

Existence of ϵ -Nash Equilibria in Nonzero-Sum Borel Stochastic Games and Equilibria of Quantized Models

Naci Saldi^{a,*}, Gürdal Arslan^b, Serdar Yüksel^c

^a*Bilkent University, Department of Mathematics, Ankara, 06800, Türkiye*

^b*University of Hawaii at Manoa, Department of Electrical Engineering, Honolulu, 96822, HI, USA*

^c*Queen's University, Department of Mathematics and Statistics, Kingston, K7L 3N6, ON, Canada*

Abstract

Establishing the existence of exact or near Markov or stationary perfect Nash equilibria in nonzero-sum Markov games over Borel spaces remains a challenging problem, with few positive results to date. In this paper, we establish the existence of approximate Markov and stationary Nash equilibria for nonzero-sum stochastic games over Borel spaces, assuming only mild regularity conditions on the model. Our approach involves analyzing a quantized version of the game, for which we provide an explicit construction under both finite-horizon and discounted cost criteria. This work has significant implications for emerging applications such as multi-agent learning. Our results apply to both compact and non-compact state spaces. For the compact state space case, we first approximate the standard Borel model with a finite state-action model. Using the existence of Markov and stationary perfect Nash equilibria for these finite models under finite-horizon and discounted cost criteria, we demonstrate that these joint policies constitute approximate Markov and stationary perfect equilibria under mild continuity conditions on the one-stage costs and transition probabilities. For the non-compact state space case, we achieve similar results by first approximating the model with a compact-state model. Compared with previous results in the literature, which we comprehensively review, we provide more general and complementary conditions, along with explicit approximation models whose equilibria are ϵ -equilibria for the original model.

*Corresponding author

Keywords: Stochastic Games, Approximate Nash Equilibrium, Quantized Game Models.

1. Introduction

The existence of Markov or stationary perfect Nash equilibria for nonzero-sum Markov games is a challenging problem due to the complexities introduced by the nonzero-sum cost structure, where the gains or losses of one player do not directly offset those of another. Research in this area builds on Shapley's foundational work on stochastic games Shapley (1953), though Shapley's original model focused on zero-sum games. In the zero-sum case, the existence of a value and optimal strategies can be established more straightforwardly through the min-max theorem, leveraging the symmetric nature of the game between the players. However, in the nonzero-sum setting, the complexity increases significantly.

Fink's extension to Shapley's model Fink (1964) demonstrated that for finite state and action spaces, stationary Nash equilibria exist even in the nonzero-sum case for the discounted cost criterion. This result is proved using the Kakutani-Fan-Glicksberg fixed point theorem (see (Aliprantis and Border, 2006, Corollary 17.55)), applied to the best-response correspondence. This result was significant as it opened the door to analyzing more realistic scenarios where players' interests are not strictly opposed. However, the approach developed by Fink cannot be applied to models with Borel state spaces, as in this case, we cannot find an appropriate topology on the set of policies to satisfy the conditions of the Kakutani-Fan-Glicksberg fixed point theorem. In contrast, for zero-sum games, it is unnecessary to use the Kakutani-Fan-Glicksberg fixed point theorem; the existence of equilibria can be established via the minimax theorem (Fan, 1953, Theorem 1), even for models with standard Borel state spaces (see Nowak (2003)). In this regard, Mamer and Schilling (1986) (see also (Balder, 1988, Theorem 3.4) and (Hogeboom-Burr and Yüksel, 2021, Theorem 3.2)) have shown that saddle-point equilibria exist under mild conditions.

Research on non-zero-sum discounted Markov games with standard Borel spaces has been fragmented. Notably, Levy (2013); Levy and McLennan (2015) reported the non-existence of stationary equilibrium policies for such games with continuous spaces. Moreover, there have been very few positive results regarding the existence of stationary equilibria in discounted Markov games with continuous state spaces. For example, Himmelberg et al. (1976)

imposed restrictive separability conditions on the transition kernels and cost functions, Parthasarathy and Sinha (1989) required state independence of transitions, and Jaśkiewicz and Nowak (2016) presented conditions for equilibrium under policies with memory. We refer the reader to (Başar and Zaccour, 2018, Chapter 6) for a more comprehensive discussion on this topic.

The situation in the finite horizon case allows for more positive results than in the infinite horizon discounted cost scenario when it comes to proving the existence of equilibria. A prominent study on this topic is conducted by Rieder Rieder (1979), who examines finite-horizon Markov games within the framework of standard Borel spaces. His approach employs a direct method based on backward induction, combined with an ingenious measurable selection argument. This analysis is specifically tailored for finite horizons, where establishing the existence of equilibria is comparatively more straightforward. In Rieder’s work, the existence of Markov policies is established, which is a common consideration in finite-horizon settings. He imposes a setwise continuity condition on the transition kernel with respect to control actions.

As is common in the literature, if an exact solution cannot be found or its existence cannot be established, one seeks approximate solutions. To this end, in this paper, we investigate the existence of ε -Nash equilibria obtained through finite state-action approximations under the most general conditions known to us. Our primary motivation for addressing this problem stems from its applications in multi-agent learning algorithms, as explicitly discussed in Yongacoglu et al. (2023, 2024). These studies highlight the convergence to ε -equilibrium policies through policy revision processes along ε -satisficing paths Yongacoglu et al. (2024) (where an agent revises a policy only when they are not ε -satisfied). Specifically, the existence of ε -equilibria is a sufficient condition for ensuring the convergence of the independent learning algorithms developed in these studies for a large class of stage games.

On the existence of ε -equilibria for games with uncountable state and action spaces, there are only a few studies. Whitt Whitt (1980) examines approximations under conditions more stringent than ours, notably requiring a uniform version of total variation convergence of the transition probability, and does so without explicitly constructing the approximating models. The lack of explicit construction in Whitt’s work arises from the fact that he conducted his analysis through the dynamic programming principle using a monotone contraction operator framework.

In contrast, Nowak Nowak (1985) imposes conditions that are comple-

mentary to ours. Notably, he requires that the state space be a countably generated measurable space (whereas we assume the state space has a metric structure) and also assumes that the transition kernel has a density with respect to a reference measure. Using this density assumption, he treats the density as an element of a function-valued L_1 -space and establishes the existence of an approximate model with countably many states via the separability of this L_1 -space. Consequently, Nowak does not explicitly construct the approximating models, and the approximate model generally has countably many elements, unlike in our case.

In (Nowak, 1985, Remark 6.1), Nowak notes that to obtain a finite approximate model, the state space must be compact and the model components must be continuous on their domains. This observation is comparable to our findings for the compact-state case, with a significant distinction: in our approach, the density assumption is not required as our construction is explicit and differs from Nowak’s method. Building on similar ideas as in Nowak (1985), Nowak and Altman (2002) address the same approximation problem for unbounded one-stage costs under discounted and average cost criteria. For the average cost criterion, they assume geometric ergodicity type conditions as it is common for the average cost criterion. Once more, a significant difference between their approach and ours lies in the construction of the approximate model and the density assumption.

Our approximation result is achieved via an explicit finite-game construction, building on recent work Saldi et al. (2017, 2018) that developed finite approximations for MDPs with standard Borel spaces. These methods were initially applied under the assumption of weak continuity for the transition probabilities and were shown to yield near-optimal approximations. In our current study, given the game-theoretic nature of the problem, we impose more stringent conditions on the kernel than weak continuity. This is because, unlike standard weakly continuous MDPs—where value functions are continuous in the state—value functions in nonzero-sum games generally lack this regularity. Indeed, as Rieder suggests, in such cases, we may need to settle for merely measurable value functions. Nonetheless, our conditions are still less restrictive than total variation continuity, while being more demanding than setwise continuity. Additionally, we extend our analysis to encompass non-compact state spaces.

Contributions of the Paper.

- (i) We present conditions on the existence of near Markov and stationary

perfect Nash equilibria for nonzero-sum stochastic games with Borel spaces, under both finite-horizon and discounted cost criteria. Our results apply to both compact and non-compact state spaces. Our conditions complement and generalize the results reported in the literature, as reviewed above.

- (ii) Furthermore, as previously noted, we establish the existence of near Markov and stationary perfect Nash equilibria through a finite state-action model approximation of the original model. Our finite model construction is explicit, based on the method developed in Saldi et al. (2017, 2018) for obtaining finite approximations of MDPs with Borel spaces. Specifically, we construct a finite Markov game model and demonstrate that, for every $\varepsilon > 0$, a sufficiently fine approximation of the original model exists. This approximation ensures the existence of an equilibrium for the finite model (as guaranteed by Fink (1964); Rieder (1979)), which serves as an ε -Nash equilibrium for the original problem. Thus, our result not only establishes existence but also provides an explicit method to compute or learn near Markov perfect equilibria. This approach has implications for multi-agent learning problems involving general spaces and information structures beyond finite models Altabaa et al. (2023); Yongacoglu et al. (2024).
- (iii) By demonstrating continuous convergence as the approximation becomes finer, our contribution also establishes a positive result on the continuous dependence of equilibria in the refinement of information structures within Markov games—a question for which there are few positive results Hogeboom-Burr and Yüksel (2023).

Notation

For a metric space \mathbf{E} , the Borel σ -algebra (the smallest σ -algebra that contains the open sets of \mathbf{E}) is denoted by $\mathcal{B}(\mathbf{E})$. We let $B(\mathbf{E})$ and $C_b(\mathbf{E})$ denote the set of all bounded Borel measurable and continuous real functions on \mathbf{E} , respectively. For any $u \in C_b(\mathbf{E})$ or $u \in B(\mathbf{E})$, let $\|u\| := \sup_{e \in \mathbf{E}} |u(e)|$ which turns $C_b(\mathbf{E})$ and $B(\mathbf{E})$ into Banach spaces. Let $\mathcal{P}(\mathbf{E})$ denote the set of all probability measures on \mathbf{E} . A sequence $\{\mu_n\}$ of probability measures on \mathbf{E} is said to converge weakly (resp., setwise) (see Hernández-Lerma and Lasserre (2003)) to a probability measure μ if $\int_{\mathbf{E}} g(e) \mu_n(de) \rightarrow \int_{\mathbf{E}} g(e) \mu(de)$ for all $g \in C_b(\mathbf{E})$ (resp., for all $g \in B(\mathbf{E})$). For any $\mu, \nu \in \mathcal{P}(\mathbf{E})$, the total variation

distance between μ and ν , denoted as $\|\mu - \nu\|_{TV}$, is equivalently defined as

$$\|\mu - \nu\|_{TV} := 2 \sup_{D \in \mathcal{B}(\mathbb{E})} |\mu(D) - \nu(D)| = \sup_{\|g\| \leq 1} \left| \int_{\mathbb{E}} g(e) \mu(de) - \int_{\mathbb{E}} g(e) \nu(de) \right|.$$

Unless otherwise specified, the term ‘measurable’ will refer to Borel measurability in the rest of the paper.

2. Nonzero Sum Stochastic Games

A discrete-time nonzero sum stochastic game can be described by a tuple

$$(\mathbf{X}, \mathbf{A}^1, \dots, \mathbf{A}^N, c^1, \dots, c^N, p),$$

where Borel spaces (i.e., Borel subsets of complete and separable metric spaces) \mathbf{X} and $\{\mathbf{A}^i\}_{i=1}^N$ denote the *state* and *action* spaces, respectively. The *stochastic kernel*

$$p : \mathbf{X} \times \mathbf{A} \ni (x, \mathbf{a}) \mapsto p(\cdot | x, \mathbf{a}) \in \mathcal{P}(\mathbf{X})$$

denotes the *transition probability* of the next state given that previous state and actions are (x, \mathbf{a}) (see Hernández-Lerma and Lasserre (1996)), where

$$\mathbf{A} := \prod_{i=1}^N \mathbf{A}^i, \quad \mathbf{a} = (a^1, \dots, a^N).$$

Hence, it satisfies: (i) $p(\cdot | x, \mathbf{a})$ is an element of $\mathcal{P}(\mathbf{X})$ for all $(x, \mathbf{a}) \in \mathbf{X} \times \mathbf{A}$, and (ii) $p(D | \cdot, \cdot)$ is a measurable function from $\mathbf{X} \times \mathbf{A}$ to $[0, 1]$ for each $D \in \mathcal{B}(\mathbf{X})$. The *one-stage cost* function c^i for player i is a measurable function from $\mathbf{X} \times \mathbf{A}$ to \mathbb{R} .

Define the history spaces

$$\mathbf{H}_0 = \mathbf{X}, \quad \mathbf{H}_t = (\mathbf{X} \times \mathbf{A})^t \times \mathbf{X}, \quad t \geq 1$$

endowed with their product Borel σ -algebras generated by $\mathcal{B}(\mathbf{X})$ and $\mathcal{B}(\mathbf{A}^i)$, $i = 1, \dots, N$. A *policy* for player i is a sequence $\pi^i = \{\pi_t^i\}$ of stochastic kernels on \mathbf{A}^i given \mathbf{H}_t . The set of all policies for player i is denoted by Π^i . Let Φ^i denote the set of stochastic kernels on \mathbf{A}^i given \mathbf{X} , and let \mathbb{F}^i denote the set of all measurable functions from \mathbf{X} to \mathbf{A}^i . A *randomized Markov* policy for player i is a sequence $\pi^i = \{\pi_t^i\}$ of stochastic kernels on \mathbf{A}^i given \mathbf{X} . A

deterministic Markov policy is a sequence of stochastic kernels $\pi^i = \{\pi_t^i\}$ on \mathbf{A}^i given \mathbf{X} such that $\pi_t^i(\cdot | x) = \delta_{f_t(x)}(\cdot)$ for some $f_t \in \mathbb{F}^i$, where δ_z denotes the point mass at z . The set of randomized and deterministic Markov policies for player i are denoted by \mathbf{RM}^i and \mathbf{M}^i , respectively. A *randomized stationary* policy for player i is a constant sequence $\pi^i = \{\pi_t^i\}$ of stochastic kernels on \mathbf{A}^i given \mathbf{X} such that $\pi_t^i(\cdot | x) = \varphi(\cdot | x)$ for all t for some $\varphi \in \Phi^i$. A *deterministic stationary* policy is a constant sequence of stochastic kernels $\pi^i = \{\pi_t^i\}$ on \mathbf{A}^i given \mathbf{X} such that $\pi_t^i(\cdot | x) = \delta_{f(x)}(\cdot)$ for all t for some $f \in \mathbb{F}^i$. The set of randomized and deterministic stationary policies for player i are identified with the sets Φ^i and \mathbb{F}^i , respectively. Hence, we have

$$\mathbb{F}^i \subset \mathbf{M}^i \subset \Pi^i, \quad \Phi^i \subset \mathbf{RM}^i \subset \Pi^i, \quad \mathbb{F}^i \subset \Phi^i, \quad \mathbf{M}^i \subset \mathbf{RM}^i.$$

According to the Ionescu Tulcea theorem (see Hernández-Lerma and Lasserre (1996)), an initial distribution μ on \mathbf{X} and a joint policy $\boldsymbol{\pi} := (\pi^1, \dots, \pi^N)$ define a unique probability measure $P_\mu^\boldsymbol{\pi}$ on $\mathbf{H}_\infty = (\mathbf{X} \times \mathbf{A})^\infty$. The expectation with respect to $P_\mu^\boldsymbol{\pi}$ is denoted by $\mathbb{E}_\mu^\boldsymbol{\pi}$. If $\mu = \delta_x$, we write $P_x^\boldsymbol{\pi}$ and $\mathbb{E}_x^\boldsymbol{\pi}$ instead of $P_{\delta_x}^\boldsymbol{\pi}$ and $\mathbb{E}_{\delta_x}^\boldsymbol{\pi}$. For player i , the cost functions to be minimized in this paper are the finite-horizon cost and the β -discounted cost, respectively given by

$$J^i(\boldsymbol{\pi}, x) = \mathbb{E}_x^\boldsymbol{\pi} \left[\sum_{t=0}^{T-1} c^i(x_t, \mathbf{a}_t) \right],$$

$$J^i(\boldsymbol{\pi}, x) = \mathbb{E}_x^\boldsymbol{\pi} \left[\sum_{t=0}^{\infty} \beta^t c^i(x_t, \mathbf{a}_t) \right].$$

Remark 1. We observe that the infinite sum $\sum_{t=0}^{\infty} \beta^t c^i(x_t, \mathbf{a}_t)$ may not be finite or well-defined in the definition of J^i if c^i is assumed only to be measurable. However, additional assumptions introduced in subsequent sections guarantee the well-defined nature of J^i .

In the definition below, the cost is either finite-horizon or discounted.

Definition 1 (Nash equilibrium). A joint policy $\boldsymbol{\pi}^*$ is said to be ε -Nash equilibrium ($\varepsilon \geq 0$) if

$$J^i(\boldsymbol{\pi}^*, x) \leq \inf_{\pi^i \in \Pi^i} J^i(\boldsymbol{\pi}^{-i}, \pi^i, x) + \varepsilon \quad \forall x \in \mathbf{X},$$

for all $i = 1, \dots, N$, where $\boldsymbol{\pi}^{-i} := \boldsymbol{\pi} \setminus \{\pi^i\}$. If $\varepsilon = 0$, it is called Nash equilibrium.

In nonzero-sum stochastic games, the primary objective is to establish the existence of (and, if possible, compute or learn) a Markov perfect Nash equilibrium for the finite-horizon case and a stationary perfect Nash equilibrium for the discounted case. This task becomes especially challenging when dealing with uncountable Borel spaces in the state and action spaces. In contrast, for finite cases—where both the state and action spaces are finite—the existence of such equilibria can be established relatively easily, as shown in Fink (1964); Rieder (1979). To address the challenge of establishing approximate Markov or stationary perfect Nash equilibria in uncountable cases, we employ a finite approximation method. By approximating our model with a finite one, we prove that the stationary or Markov perfect Nash equilibrium of the finite model serves as an approximate equilibrium for the original problem. This approach not only provides existence results for near Markov or stationary equilibria but also enables effective computation and learning of these near equilibria using the finite model. To this end, we first present the construction of the finite model.

3. δ -Approximation of N -player Game

In this section, we construct the finite approximation of the game model introduced in the previous section. We impose the assumptions below on the components of the original nonzero sum stochastic game.

Assumption 1.

- (a) The one-stage cost functions c^i are in $C_b(\mathbf{X} \times \mathbf{A})$.
- (b) The stochastic kernel $p(\cdot | x, \mathbf{a})$ is setwise continuous in (x, \mathbf{a}) .
- (c) \mathbf{X} and \mathbf{A}^i , $i = 1, \dots, N$, are compact.

Let $d_{\mathbf{X}}$ and $d_{\mathbf{A}^i}$ denote the metric on \mathbf{X} and \mathbf{A}^i , respectively, for all $i = 1, \dots, N$. Fix any $\delta > 0$. Since the state space \mathbf{X} and action spaces \mathbf{A}^i are assumed to be compact, one can find finite sets $\mathbf{X}_\delta = \{x_1, \dots, x_{k_\delta}\}$ and $\mathbf{A}_\delta^i = \{a_1^i, \dots, a_{h_\delta^i}^i\}$ such that they are δ -nets in the corresponding uncountable spaces. Define functions Q_δ and $Q_{i,\delta}$ by

$$Q_\delta(x) := \arg \min_{z \in \mathbf{X}_\delta} d_{\mathbf{X}}(x, z),$$

$$Q_{i,\delta}(a^i) := \arg \min_{b^i \in \mathbf{A}_\delta^i} d_{\mathbf{A}^i}(a^i, b^i),$$

where ties are broken so that functions are measurable. Note that Q_δ induces a partition $\{\mathcal{S}_\delta^i\}_{i=1}^{k_\delta}$ on the state space \mathbf{X} given by

$$\mathcal{S}_\delta^i = \{x \in \mathbf{X} : Q_\delta(x) = x_i\},$$

with diameter $\text{diam}(\mathcal{S}_\delta^i) \leq 2\delta$. Let ν_δ be a probability measures on \mathbf{X} satisfying

$$\nu_\delta(\mathcal{S}_\delta^i) > 0 \text{ for all } i = 1, \dots, k_\delta.$$

We let ν_δ^i be the restriction of ν_δ to \mathcal{S}_δ^i defined by

$$\nu_\delta^i(\cdot) := \frac{\nu_\delta(\cdot)}{\nu_\delta(\mathcal{S}_\delta^i)}.$$

The measures ν_δ^i will be used to define a finite game model with resolution δ . To this end, the one-stage cost functions $c_\delta^i : \mathbf{X}_\delta \times \mathbf{A} \rightarrow \mathbb{R}$ and the transition probability p_δ on \mathbf{X}_δ given $\mathbf{X}_\delta \times \mathbf{A}$ are defined by

$$c_\delta^i(x_i, \mathbf{a}) := \int_{\mathcal{S}_\delta^i} c^i(x, \mathbf{a}) \nu_\delta^i(dx), \tag{1}$$

$$p_\delta(\cdot | x_i, \mathbf{a}) := \int_{\mathcal{S}_\delta^i} Q_\delta * p(\cdot | x, \mathbf{a}) \nu_\delta^i(dx),$$

where $Q_\delta * p(\cdot | x, \mathbf{a}) \in \mathcal{P}(\mathbf{X}_\delta)$ is the pushforward of the measure $p(\cdot | x, \mathbf{a})$ with respect to Q_δ . For each δ , we define δ -approximation of the original game as a finite nonzero sum stochastic game with the following components: \mathbf{X}_δ is the state space, $\{\mathbf{A}_\delta^i\}_{i=1}^N$ are the action spaces, p_δ is the transition probability and $\{c_\delta^i\}_{i=1}^N$ are the one-stage cost functions. History spaces, policies and cost functions are defined in a similar way as in the original model. To distinguish them from the original game model, we add δ as a subscript in each object for the finite model.

4. Finite-Horizon Cost

Here, we consider the approximation problem for the finite-horizon cost criterion. Throughout this section, Assumption 1 is assumed to hold. We

first introduce the best response mappings in the original and approximate models. Then, we state the approximation result. However, before proceeding, let us define the subgame perfect Nash equilibrium for the finite-horizon cost criterion.

Definition 2 (Subgame perfect Nash equilibrium). A joint policy $\boldsymbol{\pi}^*$ is said to be subgame perfect ε -Nash equilibrium ($\varepsilon \geq 0$) if, for each $i = 1, \dots, N$, we have

$$\mathbb{E}^{\boldsymbol{\pi}^*} \left[\sum_{k=t}^{T-1} c^i(x_k, \mathbf{a}_k) \mid h_t \right] \leq \inf_{\boldsymbol{\pi}^i \in \Pi^i} \mathbb{E}^{(\boldsymbol{\pi}^*, -^i, \boldsymbol{\pi}^i)} \left[\sum_{k=t}^{T-1} c^i(x_k, \mathbf{a}_k) \mid h_t \right] + \varepsilon,$$

for all $h_t \in \mathbf{H}_t$ and $t = 0, \dots, T-1$, where in the expectations starting from time t , policies prior to time t are considered irrelevant, while other policies that utilize information preceding time t rely on a fixed historical variable, denoted as h_t . If $\varepsilon = 0$, it is called subgame perfect Nash equilibrium.

We note that it suffices to consider Markovian policies in the infimum on the right-hand side of the expression in the above definition. Consequently, conditioning on the last state x_t in the history variable h_t is adequate on the right-hand side, as the past becomes irrelevant in such cases.

Best Response Mapping

Given some fixed Markov policies $\boldsymbol{\pi}^{-i}$ of all players except player i , the player i best response is characterized via dynamic programming principle:

$$T_t^{\boldsymbol{\pi}^{-i}} J_{t+1}^*(\boldsymbol{\pi}^{-i}; \cdot) = J_t^*(\boldsymbol{\pi}^{-i}; \cdot), \text{ for } t = 0, \dots, T-1, \quad (2)$$

where the operator $T_t^{\boldsymbol{\pi}^{-i}} : B(\mathbf{X}) \rightarrow B(\mathbf{X})$ is defined as

$$\begin{aligned} T_t^{\boldsymbol{\pi}^{-i}} J(x) &:= \min_{a^i \in \mathbf{A}^i} \left[c^i(x, \boldsymbol{\pi}_t^{-i}(x), a^i) + \int_{\mathbf{X}} J(y) p(dy|x, \boldsymbol{\pi}_t^{-i}(x), a^i) \right] \\ &= \min_{\gamma^i \in \mathcal{P}(\mathbf{A}^i)} \left[c^i(x, \boldsymbol{\pi}_t^{-i}(x), \gamma^i) + \int_{\mathbf{X}} J(y) p(dy|x, \boldsymbol{\pi}_t^{-i}(x), \gamma^i) \right] \end{aligned}$$

and $J_T^*(\boldsymbol{\pi}^{-i}; \cdot) = 0$. Here, with an abuse of notation, for any collection of probability measures $(\gamma^1, \dots, \gamma^N) \in \mathcal{P}(\mathbf{A}^1) \times \dots \times \mathcal{P}(\mathbf{A}^N)$, we define

$$c^i(x, \gamma^1, \dots, \gamma^N) := \int_{\mathbf{X}} c^i(x, a^1, \dots, a^N) \gamma^1(da^1) \otimes \dots \otimes \gamma^N(da^N)$$

$$p(\cdot|x, \gamma^1, \dots, \gamma^N) := \int_{\mathbf{X}} p(\cdot|x, a^1, \dots, a^N) \gamma^1(da^1) \otimes \dots \otimes \gamma^N(da^N).$$

In above recursion, $J_t^*(\boldsymbol{\pi}^{-i}; \cdot)$ is the optimal cost-to-go at time time t of the player i , if the policies of other players are fixed as $\boldsymbol{\pi}^{-i}$:

$$J_t^*(\boldsymbol{\pi}^{-i}; x) := \inf_{\pi^i \in \Pi^i} \mathbb{E}_x^{(\boldsymbol{\pi}^{-i}, \pi^i)} \left[\sum_{l=t}^{T-1} c^i(x_l, \mathbf{a}_l) \right], \quad (3)$$

where subscript x in the expectation means that $x_t = x$ and the stochastic kernels after time t in the policies are used only. For instance, $\{\pi_l^i\}_{l=0}^{t-1}$ are irrelevant in this case. If the measurable functions $\pi_t^{*,i}(\cdot; \boldsymbol{\pi}^{-i})$ ($t = 0, \dots, T-1$) from \mathbf{X} to $\mathcal{P}(\mathbf{A}^i)$ minimizes the expression in (2) for all $x \in \mathbf{X}$, then it is known that the Markov policy $\pi^{*,i}(\cdot; \boldsymbol{\pi}^{-i}) = \{\pi_t^{*,i}(\cdot; \boldsymbol{\pi}^{-i})\}_{t=0}^{T-1}$ is the optimal solution of the optimization problem in (3) for each $t = 0, \dots, T-1$ (again for each t , functions before time t are irrelevant). Hence, we can define the best response of player i to the joint policy $\boldsymbol{\pi}^{-i}$ as

$$\text{Best}_i(\boldsymbol{\pi}^{-i}) = \left\{ \pi^{*,i}(\cdot; \boldsymbol{\pi}^{-i}) : \pi_t^{*,i}(\cdot; \boldsymbol{\pi}^{-i}) \text{ (} t = 0, \dots, T-1 \text{) minimizes (2) } \forall x \in \mathbf{X} \right\}.$$

Using this, we define the best response map of all players as follows:

$$\text{Best} : \prod_{i=1}^N \text{RM}^i \ni \boldsymbol{\pi} \mapsto \prod_{i=1}^N \text{Best}_i(\boldsymbol{\pi}^{-i}) \in 2^{\prod_{i=1}^N \text{RM}^i}.$$

Therefore, a joint policy $\boldsymbol{\pi}^*$ is Markov perfect Nash equilibrium if $\boldsymbol{\pi}^* \in \text{Best}(\boldsymbol{\pi}^*)$.

For the approximate finite model, similar definitions can be made if we replace $(\mathbf{X}, \mathbf{A}^1, \dots, \mathbf{A}^N, c^1, \dots, c^N, p)$ with $(\mathbf{X}_\delta, \mathbf{A}_\delta^1, \dots, \mathbf{A}_\delta^N, c_\delta^1, \dots, c_\delta^N, p_\delta)$ and integral with summation. In this case, we also add δ as a subscript to the operators and the optimal cost-to-go functions.

Existence of Approximate Markov Perfect Nash Equilibrium

For any $\delta > 0$, by Rieder (1979), it is known that there exists a Markov perfect Nash equilibrium $\boldsymbol{\pi}_\delta^*$ for the finite δ -approximation of the original game problem. Hence, for all $i = 1, \dots, N$, we have

$$T_{\delta,t}^{\boldsymbol{\pi}_\delta^{*, -i}} J_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) = J_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot), \text{ for } t = 0, \dots, T-1, \quad (4)$$

where the operator $T_{\delta,t}^{\pi_{\delta}^{*, -i}} : B(\mathbf{X}_{\delta}) \rightarrow B(\mathbf{X}_{\delta})$ is defined as

$$\begin{aligned} & T_{\delta,t}^{\pi_{\delta}^{*, -i}} J(x_j) \\ & := \min_{a^i \in A_{\delta}^i} \left\{ \int_{\mathcal{S}_{\delta}^j} \left[c^i(x, \pi_{\delta,t}^{*, -i}(x_j), a^i) + \int_{\mathbf{X}} \hat{J}(y) p(dy|x, \pi_{\delta,t}^{*, -i}(x_j), a^i) \right] \nu_{\delta}^j(dx) \right\} \\ & = \min_{\gamma^i \in \mathcal{P}(A_{\delta}^i)} \left\{ \int_{\mathcal{S}_{\delta}^j} \left[c^i(x, \pi_{\delta,t}^{*, -i}(x_j), \gamma^i) + \int_{\mathbf{X}} \hat{J}(y) p(dy|x, \pi_{\delta,t}^{*, -i}(x_j), \gamma^i) \right] \nu_{\delta}^j(dx) \right\}, \end{aligned}$$

$\hat{J} = J \circ Q_{\delta}$, and $J_{\delta,T}^*(\pi_{\delta}^{*, -i}; \cdot) = 0$. For each $i = 1, \dots, N$, the minimum in (4) is achieved by the policies in Markov perfect Nash equilibrium π_{δ}^* .

We now extend the definition of the operators $T_{\delta,t}^{\pi_{\delta}^{*, -i}}$ to $B(\mathbf{X})$ as follows:

$$\begin{aligned} & \hat{T}_{\delta,t}^{\pi_{\delta}^{*, -i}} J(z) \\ & := \min_{a^i \in A_{\delta}^i} \left\{ \int_{\mathcal{S}_{\delta}^{i(z)}} \left[c^i(x, \pi_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathbf{X}} \hat{J}(y) p(dy|x, \pi_{\delta,t}^{*, -i}(z), a^i) \right] \nu_{\delta}^{i(z)}(dx) \right\} \\ & = \min_{\gamma^i \in \mathcal{P}(A_{\delta}^i)} \left\{ \int_{\mathcal{S}_{\delta}^{i(z)}} \left[c^i(x, \pi_{\delta,t}^{*, -i}(z), \gamma^i) + \int_{\mathbf{X}} \hat{J}(y) p(dy|x, \pi_{\delta,t}^{*, -i}(z), \gamma^i) \right] \nu_{\delta}^{i(z)}(dx) \right\}, \end{aligned}$$

$\hat{J} = J \circ Q_{\delta}$, and with an abuse of notation, we denote the extended policy $\pi_{\delta,t}^{*, -i} \circ Q_{\delta}$ as $\pi_{\delta,t}^{*, -i}$ in order not to complicate the notation further. Here, $i : \mathbf{X} \rightarrow \{1, \dots, k_{\delta}\}$ gives the index of the bin to which z belongs. One can prove that

$$\hat{T}_{\delta,t}^{\pi_{\delta}^{*, -i}} \hat{J}_{\delta,t+1}^*(\pi_{\delta}^{*, -i}; \cdot) = \hat{J}_{\delta,t}^*(\pi_{\delta}^{*, -i}; \cdot), \text{ for } t = 0, \dots, T-1, \quad (5)$$

where " $\hat{\cdot}$ " means piece-wise constant extensions of functions defined on \mathbf{X}_{δ} to \mathbf{X} .

To prove the next result, we need to put further conditions on the transition probability in addition to Assumption 1-(b). To this end, define the stochastic kernel $p^{\delta} : \mathbf{X} \times \mathbf{A} \rightarrow \mathcal{P}(\mathbf{X}_{\delta})$ for each $\delta \geq 0$ as

$$p^{\delta}(\cdot|x, \mathbf{a}) := Q_{\delta} * p(\cdot|x, \mathbf{a}).$$

Since $p(\cdot|x, \mathbf{a})$ is setwise continuous, the conditional probability $p(\mathcal{S}_{\delta}^j|x, \mathbf{a})$ is continuous in (x, \mathbf{a}) for each $j = 1, \dots, k_{\delta}$. Hence, if $(x_n, \mathbf{a}_n) \rightarrow (x, \mathbf{a})$, then

$$\lim_{n \rightarrow \infty} \|p^{\delta}(\cdot|x_n, \mathbf{a}_n) - p(\cdot|x, \mathbf{a})\|_{TV} = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_{\delta}} |p^{\delta}(\mathcal{S}_{\delta}^j|x_n, \mathbf{a}_n) - p(\mathcal{S}_{\delta}^j|x, \mathbf{a})| = 0.$$

Hence, $p^\delta(\cdot|x, \mathbf{a})$ is continuous in total variation norm. As X and \mathbf{A} are compact, $p^\delta(\cdot|x, \mathbf{a})$ is also uniformly continuous, and therefore, we can define the modulus of continuity of $p^\delta(\cdot|x, \mathbf{a})$ as

$$\omega_\delta(r) := \sup_{d_{\mathsf{X}}(x,y) + d_{\mathbf{A}}(\mathbf{a},\mathbf{b}) \leq r} \|p^\delta(\cdot|x, \mathbf{a}) - p^\delta(\cdot|y, \mathbf{b})\|_{TV},$$

which converges to zero as $r \rightarrow 0$. In the assumption below, we want this convergence to be fast enough.

Assumption 2.

(d) We suppose that $\lim_{\delta \rightarrow 0} \omega_\delta(2\delta) = 0$.

This additional assumption is true if the original transition probability $p(\cdot|x, \mathbf{a})$ is continuous in total variation norm.

Theorem 1. *Under Assumption 1 and Assumption 2, for each $i = 1, \dots, N$, we have*

$$\lim_{\delta \rightarrow 0} \|\hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - J_t^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| = 0 \quad \forall t = 0, \dots, T.$$

Proof. We prove the result by backward induction. Since the terminal cost at time T is zero for each problem, the result trivially holds.

Suppose that the statement is true for $t + 1$ and consider t . Then, we have

$$\begin{aligned} \|\hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - J_t^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| &= \|\hat{T}_{\delta,t}^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T_t^{\boldsymbol{\pi}_\delta^{*, -i}} J_{t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ &\leq \|\hat{T}_{\delta,t}^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T_t^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ &\quad + \|T_t^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T_t^{\boldsymbol{\pi}_\delta^{*, -i}} J_{t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\|, \end{aligned}$$

where the second term in the last expression converges to zero as $\delta \rightarrow 0$ by the induction hypothesis as the operator $T_t^{\boldsymbol{\pi}_\delta^{*, -i}}$ is non-expansive. Hence, it remains to prove that the first expression converges to zero as $\delta \rightarrow 0$. To this end, we define $K := T \sup_{i=1, \dots, N} \|c^i\|$. Then, we have

$$\begin{aligned} &\|\hat{T}_{\delta,t}^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T_t^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ &= \sup_{z \in \mathsf{X}} \left| \min_{a^i \in \mathsf{A}_\delta^i} \left\{ \int_{\mathcal{S}_\delta^i(z)} \left[c^i(x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathsf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right| \end{aligned}$$

$$\begin{aligned}
& - \min_{a^i \in A^i} \left[c^i(z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right] \Big| \\
\leq & \sup_{z \in \mathbf{X}} \left| \min_{a^i \in A^i} \left\{ \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right. \\
& \left. - \min_{a^i \in A^i} \left\{ \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right| \\
+ & \sup_{z \in \mathbf{X}} \left| \min_{a^i \in A^i} \left\{ \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right. \\
& \left. - \min_{a^i \in A^i} \left[c^i(z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right] \right| \\
\leq & \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), Q_{i,\delta}(a^i)) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), Q_{i,\delta}(a^i)) \right] \nu_\delta^{i(z)}(dx) \right. \\
& \left. - \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right| \\
+ & \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} c^i(x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \nu_\delta^{i(z)}(dx) - c^i(z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right| \\
+ & \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \nu_\delta^{i(z)}(dx) \right. \\
& \left. - \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right|.
\end{aligned}$$

In the last expression, the second term converges to zero as $\delta \rightarrow 0$ by uniform continuity of the function $c^i(x, \mathbf{a})$. The last term can be written as

$$\begin{aligned}
& \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} \sum_{y \in \mathbf{X}} J_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p^\delta(y|x, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \nu_\delta^{i(z)}(dx) \right. \\
& \left. - \sum_{y \in \mathbf{X}} J_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p^\delta(y|z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right|,
\end{aligned}$$

and so, it can be upper bounded by $K \omega_\delta(2\delta)$ as $\sup_{j=1, \dots, k_\delta} \text{diam}(\mathcal{S}_\delta^j) \leq 2\delta$, which converges to zero as $\delta \rightarrow 0$ by Assumption 2. In the last expression, the first term can be upper bounded by

$$\begin{aligned}
& \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| c^i(z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), Q_{i,\delta}(a^i)) + \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), Q_{i,\delta}(a^i)) \right. \\
& \left. - c^i(z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) - \int_{\mathbf{X}} \hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_{\delta,t}^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right| \\
\leq & \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| c^i(z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), Q_{i,\delta}(a^i)) - c^i(z, \boldsymbol{\pi}_{\delta,t}^{*, -i}(z), a^i) \right|
\end{aligned}$$

$$\begin{aligned}
& + \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| \int_{\mathbf{X}} \hat{J}_{\delta, t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_{\delta, t}^{*, -i}(z), Q_{i, \delta}(a^i)) \right. \\
& \quad \left. - \int_{\mathbf{X}} \hat{J}_{\delta, t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_{\delta, t}^{*, -i}(z), a^i) \right| \\
& = \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| c^i(z, \boldsymbol{\pi}_{\delta, t}^{*, -i}(z), Q_{i, \delta}(a^i)) - c^i(z, \boldsymbol{\pi}_{\delta, t}^{*, -i}(z), a^i) \right| \\
& + \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| \sum_{y \in \mathbf{X}} J_{\delta, t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p^\delta(y|z, \boldsymbol{\pi}_{\delta, t}^{*, -i}(z), Q_{i, \delta}(a^i)) \right. \\
& \quad \left. - \sum_{y \in \mathbf{X}} J_{\delta, t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p^\delta(y|z, \boldsymbol{\pi}_{\delta, t}^{*, -i}(z), a^i) \right|
\end{aligned}$$

The first term above converges to zero as $\delta \rightarrow 0$ again by uniform continuity of the function $c^i(x, \mathbf{a})$ and the second term can be upper bounded by $K \omega_\delta(2\delta)$ as $\sup_{a^i \in \mathbf{A}^i} d_{\mathbf{A}^i}(Q_{i, \delta}(a^i), a^i) \leq \delta$, which converges to zero as $\delta \rightarrow 0$ by Assumption 2. This completes the proof. \square

We now proceed to prove the key result of this section, which implies that the Markov perfect Nash equilibrium of the finite model, when extended to the original model, serves as an approximate Markov perfect equilibrium for the original game.

Theorem 2. *Under Assumption 1 and Assumption 2, for each $i = 1, \dots, N$, we have*

$$\lim_{\delta \rightarrow 0} \|J_t^i(\boldsymbol{\pi}_\delta^*, \cdot) - J_t^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| = 0 \quad \forall t = 0, \dots, T,$$

where $J_t^i(\boldsymbol{\pi}_\delta^*, \cdot)$ is the cost-to-go of player i at time t under the joint policy $\boldsymbol{\pi}_\delta^*$.

Proof. We again prove the result by backward induction. Since the terminal cost at time T is zero, the result trivially holds.

Suppose that the statement is true for $t + 1$ and consider t . Then, we have

$$\|J_t^i(\boldsymbol{\pi}_\delta^*, \cdot) - J_t^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| = \|T_t^{\boldsymbol{\pi}_\delta^{*, i}} J_{t+1}^i(\boldsymbol{\pi}_\delta^*, \cdot) - T_t^{\boldsymbol{\pi}_\delta^{*, -i}} J_{t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\|, \quad (6)$$

where

$$T_t^{\boldsymbol{\pi}_\delta^{*, i}} J(z) := c^i(z, \boldsymbol{\pi}_{\delta, t}^*(z)) + \int_{\mathbf{X}} J(y) p(dy|z, \boldsymbol{\pi}_{\delta, t}^*(z)).$$

Then, we can bound (6) via triangle inequality as follows

$$(6) \leq \|T_t^{\pi_{\delta}^{*,i}} J_{t+1}^i(\pi_{\delta}^*, \cdot) - \hat{T}_{\delta,t}^{\pi_{\delta}^{*,i}} \hat{J}_{\delta,t+1}^*(\pi_{\delta}^{*,i}; \cdot)\| \\ + \|\hat{T}_{\delta,t}^{\pi_{\delta}^{*,i}} \hat{J}_{\delta,t+1}^*(\pi_{\delta}^{*,i}; \cdot) - T_t^{\pi_{\delta}^{*,i}} J_{t+1}^*(\pi_{\delta}^{*,i}; \cdot)\| \quad (7)$$

The second term in the last expression converges to zero as $\delta \rightarrow 0$ by Theorem 1. For the first term, the minimum is achieved in $\hat{T}_{\delta,t}^{\pi_{\delta}^{*,i}} \hat{J}_{\delta,t+1}^*(\pi_{\delta}^{*,i}; \cdot)$ by $\pi_{\delta,t}^{*,i}$ as π_{δ}^* is a Nash equilibrium in the finite game. Hence, we can write

$$\hat{T}_{\delta,t}^{\pi_{\delta}^{*,i}} \hat{J}_{\delta,t+1}^*(\pi_{\delta}^{*,i}; \cdot) \\ = \min_{a^i \in A_{\delta}^i} \left\{ \int_{\mathcal{S}_{\delta}^{i(z)}} \left[c^i(x, \pi_{\delta,t}^{*,i}(z), a^i) + \int_{\mathcal{X}} \hat{J}_{\delta,t+1}^*(\pi_{\delta}^{*,i}; y) p(dy|x, \pi_{\delta,t}^{*,i}(z), a^i) \right] \nu_{\delta}^{i(z)}(dx) \right\} \\ = \int_{\mathcal{S}_{\delta}^{i(z)}} \left[c^i(x, \pi_{\delta,t}^*(z)) + \int_{\mathcal{X}} \hat{J}(y) p(dy|x, \pi_{\delta,t}^*(z)) \right] \nu_{\delta}^{i(z)}(dx).$$

Hence, the first term in (7) can be written as

$$\sup_{z \in \mathcal{X}} \left| \int_{\mathcal{S}_{\delta}^{i(z)}} \left[c^i(x, \pi_{\delta,t}^*(z)) + \int_{\mathcal{X}} \hat{J}_{\delta,t+1}^*(\pi_{\delta}^{*,i}; y) p(dy|x, \pi_{\delta,t}^*(z)) \right] \nu_{\delta}^{i(z)}(dx) \right. \\ \left. - \left[c^i(z, \pi_{\delta,t}^*(z)) + \int_{\mathcal{X}} J_{t+1}^i(\pi_{\delta}^*, y) p(dy|z, \pi_{\delta,t}^*(z)) \right] \right|.$$

Using exactly the same arguments that we used in the proof of Theorem 1, we can establish that this term converges to zero as $\delta \rightarrow 0$. This completes the proof. \square

Now, it is time to discuss the implications of Theorem 2. Note that for each $i = 1, \dots, N$, we have

$$J_0^*(\pi_{\delta}^{*,i}; x) = \inf_{\pi \in \Pi^i} \mathbb{E}_x^{(\pi_{\delta}^{*,i}, \pi^i)} \left[\sum_{t=0}^{T-1} c^i(x_t, \mathbf{a}_t) \right]. \quad (8)$$

Hence, Theorem 2 implies that for any $\varepsilon > 0$, there exists $\delta(\varepsilon)$ such that for any $\delta \leq \delta(\varepsilon)$, the policy π_{δ}^* is Markov perfect ε -Nash equilibrium. Moreover, since Theorem 2 is true for other t values greater than zero, we can conclude that Markov perfect ε -Nash equilibrium π_{δ}^* is also subgame perfect ε -Nash equilibrium.

5. Discounted Cost

In this section, we address the approximation problem related to the discounted cost criterion. We assume that Assumption 1 remains valid throughout this section. Initially, we present the best response mappings for both the original and approximate models. Following this, we outline the approximation result. However, prior to this, we again need to define the subgame perfect Nash equilibrium for the discounted cost criterion.

Definition 3 (Subgame perfect Nash equilibrium). A joint policy $\boldsymbol{\pi}^*$ is said to be subgame perfect ε -Nash equilibrium ($\varepsilon \geq 0$) if, for each $i = 1, \dots, N$, we have

$$\mathbb{E}^{\boldsymbol{\pi}^*} \left[\sum_{k=t}^{\infty} \beta^k c^i(x_k, \mathbf{a}_k) \mid h_t \right] \leq \inf_{\boldsymbol{\pi}^i \in \Pi^i} \mathbb{E}^{(\boldsymbol{\pi}^*, \boldsymbol{\pi}^i)} \left[\sum_{k=t}^{\infty} \beta^k c^i(x_k, \mathbf{a}_k) \mid h_t \right] + \varepsilon,$$

for all $h_t \in \mathbf{H}_t$ and $t = 0, \dots$, where in the expectations starting at time t , policies prior to time t are considered irrelevant, while other policies that utilize information preceding time t rely on a fixed historical variable, denoted as h_t . If $\varepsilon = 0$, it is called subgame perfect Nash equilibrium.

Note focusing solely on stationary policies in the infimum on the right side of the expression in the aforementioned definition is adequate. Consequently, conditioning on the latest state x_t in the history variable h_t suffices on the right-hand side, as earlier states become irrelevant in such cases.

Best Response Mapping

Given some fixed stationary policies $\boldsymbol{\pi}^{-i}$ of all players except player i , the player i best response is characterized via dynamic programming principle:

$$T^{\boldsymbol{\pi}^{-i}} J^*(\boldsymbol{\pi}^{-i}; \cdot) = J^*(\boldsymbol{\pi}^{-i}; \cdot), \quad (9)$$

where the operator $T^{\boldsymbol{\pi}^{-i}} : B(\mathbf{X}) \rightarrow B(\mathbf{X})$ is defined as

$$T^{\boldsymbol{\pi}^{-i}} J(x) := \min_{a^i \in \mathbf{A}^i} \left[c^i(x, \boldsymbol{\pi}^{-i}(x), a^i) + \beta \int_{\mathbf{X}} J(y) p(dy|x, \boldsymbol{\pi}^{-i}(x), a^i) \right].$$

Here, similar to the operators defined for the finite-horizon cost criterion, one can always perform the minimization over the set of probability measures on the action spaces, i.e. $\mathcal{P}(\mathbf{A}^i)$, which yields the same operator. However, to

avoid complicating the notation, we omit explicitly stating this each time we define such an operator, but we implicitly assume its ability to be defined and minimized over the set of probability measures as well.

It is straightforward to prove that $T^{\pi^{-i}}$ is β -contraction on $B(\mathbf{X})$ with respect to sup-norm. Hence, it has a unique fixed point by Banach fixed point theorem and this unique fixed point $J^*(\pi^{-i}; \cdot)$ is the optimal cost of the player i , if the policies of other players are fixed as π^{-i} :

$$J^*(\pi^{-i}; x) := \inf_{\pi^i \in \Pi^i} \mathbb{E}_x^{\pi^{-i}, \pi^i} \left[\sum_{t=0}^{\infty} \beta^t c^i(x_t, \mathbf{a}_t) \right]. \quad (10)$$

If the measurable function $\pi^{*,i}(\cdot; \pi^{-i})$ from \mathbf{X} to $\mathcal{P}(A^i)$ minimizes the expression in (9) for all $x \in \mathbf{X}$, then it is known that the stationary policy $\pi^{*,i}(\cdot; \pi^{-i})$ is the optimal solution of the optimization problem in (10). Hence, we can define the best response of player i to the joint policy π^{-i} as

$$\text{Best}_i(\pi^{-i}) = \{ \pi^{*,i}(\cdot; \pi^{-i}) : \pi^{*,i}(\cdot; \pi^{-i}) \text{ minimizes (9) for all } x \in \mathbf{X} \}.$$

Using this, we define the best response map of all players as follows:

$$\text{Best} : \prod_{i=1}^N \Phi^i \ni \boldsymbol{\pi} \mapsto \prod_{i=1}^N \text{Best}_i(\boldsymbol{\pi}^{-i}) \in 2^{\prod_{i=1}^N \Phi^i}.$$

Therefore, a joint policy $\boldsymbol{\pi}^*$ is stationary perfect Nash equilibrium if $\boldsymbol{\pi}^* \in \text{Best}(\boldsymbol{\pi}^*)$.

For the approximate finite model, similar definitions can be made if we replace $(\mathbf{X}, \mathbf{A}^1, \dots, \mathbf{A}^N, c^1, \dots, c^N, p)$ with $(\mathbf{X}_\delta, \mathbf{A}_\delta^1, \dots, \mathbf{A}_\delta^N, c_\delta^1, \dots, c_\delta^N, p_\delta)$ and integral with summation. In this case, we also add δ as a subscript to the operators and the optimal cost function.

Existence of Approximate Stationary Perfect Nash Equilibrium

For any $\delta > 0$, by Fink (1964), it is known that there exists a stationary perfect Nash equilibrium $\boldsymbol{\pi}_\delta^*$ for the finite δ -approximation of the original game problem. Hence, for all $i = 1, \dots, N$, we have

$$T_\delta^{\boldsymbol{\pi}_\delta^{*, -i}} J_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) = J_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot), \quad (11)$$

where the operator $T_\delta^{\boldsymbol{\pi}_\delta^{*, -i}} : B(\mathbf{X}_\delta) \rightarrow B(\mathbf{X}_\delta)$ is defined as

$$T_\delta^{\pi_\delta^{*, -i}} J(x_j) := \min_{a^i \in A_\delta^i} \left\{ \int_{S_\delta^j} \left[c^i(x, \pi_\delta^{*, -i}(x_j), a^i) + \beta \int_{\mathbf{X}} \hat{J}(y) p(dy|x, \pi_\delta^{*, -i}(x_j), a^i) \right] \right\},$$

where $\hat{J} = J \circ Q_\delta$. For each $i = 1, \dots, N$, the minimum in (11) is achieved by the policies in stationary perfect Nash equilibrium π_δ^* .

We now extend the definition of the operator $T_\delta^{\pi_\delta^{*, -i}}$ to $B(\mathbf{X})$ as follows:

$$\begin{aligned} & \hat{T}_\delta^{\pi_\delta^{*, -i}} J(z) \\ & := \min_{a^i \in A_\delta^i} \left\{ \int_{S_\delta^{i(z)}} \left[c^i(x, \pi_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}(y) p(dy|x, \pi_\delta^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\}, \end{aligned}$$

where $\hat{J} = J \circ Q_\delta$, and with an abuse of notation, we denote the extended policy $\pi_\delta^{*, -i} \circ Q_\delta$ as $\pi_\delta^{*, -i}$ in order not to complicate the notation further. Recall that $i : \mathbf{X} \rightarrow \{1, \dots, k_\delta\}$ gives the index of the bin to which z belongs. One can prove that

$$\hat{T}_\delta^{\pi_\delta^{*, -i}} \hat{J}_\delta^*(\pi_\delta^{*, -i}; \cdot) = \hat{J}_\delta^*(\pi_\delta^{*, -i}; \cdot), \quad (12)$$

where we recall that " $\hat{\cdot}$ " means piece-wise constant extensions of functions defined on X_δ to \mathbf{X} . To prove the next result, we again need to suppose that Assumption 2 holds.

Theorem 3. *Under Assumption 1 and Assumption 2, for each $i = 1, \dots, N$, we have*

$$\lim_{\delta \rightarrow 0} \|\hat{J}_\delta^*(\pi_\delta^{*, -i}; \cdot) - J^*(\pi_\delta^{*, -i}; \cdot)\| = 0.$$

Proof. To prove the result, we use contraction property of the operators $\hat{T}_\delta^{\pi_\delta^{*, -i}}$ and $T^{\pi_\delta^{*, -i}}$. Indeed, by Banach fixed point theorem, if we start with a common initial function $J_0 \in B(\mathbf{X})$, then

$$\left(\hat{T}_\delta^{\pi_\delta^{*, -i}} \right)^t J_0 =: \hat{J}_{\delta, t}^*(\pi_\delta^{*, -i}; \cdot) \rightarrow \hat{J}_\delta^*(\pi_\delta^{*, -i}; \cdot), \quad \left(T^{\pi_\delta^{*, -i}} \right)^t J_0 =: J_t^*(\pi_\delta^{*, -i}; \cdot) \rightarrow J^*(\pi_\delta^{*, -i}; \cdot)$$

in sup-norm as $t \rightarrow \infty$. Hence, by using induction, we first prove that

$$\lim_{\delta \rightarrow 0} \|\hat{J}_{\delta, t}^*(\pi_\delta^{*, -i}; \cdot) - J_t^*(\pi_\delta^{*, -i}; \cdot)\| = 0$$

for all $t \geq 0$. Then, the result follows from the triangle inequality.

Since the function J_0 at time zero is common, the result trivially holds. Suppose that the statement is true for t and consider $t + 1$. Then, we have

$$\begin{aligned} & \|\hat{J}_{\delta,t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - J_{t+1}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| = \|\hat{T}_\delta^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T^{\boldsymbol{\pi}_\delta^{*, -i}} J_t^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ & \leq \|\hat{T}_\delta^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ & \quad + \|T^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T^{\boldsymbol{\pi}_\delta^{*, -i}} J_t^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\|, \end{aligned}$$

where the second term in the last expression converges to zero as $\delta \rightarrow 0$ by the induction hypothesis as the operator $T^{\boldsymbol{\pi}_\delta^{*, -i}}$ is contraction. Hence, it remains to prove that the first expression converges to zero as $\delta \rightarrow 0$. To this end, we define $K := \sup_{i=1, \dots, N} \|c^i\|$. Then, we have

$$\begin{aligned} & \|\hat{T}_\delta^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - T^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ & = \sup_{z \in \mathbf{X}} \left| \min_{a^i \in A_\delta^i} \left\{ \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right. \\ & \quad \left. - \min_{a^i \in A^i} \left[c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right] \right| \\ & \leq \sup_{z \in \mathbf{X}} \left| \min_{a^i \in A_\delta^i} \left\{ \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right. \\ & \quad \left. - \min_{a^i \in A^i} \left\{ \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right| \\ & + \sup_{z \in \mathbf{X}} \left| \min_{a^i \in A^i} \left\{ \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right\} \right. \\ & \quad \left. - \min_{a^i \in A^i} \left[c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right] \right| \\ & \leq \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) \right] \nu_\delta^{i(z)}(dx) \right. \\ & \quad \left. - \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right] \nu_\delta^{i(z)}(dx) \right| \\ & + \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} c^i(x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \nu_\delta^{i(z)}(dx) - c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right| \\ & + \sup_{z \in \mathbf{X}} \sup_{a^i \in A^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \nu_\delta^{i(z)}(dx) \right. \\ & \quad \left. - \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right|. \end{aligned}$$

In the last expression, the second term converges to zero as $\delta \rightarrow 0$ by uniform

continuity of the function $c^i(x, \mathbf{a})$. The last term can be written as

$$\beta \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| \int_{\mathcal{S}_\delta^{i(z)}} \sum_{y \in \mathbf{X}} J_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p^\delta(y|x, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \nu_\delta^{i(z)}(dx) - \sum_{y \in \mathbf{X}} J_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p^\delta(y|z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right|,$$

and so, it can be upper bounded by $K \omega_\delta(2\delta)$ as $\sup_{j=1, \dots, k_\delta} \text{diam}(\mathcal{S}_\delta^j) \leq 2\delta$, which converges to zero as $\delta \rightarrow 0$ by Assumption 2. In the last expression, the first term can be upper bounded by

$$\begin{aligned} & \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) + \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) \right. \\ & \quad \left. - c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) - \beta \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right| \\ & \leq \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) - c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right| \\ & \quad + \beta \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) \right. \\ & \quad \quad \left. - \int_{\mathbf{X}} \hat{J}_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p(dy|z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right| \\ & = \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) - c^i(z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right| \\ & \quad + \beta \sup_{z \in \mathbf{X}} \sup_{a^i \in \mathbf{A}^i} \left| \sum_{y \in \mathbf{X}} J_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p^\delta(y|z, \boldsymbol{\pi}_\delta^{*, -i}(z), Q_{i,\delta}(a^i)) \right. \\ & \quad \quad \left. - \sum_{y \in \mathbf{X}} J_{\delta,t}^*(\boldsymbol{\pi}_\delta^{*, -i}; y) p^\delta(y|z, \boldsymbol{\pi}_\delta^{*, -i}(z), a^i) \right| \end{aligned}$$

The first term above converges to zero as $\delta \rightarrow 0$ again by uniform continuity of the function $c^i(x, \mathbf{a})$ and the second term can be upper bounded by $K \omega_\delta(2\delta)$ as $\sup_{a^i \in \mathbf{A}^i} d_{\mathbf{A}^i}(Q_{i,\delta}(a^i), a^i) \leq \delta$, which converges to zero as $\delta \rightarrow 0$ by Assumption 2. This completes the proof. \square

We will now demonstrate the main result of this section, which indicates that the stationary perfect Nash equilibrium from the finite model, when

applied to the original model, functions as an approximate stationary perfect equilibrium for the original game.

Theorem 4. *Under Assumption 1 and Assumption 2, for each $i = 1, \dots, N$, we have*

$$\lim_{\delta \rightarrow 0} \|J^i(\boldsymbol{\pi}_\delta^*, \cdot) - J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| = 0,$$

where $J^i(\boldsymbol{\pi}_\delta^*, \cdot)$ is the cost of player i under the joint policy $\boldsymbol{\pi}_\delta^*$.

Proof. We prove the result by contraction property of the operators. We first define the following operator on $B(\mathbf{X})$:

$$T^{\boldsymbol{\pi}_\delta^*, i} J(z) := c^i(z, \boldsymbol{\pi}_\delta^*(z)) + \beta \int_{\mathbf{X}} J(y) p(dy|z, \boldsymbol{\pi}_\delta^*(z)).$$

It is trivial to prove that $T^{\boldsymbol{\pi}_\delta^*, i}$ is β -contraction and the unique fixed point of it is $J^i(\boldsymbol{\pi}_\delta^*, \cdot)$. We also define the following operator on $B(\mathbf{X})$:

$$\hat{T}^{\boldsymbol{\pi}_\delta^*, i} J(z) := \int_{\mathcal{S}_\delta^{i(z)}} \left[c^i(x, \boldsymbol{\pi}_\delta^*(z)) + \beta \int_{\mathbf{X}} J(y) p(dy|x, \boldsymbol{\pi}_\delta^*(z)) \right] \nu_\delta^{i(z)}(dx).$$

This operator is also β -contraction and the unique fixed point of it is $\hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)$ since the minimum is achieved in $\hat{T}_\delta^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)$ by $\pi_\delta^{*, i}$ as $\boldsymbol{\pi}_\delta^*$ is a Nash equilibrium in the finite game; that is

$$\hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) = \hat{T}_\delta^{\boldsymbol{\pi}_\delta^{*, -i}} \hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) = \hat{T}^{\boldsymbol{\pi}_\delta^*, i} \hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot).$$

With these observations, we then have

$$\begin{aligned} & \|J^i(\boldsymbol{\pi}_\delta^*, \cdot) - J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ & \leq \|T^{\boldsymbol{\pi}_\delta^*, i} J^i(\boldsymbol{\pi}_\delta^*, \cdot) - T^{\boldsymbol{\pi}_\delta^*, i} J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| + \|T^{\boldsymbol{\pi}_\delta^*, i} J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - \hat{T}^{\boldsymbol{\pi}_\delta^*, i} J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ & \quad + \|\hat{T}^{\boldsymbol{\pi}_\delta^*, i} J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - \hat{T}^{\boldsymbol{\pi}_\delta^*, i} \hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| + \|\hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ & \leq \beta \|J^i(\boldsymbol{\pi}_\delta^*, \cdot) - J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| + \|T^{\boldsymbol{\pi}_\delta^*, i} J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - \hat{T}^{\boldsymbol{\pi}_\delta^*, i} J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \\ & \quad + (1 + \beta) \|J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot) - \hat{J}_\delta^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\| \end{aligned}$$

Hence we obtain

$$\|J^i(\boldsymbol{\pi}_\delta^*, \cdot) - J^*(\boldsymbol{\pi}_\delta^{*, -i}; \cdot)\|$$

$$\leq \frac{\|T^{\pi_\delta^*, i} J^*(\pi_\delta^{*, -i}; \cdot) - \hat{T}^{\pi_\delta^*, i} J^*(\pi_\delta^{*, -i}; \cdot)\| + (1 + \beta) \|J^*(\pi_\delta^{*, -i}; \cdot) - \hat{J}_\delta^*(\pi_\delta^{*, -i}; \cdot)\|}{1 - \beta}$$

The second term in the last expression converges to zero as $\delta \rightarrow 0$ by Theorem 3. For the first term, using exactly the same arguments that we used in the proof of Theorem 3, we can establish that this term converges to zero as $\delta \rightarrow 0$. This completes the proof. \square

Now, it is time to discuss the implications of Theorem 4. Note that for each $i = 1, \dots, N$, we have

$$J^*(\pi_\delta^{*, -i}; x) = \inf_{\pi \in \Pi^i} \mathbb{E}_x^{(\pi_\delta^{*, -i}, \pi^i)} \left[\sum_{t=0}^{\infty} \beta^t c^i(x_t, \mathbf{a}_t) \right]. \quad (13)$$

Hence, Theorem 4 implies that for any $\varepsilon > 0$, there exists $\delta(\varepsilon)$ such that for any $\delta \leq \delta(\varepsilon)$, the policy π_δ^* is stationary perfect ε -Nash equilibrium. Moreover, since Theorem 4 is still true if the initial time is greater than zero, we can conclude that stationary perfect ε -Nash equilibrium π_δ^* is also subgame perfect ε -Nash equilibrium.

6. Extension to Non-Compact State Spaces

In this section, we briefly explain how the results established in the previous sections can be extended to non-compact state stochastic games. We employ the following strategy: (i) first, we define a sequence of compact-state games to approximate the original game; (ii) then, we use the previous results to approximate the compact-state games with finite-state models; and (iii) finally, we prove the convergence of the finite-state models to the original model. Notably, steps (ii) and (iii) will be accomplished simultaneously.

We impose the assumptions below on the components of the stochastic game. With the exception of the local compactness of the state space, these are the same with Assumption 1.

Assumption 3.

- (a) The one-stage cost functions c^i are in $C_b(\mathbf{X} \times \mathbf{A})$.
- (b) The stochastic kernel $p(\cdot | x, \mathbf{a})$ is setwise continuous in (x, \mathbf{a}) .
- (c) \mathbf{X} is locally compact and $\{\mathbf{A}_i\}_{i=1}^N$ are compact.

6.1. Compact Approximation of N -player Game

Since \mathbf{X} is locally compact separable metric space, there exists a nested sequence of compact sets $\{K_n\}$ such that $K_n \subset \text{int } K_{n+1}$ and $\mathbf{X} = \bigcup_{n=1}^{\infty} K_n$ (Aliprantis and Border, 2006, Lemma 2.76, p. 58). Let $\{\nu_n\}$ be a sequence of probability measures and for each $n \geq 1$, $\nu_n \in \mathcal{P}(K_n^c)$. Similar to the finite-state game construction in Section 3, we define a sequence of compact-state games to approximate the original model.

To this end, for each n , let $\mathbf{X}_n = K_n \cup \{\Delta_n\}$, where $\Delta_n \in K_n^c$ is a so-called pseudo-state. We define the transition probability p_n on \mathbf{X}_n given $\mathbf{X}_n \times \mathbf{A}$ and the one-stage cost functions $c_n^i : \mathbf{X}_n \times \mathbf{A} \rightarrow \mathbb{R}$ by

$$p_n(\cdot | x, \mathbf{a}) = \begin{cases} p(\cdot \cap K_n | x, \mathbf{a}) + p(K_n^c | x, \mathbf{a}) \delta_{\Delta_n}, & \text{if } x \in K_n \\ \int_{K_n^c} (p(\cdot \cap K_n | z, \mathbf{a}) + p(K_n^c | z, \mathbf{a}) \delta_{\Delta_n}) \nu_n(dz), & \text{if } x = \Delta_n, \end{cases}$$

$$c_n^i(x, \mathbf{a}) = \begin{cases} c^i(x, \mathbf{a}), & \text{if } x \in K_n \\ \int_{K_n^c} c^i(z, \mathbf{a}) \nu_n(dz), & \text{if } x = \Delta_n. \end{cases}$$

With these definitions, compact-state non-zero sum stochastic game is defined as a stochastic game with the components $(\mathbf{X}_n, \mathbf{A}, p_n, c_n^1, \dots, c_n^N)$. History spaces, policies and cost functions are defined in a similar way as in the original model. To distinguish them from the original game model, we add n as a subscript in each object for the compact model.

In addition to Assumption 3, we suppose that the following is true.

Assumption 4. For each $n \geq 1$, the transition probability p_n satisfies Assumption 2.

This additional assumption is true if the original transition probability $p(\cdot | x, \mathbf{a})$ is continuous in total variation norm.

Note that under Assumption 3 and Assumption 4, for each $n \geq 1$, compact-state game model with state space \mathbf{X}_n satisfies Assumption 1 and Assumption 2. Hence, approximation results established in the previous sections are applicable to this game model. In the rest of this section, we will concentrate on the discounted cost criterion. However, a similar analysis can be applied to the finite-horizon cost criterion under the same set of assumptions. To avoid repetition, we will not include that analysis here.

For each $n \geq 1$, Theorem 4 guarantees the existence of a stationary perfect $\varepsilon(n)$ -Nash equilibrium $\boldsymbol{\pi}_n^*$ for a compact-state game with state space \mathbf{X}_n , derived from some finite game model, where $\varepsilon(n) \rightarrow 0$ as $n \rightarrow \infty$. Hence, we have the following:

$$\|J_n^i(\boldsymbol{\pi}_n^*, \cdot) - J_n^*(\boldsymbol{\pi}_n^{*,-i}, \cdot)\| \leq \varepsilon(n) \quad \forall i = 1, \dots, N.$$

Note that, for all $i = 1, \dots, N$, we have

$$T_n^{\boldsymbol{\pi}_n^{*,-i}} J_n^*(\boldsymbol{\pi}_n^{*,-i}, \cdot) = J_n^*(\boldsymbol{\pi}_n^{*,-i}, \cdot),$$

where the operator $T_n^{\boldsymbol{\pi}_n^{*,-i}} : B(\mathbf{X}_n) \rightarrow B(\mathbf{X}_n)$ is defined as

$$T_n^{\boldsymbol{\pi}_n^{*,-i}} J(x) := \min_{a^i \in \mathbf{A}^i} \left[c_n^i(x, \boldsymbol{\pi}_n^{*,-i}(x), a^i) + \beta \int_{\mathbf{X}_n} J(y) p_n(dy|x, \boldsymbol{\pi}_n^{*,-i}(x), a^i) \right].$$

We now extend the definition of the operator $T_n^{\boldsymbol{\pi}_n^{*,-i}}$ to $B(\mathbf{X})$ as follows:

$$\hat{T}_n^{\boldsymbol{\pi}_n^{*,-i}} J(x) := \min_{a^i \in \mathbf{A}^i} \left[\hat{c}_n^i(x, \boldsymbol{\pi}_n^{*,-i}(x), a^i) + \beta \int_{\mathbf{X}} J(y) \hat{p}_n(dy|x, \boldsymbol{\pi}_n^{*,-i}(x), a^i) \right],$$

where

$$\hat{p}_n(\cdot|x, a) = \begin{cases} p(\cdot|x, \mathbf{a}), & \text{if } x \in K_n \\ \int_{K_n^c} p(\cdot|z, \mathbf{a}) \nu_n(dz), & \text{if } x \in K_n^c, \end{cases}$$

$$\hat{c}_n^i(x, \mathbf{a}) = \begin{cases} c^i(x, \mathbf{a}), & \text{if } x \in K_n \\ \int_{K_n^c} c^i(z, \mathbf{a}) \nu_n(dz), & \text{if } x \in K_n^c. \end{cases}$$

One can prove that

$$\hat{T}_n^{\boldsymbol{\pi}_n^{*,-i}} \hat{J}_n^*(\boldsymbol{\pi}_n^{*,-i}, \cdot) = \hat{J}_n^*(\boldsymbol{\pi}_n^{*,-i}, \cdot),$$

where, in this case, " $\hat{\cdot}$ " means extensions of functions defined on \mathbf{X}_n to \mathbf{X} as follows:

$$\hat{J}(x) = J(x) \text{ if } x \in K_n, \quad \hat{J}(x) = J(\Delta_n) \text{ if } x \in K_n^c.$$

We can also extend policies in a similar manner, but to avoid complicating the notation, we will not use the notation " $\hat{\cdot}$ " in this case.

Theorem 5. *Under Assumption 3 and Assumption 4, for each $i = 1, \dots, N$, we have*

$$\lim_{n \rightarrow \infty} \|\hat{J}_n^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) - J^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K = 0$$

for any compact $K \subset \mathsf{X}$, where $\|\cdot\|_K$ is the sup-norm on the set K .

Proof. To prove the result, we use contraction property of the operators $\hat{T}_n^{\boldsymbol{\pi}_n^{*, -i}}$ and $T^{\boldsymbol{\pi}_n^{*, -i}}$. Indeed, by Banach fixed point theorem, if we start with a common initial function $J_0 \in B(\mathsf{X})$, then

$$\left(\hat{T}_n^{\boldsymbol{\pi}_n^{*, -i}}\right)^t J_0 =: \hat{J}_{n,t}^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) \rightarrow \hat{J}_n^*(\boldsymbol{\pi}_n^{*, -i}; \cdot), \quad \left(T^{\boldsymbol{\pi}_n^{*, -i}}\right)^t J_0 =: J_t^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) \rightarrow J^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)$$

in sup-norm as $t \rightarrow \infty$ (and so, in sup-norm on any compact set K as $t \rightarrow \infty$). Hence, by using induction, we first prove that

$$\lim_{n \rightarrow \infty} \|\hat{J}_{n,t}^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) - J_t^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K = 0$$

for all $t \geq 0$ and for any compact $K \subset \mathsf{X}$. Then, the result follows from the triangle inequality.

Since the function J_0 at time zero is common, the result trivially holds. Suppose that the statement is true for t and consider $t+1$. Fix any compact $K \subset \mathsf{X}$. By definition of \hat{p}_n and \hat{c}_n^i , there exists $n_0 \geq 1$ such that for all $n \geq n_0$, we have $\hat{p}_n = p$ and $\hat{c}_n^i = c^i$ on K as $K \subset K_n$. With this observation, for each $n \geq n_0$, we have

$$\begin{aligned} & \|\hat{J}_{n,t+1}^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) - J_{t+1}^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K \\ &= \sup_{x \in K} \left| \min_{a^i \in \mathsf{A}^i} \left[c^i(x, \boldsymbol{\pi}_n^{*, -i}(x), a^i) + \beta \int_{\mathsf{X}} \hat{J}_{n,t}^*(\boldsymbol{\pi}_n^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_n^{*, -i}(x), a^i) \right] \right. \\ & \quad \left. - \min_{a^i \in \mathsf{A}^i} \left[c^i(x, \boldsymbol{\pi}_n^{*, -i}(x), a^i) + \beta \int_{\mathsf{X}} J_t^*(\boldsymbol{\pi}_n^{*, -i}; dy) p(dy|x, \boldsymbol{\pi}_n^{*, -i}(x), a^i) \right] \right| \\ & \leq \beta \sup_{(x, a^i) \in K \times \mathsf{A}^i} \left| \int_{\mathsf{X}} \hat{J}_{n,t}^*(\boldsymbol{\pi}_n^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_n^{*, -i}(x), a^i) - \int_{\mathsf{X}} J_t^*(\boldsymbol{\pi}_n^{*, -i}; y) p(dy|x, \boldsymbol{\pi}_n^{*, -i}(x), a^i) \right|. \end{aligned} \tag{14}$$

Note that since p is setwise continuous, it is also weakly continuous. Therefore, the set of probability measures $\{p(\cdot|x, \boldsymbol{\pi}_n^{*, -i}(x), a^i)\}_{(x, n, a^i) \in K \times \mathbb{N} \times \mathsf{A}^i}$ is tight. Hence, for any $\epsilon > 0$, there exists a compact set $K_\epsilon \subset \mathsf{X}$ such that

$$\sup_{(x, n, a^i) \in K \times \mathbb{N} \times \mathsf{A}^i} p(K_\epsilon^c|x, \boldsymbol{\pi}_n^{*, -i}(x), a^i) \leq \epsilon.$$

Let $M := \sup_{i=1,\dots,N} \|c^i\|$. One can prove that

$$\|\hat{J}_{n,t}^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|, \|J_t^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\| \leq \frac{M}{1-\beta}.$$

In view of this, we can obtain

$$(14) \leq \beta \|\hat{J}_{n,t}^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) - J_t^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_{K_\epsilon} + \beta \frac{2M}{1-\beta} \epsilon.$$

The first term in the last expression converges to zero as $n \rightarrow \infty$ by the induction hypothesis. Since ϵ is arbitrary, this completes the proof. \square

We will now establish the central result of this section, which shows that the stationary perfect Nash equilibrium derived from the finite model, when applied to the original model, acts as an approximate stationary perfect equilibrium for the original game.

Theorem 6. *Under Assumption 3 and Assumption 4, for each $i = 1, \dots, N$, we have*

$$\lim_{n \rightarrow \infty} \|J^i(\boldsymbol{\pi}_n^*, \cdot) - J^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K = 0$$

for any compact $K \subset \mathbf{X}$, where $J^i(\boldsymbol{\pi}_n^*, \cdot)$ is the cost of player i under the joint policy $\boldsymbol{\pi}_n^*$.

Proof. We prove the result by contraction property of the operators. We first define the following operator on $B(\mathbf{X})$:

$$T^{\boldsymbol{\pi}_n^*, i} J(x) := c^i(x, \boldsymbol{\pi}_n^*(x)) + \beta \int_{\mathbf{X}} J(y) p(dy|x, \boldsymbol{\pi}_n^*(x)).$$

It is trivial to prove that $T^{\boldsymbol{\pi}_n^*, i}$ is β -contraction and the unique fixed point of it is $J^i(\boldsymbol{\pi}_n^*, \cdot)$. We also define the following operator on $B(\mathbf{X})$:

$$\hat{T}^{\boldsymbol{\pi}_n^*, i} J(x) := \hat{c}_n^i(x, \boldsymbol{\pi}_n^*(x)) + \beta \int_{\mathbf{X}} J(y) \hat{p}_n(dy|x, \boldsymbol{\pi}_n^*(x)).$$

This operator is also β -contraction and the unique fixed point of it is $\hat{J}_n^i(\boldsymbol{\pi}_n^*, \cdot)$, which is the extension of the cost $J_n^i(\boldsymbol{\pi}_n^*, \cdot)$ of player i in the compact-state game to the whole state space \mathbf{X} . Using exactly the same arguments that we used in the proof of Theorem 5, we can establish that

$$\lim_{n \rightarrow \infty} \|J^i(\boldsymbol{\pi}_n^*, \cdot) - \hat{J}_n^i(\boldsymbol{\pi}_n^*, \cdot)\|_K = 0$$

for any compact $K \subset \mathsf{X}$.

With these observations, we then have

$$\begin{aligned} & \|J^i(\boldsymbol{\pi}_n^*, \cdot) - J^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K \\ & \leq \|J^i(\boldsymbol{\pi}_n^*, \cdot) - \hat{J}_n^i(\boldsymbol{\pi}_n^*, \cdot)\|_K + \|\hat{J}_n^i(\boldsymbol{\pi}_n^*, \cdot) - \hat{J}_n^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K + \|\hat{J}_n^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) - J^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K \\ & \leq \|J^i(\boldsymbol{\pi}_n^*, \cdot) - \hat{J}_n^i(\boldsymbol{\pi}_n^*, \cdot)\|_K + \varepsilon(n) + \|\hat{J}_n^*(\boldsymbol{\pi}_n^{*, -i}; \cdot) - J^*(\boldsymbol{\pi}_n^{*, -i}; \cdot)\|_K \end{aligned}$$

The third term in the last expression converges to zero as $n \rightarrow \infty$ by Theorem 5. The first term converges to zero by above argument. By assumption $\varepsilon(n) \rightarrow 0$ as $n \rightarrow \infty$ as well. This completes the proof. \square

Now, it is time to discuss the implications of Theorem 6. Note that for each $i = 1, \dots, N$, we have

$$J^*(\boldsymbol{\pi}_n^{*, -i}; x) = \inf_{\pi \in \Pi^i} \mathbb{E}_x^{\pi_n^{*, -i}, \pi^i} \left[\sum_{t=0}^{\infty} \beta^t c^i(x_t, \mathbf{a}_t) \right]. \quad (15)$$

Hence, Theorem 6 implies that for any compact $K \subset \mathsf{X}$ and for any $\varepsilon > 0$, there exists $n(K, \varepsilon)$ such that for any $n \geq n(K, \varepsilon)$, the policy $\boldsymbol{\pi}_n^*$ is stationary perfect ε -Nash equilibrium if the initial points are in K . Moreover, since Theorem 6 is still true if the initial time is greater than zero, we can conclude that stationary perfect ε -Nash equilibrium $\boldsymbol{\pi}_n^*$ is also subgame perfect ε -Nash equilibrium if the initial points are in K .

7. Conclusion

In this paper, we have established the existence of near Markov and stationary perfect Nash equilibria for nonzero-sum stochastic games using finite state-action approximation method, under both finite-horizon and discounted cost criteria, addressing both compact and non-compact state spaces. For compact state spaces, our approach involves initial approximation using a finite state-action model. Leveraging the existence of Markov and stationary perfect Nash equilibria within these finite models, under the respective finite-horizon and discounted cost criteria, we have demonstrated that these joint policies serve as approximate Markov and stationary perfect equilibria, subject to specific continuity conditions on the one-stage costs and transition probabilities. In the case of non-compact state spaces, we have introduced a sequence of compact-state games to approximate the original game, subsequently employing prior findings to approximate these compact-state games

with finite-state models. Ultimately, we have validated the convergence of these finite-state models to the original model.

Building on our findings in establishing approximate Markov and stationary perfect Nash equilibria for nonzero-sum stochastic games, our next step is to integrate these results into the framework of multi-agent learning algorithms, as discussed in Yongacoglu et al. (2023, 2024). These algorithms emphasize the convergence towards ε -equilibrium policies through policy revision processes along ε -satisficing paths Yongacoglu et al. (2024). Our established existence of ε -equilibria serves as a foundational condition for ensuring the convergence of these independent learning algorithms across a broad spectrum of stage games. Moving forward, we aim to validate and enhance the practical applicability of our theoretical findings by implementing them within these learning algorithms.

References

- Aliprantis, C., Border, K., 2006. Infinite Dimensional Analysis. Berlin, Springer, 3rd ed.
- Altabaa, A., Yongacoglu, B., Yüksel, S., 2023. Decentralized multi-agent reinforcement learning for continuous-space stochastic games. arXiv preprint arXiv:2303.13539 (2023 American Control Conference) .
- Balder, E., 1988. Generalized equilibrium results for games with incomplete information. *Mathematics of Operations Research* 13, 265–276.
- Başar, T., Zaccour, G., 2018. Handbook of dynamic game theory. Springer.
- Fan, K., 1953. Minimax theorems. *Proceedings of the National Academy of Sciences* 39, 42–47.
- Fink, A.M., 1964. Equilibrium in a stochastic n -person game. *Journal of Science of the Hiroshima University, Series A-I (Mathematics)* 28, 89 – 93.
- Hernández-Lerma, O., Lasserre, J., 1996. Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer.
- Hernández-Lerma, O., Lasserre, J., 2003. Markov Chains and Invariant Probabilities. Birkhauser.

- Himmelberg, C., Parthasarathy, T., Raghavan, T., Vleck, F.V., 1976. Existence of p -equilibrium and optimal stationary strategies in stochastic games. *Proceedings of the American Mathematical Society* 60, 245–251.
- Hogeboom-Burr, I., Yüksel, S., 2021. Comparison of information structures for zero-sum games and a partial converse to Blackwell ordering in standard Borel spaces. *SIAM Journal on Control and Optimization* 59, 1781–1803.
- Hogeboom-Burr, I., Yüksel, S., 2023. Continuity properties of value functions in information structures for zero-sum and general games and stochastic teams. *SIAM Journal on Control and Optimization* 61.
- Jaśkiewicz, A., Nowak, A., 2016. Stationary almost Markov perfect equilibria in discounted stochastic games. *Math. Oper. Res.* 41, 430–441.
- Levy, Y., 2013. Discounted stochastic games with no stationary nash equilibrium: two examples. *Econometrica* 81, 1973–2007.
- Levy, Y.J., McLennan, A., 2015. Corrigendum to "discounted stochastic games with no stationary nash equilibrium: two examples". *Econometrica* 83, 1237–1252.
- Mamer, J.W., Schilling, K.E., 1986. A zero-sum game with incomplete information and compact action spaces. *Mathematics of Operations Research* 11, 627–631.
- Nowak, A., 1985. Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space. *Journal of Optimization Theory and Applications* 45, 591–602.
- Nowak, A., 2003. Zero-sum stochastic games with borel state spaces, in: *Stochastic games and applications*. Springer, pp. 77–91.
- Nowak, A.S., Altman, E., 2002. ε -equilibria for stochastic games with uncountable state space and unbounded costs. *SIAM Journal on Control and Optimization* 40, 1821–1839.
- Parthasarathy, T., Sinha, S., 1989. Existence of stationary equilibrium strategies in non-zero sum discounted stochastic games with uncountable state space and state-independent transitions. *International Journal of Game Theory* 18, 189–194.

- Rieder, U., 1979. Equilibrium plans for non-zero-sum markov games. *Game theory and related topics* , 91–101.
- Saldi, N., Linder, T., Yüksel, S., 2018. *Finite Approximations in Discrete-Time Stochastic Control: Quantized Models and Asymptotic Optimality*. Springer, Cham.
- Saldi, N., Yüksel, S., Linder, T., 2017. On the asymptotic optimality of finite approximations to Markov decision processes with Borel spaces. *Mathematics of Operations Research* 42, 945–978.
- Shapley, L.S., 1953. Stochastic games. *Proceedings of the National Academy of Sciences* 39, 1095–1100.
- Whitt, W., 1980. Representation and approximation of noncooperative sequential games. *SIAM Journal on Control and Optimization* 18, 33–48.
- Yongacoglu, B., Arslan, G., Yüksel, S., 2023. Satisficing paths and independent multi-agent reinforcement learning in stochastic games. *SIAM Journal on Mathematics of Data Science* (arXiv:2110.04638) .
- Yongacoglu, B., Arslan, G., Yüksel, S., 2024. Independent learning in mean-field games: Satisficing paths and convergence to subjective equilibria. to appear in *Journal of Machine Learning Research* (arXiv:2209.05703) .