphi(da) \Psi(x)(db). \quad (3.9)$$

In addition, a strategy $\Phi^* \in S_1$ is an optimal strategy of player 1 given player 2 chooses Ψ if and only if (3.9) attains point-wise minimum at Φ^* . Moreover,

$$\rho_1^* = \inf_{\pi^1 \in \Pi_1} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln E_x^{\pi^1, \Psi} [e^{\sum_{t=0}^{n-1} c_1(x_t, a_t, b_t)}]. \quad (3.10)$$

$$\left(\rho_1^{*\Psi} = \inf_{\pi^1 \in \Pi_1} J_1^{\pi^1, \Psi} \right)$$

4 Existence of Nash equilibrium

In this section we establish the existence of a pair of stationary equilibrium strategies for a nonzero-sum game. To this end we first outline a standard procedure for establishing the existence of a Nash equilibrium. From Theorem 3.2 it follows that given that player 2 is using the strategy $\Psi \in S_2$, we can find an optimal response $\Phi^* \in S_1$ for player 1. Clearly Φ^* depends on Ψ and moreover there may be several optimal responses for player 1 in S_1 . Analogous results holds for player 2 if player 1 announces that he is going to use a strategy $\Phi \in S_1$. Hence given a pair of strategies $(\Phi, \Psi) \in S_1 \times S_2$, we can find a set of pairs of optimal responses $\{(\Phi^*, \Psi^*) \in S_1 \times S_2\}$ via the appropriate pair of optimality equations described above. This defines a set-valued map. Clearly any fixed point of this set-valued map is a Nash equilibrium.

To ensure the existence of a Nash equilibrium, we first take the following separability assumptions.

Assumption 4.1.

(i) There exist two sub stochastic kernels

$$P_1 : X \times A \rightarrow \mathcal{P}(X), \quad P_2 : X \times B \rightarrow \mathcal{P}(X)$$

such that

$$P(\cdot \mid x, a, b) = P_1(\cdot \mid x, a) + P_2(\cdot \mid x, b), \quad x \in X, \quad a \in A, \quad b \in B.$$

Since $P \ll \lambda$, we have $P_1 \ll \lambda$ and $P_2 \ll \lambda$. Let h_1 and h_2 be the respective densities. We assume that for each $x, y \in X$, $h_1(x, \cdot, y)$ and $h_2(x, \cdot, y)$ are continuous.

(ii) For each $x \in X$,

$$\int_X \left[\sup_a |h_1(x, a, y)| + \sup_b |h_1(x, b, y)| \right] \lambda(dy) < \infty$$

Assumption 4.2. The reward functions $c_i, i = 1, 2$, are separable in action variables, i.e., there exist bounded continuous(in the second variable) functions

$$c_{i1} : X \times A \rightarrow \mathbb{R}, \quad c_{i2} : X \times B \rightarrow \mathbb{R}, \quad i = 1, 2,$$

such that

$$c_i(x, a, b) = c_{i1}(x, a) + c_{i2}(x, b), \quad x \in X, \quad a \in A, \quad b \in B.$$

Following [14] and [18] we topologize the spaces S_1 and S_2 with the topology of relaxed controls introduced in [20]. We identify two elements $\Phi, \hat{\Phi} \in S_1$ if $\Phi = \hat{\Phi}$ a.e. λ (where λ is as in Assumption 3.1). Let

$Y_1 = \{f : X \times A \rightarrow \mathbb{R} \mid f \text{ is measurable in the first argument and continuous in the second and there exists } g \in L^1(\lambda) \text{ such that } |f(x, a)| \leq g(x) \text{ for every } a \in A\}.$

Then Y_1 is a Banach space with norm [20]

$$\|f\|_W = \int_X \sup_a |f(x, a)| \lambda(dx).$$

Every $\Phi \in S_1$ (with the λ -a.e. equivalence relation) can be identified with the element $\Lambda_\Phi \in Y_1^*$ (the dual of Y_1) defined as

$$\Lambda_\Phi(f) = \int_X \int_A f(x, a) \Phi(x)(da) \lambda(dx).$$

Thus S_1 can be identified with a subset of Y_1^* . Equip S_1 with the weak-star topology. Then it can be shown as in [18] that S_1 is compact and metrizable. S_2 can be topologized analogously.

Next, we present the following lemmas, which play a pivotal role to show upper semi-continuity of a specific set-valued map which we have mentioned earlier.

Lemma 4.1. *Let, $\Phi_m \rightarrow \Phi \in S_1$ and $\Psi_m \rightarrow \Psi \in S_2$ in the weak star topology. Then under Assumption 4.2 for $i = 1, 2$, $\hat{c}_i(x, \Phi_m, \Psi_m) \rightarrow \hat{c}_i(x, \Phi, \Psi)$ as $m \rightarrow \infty$.*

Proof. We have for $i = 1, 2$,

$$\begin{aligned} \hat{c}_i(x, \Phi_m, \Psi_m) &= \ln \tilde{c}_i(x, \Phi_m, \Psi_m) \\ &= \ln \int_B \int_A e^{c_i(x, a, b)} \Phi_m(x)(da) \Psi_m(x)(db) \\ &= \ln \int_A e^{c_{i1}(x, a)} \Phi_m(x)(da) + \ln \int_B e^{c_{i2}(x, b)} \Psi_m(x)(db) \\ &= \ln \int_X 1 \cdot \lambda(dx) \int_A e^{c_{i1}(x, a)} \Phi_m(x)(da) + \ln \int_X 1 \cdot \lambda(dx) \int_B e^{c_{i2}(x, b)} \Psi_m(x)(db) \\ &= \ln \int_X \int_A \left(1 \cdot e^{c_{i1}(x, a)} \right) \Phi_m(x)(da) \lambda(dx) + \ln \int_X \int_B \left(1 \cdot e^{c_{i2}(x, b)} \right) \Psi_m(x)(db) \lambda(dx). \end{aligned}$$

Now by Assumption 4.2 and since $\Phi_m \rightarrow \Phi$ and $\Psi_m \rightarrow \Psi$ in the weak star topology, the result is immediate. $\square$

Lemma 4.2. *Suppose Assumptions 2.1, 3.1, 4.1 and 4.2 hold. Let $\{v_m\}$ be a uniformly bounded sequence in $\mathbb{B}(X)$ and $v \in \mathbb{B}(X)$ be a weak star limit point of $\{v_m\}$. If $\Phi_m \rightarrow \Phi \in S_1$ and $\Psi_m \rightarrow \Psi \in S_2$ in the weak star topology, then for each $x \in X$ and $i = 1, 2$,*

$$\int_X v_m(y) \hat{P}_i(dy|x, \Phi_m, \Psi_m) \rightarrow \int_X v(y) \hat{P}_i(dy|x, \Phi, \Psi) \quad \text{as } m \rightarrow \infty.$$

Proof. Note that

$$\begin{aligned} & \left| \int_X v_m(y) \hat{P}_i(dy|x, \Phi_m(x), \Psi_m(x)) - \int_X v(y) \hat{P}_i(dy|x, \Phi(x), \Psi(x)) \right| \\ & \leq \left| \int_X v_m(y) \hat{P}_i(dy|x, \Phi_m(x), \Psi_m(x)) - \int_X v(y) \hat{P}_i(dy|x, \Phi_m(x), \Psi_m(x)) \right| \\ & \quad + \left| \int_X v(y) \hat{P}_i(dy|x, \Phi_m(x), \Psi_m(x)) - \int_X v(y) \hat{P}_i(dy|x, \Phi(x), \Psi(x)) \right|. \end{aligned} \quad (4.1)$$

We claim that

$$\int_X v(y) \hat{P}_i(dy|x, \Phi_m, \Psi_m) \rightarrow \int_X v(y) \hat{P}_i(dy|x, \Phi, \Psi) \quad \text{as } m \rightarrow \infty. \quad (4.2)$$

Observe that, under Assumption 4.1(i), Assumption 4.2 and using (2.4) we get

$$\begin{aligned} & \int_X v(y) \hat{P}_i(dy|x, \Phi_m, \Psi_m) \\ & = \frac{\int_B \int_A \int_X v(y) e^{c_{i1}(x,a)} e^{c_{i2}(x,b)} P_1(dy|x, a) \Phi_m(x)(da) \Psi_m(x)(db)}{c_i(x, \Phi_m(x), \Psi_m(x))} + \\ & \quad \frac{\int_B \int_A \int_X v(y) e^{c_{i1}(x,a)} e^{c_{i2}(x,b)} P_2(dy|x, a) \Phi_m(x)(da) \Psi_m(x)(db)}{c_i(x, \Phi_m(x), \Psi_m(x))} \\ & = \frac{\int_A \int_X v(y) e^{c_{i1}(x,a)} h_1(x, a, y) \lambda(dy) \Phi_m(x)(da) \int_B e^{c_{i2}(x,b)} \Psi_m(x)(db)}{c_i(x, \Phi_m(x), \Psi_m(x))} + \\ & \quad \frac{\int_B \int_X v(y) e^{c_{i2}(x,b)} h_2(x, b, y) \lambda(dy) \Psi_m(x)(db) \int_A e^{c_{i1}(x,a)} \Phi_m(x)(da)}{c_i(x, \Phi_m(x), \Psi_m(x))}. \end{aligned}$$

By using Lemma 4.1 and since $\Phi_m \rightarrow \Phi \in S_1$ and $\Psi_m \rightarrow \Psi \in S_2$, under Assumption 4.1(i), (4.2) holds true, i.e. the second term in the right hand side of (4.1) goes to zero as $m \rightarrow \infty$. Now we show that the first one also goes to zero.

Again note that,

$$\begin{aligned}
& \left| \int_X v_m(y) \hat{P}_i(dy|x, \Phi_m(x), \Psi_m(x)) - \int_X v(y) \hat{P}_i(dy|x, \Phi_m(x), \Psi_m(x)) \right| \\
&= \left| \int_X \int_B \int_A \frac{v_m(y) - v(y)}{\tilde{c}_i(x, \Phi_m(x), \Psi_m(x))} e^{c_i(x, a, b)} P(dy|x, a, b) \Phi_m(x)(da) \Psi_m(x)(db) \right| \\
&\leq e^{\bar{c}} \int_B \int_A \left| \int_X (v_m(y) - v(y)) P(dy|x, a, b) \right| \Phi_m(x)(da) \Psi_m(x)(db) \\
&\leq e^{\bar{c}} \sup_{b \in B} \sup_{a \in A} \left| \int_X (v_m(y) - v(y)) P(dy|x, a, b) \right| \\
&= e^{\bar{c}} \sup_{b \in B} \sup_{a \in A} \left| \int_X (v_m(y) - v(y)) h(x, a, b, y) \lambda(dy) \right|,
\end{aligned}$$

where $h = h_1 + h_2$. From the compactness of A, B and the continuity of $h(x, \cdot, \cdot, y)$, it follows that for $m \in \mathbb{N}$,

$$V_m(x) := \left| \int_X (v_m(y) - v(y)) h(x, a_m, b_m, y) \lambda(dy) \right| = \sup_{b \in B} \sup_{a \in A} \left| \int_X (v_m(y) - v(y)) h(x, a, b, y) \lambda(dy) \right|,$$

for some sequences $\{a_m\} \in A$ and $\{b_m\} \in B$. We now prove that $V_m(x) \rightarrow 0$ as $m \rightarrow \infty$.

Since A and B are compact, without loss of generality, we can assume that

$$a_m \rightarrow a_0 \quad \text{and} \quad b_m \rightarrow b_0, \quad \text{for some } a_0 \in A \text{ and } b_0 \in B.$$

Note that, for each m , we have

$$\begin{aligned}
V_m(x) &\leq \left| \int (v_m(y) - v(y)) (h(x, a_m, b_m, y) - h(x, a_0, b_0, y)) \lambda(dy) \right| + \\
&\quad \left| \int (v_m(y) - v(y)) h(x, a_0, b_0, y) \lambda(dy) \right|. \tag{4.3}
\end{aligned}$$

Moreover,

$$\begin{aligned}
& \left| \int (v_m(y) - v(y)) (h(x, a_m, b_m, y) - h(x, a_0, b_0, y)) \lambda(dy) \right| \leq \\
& \|v_m(y) - v(y)\| \|h(x, a_m, b_m, \cdot) - h(x, a_0, b_0, \cdot)\|_{L^1(\lambda)}.
\end{aligned}$$

By Assumption 4.1(ii) and by the boundedness of $\{\|v_m(y) - v(y)\|\}$, from the last inequality we get that, first term on the right-hand side of (4.3) goes to zero as $m \rightarrow \infty$. Since, v is a weak star limit of $\{v_m\}$ and $h(x, a_0, b_0, \cdot) \in L_1(\lambda)$, so the second term on the right-hand side of (4.3) also goes to zero as $m \rightarrow \infty$. Thus, we have shown that $V_m(x) \rightarrow 0$ as $m \rightarrow \infty$. Hence the result followed. $\square$

Lemma 4.3. *Let for $M > 0$, $\{v_m\}$ be any sequence in $\mathbb{B}_M(X)$ with $v_m(x_0) = 0$ for all $m \in \mathbb{N}$. Then $\{v_m\}$ is uniformly bounded.*

Proof. Now for each fixed $m \in \mathbb{N}$ as,

$$\inf_{x \in X} v_m(x) \leq v_m(x_0) = 0$$

and $\|v_m(x)\|_{sp} \leq M$, using (2.8) we have

$$\sup_{x \in X} v_m(x) \leq M \quad (4.4)$$

Again as $\sup_{x \in X} v_m(x) \geq v_m(x_0)$ and $\|v_m(x)\|_{sp} \leq M$, using (2.8) we have

$$\inf_{x \in X} v_m(x) \geq -M \quad (4.5)$$

Now from, (4.4) and (4.5) for fixed $m \in \mathbb{N}$ and each $x \in X$ we get

$$-M \leq v_m(x) \leq M$$

Therefore, $\{v_m\}$ is a uniformly bounded sequence in $\mathbb{B}_M(X)$. □

For a fix $\Phi \in S_1$ let,

$$\begin{aligned} H(\Phi) &= \left\{ \Psi^* \in S_2 : \hat{c}_2(x, \Phi(x), \Psi^*(x)) + \ln \int_X e^{v_2^{*\Phi}(y)} \hat{P}_2(dy|x, \Phi(x), \Psi^*(x)) \right. \\ &\quad \left. = \inf_{\psi \in \mathcal{P}(B)} \left[\hat{c}_2(x, \Phi(x), \psi) + \ln \int_X e^{v_2^{*\Phi}(y)} \hat{P}_2(dy|x, \Phi(x), \psi) \right] \right\}, \end{aligned}$$

where $v_2^{*\Phi}$ is the unique solution of (3.5) corresponding to the strategy $\Phi \in S_1$.

Similarly for a fix $\Psi \in S_2$ let,

$$\begin{aligned} H(\Psi) &= \left\{ \Phi^* \in S_1 : \hat{c}_1(x, \Phi^*(x), \Psi(x)) + \ln \int_X e^{v_1^{*\Psi}(y)} \hat{P}_1(dy|x, \Phi^*(x), \Psi(x)) \right. \\ &\quad \left. = \inf_{\varphi \in \mathcal{P}(A)} \left[\hat{c}_1(x, \varphi, \Psi(x)) + \ln \int_X e^{v_1^{*\Psi}(y)} \hat{P}_1(dy|x, \varphi, \Psi(x)) \right] \right\}, \end{aligned}$$

where $v_1^{*\Psi}$ is the unique solution of (3.9) corresponding to the strategy $\Psi \in S_2$.

Remark 4.1. Since exponential and logarithmic functions are increasing functions, so $H(\Phi)$ and $H(\Psi)$ also have the following expressions:

$$\begin{aligned} H(\Phi) &= \left\{ \Psi^* \in S_2 : \int_B \int_A e^{c_2(x,a,b)} \int_X e^{v_2^{*\Phi}(y)} P(dy|x, a, b) \Phi(x)(da) \Psi^*(x)(db) \right. \\ &\quad \left. = \inf_{\psi \in \mathcal{P}(B)} \int_B \int_A e^{c_2(x,a,b)} \int_X e^{v_2^{*\Phi}(y)} P(dy|x, a, b) \Phi(x)(da) \psi(db) \right\}. \end{aligned}$$

$$\begin{aligned} H(\Psi) &= \left\{ \Phi^* \in S_1 : \int_B \int_A e^{c_1(x,a,b)} \int_X e^{v_1^{*\Psi}(y)} P(dy|x, a, b) \Phi^*(x)(da) \Psi(x)(db) \right. \\ &\quad \left. = \inf_{\varphi \in \mathcal{P}(A)} \int_B \int_A e^{c_1(x,a,b)} \int_X e^{v_1^{*\Psi}(y)} P(dy|x, a, b) \varphi(da) \Psi(x)(db) \right\}. \end{aligned}$$

Next set

$$H(\Phi, \Psi) = H(\Psi) \times H(\Phi).$$

Lemma 4.4. *Under Assumptions 2.1, 3.1, 4.1 and 4.2, $H(\Phi, \Psi)$ is a non-empty compact convex subset of $S_1 \times S_2$.*

Proof. From Remark 2.1, we know that $\int_X e^{v(y)} \hat{P}_2(dy|x, \Phi(x), \psi)$ is continuous on $\mathcal{P}(A) \times \mathcal{P}(B)$ for each $x \in X$. As B is compact, $\mathcal{P}(B)$ is also compact. Then it is easy to see that $H(\Phi)$ is non-empty. Let $\Psi_m^* \in H(\Phi)$ and as S_2 is compact, $\{\Psi_m^*\}$ has a convergent subsequence (denoted by the same sequence by abuse of notation) such that $\Psi_m^* \rightarrow \hat{\Psi} \in S_2$. Now for any $\psi \in \mathcal{P}(B)$

$$\begin{aligned} \hat{c}_2(x, \Phi(x), \Psi_m^*(x)) + \ln \int_X e^{v_2^{*\Phi}(y)} \hat{P}_2(dy|x, \Phi(x), \Psi_m^*(x)) \\ \leq \hat{c}_2(x, \Phi(x), \psi) + \ln \int_X e^{v_2^{*\Phi}(y)} \hat{P}_2(dy|x, \Phi(x), \psi). \end{aligned} \quad (4.6)$$

Using Lemma 4.1 and 4.2, from (4.6) we get for any $\psi \in \mathcal{P}(B)$

$$\begin{aligned} \hat{c}_2(x, \Phi(x), \hat{\Psi}(x)) + \ln \int_X e^{v_2^{*\Phi}(y)} \hat{P}_2(dy|x, \Phi(x), \hat{\Psi}(x)) \\ \leq \hat{c}_2(x, \Phi(x), \psi) + \ln \int_X e^{v_2^{*\Phi}(y)} \hat{P}_2(dy|x, \Phi(x), \psi) \quad \lambda - a.e. \end{aligned}$$

Hence it follows that $\hat{\Psi} \in H(\Phi)$ and therefore $H(\Phi)$ is closed. Since S_2 is a compact metric space, it follows that $H(\Phi)$ is also compact. Using Remark 4.1 the convexity of $H(\Phi)$ and $H(\Psi)$ follows easily. By analogous arguments, $H(\Psi)$ is also non-empty compact subset of S_2 . Hence $H(\Phi, \Psi)$ is a non-empty compact convex subset of $S_1 \times S_2$. $\square$

Next lemma proves the upper semi-continuity of a certain set valued map. This result will be useful in establishing the existence of a Nash equilibrium in the space of stationary Markov strategies.

Lemma 4.5. *Under Assumptions 2.1, 2.2, 3.1, 3.2, 4.1 and 4.2 the map $(\Phi, \Psi) \rightarrow H(\Phi, \Psi)$ from $S_1 \times S_2 \rightarrow 2^{S_1 \times S_2}$ is upper semi-continuous.*

Proof. Let $\Psi_m^* \in H(\Phi_m)$. $\{\Phi_m\}$ has a convergent subsequence (denoted by the same sequence by abuse of notation) such that $\Phi_m \rightarrow \bar{\Phi} \in S_1$ and similarly $\{\Psi_m^*\}$ has a subsequence too such that $\Psi_m^* \rightarrow \hat{\Psi} \in S_2$. Since, $\{v_2^{*\Phi_m}\} \in \mathbb{B}_L(X)$ and $\{\rho_2^{*\Phi_m}\}$ is bounded so without loss of generality let $v_2^{*\Phi_m} \rightarrow v_2$ in the weak star sense and $\rho_2^{*\Phi_m} \rightarrow \rho_2$. Then since,

$$\rho_2^{*\Phi_m} + v_2^{*\Phi_m} = \hat{c}_2(x, \Phi_m(x), \Psi_m^*(x)) + \ln \int_X e^{v_2^{*\Phi_m}(y)} \hat{P}_2(dy|x, \Phi_m(x), \Psi_m^*(x)) \quad (4.7)$$

Using Lemma 4.1, 4.2 and 4.3 it follows that

$$\rho_2 + v_2(x) = \hat{c}_2(x, \bar{\Phi}(x), \hat{\Psi}(x)) + \ln \int_X e^{v_2(y)} \hat{P}_2(dy|x, \bar{\Phi}(x), \hat{\Psi}(x)) \quad \lambda - a.e. \quad (4.8)$$

From (4.7) for any $\psi \in \mathcal{P}(B)$ we get

$$\rho_2^{*\Phi_m} + v_2^{*\Phi_m} \leq \hat{c}_2(x, \Phi_m(x), \psi) + \ln \int_X e^{v_2^{*\Phi_m}(y)} \hat{P}_2(dy|x, \Phi_m(x), \psi).$$

Again using Lemma 4.1, 4.2 and 4.3 it follows that

$$\rho_2 + v_2(x) \leq \hat{c}_2(x, \bar{\Phi}(x), \psi) + \ln \int_X e^{v_2(y)} \hat{P}_2(dy|x, \bar{\Phi}(x), \psi) \quad \lambda - a.e. \quad (4.9)$$

Let $v_2^*(x) = v_2(x) - v_2(x_0)$. Then from (4.9) we get, for any $\psi \in \mathcal{P}(B)$

$$\rho_2 + v_2^*(x) \leq \hat{c}_2(x, \bar{\Phi}(x), \psi) + \ln \int_X e^{v_2^*(y)} \hat{P}_2(dy|x, \bar{\Phi}(x), \psi) \quad \lambda - a.e. \quad (4.10)$$

and from (4.8) we get

$$\left. \begin{aligned} \rho_2 + v_2^*(x) &= \hat{c}_2(x, \bar{\Phi}(x), \hat{\Psi}(x)) + \ln \int_X e^{v_2^*(y)} \hat{P}_2(dy|x, \bar{\Phi}(x), \hat{\Psi}(x)) \\ &\geq \inf_{\psi \in \mathcal{P}(B)} \left[\hat{c}_2(x, \bar{\Phi}(x), \psi) + \ln \int_X e^{v_2^*(y)} \hat{P}_2(dy|x, \bar{\Phi}(x), \psi) \right] \end{aligned} \right\} \lambda - a.e. \quad (4.11)$$

Since (4.10) holds for every $\psi \in \mathcal{P}(B)$, from (4.10) and (4.11) we get

$$\rho_2 + v_2^*(x) = \inf_{\psi \in \mathcal{P}(B)} \left[\hat{c}_2(x, \bar{\Phi}(x), \psi) + \ln \int_X e^{v_2^*(y)} \hat{P}_2(dy|x, \bar{\Phi}(x), \psi) \right] \quad \lambda - a.e. \quad (4.12)$$

with $v_2^*(x_0) = 0$. Now by Theorem 3.1 we can say that (4.12) has unique solution $(\rho_2^{*\bar{\Phi}}, v_2^{*\bar{\Phi}})$ (corresponds to $\bar{\Phi} \in S_1$) satisfying $v_2^{*\bar{\Phi}}(x_0) = 0$. Therefore, $\rho_2 = \rho_2^{*\bar{\Phi}}$ and $v_2^* = v_2^{*\bar{\Phi}}$. Thus, from (4.11) and (4.12) it follows that $\hat{\Psi} \in H(\bar{\Phi})$.

Suppose $\Phi_m^* \in H(\Psi_m)$ and along a suitable subsequence $\Phi_m^* \rightarrow \hat{\Phi} \in S_1$ and $\Psi_m \rightarrow \bar{\Psi} \in S_2$. Then by similar arguments one can show that $\hat{\Phi} \in H(\bar{\Psi})$. This proves that $(\hat{\Phi}, \hat{\Psi}) \in H(\bar{\Phi}, \bar{\Psi})$. Hence the map $(\Phi, \Psi) \rightarrow H(\Phi, \Psi)$ is upper semi-continuous.

Now we are now all set to show the existence is of the Nash equilibrium which is directly follows by using Fan's fixed point theorem [9]. $\square$

Theorem 4.1. *Suppose that the Assumptions 2.1, 2.2, 3.1, 3.2, 4.1 and 4.2 are satisfied. Then there exists a Nash equilibrium in the space of stationary strategies $S_1 \times S_2$.*

Proof. Using Lemma 4.4 and 4.5 from Fan's fixed point theorem, it follows that there exists a fixed point $(\Phi^*, \Psi^*) \in (S_1 \times S_2)$ for the map $(\Phi, \Psi) \rightarrow H(\Phi, \Psi)$ from $S_1 \times S_2 \rightarrow 2^{S_1 \times S_2}$. This implies that $(\rho_1^{*\Psi^*}, v_1^{*\Psi^*}), (\rho_2^{*\Phi^*}, v_2^{*\Phi^*})$ satisfy the following coupled optimality equations:

$$\begin{aligned}
\rho_1^{*\Psi^*} + v_1^{*\Psi^*} &= \inf_{\varphi \in \mathcal{P}(A)} \left[\hat{c}_1(x, \varphi, \Psi^*(x)) + \ln \int_X e^{v_1^{*\Psi^*}}(y) \hat{P}_1(dy|x, \varphi, \Psi^*(x)) \right] \\
&= \left[\hat{c}_1(x, \Phi^*(x), \Psi^*(x)) + \ln \int_X e^{v_1^{*\Psi^*}}(y) \hat{P}_1(dy|x, \Phi^*(x), \Psi^*(x)) \right],
\end{aligned} \tag{4.13}$$

and

$$\begin{aligned}
\rho_2^{*\Phi^*} + v_2^{*\Phi^*} &= \inf_{\psi \in \mathcal{P}(B)} \left[\hat{c}_2(x, \Phi^*(x), \psi) + \ln \int_X e^{v_2^{*\Phi^*}}(y) \hat{P}_2(dy|x, \Phi^*(x), \psi) \right] \\
&= \left[\hat{c}_2(x, \Phi^*(x), \Psi^*(x)) + \ln \int_X e^{v_2^{*\Phi^*}(x)}(y) \hat{P}_1(dy|x, \Phi^*(x), \Psi^*(x)) \right].
\end{aligned} \tag{4.14}$$

Now by Theorem 3.1, from (4.13), it follows that

$$\rho_1^{*\Psi^*} = \inf_{\pi^1 \in \Pi_1} J_1^{\pi^1, \Psi^*} = J_1^{\Phi^*, \Psi^*}. \tag{4.15}$$

Similarly, by Theorem 3.2, from (4.14), it follows that

$$\rho_2^{*\Phi^*} = \inf_{\pi^2 \in \Pi_2} J_2^{\pi^2, \Phi^*} = J_2^{\Phi^*, \Psi^*}. \tag{4.16}$$

Thus, from equations (4.15) and (4.16), we get

$$\left. \begin{aligned} J_1^{\pi^1, \Psi^*} &\geq J_1^{\Phi^*, \Psi^*}, \forall \pi^1 \in \Pi_1, \\ J_2^{\pi^2, \Phi^*} &\geq J_2^{\Phi^*, \Psi^*}, \forall \pi^2 \in \Pi_2. \end{aligned} \right\} \tag{4.17}$$

Hence $(\Phi^*, \Psi^*) \in S_1 \times S_2$ obviously forms a λ -equilibrium stationary strategy (i.e., (4.17) holds in a set of λ -measure 1). Then by a construction analogous to Theorem 1 of [18] the existence of the desired Nash equilibrium follows. $\square$

References

- [1] Tansu Alpcan and Tamer Başar. *Network security*. Cambridge University Press, Cambridge, 2011. A decision and game-theoretic approach.
- [2] Arnab Basu and Mrinal K Ghosh. Zero-sum risk-sensitive stochastic games on a countable state space. *Stochastic Process. Appl.*, 124(1):961–983, 2014.
- [3] Arnab Basu and Mrinal K Ghosh. Nonzero-sum risk-sensitive stochastic games on a countable state space. *Mathematics of Operations Research*, 43(2):516–532, 2018.

- [4] N. Bauerle and U. Rieder. Zero-sum risk sensitive stochastic games. *Stoch. Processes and Their Appl*, 127:622–642, 2017.
- [5] Dimitri Bertsekas and Steven E Shreve. *Stochastic optimal control: the discrete-time case*, volume 5. Athena Scientific, 1996.
- [6] Tomasz R. Bielecki, Stanley R. Pliska, and Shuenn-Jyi Sheu. Risk sensitive portfolio management with Cox-Ingersoll-Ross interest rates: the HJB equation. *SIAM J. Control Optim.*, 44(5):1811–1843, 2005.
- [7] Anup Biswas and Vivek S. Borkar. Ergodic risk-sensitive control—a survey. *Annu. Rev. Control*, 55:118–141, 2023.
- [8] Giovanni B Di Masi and Lukasz Stettner. Risk-sensitive control of discrete-time markov processes with infinite horizon. *SIAM Journal on Control and Optimization*, 38(1):61–78, 1999.
- [9] Ky Fan. Fixed-point and minimax theorems in locally convex topological linear spaces. *Proceedings of the National Academy of Sciences*, 38(2):121–126, 1952.
- [10] Mrinal K Ghosh and Arunabha Bagchi. Stochastic games with average payoff criterion. *Applied Mathematics and Optimization*, 38:283–301, 1998.
- [11] Mrinal K Ghosh, Subrata Golui, Chandan Pal, and Somnath Pradhan. Discrete-time zero-sum games for markov chains with risk-sensitive average cost criterion. *Stochastic Processes and their Applications*, 158:40–74, 2023.
- [12] Daniel Hernández-Hernández and Steven I Marcus. Risk sensitive control of markov processes in countable state space. *Systems & control letters*, 29(3):147–155, 1996.
- [13] Onésimo Hernández-Lerma. *Adaptive Markov control processes*, volume 79. Springer Science & Business Media, 2012.
- [14] CJ Himmelberg, Thiruvengkatachari Parthasarathy, TES Raghavan, and FS Van Vleck. Existence of p-equilibrium and optimal stationary strategies in stochastic games. *Proceedings of the American Mathematical Society*, 60(1):245–251, 1976.
- [15] Ronald A. Howard and James E. Matheson. Risk-sensitive Markov decision processes. *Management Sci.*, 18:356–369, 1971/72.
- [16] Matthew O. Jackson. *Social and economic networks*. Princeton University Press, Princeton, NJ, 2008.

- [17] Andrzej S. Nowak and Eitan Altman. ϵ -equilibria for stochastic games with uncountable state space and unbounded costs. *SIAM J. Control Optim.*, 40(6):1821–1839, 2002.
- [18] T Parthasarathy. Existence of equilibrium stationary strategies in discounted stochastic games. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 114–127, 1982.
- [19] Tim Roughgarden. *Twenty Lectures on Algorithmic Game Theory*. Cambridge University Press, 2016.
- [20] Jack Warga. Functions of relaxed controls. *SIAM Journal on Control*, 5(4):628–641, 1967.
- [21] Qingda Wei and Xian Chen. Nonzero-sum expected average discrete-time stochastic games: the case of uncountable spaces. *SIAM J. Control Optim.*, 57(6):4099–4124, 2019.
- [22] Qingda Wei and Xian Chen. Risk-sensitive average equilibria for discrete-time stochastic games. *Dynamic Games and Applications*, 9:521–549, 2019.
- [23] Qingda Wei and Xian Chen. Nonzero-sum risk-sensitive average stochastic games: the case of unbounded costs. *Dynamic games and Applications*, 11:835–862, 2021.
- [24] Peter Whittle. Risk-sensitive linear/quadratic/gaussian control. *Advances in Applied Probability*, 13(4):764–777, 1981.