

On Non-Zero Sum Game Considered on Solutions of a Hybrid System with Frequent Random Jumps

Ilaria Brunetti, Vladimir Gaitsgory, Eitan Altman

▶ **To cite this version:**

Ilaria Brunetti, Vladimir Gaitsgory, Eitan Altman. On Non-Zero Sum Game Considered on Solutions of a Hybrid System with Frequent Random Jumps. Dynamic Games and Applications, 2016. hal-01305355

HAL Id: hal-01305355
https://inria.hal.science/hal-01305355

Submitted on 20 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Non-Zero Sum Game Considered on Solutions of a Hybrid System with Frequent Random Jumps

Ilaria Brunetti*[1], Vladimir Gaitsgory†[2] and Eitan Altman‡[3]

[1]CERI/LIA, University of Avignon, 339 Chemin des Meinajaries, Avignon and INRIA Sophia Antipolis, 2004 Route des Lucioles, 06902, France
[2]Department of Mathematics, Macquarie University, Sydney, NSW 2109
[3]INRIA Sophia Antipolis, 2004 Route des Lucioles, 06902 and LINCS, 23 avenue d'Italie, 75013 Paris, France

Abstract

We study a non-zero sum game considered on the solutions of a hybrid dynamical system that evolves in continuous time and that is subjected to abrupt changes of parameters. The changes of the parameters are synchronized with (and determined by) the changes of the states/actions of two Markov decision processes, each of which is controlled by a player that aims at minimizing his or her objective function. The lengths of the time intervals between the "jumps" of the parameters are assumed to be small. We show that an asymptotic Nash equilibrium of such hybrid game can be constructed on the basis of a Nash equilibrium of a deterministic averaged dynamic game.

1 Introduction

In this paper, we study a non-zero sum game considered on the solutions of a hybrid dynamical system that evolves in continuous time and that is subjected to abrupt changes of parameters. The changes of the parameters are synchronized with (and determined by) the changes of the states/actions of two Markov decision processes (MDPs), each of which is controlled by a player that aims at minimizing his or her objective function. The lengths of the time intervals between the "jumps" of the parameters are assumed to be equal to ϵ (a small positive parameter). That is, the parameters are assumed to change their values frequently (the smaller is the parameter ϵ, the higher is the frequency).

Hybrid systems of this type arise in modelling admission control into telecommunication networks (see [16]). The states of the continuous time system may be taken to be the amounts of workload (i.e., the transmission time required by packets awaiting transmission) in the different nodes. The dynamics of these states is determined by the number, the routes and the types of packets that are present in the network. These can be described by a Markov chain

*ilaria.brunetti@inria.fr; An essential part of this paper was written while Ilaria Brunetti was visiting the Department of Mathematics of Macquarie University, Sydney, Australia

†vladimir.gaitsgory@mq.edu.au; the work of this author was supported by the ARC Discovery Grants DP130104432 and DP150100618

‡eitan.altman@inria.fr

with a finite number of states (parameters). The transition between the states (parameters) occur at the beginning or end of a packet transmission. The actions controlling these transitions are to accept or reject a new coming call and (if it is not rejected) to impose certain requirements for its routing, bandwidth and duration.

The fact that the parameters change their values on much faster scale than the continuously changing state variables allows one to employ an averaging method to show that the state trajectory of the hybrid system are approximated by solutions of a certain deterministic control system. In fact, the main result of this paper establishes that an asymptotic Nash equilibrium of the game considered on the solutions of the hybrid system (see Definition 2 in Section 2 below) can be constructed on the basis of a Nash equilibrium of the game considered on the solutions of the deterministic averaged system.

Note that problems of optimal control and zero-sum games considered on the solutions of the hybrid system similar to one we are dealing with have been studied in [2], [3], [4], [13], [16] (see also related results in [1], [5], [8] and [15]). The present paper continues this line of research by developing a technique allowing one to deal with non-zero sum games. An important feature of our result is that the information structures of the hybrid and the averaged games are different (this is in contrast to a result obtained for a zero-sum game with full information; see [13]). In fact, we show that an asymptotic Nash equilibrium of the hybrid game, in which each of the players chooses his/her actions on the basis of the full information of his or her states/actions histories can be constructed on the basis of an open loop Nash equilibrium controls of the averaged game.

From a broader perspective, although problems of optimal control and zero-sum games posed on solutions of dynamical systems evolving in different time scales (singularly perturbed systems) received a great deal of attention in the literature (see, e.g., [6], [11], [12] and references therein), to the best of our knowledge, no results on non-zero sum games for such systems have been obtained. Results presented in this paper indicate the way how to approach non-zero sum games for other classes of systems characterized by the presence of slow/fast dynamics separation.

The paper is organized as follows. In Sections 2 and 3, the hybrid and, respectively, averaged deterministic games are introduced. In section 4, the main results are stated (see Propositions 1 and 2), and in Section 5, these are proved.

2 Hybrid Game

Let the dynamics of the state vector $Z(t) \in \mathbb{R}^N$ be described by the equation

$$\dot{Z}(t) = f^1(Z(t), Y^1(t)) + f^2(Z(t), Y^2(t)), \quad Z(0) = z_0, \qquad t \in [0, 1], \tag{1}$$

where $f^i(\cdot,\cdot) : \mathbb{R}^N \times \mathbb{R}^{N_i} \to \mathbb{R}^N$, $i = 1, 2$, are continuous functions satisfying Lipschitz conditions in z (see Assumption 2 below). The functions $Y^i(t) \in \mathbb{R}^{N_i}$, $i = 1, 2$, are "controls" defined by two players. These controls are not chosen directly by the players. They are obtained as the result of the players controlling the transition probabilities of two associated stochastic discrete event systems described as follows.

The system i ($i = 1, 2$) has a finite state space X^i and it changes its states at discrete moments of time $t_j = j\epsilon$, $j = 0, 1, \ldots, \lfloor \epsilon^{-1} \rfloor$, where $\epsilon > 0$ is a small parameter representing the time unit and $\lfloor b \rfloor$ stands for the greatest integer which is smaller than or equal to b. Player i has a finite action space A^i, and if he/she chooses an action $a \in A^i$, then, provided that

the current state of the system is $x \in X^i$, its next state will be $y \in X^i$ with the probability $P^i_{xay} \geq 0$ ($\sum_{y\in X} P^i_{xay} = 1$).

A policy $u^i = \{u^i_0, u^i_1, \ldots, u^i_j, \ldots\ , u^i_{\lfloor \epsilon^{-1} \rfloor}\}$ of the player i is a sequence of probability measures on A^i chosen as functions of the present state of the player and of his or her states/actions history. That is, $u^i_0 = u^i_0(x^i_0)$ and $u^i_{j+1} = u^i_{j+1}(h^{j,i}_0, x^i_{j+1})$, where $h^{j,i}_0$ is the states/actions history of the i^{th} system/player,

$$h^{j,i}_0 := (x^i_l, a^i_l,\ l = 0, ..., j), \quad j = 1, ..., \lfloor \epsilon^{-1} \rfloor, \quad i = 1, 2. \tag{2}$$

Let $\mathcal{U}_i$ stand for the set of all such policies and let $\mathcal{F}^i$ be the discrete $\sigma-$algebra of all subsets of the set $\mathbf{H}_i$ of all possible states and actions histories that can be observed until the moment $[\epsilon^{-1}]$,

$$\mathbf{H}_i = \cup\{h_0^{[\epsilon^{-1}],i}\}.$$

Each initial distribution β of the initial states (x^1_0, x^2_0) and a pair of policies $(u^1, u^2) \in \mathcal{U}_1 \times \mathcal{U}_2$ uniquely defines a probability measure $\mathbb{P}^{(u^1,u^2)}_\beta$ over the space of samples $\mathbf{H} := \mathbf{H}_1 \times \mathbf{H}_2$ equipped with the discrete $\sigma-$algebra $\mathcal{F} := \mathcal{F}^1 \otimes \mathcal{F}^2$. Denote by $\mathbb{E}^{(u^1,u^2)}_\beta$ the corresponding expectation operator. When the distribution of the initial states is concentrated on a single states pair $\zeta = (\zeta^1, \zeta^2)$, we shall denote the corresponding probability measure and the mathematical expectation operator as $\mathbb{P}^{(u^1,u^2)}_\zeta$ and $\mathbb{E}^{(u^1,u^2)}_\zeta$.

Let $g^i : X^i \times A^i \to \mathbb{R}^{N_i}$ be a given vector function and let X^i_j and A^i_j, $j = 0, 1, \ldots, \lfloor \epsilon^{-1} \rfloor$, be the state-action processes of the system/player i. Then $Y^i(t)$ in (1) are defined by the equations

$$Y^i(t) = g^i(X^i_{\lfloor t/\epsilon \rfloor}, A^i_{\lfloor t/\epsilon \rfloor}), \quad i = 1, 2. \tag{3}$$

Note that the dynamics of the state vector $Z(t)$ is fully determined by the states/actions realizations $\{X^i_j, A^i_j\}$, $i = 1, 2$. For convenience, $Z(t)$ will be referred to as a "macro state" vector of the system and $\{X^i_j\}$ will be referred to as "micro states" of the players $i = 1, 2$.

Along with the class of policies $\mathcal{U} := \mathcal{U}_1 \times \mathcal{U}_2$ described above, we will be dealing with two other classes of policies $\mathcal{U}_M := \mathcal{U}_{M_1} \times \mathcal{U}_{M_2}$ and $\mathcal{U}_S := \mathcal{U}_{S_1} \times \mathcal{U}_{S_2}$, where $\mathcal{U}_{M_i}$ and $\mathcal{U}_{S_i}$ ($i = 1, 2$) are defined as follows:

- $\mathcal{U}_{M_i}$ is the set of policies of the player i such that, at every moment $t_j = j\epsilon$ of decision making, the probability measure u^i_j on A_i is chosen as a function of two arguments, one being the current moment of time and the other being the current micro state of the player i.

- $\mathcal{U}_{S_i}$ is the set of stationary policies of the player i. That is, it is the set of policies such that at any moment $t_j = j\epsilon$, the probability measure u^i_j on A_i is chosen as a function of the current micro state of the player i.

Note that, as follows from the definitions above, $\mathcal{U}_S \subset \mathcal{U}_M \subset \mathcal{U}$.

Assumption 1. *Under any stationary policy, the state space of the stochastic processes $\{X^i_j, A^i_j\}$ ($i = 1, 2$) forms an aperiodic Markov chain such that all states communicate (regular Markov chain).*

As mentioned above, we make the following assumption about the functions f^1 and f^2.

Assumption 2. *The functions $f_i(\cdot,\cdot)$, $i=1,2$ are continuous and satisfy Lipschitz conditions in the first argument. That is, there exist positive constants C_1, C_2 such that*

$$\|f^i(z,y^i)-f^i(\tilde{z},y^i)\|\le C_i\|z-\tilde{z}\|,\quad i=1,2 \tag{4}$$

for arbitrary z and $\tilde{z}$ from a sufficiently large subset of $\mathbb{R}^N$.

Remark 1. *According to their definitions, the processes $Y^1(t)$ and $Y^2(t)$ take values in some finite subsets D_1 and D_2 of $\mathbb{R}^{N_1}$ and $\mathbb{R}^{N_2}$, that is, $Y^i(t)\in D_i$, $i=1,2$. It implies that $Z(t)\in D$ $\forall t\in[0,1]$, where D is a compact subset of $\mathbb{R}^N$ (remind that the initial condition in (1) is fixed). Note that, since $f_i(\cdot,\cdot)$, $i=1,2$ are continuous, there exist positive constants M_i, $i=1,2$, such that*

$$\|f^i(z,y^i)\|\le M_i,\quad \forall(z,y^i)\in D\times D_i,\quad i=1,2. \tag{5}$$

Assume that the player i wishes to minimize his/her cost that depends only on the final value of the macro state. More specifically, assume that the player 1 aims at minimizing the payoff function $\mathbb{E}_\zeta^{(u^1,u^2)}[G^1(Z(1))]$ over $u^1\in\mathcal{U}_1$ and the player 2 aims at minimizing the payoff function $\mathbb{E}_\zeta^{(u^1,u^2)}[G^2(Z(1))]$ over $u^2\in\mathcal{U}_2$, where $G^i(\cdot):\mathbb{R}^N\to\mathbb{R}$ $(i=1,2)$ (it is assumed that the initial micro states $\zeta=(\zeta_1,\zeta_2)$ are fixed and are known to the players). This game will be referred to as *hybrid game*.

DEFINITION 1. *A pair of policies $u_\epsilon^*=(u_\epsilon^{1*},u_\epsilon^{2*})\in\mathcal{U}$ is a* ***Nash Equilibrium (NE) of the hybrid game*** *if*

$$\begin{array}{ll}\mathbb{E}_\zeta^{(u_\epsilon^{1*},u_\epsilon^{2*})}[G^1(Z(1))]\le\mathbb{E}_\zeta^{(u_\epsilon^1,u_\epsilon^{2*})}\{G^1(Z(1))\} & \forall u_\epsilon^1\in\mathcal{U}_1,\\ \mathbb{E}_\zeta^{(u_\epsilon^{1*},u_\epsilon^{2*})}\{G^2(Z(1))\}\le\mathbb{E}_\zeta^{(u_\epsilon^{1*},u_\epsilon^2)}\{G^2(Z(1))\} & \forall u_\epsilon^2\in\mathcal{U}_2.\end{array}$$

Note that the hybrid game can be shown to be equivalent to a game with finitely many states. The number of such states will, however, be very large if ϵ is small (this number tending to infinity with ϵ approaching zero). This makes finding an equilibrium extremely difficult, and we do not follow this path in the paper. Instead, we introduce the following asymptotic version of the equilibrium.

DEFINITION 2. *A pair of policies $u_\epsilon^*=(u_\epsilon^{1*},u_\epsilon^{2*})\in\mathcal{U}$ is an* ***Asymptotic Nash Equilibrium (ANE) of the hybrid game*** *if*

$$\begin{cases}\lim_{\epsilon\to0}\mathbb{E}_\zeta^{(u_\epsilon^{1*},u_\epsilon^{2*})}[G^1(Z(1))]\le\underline{\lim}_{\epsilon\to0}\mathbb{E}_\zeta^{(u_\epsilon^1,u_\epsilon^{2*})}\{G^1(Z(1))\} & \forall u_\epsilon^1\in\mathcal{U}_1,\\ \lim_{\epsilon\to0}\mathbb{E}_\zeta^{(u_\epsilon^{1*},u_\epsilon^{2*})}\{G^2(Z(1))\}\le\underline{\lim}_{\epsilon\to0}\mathbb{E}_\zeta^{(u_\epsilon^{1*},u_\epsilon^2)}\{G^2(Z(1))\} & \forall u_\epsilon^2\in\mathcal{U}_2,\end{cases} \tag{6}$$

where limits in the left-hand-sides are assumed to exist.

The fact that the changes of the micro states/actions occur frequently (every moment $t_j=j\epsilon$) means that the processes $Y_i(t)$, $i=1,2$, change their values on a much faster scale than does the macro state $Z(t)$. This allows one to approximate the solutions of the hybrid system (1) with the solutions of the deterministic averaged control system introduced in the next section. The main result of the paper is the construction of ANE of the stochastic hybrid game on the basis of the Nash equilibrium of the deterministic game considered on the trajectories of the averaged system (see Sections 3-5 below).

To simplify the presentation, we assume that the functions $G_i(\cdot)$ $i=1,2$, defining the payoff functions satisfy Lipschitz conditions on D (although the continuity of these functions would suffice our purposes). Thus, we make the following assumption.

Assumption 3. *There exist positive constants* $C_G^i,\ i=1,2,$ *such that*

$$|G^i(z)-G^i(\tilde{z})|\le C_G^i\|z-\tilde{z}\| \quad \forall\, z,\tilde{z}\in D, \qquad i=1,2. \tag{7}$$

3 Averaged Dynamic Game

Let $\omega_i(s^i;x,a)$ be the vector of steady state probabilities of the micro state-action pair (x,a) of the player i when he/she uses a stationary policy $s^i\in\mathcal{U}_{S_i}$. That is,

$$\omega^i(s^i;x,a):=\lim_{j\to\infty}\mathbb{P}_\zeta^{s^i}(X_j^i=x,A_j^i=a), \quad \omega^i(s^i)=\{\omega^i(s^i;x,a)\} \tag{8}$$

(due to the ergodicity Assumption 1, the limit value $\omega^i(s^i;x,a)$ is independent of the initial conditions). Define the sets W_i by the equations

$$W_i=\bigcup_{s^i\in\mathcal{U}_{S_i}}\{\omega^i(s^i)\}, \quad i=1,2. \tag{9}$$

Note that the sets W_i are polyhedrons (see, e.g., [9], pp. 93-95).

Consider a deterministic system, in which the dynamics of the state vector $z(t)$ is described by the equation

$$\dot{z}(t)=\hat{f}^1(z(t),\omega^1(t))+\hat{f}^2(z(t),\omega^2(t)), \qquad z(0)=z_0 \quad t\in[0,1], \tag{10}$$

where

$$\hat{f}^i(z,\omega^i):=\sum_{x,a}f^i(z,g^i(x,a))\omega^i(x,a) \quad \forall\,\omega^i\in W_i, \quad i=1,2. \tag{11}$$

The functions $\omega^i(\cdot),\ i=1,2,$ are controls chosen by the players. These are assumed to be measurable functions of t that satisfy the inclusions $\omega^i(t)\in W_i\ \forall t\in[0,1],\ i=1,2.$

As one can readily see, from Assumption 2 it follows that

$$\|\hat{f}^i(z,\omega^i)-\hat{f}^i(\tilde{z},\omega^i)\|\le C_i\|z-\tilde{z}\|, \quad \forall z,\tilde{z}\in D,\ \forall\omega^i\in W_i, \tag{12}$$

and from Remark 1 it follows that

$$\|\hat{f}^i(z,\omega^i)\|\le M_i, \quad \forall(z,\omega^i)\in D\times W_i. \tag{13}$$

Note that from (12) it follows that the solution of (10) exists and is unique with any choice of controls $\omega_i(\cdot),\ i=1,2.$

Assume that the player i wishes to minimize the terminal cost function $G_i(z(1))$, where $G_i(\cdot)$ is the same as in the previous section $(i=1,2)$. Given a pair of controls $(\omega^1(t),\omega^2(t))$, let $J_i(\omega^1,\omega^2)$ stand for the cost function of the payer i obtained with the players adopting these controls. That is,

$$J_i(\omega^1,\omega^2):=G^i(z(1)), \tag{14}$$

where $z(t)$ is the solution of (10) obtained applying $(\omega^1(t),\omega^2(t))$.

DEFINITION 3. *A pair of controls* $(\omega^{1*}(\cdot),\omega^{2*}(\cdot))$ *is a* ***Nash equilibrium of the averaged game*** *if*

$$\begin{cases} J_1(\omega^{1*},\omega^{2*})\le J_1(\omega^1,\omega^{2*}),\\ J_2(\omega^{1*},\omega^{2*})\le J_2(\omega^{1*},\omega^2), \end{cases} \tag{15}$$

for any $\omega^1(\cdot)$ *(resp.* $\omega^2(\cdot)$*).*

Note that the Nash equilibrium of the averaged game is defined in the open loop setting.

Remark 2. *Due to linearity of the functions $\hat{f}^i(z,\omega^i)$ in ω^i, $i=1,2$, and due to convexity and compactness of the sets W_i, $i=1,2$, a Nash equilibrium of the averaged game exists (see Theorem 1.1 in [14]).*

In the next section, we will show how ANE policies of the hybrid game can be constructed on the basis of a Nash equilibrium pair of the averaged game.

4 Construction of ANE Policies - Main Results

Let $\omega^i(\cdot)$ be a control of the player i in the averaged game. Partition the time interval $[0,1]$ by the points

$$\tau_l := l\Delta(\epsilon), \quad l=0,1,2,\dots,\ell(\epsilon), \quad \ell(\epsilon) := \lfloor \Delta(\epsilon)^{-1} \rfloor, \quad \tau_{\ell(\epsilon)+1} = 1, \tag{16}$$

where $\Delta(\epsilon) > 0$ is a function of ϵ such that

$$\lim_{\epsilon\to\infty} \Delta(\epsilon) = 0, \qquad \lim_{\epsilon\to 0} \frac{\Delta(\epsilon)}{\epsilon} = \infty. \tag{17}$$

On each interval $[\tau_l, \tau_{l+1}]$ $(l=0,1,\dots,\lfloor\Delta(\epsilon)^{-1}\rfloor - 1)$, define the time averages ω_l^i,

$$\omega_l^i := \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} \omega^i(t)dt \ \in W_i, \quad i=1,2, \tag{18}$$

(the validity of the last inclusions follows from the convexity of W_i, $i=1,2$). Note that, from the fact that $\omega_l^i \in W_i$, it follows that there exists a stationary policy $s^i \in \mathcal{U}_{S_i}$ of the player i such that

$$\omega_l^i = \omega_l^i(s_l^i)$$

(see (8) and (9)). Let us define a policy of the player i in the hybrid game that consists of:

1. Applying, at each $j = \lfloor\tau_l/\epsilon\rfloor, \lfloor\tau_l/\epsilon\rfloor + 1, \dots, \lfloor\tau_{l+1}/\epsilon\rfloor - 1$, the policy s_l^i, for any $l = 0,1,2,\dots,\ell(\epsilon)-1$;

2. Applying an arbitrary stationary policy for $\lfloor\tau_{\ell(\epsilon)}/\epsilon\rfloor, \lfloor\tau_{\ell(\epsilon)}/\epsilon\rfloor + 1, \dots, \lfloor\epsilon^{-1}\rfloor$.

Let us denote this policy as $u_\epsilon^i(\omega^i)$. Note that, by construction, $u_\epsilon^i(\omega^i) \in \mathcal{U}_{M_i}$, $i=1,2$. The main results of the paper are Propositions 1 and 2 stated below.

Proposition 1. *Let $\omega(t) = (\omega^1(t), \omega^2(t))$ be a pair of controls and let $z(t)$ be the corresponding solution of (10). Let also $u_\epsilon(\omega) = (u_\epsilon^1(\omega^1), u_\epsilon^2(\omega^2))$ be the pair of policies defined above and let $Z(t)$ be the random trajectory of system (1) obtained with the players using these policies. Then*

$$\max_{t\in[0,1]} \mathbb{E}_\zeta^{u_\epsilon(\omega)} \|Z(t) - z(t)\| \le \gamma(\epsilon), \tag{19}$$

where $\lim_{\epsilon\to 0}\gamma(\epsilon) = 0$. Also,

$$\lim_{\epsilon\to\ 0} \mathbb{E}_\zeta^{u_\epsilon(\omega)} G^i(Z(1)) = J^i(\omega^1,\omega^2), \quad i=1,2. \tag{20}$$

Proposition 2. *Let $\omega^*(t) := (\omega^{1*}(t), \omega^{2*}(t))$ be a Nash equilibrium of the averaged game. Let also $u_\epsilon^*(\omega^*) = (u_\epsilon^{1*}(\omega^{1*}), u_\epsilon^{2*}(\omega^{2*}))$ be defined as above (with the use of $\omega^*(t)$ instead of $\omega(t)$). Then $u_\epsilon^*(\omega^*)$ is an asymptotic Nash equilibrium of the hybrid game.*

The proofs of Propositions 1 and 2 are given in the next section.

5 Proofs of Propositions 1 and 2

Let us first recall some results from the MDP theory that are needed for the proofs of the main results. Let h^i stand for the full micro states-actions history of the player i (that is, $h^i = h_0^{\lfloor \epsilon^{-1} \rfloor, i}$; see (2)) and let $h_m^{K,i}$ be a part of this history corresponding to the interval $[t_m, t_{m+K}]$,

$$h_m^{K,i} := \{x_m^i, a_m^i, \dots, x_{m+K}^i, a_{m+K}^i\}.$$

Denote by $\phi_m^{K,i}(h^i; x, a)$ the frequency of appearance of the micro state-action pair $(x, a) \in X^i \times A^i$ on the interval $[t_m, t_{m+K}]$,

$$\phi_m^{K,i}(h^i; x, a) := \frac{1}{K} \sum_{j=m}^{m+K-1} \mathbb{1}_{\{x_j^i = x, a_j^i = a\}}, \quad \phi_m^{K,i}(h^i) = \{\phi_m^{K,i}(h^i, x, a)\}, \tag{21}$$

where $\mathbb{1}_{\{x_n^i = x, a_n^i = a\}}$ is the indicator function. If H^i is a random realization of h^i, we denote:

$$\phi_m^{K,i}(x, a) = \phi_m^{K,i}(H^i, x, a), \quad \phi_m^{K,i} = \{\phi_m^{K,i}(x, a)\}. \tag{22}$$

Lemma 1. *The following relationships are valid:*

$$\lim_{K \to \infty} \sup_{\zeta} \sup_{s^i \in \mathcal{U}_{S_i}} \mathbb{E}_\zeta^{s^i} \|\phi_0^{K,i} - \omega^i(s^i)\| = 0, \tag{23}$$

and also

$$\lim_{K \to \infty} \sup_{\zeta} \sup_{u^i \in \mathcal{U}_i} \mathbb{E}_\zeta^{u^i} d_0^{K,i} = 0, \tag{24}$$

where

$$d_0^{K,i} := dist(\phi_0^{K,i}, W_i) = \inf_{\omega^i \in W_i} \|\phi_0^{K,i} - \omega^i\|.$$

Proof. The proof is similar to that of Theorem 4.1 in [4]. □

Proof of Proposition 1. Let $\omega(t) = (\omega^1(t), \omega^2(t))$ be as in the statement of Proposition 1 and let $z(t)$ be the corresponding solution of (10). The proof proceeds as follows. Firstly, we approximate $z(t)$ by the solution of the difference equation (25) (see below). We then evaluate the mathematical expectation of the norm of the difference between the latter and the solution of the stochastic difference equation (28) (see the estimate (38)). Finally, we show that the solution of the stochastic difference equation (28) approximates the solution of (1) (see the estimate (43)). We will now discuss these steps in more detail.

Define the sequence of vectors ξ_l, $l = 0, \dots, \ell(\epsilon)$, as the solution of the following difference equation

$$\xi_{l+1} = \xi_l + \Delta(\epsilon) \left[\hat{f}^1(z(\tau_l), \omega_l^1) + \hat{f}^2(z(\tau_l), \omega_l^2)\right], \quad \xi_0 := z_0. \tag{25}$$

By definition

$$z(\tau_{l+1}) = z(\tau_l) + \int_{\tau_l}^{\tau_{l+1}} \hat{f}^1(z(t), \omega^1(t))dt + \int_{\tau_l}^{\tau_{l+1}} \hat{f}^2(z(t), \omega^2(t))dt. \tag{26}$$

Hence,

$$\|z(\tau_{l+1}) - \xi_{l+1}\| \le \|z(\tau_l) - \xi_l\| + \Delta(\epsilon) \sum_{i=1}^{2} \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} \hat{f}^i(z(t), \omega^i(t))dt - \hat{f}^i(z(\tau_l), \omega_l^i)) \right\|.$$

As one can see,

$$\left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} \hat{f}^i(z(t), \omega^i(t))dt - \hat{f}^i(z(\tau_l), \omega_l^i)) \right\|$$

$$\le \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} \hat{f}^i(z(t), \omega^i(t))dt - \hat{f}^i(z(\tau_l), \omega^i(t)) \right\| + \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} \hat{f}^i(z(\tau_l), \omega^i(t))dt - \hat{f}^i(z(\tau_l), \omega_l^i)) \right\|$$

$$\le \tilde{M}_i \Delta(\epsilon),$$

where M_i is a constant. The validity of the last inequality follows from the Lipschitz continuity of the function $\hat{f}^i(z, \omega^i)$ in z and from its linearity in ω^i, the latter implies (see (18))

$$\frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} \hat{f}^i(z(\tau_l), \omega^i(t))dt = \hat{f}^i(z(\tau_l), \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} \omega(t)dt) = \hat{f}^i(z(\tau_l), \omega_l^i)).$$

Consequently,

$$\|z(\tau_{l+1}) - \xi_{l+1}\| \le \|z(\tau_l) - \xi_l\| + \Delta(\epsilon)^2 \tilde{M}, \quad \tilde{M} = \tilde{M}^1 + \tilde{M}^2,$$

and, for any $l = 1, \dots, \ell(\epsilon)$,

$$\|z(\tau_l) - \xi_l\| \le \ell(\epsilon)\Delta(\epsilon)^2 \tilde{M} \le \Delta(\epsilon)\tilde{M}. \tag{27}$$

Let $Z_l,\ l = 0, 1 \dots, \ell(\epsilon)$, be the sequence of random vectors defined by the equation

$$Z_{l+1} = Z_l + \int_{\tau_l}^{\tau_{l+1}} f^1(Z_l, Y^1(t))dt + \int_{\tau_l}^{\tau_{l+1}} f^2(Z_l, Y^2(t))dt, \quad l = 0, 1, \dots, \ell(\epsilon) - 1. \tag{28}$$

By subtracting the latter from (25) and taking the expectation over the probability measure corresponding to the policies $u_\epsilon(\omega) = (u_\epsilon^1(\omega^1), u_\epsilon^2(\omega^2))$, we obtain

$$\mathbb{E}_\zeta^{u_\epsilon(\omega)} \|\xi_{l+1} - Z_{l+1}\| \le \mathbb{E}_\zeta^{u_\epsilon(\omega)} \|\xi_l - Z_l\| + \Delta(\epsilon) \sum_{i=1}^{2} \mathbb{E}_\zeta^{u_\epsilon(\omega)} \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(Z_l, Y^i(t))dt - \hat{f}^i(z(\tau_l), \omega_l^i) \right\|$$

$$\le \mathbb{E}_\zeta^{u_\epsilon(\omega)} \|\xi_l - Z_l\| + \Delta(\epsilon) \sum_{i=1}^{2} \mathbb{E}_\zeta^{u_\epsilon(\omega)} \left\{ \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(Z_l, Y^i(t))dt - \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l), Y^i(t))dt \right\| \right.$$

$$\left. + \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l), Y^i(t))dt - \hat{f}^i(z(\tau_l), \omega_l^i)) \right\| \right\}. \tag{29}$$

Using the Lipschitz continuity of $f^i(z, \omega^i)$ in z and the estimate (27), one can obtain that

$$\left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(Z_l, Y^i(t))dt - \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l), Y^i(t))dt \right\| \le \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} C_i \|Z_l - z(\tau_l)\|dt$$

$$\le C_i(\|Z_l - \xi_l\| + \|\xi_l - z(\tau_l))\| \le C_i(\|Z_l - \xi_l\| + \tilde{M}\Delta(\epsilon)). \tag{30}$$

By substituting the latter inequality in (29), one obtains

$$\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\|\xi_{l+1}-Z_{l+1}\| \le \mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\|\xi_l - Z_l\| + (C_1+C_2)\tilde{M}\Delta(\epsilon)^2 + \Delta(\epsilon)\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\left\{\sum_{i=1}^{2} C_i\|Z_l-\xi_l\| + \left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l),Y^i(t))dt - \hat{f}^i(z(\tau_l),\omega_l^i))\right\|\right\}. \tag{31}$$

Let $K(\epsilon)=\min_{l=0,1,\ldots,l(\epsilon)-1}(\lfloor\tau_{l+1}/\epsilon\rfloor-\lfloor\tau_l/\epsilon\rfloor)$. Note that (see [16])

$$2 \ge \lfloor\tau_{l+1}/\epsilon\rfloor - \lfloor\tau_l/\epsilon\rfloor - K(\epsilon), \qquad \left|K(\epsilon)-\frac{\Delta(\epsilon)}{\epsilon}\right| \le 1;$$

the latter, as can be readily verified, implies

$$\left|K(\epsilon)^{-1}-\frac{\epsilon}{\Delta(\epsilon)}\right| \le \frac{\epsilon^2}{\Delta(\epsilon)^2}\left(\frac{1}{1-\epsilon/\Delta(\epsilon)}\right). \tag{32}$$

Using the estimates above, one can show that there exist positive constants L_1^i,L_2^i such that, for $i=1,2$, and $l=1,\ldots,\ell(\epsilon)$,

$$\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l),Y^i(t))dt - \frac{\epsilon}{\Delta(\epsilon)}\sum_{n=\lfloor\tau_l/\epsilon\rfloor}^{\lfloor\tau_l/\epsilon\rfloor+K(\epsilon)-1} f^i(z(\tau_l),g^i(X_n^i,A_n^i))\right\| \le L_1^i\frac{\epsilon}{\Delta(\epsilon)}, \tag{33}$$

$$\left\|\frac{\epsilon}{\Delta(\epsilon)}\sum_{n=\lfloor\tau_l/\epsilon\rfloor}^{\lfloor\tau_l/\epsilon\rfloor+K(\epsilon)-1} f^i(z(\tau_l),g^i(X_n^i,A_n^i)) - \frac{1}{K(\epsilon)}\sum_{n=\lfloor\tau_l/\epsilon\rfloor}^{\lfloor\tau_l/\epsilon\rfloor+K(\epsilon)-1} f^i(z(\tau_l),g^i(X_n^i,A_n^i)\right\| \le L_2^i\frac{\epsilon}{\Delta(\epsilon)}. \tag{34}$$

In fact, (33) follows from the fact that

$$\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l),Y^i(t))dt - \frac{1}{\Delta(\epsilon)}\int_{\lfloor\tau_l/\epsilon\rfloor\epsilon}^{(\lfloor\tau_l/\epsilon\rfloor+K(\epsilon))\epsilon} f^i(z(\tau_l),Y^i(t))dt\right\|$$

$$\le M_i\left(\frac{\tau_l-\lfloor\tau_l/\epsilon\rfloor\epsilon}{\Delta(\epsilon)} + \frac{(\tau_{l+1}-(\lfloor\tau_l/\epsilon\rfloor+K(\epsilon))\epsilon}{\Delta(\epsilon)}\right) \le 4M_i\frac{\epsilon}{\Delta(\epsilon)},$$

and also from the fact that

$$\frac{1}{\Delta(\epsilon)}\int_{\lfloor\tau_l/\epsilon\rfloor\epsilon}^{(\lfloor\tau_l/\epsilon\rfloor+K(\epsilon))\epsilon} f^i(z(\tau_l),Y^i(t))dt = \frac{\epsilon}{\Delta(\epsilon)}\sum_{n=\lfloor\tau_l/\epsilon\rfloor}^{\lfloor\tau_l/\epsilon\rfloor+K(\epsilon)-1} f^i(z(\tau_l),g^i(X_n^i,A_n^i)).$$

That is, (33) is valid with $L_1^i=4M_i$, where $M_i,\ i=1,2$, are as in (5). To verify (34), it is sufficient to note that the right-hand-side in (34) is majorized by the following expression

$$M_iK(\epsilon)\left|\frac{\epsilon}{\Delta(\epsilon)}-\frac{1}{K(\epsilon)}\right| \le M_i\epsilon.$$

Thus, (34) is valid with $L_2^i=M_i,\ i=1,2$.

Let (as above) $\phi_m^{K,i} = \{\phi_m^{K,i}(x,a)\}$ stand for the state-action frequencies of the controller i that corresponds to a random realization of the history H^i (see (22)). Then

$$\frac{1}{K(\epsilon)} \sum_{n=\lfloor \tau_l/\epsilon \rfloor}^{\lfloor \tau_l/\epsilon \rfloor + K(\epsilon)-1} f^i(z(\tau_l), g^i(X_n^i, A_n^i)) = \sum_{(x,a)\in X^i \times A^i} \phi_{\lfloor \tau_l/\epsilon \rfloor}^{K(\epsilon),i}(x,a) f^i(z(\tau_l), g^i(x,a)).$$

Note that from (23) (see Lemma 1) it follows that there exists a function $\mu_i : \mathbb{N} \to \mathbb{R}$, with $\lim_{K\to\infty} \mu_i(K) = 0$ (implying $\lim_{\epsilon\to 0} \mu_i(K(\epsilon)) = 0$) such that

$$\mathbb{E}^{s^i}_{X^i_{\lfloor \tau_l/\epsilon \rfloor}} \left\{ \max_{(x,a)} |\phi_{\lfloor \tau_l/\epsilon \rfloor}^{K(\epsilon),i}(x,a) - \omega_l^i(x,a)| \right\} \le \mu_i(K(\epsilon)). \tag{35}$$

Since

$$\begin{aligned}
&\sum_{i=1}^2 \mathbb{E}_\zeta^{u_\epsilon(\omega)} \left\{ \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l), Y^i(t))dt - \hat{f}^i(z(\tau_l), \omega_l^i)) \right\| \right\} \\
&\le \sum_{i=1}^2 \mathbb{E}_\zeta^{u_\epsilon(\omega)} \left\{ \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l), Y^i(t))dt - \frac{\epsilon}{\Delta(\epsilon)} \sum_{n=\lfloor \tau_l/\epsilon \rfloor}^{\lfloor \tau_l/\epsilon \rfloor + K(\epsilon)-1} f^i(z(\tau_l), g^i(X_n^i, A_n^i)) \right\| \right. \\
&\quad + \left\| \frac{\epsilon}{\Delta(\epsilon)} \sum_{n=\lfloor \tau_l/\epsilon \rfloor}^{\lfloor \tau_l/\epsilon \rfloor + K(\epsilon)-1} f^i(z(\tau_l), g^i(X_n^i, A_n^i)) - \frac{1}{K(\epsilon)} \sum_{n=\lfloor \tau_l/\epsilon \rfloor}^{\lfloor \tau_l/\epsilon \rfloor + K(\epsilon)-1} f^i(z(\tau_l), g^i(X_n^i, A_n^i)) \right\| \\
&\quad \left. + \left\| \sum_{x,a} \phi_{\lfloor \tau_l/\epsilon \rfloor,i}^{K(\epsilon)}(H^i; x, a) f^i(z(\tau_l), g^i(x,a)) - \hat{f}^i(z(\tau_l), \omega_l^i) \right\| \right\}
\end{aligned}$$

and since, by definition, $\hat{f}^i(z(\tau_l), \omega_l^i) := \sum_{x,a} \omega_l^i(x,a) f^i(z(\tau_l), g^i(x,a))$, one can obtain, using (35), (33), (34),

$$\begin{aligned}
&\sum_{i=1}^2 \mathbb{E}_\zeta^{u_\epsilon(\omega)} \left\{ \left\| \frac{1}{\Delta(\epsilon)} \int_{\tau_l}^{\tau_{l+1}} f^i(z(\tau_l), Y^i(t))dt - \hat{f}^i(z(\tau_l), \omega_l^i) \right\| \right\} \le (\tilde{L}_1 + \tilde{L}_2) \frac{\epsilon}{\Delta(\epsilon)} \\
&\quad + \sum_{i=1}^2 \mathbb{E}_\zeta^{u_\epsilon(\omega)} \left\{ \sum_{x,a} \left| \phi_{\lfloor \tau_l/\epsilon \rfloor,i}^{K(\epsilon)}(H^i; x, a) - \omega_l^i(x,a) \right| \|f^i(z(\tau_l), g^i(x,a))\| \right\} \\
&\le (\tilde{L}_1 + \tilde{L}_2) \frac{\epsilon}{\Delta(\epsilon)} + \sum_{i=1}^2 M_i \mathbb{E}_\zeta^{u_\epsilon^i(\omega)} \left\{ \sum_{x,a} \left| \phi_{\lfloor \tau_l/\epsilon \rfloor,i}^{K(\epsilon)}(H^i; x, a) - \omega_l^i(x,a) \right| \right\} \qquad (36) \\
&\le (\tilde{L}_1 + \tilde{L}_2) \frac{\epsilon}{\Delta(\epsilon)} + \sum_{i=1}^2 M_i \mathbb{E}_\zeta^{u_\epsilon^i(\omega)} \left\{ \mathbb{E}^{s_l^i}_{X^i_{\lfloor \tau_l/\epsilon \rfloor}} \sum_{x,a} \left| \phi_{\lfloor \tau_l/\epsilon \rfloor,i}^{K(\epsilon)}(H^i; x, a) - \omega_l^i(x,a) \right| \right\} \\
&\le (\tilde{L}_1 + \tilde{L}_2) \frac{\epsilon}{\Delta(\epsilon)} + \sum_{i=1}^2 M_i \mu_i(K(\epsilon)),
\end{aligned}$$

where $\tilde{L}_i = L_1^i + L_2^i$ ($L_j^i,\ j = 1,2$, being the constants from (33) and (34)) and $M_i,\ i = 1,2$, are the constants from (5). The substitution of the latter into (31) leads to

$$\mathbb{E}_\zeta^{u_\epsilon(\omega)} [\|\xi_{l+1} - Z_{l+1}\|] \le \mathbb{E}_\zeta^{u_\epsilon(\omega)} [\|\xi_l - Z_l\|] + \tilde{M}\Delta(\epsilon) \mathbb{E}_\zeta^{u_\epsilon(\omega)} [\|\xi_l - Z_l\|] + \Delta(\epsilon)\kappa(\epsilon), \tag{37}$$

where $\lim_{\epsilon\to 0}\kappa(\epsilon)=0$ and $\tilde{M}$ is an appropriately chosen constant. This implies (see Proposition 5.1 in [10]) that

$$\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\left[\|\xi_l - Z_l\|\right] \le \nu(\epsilon), \quad l=1,\dots,\ell(\epsilon), \tag{38}$$

with $\lim_{\epsilon\to 0}\nu(\epsilon)=0$.

By definition,

$$Z(\tau_{l+1}) = Z(\tau_l) + \int_{\tau_l}^{\tau_{l+1}} f^1(Z(t),Y^1(t))dt + \int_{\tau_l}^{\tau_{l+1}} f^2(Z(t),Y^2(t))dt \quad l=1,\dots,\ell(\epsilon)-1. \tag{39}$$

By subtracting (28) from (39), one obtains

$$\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\|Z(\tau_{l+1}) - Z_{l+1}\| \le \mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\left[\|Z(\tau_l)-Z_l\| + (C_1+C_2)\int_{\tau_l}^{\tau_{l+1}}\|Z(t)-Z_l\|dt\right]. \tag{40}$$

Note that, due to (5),

$$\|Z(t)-Z(\tau_l)\| \le (M_1+M_2)\Delta(\epsilon) \ \ \forall t\in[\tau_l,\tau_{l+1}]. \tag{41}$$

Hence,

$$\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\|Z(\tau_{l+1}) - Z_{l+1}\| \le \mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\left[\|Z(\tau_l)-Z_l\| + C\Delta(\epsilon)\|Z(\tau_l)-Z_l\|\right] + C(M_1+M_2)\Delta(\epsilon)^2. \tag{42}$$

Due to Proposition 5.1 from [10], the latter implies that there exists an appropriately chosen positive constant $\tilde{M}$ such that

$$\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\|Z(\tau_l)-Z_l\| \le \tilde{M}\Delta(\epsilon), \ \ l=0,1,\dots,\ell(\epsilon). \tag{43}$$

By combining (27), (38), (43), we may conclude that

$$\begin{aligned}\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\|z(\tau_l)-Z(\tau_l)\| &\le \mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\{\|z(\tau_l)-\xi_l\| + \|\xi_l - Z_l\| + \|Z_l - Z(\tau_l)\|\} \\ &\le \nu(\epsilon) + \Delta(\epsilon)\tilde{M}, \ \ l=0,1,\dots,\ell(\epsilon).\end{aligned} \tag{44}$$

This and the fact that $\|z(t)-z(\tau_l)\| \le M\Delta(\epsilon)\ \forall t\in[\tau_l,\tau_{l+1}]$ imply the validity of (19) with some $\gamma(\epsilon)$ such that $\lim_{\epsilon\to 0}\gamma(\epsilon)=0$. The validity of (20) follows from (19) since (see (7))

$$|\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}G^i(Z(1)) - G^i(z(1))| \le C_G^i\mathbb{E}_{\zeta}^{u_\epsilon(\omega)}\|Z(1)-z(1)\| \le C_G^i\gamma(\epsilon).$$

Thus, the proof of the proposition is completed. □

Proof of Proposition 2. Let $h^1 = h_0^{\lfloor\epsilon^{-1}\rfloor,1} = (x_0^1, a_0^1, \dots, x_{\lfloor\epsilon^{-1}\rfloor}^1, a_{\lfloor\epsilon^{-1}\rfloor}^1)$ be a realization of a state-action trajectory of player 1 and let

$$y^1(t,h^1) := g^1(x_{\lfloor t/\epsilon\rfloor}^1, a_{\lfloor t/\epsilon\rfloor}^1). \tag{45}$$

Define the projection of the vector of the state action frequencies $\phi_{\lfloor\tau_l/\epsilon\rfloor}^{K(\epsilon),1}(h^1)$ (see (21)) onto W_1 by the equation:

$$\omega_l^1(h^1) := \arg\min_{\omega^1\in W_1}\left\|\phi_{\lfloor\tau_l/\epsilon\rfloor}^{K(\epsilon),1}(h^1) - \omega^1\right\|, \tag{46}$$

where $K(\epsilon)$ is as in Proposition 1. Define the function

$$\omega^1(t,h^1) := \{\omega_l^1(h^1) \text{ for } t \in [\tau_l, \tau_{l+1}],\ l = 1,\dots,\ell(\epsilon)\},$$

and denote by $z(t,h^1)$ the solution of the differential equation

$$\dot{z}(t,h^1) = \hat{f}_1(z(t,h^1),\omega^1(t,h^1)) + \hat{f}_2(z(t,h^1),\omega^{2*}(t)). \tag{47}$$

Define the piecewise constant function

$$\tilde{\omega}^{2*}(t) = \{\omega_l^{2*} \text{ for } t \in [\tau_l,\tau_{l+1}], l = 1,\dots,\ell(\epsilon)\},$$

where $\omega_l^{2*} := \frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} \omega^{2*}(t)dt$. Let $\tilde{z}(t,h^1)$ be the solution of the following differential equation

$$\dot{\tilde{z}}(t,h^1) = \hat{f}_1(\tilde{z}(t,h^1),\omega^1(t,h^1)) + \hat{f}_2(\tilde{z}(t,h^1),\tilde{\omega}^{2*}(t)). \tag{48}$$

By subtracting (47) from (48) (and having in mind linearity of $\hat{f}^2(z,\omega^2)$ in ω^2)), one obtains

$$\|\tilde{z}(\tau_{l+1},h^1) - z(\tau_{l+1},h^1)\| \le \|\tilde{z}(\tau_l,h^1) - z(\tau_l,h^1)\| + \tilde{M}\Delta\|\tilde{z}(t,h^1) - z(t,h^1)\| + \tilde{M}\Delta^2(\epsilon),$$

where $\tilde{M}$ is an appropriate positive constant. Due to Proposition 5.1 in [10], the latter implies that

$$\|\tilde{z}(\tau_l,h^1) - z(\tau_l,h^1)\| \le \tilde{M}\Delta(\epsilon) \qquad \forall l = 1\dots,\ell(\epsilon), \tag{49}$$

which, in turn, implies that

$$\max_{t\in[0,1]} \|\tilde{z}(t,h^1) - z(t,h^1)\| \le \tilde{M}\Delta(\epsilon), \tag{50}$$

and, by (7),

$$\|G^1(\tilde{z}(1,h^1)) - G^1(z(1,h^1))\| \le C_G^1\|\tilde{z}(1,h^1) - z(1,h^1)\| \le C_G^1\tilde{M}\Delta(\epsilon). \tag{51}$$

Due to the definition of the Nash equilibrium (see (15)),

$$G^1(z(1,h^1)) \ge G^{1*}. \tag{52}$$

Consequently,

$$G^1(\tilde{z}(1,h^1)) \ge G^{1*} - C_G^1\tilde{M}\Delta(\epsilon). \tag{53}$$

Let H^1 be a random realization of h^1. Using the fact that $\hat{f}^i(z,\omega^i)$ $(i=1,2)$ are Lipschitz continuous in z and linear in ω^i, one can obtain

$$\begin{aligned}\tilde{z}(\tau_{l+1},H^1) &= \tilde{z}(\tau_l,H^1) + \int_{\tau_l}^{\tau_{l+1}} \hat{f}^1(\tilde{z}(t,H^1),\omega_1(t,H^1))dt + \int_{\tau_l}^{\tau_{l+1}} \hat{f}^2(\tilde{z}(t,H^1),\tilde{\omega}_2^*(t))dt \\ &= \tilde{z}(\tau_l,H^1) + \Delta(\epsilon)\hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1)) + \Delta(\epsilon)\hat{f}^2(\tilde{z}(\tau_l,H^1),\omega_l^{2*})) + O(\Delta(\epsilon)^2).\end{aligned} \tag{54}$$

By subtracting (54) from (39), one obtains

$$\begin{aligned}&\|Z(\tau_{l+1}) - \tilde{z}(\tau_{l+1},H^1)\| \le \|Z(\tau_l) - \tilde{z}(\tau_l,H^1)\| \\ &+ \Delta(\epsilon)\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(Z(t),Y^1(t)dt - \hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1))\right\| \\ &+ \Delta(\epsilon)\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(Z(t),Y^2(t)dt - \hat{f}^2(\tilde{z}(\tau_l,H^1),\omega_l^{2*})\right\| + O(\Delta(\epsilon)^2).\end{aligned} \tag{55}$$

By (4),

$$\begin{aligned}
&\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(Z(t),Y^1(t))dt-\hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1))\right\| \\
&\le \left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(Z(t),Y^1(t))dt-\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(\tilde{z}(\tau_l,H^1),Y^1(t))\right\| \\
&+\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(\tilde{z}(\tau_l,H^1),Y^1(t))-\hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1))\right\| \\
&\le \left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(\tilde{z}(\tau_l,H^1),Y^1(t))dt-\hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1))\right\| \\
&+C_1\Delta(\epsilon)\left\|Z(\tau_l)-\tilde{z}(\tau_l,H^1)\right\|+O(\Delta(\epsilon)^2).
\end{aligned} \tag{56}$$

Analogously,

$$\begin{aligned}
&\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(Z(t),Y^2(t))dt-\hat{f}^2(\tilde{z}(\tau_l,H^1),\omega_l^{2*})\right\| \\
&\le \left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(Z(t),Y^2(t))dt-\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(\tilde{z}(\tau_l,H^1),Y^2(t))dt\right\| \\
&+\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(\tilde{z}(\tau_l,H^1),Y^2(t))dt-\hat{f}^2(\tilde{z}(\tau_l,H^1),\omega_l^{2*})\right\| \\
&\le \left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(\tilde{z}(\tau_l,H^1),Y^2(t))-\hat{f}^2(\tilde{z}(\tau_l,H^1),\omega_l^{2*})\right\| \\
&+C_2\Delta(\epsilon)\left\|Z(\tau_l)-\tilde{z}(\tau_l,H^1)\right\|+O(\Delta(\epsilon)^2).
\end{aligned} \tag{57}$$

For brevity, let us re-denote $u_\epsilon^{i*}(\omega^{i*})$ as u_ϵ^{i*} $(i=1,2)$. From (55), (56) and (57) it follows that, for any $u_\epsilon^1\in\mathcal{U}_1$,

$$\begin{aligned}
&\mathbb{E}_\zeta^{u_\epsilon^1,u_\epsilon^{2*}}\|Z(\tau_{l+1})-\tilde{z}(\tau_{l+1},H^1)\|\le\mathbb{E}_\zeta^{u_\epsilon^1,u_\epsilon^{2*}}\left\{\|Z(\tau_l)-\tilde{z}(\tau_l,H^1)\|(1+(C_1+C_2)\Delta(\epsilon))\right\} \\
&+\Delta(\epsilon)\mathbb{E}_\zeta^{u_\epsilon^1,u_\epsilon^{2*}}\left\{\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(\tilde{z}(\tau_l,H^1),Y^1(t))dt-\hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1))\right\|\right. \\
&\left.+\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(\tilde{z}(\tau_l,H^1),Y^2(t))dt-\hat{f}^2(\tilde{z}(\tau_l,H^1),\omega_l^{2*})\right\|\right\}+O(\Delta(\epsilon)^2).
\end{aligned} \tag{58}$$

Similarly to (36), one can obtain

$$\begin{aligned}
&\mathbb{E}_\zeta^{u_\epsilon^1,u_\epsilon^{2*}}\left\{\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^2(\tilde{z}(\tau_l,H^1),Y^2(t))dt-\hat{f}^2(\tilde{z}(\tau_l,H^1),\omega_l^{2*})\right\|\right\} \\
&\le\tilde{L}_2\frac{\epsilon}{\Delta(\epsilon)}+M_2\mathbb{E}_\zeta^{u_\epsilon^{2*}}\left\{\sum_{x,a}\left|\phi^{K(\epsilon)}_{\lfloor\tau_l/\epsilon\rfloor,2}(H^2;x,a)-\omega_l^{2*}(x,a))\right|\right\} \\
&\le\tilde{L}_2\frac{\epsilon}{\Delta(\epsilon)}+M_2\mathbb{E}_\zeta^{u_\epsilon^{2*}}\left\{\mathbb{E}^{s_l^2}_{X^2_{\lfloor\tau_l/\epsilon\rfloor}}\sum_{x,a}\left|\phi^{K(\epsilon)}_{\lfloor\tau_l/\epsilon\rfloor,2}(H^2;x,a)-\omega_l^{2*}(x,a))\right|\right\} \\
&\le\tilde{L}_2\frac{\epsilon}{\Delta(\epsilon)}+M_2\mu_2(K(\epsilon)),
\end{aligned} \tag{59}$$

where $\tilde{L}_2$ and M_2 are the same constants as in (36). In an analogous way,

$$\begin{aligned}&\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\left\{\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(\tilde{z}(\tau_l,H^1),Y^1(t))dt-\hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1))\right\|\right\}\\&\leq \tilde{L}_1\frac{\epsilon}{\Delta(\epsilon)}+M_1\mathbb{E}_{\zeta}^{u_\epsilon^1}\left\{\sum_{x,a}\left|\phi^{K(\epsilon)}_{\lfloor\tau_l/\epsilon\rfloor,1}(H^1;x,a)-\omega_l^1(x,a;H^1)\right|\right\},\end{aligned}\tag{60}$$

where M_1 is the constant from (5). From the definition of $\omega_l^1(H^1)=\{\omega_l^1(x,a;H^1)\}$ as $argmin$ in (46) and from (24) (see Lemma 1) it follows that there exists a function $\tilde{\mu}_1:\mathbb{N}\to\mathbb{R}$, with $\lim_{K\to\infty}\tilde{\mu}_1(K)=0$ (implying $\lim_{\epsilon\to 0}\tilde{\mu}_1(K(\epsilon))=0$) such that

$$\mathbb{E}_{\zeta}^{u_\epsilon^1}\left\{\sum_{x,a}\left|\phi^{K(\epsilon)}_{\lfloor\tau_l/\epsilon\rfloor,1}(H^1;x,a)-\omega_l^1(x,a;H^1))\right|\right\}\leq\tilde{\mu}_1(K(\epsilon)).\tag{61}$$

Hence, by (60),

$$\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\left\{\left\|\frac{1}{\Delta(\epsilon)}\int_{\tau_l}^{\tau_{l+1}} f^1(\tilde{z}(\tau_l,H^1),Y^1(t))dt-\hat{f}^1(\tilde{z}(\tau_l,H^1),\omega_l^1(H^1))\right\|\right\}\leq\tilde{L}_1\frac{\epsilon}{\Delta(\epsilon)}+M_1\tilde{\mu}_1(K(\epsilon)).\tag{62}$$

By substituting (59) and (62) into (58), one obtains

$$\begin{aligned}&\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\|Z(\tau_{l+1})-\tilde{z}(\tau_{l+1},H^1)\|\leq(1+C\Delta(\epsilon))\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\left\{\|Z(\tau_l)-\tilde{z}(\tau_l,H^1)\|\right\}\\&+\Delta(\epsilon)\tilde{\nu}(\epsilon),\end{aligned}\tag{63}$$

with $\lim_{\epsilon\to 0}\tilde{\nu}(\epsilon)=0$. By virtue of Proposition 5.1 in [10], the latter implies that there exists a function $\eta(\epsilon)$, with $\eta(\epsilon)\to 0$, such that $\forall l=1,\dots,\ell(\epsilon)$,

$$\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\|Z(\tau_l)-\tilde{z}(\tau_l,H^1)\|\leq\eta(\epsilon).\tag{64}$$

Thus,

$$\left|\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\{G^1(Z(1))\}-\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\{G^1(\tilde{z}(1,H^1))\}\right|\leq C_G^1\eta(\epsilon)\qquad\forall u_\epsilon^1\in\mathcal{U}_1.\tag{65}$$

By (53),

$$\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\left\{G^1(\tilde{z}(1,H^1))\right\}\geq G^{1*}+O(\Delta(\epsilon)),$$

and from (65) it follows that

$$\underline{\lim}_{\epsilon\to 0}\,\mathbb{E}_{\zeta}^{u_\epsilon^1,u_\epsilon^{2*}}\{G^1(Z(1))\}\geq G^{1*}\quad\forall u_\epsilon^1\in\mathcal{U}_1.$$

A similar procedure can be applied to show that

$$\underline{\lim}_{\epsilon\to 0}\,\mathbb{E}_{\zeta}^{u_\epsilon^{1*},u_\epsilon^2}\{G^2(Z(1))\}\geq G^{2*}.$$

Since, by Proposition (1),

$$G_i^*=\lim_{\epsilon\to 0}\mathbb{E}_x^{u_\epsilon^{1*},u_\epsilon^{2*}}\{G^i(Z(1))\},\quad i=1,2,$$

the proposition is proved.

□

References

[1] E. Altman, K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi and D. S. Menasche, *Constrained Cost-Coupled Stochastic Games with Independent State Processes*, Operations Research Letters, 36 (2008), pp. 160-164.

[2] E. Altman and V. Gaitsgory, *Control of a Hybrid Stochastic System*, Sistem & Control Letters 20(1993):4, pp. 307-314.

[3] E. Altman and V. Gaitsgory, *A Hybrid Differential Stochastic Zero Sum Game with Fast Stochastic Part*, Annals of the International Society of Dynamic Games , 3 (1995), pp. 47-59.

[4] E. Altman and V. Gaitsgory, *Asymptotic Optimization of a Nonlinear Hybrid System Controlled by a Markov Decision Process*, SIAM J., Control Optim. 35 (1997): 6, pp. 2070-2085.

[5] E. Altman and Y. Hayel, *Markov Decision Evolutionary Games*, in IEEE Transactions on Automatic Control, 55 (2010):7.

[6] O. Alvarez and M. Bardi *Ergodicity, stabilization, and singular perturbations for Bellman- Isaacs equations*, Mem. Amer. Math. Soc. 204 (2010), pp. 1-88.

[7] T. Basar and G.J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. SIAM, Philadelphia, PA, 1999.

[8] I. Brunetti, Y. Hayel and E. Altman, *State Policy Couple Dynamics in Evolutionary Games*, ACC Conference, Chicago, 2015.

[9] C. Derman, *Finite State Markovian Decision Processes*, Academic Press, New York, 1970

[10] V. Gaitsgory, *Suboptimization of Singularly Perturbed Control Problems*, SIAM J. Control Optim. 30 (1992):5, pp. 1228-1240.

[11] V. Gaitsgory, *Limit Hamilton-Jacobi-Isaacs Equations for Singularly Perturbed Zero-Sum Differential Games*, J. Math. Anal. Appl. 202 (1996):3, pp. 862-899.

[12] V. Gaitsgory and S. Rossomakhne, *Averaging and Linear Programming in Some Singularly Perturbed Problems of Optimal Control*, Journal of Applied Mathematics and Optimization, 71(2015), pp. 195-276.

[13] M. T. Nguyen, E. Altman and V. Gaitsgory, *On Stochastic Hybrid Zero-Sum Games with Nonlinear Slow Dynamics*, Annals of the International Society of Dynamic Games, 6 (2001), pp 129-145.

[14] R.C. Scalzo and S.A. Williams, *On the Existence of a Nash Equilibrium Point for N-Person Differential Games* , Applied Mathematics and Optimization, 2 (1976):3, pp. 271-278

[15] M. Sirbu *Stochastic Perron's Method and Elementary Strategies for Zero-Sum Differential Games*, SIAM J. Control Optim., 52 (2014):3, pp. 1693-1711.

[16] P. Shi, E. Altman and V. Gaitsgory, *On Asymptotic Optimization of a Class of Nonlinear Stochastic Hybrid Systems*, in Mathematical Methods of Operations Research 47(1998), pp.289-315.