Dyn Games Appl (2017) 7:1–20
DOI 10.1007/s13235-015-0173-z

Extremal Shift Rule for Continuous-Time Zero-Sum Markov Games

Yurii Averboukh[1,2]

Published online: 29 October 2015

Abstract In the paper we consider the controlled continuous-time Markov chain describing the interacting particles system with the finite number of types. The system is controlled by two players with the opposite purposes. This Markov game converges to a zero-sum differential game when the number of particles tends to infinity. Krasovskii–Subbotin extremal shift provides the optimal strategy in the limiting game. The main result of the paper is the near optimality of the Krasovskii–Subbotin extremal shift rule for the original Markov game.

Keywords Continuous-time Markov games · Differential games · Extremal shift rule · Control with guide strategies

Mathematics Subject Classification 91A15 · 91A23 · 91A05

1 Introduction

The paper is concerned with the construction of near-optimal strategies for zero-sum two-player continuous-time Markov game based on deterministic game. The term 'Markov game' is used for a Markov chain with the Kolmogorov matrix depending on controls of players. These games are also called continuous-time stochastic games. Continuous-time Markov games were first studied by Zachrisson [20]. The recent progress in the theory of continuous-time Markov games can be found in [15,17] and references therein.

We consider the case when the continuous-time Markov chain describes the interacting particle system. The interacting particle system converges to the deterministic system as the number of particles tends to infinity [7,8] (see also [3,4]). The value function of the controlled Markov chain converges to the value function of the limiting control system [8]

✉ Yurii Averboukh
ayv@imm.uran.ru

1 Krasovskii Institute of Mathematics and Mechanics UrB RAS, Yekaterinburg, Russia

2 Ural Federal University, Yekaterinburg, Russia

(see the corresponding result for the discrete-time systems in [6]). This result is extended to the case of zero-sum games as well as to the case of nonzero-sum games in [8]. If the nonanticipative strategy is optimal for the differential game, then it is near optimal for the Markov game [8]. However, the nonanticipative strategies require the knowledge of the control of the second player. Often this information is inaccessible, and the player has only the information about the current position. In this case, one can use feedback strategies or control with guide strategies.

Control with guide strategies were proposed by Krasovskii and Subbotin to construct the solution of the deterministic differential game under informational disturbances [13]. Note that feedback strategies do not provide the stable solution of the differential game. If the player uses control with guide strategy, then the control is formed stepwise, and the player has a model of the system that is used to choose an appropriate control on each step. The easiest way to construct a control with guide strategy is the extremal shift rule. The value function is achieved in the limit when the time between control corrections tends to zero. In the original work by Krasovskii and Subbotin, the motion of the model is governed by the system that is a copy of the original system and the motion of the original system is close to the motion of the model. Therefore the model can be called guide. Formally, control with guide strategy is a strategy with memory. However, it suffices to store only the finite number of vectors. Moreover, the player does not require the information on second player's control.

The control with guide strategies realizing the extremal shift was used for the differential games without Lipschitz continuity assumption on the system dynamics in [14] and for the games governed by delay differential equations in [12,16]. Krasovskii and Kotelnikova proposed the stochastic control with guide strategies [9–11]. In that case, the real motion of the deterministic system is close to the auxiliary stochastic process generated by optimal control for the stochastic differential game. The Nash equilibrium for two-player game in the class of control with guide strategies was constructed via extremal shift in [1].

In this paper, we let the player use the control with guide strategy realizing the extremal shift rule in the Markov game. We assume that the motion of the guide is given by the limiting deterministic differential game. We estimate the expectation of the distance between the Markov chain and the motions of the model (guide). This leads to the estimate between the outcome of the player in the Markov game and the value function of the limiting differential game.

The paper is organized as follows. In preliminary Sect. 2, we introduce the Markov game describing the interacting particle system and the limiting deterministic differential game. In Sect. 3, we give the explicit definition of control with guide strategies and formulate the main results. Section 4 is devoted to a property of transition probabilities. In Sect. 5, we estimate the expectation of distance between the Markov chain and the deterministic guide. Section 6 provides the proofs of the statements formulated in Sect. 3. Finally, in Sect. 7, we illustrate the theoretical results by a simulation of some Markov chain.

2 Preliminaries

We consider the system of the finite number of particles. Each particle can be of type i, $i \in \{1, \dots, d\}$. The type of each particle is a random variable governed by a Markov chain. To specify this chain, consider the Kolmogorov matrix $Q(t, x, u, v) = (Q_{ij}(t, x, u, v))_{i,j=1}^{d}$. That means that the elements of matrix $Q(t, x, u, v)$ satisfy the following properties

– $Q_{ij}(t, x, u, v) \geq 0$ for $i \neq j$;

–

$$Q_{ii}(t,x,u,v) = -\sum_{j\neq i} Q_{ij}(t,x,u,v). \tag{1}$$

Here $t \in [0,T]$, $x \in \Sigma_d = \{(x_1,\dots,x_d) : x_i \geq 0, x_1+\cdots+x_n = 1\}$, $u \in U$, $v \in V$. The parameter x is used below for the state of the interacting particle system. We assume that $x = (x_1,\dots,x_d)$ is a row vector. The variables u and v are controlled by the first and the second players, respectively.

Additionally we assume that

– U and V are compact sets;
– Q is a continuous function;
– for any t, u and v, the function $x \mapsto Q(t,x,u,v)$ is Lipschitz continuous;
– for any $t \in [0,T]$, $\xi, x \in \mathbb{R}^n$, the following equality holds true.

$$\min_{u\in U}\max_{v\in V}\langle \xi, xQ(t,x,u,v)\rangle = \max_{v\in V}\min_{u\in U}\langle \xi, xQ(t,x,u,v)\rangle \tag{2}$$

Condition (2) is an analog of well-known Isaacs condition.

For fixed parameters $x \in \Sigma_d$, $u \in U$, and $v \in V$, the type of each particle is determined by the Markov chain with the generator

$$(Q(t,x,u,v)f)_i = \sum_{j\neq i} Q_{ij}(t,x,u,v)(f_j - f_i), \quad f = (f_1,\dots,f_d).$$

The another way to specify the Markov chain is the Kolmogorov forward equation

$$\frac{\mathrm{d}}{\mathrm{d}t}P(s,t,x) = P(s,t,x)Q(t,x,u,v).$$

Here $P(s,t,x) = (P_{ij}(s,t,x))_{ij=1}^d$ is the matrix of the transition probabilities.

Now we consider the controlled mean-field interacting particle system (see for details [8]). Let n_i be a number of particles of the type i. The vector $N = (n_1,\dots,n_d) \in \mathbb{Z}_+^d$ is the state of the system consisting of $|N| = n_1+\cdots+n_d$ particles. For $i \neq j$ and a vector $N = (n_1,\dots,n_d)$ denote by $N^{[ij]}$ the vector obtained from N by removing one particle of type i and adding one particle of type j, i.e., we replace the ith coordinate with $n_i - 1$ and the jth coordinate with $n_j + 1$. The mean-field interacting particle system is a Markov chain with the generator

$$\sum_{i,j=1}^{d} n_i Q_{ij}(t, N/|N|, u(t), v(t))\left[f\left(N^{[ij]}\right) - f(N)\right].$$

The purpose of the first (respectively, second) player is to minimize (respectively, maximize) the expectation of $\sigma(N/|N|)$.

Denote the inverse number of particles by $h = 1/|N|$. Normalizing the states of the interacting particle system, we get the generator (see [8])

$$L_t^h[u,v]f(N/|N|) = \sum_{i,j=1}^{d} \frac{1}{h}\frac{n_i}{|N|} Q_{ij}(t, N/|N|, u(t), v(t))\left[f\left(\frac{N^{[ij]}}{|N|}\right) - f\left(\frac{N}{|N|}\right)\right]. \tag{3}$$

Denote the vector $N/|N|$ by $x = (x_1, \dots, x_d)$. Thus, we have that

$$L_t^h[u, v]f(x) = \sum_{i,j=1}^{d} \frac{1}{h} x_i Q_{ij}(t, x, u(t), v(t))[f(x - he^i + he^j) - f(x)].$$

Here e^i is the ith coordinate vector. The vector x belongs to the set

$$\Sigma_d^h = \{(x_1, \dots, x_d) : x_i \in h\mathbb{Z}, \ \ x_1 + \cdots + x_d = 1\} \subset \Sigma_d.$$

Further, let $\mathcal{U}_{\text{det}}[s]$ (respectively, $\mathcal{V}_{\text{det}}[s]$) denote the set of deterministic controls of the first (respectively, second) player on $[s, T]$, i.e.,

$$\mathcal{U}_{\text{det}}[s] = \{u : [s, T] \to U \text{ measurable}\},$$
$$\mathcal{V}_{\text{det}}[s] = \{v : [s, T] \to V \text{ measurable}\}.$$

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ be a filtered probability space. Extending the definition given in [5, p. 135] to the stochastic game case, we say that the pair of stochastic processes u and v on $[s, T]$ is an admissible pair of controls if

1. $u(t) \in U$, $v(t) \in V$;
2. the processes u and v are progressive measurable;
3. for any $y \in \Sigma_d^h$, there exists an unique $\{\mathcal{F}_t\}_{t\in[s,T]}$-adapted càdlàg stochastic process $X^h(t, s, y, u, v)$ taking values in Σ_d^h, starting at y at time s and satisfying the following condition for any $f \in C(\Sigma_d^h)$

$$f\left(X^h(t, s, y, u, v)\right) - \int_s^t L_t^h[u(\tau), v(\tau)]f\left(X^h(\tau, s, y, u, v)\right) \mathrm{d}\tau$$

is a martingale.

In particular, the third condition means that

$$\mathbb{E}_{sy}^h f\left(X^h(t, s, y, u, v)\right) - f(y) = \int_s^t \mathbb{E}_{sy}^h L_t^h[u(\tau), v(\tau)]f\left(X^h(\tau, s, y, u, v)\right) \mathrm{d}\tau. \quad (4)$$

Here $\mathbb{E}_{sy}^h$ denotes the conditional expectation of corresponding stochastic processes.

The purposes of the players can be reformulated in the following way. The first (respectively, second) player wishes to minimize (respectively, maximize) the payoff

$$\mathbb{E}_{sy}^h \sigma\left(X^h(T, s, y, u, v)\right).$$

Let $\mathcal{U}^h[s]$ be a set of stochastic processes u taking values in U such that the pair (u, v) is admissible for any $v \in \mathcal{V}_{\text{det}}[s]$. Analogously, let $\mathcal{V}^h[s]$ be a set of stochastic processes v taking values in V such that the pair (u, v) is admissible for any $u \in \mathcal{U}_{\text{det}}[s]$.

Denote by $P_{sy}^h(A)$ the conditional probability of an event A under condition that the Markov chain corresponding to the parameter h starts at y at time s, i.e.,

$$P_{sy}^h(A) = \mathbb{E}_{sy}^h \mathbf{1}_A$$

Further, let $p^h(s, y, t, z, u, v)$ denote the transition probability, i.e.,

$$p^h(s, y, t, z, u, v) = P_{sy}^h\left(X^h(t, s, y, u, v) = z\right) = \mathbb{E}_{sy}^h \mathbf{1}_{\{z\}}\left(X^h(t, s, y, u, v)\right).$$

The substituting $\mathbf{1}_{\{z\}}$ for f in (3) and (4) gives that

$$p^h(s,y,t,z,u,v) = p^h(s,y,s,z,u,v) + \frac{1}{h}\int_s^t \mathbb{E}^h_{sy} \sum_{i,j=1}^d X^h_i(\tau,s,y,u,v) Q_{ij}\left(\tau, X^h(\tau,s,y,u,v), u(\tau), v(\tau)\right) \times \left[\mathbf{1}_z\left(X^h(\tau,s,y,u,v) - he^i + he^j\right) - \mathbf{1}_z\left(X^h(\tau,s,y,u,v)\right)\right] \mathrm{d}\tau. \tag{5}$$

Here $X^h_i(\tau,s,y,u,v)$ denotes the ith component of $X^h(\tau,s,y,u,v)$.

Recall (see [8]) that if $h \to 0$, then the generator $L^h_t[u,v]$ converges to the generator

$$\Lambda_t[u,v]f(x) = \sum_{i=1}^d \sum_{j\neq i} x_i Q_{ij}(t,x,u(t),v(t))\left[\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_i}(x)\right] = \sum_{k=1}^d \sum_{i\neq k} [x_i Q_{ik}(t,x,u(t),v(t)) - x_k Q_{ki}(t,x,u(t),v(t))]\frac{\partial f}{\partial x_k}(x).$$

For controls $u \in \mathcal{U}_{\mathrm{det}}[s]$ and $v \in \mathcal{V}_{\mathrm{det}}[s]$, the deterministic evolution generated by the generator $\Lambda_t[u(t),v(t)]$ is described by the equation

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t(x) = \sum_{k=1}^d \sum_{i\neq k} [x_i Q_{ik}(t,x,u(t),v(t)) - x_k Q_{ki}(t,x,u(t),v(t))]\frac{\partial f_t}{\partial x_k}(x), \quad f_s(x) = f(x), \quad x = (x_1,\ldots,x_d) \in \Sigma_d. \tag{6}$$

Here the function $f_t(y)$ is equal to $f(x(t))$ when $x(s) = y$. The characteristics of (6) solve the ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t} x_k(t) = \sum_{i\neq k} [x_i(t) Q_{ik}(t,x(t),u(t),v(t)) - x_k(t) Q_{ki}(t,x(t),u(t),v(t))] = \sum_{i=1}^d x_i(t) Q_{ik}(t,x(t),u(t),v(t)), \quad k = \overline{1,d}.$$

One can rewrite this system in the vector form

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) = x(t) Q(t,x(t),u(t),v(t)), \quad t \in [0,T], \quad x(t) \in \mathbb{R}^d, \quad u(t) \in U, \quad v(t) \in V. \tag{7}$$

For given $u \in \mathcal{U}_{\mathrm{det}}[s]$, $v \in \mathcal{V}_{\mathrm{det}}[s]$ denote the solution of the initial value problem for (7) and the condition $x(s) = y$ by $x(\cdot,s,y,u,v)$.

Consider the deterministic zero-sum game with the dynamics given by (7) and terminal payoff equal to $\sigma(x(T,s,y,u,v))$. This game has a value in the class of feedback strategies [13] that is a continuous function of the position. Denote it by $\mathrm{Val}(s,y)$. Recall (see [18]) that the function $\mathrm{Val}(s,y)$ is a minimax (viscosity) solution of the Hamilton–Jacobi PDE

$$\frac{\partial W}{\partial t} + H(t,x,\nabla W) = 0, \quad W(T,x) = \sigma(x). \tag{8}$$

Here the Hamiltonian H is defined by the rule

$$H(t,x,\xi) = \min_{u\in U}\max_{v\in V}\langle \xi, xQ(t,x,u,v)\rangle.$$

Remark 1 The approaches based on control with guide strategies and nonanticipative strategies (see [2] for detailed presentation of this approach) are equivalent to the feedback formalization [18]. The value function in these cases is also equal to Val.

3 Control with Guide Strategies

In this section, we introduce the control with guide strategies for the Markov game. It is assumed that the control is formed stepwise and the player has an information about the current state of the system, i.e., the vector x is known. Additionally, we assume that the player can evaluate the expected state and the player's control depends on current state of the system and on the evaluated state. This evaluation is called guide. At each time of control correction, the player computes the value of the guide and the control that is used up to the next time of control correction.

Formally (see [19]), control with guide strategy of player 1 is a triple $\mathfrak{u} = (u(t, x, w), \psi_1(t_+, t, x, w), \chi_1(s, y))$. Here the function $u(t, x, w)$ is equal to the control implemented after time t if at time t the state of the system is x and the state of the guide is w. The function $\psi_1(t_+, t, x, w)$ determines the state of the guide at time t_+ under the condition that at time t the state of the system is x and the state of the guide is w. The function χ_1 initializes the guide, i.e., $\chi_1(s, y)$ is the state of the guide in the initial position (s, y).

We use the control with guide strategies for the Markov game with the generator L_t^h. Here we assume that $h > 0$ is fixed. Let (s, y) be an initial position, $s \in [0, T]$, $y \in \Sigma_d^h$. Assume that player 1 chooses the control with guide strategy $\mathfrak{u}$ and the partition $\Delta = \{t_k\}_{k=0}^m$ of the time interval $[s, T]$, whereas player 2 chooses the control $v \in \mathcal{V}^h[s]$. This control can be also formed stepwise using some second player's control with guide strategy.

We say that the stochastic process $\mathcal{X}_1^h[\cdot, s, y, \mathfrak{u}, \Delta, v]$ is generated by strategy $\mathfrak{u}$, partition Δ and the second player's control v if for $t \in [t_k, t_{k+1})$ $\mathcal{X}_1^h[t, s, y, \mathfrak{u}, \Delta, v] = X^h(t, t_k, x_k, u_k, v)$, where

- $x_0 = y$, $w_0 = \chi_1(t_0, x_0)$, $u_0 = u(t_0, x_0, w_0)$;
- for $k = \overline{1, r}$ $x_k = X^h(t_k, t_{k-1}, x_{k-1}, u_{k-1}, v)$, $w_k = \psi_1(t_k, t_{k-1}, x_{k-1}, w_{k-1})$, $u_k = u(t_k, x_k, w_k)$.

Note that even though the state of the guide w_k is determined by the deterministic function, it depends on the random variable x_{k-1}. Thus, w_k is a random variable.

Below we define the first player's control with guide strategy that realizes the extremal shift rule (see [13]). Let φ be a supersolution of Eq. (8). That means (see [18]) that for any $(t_*, x_*) \in [0, T] \times \Sigma_d$, $t_+ > t_*$ and $v_* \in V$, there exists a solution $\zeta_1(\cdot, t_+, t_*, x_*, v_*)$ of the differential inclusion

$$\dot{\zeta}_1(t) \in \operatorname{co}\{\zeta_1(t) Q(t, \zeta_1(t), u, v_*) : u \in U\}$$

satisfying the conditions

$$\zeta_1(t_*, t_+, t_*, x_*, v_*) = x_*, \quad \varphi(t_+, \zeta_1(t_+, t_+, t_*, x_*, v_*)) \le \varphi(t_*, x_*).$$

Define the control with guide strategy $\hat{\mathfrak{u}} = (\hat{u}, \hat{\psi}_1, \hat{\chi}_1)$ by the following rules. If $t_*, t_+ \in [0, T]$, $t_+ > t_*$, $x_*, w_* \in \Sigma_d$, then choose u_*, v_* by the rules

$$\min_{u \in U} \max_{v \in V} \langle x_* - w_*, x_* Q(t_*, x_*, u, v) \rangle = \max_{v \in V} \langle x_* - w_*, x_* Q(t_*, x_*, u_*, v) \rangle, \tag{9}$$

$$\max_{v \in V} \min_{u \in U} \langle x_* - w_*, x_* Q(t_*, x_*, u, v) \rangle = \min_{u \in U} \langle x_* - w_*, x_* Q(t_*, x_*, u, v_*) \rangle. \tag{10}$$

Put

(u1) $\hat{u}(t_*, x_*, w_*) = u_*$,
(u2) $\hat{\psi}_1(t_+, t_*, x_*, w_*) = \zeta_1(t_+, t_+, t_*, w_*, v_*)$,
(u3) $\hat{\chi}_1(s, y) = y$.

Note that if the first player uses the strategy $\hat{\mathfrak{u}}$ in the differential game with the dynamics given by (7), then she guarantees the limit outcome not greater than φ (see [13,18]). If additionally $\varphi = \mathrm{Val}$, then the strategy $\hat{\mathfrak{u}}$ is optimal in the deterministic game.

The main result of the paper is the following.

Theorem 1 *Assume that σ is Lipschitz continuous with a constant R, and the function φ is a supersolution of (8). If the first player uses the control with guide strategy $\hat{\mathfrak{u}}$ determined by (u1)–(u3), then*

(i)

$$\limsup_{\delta\downarrow 0}\left\{\mathbb{E}^h_{sy}\left(\sigma\left(\mathcal{X}^h_1[T, s, y, \hat{\mathfrak{u}}, \Delta, v]\right)\right) : d(\Delta)\le\delta, v\in\mathcal{V}^h[s]\right\} \le \varphi(s, y) + R\sqrt{Dh}.$$

(ii)

$$\limsup_{\delta\downarrow 0}\left\{P^h_{sy}\left(\sigma\left(\mathcal{X}^h_1[T, s, y, \hat{\mathfrak{u}}, \Delta, v]\right)\ge\varphi(s, y) + R\sqrt[3]{Dh}\right) : d(\Delta)\le\delta,\ \ v\in\mathcal{V}^h[s]\right\} \le \sqrt[3]{Dh}.$$

Here D is a constant not dependent on φ and σ.

The theorem is proved in Sect. 6.

Now let us consider the case when the second player uses control with guide strategies. The control with guide strategy of the second player is a triple $\mathfrak{v} = (v(t, x, w), \psi_2(t_+, t, x, w), \chi_2(s, y))$. Here w denotes the state of the second player's guide. The control in this case is formed also stepwise. If (s, y) is an initial position, Δ is a partition of time interval $[s, T]$ and $u\in\mathcal{U}^h[s]$ is a control of player 1, then denote by $\mathcal{X}^h_2[\cdot, s, y, \mathfrak{v}, \Delta, u]$ the corresponding stochastic process.

Let ϕ be a subsolution of Eq. (8). That means (see [18]) that for any $(t_*, x_*)\in[0, T]\times\Sigma_d$, $t_+ > t_*$ and u^* there exists a trajectory $\zeta_2(\cdot, t_+, t_*, x_*, u^*)$ of the differential inclusion

$$\dot{\zeta}_2(t)\in\mathrm{co}\{\zeta_2(t)Q(t, \zeta_2(t), u^*, v) : v\in V\},\quad \zeta_2(t_*) = x_*$$

satisfying the condition $\phi(t_+, \zeta_2(t_+, t_+, t_*, x_*, u^*))\ge\phi(t_*, x_*)$.

Define the strategy $\hat{\mathfrak{v}}$ by the following rule. If (t_*, x_*) is a position, $t_+ > t_*$ and $w_*\in\Sigma_d$ is a state of the guide, then choose v^* and u^* by the rules

$$\min_{v\in V}\max_{u\in U}\langle x_* - w_*, x_*Q(t_*, x_*, u, v)\rangle = \max_{u\in U}\langle x_* - w_*, x_*Q(t_*, x_*, u, v^*)\rangle,$$
$$\max_{u\in U}\min_{v\in V}\langle x_* - w_*, x_*Q(t_*, x_*, u, v)\rangle = \min_{v\in V}\langle x_* - w_*, x_*Q(t_*, x_*, u^*, v)\rangle.$$

Put

(v1) $v(t_*, x_*, w_*) = v^*$,
(v2) $\psi_2(t_+, t_*, x_*, w_*) = \zeta_2(t_+, t_+, t_*, x_*, u^*)$
(v3) $\chi_2(s, y) = y$.

Corollary 1 *Let ϕ be a subsolution of (8). If the second player uses the control with guide strategy $\hat{\mathfrak{v}}$ determined by (v1)–(v3), then*

(i)

$$\liminf_{\delta\downarrow 0}\left\{\mathbb{E}^h_{sy}\left(\sigma\left(\mathcal{X}^h_2[T,s,y,\hat{\mathfrak{u}},\Delta,v]\right)\right) : d(\Delta)\le\delta, u\in\mathcal{U}^h[s]\right\} \ge \phi(s,y) - R\sqrt{Dh}.$$

(ii)

$$\limsup_{\delta\downarrow 0}\left\{P^h_{sy}\left(\sigma\left(\mathcal{X}^h_2[T,s,y,\hat{\mathfrak{v}},\Delta,u]\right)\le\phi(s,y)-R\sqrt[3]{Dh}\right) : d(\Delta)\le\delta,\ \ u\in\mathcal{U}^h[s]\right\}\le\sqrt[3]{Dh}.$$

The corollary is also proved in Sect. 6.

4 Properties of Transition Probabilities

Now we prove the following.

Lemma 1 *There exists a function $\alpha^h(\delta)$ such that $\alpha^h(\delta)\to 0$ as $\delta\to 0$ and for any $t_*, t_+\in[0,T]$, $\xi,\eta\in\Sigma_d$, $\xi=(\xi_1,\dots,\xi_d)$, $\bar{u}\in U$, $\bar{v}\in\mathcal{V}^h[t_*]$ the following properties hold true*

1. *if $\eta=\xi$, then*

$$p^h(t_*,\xi,t_+,\eta,\bar{u},\bar{v}) \le 1+\frac{1}{h}\sum_{k=1}^{d}\int_{t_*}^{t_+}\int_V \xi_k Q_{kk}(t_*,\xi,\bar{u},v)\nu_\tau(dv)d\tau + \alpha^h(t_+-t_*)\cdot(t_+-t_*);$$

2. *if $\eta=\xi-he^i+he^j$, then*

$$p^h(t_*,\xi,t_+,\eta,u,v) \le \frac{1}{h}\int_{t_*}^{t_+}\int_V \xi_i Q_{ij}(t_*,\xi,\bar{u},v)\nu_\tau(dv)d\tau+\alpha^h(t_+-t_*)\times(t_+-t_*);$$

3. *if $\eta\ne\xi$ and $\eta\ne\xi-he^i+he^j$, then*

$$p^h(t_*,\xi,t_+,\eta,u,v)\le\alpha^h(t_+-t_*)\times(t_+-t_*);$$

Here ν_τ is a measure on V depending on t_*, t_+, ξ, η, $\bar{u}$ and $\bar{v}$.

Proof First denote

$$K=\sup\{|Q_{ij}(t,x,u,v)| : i,j=\overline{1,d},\ \ t\in[0,T],\ \ x\in\Sigma_d,\ \ u\in U,\ \ v\in Q\}. \tag{11}$$

For any $x\in\Sigma_d$, $t\in[0,T]$, $u\in U$, $v\in V$, the following estimates hold true

$$\|x\|\le\sqrt{d},\ \ \left|\sum_{i=1}^{n}x_i Q_{ij}(t,x,u,v)\right|\le K,\ \ \|xQ(t,x,u,v)\|\le K\sqrt{d}. \tag{12}$$

Further, let $\gamma(\delta)$ be a common modulus of continuity with respect to t of the functions Q_{ij}, i.e., for all $i, j, t', t'' \in [0, T]$, $x \in \Sigma_d$, $u \in U$, $v \in Q$

$$|Q_{ij}(t', x, u, v) - Q_{ij}(t'', x, u, v)| \le \gamma(t'' - t') \tag{13}$$

and $\gamma(\delta) \to 0$ as $\delta \to 0$. From (5) and (12), we obtain that

$$p^h(t_*, \xi, t, \eta, u, v) \le p^h(t_*, \xi, t_*, \eta, u, v) + \frac{2Kd}{h}(t - t_*). \tag{14}$$

Further, for a given control $\bar{v} \in \mathcal{V}^h[t_*]$, let $\mathbb{E}^h_{t_*\xi;\tau x}$ denote the expectation under conditions $X^h(t_*, t_*, \xi, \bar{u}, \bar{v}) = \xi$, and $X^h(\tau, t_*, \xi, \bar{u}, \bar{v}) = x$.

We have that

$$\mathbb{E}^h_{t_*\xi} f = \sum_{x \in \Sigma^h_d} \mathbb{E}^h_{t_*\xi;\tau x} f \times p^h(t_*, \xi, \tau, x, \bar{u}, \bar{v}).$$

From this and (5), we get

$$\begin{aligned}
p^h(t_*, \xi, t_+, \eta, \bar{u}, \bar{v}) &= p^h(t_*, \xi, t_*, \eta, \bar{u}, \bar{v}) \\
&+ \frac{1}{h}\int_{t_*}^{t_+} \sum_{x \in \Sigma^h_d} \mathbb{E}^h_{t_*\xi;\tau x} \sum_{i,j=1}^{d} x_i Q_{i,j}(\tau, x, \bar{u}, \bar{v}(\tau)) [\mathbf{1}_\eta(x - he^i + he^j) - \mathbf{1}_\eta(x)] \\
&\times p^h(t_*, \xi, \tau, x, \bar{u}, \bar{v}) \mathrm{d}\tau \le p^h(t_*, \xi, t_*, \eta, \bar{u}, \bar{v}) \\
&+ \frac{1}{h}\int_{t_*}^{t_+} \sum_{x \in \Sigma^h_d} \mathbb{E}^h_{t_*\xi;\tau x} \sum_{i,j=1}^{d} x_i Q_{i,j}(\tau, x, \bar{u}, \bar{v}(\tau)) [\mathbf{1}_\eta(x - he^i + he^j) - \mathbf{1}_\eta(x)] \\
&\times p^h(t_*, \xi, t_*, x, \bar{u}, \bar{v}) \mathrm{d}\tau + \frac{2K^2 d^2}{h}(t - t_*)^2.
\end{aligned}$$

Here x_i denotes the ith component of the vector x.

We have that $p^h(t_*, \xi, t_*, x, \bar{u}, \bar{v}) = 1$ for $x = \xi$ and $p^h(t_*, \xi, t_*, x, \bar{u}, \bar{v}) = 0$ for $x \ne \xi$. Thus,

$$\begin{aligned}
p^h(t_*, \xi, t_+, \eta, \bar{u}, \bar{v}) &\le p^h(t_*, \xi, t_*, \eta, \bar{u}, \bar{v}) \\
&+ \frac{1}{h}\int_{t_*}^{t_+} \mathbb{E}^h_{t_*\xi;\tau\xi} \sum_{i,j=1}^{d} \xi_i Q_{i,j}(\tau, \xi, \bar{u}, \bar{v}(\tau)) [\mathbf{1}_\eta(\xi - he^i + he^j) - \mathbf{1}_\eta(\xi)] \\
&+ \frac{2K^2 d^2}{h}(t - t_*)^2 \\
&\le p^h(t_*, \xi, t_*, \eta, \bar{u}, \bar{v}) + \frac{1}{h}\int_{t_*}^{t_+} \mathbb{E}^h_{t_*\xi;\tau\xi} \\
&\sum_{i,j=1}^{d} \xi_i Q_{i,j}(t_*, \xi, \bar{u}, \bar{v}(\tau)) [\mathbf{1}_\eta(\xi - he^i + he^j) - \mathbf{1}_\eta(\xi)] \mathrm{d}\tau \\
&+ \frac{2K^2 d^2}{h}(t - t_*)^2 + \frac{2d}{h}\gamma(t - t_*) \times (t - t_*).
\end{aligned}$$

Recall that for each τ $\bar{v}(\tau)$ is a random variable with values in V. Define the measure ν_τ on V as the image measure of $P_{t_*\xi;\tau\xi}$ by $\bar{v}(\tau)$, where for $A \in \mathcal{F}$

$$P_{t_*\xi;\tau\xi}(A) = \mathbb{E}_{t_*\xi;\tau\xi} \mathbf{1}_A.$$

We have that

$$\mathbb{E}_{t_*\xi;\tau\xi} Q_{ij}(t_*,\xi,\bar{u},\bar{v}(\tau)) = \int_\Omega Q_{ij}(t_*,\xi,\bar{u},\bar{v}(\tau,\omega))P_{t_*\xi;\tau\xi}(\mathrm{d}\omega)$$
$$= \int_V Q_{ij}(t_*,\xi,\bar{u},v)\nu_\tau(\mathrm{d}v).$$

Consequently,

$$p^h(t_*,\xi,t_+,\eta,\bar{u},\bar{v}) \le p^h(t_*,\xi,t_*,\eta,\bar{u},\bar{v})$$
$$+\frac{1}{h}\int_{t_*}^{t_+}\int_V \sum_{i,j=1}^d \xi_i Q_{i,j}(t_*,\xi,\bar{u},v)[\mathbf{1}_\eta(\xi - he^i + he^j) - \mathbf{1}_\eta(\xi)]\nu_\tau(\mathrm{d}v)\mathrm{d}\tau$$
$$+\alpha^h(t-t_*)\times(t-t_*). \tag{15}$$

Here we denote

$$\alpha^h(\delta) = \frac{2K^2d^2}{h}\delta + \frac{2d}{h}\gamma(\delta).$$

From (15) the second and third statements of the Lemma follow. To derive the first statement, use the property of Kolmogorov matrixes (1). We have that

$$p^h(t_*,\xi,t_+,\xi,\bar{u},\bar{v}) \le p^h(t_*,\xi,t_*,\eta,\bar{u},\bar{v})$$
$$-\frac{1}{h}\int_{t_*}^{t_+}\int_V \sum_{i=1}^d\sum_{j\ne i}\xi_i Q_{i,j}(t_*,\xi,\bar{u},v)\nu_\tau(\mathrm{d}v)\mathrm{d}\tau + \alpha^h(t-t_*)\times(t-t_*)$$
$$= 1 + \frac{1}{h}\int_{t_*}^{t_+}\int_V \sum_{i=1}^d \xi_i Q_{i,i}(t_*,\xi,\bar{u},v)\nu_\tau(\mathrm{d}v)\mathrm{d}\tau + \alpha^h(t-t_*)\times(t-t_*).$$

□

5 Key Estimate

This section provides the estimate of the distance between the controlled Markov chain and the guide. This estimate is an analog of [13, Lemma 2.3.1].

Lemma 2 *There exist constants $\beta, C > 0$, and a function $\varkappa^h(\delta)$ such that $\varkappa^h(\delta)\to 0$ as $\delta\to 0$ and the following property holds true.*
If

1. $(t,x)\in[0,T]\times\Sigma_d^h$, $w_*\in\Sigma_d$, $t_+ > t_*$,
2. *The controls u_*, v_* are chosen by rules (9) and (10) respectively,*
3. $w_+ = \zeta_1(t_+,t_+,t_*,w_*,v_*)$,

then for any $v\in\mathcal{V}^h[t_*]$

$$\mathbb{E}^h_{t_*x_*}\left(\|\mathcal{X}(t_+,t_*,x_*,u_*,v) - w_+\|^2\right)$$
$$\le (1+\beta(t_+-t_*))\|x_*-w_*\|^2 + Ch(t_+-t_*) + \varkappa^h(t_+-t_*)\times(t-t_*).$$

Proof Denote the ith component of vector x_* by x_{*i}.

We have that

$$\mathbb{E}^h_{t_* x_*}(\|\mathcal{X}(t_+, t_*, x_*, u_*, v) - w_+\|^2) = \sum_{z\in\Sigma^h_d} \|z - w_+\|^2 p^h(t_*, x_*, t_+, z, u_*, v). \tag{16}$$

Further,

$$\begin{aligned}\|z - w_+\|^2 &= \|(z - x_*) + (x_* - w_*) + (w_* - w_+)\|^2 \\ &= \|x_* - w_*\|^2 + 2\langle x_* - w_*, z - x_*\rangle \\ &\quad - 2\langle x_* - w_*, w_+ - w_*\rangle + \|z - x_*\|^2 + \|w_+ - w_*\|^2.\end{aligned}$$

It follows from (12) that

$$\left\|\frac{\mathrm{d}}{\mathrm{d}t}\zeta_1(t_+, t, t_*, w_*, v_*)\right\| \le K\sqrt{d}, \quad \|w_+ - w_*\|^2 \le K^2 d(t_+ - t_*)^2. \tag{17}$$

From Lemma 1, it follows that

$$\begin{aligned}&\sum_{z\in\Sigma^h_d} \|z - x_*\|^2 p^h(t_*, x_*, t_+, z, u_*, v) \\ &\quad\le \sum_{i=1}^{d}\sum_{j\ne i} \| - he^i + he^j\|^2 \frac{1}{h}\int_{t_*}^{t_+}\int_V Q_{ij}(t_*, x_*, u_*, v)\nu_\tau(\mathrm{d}v)\mathrm{d}\tau \\ &\qquad + 2d^3\alpha^h(t_+ - t_*)\times(t_+ - t_*) \\ &\quad\le 2hd^2K(t_+ - t_*) + 2d^3\alpha^h(t_+ - t_*)\times(t_+ - t_*).\end{aligned} \tag{18}$$

For simplicity denote $\zeta_*(t) = \zeta_1(t, t_+, t_*, w_*, u, v_*)$. We have that for each t there exists a probability μ_t on U such that

$$\frac{\mathrm{d}\zeta_*}{\mathrm{d}t}(t) = \int_U \zeta_*(t)Q(t, \zeta_*(t), u, v_*)\mu_t(\mathrm{d}u).$$

Therefore,

$$\begin{aligned}&\sum_{z\in\Sigma^h_d} \langle x_* - w_*, w_+ - w_*\rangle p^h(t_*, x_*, t_+, z, u_*, v) \\ &\quad= \left\langle x_* - w_*, \int_{t_*}^{t_+}\int_{u\in U} \zeta_*(t)Q(t, \zeta_*(t), u, v_*)\mu_t(\mathrm{d}u)\mathrm{d}t\right\rangle.\end{aligned} \tag{19}$$

Define

$$\begin{aligned}\varrho(\delta) \triangleq \sup\Big\{&|y''Q(t'', y'', u, v) - y'Q(t', y', u, v)| : \\ &t', t'' \in [0, T], \quad y', y'' \in \Sigma_d, \quad u \in U, \quad v \in V, \\ &|t' - t''| \le \delta, \quad \|y' - y''\| \le \delta K\sqrt{d}\Big\}.\end{aligned} \tag{20}$$

We have that $\varrho(\delta) \to 0$, as $\delta \to 0$. From (17), (19), and (20), it follows that

$$\begin{aligned}\sum_{z\in\Sigma_d^h} &\langle x_* - w_*, w_+ - w_*\rangle p^h(t_*, x_*, t_+, z, u_*, v)\\ &\geq \left\langle x_* - w_*, \int_{t_*}^{t_+}\int_{u\in U} w_* Q(t_*, w_*, u, v_*)\mu_t(\mathrm{d}u)\mathrm{d}t\right\rangle\\ &\quad - \sqrt{2d}\varrho(t_+ - t_*)\times(t_+ - t_*).\end{aligned} \tag{21}$$

Using Lemma 1 one more time we get the inequality

$$\begin{aligned}\sum_{z\in\Sigma_d^h} &\langle x_* - w_*, z - x_*\rangle p^h(t_*, x_*, t_+, z, u_*, v)\\ &\leq \sum_{i=1}^d\sum_{j\neq i}\langle x_* - w_*, -he^i + he^j\rangle\frac{1}{h}\int_{t_*}^{t_+}\int_V x_{*i} Q_{ij}(t_*, x_*, u_*, v)\nu_t(\mathrm{d}v)\mathrm{d}t\\ &\quad + 2d^3\alpha^h(t_+ - t_*)\times(t_+ - t_*).\end{aligned} \tag{22}$$

The first term in the right-hand side of (22) can be transformed as follows. Denote for simplicity

$$\widehat{Q}_{ij} = \int_{t_*}^{t_+}\int_V Q_{ij}(t_*, x_*, u_*, v)\nu_t(\mathrm{d}v)\mathrm{d}t.$$

Note that $\widehat{Q} = (\widehat{Q}_{ij})_{i,j=1}^d$ is a Kolmogorov matrix. We have that

$$\begin{aligned}\sum_{i=1}^d\sum_{j\neq i}&(-he^i + he^j)\frac{1}{h}\int_{t_*}^{t_+}\int_V x_{*i} Q_{ij}(t_*, x_*, u_*, v)\nu_t(\mathrm{d}v)\mathrm{d}t\\ &= \sum_{i=1}^d\sum_{j\neq i} e^j x_{*,i}\widehat{Q}_{ij} - \sum_{i=1}^d x_{*,i}e^i\sum_{j\neq i}\widehat{Q}_{ij}\\ &= \sum_{i=1}^d\sum_{j=1}^d e^j x_{*,i}\widehat{Q}_{ij} = \sum_{j=1}^d\left[\sum_{i=1}^d x_{*,i}\widehat{Q}_{ij}\right]e^j = x_*\widehat{Q}.\end{aligned}$$

This and (22) yield the estimate

$$\begin{aligned}\sum_{z\in\Sigma_d^h} &\langle x_* - w_*, z - x_*\rangle p^h(t_*, x_*, t_+, z, u_*, v)\\ &\leq \left\langle x_* - w_*, \int_{t_*}^{t_+}\int_V x_* Q(t_*, x_*, u_*, v)\nu_t(\mathrm{d}v)\mathrm{d}t\right\rangle\\ &\quad + 2d^3\alpha^h(t_+ - t_*)\times(t_+ - t_*).\end{aligned} \tag{23}$$

Substituting (17)–(21), (23) in (16), we get the inequality

$$\begin{aligned}\mathbb{E}^h_{t_* x_*}(\|\mathcal{X}(t_+, t_*, x_*, u_*, v) - w_+\|^2) &\le \|x_* - w_*\|^2 \\ &+ 2\left\langle x_* - w_*, \int_{t_*}^{t_+}\int_V x_* Q(t_*, x_*, u_*, v)\nu_t(\mathrm{d}v)\mathrm{d}t\right\rangle \\ &- 2\left\langle x_* - w_*, \int_{t_*}^{t_+}\int_{u\in U} w_* Q(t_*, w_*, u, v_*)\mu_t(\mathrm{d}u)\mathrm{d}t\right\rangle \\ &+ 2Kd^2 h(t_+ - t_*) + \left(6d^3\alpha^h(t_+ - t_*) + \sqrt{2d}\varrho(t_+ - t_*)\right)\times(t_+ - t_*).\end{aligned} \tag{24}$$

Let Υ be a Lipschitz constant of the function $y \mapsto yQ(t, y, u, v)$, i.e., for all $y', y'' \in \Sigma_d$, $t \in [0, T]$, $u \in U$, $v \in Q$

$$\|y'Q(t, y', u, v) - y''Q(t, y'', u, v)\| \le \Upsilon\|y' - y''\|.$$

We have that

$$\begin{aligned}&2\left\langle x_* - w_*, \int_{t_*}^{t_+}\int_V x_* Q(t_*, x_*, u_*, v)\nu_t(\mathrm{d}v)\mathrm{d}t\right\rangle \\ &\quad - 2\left\langle x_* - w_*, \int_{t_*}^{t_+}\int_{u\in U} w_* Q(t_*, w_*, u, v_*)\mu_t(\mathrm{d}u)\mathrm{d}t\right\rangle \\ &\le 2\int_{t_*}^{t_+}\int_{u\in U}\int_{v\in V}\Big[\left\langle x_* - w_*, x_* Q(t_*, x_*, u_*, v)\right\rangle \\ &\quad - \left\langle x_* - w_*, x_* Q(t_*, x_*, u, v_*)\right\rangle\Big]\nu_t(\mathrm{d}v)\mu_t(\mathrm{d}u)\mathrm{d}t \\ &\quad + 2\Upsilon\|x_* - w_*\|^2(t_+ - t_*).\end{aligned}$$

The choice of u_* and v_* gives that for all $u \in U$, $v \in V$

$$\langle x_* - w_*, x_* Q(t_*, x_*, u_*, v)\rangle \le \langle x_* - w_*, x_* Q(t_*, x_*, u, v_*)\rangle.$$

Consequently, we get the estimate

$$\begin{aligned}&2\left\langle x_* - w_*, \int_{t_*}^{t_+}\int_V x_* Q(t_*, x_*, u_*, v)\nu_t(\mathrm{d}v)\mathrm{d}t\right\rangle \\ &\quad - 2\left\langle x_* - w_*, \int_{t_*}^{t_+}\int_{u\in U} w_* Q(t_*, w_*, u, v_*)\mu_t(\mathrm{d}u)\mathrm{d}t\right\rangle \\ &\le 2\Upsilon\|x_* - w_*\|^2(t_+ - t_*).\end{aligned}$$

From this and (24), the conclusion of the Lemma follows for $\beta = 2\Upsilon$, $C = 2d^2K$, $\varkappa^h(\delta) = 6d^3\alpha^h(\delta) + \sqrt{2d}\varrho(\delta)$. □

6 Near-Optimal Strategies

In this section, we prove Theorem 1 and Corollary 1.

Proof of Theorem 1 Let $v \in \mathcal{V}^h[s]$ be a control of the second player. Consider a partition $\Delta = \{t_k\}_{k=1}^m$ of the time interval $[s, T]$. If $x_0, x_1, \ldots, x_m$ are vectors, $x_0 = y$, then denote

by $\hat{p}_r^h(x_1, \dots, x_r, \Delta)$ the probability of the event $\mathcal{X}_1^h[t_k, s, y, \hat{u}, \Delta, v] = x_k$ for $k = \overline{1, r}$. Define vectors $w_0, \dots, w_m$ recursively in the following way. Put

$$w_0 \triangleq \hat{\chi}_1(s, y) = y, \tag{25}$$

for $k > 0$ put

$$w_k \triangleq \hat{\psi}_1(t_k, t_{k-1}, x_{k-1}, w_{k-1}). \tag{26}$$

If $w_0, \dots, w_m$ are defined by rules (25), (26) and $r \in \overline{1, m}$, we set

$$g_r(x_0, \dots, x_{r-1}, \Delta) \triangleq (w_0, \dots, w_r).$$

In addition, put $g_0(\Delta) \triangleq y$.

Below we use the transformation $G(\cdot, \mathcal{X}_1^h[\cdot, s, y, \hat{u}, \Delta, v])$ of the stochastic process $\mathcal{X}_1^h[\cdot, s, y, \hat{u}, \Delta, v]$ defined in the following way. If x_i are values of $\mathcal{X}_1^h[t_i, s, y, \hat{u}, \Delta, v]$, $i = 0, \dots, r$, and $(w_0, \dots, w_r) = g_r(x_0, \dots, x_{r-1}, \Delta)$, then we put

$$G(t_r, \mathcal{X}_1^h[\cdot, s, y, \hat{u}, \Delta, v]) \triangleq w_r.$$

Generally, the stochastic process $G(\cdot, \mathcal{X}_1^h[\cdot, s, y, \hat{u}, \Delta, v])$ is non-Markov.

Further, if $u_i = \hat{u}(t_i, x_i, w_i)$, $i = 0, \dots, r$, and

$$(w_0, \dots, w_r) = g_r(x_0, \dots, x_{r-1}, \Delta),$$

we write $\varsigma_r(x_0, \dots, x_r, \Delta) \triangleq u_r$.

We have that for any $r \in \overline{1, m}$

$$\begin{aligned}
&\mathbb{E}_{sy}^h \left(\|\mathcal{X}_1^h[t_r, s, y, \hat{u}, \Delta, v] - G(t_r, \mathcal{X}_1^h[\cdot, s, y, \hat{u}, \Delta, v])\|^2 \right) \\
&= \sum_{x_1, \dots, x_r} \|x_r - g_r(x_0, \dots, x_{r-1}, \Delta)\|^2 \hat{p}_r(x_0, \dots, x_r, \Delta) \\
&= \sum_{x_1, \dots, x_{r-1}} \hat{p}_{r-1}(x_0, \dots, x_{r-1}, \Delta) \cdot \sum_{x_r} \|x_r - g_r(x_0, \dots, x_{r-1}, \Delta)\|^2 \\
&\quad \times P^h_{t_{r-1} x_{r-1}}(X(t_r, t_{r-1}, x_{r-1}, \varsigma_{r-1}(x_0, \dots, x_{r-1}), v) = x_r).
\end{aligned} \tag{27}$$

By Lemma 2, we have that

$$\begin{aligned}
&\sum_{x_r} \|x_r - g_r(x_1, \dots, x_{r-1}, \Delta))\|^2 \\
&\qquad \times P^h_{t_{r-1} x_{r-1}}(X(t_r, t_{r-1}, x_{r-1}, \varsigma_{r-1}(x_0, \dots, x_{r-1}), v) = x_r) \\
&\quad \le (1 + \beta(t_r - t_{r-1}))\|x_{r-1} - g_{r-1}(x_0, \dots, x_{r-2}, \Delta)\|^2 \\
&\qquad + Ch \times (t_r - t_{r-1}) + \varkappa^h(t_r - t_{r-1}) \times (t_r - t_{r-1}).
\end{aligned}$$

From this and (27), it follows that

$$\begin{aligned}
&\mathbb{E}_{sy}^h \left(\|\mathcal{X}_1^h[t_r, s, y, \hat{u}, \Delta, v] - G(t_r, \mathcal{X}_1^h[\cdot, s, y, \hat{u}, \Delta, v])\|^2 \right) \\
&\quad \le (1 + \beta(t_r - t_{r-1}))\mathbb{E}_{sy}^h(\|x_{r-1} - g_{r-1}(x_0, \dots, x_{r-2})\|^2) \\
&\qquad + Ch \times (t_r - t_{r-1}) + \varkappa^h(t_r - t_{r-1}) \times (t_r - t_{r-1}).
\end{aligned} \tag{28}$$

Applying this inequality recursively, we get

$$\mathbb{E}^h_{sy}\left(\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|^2\right)$$
$$\le \exp(\beta(T-s))\mathbb{E}^h_{sy}(\|x_0-g_0(\Delta)\|^2)$$
$$+Ch\times(T-s)+\varkappa^h(d(\Delta))\times(T-s).$$

Taking into account the equality $x_0=y=g_0(\Delta)$, we conclude that

$$\mathbb{E}^h_{sy}\left(\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|^2\right)\le\epsilon(h,d(\Delta)). \tag{29}$$

Here we denote

$$\epsilon(h,\delta)\triangleq Dh+T\varkappa^h(\delta),\quad D\triangleq CT.$$

Note that for any h $\epsilon(h,\delta)\to Dh$, as $\delta\to 0$. Using (29) and Jensen's inequality, we get

$$\mathbb{E}^h_{sy}\left(\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|\right)\le\sqrt{\epsilon(d(\Delta),h)}. \tag{30}$$

By construction of control with guide strategy $\hat{u}$, the following inequalities hold true

$$\varphi(s,y)=\varphi(t_0,g_0(\Delta))\ge\varphi(t_1,g_1(x_0,\Delta))\ge\dots$$
$$\ge\varphi(t_m,g_m(x_0,\dots,x_{m0-1},\Delta))=\sigma(g_m(x_0,\dots,x_{m0-1},\Delta)).$$

Hence,

$$\sigma(G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v]))\le\varphi(s,y). \tag{31}$$

Since σ is Lipschitz continuous with the constant R, we have that for any partition Δ and the second player's control v

$$\sigma(\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v])$$
$$\le\varphi(s,y)+R\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|.$$

This and (30) yield the inequality

$$\mathbb{E}^h_{sy}\sigma(\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v])\le\varphi(s,y)+R\sqrt{\epsilon(d(\Delta),h)}.$$

Passing to the limit as $d(\Delta)\to 0$ and taking into account the property $\epsilon(\delta,h)\to Dh$, as $\delta\to 0$, we obtain the first statement of the theorem.

Now let us prove the second statement of the theorem. Using Markov inequality and (29), we get

$$P\left(\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|\ge[\epsilon(h,d(\Delta))]^{1/3}\right)$$
$$=P\left(\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|^2\ge[\epsilon(h,d(\Delta))]^{2/3}\right)$$
$$\le\frac{\mathbb{E}^h_{sy}(\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|^2)}{[\epsilon(h,d(\Delta))]^{2/3}}$$
$$\le\sqrt[3]{\epsilon(h,d(\Delta))}.$$

Lipschitz continuity of the function σ and (31) yield the following inclusion

$$\{\sigma(\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v])\ge\varphi(s,y)+R[\epsilon(h,d(\Delta))]^{1/3}\}\subset$$
$$\left\{\|\mathcal{X}^h_1[T,s,y,\hat{u},\Delta,v]-G(T,\mathcal{X}^h_1[\cdot,s,y,\hat{u},\Delta,v])\|\ge[\epsilon(h,d(\Delta))]^{1/3}\right\}.$$

Finally, for any partition Δ and any second player's control $v \in \mathcal{V}^h[s]$, we have that

$$P\{\sigma(\mathcal{X}_1^h[T, s, y, \hat{u}, \Delta, v]) \geq \varphi(s, y) + R[\epsilon(h, d(\Delta))]^{1/3}\} \leq [\epsilon(h, d(\Delta))]^{1/3}.$$

From this, the second statement of the theorem follows. □

To prove Corollary 1, it suffices to replace the payoff function with $-\sigma$ and interchange the players.

7 Example

In this section we illustrate the theory developed above with a simulation study. Let $d = 2$,

$$Q(t, x, u, v) = \begin{pmatrix} -u & u \\ v & -v \end{pmatrix},$$

$u, v \in [0, 1]$, $\sigma(x_1, x_2) = \frac{1}{2}|x_1 - x_2|$. In this case the generator $L_t^h[u, v]$ (see (3)) takes the form

$$\begin{aligned} &L_t^h[u, v]f(x_1, x_2) \\ &\quad = \frac{1}{h}[u(f(x_1 - h, x_2 + h) - f(x_1, x_2)) + v(f(x_1 + h, x_2 - h) - f(x_1, x_2))], \end{aligned} \tag{32}$$

whereas the limiting dynamics (7) takes the form

$$\begin{cases} \dot{x}_1 = -x_1 u + x_2 v, \\ \dot{x}_2 = \ x_1 u - x_2 v. \end{cases} \tag{33}$$

Using equality $x_1 + x_2 = 1$, we can reduce the system (33) to the ODE

$$\dot{x}_2 = (1 - x_2)u - x_2 v. \tag{34}$$

The corresponding Hamiltonian for $x_2 \in [0, 1]$ is equal to

$$H(t, x_2, \xi) = \begin{cases} 0, & \xi \geq 0, \\ (1 - 2x_2)\xi, & \xi \leq 0. \end{cases} \tag{35}$$

Further, we replace the payoff function σ with the equivalent payoff function $\tilde{\sigma} = |1/2 - x_2|$. The solution of the Hamilton–Jacobi PDE (8) with the Hamiltonian (35) and boundary condition given by the function $\tilde{\sigma}$ is

$$\varphi(t, x_2) = \begin{cases} x_2 - 1/2 & x_2 \geq 1/2, \\ (1/2 - x_2)e^{-2(T-t)}, & x_2 \leq 1/2. \end{cases}$$

We have that the controls u_* and v_* satisfying condition (9) and (10) are

$$u_* = v_* = \begin{cases} 0, & x_2 \geq w, \\ 1, & x_2 < w, \end{cases}$$

Let $t_*, t_+ \in [0, T]$, $x_* \in [0, 1]$, $v_* \in [0, 1]$. If $x_* < 1/2$, then put

$$\tilde{\zeta}_1(t, t_+, t_*, x_*, v_*) = \frac{1}{1 + v_*} + \left(x_* - \frac{1}{1 + v_*}\right) e^{-(1+v_*)(t-t_0)},$$

if $x_* \geq 1/2$ then put

$$\tilde{\zeta}_1(t, t_+, t_*, x_*, v_*) = 1/2 + (x_* - 1/2)e^{-2v_*(t-t_0)}.$$

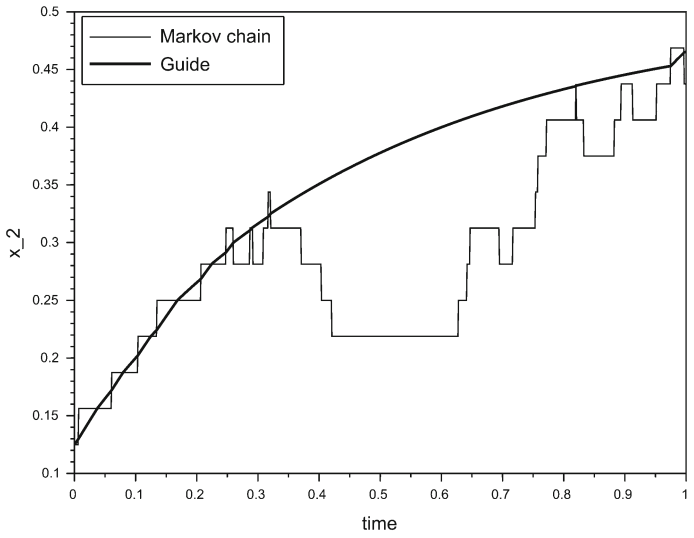

Fig. 1 Sample path of the Markov chain with $N = 32$

Note that $\tilde{\zeta}_1(\cdot, t_+, t_*, x_*, v_*)$ is a solution of the differential inclusion

$$\dot{y} \in \mathrm{co}\{(1-y)u - yv_* : u \in [0,1]\}.$$

Moreover,

$$\tilde{\zeta}_1(t_*, t_+, t_*, x_*, v_*) = x_*,$$

and $\varphi(t_*, x_*) \le \varphi(t_+, \tilde{\zeta}_1(t_+, t_+, t_*, x_*, v_*))$. The function $\tilde{\zeta}_1$ is a second coordinate of the function ζ defined in Sect. 3. Thus, the control with guide strategy determined by (u1)–(u3) takes the form

$$\hat{u}(t, x_2, w) = \begin{cases} 0, & x_2 \ge w, \\ 1, & x_2 < w, \end{cases}$$

$\hat{\psi}_1(t_+, t, x_2, w) = \tilde{\zeta}(t_+, t_+, t, w, v_*)$ where

$$v_* = \begin{cases} 0, & x_2 \ge w, \\ 1, & x_2 < w, \end{cases}$$

$\hat{\chi}_1(s, y_2) = y_2$.

To simulate the Markov chain with the generator (32), we consider the discrete-time Markov chain defined for $t_k = k\Delta t$ with the transition probabilities

$$p^h_{\Delta t}(t_k, y_1, y_2, t_{k+1}, x_1, x_2, u, v) = \begin{cases} u\Delta t, & x_1 = y_1 - h, x_2 = y_2 + h, \\ v\Delta t, & x_1 = y_1 + h, x_2 = y_2 - h, \\ 1 - (u+v)\Delta t, & x_1 = y_y, x_2 = y_2, \\ 0, & \text{otherwise.} \end{cases}$$

The state of the guide was computed analytically. The results of the simulation for $N = 32$, $t_0 = 0$, $x_0 = (7/8, 1/8)$, $T = 1$, $\Delta t = 10^{-3}$,

$$v(t, x_1, x_2) = \begin{cases} 0, & x_2 \le 1/2, \\ 1, & x_2 > 1/2 \end{cases}$$

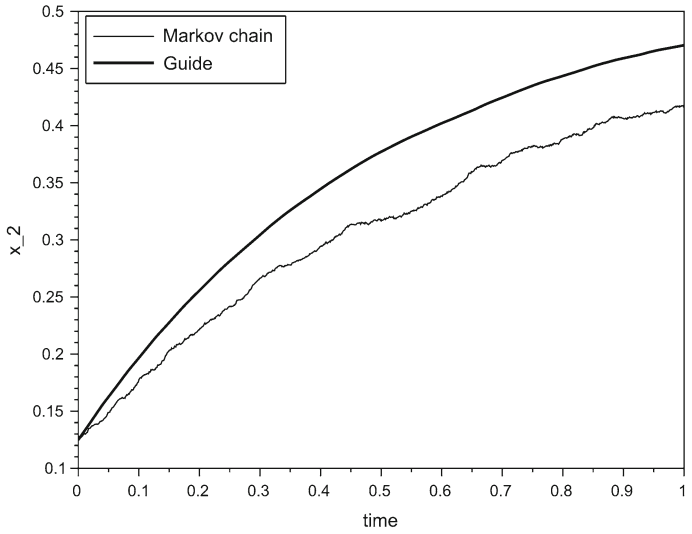

Fig. 2 Markov chain with $N = 32$ averaged over 100 simulations

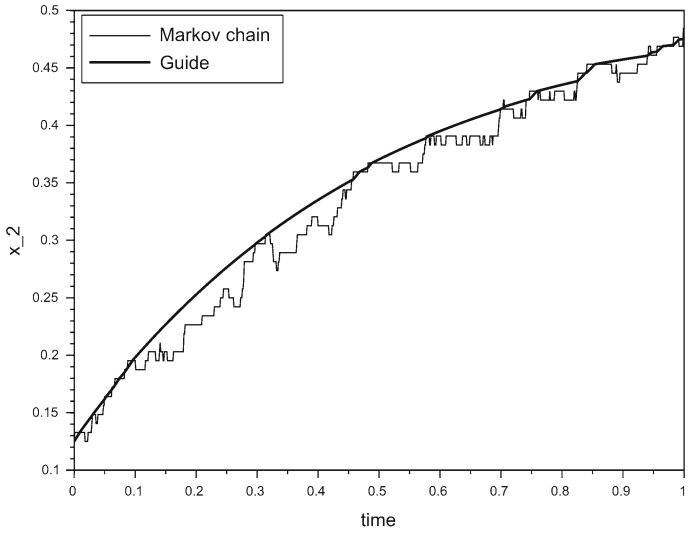

Fig. 3 Sample path of the Markov chain with $N = 128$

and 100 trials are presented in Figs. 1 and 2.

Note that the value of the game is 0.0017119, whereas the average distance between the state of the Markov chain and 1/2 is 0.08375, and the average distance between state of the guide and 1/2 is 0.0295650. Moreover the standard deviation between state of the Markov chain and the state of the guide is 0.0074127.

The results of the simulation of Markov chain with $N = 128$ particles are presented in Figs. 3 and 4. In this case the average distance between the state of the Markov chain and 1/2 is 0.0678906, the average distance between state of the guide and 1/2 is 0.0379969, the standard deviation between state of the Markov chain and the state of the guide is 0.0017119.

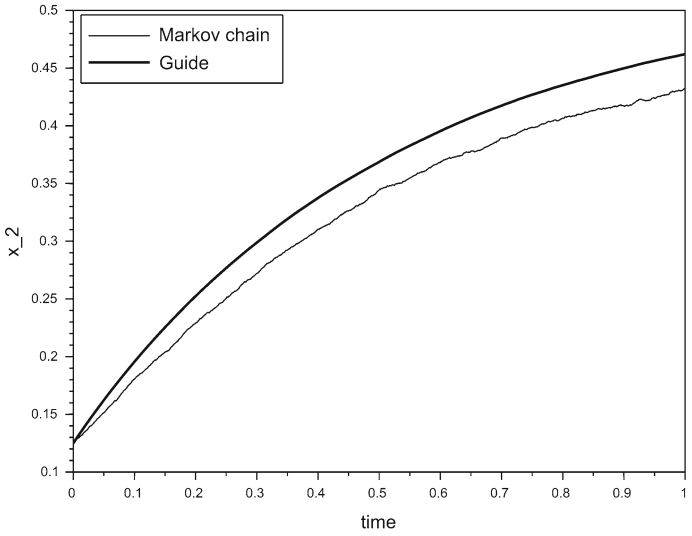

Fig. 4 Markov chain with $N = 128$ averaged over 100 simulations

Note that the ratio of the standard deviations is less than ratio of the number of particles. This result corresponds to estimate (29).

8 Conclusion

A continuous-time Markov game describing an interacting particle system is considered. A near-optimal guaranteeing strategy in the above Markov game is designed on the basis of an auxiliary deterministic game derived from the Markov one. The extreme shift rule proposed by Krasovskii and Subbotin is used for the design. The above strategy uses the model of the Markov game defined by deterministic relations called the guide. However, it receives state estimates of the Markov game to the input. Thus, the strategy used in the paper is a stochastic and memory strategy. The question whether the auxiliary game has purely feedback strategy which is near optimal in the Markov game remains open.

The study of the paper is restricted to Markov games describing interacting particle systems. The extension of the results obtained to general Markov games is the subject of future work.

Acknowledgements The author would like to thank Vassili Kolokoltsov for insightful discussions and the anonymous reviewer for the valuable comments. The research was supported by Russian Foundation for Basic Research (Project N15-01-07909).

References

1. Averboukh Y (2015) Universal Nash equilibrium strategies for differential games. J Dyn Control Syst 21:329–350
2. Bardi M, Capuzzo Dolcetta I (1997) Optimal control and viscosity solutions of Hamilton–Jacobi–Bellman equations. Birkhäuser, Basel

3. Benaïm M, Le Boudec J-Y (2008) A class of mean field interaction models for computer and communication systems. Perform Eval 65:823–838
4. Darling RWR, Norris JR (2010) Differential equation approximations for Markov chains. Probab Surv 5:37–79
5. Fleming WH, Soner HM (2006) Controlled Markov processes and viscosity solutions. Springer, New York
6. Gast N, Gaujal B, Le Boudec J-Y (2010) Mean field for Markov decision processes: from discrete to continuous optimization. INRIA report no. 7239
7. Kolokoltsov VN (2010) Nonlinear Markov process and kinetic equations. Cambridge University Press, Cambridge
8. Kolokoltsov VN (2013) Nonlinear Markov games on a finite state space (mean-field and binary interactions). Int J Stat Probab 1:77–91
9. Krasovskii NN, Kotelnikova AN (2009) Unification of differential games, generalized solutions of the Hamilton–Jacobi equations, and a stochastic guide. Differ Equ 45:1653–1668
10. Krasovskii NN, Kotelnikova AN (2010) An approach-evasion differential game: stochastic guide. Proc Steklov Inst Math 269:S191–S213
11. Krasovskii NN, Kotelnikova AN (2010) On a differential interception game. Proc Steklov Inst Math 268:161–206
12. Krasovskii NN, Kotelnikova AN (2012) Stochastic guide for a time-delay object in a positional differential game. Proc Steklov Inst Math 277:S145–S151
13. Krasovskii NN, Subbotin AI (1988) Game-theoretical control problems. Springer, New York
14. Kriazhimskii AV (1978) On stable position control in differential games. J Appl Math Mech 42(6):1055–1060
15. Levy Y (2013) Continuous-time stochastic games of fixed duration. Dyn Games Appl 3:279–312
16. Lukoyanov NYu, Plaksin AR (2013) Finite-dimensional modeling guides in time-delay systems. Trudy Inst Mat i Mekh UrO RAN 19:182–195 (in Russian)
17. Neyman A (2012) Continuous-time stochastic games. DP #616. Center for the Study of Rationality, Hebrew University, Jerusalem
18. Subbotin AI (1995) Generalized solutions of first-order PDEs. The dynamical perspective. Birkhaüser, Boston
19. Subbotin AI, Chentsov AG (1981) Optimization of guarantee in control problems. Nauka, Moscow (in Russian)
20. Zachrisson LE (1964) Markov games. In: Dresher M, Shapley LS, Tucker AW (eds) Advances in game theory. Princeton University Press, Princeton, pp 211–253