

# Differential Game with Discrete Stopping Time

D. V. Khlopin<sup>1\*</sup>

<sup>1</sup>*Krasovskii Institute of Mathematics and Mechanics, Ural Branch,  
Russian Academy of Sciences, Yekaterinburg, 620108 Russia  
e-mail: \*khlopin@imm.uran.ru*

Received June 30, 2021; revised October 27, 2021; accepted December 10, 2021

**Abstract**—In this paper, we consider a zero-sum differential game that can be stopped by any of the participants at one of time points known in advance. The cost functional may depend both on the time when the game ends and the position of the system at that time and on the player who initiates the game stopping. When optimizing the expectation of this functional, each player, based on the information the player has about the realized trajectory of the system, makes decisions both about his/her probability of stopping the game and about his/her own control of the conflict-controlled system; however, nondeterministic distribution rules are also allowed. The existence of the game value is shown under the assumption of a saddle point condition in the small game (the Isaacs condition). The construction of a stochastic guide whose model is a continuous-time Markov chain was used to construct approximately optimal strategies for the players.

*Keywords:* zero-sum two-person game, Dynkin's stopping game, differential game, stochastic guide, extremal shift, continuous-time Markov chain

**DOI:** 10.1134/S0005117922040105

Game settings in which each player (or one of the players) by his/her actions can choose a convenient time for stopping the entire game are due to numerous applications, primarily in economics; see, e.g., [9, 15, 17]. The game setting in which the player (players) controls the termination of the process is usually associated with the fundamental paper by Dynkin [2]. In such games, the dynamics is usually assumed to be nondeterministic and can be given by a stochastic differential equation [13, 14], a Lévy–Feller process [12], or a Markov chain with continuous [20] or discrete [7] time.

In the present paper, the dynamics of the game is given by a deterministic conflict-controlled system, and the distribution of the stopping time is not considered to be known to the players in advance as, for example, in [6, 19] and is not considered an absolutely continuous random variable, albeit depending on the players' actions as in [8]. Each player can try to stop the game only at some discrete times by assigning a conditional stopping probability (from subsets of the compact set  $[0; 1]$  known in advance). The sets of time points at which players can initiate the end of the game do not overlap and are also known in advance. The case in which the game will not be stopped is not excluded (see condition (1.4)). The cost functional is a random variable and depends both on the time when the game ends and the state of the system at that time as well as on the player who initiated the early stopping of the game; the task of the players is to optimize the expected value of this random variable. To this end, each player, based on the information he/she has about the realized trajectory of the system, makes decisions both about the conditional probability of the end of the game and about his/her own control of the conflict-controlled system. Here the strategies are actually assumed to be random processes that depend in a nonanticipating way on the

realized trajectory of the system. In particular, each player has the right to use both classical deterministic strategies—piecewise continuous programming strategies or piecewise constant positional strategies—and their arbitrary distributions, including strategies with a strategic guide.

The control scheme with a guide was proposed by Krasovskii and Subbotin (see [4]) as a control method resistant to small information noise. A little later, its stochastic version was proposed in [3]. In [5], the stochastic guide was used in the framework of the vanishing slackness method for the Hamilton–Jacobi equations. In [10, 11], a stochastic guide was constructed based on a continuous-time Markov game. In [8], a version of such a stochastic guide was used to construct approximately optimal strategies in a differential game on a finite interval the early stopping of which was a Poisson process with intensity determined by the players.

Assuming the saddle point condition in the small game and the insignificance of infinity (see (1.4)), we show the existence of a game value and propose a procedure based on a stochastic guide that realizes the game value with arbitrary accuracy. The corresponding procedure does not require any knowledge (accumulation of knowledge) about the opponent’s actions or memory about the entire realized trajectory. As part of this procedure, the player changes his/her control  $\bar{u}$  at some times  $t_k$  subject to a Poisson process using the rule

$$\bar{u}(t) \equiv u^{\text{to } z(t_k)}(t_k, y(t_k)) \quad \forall t \in [t_k; t_{k+1});$$

here  $y(t_k)$  and  $z(t_k)$  are the positions of the original game and the model game realized at that time, and  $u^{\text{to } z(t_k)}$  is the control that maximally shifts the position of the original system to the point  $z(t_k)$ . The process  $z(\cdot)$  will be given by some conflict-controlled continuous-time Markov chain in which the fictitious opponent player moves the position of the chain  $z(t)$  as far as possible from the latest available observation of the position of the original system, position  $y(t_k)$ , and the fictitious ally player plays optimally. Note that his/her optimal strategy can be constructed exactly by solving a finite system of equations (3.2).

The article is organized as follows. In the first section, we formalize the original differential game and state the main result—the existence of a saddle point and approximately optimal strategies realizing it for each player. In the second section, the necessary constants are selected for each accuracy parameter and a set of states is specified for the model based on a Markov game; in the same section, we state temporary assumptions for the times and probabilities of the end of the game that are convenient for proofs. The third section describes a continuous-time Markov game and its optimal strategies and proves a number of estimates for the divergence of trajectories and values of the functional in this game. In the fourth section, a stochastic guide is constructed and the existence of a game value is proved under the temporary assumptions. In the last section, this result is extended to the general case.

## 1. ORIGINAL DIFFERENTIAL GAME

Consider the following conflict-controlled system operating in  $\mathbb{R}^d$ :

$$\begin{aligned} \frac{d}{dt}y(t) &= f(y(t), u(t), v(t)), & y(0) &= x_* \in \mathbb{R}^d, \\ t &\geq 0, & u(t) &\in \mathbb{U}, & v(t) &\in \mathbb{V}. \end{aligned} \tag{1.1}$$

In what follows, we assume that  $\mathbb{U}$  and  $\mathbb{V}$  are metric compact sets and the function  $f: \mathbb{R}^d \times \mathbb{U} \times \mathbb{V} \rightarrow \mathbb{R}^d$  is  $L$ -Lipschitz in  $x$  for some constant  $L > 1$  and is bounded by the same constant. We will also

require that the saddle point condition in the small game (the Isaacs condition) be satisfied: for all  $x, w \in \mathbb{R}^d$  one has

$$\min_{u \in \mathbb{U}} \max_{v \in \mathbb{V}} \langle w, f(x, u, v) \rangle = \max_{v \in \mathbb{V}} \min_{u \in \mathbb{U}} \langle w, f(x, u, v) \rangle. \tag{1.2}$$

In the setting we are considering, the game can end on the initiative of any of the players, and the end time  $\theta_{\min}$  of the game is nondeterministic and determined by the following procedure. Players have infinitely increasing sets of times  $(t_k^I)_{k \in \mathbb{N}}$ ,  $(t_k^{II})_{k \in \mathbb{N}}$  and sequences of intervals  $[\phi_k^-; \phi_k^+]$  and  $[\psi_k^-; \psi_k^+]$  nested in  $[0; 1]$ . We will assume that the sets of time points are disjoint. Each of the players at their own times ( $t_k^I$  or  $t_k^{II}$ , respectively), if the game has not yet been stopped, assigns an element of the corresponding interval,  $\phi_k$  or  $\psi_k$ , respectively, as the conditional probability of stopping the game at this time on his/her initiative. In view of the above, the time  $\theta_{\min}$  of the end of the game has the distribution

$$\begin{aligned} \mathbb{P}(\theta_{\min} = t_k^I) &= \phi_k \cdot \prod_{i < k} (1 - \phi_i) \cdot \prod_{t_k^{II} < t_k^I} (1 - \psi_i), \\ \mathbb{P}(\theta_{\min} = t_k^{II}) &= \psi_k \cdot \prod_{i < k} (1 - \psi_i) \cdot \prod_{t_k^I < t_k^{II}} (1 - \phi_i). \end{aligned}$$

Thus, except for  $\phi_i$  and  $\psi_i$  the players do not have any other instruments of influence upon the time of the termination of the game. In particular, for the same  $\phi_i$  and  $\psi_i$ , any probabilities of events about the time of the termination of the game and about its initiating player do not depend on the realized trajectory; for example, if any  $\phi_k^-$  or  $\psi_k^-$  are equal to one, then the game always ends at this time; if  $\phi_k^+$  or  $\psi_k^+$  are zero, then the game cannot end at this time.

Let us also denote by  $\theta_1$  a random element equal to the game end time if it were initiated by the first player and equal to  $+\infty$  otherwise; similarly, by  $\theta_2$  we denote the game end time if it were initiated by the second player and assume  $\theta_2$  to be equal to  $+\infty$  otherwise. Now we also have  $\theta_{\min} = \min(\theta_1, \theta_2)$  and

$$\begin{aligned} \mathbb{P}(\theta_1 \geq t_k^I) &= \prod_{i < k} (1 - \phi_i), \\ \mathbb{P}(\theta_2 \geq t_k^{II}) &= \prod_{i < k} (1 - \psi_i). \end{aligned} \tag{1.3}$$

### 1.1. Strategies

We will assume that, in addition to public information about the dynamics of system (1.1), the initial position of the game, the possible times  $\theta_I^k$  and  $\theta_{II}^k$  of the end of the game, and conditional probabilities of the game stopping at these times, the players also have information about the realized trajectory of system (1.1). They have the right to submit both their control to system (1.1) and the conditional probabilities of the game stopping, which depend on the realized trajectory in a nonanticipatory way. Finally, we will allow the corresponding rules to be nondeterministic.

Recall that càdlàg functions are functions continuous on the right and having limit on the left; the set of such functions from the set  $A$  into the set  $B$  will be called Skorokhod's space (see [1]).

By  $\mathfrak{D}^I$  we denote the set of all measurable mappings  $U$  of  $D(\mathbb{R}_+, \mathbb{R}^d)$  into the space of probability measures over  $D(\mathbb{R}_+, \mathbb{U})$  such that for any time  $t > 0$  there exists a time  $t' > t$  such that for all  $x, x' \in D(\mathbb{R}_+, \mathbb{R}^d)$  the equality of the restrictions  $x|_{[0;t]} = x'|_{[0;t]}$  implies the equality of the

induced probabilities  $U(x)$  and  $U(x')$  on  $D([0; t'], \mathbb{U})$ . In a similar way, we define the set  $\mathfrak{D}^{II}$ . The elements of the sets  $\mathfrak{D}^I$  and  $\mathfrak{D}^{II}$  will be called admissible strategies.

Let us discuss what admissible strategies might be, for example, for the first player. These may turn out to be constant controls (elements of  $\mathbb{U}$ ); such controls do not depend on time or on the realized trajectory. We can consider programming controls (controls from  $B(\mathbb{R}_+, \mathbb{U})$  continuous on the right and having a limit on the left); such controls depend only on time but in no way depend on the already realized trajectory  $y$ . Within the framework of the classical formalization of the theory of differential games, positional controls of the following form are usually considered:  $U_t = u(t_i, y(t_i))$  for each  $t$  in the interval  $[t_i; t_{i+1})$ ; moreover, both the mapping  $u : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{U}$  and the infinitely increasing sequence of times  $t_i$  can be chosen arbitrarily. Note that positional strategies already use information about the realized trajectory; moreover, for each time  $t \in [t_i; t_{i+1})$ , the control is known at least up to the time  $t_{i+1}$  and positional strategies are thus also valid. One can use the control with a guide: the motion of the model  $Z$  is constructed in a nonpredictive way along the real-time trajectory, after which the control is selected by the rule  $U_t = u(t_i, Z(t_i))$  for all  $t \in [t_i; t_{i+1})$ ; the resulting strategy is also admissible. All of the strategies listed above were deterministic, but if at each time  $t_i$  we choose not just one element from  $\mathbb{U}$ , but some distribution of them, then the corresponding rule will specify, generally speaking, a nondeterministic admissible strategy. One way to specify the appropriate distributions is to make the model motion of  $Z$  a random process consistent with some filtering containing the original one and then take a marginal distribution on  $\mathbb{R}^d$  due to this process.

For the stochastic basis, we take Skorokhod's space  $D(\mathbb{R}_+, \mathbb{R}^d)$  (see [1]) with the Borel sigma-algebra  $\mathcal{B}(D(\mathbb{R}_+, \mathbb{R}^d))$  and the canonical filtration—the flow of the sigma-algebras  $\mathcal{F}_t \triangleq \sigma(y|_{[0; t]})$ ,  $t \geq 0$ . Let us show that each pair of admissible strategies  $(U, V) \in \mathfrak{D}^I \times \mathfrak{D}^{II}$  determines the unique probability  $\mathbb{P}_{U, V}$  on this stochastic basis, thereby determining a random process  $\hat{Y}$  on the solutions of system (1.1).

Note first that at the initial time the process is given, since the position of the process  $\hat{Y}$  is known, and hence the control distributions corresponding to it at this time are also known. Then for some positive  $t'$  the induced probabilities  $U(x) \otimes V(x)$  over  $D([0; t'], \mathbb{U}) \times D([0; t'], \mathbb{V})$  are known, since each pair of elements from  $D([0; t'], \mathbb{U}) \times D([0; t'], \mathbb{U})$  uniquely recovers some solution of system (1.1), the restriction of  $P_{U, V}$  over  $D([0; t'], \mathbb{R})$  is uniquely recovered as the measure image by virtue of this mapping, and hence so is the process  $\hat{Y}$  on the interval  $[0; t']$ .

Therefore, the set of intervals containing the time 0 and contained in the semiaxis on which the process  $\hat{Y}$  is defined and uniquely determined is nonempty. If this set contains the semiaxis, then everything has been shown. Assume it does not. Then we find the largest element in our set—the interval  $\mathcal{T}$ . Owing to the continuity of the trajectories of system (1.1) and hence of  $\hat{Y}|_{\mathcal{T}}$ , we can uniquely reconstruct the trajectories of  $\hat{Y}$  on the closure of  $\mathcal{T}$ —some interval  $[0; t]$ . Then there exists a positive time  $t''$  such that for each realization of the process  $\hat{Y}$  on  $[0; t]$ , the distribution of controls on  $[0; t'')$  is known. Knowing these controls for each such implementation, it is possible to construct the distribution of trajectories of system (1.1) up to the time  $t''$ ; thereby the process  $\hat{Y}$  is uniquely reconstructed already on the interval  $[0; t'')$  containing both  $\mathcal{T}$  and its closure  $[0; t]$ . However, this contradicts the choice of  $\mathcal{T}$  as the largest interval on which the process  $\hat{Y}$  is defined and uniquely determined. Therefore, this process is uniquely determined on the entire semiaxis.

### 1.2. Stopping Times

Apart from controls of system (1.1), the players also control the probabilities  $\phi_i, \psi_i$  of the termination of the game. To describe the possibilities of the first player admissible with such a choice, we demand that each  $\phi_k$  were an  $\mathcal{F}_{\theta_k^I} \otimes \mathcal{F}_{\theta_k^U}$ -measurable random variable (on  $D(\mathbb{R}_+, \mathbb{R}^m) \times D(\mathbb{R}_+, \mathbb{U})$ ) ranging in  $[\phi_k^-; \phi_k^+]$ , in particular, depending in a nonanticipating way on a pair (trajectory, control) in  $D(\mathbb{R}_+, \mathbb{R}^m) \times D(\mathbb{R}_+, \mathbb{U})$ .

In particular, they will turn out to be random variables with values in  $[\phi_k^-; \phi_k^+]$  that are independent of the trajectory being realized, as analogs of programming controls. Admissible sequences will also be arbitrary probabilities that satisfy the same probability rules and depend on the state of the process at the time  $t_k^I$  (or earlier). Finally, nothing prevents us from specifying the probabilities satisfying these rules in terms of the trajectories of some other random process  $Z$  adapted to the filtration containing the original one and responsible for the strategy  $U$ .

Note that rule (1.3) associates each such sequence  $\phi_k$  with a distribution (probability)  $\theta_1$  on  $\{\theta_k^I | i \in \mathbb{N}\} \cup \{+\infty\}$ . Moreover, in this case, we can assume that there exists some standard probability space  $\Omega^I$  (for example, the instance  $([0; 1], \mathcal{B}([0; 1]), \lambda)$  was used for this purpose in [8]), and for each element in  $D(\mathbb{R}_+, \mathbb{R}^m) \times D(\mathbb{R}_+, \mathbb{U})$  its value  $\theta_1$  is the distribution of some random variable given in this space and denoted by the same letter for simplicity. We will call such random variables (on  $\Omega^I$ ) depending on  $D(\mathbb{R}_+, \mathbb{R}^m) \times D(\mathbb{R}_+, \mathbb{U})$  and ranging in  $\{\theta_k^I | i \in \mathbb{N}\} \cup \{+\infty\}$  the admissible times for the first player to exit. Let us place all such random variables into a set  $\mathfrak{Q}^I$ . Similarly, for the second player we fix a standard probability space  $\Omega^{II}$  and put each sequence of  $\mathcal{F}_{\theta_k^I} \otimes \mathcal{F}_{\theta_k^V}$ -measurable random variables with values in  $[\phi_k^-; \phi_k^+]$  in correspondence with a random variable  $\theta_2$  given on  $\Omega^{II}$  whose distribution obeys (1.3). Such random variables (on  $\Omega^{II}$ ) depending on  $D(\mathbb{R}_+, \mathbb{R}^m)$  and ranging in  $\{\theta_k^I | i \in \mathbb{N}\} \cup \{+\infty\}$  will be called the admissible times for the second player to exit. These times form a set  $\mathfrak{Q}^{II}$ .

Now on the probability space  $\Omega \triangleq \Omega^I \times \Omega^{II}$  we can correctly define a random variable  $\theta_{\min} = \min(\theta_1, \theta_2)$  depending on  $D(\mathbb{R}_+, \mathbb{R}^m) \times D(\mathbb{R}_+, \mathbb{U}) \times D(\mathbb{R}_+, \mathbb{V})$ . Moreover, any pair of admissible strategies  $(U, V)$  also uniquely determines the probability  $P_{\Omega, U, V}$  on  $\Omega^I \times \Omega^{II} \times D(\mathbb{R}_+, \mathbb{R}^m)$ .

### 1.3. Player's Objectives

Now let us define the objectives of the players. Let the functions  $\sigma_1 : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow [-1; 1]$  and  $\sigma_2 : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow [-1; 1]$  be given. We will also assume, increasing  $L$  if needed, that they are  $L$ -Lipschitz in  $x, t$ . The number  $\sigma_1(\theta_1, y(\theta_1))$  will be the payment of the first player to the second if the first player initiated the end of the game at time  $\theta_1$ . The number  $\sigma_2(\theta_2, y(\theta_2))$  will become the payment of the first player to the second if the second player initiated the end of the game at time  $\theta_2$ .

Now in the case of stopping the game, the objective function is defined at the final time. Since there is, possibly zero, probability that the game will never be stopped, we require at least one of three conditions to be met:

1. The objective functions  $\sigma$  and the controlled system (1.1) are such that  $\sigma_1(t, y(t))$  and  $\sigma_2(t, y(t))$  tend to a common limit as  $t \rightarrow \infty$  with the convergence being uniform over all possible trajectories  $y$  of system (1.1). In other words, for any positive  $\varepsilon$  there exists a positive integer  $N$  such that for all  $t > N$  and all trajectories  $y$  of system (1.1) one has

$$\left| \sigma_1(t, y(t)) - \sigma_2(t, y(t)) \right| + \left| \sigma_1(N, y(N)) - \sigma_1(t, y(t)) \right| < \varepsilon.$$

2. At least one of the tuples  $(\phi_k^-)_{k \in \mathbb{N}}$  and  $(\psi_k^-)_{k \in \mathbb{N}}$  contains one.
3. At least one of the series  $\sum_{k \in \mathbb{N}} \phi_k^-$  and  $\sum_{k \in \mathbb{N}} \psi_k^-$  is diverging.

Both the second and the third conditions guarantee that with probability 1, regardless of the choice of admissible strategies  $U, V$  and terminal times  $\theta_1, \theta_2$  of players, the game ends in finite time. Moreover, by virtue of the validity of any of the three conditions above, there exists a sufficiently large positive integer  $N_\varepsilon$  such that for each  $i \in \{1, 2\}$ , for any admissible strategies  $U, V$  and terminal times  $\theta_1, \theta_2$  of the players, the end time of the game  $\theta_{\min} = \min(\theta_1, \theta_2)$  satisfies

$$\begin{aligned} & \mathbb{P}_{\Omega, U, V}(\theta_{\min} \geq N_\varepsilon) \sup_{t \geq N_\varepsilon} \mathbb{E}_{U, V} \left( \sigma_i(\theta_{\min}, y(\theta_{\min})) \mid \theta_{\min} \geq t \right) - \varepsilon \\ & < \mathbb{P}_{\Omega, U, V}(\theta_{\min} \geq N_\varepsilon) \mathbb{E}_{\Omega, U, V} (J(y, \theta_1, \theta_2) \mid \theta_{\min} \geq N_\varepsilon) \\ & < \mathbb{P}_{\Omega, U, V}(\theta_{\min} \geq N_\varepsilon) \inf_{t \geq N_\varepsilon} \mathbb{E}_{\Omega, U, V} \left( \sigma_i(\theta_{\min}, y(\theta_{\min})) \mid \theta_{\min} \geq t \right) + \varepsilon. \end{aligned} \quad (1.4)$$

Indeed, if the first condition in this formula is satisfied, then all expectations differ little from each other, the second condition nullifies the probabilities in this formula with sufficiently large  $N_\varepsilon$ , and the third condition allows at least one of the expressions

$$- \sum_{t_k^I < N_\varepsilon} \ln(1 - \phi_k^-), \quad \text{or} \quad - \sum_{t_k^I < N_\varepsilon} \ln(1 - \psi_k^-)$$

to be made arbitrarily large, thus making the probabilities in this formula tend to zero.

Now, if the game has not been stopped in finite time and the probability of such an event is positive, it is correct to write

$$\sigma_1(+\infty, y(+\infty)) = \sigma_2(+\infty, y(+\infty)),$$

and hence the payment of the first player to the second is also defined in this case,  $\sigma_1(\theta_1, y(\theta_1)) = \sigma_2(\theta_2, y(\theta_2))$ . However, then, subject to any of the three conditions above, the payment of the first player to the second is equal almost everywhere to

$$J(y, \theta_1, \theta_2) = \begin{cases} \sigma_1(\theta_1, y(\theta_1)), & \theta_1 = \theta_{\min} \\ \sigma_2(\theta_2, y(\theta_2)), & \theta_2 = \theta_{\min}. \end{cases}$$

#### 1.4. Game Value

Assume that the players have chosen some admissible strategies  $U$  and  $V$ . As has been shown above, one can unambiguously recover the distribution of the process  $\hat{Y}$  on solutions of (1.1); thus, on the set of all possible solutions of (1.1) one can unambiguously reconstruct the probability  $\mathbb{P}_{U, V}$ ; in turn, this defines the corresponding expectation  $\mathbb{E}_{U, V}$ . Now, with each choice of admissible terminal times  $\theta_1, \theta_2$  one can consider the distribution of the random variable  $J(y, \theta_1, \theta_2)$ . We assume that by their actions the players try to optimize its expectation, the number

$$\mathbb{E}_{\Omega, U, V} J(y, \theta_1, \theta_2). \quad (1.5)$$

Depending on which of the players is informationally discriminated, we obtain the following two game values:

$$\mathcal{V}^+ \triangleq \inf_{U \in \mathcal{D}^I, \theta_1 \in \Omega^I} \sup_{V \in \mathcal{D}^{II}, \theta_2 \in \Omega^{II}} \mathbb{E}_{\Omega, U, V} J(y, \theta_1, \theta_2),$$

$$\mathcal{V}^- \triangleq \sup_{V \in \mathfrak{D}^{II}, \theta_2 \in \Omega^{II}} \inf_{U \in \mathfrak{D}^I, \theta_1 \in \Omega^I} \mathbb{E}_{\Omega, U, V} J(y, \theta_1, \theta_2).$$

The main purpose of this paper is to prove the following assertion.

**Theorem 1.1.** *Under condition (1.4), one has the equality  $\mathcal{V}^- = \mathcal{V}^+$ . Moreover, for each positive  $\varepsilon$  there exist players' admissible strategies  $U^\varepsilon \in \mathfrak{D}^I$  and  $V^\varepsilon \in \mathfrak{D}^{II}$  and terminal times  $\theta_1^\varepsilon \in \Omega^I, \theta_2^\varepsilon \in \Omega^{II}$  such that*

$$-\varepsilon + \sup_{V \in \mathfrak{D}^{II}, \theta_2 \in \Omega^{II}} \mathbb{E}_{\Omega, U^\varepsilon, V} J(y, \theta_1^\varepsilon, \theta_2) \leq \mathcal{V}^+ = \mathcal{V}^- \leq \varepsilon + \inf_{U \in \mathfrak{D}^I, \theta_1 \in \Omega^I} \mathbb{E}_{\Omega, U, V^\varepsilon} J(y, \theta_1, \theta_2^\varepsilon).$$

Let us describe the scheme of proof. Since  $\mathcal{V}^- \leq \mathcal{V}^+$ , from the symmetry of the players, for each positive  $\varepsilon$  it is required to describe necessary admissible strategies  $U_\varepsilon$  only for the first player and a terminal time  $\theta_1^\varepsilon$  satisfying

$$-\varepsilon + \sup_{V \in \mathfrak{D}^{II}, \theta_2 \in \Omega^{II}} \mathbb{E}_{\Omega, U^\varepsilon, V} J(y, \theta_1^\varepsilon, \theta_2) \leq \mathcal{V}^+. \tag{1.6}$$

To this end, we first make temporary assumptions about admissible terminal times of the players convenient for the proof and introduce the necessary constants, in particular, by passing from  $t_k^I, t_k^{II}$  to  $t_n = nh$ ; then we consider a continuous-time Markov game as a model game, indicate a series of estimates for it, and construct optimal strategies in this game. On the basis of these optimal strategies, we will further describe the behavior of the guide model and, based on it, the construction of the strategy for the original game, show the mean-square convergence of the corresponding trajectories, and hence the values, on average, of the objective function. In the last section, we remove the temporary assumptions on admissible terminal times for the players.

## 2. TEMPORARY ASSUMPTIONS AND INTRODUCTION OF CONSTANTS

Let us take some positive number  $\varepsilon < \min(1/16L\pi, 6/\sqrt{d})$ , and, together with it, also some positive integer  $N_{\varepsilon/12}$  (see (1.4)). Then one can also find some subinterval  $[T_-; T_+) \subset [N_{\varepsilon/12}; +\infty)$  that does not contain elements of unboundedly increasing sequences  $(t_k^I)_{k \in \mathbb{N}}$  and  $(t_k^{II})_{k \in \mathbb{N}}$ . Since there are finitely many elements of such sequences in  $[0; T_+)$ , there exists a positive  $h < \frac{T_+ - T_-}{2}$  such that, for all positive integers  $k$  and  $l$ , the inequality  $t_k^I, t_l^{II} \leq T_+$  implies that

$$t_k^I > 2h, \quad t_l^{II} > 2h, \quad |t_k^I - t_l^{II}| > 2h, \quad t_k^I/h \notin \mathbb{Z}, \quad t_l^{II}/h \notin \mathbb{Z}.$$

In this case, decreasing  $h$  if necessary, if  $t_k^I \leq T_+$  and  $\phi_k^- > 1 - h$ , then one can also guarantee that  $\phi_k^- = 1$ . Moreover, the same property can also be ensured for pairs  $(t_k^{II}, \psi_k^+)$ .

Take the smallest element of  $h\mathbb{Z}$  greater than  $T_-$  for  $T$ . Now  $T \in [T_-; T_+) \cap h\mathbb{Z}$ , and the interval  $(T - h; T + h)$  does not contain elements of the sequences  $(t_k^I)_{k \in \mathbb{N}}$  and  $(t_k^{II})_{k \in \mathbb{N}}$ .

Now let us make the following temporary assumptions about  $T$  and  $h$  as well as the sequences  $(t_k^I)_{k \in \mathbb{N}}, (t_k^{II})_{k \in \mathbb{N}}$  and the corresponding intervals  $[\phi_k^-; \phi_k^+], [\psi_k^-; \psi_k^+]$ :

1. For all  $k \in \mathbb{N}, t_k^I \leq T$  implies  $\phi_k^+ < 1 - h$  and  $t_k^{II} \leq T$  implies  $\psi_k^+ < 1 - h$ .
2. All elements of sequences  $(t_k^I)_{k \in \mathbb{N}}$  and  $(t_k^{II})_{k \in \mathbb{N}}$  not exceeding  $T$  lie in  $h\mathbb{N}$ .

The first assumption guarantees that the probability of not stopping the game by time  $T$  is separated from zero by a constant independent of the actions of the players. The second assumption means that all possible times of the end of the game up to time  $T$  inclusive lie in some lattice.

In addition to the temporary assumptions above, taking  $h$  equal to  $h/k$  if necessary for sufficiently large  $k \in \mathbb{N}$ , we assume here and in the sequel that the following inequality is guaranteed everywhere:

$$h < \min \left( \frac{\varepsilon^2 e^{-8LT}}{288L^2 d}, \frac{\varepsilon}{12T(1 + 16^{1/3}T^{-2/3})^2}, \frac{\varepsilon}{6(L + 4)} \right);$$

since  $L > 1$  and  $\varepsilon < \min(1/16\pi L, 6/\sqrt{d}) < 192(\sqrt[3]{2} - 1)$ , we see that this also guarantees that

$$6h(L + 4) < \varepsilon, \quad h < \frac{1}{32\pi L(L + 1)} < \frac{1}{L(2L/3 + 5)}, \quad Lh\sqrt{d} < 1, \tag{2.1}$$

$$3h < 32\pi h(hL\sqrt{d} + L^2) < 1, \quad Le^{4LT}\sqrt{2d}\sqrt{h} < \varepsilon/12, \tag{2.2}$$

$$h < \sqrt[3]{32T} - \sqrt[3]{16T}, \quad hT(1 + 16^{1/3}T^{-2/3})^2 < \varepsilon/12. \tag{2.3}$$

By construction, for any positive integer  $n$  there exists at most one element of sequences  $(t_k^I)_{k \in \mathbb{N}}$  and  $(t_k^{II})_{k \in \mathbb{N}}$  on the interval  $[nh; nh + h)$ . If there exists some  $t_k^I$  on this interval, then we assign  $\Phi_n \triangleq [\phi_k^-/h; \phi_k^+/h]$ ; otherwise, we take  $\Phi_n \triangleq \{0\}$ . If this interval also contains some  $t_k^{II}$ , then we assign  $\Psi_n \triangleq [\psi_k^-; \psi_k^+]$ , otherwise, we set  $\Psi_n \triangleq \{0\}$ .

Note that, owing to the boundedness of the dynamics of system (1.1), there exists a compact set  $K_< \subset \mathbb{R}^d$  beyond which no solution  $y$  of Eq. (1.1) with the initial condition  $y(0) = x_*$  can go. Extending this compact set up to  $K_>$  if necessary, we can assume that any motion in (1.1) at any time in  $[0; T]$  is separated from the boundaries of  $K_>$  by at least a distance  $L$ .

Fix some infinitely smooth monotone nonincreasing scalar function  $a : \mathbb{R} \rightarrow [0; 1]$  so that  $a(0) = 1$ ,  $a(L) = 0$ , and its Lipschitz constant does not exceed 1. Let us also introduce

$$\hat{f}(x, u, v) \triangleq a(\text{dist}(x; K_<))f(x, u, v)$$

for all  $(x, u, v) \in \mathbb{R}^d \times \mathbb{U} \times \mathbb{V}$ . Remaining continuous everywhere, this function is equal to zero outside  $K_>$ . Now, following  $f$ , the function  $\hat{f}|_{K_> \times \mathbb{U} \times \mathbb{V}}$  is Lipschitz in  $x$  with the constant  $2L$  and its norm is also at most  $L$ .

Only now we finally take  $\mathcal{Z} \triangleq h((\mathbb{N} \cup \{0\}) \times \mathbb{Z}^{d+1}) \subset \mathbb{R}^{d+2}$  and  $\mathcal{Z}_< \triangleq \mathcal{Z} \cap (\mathbb{R} \times \mathbb{K}_> \times ((-\infty; -1] \cup [1; +\infty) \cup \{0\})) \subset \mathbb{R}^{d+2}$ .

Denote the basis in  $\mathbb{R}^{d+2}$  by  $(\pi_0, \pi_1, \dots, \pi_d, \pi_{d+1})$ . For brevity, we write the coordinates of a vector  $z \in \mathbb{R}^{d+2}$  in the form

$$(\pi_0 z, \pi_1 z, \dots, \pi_d z, \pi_{d+1} z).$$

In this case, the zero coordinate of each point will correspond to time, and the last one is intended to fix the game stopping time. Since the coordinates  $(\pi_1 z, \dots, \pi_d z)$  track the trajectory of the original system, we will assume that  $\mathbb{R}^d$  coincides with the subset

$$\{z \in \mathbb{R}^{d+2} \mid \pi_0 z = \pi_{d+1} z = 0\} \subset \mathbb{R}^{d+2};$$

consequently, we can subtract elements of  $\mathbb{R}^d$  and  $\mathbb{R}^{d+2}$  from each other to obtain an element of  $\mathbb{R}^{d+2}$ . Moreover, any element  $z \in \mathbb{R}^{d+2}$  is projected onto  $\mathbb{R}^d$  according to the rule

$$z \mapsto z - (\pi_0 z)\pi_0 - (\pi_{d+1} z)\pi_{d+1}.$$

Then we can apply the inner product and norm on  $\mathbb{R}^d$  to vectors in  $\mathbb{R}^{d+2}$ , denoting them by  $\langle \cdot, \cdot \rangle_d$  and  $\| \cdot \|_d$ , respectively.



3. MODEL GAME

For a guide we will use the trajectories of a specially selected model game—a conflict-controlled continuous-time Markov chain with the phase space  $\mathcal{Z}_<$ .

3.1. Randomized Strategies in the Markov Game

For the players' strategies we take time-invariant randomized strategies.

By a time-invariant randomized strategy  $\bar{\mu}$  of the first player we understand a pair  $(\mu, \phi_\mu)$  of mappings that take each  $w = (nh, x, s) \in \mathcal{Z}_<$  to some probability measure  $\mu[w]$  on  $\mathbb{U}$  and some number  $\phi_\mu[w] \in \Phi_n$ . By a randomized strategy  $\bar{\nu}$  of the second player we mean a pair  $(\nu, \psi_\nu)$  of mappings that take each  $w = (nh, x, s) \in \mathcal{Z}_<$  to some probability measure  $\nu[w]$  on  $\mathbb{V}$  and some number  $\psi_\nu[w] \in \Psi_n$ . Denote the set of all time-invariant strategies of the first and second players by  $\check{\mathbb{U}}_\zeta$  and  $\check{\mathbb{V}}_\zeta$ . By  $\mathbb{U}_\zeta$  and  $\mathbb{V}_\zeta$  we denote their projections—the sets of mappings  $w \mapsto \mu[w]$  and  $w \mapsto \nu[w]$ , respectively.

Define a mapping  $\check{f} : \mathcal{Z}_< \times \mathbb{U}_\zeta \times \mathbb{V}_\zeta \rightarrow \mathbb{R}^d$  by the following rule: for every pair of strategies  $(\mu, \nu) \in \mathbb{U}_\zeta \times \mathbb{V}_\zeta$  and a point  $w = (t, x, s) \in \mathcal{Z}_<$ ,

$$\check{f}(w, \mu, \nu) \triangleq 1_{\{0\}}(s) \int_{\mathbb{U}} \int_{\mathbb{V}} \hat{f}(x, u, v) \mu[w](du) \nu[w](dv).$$

For each  $i \in [1 : d]$ , we define its projections  $\pi_i^+ \check{f}$ ,  $\pi_i^- \check{f}$  by the rules

$$\begin{aligned} \pi_i^+ \check{f}(w, \mu, \nu) &\triangleq 1_{\{0\}}(s) \int_{\mathbb{U}} \int_{\mathbb{V}} \max(0, \pi_i \hat{f}(x, u, v)) \mu[w](du) \nu[w](dv), \\ \pi_i^- \check{f}(w, \mu, \nu) &\triangleq 1_{\{0\}}(s) \int_{\mathbb{U}} \int_{\mathbb{V}} \min(0, \pi_i \hat{f}(x, u, v)) \mu[w](du) \nu[w](dv). \end{aligned}$$

We will also need time-dependent strategies—mappings in  $D(\mathbb{R}_+, \check{\mathbb{U}}_\zeta)$  and  $D(\mathbb{R}_+, \check{\mathbb{V}}_\zeta)$ . Denote their sets by  $\check{\mathbb{U}}_\omega$  and  $\check{\mathbb{V}}_\omega$ , respectively. The mappings  $\check{f}$ ,  $\pi_i^+ \check{f}$ , and  $\pi_i^- \check{f}$  are continued to these sets using the same formulas.

3.2. Dynamics of the Markov Game

For all points  $w = (t, x, s) \in \mathcal{Z}_<$  and time-invariant randomized strategies  $\bar{\mu}, \bar{\nu}$  of the players, we define the Lévy measure  $\check{\eta}(w, \bar{\mu}, \bar{\nu}; \cdot)$  by setting  $\check{\eta}(w, \bar{\mu}, \bar{\nu}; A)$  for each subset  $A \subset \mathcal{Z}_<$  to be equal to

$$\begin{aligned} \frac{1}{h} \delta_{\pi_0}(A) + \frac{1}{h} \sum_{i=1}^d \left[ \pi_i^+ \check{f}(w, \mu, \nu) \delta_{h\pi_i}(A) + \pi_i^- \check{f}(w, \mu, \nu) \delta_{-h\pi_i}(A) \right] \\ + \frac{1_{\{0\}}(\pi_{d+1} w)}{h} \left[ \frac{\phi_\nu[w]}{1 - \phi_\nu[w]} \delta_{(s+1)\pi_{d+1}}(A) + \frac{\psi_\nu[w]}{1 - \psi_\nu[w]} \delta_{-(s+1)\pi_{d+1}}(A) \right]. \end{aligned}$$

Note that with fixed strategies  $\mu$  and  $\nu$  the mappings

$$\begin{aligned} w &\mapsto \check{f}(w, \mu, \nu), \\ w &\mapsto (\pi_i^+ \check{f}(w, \mu, \nu), \pi_i^- \check{f}(w, \mu, \nu)) \end{aligned}$$

inherit the  $2L$ -Lipschitz continuity in  $x$  and the boundedness of the norm by the constant  $L$  from the function  $f$ . We have also thereby shown the following inequalities: for all  $w \in \mathcal{Z}_<$ ,

$$\begin{aligned} \left\| \int_{\mathbb{R}^{d+2}} \sum_{i=1}^d \pi_i y \eta(w, \bar{\mu}, \bar{\nu}; dy) \right\| &\leq L, \\ \sum_{i=1}^d \left| \pi_i^+ \check{f}(w, \mu, \nu) + \pi_i^- \check{f}(w, \mu, \nu) \right| &\leq L\sqrt{d}. \end{aligned} \tag{3.1}$$

For all randomized strategies  $\bar{\mu}, \bar{\nu}$  for such a measure  $\check{\eta}$ , we consider the continuous-time Markov chain corresponding to the Kolmogorov matrix  $(\check{Q}_{wy}(\bar{\mu}, \bar{\nu}))_{w,y \in \mathcal{Z}_<}$  introduced by the rule

$$\left\{ \begin{array}{ll} \frac{1}{h}, & y = w + h\pi_0 \\ \frac{1}{h} \pi_i^+ \check{f}(w, \mu, \nu), & y = w + h\pi_i, \quad i \in [1 : d] \\ \frac{1}{h} \pi_i^- \check{f}(w, \mu, \nu), & y = w - h\pi_i, \quad i \in [1 : d] \\ \frac{1_{\{0\}}(s)}{h} \cdot \frac{\phi_\nu[w]}{1 - \phi_\nu[w]}, & y = w + (1 + \pi_0 w)\pi_{d+1} \\ \frac{1_{\{0\}}(s)}{h} \cdot \frac{\psi_\nu[w]}{1 - \psi_\nu[w]}, & y = w - (1 + \pi_0 w)\pi_{d+1} \\ -\frac{1}{h} - \frac{1}{h} \sum_{i=1}^d (\pi_i^+ \check{f}(w, \mu, \nu) + \pi_i^- \check{f}(w, \mu, \nu)) \\ \quad - \frac{1_{\{0\}}(s)}{h} \left( \frac{\phi_\nu[w]}{1 - \phi_\nu[w]} + \frac{\psi_\nu[w]}{1 - \psi_\nu[w]} \right), & y = w \\ 0, & \text{otherwise.} \end{array} \right.$$

*Remark 3.1.* By the construction of the Markov chain, each of its stopping times

$$\tau_k \triangleq \{t \mid \pi_0 \check{Y}(t) = kh\}$$

is finite almost everywhere; moreover, being the sum of random variables collectively independent and distributed exponentially with parameter  $1/h$ , each of them does not depend in any way on the players' actions, in particular, on the past and/or future values of  $\mu, \nu, \phi$ , and  $\psi$ .

*Remark 3.2.* Note also that, by the construction of the Markov chain, the probability of jumping along the last coordinate (in  $\pi_{d+1} \check{Y}$ ) between the stopping times  $\tau_{k-1}$  and  $\tau_k$  for intensities  $(\phi, \psi)$  coincides with the probability of ending the original game at time  $t_k = kh$  for the same intensities of the players.

### 3.3. Game Value and Optimal Strategies in the Markov Game

Since the matrices  $\check{Q}_{wy}^h$ , as well as the lengths of the jumps, are uniformly bounded, all the assumptions in [16, Remark 4.2(b)] are satisfied. Then, as shown in [16, Proposition 3.1(a)], for each pair of time-depending randomized strategies  $(\bar{\mu}, \bar{\nu}) \in \check{\mathcal{U}}_\varpi \times \check{\mathcal{V}}_\varpi$  and initial conditions  $z_0$  (at time 0) there exists a process  $(\check{Y}(t))_{t \geq 0}$  generated by them and hence also the probability  $\check{\mathbb{P}}_{\bar{\mu}, \bar{\nu}}[z_0]$  on all possible càdlàg mappings of  $\mathbb{R}_+$  into  $\mathcal{Z}_<$ .

Now we introduce the current payment of this Markov game. For every  $w = (t, x, s) \in \mathbb{R}^{d+1} \times ([-1 - T; -1] \cup \{0\} \cup [1; 1 + T])$ , we define

$$\check{\sigma}(w) = \begin{cases} \frac{1}{2}(\sigma_1(T, x) + \sigma_2(T, x)), & s = 0 \\ \sigma_1(s - 1, x), & s \geq 1 \\ \sigma_2(-s - 1, x), & s \leq -1. \end{cases}$$

Just as  $\sigma_i$ , this function is bounded by one and is  $L$ -Lipschitz. For each initial position  $z_0 \in \mathcal{Z}_<$  at time 0, the players can ensure one of the values depending on which one of them is informationally discriminated,

$$\begin{aligned} \check{V}^+(z_0) &= \inf_{\bar{\mu} \in \check{U}_\infty} \sup_{\bar{\nu} \in \check{V}_\infty} \check{\mathbb{E}}_{\bar{\mu}, \bar{\nu}}[z_0] \int_0^\infty h e^{-ht} \check{\sigma}(z(t)) dt, \\ \check{V}^-(z_0) &= \sup_{\bar{\nu} \in \check{V}_\infty} \inf_{\bar{\mu} \in \check{U}_\infty} \check{\mathbb{E}}_{\bar{\mu}, \bar{\nu}}[z_0] \int_0^\infty h e^{-ht} \check{\sigma}(z(t)) dt. \end{aligned}$$

Now, according to [16, Theorem 5.1] and [21, Theorem 2], the system of equations (see, [16, (5.4)] and [21, (11)])

$$\begin{aligned} \inf_{\bar{\mu}[w]} \sup_{\bar{\nu}[w]} \sum_{y \in \mathcal{Z}_<} \bar{Q}_{wy}(\bar{\mu}, \bar{\nu}) \check{V}(y) &= h(\check{V}(w) - e^{-h\pi_0 w} \check{\sigma}(w)) \\ &= \sup_{\bar{\nu}[w]} \inf_{\bar{\mu}[w]} \sum_{y \in \mathcal{Z}_<} \bar{Q}_{wy}(\bar{\mu}, \bar{\nu}) \check{V}(y), \quad w \in \mathcal{Z}_<, \end{aligned} \tag{3.2}$$

has a unique solution by virtue of the saddle point condition in the small game, and this solution coincides with  $\check{V} \triangleq \check{V}^- = \check{V}^+$ . Moreover, any time-invariant strategies  $\bar{\mu}^{\text{opt}}$  and  $\bar{\nu}^{\text{opt}}$  of the players that realize a saddle point in this system for all  $w \in \mathcal{Z}_<$  are optimal strategies in this problem (see, e.g., [21, (6)] and [16, (5.5)–(5.7)]). Let us choose some  $\bar{\mu}^{\text{opt}}$  and  $\bar{\nu}^{\text{opt}}$  with such a property.

### 3.4. Estimates for the Objective Functional in the Markov Game

Note first that, just as in [8, (4.4)], one can show that the inequality

$$\check{\mathbb{E}} \max_{t \in [0; T+r], t/h \in \mathbb{N}} |\pi_0 \check{Y}(t) - t|^2 \leq 4\check{\mathbb{E}} |\pi_0 \check{Y}(T+r) - T - r|^2 = 8h(T+r)$$

holds for all positive integer  $r/h$ . Now, by virtue of the Markov inequality, the probability of the maximum value of  $|\pi_0 \check{Y}(t) - t|^2$  being greater than  $r^2$  on this interval is at most  $8h(T+r)/r^2$ . Then the probability of the stopping time  $\tau_{T/h}$  corresponding to time  $T$  to be greater than  $T+r$  is at most  $8h(T+r)/r^2$ .

Consider an arbitrary trajectory  $\check{Y}$  of the Markov game. Note that by construction, beginning from the stopping time  $\tau_{T/h}$ , this trajectory changes only in the zero coordinate; in particular,  $\check{\sigma}(\check{Y}(t)) = \check{\sigma}(\check{Y}(\tau_{T/h}))$  for all  $t \geq \tau_{T/h}$ . By virtue of the boundedness of  $\check{\sigma}$ , for all trajectories  $\check{Y}$  it follows from  $\tau_{T/h} \leq T+r$  that

$$\left| \int_0^\infty h e^{-ht} \check{\sigma}(\check{Y}(t)) dt - \check{\sigma}(\check{Y}(\tau_{T/h})) \right| \leq h \int_0^{\tau_{T/h}} e^{-ht} dt \leq h(T+r),$$

and, in the case of  $\tau_{(T+r)/h} > T + r$ , this modulus is estimated from above by the number 2. Since the probability of the event  $\tau_{T/h} > T + r$ , as noted above, is at most  $8h(T + r)/r^2$ , we have shown that

$$\mathbb{E} \left| \int_0^\infty h e^{-ht} \check{\sigma}(\check{Y}(t)) dt - \check{\sigma}(\check{Y}(\tau_{T/h})) \right| \leq h(T + r)(1 + 16/r^2). \tag{3.3}$$

The right-hand side of (3.3) decreases with respect to  $r$  for  $r \leq \sqrt[3]{32T}$ . Take a positive integer in the interval  $[\sqrt[3]{16T}/h; \sqrt[3]{16T}/h + 1)$  for  $r/h$ ; by virtue of (2.3), it is ensured that  $r \leq \sqrt[3]{32T}$ . Now, estimating the right-hand side in (3.3) by

$$h(T + r)(1 + 16/r^2) \leq hT \left(1 + \sqrt[3]{16T}/T\right)^2 \stackrel{(2.3)}{\leq} \varepsilon/12,$$

we obtain an estimate independent of the actions of the players,

$$\mathbb{E} \left| \int_0^\infty h e^{-ht} \check{\sigma}(\check{Y}(t)) dt - \check{\sigma}(\check{Y}(\tau_{T/h})) \right| \leq \varepsilon/12. \tag{3.4}$$

3.5. Estimates on a Trajectory of the Markov Game

Fix some pair of randomized strategies  $\bar{\mu} = (\mu, \phi_\mu)$ ,  $\bar{\nu} = (\nu, \psi_\nu)$ ; hence the distribution  $\check{\mathbb{P}} \triangleq \check{\mathbb{P}}_{\bar{\mu}, \bar{\nu}}[z_0]$  and the Lévy measure  $\eta(z; \cdot) \triangleq \check{\eta}(z, \bar{\mu}, \bar{\nu}; \cdot)$ . Such a Lévy measure is associated with the Lévy–Khintchine generator (see, e.g., [18, (5.1)] and [22, (2.14)]) that takes each function  $g \in C_c^2(\mathcal{Z}_<)$  to a mapping  $x \mapsto \Lambda g(x)$  according to the rule

$$\check{\Lambda}g(x) = \int_{\mathbb{X}} [g(x + y) - g(x)] \eta(x; dy) \quad \forall x \in \mathcal{Z}_<. \tag{3.5}$$

In this case, one has Dynkin’s formula [22, Proposition 2.3]

$$\mathbb{E} \left( g(\check{Y}(t'')) - \int_{t'}^{t''} \Lambda g(\check{Y}(s)) ds \middle| \check{Y}(t') \right) = g(\check{Y}(t')) \tag{3.6}$$

for all  $g \in C_c^2(\mathcal{Z}_<)$  and at all times  $t'$  and  $t''$  ( $t'' \geq t'$ ). Then the following process is a martingale:

$$g(\check{Y}(t'')) - \int_{t'}^{t''} \Lambda g(\check{Y}(s)) ds.$$

Consider the behavior of the Markov chain if  $t' = 0$  and the initial state  $w' \triangleq \check{Y}(0)$  is known. Assume that  $M \triangleq \sqrt{hL\sqrt{d} + L^2}$ . Applying the function

$$w \mapsto \|w - w'\|_d^2 \triangleq \sum_{i=1}^d (\pi_i w - \pi_i w')^2 \quad \forall w \in \mathbb{R}^{d+2}$$

to the above martingale as  $g$ , we can find from the estimates (3.1) (see [8, (4.14)]) that

$$\mathbb{E} \|\check{Y}(t) - \check{Y}(0)\|_d^2 \leq hL\sqrt{d}t + \frac{4LM}{3}(e^t - 1)^{3/2}.$$

Now, since  $e^r - 1 \leq re^r$  for  $r \geq 0$ , we have  $(e^r - 1)^{3/2} \leq \sqrt{r}(e^{3r/2} - e^{r/2})$  and

$$\mathbb{E} \|\check{Y}(t) - \check{Y}(0)\|_d^2 \leq hL\sqrt{dt} + \frac{4LM\sqrt{t}}{3}(e^{3t/2} - e^{t/2}).$$

In particular, for the stopping time  $\tau = \inf\{t \geq 0 \mid \pi_0\check{Y}(t) \neq \pi_0\check{Y}(0)\}$ , by virtue of its density being exponential with indicator  $1/h$  and in view of the equality  $\int_0^\infty \sqrt{r}e^{-\nu r} dr = \sqrt{\frac{\pi}{4\nu^3}}$ , we have

$$\begin{aligned} \mathbb{E} \|\check{Y}(\tau) - \check{Y}(0)\|_d^2 &\leq hL\sqrt{d}\mathbb{E}\tau + \frac{4LM}{3} \mathbb{E} \int_0^\infty \sqrt{s}(e^{3s/2} - e^{s/2})e^{-s/h} ds \\ &= h^2L\sqrt{d} + \frac{2LM\sqrt{\pi}}{3} ((1/h - 3/2)^{-3/2} - (1/h - 1/2)^{-3/2}) \\ &\leq h^2L\sqrt{d} + LM\sqrt{\pi}(1/h - 3/2)^{-5/2} \leq h^2L(\sqrt{d} + 1); \end{aligned}$$

at the last step, we have also used the condition  $3h < 2^5M^2\pi h < 1$  (see (2.2)). In a similar way, the equality  $\int_0^\infty r^{3/2}e^{-\nu r} dr = \frac{3}{4}\sqrt{\frac{\pi}{\nu^5}}$  also implies

$$\begin{aligned} \mathbb{E} \int_0^\tau \|\check{Y}(t) - \check{Y}(0)\|_d^2 dt &\leq \int_0^\infty \int_0^s \left[ hL\sqrt{dt} + \frac{4LM\sqrt{t}}{3}(e^{3t/2} - e^{t/2}) \right] dt e^{-s/h} ds \\ &\leq hL\sqrt{d} \int_0^\infty \frac{s^2}{2} e^{-s/h} ds + \frac{4LM}{3} \int_0^\infty s^{3/2}(e^{3s/2} - e^{s/2})e^{-s/h} dt ds \\ &\leq h^3L\sqrt{d} + LM\sqrt{\pi}((1/h - 3/2)^{-5/2} - (1/h - 1/2)^{-5/2}) \\ &\leq h^3L\sqrt{d} + \frac{5LM\sqrt{\pi}}{2}(1/h - 3/2)^{-7/2} \leq h^3L(\sqrt{d} + 5). \end{aligned}$$

Since the strategy is time-invariant and the considered process is strongly Markov, the same estimates hold if we consider an arbitrary stopping time  $\tau'$  instead of the initial time, take any stopping time  $\tau'' \geq \tau'$  not exceeding

$$\inf \{t \geq 0 \mid \pi_0\check{Y}(t) \neq \pi_0\check{Y}(\tau'), t \geq \tau'\}$$

for  $\tau$ , and finally, replace the expectation by the conditional expectation at the stopping time  $\tau'$ . Thus,

$$\begin{aligned} \mathbb{E}_{\tau'} \|\check{Y}(\tau'') - \check{Y}(\tau')\|_d^2 &\leq h^2L(\sqrt{d} + 1), \\ \mathbb{E}_{\tau'} \int_{\tau'}^{\tau''} \|\check{Y}(t) - \check{Y}(\tau')\|_d^2 dt &\leq h^3L(\sqrt{d} + 5). \end{aligned} \tag{3.7}$$

#### 4. GUIDE SCHEME AND DOUBLE GAME

##### 4.1. Aiming

When constructing the scheme with a guide, we need to be able to aim in a given direction (and deviate from it). First, we introduce functions necessary for this in a deterministic system. To

this end, note that, owing to (1.2), for any vectors  $x, z \in \mathbb{R}^d$  there exist controls  $u^{\text{to } z}(x) \in \mathbb{U}$  and  $v^{\text{from } z}(x) \in \mathbb{V}$  such that

$$\begin{aligned} \min_{u \in \mathbb{U}} \max_{v \in \mathbb{V}} \langle x - z, \hat{f}(x, u, v) \rangle &= \max_{v \in \mathbb{V}} \langle x - z, \hat{f}(t, x, u^{\text{to } z}(x), v) \rangle \\ &= \max_{v \in \mathbb{V}} \min_{u \in \mathbb{U}} \langle x - z, \hat{f}(x, u, v) \rangle = \min_{u \in \mathbb{U}} \langle x - z, \hat{f}(x, u, v^{\text{from } z}(x)) \rangle. \end{aligned}$$

For convenience, we also assume that  $u^{\text{to } w'} \equiv u^{\text{to } z'}$  for all  $w' = (t, z', s) \in \mathbb{R}^{m+2}$ . Likewise, for each  $x' \in \mathbb{R}^d$  we equip the second player with a time-invariant randomized strategy  $\bar{\nu}^{\text{from } x'}$ ,

$$\mathcal{Z}_< \ni w \triangleq (t, z, s) \mapsto \bar{\nu}^{\text{from } x'}(w) \triangleq (v^{\text{from } x'}(z), \psi^{\text{opt}}[w]).$$

In this case, for all  $x, z \in \mathbb{R}^d$  and  $w = (\tau, z, 0) \in \mathbb{R}^{d+2}$ , by virtue of the  $2L$ -Lipschitz continuity of  $\hat{f}$  and  $\check{f}$ , we have

$$\begin{aligned} &\langle x - w, f(x, u^{\text{to } w}(x), v) - \check{f}(w, \mu^{\text{opt}}(w), \nu^{\text{from } x}(w)) \rangle_d \\ &\leq \max_u \langle x - w, f(x, u, v) - f(x, u, v^{\text{from } x}(z)) \rangle_d \\ &\quad + \max_{\bar{\mu}} \langle x - w, \check{f}(w, \delta_{u^{\text{to } w}(x)}, \nu^{\text{from } x}(w)) - \check{f}(w, \bar{\mu}, \nu^{\text{from } x}(w)) \rangle_d \\ &\quad + 2L\|x - w\|_d^2 \leq 2L\|x - w\|_d^2. \end{aligned} \tag{4.1}$$

Fix some  $x' \in \mathbb{R}^d$  and  $w' \in \mathbb{R}^{d+2}$ . Consider the mapping  $R_{x', w'} : \mathbb{R}^{d+2} \times \mathbb{R}^{d+2} \rightarrow \mathbb{R}$  independent of the zero and last coordinates of the arguments and given by the following rule: for all  $x \in \mathcal{Z}_<$ ,

$$\hat{R}_{x', w'}(x) = \sum_{i=1}^d (\pi_i x' - \pi_i w') (\pi_i x - \pi_i x') = \langle x' - w', x - x' \rangle_d; \tag{4.2}$$

for the argument of this function we can also use the elements of  $\mathbb{R}^d$  by virtue of the embedding  $\mathbb{R}^d \subset \mathbb{R}^{d+2}$ . By a straightforward calculation of the generator, using the  $2L$ -Lipschitz continuity of  $\check{f}$ , we conclude (see [8, (5.2)]) that in the case of  $x \in \mathbb{R}^d$  and  $w \in \mathbb{R}^{d+2}$  we have

$$\begin{aligned} \hat{\Lambda}[\bar{\mu}^{\text{opt}}, \bar{\nu}^{\text{from } x'}] R_{w', x'}(w) &= -\langle x' - w', \check{f}(w, \mu^{\text{opt}}, \nu^{\text{from } x'}) \rangle \\ &\leq -\langle x' - w', \check{f}(w', \mu^{\text{opt}}, \nu^{\text{from } x'}) \rangle_d + 2L\|x' - w'\|_d \cdot \|w - w'\|_d \\ &\leq -\langle x' - w', \check{f}(w', \mu^{\text{opt}}, \nu^{\text{from } x'}) \rangle_d + L\|x' - w'\|_d^2 + L\|w - w'\|_d^2. \end{aligned} \tag{4.3}$$

Such a representation will prove useful when substituting the generator into Dynkin's formula (3.6).

Considering now the random process in the original game, we can verify (see [8, (5.4)]) that for an arbitrary strategy of the second player  $v$ , an analog of Dynkin's formula holds for the generator

$$K_< \ni x \mapsto \hat{\Lambda}[u^{\text{to } w'}(x'), v] R_{x', w'}(x) = \langle x' - w', f(x, u^{\text{to } w'}(x'), v) \rangle_d; \tag{4.4}$$

namely, for any stopping times  $\tau'$  and  $\tau''$  ( $\tau' \leq \tau''$ ) one has

$$R_{x', w'}(y(\tau')) = \mathbb{E}_{\tau'} \left[ R_{x', w'}(y(\tau'')) - \int_{\tau'}^{\tau''} \hat{\Lambda}[u^{\text{to } w'}(x'), v] R_{x', w'}(y(t)) dt \right]. \tag{4.5}$$

In particular, applying the  $2L$ -Lipschitz continuity of  $f$  to (4.4), evaluating the product by half the sum of squares, and using (4.1), we obtain

$$\begin{aligned} & \hat{\Lambda}[\bar{u}^{to w'}(x'), \bar{v}(t)] R_{x', w'}(x) \\ & \leq \left\langle x' - w', f(x', u^{to w'}(x'), v[t]) \right\rangle_d + 2L\|x - x'\|_d \cdot \|x' - w'\|_d \\ & \stackrel{(4.1)}{\leq} \left\langle x' - w', \check{f}(w', \mu^{opt}, v^{from x'}) \right\rangle_d + 3L\|x' - w'\|_d^2 + L\|x - x'\|_d^2. \end{aligned} \tag{4.6}$$

4.2. Construction of the Guide for the First Player

Now we will construct a càdlàg random process in the phase space  $K_{<} \times \mathcal{Z}_{<}$ ; we will construct this process up to the end of the game. Here we assume that some admissible strategy  $V$  and terminal time  $\theta_2$  are chosen by the second player but are not known to the first one. For each choice of  $V$  on trajectories in the phase space  $K_{<} \times \mathcal{Z}_{<}$ , we will obtain its own probability distribution; fixing the choice of  $V$ , we also fix the distribution  $\tilde{\mathbb{P}}$  and the corresponding expectation  $\tilde{\mathbb{E}}$ . The projection  $\hat{Y}$  of this distribution with trajectories with values  $K_{<} \times \mathcal{Z}_{<}$  onto the first component will give the distribution  $\mathbb{P}_{U, V}$  of trajectories  $y$  of the original game, and indeed, as we will see later, it will be generated by some admissible strategy  $U$  of the first player. The projection  $\check{Y}$  of this distribution onto the second component will be responsible for the distribution of the guide and will be constructed by the first player based on the distributions  $\check{\mathbb{P}}$  generated by the time-invariant strategies  $(\mu^{opt}, \nu^{from y(t_k)})$ .

Thus, suppose that at predetermined times  $(t_k = kh)_{k \in \mathbb{N}}$  the first player reads the position  $y(\cdot)$  of the real game; we can also assume that  $t_0 = 0$ . Let us determine the stopping times  $\tau_k \triangleq \min\{t \mid \pi_0 \check{Y}(t) = kh\}$  for the random process  $\check{Y}$ . As noted in Remark 3.1, by the construction of the Markov chain, each of these stopping times is almost everywhere finite; moreover, being the sum of independent random variables distributed exponentially with parameter  $1/h$ , each of these times does not depend in any way on the actions of the players. Moreover, their entire sequence can be considered generated even before the start of the game. In this way, we already know the zero coordinate of the model,  $\pi_0 \check{Y}(t) \triangleq |\{i \in \mathbb{N} \cup \{0\} \mid \tau_i \leq t\}|h$ .

To generate the first to  $d$ th model coordinates up to the stopping time  $\check{\theta}_{\min}$ , we will use the projection of the same measure  $\check{\eta}$  but only onto the first  $d$  coordinates,

$$\eta(w, \bar{\mu}, \bar{\nu}; A) \triangleq \frac{1}{h} \sum_{i=1}^d [\pi_i^+ \check{f}(w, \mu, \nu) \delta_{h\pi_i} + \pi_i^- \check{f}(w, \mu, \nu) \delta_{-h\pi_i}];$$

this corresponds to the Kolmogorov matrix  $(Q_{wy}(\bar{\mu}, \bar{\nu}))_{w, y \in \mathcal{Z}_{<}}$  introduced by the rule

$$\left\{ \begin{array}{ll} \pi_i^+ \check{f}(w, \mu, \nu)/h, & y = w + h\pi_i, i \in [1 : d] \\ \pi_i^- \check{f}(w, \mu, \nu)/h, & y = w - h\pi_i, i \in [1 : d] \\ - \sum_{i=1}^d (\pi_i^+ \check{f}(w, \mu, \nu) + \pi_i^- \check{f}(w, \mu, \nu))/h, & y = w, \\ 0, & \text{otherwise.} \end{array} \right.$$

Finally, for the first player, on each interval  $[t_{k-1}; t_k)$  we describe the control procedure for the initial conflict-controlled system if the game has not yet been stopped at time  $t_k$  and he/she has already constructed the random process  $Y$  until the stopping time  $\tau_k$ .

At time  $t_0 = 0$ , the player initializes a random process  $Y$  by assigning  $\hat{Y}(0) = y(0) = x_*$  and  $\check{Y}(0) = w_0$ , where  $w_0$  is the point of the lattice  $h\mathbb{R}^{d+2}$  closest to  $(0, y(0), 0)$ , and assigns the control  $u^{to w_0}$  on the interval  $[0; h)$ . Then the first player calculates the trajectory  $\check{Y}$  of the process with the Lévy measure  $\eta$  due to the pair of strategies  $(\mu^{opt}, \nu^{from y(0)})$  until the stopping time  $\tau_1$ , after which he/she assigns  $w_1 = \check{Y}(\tau_1)$ . Since, according to the condition, it is impossible to terminate the game at the initial time, the only possible value for the first (and for the second) player for  $\phi_0$  (and for  $\psi_0$ ) is zero.

The first player does not know how the second player is controlled on the interval  $[0; h)$ , however, if the game has not yet ended, then at time  $t_1 = h$  the first player, knowing the position  $y(t_1) = \hat{Y}(t_1)$ , assigns  $u^{to w_1}$  as his/her control in the original game on the interval  $[h; 2h)$ , and for the model  $\check{Y}$ , assigns time-invariant strategies  $(\mu^{opt}, \nu^{from y(t_1)})$  for both players on the interval  $[\tau_1; \tau_2)$ . He/she will also take  $\phi_1 = \phi^{opt}(\check{Y}(\tau_1)) \in \Phi_1$  as the probability of stopping of the original game initiated by him/her.

Thus, suppose that we have constructed the control in the original game and the trajectory of the Markov process up to the time  $t_k$  and the stopping time  $\tau_k$ , respectively. Let us show how the first player is controlled until the time  $t_{k+1}$  and the stopping time  $\tau_{k+1}$  in the case where the original game has not yet ended on the interval  $[t_{k-1}; t_k)$  and the trajectory of the original game is known until the time  $t_k$  inclusive.

At time  $t_k = kh$ , knowing the positions  $y(t_k) = \hat{Y}(t_k)$  and  $w_k \triangleq \check{Y}(\tau_k)$ , the first player assigns the function  $u^{to w_k}$  as his/her control in the original game on the interval  $[t_k; t_{k+1})$  and assigns time-invariant strategies  $(\mu^{opt}, \nu^{from y(t_k)})$  for both players for the model on the interval  $[\tau_k; \tau_{k+1})$ . He/she will also take  $\phi_k = \phi^{opt}(Y(\tau_k))$ . As before, he/she will assume that only the termination of the original game during  $[t_k; t_{k+1})$  means the termination of the model game; thus, he/she still does not need information about the probability of termination initiated by the second player neither for calculating the Markov chain up to stopping time  $\tau_{k+1}$  nor for the trajectory of the original game up to time  $t_{k+1}$ . The trajectories up to time  $t_{k+1}$  and the stopping time  $\tau_{k+1}$  are constructed, and hence the induction hypothesis is proved.

Thus, we have described both the construction of the trajectory  $y(\cdot) = \hat{Y}(\cdot)$  of the original game up to the time of its end  $\theta_{min}$  and the construction of all, except for the last one, coordinates of the model  $\check{Y}$  until time  $\check{\theta}_{min}$ . Let us do this for the last coordinate, too.

No matter how the original game develops, it can be finished ahead of time only at one of the times  $t_n^I, t_n^{II}$ , and on each interval  $[t_k; t_{k+1})$  there is at most one such time. Then we can assume that the original game ended at the time  $\theta_{min}$  in  $[t_k; t_{k+1})$  exactly when the model game ends on the interval  $[\tau_k; \tau_{k+1})$ , that is, one of the events  $\check{\theta}_i \in [\tau_k; \tau_{k+1})$  will be realized. By virtue of Remark 3.2, the same probabilities of termination for each player in these games imply that the probability of this event coincides with the probability of transfer along the last coordinate in the Markov game. Then on each time interval  $[0; \tau_k)$  we have the opportunity to reconstruct the last coordinate of the model:  $\pi_{d+1}\check{Y}(t) \triangleq 0$  until the time  $\check{\theta}_{min}$  of stopping of the model game,  $\pi_{d+1}\check{Y}(t) \triangleq 1 + \check{\theta}_{min}$  for  $t \geq \check{\theta}_{min} = \check{\theta}_1$ , and  $\pi_{d+1}\check{Y}(t) \triangleq -1 - \check{\theta}_{min}$  for  $t \geq \check{\theta}_{min} = \check{\theta}_2$ . On the other hand, by prescribing the probabilities  $\phi_k = \phi^{opt}(\check{Y}(\tau_k))$  the first player has determined not only the stopping time  $\check{\theta}_1$  of the Markov game but, by virtue of the same remark, has also prescribed the terminal time  $\theta_1$  in the original game.



Thus, for each choice of strategy and the time of termination by the second player on the trajectories of the process  $(\hat{Y}, \check{Y})$  in the phase space  $K_{<} \times \mathcal{Z}_{<}$  we have constructed the probability distribution  $\tilde{\mathbb{P}}$  and the corresponding expectation  $\tilde{\mathbb{E}}$ . Disintegrating it first over  $\bar{\phi}^{\text{opt}}(\check{Y}|_{[0; \tau_{k-1}]})$  and then over  $y|_{[0; t_{k-1}]} = \hat{Y}|_{[0; t_{k-1}]}$  and  $u^{\text{toy}}|_{[0; t_{k-1}]}$ , we obtain the distribution of the first player's controls up to time  $t_{k+1}$ , and so the strategy presented in this section is admissible. In a similar way, each probability  $\phi_k$  of termination by the first (second) player is determined as a marginal probability with the already known, realized part of the trajectory and control and is also admissible.

Note that the indicated procedure for constructing an admissible strategy  $U$  and a terminal time  $\theta_1$  did not use time assumptions anywhere. Indeed, we were not interested anywhere at what exact time in the interval  $[t_k = kh; t_{k+1} = kh + h)$  the original game could end; we were only interested in whether it ended on this interval. Further, the probability  $\phi^{\text{opt}}(Y(\tau_k))$  given by us certainly falls into the desired compact set by the construction of  $\Phi_k$ , and since we have used only the information available at time  $t_k$ , which means that it is certainly present at any time in the interval  $[t_k; t_{k+1})$ , the resulting stopping time  $\theta_1$  is admissible. Finally, this procedure used the information about whether the game stopped only at the end of each interval  $[t_k; t_{k+1})$ . In particular, it did not prohibit the second player to decide on his/her next probability to end the game during the interval  $[t_k; t_{k+1})$  only at the time  $t_{k+1}$ . Thus, this procedure allows not only all admissible stopping times for the second player but also all stopping times  $\theta_2$  of the second player for which, with all  $n \in \mathbb{N}$  and  $s \in [0; 1)$ , the probabilities of the events  $\{\theta_2 < nh + sh\}$  are only  $\mathcal{F}_{nh}$ -measurable. Let us fix a stopping time  $\theta_2$  with this property.

4.3. Divergence of Trajectories of Original and Markov Games

Let us introduce stopping times  $\theta_{\min} = \min(\theta_1, \theta_2)$  and  $\check{\theta}_{\min} = \min(\check{\theta}_1, \check{\theta}_2)$  in the original and Markov games, respectively.

Fix some  $k \in \mathbb{N}$  and, together with it, times  $t_{k-1} = (k - 1)h, t_k = kh \leq T$  from the partition. Now we can also determine the stopping times

$$\begin{aligned} t' &\triangleq \min(t_{k-1}, \theta_{\min}), & t'' &\triangleq \min(t_k, \theta_{\min}), \\ \tau' &\triangleq \min(\tau_{k-1}, \check{\theta}_{\min}), & \tau'' &\triangleq \min(\tau_k, \check{\theta}_{\min}). \end{aligned}$$

By  $x'$  we denote the position  $y(t')$  and by  $w'$ , a point with the coordinates  $(\pi_1 \check{Y}(\tau'), \dots, \pi_d \check{Y}(\tau'))$ . In particular, for all  $t$  in  $\tau' \leq t < \tau''$  it follows that  $\pi_0 \check{Y}(t) = (k - 1)h$ ; moreover, the estimates (3.7) hold.

Let us disintegrate the probability  $\tilde{\mathbb{P}}$  over  $t'$  and  $\tau'$ . We denote the conditional expectation with respect to the sigma-algebra  $\mathcal{F}_{t'} \otimes \check{\mathcal{F}}_{\tau'}$  by  $\tilde{\mathbb{E}}_{t', \tau'}$ ; note that the pair  $(w', x')$  is  $\mathcal{F}_{t'} \otimes \check{\mathcal{F}}_{\tau'}$ -measurable.

For all  $w, x \in \mathbb{R}^{d+2}$ , the definitions of  $R_{x', w'} = \langle \cdot - x', x' - w' \rangle_d$  and  $R_{w', x'}$  and the equality

$$\|w - x\|_d^2 - \|w - x + x' - w'\|_d^2 = 2R_{x', w'}(x) + 2R_{w', x'}(w) + \|x' - w'\|_d^2$$

imply the inequalities

$$\frac{\|x - w\|_d^2 - \|x' - w'\|_d^2}{2} \leq \|x - x'\|_d^2 + \|w - w'\|_d^2 + R_{x', w'}(x) + R_{w', x'}(w)$$

and

$$\begin{aligned} \tilde{\mathbb{E}}_{t', \tau'} \|y(t'') - \check{Y}(\tau'')\|_d^2 - \|x' - w'\|_d^2 &\leq 2\tilde{\mathbb{E}}_{t', \tau'} \|\check{Y}(\tau'') - w'\|_d^2 \\ &+ 2\tilde{\mathbb{E}}_{t', \tau'} \|y(t'') - x'\|_d^2 + 2\tilde{\mathbb{E}}_{t', \tau'} R_{x', w'}(y(t'')) + 2\tilde{\mathbb{E}}_{t', \tau'} R_{w', x'}(\check{Y}(\tau'')). \end{aligned} \tag{4.7}$$

The first term on the right-hand side has already been estimated in (3.7),

$$2\tilde{\mathbb{E}}_{t',\tau'}\|\check{Y}(t'') - w'\|_d^2 = 2\check{\mathbb{E}}_{\tau'}\|\check{Y}(\tau'') - w'\|_d^2 \leq 2h^2L(\sqrt{d} + 1). \tag{4.8}$$

The second term on the right-hand side in (4.7) is easy to estimate by virtue of the boundedness of the dynamics of (1.1),

$$2\tilde{\mathbb{E}}_{t',\tau'}\|y(t'') - x'\|_d^2 \leq 2L^2\tilde{\mathbb{E}}_{t',\tau'}(t'' - t')^2 \leq 2L^2h^2. \tag{4.9}$$

For the third term on the right-hand side in (4.7), using Dynkin’s formula (4.5), from (4.6),  $\|y(s) - x'\|_d^2 \leq L(s - t')$ , and  $t'' - t' \leq h$  we conclude that

$$\begin{aligned} 2\tilde{\mathbb{E}}_{t',\tau'}R_{x',w'}(y(t'')) &= 2\tilde{\mathbb{E}}_{t',\tau'}\left[R_{x',w'}(y(t'')) - R_{x',w'}(y(t'))\right] \\ &\leq 2h\tilde{\mathbb{E}}_{t',\tau'}\langle x' - w', \check{f}(x', \mu^{\text{opt}}, \nu^{\text{from } x'}) \rangle_d \\ &\quad + 6L\|x' - w'\|_d^2 + 4L\mathbb{E}_{t'}\int_{t'}^{t''}\|y(s) - x'\|_d^2 ds \\ &\leq 2\langle x' - w', \check{f}(x', \mu^{\text{opt}}, \nu^{\text{from } x'}) \rangle_d h + 6L\|x' - w'\|_d^2 h + 4L^3h^3/3. \end{aligned} \tag{4.10}$$

By virtue of Dynkin’s formula (3.6), the estimates (4.3) and (3.7), and

$$\tilde{\mathbb{E}}_{t',\tau'}(\tau'' - \tau') \leq \check{\mathbb{E}}_{\tau'}(\tau_k - \tau_{k-1}) = h,$$

the last term in (4.7) can be estimated via

$$\begin{aligned} 2\tilde{\mathbb{E}}_{t',\tau'}R_{w',x'}(\check{Y}(\tau'')) &\leq 2\tilde{\mathbb{E}}_{t',\tau'}\int_{\tau'}^{\tau''}\hat{\Lambda}[u^{\text{to } w'}(x'), v(s)]R_{x',w'}(\check{Y}(s)) ds \\ &\leq 2\left(L\|x' - w'\|_d^2 - \langle x' - w', \check{f}(x', \mu^{\text{opt}}, \nu^{\text{from } x'}) \rangle_d\right)\check{\mathbb{E}}_{\tau'}(\tau'' - \tau') \\ &\quad + 2L\int_{\tau'}^{\tau''}\check{\mathbb{E}}_{\tau'}\|\check{Y}(s) - w'\|_d^2 ds \\ &\stackrel{(3.7)}{\leq} 2L\|x' - w'\|_d^2 h - 2\langle x' - w', \check{f}(x', \mu^{\text{opt}}, \nu^{\text{from } x'}) \rangle_d h + 2h^3L^2(\sqrt{d} + 5). \end{aligned} \tag{4.11}$$

Substituting the estimates (4.9)–(4.11) into (4.7) and using the inequalities  $L\sqrt{d}h < 1$  and  $Lh(2L/3 + 5) < 1$  (see (2.1)), we conclude that

$$\begin{aligned} \tilde{\mathbb{E}}_{t',\tau'}\|y(t'') - \check{Y}(\tau'')\|_d^2 &\leq (1 + 8Lh)\|x' - w'\|_d^2 + 2h^2L(\sqrt{d} + 1 + 2L^2h/3 + L\sqrt{d}h + 5Lh) \\ &\leq \|y(t') - \check{Y}(\tau')\|_d^2(1 + 8Lh) + 6h^2L(1 + \sqrt{d}). \end{aligned}$$

For brevity, for all positive integer  $i$  we introduce the stopping times  $t_i^\theta \triangleq \min(t_i, \theta_{\min})$  and  $\tau_i^\theta \triangleq \min(\tau_i, \check{\theta}_{\min})$ . Now for  $i \leq T/h$  the above entails

$$\tilde{\mathbb{E}}_{t_{i-1}^\theta, \tau_{i-1}^\theta}\|y(t_i^\theta) - \check{Y}(\tau_i^\theta)\|_d^2 \leq \|y(t_{i-1}^\theta) - \check{Y}(\tau_{i-1}^\theta)\|_d^2 e^{8Lh} + 6Lh^2(1 + \sqrt{d}). \tag{4.12}$$

Iterating (4.12), for all values  $i$  less than  $k \triangleq T/h$  we also obtain

$$\begin{aligned} & \tilde{\mathbb{E}}_{t_{k-1}^\theta, \tau_{k-1}^\theta} \|y(t_k^\theta) - \check{Y}(\tau_k^\theta)\|_d^2 \\ & \leq \tilde{\mathbb{E}}_{t_{k-2}^\theta, \tau_{k-2}^\theta} \|y(t_{k-1}^\theta) - \check{Y}(\tau_{k-1}^\theta)\|_d^2 e^{2 \cdot 8Lh} + 6Lh^2 (1 + \sqrt{d}) (1 + e^{8Lh}) \\ & \leq \tilde{\mathbb{E}} \|y(t_0) - \check{Y}(\tau_0)\|_d^2 e^{8kLh} + 6Lh^2 (1 + \sqrt{d}) \sum_{i=0}^{k-1} e^{8iLh} \\ & \leq \left( \frac{hd}{2} + \frac{6Lh^2 (1 + \sqrt{d})}{8Lh} \right) e^{8LT} \leq 2hde^{8LT}. \end{aligned}$$

Here, apart from  $\|y(t_0) - \check{Y}(\tau_0)\|_d^2 = \|x_0 - x_*\|^2 \leq hd/2$ , at the last step we also used the inequality  $1 + \sqrt{d} \leq 2d$ . Thus, it has been shown that

$$\tilde{\mathbb{E}} \left\| y(\min(T, \theta_{\min})) - \check{Y}(\min(\tau_{T/h}, \check{\theta}_{\min})) \right\|_d \leq e^{4LT} \sqrt{2d} \sqrt{h}. \tag{4.13}$$

4.4. Estimating the Difference of Payments

Now our task is to estimate the difference of  $J(y, \theta_1, \theta_2)$  and  $\check{\sigma}(\check{Y}(\tau_{T/h}))$ . To this end, we consider three cases:  $\theta_{\min} = \theta_1 \leq T$ ,  $\theta_{\min} = \theta_2 \leq T$ , and  $\theta_{\min} > T$ .

In the first case, we have  $\theta_{\min} = \theta_1 \leq T$ ; then  $J(y, \theta_1, \theta_2) = \sigma_1(y(\theta_{\min}))$  and  $\check{\theta}_{\min} \leq \tau_{T/h}$ ; since

$$\check{\sigma}(\check{Y}(T/h)) = \check{\sigma}(\check{Y}(\theta_{\min})) = \sigma_1(\pi_0 \check{Y}(\theta_{\min}), \dots, \pi_d \check{Y}(\theta_{\min})),$$

we have

$$\left| J(y, \theta_1, \theta_2) - \check{\sigma}(\check{Y}(\tau_{T/h})) \right| = \left| \sigma_1(y(\theta_{\min})) - \sigma_1(\pi_1 \check{Y}(\theta_{\min}), \dots, \pi_d \check{Y}(\theta_{\min})) \right|.$$

Now, by virtue of the  $L$ -Lipschitz continuity of  $\sigma_1$ , from (4.13) we obtain

$$\tilde{\mathbb{E}} \left( \left| J(y, \theta_1, \theta_2) - \check{\sigma}(\check{Y}(\tau_{T/h})) \right| \mid \theta_{\min} = \theta_1 \leq T \right) \leq Le^{4LT} \sqrt{2d} \sqrt{h}.$$

In the second case, a similar estimate follows from the  $L$ -Lipschitz continuity of  $\sigma_2$ .

For the third case of  $\theta_{\min} > T$ , note that, by virtue of the choice of  $T > N_{\varepsilon/12}$ , condition (1.4) for  $\sigma_i$  implies, now for  $\check{\sigma}$ , the inequalities

$$\begin{aligned} & \tilde{\mathbb{P}}(\theta_{\min} \geq T) \tilde{\mathbb{E}} \left( \check{\sigma}(T, y(T), 0) \mid \theta_{\min} \geq T \right) - \varepsilon/12 \\ & < \tilde{\mathbb{P}}(\theta_{\min} \geq T) \tilde{\mathbb{E}} \left( J(y, \theta_1, \theta_2) \mid \theta_{\min} \geq T \right) \\ & < \tilde{\mathbb{P}}(\theta_{\min} \geq T) \tilde{\mathbb{E}} \left( \check{\sigma}(T, y(T), 0) \mid \theta_{\min} \geq T \right) + \varepsilon/12. \end{aligned}$$

Now from the  $L$ -Lipschitz continuity of  $\sigma_i$  we have

$$\tilde{\mathbb{P}}(\theta_{\min} \geq T) \tilde{\mathbb{E}} \left( \left| J(y, \theta_1, \theta_2) - \check{\sigma}(\check{Y}(\tau_{T/h})) \right| \mid \theta_{\min} \geq T \right) < \varepsilon/12 + Le^{4LT} \sqrt{2d} \sqrt{h} \stackrel{(2.2)}{\leq} \varepsilon/6.$$

Combining all these cases, we obtain

$$\tilde{\mathbb{E}} \left| J(y, \theta_1, \theta_2) - \check{\sigma}(\check{Y}(\tau_{T/h})) \right| \leq \varepsilon/6 + Le^{4LT} \sqrt{2d} \sqrt{h} \stackrel{(2.2)}{\leq} \varepsilon/4.$$

It remains to note that now (3.4) implies

$$\check{\mathbb{E}} \left| \int_0^\infty h e^{-ht} \check{\sigma}(\check{Y}(t)) dt - J(y, \theta_1, \theta_2) \right| \leq \varepsilon/4 + \varepsilon/12 = \varepsilon/3.$$

This estimate is independent of the strategy chosen by the second player and the strategies  $\mu^{opt}, \phi^{opt}$  chosen by the first player entail the expected value  $\check{\mathbb{E}} \int_0^\infty h e^{-ht} \check{\sigma}(\check{Y}(t)) dt$  no more than  $\check{V}(0, z_0, 0)$  regardless of the strategy of the second player, therefore we have shown that

$$\check{\mathbb{E}} J(y, \theta_1, \theta_2) \geq \check{V}(0, z_0, 0) - \varepsilon/3.$$

Note that  $J(y, \theta_1, \theta_2)$  has been determined in the original game; then

$$\mathbb{E}_{\Omega, U, V} J(y, \theta_1, \theta_2) \geq \check{V}(0, z_0, 0) - \varepsilon/3.$$

Since the admissible strategy  $V$  and the terminal time  $\theta_2$  of the second player were arbitrary, it has been shown that

$$\mathcal{V}_- \geq \check{V}(0, z_0, 0) - \varepsilon/3.$$

Considering the guide for the second player based on the same Markov game, we obtain the symmetric estimate  $\mathcal{V}_+ \leq \check{V}(0, z_0, 0) + \varepsilon/3$ , based on which,

$$|\mathcal{V}_- - \mathcal{V}_+| \leq 2\varepsilon/3; \tag{4.14}$$

now the pair  $(U, \theta_1)$  constructed in Sec. 4.2 is  $2\varepsilon/3$ -optimal in the upper game; in particular, (1.6) also holds for it. Thus, the theorem is proved if the temporary assumptions are also satisfied in addition to condition (1.4).

### 5. GETTING RID OF TEMPORARY ASSUMPTIONS

Now we will show how to reduce the original game to a game with these assumptions. To this end we need the following assertion.

**Lemma 1.** *Suppose that players 1 and 2, in addition to the sets of admissible strategies,  $\mathfrak{U}$  and  $\mathfrak{V}$ , respectively, have two classes of terminal times,  $\mathfrak{Q}_b^I, \mathfrak{Q}_\sharp^I$  and  $\mathfrak{Q}_b^{II}, \mathfrak{Q}_\sharp^{II}$ , respectively. Suppose that we have the mappings*

$$\begin{aligned} \mathfrak{Q}_\sharp^I \ni \theta_1 &\mapsto \theta_1^\flat \in \mathfrak{Q}_b^I, \\ \mathfrak{Q}_b^{II} \ni \theta_2 &\mapsto \theta_2^\sharp \in \mathfrak{Q}_\sharp^{II} \end{aligned}$$

and nonnegative numbers  $r'$  and  $r''$  such that with any choice of terminal times  $\theta_1 \in \mathfrak{Q}_\sharp^I, \theta_2 \in \mathfrak{Q}_b^{II}$  and with any choice of the strategies of the players, the probabilities of the events  $|\theta_1 - \theta_1^\flat| > r'$  and  $|\theta_2 - \theta_2^\sharp| > r'$  are at most  $r''$ .

Then the inequality  $|\mathcal{V}_-^\flat - \mathcal{V}_-^\sharp| \leq 4r'' + Lr'$  holds, here  $\mathcal{V}_-^\sharp$  is the lower game value where the terminal times from the classes  $\mathfrak{Q}_\sharp^I$  and  $\mathfrak{Q}_\sharp^{II}$  are admissible and  $\mathcal{V}_-^\flat$  is the lower game value where the terminal times from the classes  $\mathfrak{Q}_b^I$  and  $\mathfrak{Q}_b^{II}$  are admissible.

**Proof.** First, note that by assumption, with any choice of admissible strategies  $U$  and  $V$  and admissible terminal times  $\theta_1 \in \mathfrak{Q}_\sharp^I$  and  $\theta_2 \in \mathfrak{Q}_b^{II}$ , the probability of the event  $|\min(\theta_1, \theta_2^\sharp) -$

$\min(\hat{\theta}_1^b, \hat{\theta}_2) > r'$  is at most  $2r''$ . Since  $J$  is bounded in absolute value by 1, it follows that the expectation generated by the strategies  $U$  and  $V$  for  $|J(y, \theta_1, \hat{\theta}_2^\sharp) - J(y, \theta_1^b, \hat{\theta}_2)|$  is at most  $4r'' + Lr'$ .

Let an admissible strategy  $U$  with terminal time  $\theta_1 \in \Omega_\sharp^I$  be an  $s$ -optimal pair in the lower game with  $\Omega_\sharp^I$  and  $\Omega_\sharp^{II}$ , and assume that there exists

$$\mathbb{E}_{\Omega,U,V} J(y, \theta_1, \theta_2) \leq \mathcal{V}_-^\sharp + s$$

for every admissible strategy  $V$  and admissible terminal time  $\theta_2 \in \Omega_\sharp^{II}$ . Then for every admissible strategy  $V$  and admissible terminal time  $\theta_2 \in \Omega_\flat^{II}$  we have

$$\mathbb{E}_{\Omega,U,V} J(y, \theta_1^b, \theta_2) - 4r'' - Lr' \leq \mathbb{E}_{\Omega,U,V} J(y, \theta_1, \theta_2^\sharp) \leq \mathcal{V}_-^\sharp + s;$$

thus, since  $\theta_1^b \in \Omega_\flat^I$ , it has been shown that  $\mathcal{V}_-^b \leq \mathcal{V}_-^\sharp + 4r'' + Lr' + s$ . By virtue of the arbitrariness of the positive  $s$ , the inequality  $\mathcal{V}_-^b - \mathcal{V}_-^\sharp \leq 4r'' + Lr'$  is also proved. The symmetry  $(b, 1) \leftrightarrow (\sharp, 2)$  implies the opposite inequality. The proof of the lemma is complete. ■

5.1. Getting Rid of the First Assumption

It was shown earlier that under the temporary assumptions, (4.14) holds and (1.6) is valid for some  $2\varepsilon/3$ -optimal  $U$  and  $\theta_1$ . Let us show that if there is a pair  $(T, h)$  for the original game that satisfies only the second temporary assumption, then the pair  $(U, \theta_1)$  constructed for it in Sec. 4.2 is  $2\varepsilon/3 + 4h$ -optimal in the upper game; in particular, (1.6) holds, and moreover,

$$\mathcal{V}_- \geq \mathcal{V}_+ - 2\varepsilon/3 - 4h \stackrel{(2.1)}{\geq} \mathcal{V}_+ - \varepsilon. \tag{5.1}$$

Consider a modified game—a game with the same dynamics, the same objective function, and the same sequences  $t_k^I$  and  $t_k^I$  but satisfying

$$\begin{aligned} \hat{\phi}_k^- &\triangleq \min(1 - h, \phi_k^-), & \hat{\phi}_k^+ &\triangleq \min(1 - h, \phi_k^+), \\ \hat{\psi}_l^- &\triangleq \min(1 - h, \psi_l^-), & \hat{\psi}_l^+ &\triangleq \min(1 - h, \psi_l^+) \end{aligned}$$

in the case of  $t_k^I \leq T$ ,  $t_l^{II} \leq T$  and  $\hat{\phi}_k^- \triangleq \phi_k^-$ ,  $\hat{\phi}_k^+ \triangleq \phi_k^+$ ,  $\hat{\psi}_k^- \triangleq \psi_k^-$ , and  $\hat{\psi}_k^+ \triangleq \psi_k^+$  otherwise. This game satisfies the first assumption. By  $\Omega_\sharp^I$  and  $\Omega_\sharp^{II}$  we denote the classes of terminal times of the first and second players, respectively, admissible in the modified game. We also adopt  $\Omega_\flat^I \triangleq \Omega^I$  and  $\Omega_\flat^{II} \triangleq \Omega^{II}$ .

Now consider an arbitrary terminal time  $\theta_2 \in \Omega_\flat^{II} = \Omega_\flat^{II}$  of the second player admissible in the original game. It is generated by  $\mathcal{F}_{t_{k+1}^{II}}$ -measurable random variables  $\psi[k]$  ranging in  $[\psi_k^-; \psi_k^+]$  on  $D(\mathbb{R}_+, \mathbb{R}^d)$ . Then for all  $k \in \mathbb{N}$  we take  $\hat{\psi}[k] \triangleq \min(1 - h, \psi[k])$ . This guarantees the values of the random variables  $\hat{\psi}[k]$  in  $[\hat{\psi}_k^-; \hat{\psi}_k^+]$  with  $|\psi[k] - \hat{\psi}[k]| \leq h$ . Now the rules  $\hat{\psi}[k]$  restore the terminal time  $\theta_2^\sharp \in \Omega_\sharp^{II}$  now admissible in the modified game.

Now consider a terminal time  $\hat{\theta}_1 \in \Omega_\flat^I$  of the first player admissible in the modified game and the corresponding sequence of random variables  $\hat{\phi}[k]$ . Note that the value of  $\hat{\phi}[k]$  outside the interval  $[\phi_k^-; \phi_k^+]$  is possible only for  $t_{k+1}^I < T$  and either  $1 - h \leq \phi_k^-$  or  $h \geq \phi_k^-$ ; based on this, by the choice of  $h$ , we have  $\phi_k^- = \phi_k^+ = 1$  and  $\hat{\phi}_k^- = 1 - h$ . Then we take  $\phi[k] = \hat{\phi}[k]$  for  $\hat{\phi}[k] \in [\phi_k^-; \phi_k^+]$  and  $\phi[k] = \phi_k^+$  otherwise; this guarantees  $|\hat{\phi}[k] - \phi[k]| \leq h$ . These rules reconstruct the terminal time  $\hat{\theta}_1^b \in \Omega_\flat^I$  admissible in the original game for the first player.

Due to the same dynamics, the sets  $\mathfrak{D}^I$  and  $\mathfrak{D}^{II}$  are the sets of all admissible strategies both in the original and in the modified game. In particular, any choice of an admissible pair  $(U, V) \in \mathfrak{D}^I \times \mathfrak{D}^{II}$  defines the same probability on all possible trajectories in each of the games. By construction, the probabilities of the events  $\theta_1 \neq \theta_1^\flat$  and  $\theta_2 \neq \theta_2^\sharp$  following the events  $|\psi[k] - \hat{\psi}[k]| \leq h$  are at most  $h$ .

All conditions in the lemma are satisfied for  $r'' = h$  and  $r' = 0$ , hence the lower values of the original and modified games differ by no more than  $4h$ ; now if the original game satisfies the second assumption, then the modified game satisfies both and its lower value is (4.14); so (5.1) is proved for the lower value of the original game.

Moreover, if the original game satisfies the second assumption, then the strategy  $U$  and terminal time  $\theta_1$  specified in Sec. 4.2 coincide with those constructed by the same procedure in the modified game. In particular, they are admissible and  $2\varepsilon/3 + 4h$ -optimal for the upper modified game. Therefore, in the original problem they are  $2\varepsilon/3 + 8h$ -optimal; in particular, they satisfy (1.6). Thus, (5.1) and (1.6) are proved if the original game satisfies the second assumption.

### 5.2. Getting Rid of the Second Assumption

We first need two unary operations on  $\mathbb{R}_+$ :  $\lceil \cdot \rceil$  and  $\lfloor \cdot \rfloor$ . For each  $R \in [0; T/h]$ , let  $\lceil R \rceil$  denote the least element of the lattice  $h\mathbb{N}$  greater than  $R$ ;  $\lfloor R \rfloor$  designates the largest element of  $h\mathbb{N}$  lattice not greater than  $R$ . For every  $R \notin [0; T/h]$ , set  $\lceil R \rceil = \lfloor R \rfloor = R$ . Moreover, since any subinterval of the interval  $[0; T]$  of length  $2h$  does not contain more than one element of the sequences  $(t_k^I)_{k \in \mathbb{N}}$  and  $(t_k^{II})_{k \in \mathbb{N}}$ , it follows that there are no such elements on  $[T - h; T + h]$  either; then it follows from  $0 < \lceil t \rceil - t$ ,  $t - \lfloor t \rfloor < h$  that these unary operations preserve the order between the elements  $t_k^I$  and  $t_k^{II}$ ; moreover, the order is not violated if the operation is applied to only one of the elements.

Now introduce two auxiliary games with the same dynamics, the same objective function, and the same compact sets  $[\phi_k^-; \phi_k^+]$  and  $[\psi_k^-; \psi_k^+]$  but differing from the original game by possible terminal times. In the  $\lceil \cdot \rceil$ -game, these sequences will be formed by the elements  $\lceil t_k^I \rceil$  for the first player and the elements  $\lfloor t_k^{II} \rfloor$  for the second one. In the  $\lfloor \cdot \rfloor$ -game, this will be the other way round:  $\lfloor t_k^I \rfloor$  for the first player and  $\lceil t_k^{II} \rceil$  for the second one. Denote the corresponding classes of terminal times admissible for the players by  $\lceil \mathcal{Q}^I \rceil$ ,  $\lfloor \mathcal{Q}^{II} \rfloor$  and  $\lfloor \mathcal{Q}^I \rfloor$ ,  $\lceil \mathcal{Q}^{II} \rceil$  for the  $\lceil \cdot \rceil$ -game and the  $\lfloor \cdot \rfloor$ -game. Note that in this case the unary operations  $\lceil \cdot \rceil$  and  $\lfloor \cdot \rfloor$  send  $\lfloor \mathcal{Q}^I \rfloor$  and  $\lceil \mathcal{Q}^I \rceil$ , as well as  $\lfloor \mathcal{Q}^{II} \rfloor$  and  $\lceil \mathcal{Q}^{II} \rceil$ , to each other.

Now consider an arbitrary terminal time  $\theta_i$  of one of the players admissible in the original game; in particular, for every  $t$  the events  $\theta_i \leq t$  are  $\mathcal{F}_t$ -measurable. Then at the terminal time  $\lceil \theta_i \rceil$  for every  $t$  the events  $\lceil \theta_i \rceil \leq t$  are  $\mathcal{F}_t$ -measurable as well. Therefore, the terminal time  $\lceil \theta_1 \rceil$  will be admissible in the  $\lceil \cdot \rceil$ -game and the time  $\lfloor \theta_2 \rfloor$ , in the  $\lfloor \cdot \rfloor$ -game as the times having suitable sets of values. The same reasoning leads to the same conclusion if  $\theta_1$  is taken to be a terminal time allowed in the  $\lfloor \cdot \rfloor$ -game and  $\theta_2$  is a terminal time allowed in the  $\lceil \cdot \rceil$ -game.

Again, owing to the same dynamics, the sets of admissible strategies are common to all the games under consideration, and each such admissible strategy creates the same distribution for all games on the trajectories  $y$ . By construction, the probabilities of events  $|\theta_i - \lceil \theta_1 \rceil| > h$  and  $|\theta_i - \lfloor \theta_1 \rfloor| > h$  are zero. By virtue of Lemma 1, for  $r' = h$  and  $r'' = 0$  the lower values of the  $\lceil \cdot \rceil$ -game and the  $\lfloor \cdot \rfloor$ -game satisfy the inequality  $|\mathcal{V}_-^{\lceil \cdot \rceil} - \mathcal{V}_-^{\lfloor \cdot \rfloor}| \leq Lh$ . Now, using the proved embeddings for terminal times in the original game, we have

$$\mathcal{V}_- = \sup_{V \in \mathfrak{D}^{II}, \theta_2 \in \mathfrak{Q}^{II}} \inf_{U \in \mathfrak{D}^I, \theta_1 \in \mathfrak{Q}^I} \mathbb{E}_{\Omega, U, V} J(y, \theta_1, \theta_2)$$

$$\begin{aligned} &\leq Lh + \sup_{V \in \mathfrak{D}^{II}, \theta_2 \in \Omega^{II}} \inf_{U \in \mathfrak{D}^I, \theta_1 \in \Omega^I} \mathbb{E}_{\Omega, U, V} J(y, [\theta_1], [\theta_2]) \\ &\leq Lh + \sup_{V \in \mathfrak{D}^{II}, \theta'_2 \in [\Omega^{II}]} \inf_{U \in \mathfrak{D}^I, \theta'_1 \in [\Omega^I]} \mathbb{E}_{\Omega, U, V} J(y, \theta'_1, \theta'_2) = Lh + \mathcal{V}_-^{[I]}. \end{aligned}$$

In a similar way, we can prove the inequality  $\mathcal{V}_- \geq Lh - \mathcal{V}_-^{[I]}$ . Thus, we have proved that

$$|\mathcal{V}_-^{[I]} - \mathcal{V}_-| \leq 2Lh.$$

Since the second assumption holds for the constructed  $[I]$ -game, it follows that the strategy  $U$  and the terminal time  $\theta_1^{[I]}$  constructed for it in Sec. 4.2 satisfy (1.6). In addition, (5.1) holds for this game, whence for the lower value  $\mathcal{V}_-$  of the original game it follows that

$$\mathcal{V}_- \geq \mathcal{V}_+ - 2\varepsilon/3 - 2h(L+2) \stackrel{(2.1)}{\geq} \mathcal{V}_+ - \varepsilon.$$

It remains to note that the strategy  $U$  does not change when going to the original game, and the terminal time  $\theta_1$  constructed in Sec. 4.2 for the original game is associated with  $\theta_1^{[I]}$  by the relation  $\theta_1 = [\theta_1^{[I]}]$ . Now the  $2\varepsilon/3 + 4h$ -optimality of the pair  $(U, \theta_1^{[I]})$  for the upper  $[I]$ -game guarantees the  $2\varepsilon/3 + 2h(L+4)$ -optimality of  $(U, \theta_1)$  for the original game. By virtue of (2.1), its  $\varepsilon$ -optimality has also been proved; i.e., (1.6) is also shown to hold true. Thus, the strategy  $U$  and the terminal time  $\theta_1$  constructed in Sec. 4.2 for this game satisfy (1.6) already without any assumptions on the original game other than property (1.4).

The proof of the theorem is complete. ■

#### FUNDING

This work was supported by the Russian Science Foundation, project no. 17-11-01093.

#### REFERENCES

1. Billingsley, P., *Convergence of Probability Measures*, New York–London–Sydney–Toronto: John Wiley & Sons, 1968. Translated under the title: *Skhodimost' veroyatnostnykh mer*, Moscow: Nauka, 1977.
2. Dynkin, E.B., Game variant of a problem on optimal stopping, *Sov. Math. Dokl.*, 1969, no. 10, pp. 270–274.
3. Krasovskii, N.N., A convergence-evasion game with stochastic guide, *Sov. Math. Dokl.*, 1977, no. 5, pp. 1020–1023.
4. Krasovskii, N.N. and Subbotin, A.I., *Pozitsionnye differentsial'nye igry* (Positional Differential Game), Moscow: Nauka, 1974.
5. Krasovskii, N.N. and Kotelnikova, A.N., An approach-evasion differential game: stochastic guide, *Proc. Steklov Inst. Math.*, 2010, vol. 269, no. 1, suppl. 1, pp. 191–213.
6. Petrosyan, L.A. and Shevkopyas, E.V., Cooperative differential games with random duration, *Vestn. S.-Peterburg. Univ. Ser. 1: Mat. Mekh. Astron.*, 2000, no. 4, pp. 14–18.
7. Seregina, T.V., Ivashko, A.A., and Mazalov, V.V., Optimal stopping strategies in the game “The Price Is Right”, *Tr. Inst. Mat. Mekh. UrO RAN*, 2019, vol. 25, no. 3, pp. 217–231.
8. Khlopin, D.V., Differential game with the possibility of early termination, *Tr. Inst. Mat. Mekh. UrO RAN*, 2021, vol. 27, no. 4.
9. Amir, R., Evstigneev, I.V., and Schenk-Hoppé, K.R., Asset market games of survival: a synthesis of evolutionary and dynamic games, *Ann. Finance*, 2013, vol. 9, no. 2, pp. 121–144.  
<https://doi.org/10.1007/s10436-012-0210-5>

10. Averboukh, Y., Approximate solutions of continuous-time stochastic games, *SIAM J. Control Optim.*, 2016, vol. 54, no. 5, pp. 2629–2649. <https://doi.org/10.1137/16M1062247>
11. Averboukh, Y., Approximate public-signal correlated equilibria for nonzero-sum differential games, *SIAM J. Control Optim.*, 2019, vol. 57, no. 1, pp. 743–772. <https://doi.org/10.1137/17M1161403>
12. Basu, A. and Stettner, L., Zero-sum Markov games with impulse controls, *SIAM J. Control Optim.*, 2020, vol. 58, no. 1, pp. 580–604. <https://doi.org/10.1137/18M1229365>
13. Bensoussan, A. and Friedman, A., Nonlinear variational inequalities and differential games with stopping times, *J. Funct. Anal.*, 1974, vol. 16, no. 1, pp. 305–352. [https://doi.org/10.1016/0022-1236\(74\)90076-7](https://doi.org/10.1016/0022-1236(74)90076-7)
14. Bensoussan, A. and Friedman, A., Nonzero-sum stochastic differential games with stopping times and free boundary problems, *Trans. Am. Math. Soc.*, 1977, vol. 231, no. 2, pp. 275–327. <https://doi.org/10.1090/S0002-9947-1977-0453082-7>
15. Bielecki, T.R., Crépey, S., Jeanblanc, M., and Rutkowski, M., Arbitrage pricing of defaultable game options with applications to convertible bonds, *Quantitative Finance*, 2008, vol. 8, no. 8, pp. 795–810. <https://doi.org/10.1080/14697680701401083>
16. Guo, X. and Hernández-Lerma, O., Zero-sum continuous-time Markov games with unbounded transition and discounted payoff rates, *Bernoulli*, 2005, vol. 11, no. 6, pp. 1009–1029. <https://doi.org/10.3150/bj/1137421638>
17. Hamadéne S. Mixed zero-sum stochastic differential game and American game options, *SIAM J. Control Optim.*, 2006, vol. 45, no. 2, pp. 496–518. <https://doi.org/10.1137/S036301290444280X>
18. Kolokoltsov, V.N., *Markov Processes, Semigroups and Generators. De Gruyter Studies in Mathematics 38*, Berlin: De Gruyter, 2011.
19. Laraki, R. and Solan, E., The value of zero-sum stopping games in continuous time, *SIAM J. Control Optim.*, 2005, vol. 43, no. 5, pp. 1913–1922. <https://doi.org/10.1137/S0363012903429025>
20. Marin-Solano, J. and Shevkoplyas, E., Non-constant discounting and differential games with random time horizon, *Automatica*, 2011, vol. 47, no. 12, pp. 2626–2638. <https://doi.org/10.1016/j.automatica.2011.09.010>
21. Neyman, A., Continuous-time stochastic games, *Games Econ. Behav.*, 2017, vol. 104, pp. 92–130. <https://doi.org/10.1016/j.geb.2017.02.004>
22. Prieto-Rumeau, T. and Hernández-Lerma, O., *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games*. Vol. 5 of *ICP Advanced Texts in Mathematics*, London: Imperial College Press, 2012.