[Communications on Stochastic Analysis](https://repository.lsu.edu/cosa)

Volume 11 | Number 2 Article 2

6-2017

Mixed Strategies for Deterministic Differential Games

Wendell H. Fleming
Brown University, wendell_fleming@brown.edu

Daniel Hernandez-Hernandez
Centro de Investigacion en Matematicas, Guanajuato, dher@cimat.mx

Communications on Stochastic Analysis
Vol. 11, No. 2 (2017) 137-156

Serials Publications
www.serialspublications.com

MIXED STRATEGIES FOR DETERMINISTIC DIFFERENTIAL GAMES

WENDELL H. FLEMING AND DANIEL HERNÁNDEZ–HERNÁNDEZ*

Abstract. This paper considers mixed strategies for two player, zero sum differential games on a finite time interval. Motivated by the classical result in static games that ensures the existence of a saddle point when mixed strategies are allowed for both players, we extend those ideas to differential games, introducing a Law of Large Numbers game, for which the value function coincides with the limit of discrete time Markov games. Finally, of particular interest is the class of strategies called approximately optimal Markov strategies, which are studied in the last part.

1. Introduction

This paper concerns the zero-sum game of a system represented by a differential equation of the form

$$\frac{d}{ds}x_s = f(s, x_s, u_s, z_s), \tag{1.1}$$

where $u., z.$ are the controls chosen by the two players. Player I (using control $u.$) is attempting to minimize a payoff P of the form (3.2), while the Player II is trying to maximize it with the control $z.$. The value of this game depends on the information pattern that is considered. For instance, Elliott-Kalton [8] introduced the definition of the upper and lower value, using the notion that one of the players may have some advantage on the available information.

The pioneering work on these differential games was done in the early 1950s by Rufus Isaacs [18], and was based on the solution of what is called now the Isaacs PDE by the method of characteristics. The information advantage of each player is somehow reflected in the structure of this PDE, and it is known that when the *Isaacs minimax condition* holds in the Hamiltonian, the upper and lower values of the game coincide. In this paper we consider differential games for which the Isaacs condition does not hold, and allow both players to use time-varying mixed strategies. It is a standard result in the theory of static games that, when it is allowed that both players select their decisions following a fixed probability distribution on the set of actions, there is a pair of *equilibrium strategies* μ^*, ν^*.

For static games in which upper and lower values are different, it is well known that a saddle point exists in terms of mixed strategies (Section 2 below). These

Received 2016-11-14; Communicated by T. Duncan.
2010 *Mathematics Subject Classification.* Primary 91A05, 91A15 ; Secondary 91A25.
Key words and phrases. Deterministic differential games, mixed strategies, Elliott-Kalton strategies, saddle point property, Law of Large Numbers game.
* This research is supported by Conacyt Grant 254166.

are probability measures μ and ν on the spaces U and Z on which the payoff $P(u,z)$ is defined. Early work in the 1960s on mixed strategies for differential games included [11], [20, Chap. 9]. To the best of our knowledge, this topic seems then to have been neglected until the recent paper [5].

At a formal level, in a mixed strategy differential game the players choose time varying probability measures μ_s and ν_s from which the controls u_s and z_s in (1.1) are chosen by random sampling. Since it is difficult to formulate this idea precisely, we consider in Section 4 discrete time approximations. It is shown that the value functions for these approximating discrete time games tend to a limit $v(t,x)$ as the mesh of the partitions tends to zero. This limit function satisfies the corresponding Isaacs PDE (4.1) in the sense of Crandall-Lions viscosity solutions [7]. This result is formulated as Theorem 4.4, and it has been obtained before by Buckdahn, Li and Quincampoix [5]. Nevertheless, this result is important in order to present the results of this paper.

In the study of differential games, the definition of value function as a (suitably interpreted) solution to the Isaacs PDE represents only one part of the theory. From an applications viewpoint, an equally important question is to find optimal (or approximately optimal) control strategies for the two players which depend on the game state and time. Such control strategies are called Markov. For mixed strategy differential games, the Isaacs PDE (4.1)-(4.2) provides formally a recipe for obtaining Markov control strategies, which are measure valued functions of state and time. Unfortunately this formalism can rarely be made rigorous. Instead, we seek control strategies which are approximately Markov, as explained in Section 6. Another feature of our paper is the idea of coarse and fine partitions of the time interval in which the mixed strategy differential game is played (Section 7).

We call $v(t,x)$ the value of the mixed strategy differential game, for initial time t and initial game state x. The main contribution of this paper is not a study of the value function itself. Instead, we focus on approximately optimal control strategies which can be obtained from the value function. This work is related with our paper [12] on approximately optimal strategies and saddle points for differential games (deterministic or stochastic), and to the earlier papers [11] and [15]. As a first step, in Section 5 auxiliary differential games called Law of Large Numbers (LLN) games are introduced. In these games, the dynamics and running cost are averaged with respect to measures μ and ν on the control spaces U and Z, in a way similar to a technique used in *relaxed control theory*. See (5.1) and (5.2). The Isaacs minimax condition holds for the LLN game, and the value function $V_M(t,x)$ agrees with the limit $v(t,x)$ in Section 4 .

The differential games which we consider have a kind of *Markovian* structure. This suggests that the game state at any time s should contain enough information to choose control strategies for both players in the LLN game which are (approximately) optimal. See Theorem 6.3. This result cannot be applied directly to obtain approximately optimal strategies for the mixed strategy game itself. To avoid this difficulty, in Section 7, we modify the stochastic difference game model in Section 4. The modified model has two steps. At the first step, approximately optimal Markov strategies are chosen as in Section 6, for a *coarse* partition π of the time interval $[t,T]$ of play. Then controls u_s, z_s which are constant on much smaller subintervals of π are chosen by random sampling. This model is suggested by the intuitive idea that controls u_s and z_s chosen by random sampling should vary with time more rapidly than the probability measures μ_s, ν_s from which the sampling occurs. The main result of this paper is Theorem 7.2, which provides an approximate saddle point for the mixed strategy differential game, in terms of modified stochastic difference games of the kind described in Section 7 and the Appendix.

As this paper was being written, we learned of Sirbu's paper [22] on mixed strategy stochastic differential games. His work also is based on discrete time approximations. In relation with stochastic differential games and mixed strategies, there is also a recent work by Buckdahn, Li and Quincampoix [6], where they use BSDEs techniques to analyse the existence of value of the game. When the Isaacs minimax condition (3.6) does not hold, the mixed strategy value $v(t,x)$ may be strictly above the lower game value $V_-(t,x)$ or strictly below the upper value game $V_+(t,x)$. Kaise and Sheu [19] described a different model which also leads to value functions with this property. Another way to get intermediate value functions is using a randomised structure in the order in which decisions are taken by the players; these ideas are elaborated in the recent paper [17].

2. Static Games

It is a classical result in game theory that a saddle point is obtained by introducing mixed strategies. In order to fix ideas, consider a static game with continuous payoff $P(u,z)$ with $u \in U,\ z \in Z$ and $U,\ Z$ compact sets in a metric space. The upper and lower values are defined as

$$V_+ = \min_{u\in U} \max_{z\in Z} P(u,z), \quad V_- = \max_{z\in Z} \min_{u\in U} P(u,z).$$

In the definition of V_+, the maximizer has what we call an information advantage, since z can be chosen as a function of u. Similarly, the minimizer has an information advantage in the definition of V_-, choosing u as a function of z. Neither player has an information advantage if $V_+ = V_-$, and the game has a saddle point. Moreover, $V_+ = V_- = val(P)$ is called the *value of the game*.

When $V_- < V_+$, the value of the game and saddle points are defined in terms of mixed strategies, as follows. Denote by $\mathcal{P}(U)$ and $\mathcal{P}(Z)$ the set of probability measures on the $\sigma-$algebra of Borel subsets $\mathcal{B}(U)$ of U, and $\mathcal{B}(Z)$ of Z, respectively. These spaces of probability measures are endowed with the weak topology; it is well known that there exists a metric on $\mathcal{P}(U)$ (or $\mathcal{P}(Z)$) which is compatible with the weak topology, and with respect to this topology $\mathcal{P}(U)$ and $\mathcal{P}(Z)$ are compact [9, p. 96].

We define a mixed strategy for the minimizer player as a probability measure μ on $\mathcal{B}(U)$ and, similarly, a mixed strategy for the maximizer player is a probability measure ν on $\mathcal{B}(Z)$. The players choose mixed strategies $\mu,\ \nu$ with payoff

$$\hat{P}(\mu,\nu) = \int_U \int_Z P(u,z)\mu(du)\nu(dz).$$

Since this transformation is bilinear, classical minimax theorems (see, e.g. [1, p. 218]) guarantee the existence of a *saddle point* $(\mu^*,\ \nu^*)$, that is, for every pair of probability measures $(\mu,\ \nu)$,

$$\hat{P}(\mu^*,\nu) \le \hat{P}(\mu^*,\nu^*) \le \hat{P}(\mu,\nu^*). \tag{2.1}$$

The value of the mixed strategy game, denoted by $val(P)$, is defined as $\hat{P}(\mu^*,\nu^*)$. Note that (2.1) is equivalent to the following inequalities:

$$\begin{aligned} \int_U P(u,z)\mu^*(du) &\le val(P), \text{ for all } z \in Z, \\ \int_Z P(u,z)\nu^*(dz) &\ge val(P), \text{ for all } u \in U. \end{aligned}$$

To achieve a *saddle point* for static games with mixed strategies, many independent plays of the game should be considered. Suppose that on each play, the minimizer chooses u randomly, with probability distribution $\mu(du)$, and the maximizer chooses z randomly with probability distribution $\nu(dz)$. For the optimal μ^* the minimizer achieves average payoff no more than the mixed strategy value $val(P)$ as the number of plays tends to infinity. Similar arguments hold for the maximizer player.

3. Differential Game Formulations

Given $T > 0$ a finite time horizon and $t \in [0,T)$, consider a controlled dynamical system for which the state process x_s at time $s \in [t,T]$ evolves according to the ordinary differential equation

$$\frac{d}{ds}x_s = f(s,x_s,u_s,z_s), \tag{3.1}$$

with initial condition $x_t = x \in \mathbb{R}^d$. The controls adopted by Players I and II are, respectively, the control processes u_s and z_s taking values in some compact subsets U and Z of $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$.

We make the following assumptions on the coefficients of the above ODE: The function $f : [0,T] \times \mathbb{R}^d \times U \times Z \to \mathbb{R}^d$ is bounded, continuous, and uniformly Lipschitz continuous with respect to t, x for $(u,z) \in U \times Z$.

The game payoff is defined by

$$P(t,x;u.,z.) = \int_t^T L(s,x_s,u_s,z_s)ds + g(x_T), \tag{3.2}$$

with $L : [0,T] \times \mathbb{R}^d \times U \times Z \to \mathbb{R}$ being bounded, continuous, and uniformly Lipschitz continuous with respect to t, x for $(u,z) \in U \times Z$ and $g : \mathbb{R}^d \to \mathbb{R}$ being bounded and Lipschitz continuous. In this zero sum differential game Player 1 is the minimizing controller, while Player II is trying to maximize P.

Remark 3.1. A basic reference on this subject is the book of Fleming-Soner [14, Chapter 11, Second edition]. We follow the notation and assumptions there (except that G on p. 377 in [14] is now f). We also notice that at the bottom of p. 377 in [14], U and Z are compact subsets of Euclidean spaces. However, the results remain true (with no changes in the proofs) if U and Z are compact metric spaces. In particular, this observation applies to the Law of Large Numbers game in Section

5. Another excellent introduction to differential games and viscosity solution is [2, Chap. 8].

In order to have a complete description of the differential game we need to introduce *upper and lower values* of the game. This procedure can be formulated in terms of anticipative strategies, using the Elliott-Kalton definition; see, for instance, [14, Chapter 11] and [2, Chapter VIII] for the precise definitions.

In fact, it turns out that the *upper value* $V_+(t,x)$ can be characterized as the unique bounded, Lipschitz viscosity solution of the following Isaacs PDE, see [10], [14, Thm. 11.6.1],

$$v_t + H^+(Dv(t,x),x,t) = 0, \quad v(T,x) = g(x), \tag{3.3}$$

with Dv the gradient of $v(t,\cdot)$ and

$$H^+(p,x,t) := \min_{u\in U}\max_{z\in Z} F(p,x,t;u,z), \tag{3.4}$$

$$F(p,x,t;u,z) := [f(t,x,u,z)\cdot p + L(t,x,u,z)]\,. \tag{3.5}$$

Similarly, the *lower value* $V_-(t,x)$ has associated an Isaacs PDE, replacing the second term on the left side of (3.3) by $H^-(p,x,t) := \max_{z\in Z}\min_{u\in U} F(p,x,t;u,z)$.

At an intuitive level, the *upper value of the game* shall describe a certain advantage for the maximizing player in the information available to both players. It can be thought as if at each time $s\in[t,T]$ the maximizing player had information of the decision taken by the other player u_s before he chooses z_s. The *lower value of the game* can be defined similarly, establishing the advantage of the information to the minimizing player. If for all $t\in[0,T)$, $x\in\mathbb{R}^d$, and $p\in\mathbb{R}^d$ the identity

$$H^+(p,x,t) = \min_{u\in U}\max_{z\in Z} F(p,x,t;u,z) = \max_{z\in Z}\min_{u\in U} F(p,x,t;u,z) = H^-(p,x,t) \tag{3.6}$$

holds, it is said that the *Isaacs minimax condition* holds, and it can be verified that in that case $V_+(t,x) = V_-(t,x)$. It is clear that in general $H^-(p,x,t)\le H^+(p,x,t)$, and using a comparison principle we can conclude that $V_-(t,x)\le V_+(t,x)$.

The idea of information advantage to the stronger player in a differential game can be made more explicit for discrete time approximations to the game, using what is called in [12] approximate Markov strategies. See also Remark 6.4.

4. Mixed Strategy Differential Games

In the previous section were recalled the differential games when one of the players has some advantage on the information available when decisions are taken. Another important case is when both players choose simultaneously their decisions. A discrete time version of this differential game was analyzed by Fleming [11, p.199], and both players choose simultaneously at each step, and neither player knows the opponent's choice at the current step. In that paper the game dynamics (3.1) are discretized. By formally taking a continuous limit, an Isaacs-type PDE is obtained (see equation (1.6) in [11]):

$$v_t + H(Dv(t,x),x,t) = 0, \quad v(T,x) = g(x), \tag{4.1}$$

with

$$H(p,x,t) = val_{u,z}F(p,x,t;u,z) = val_{u,z}\left[f(t,x,u,z)\cdot p + L(t,x,u,z)\right]. \tag{4.2}$$

Here $val_{u,z}F(p,x,t;u,z)$ is understood as the common value $val(F)$ described in (2.1) with $P(u,z) = F(p,x,t;u,z)$, for t,x and p fixed. It is clear that

$$H^- \leq H \leq H^+.$$

Example 4.1. The following example is useful to illustrate our results. Let $d = 1$, and define the corresponding set of controls for each player as $U = \{-1,1\}$, $Z = \{-1,1\}$, and one-dimensional vector field $f(t,x,u,z) = uz$, with cost function $L(t,x,u,z) = 0$. In this case, it is easy to compute $H^{\pm}(p,x,t)$ as

$$H^+(p,x,t) = 0, \quad H^-(p,x,t) = -|p| \quad \text{and} \quad H(p,x,t) = 0.$$

Remark 4.2. The equation (4.1) with boundary condition at time T has a unique bounded, uniformly continuous viscosity solution. This follows from the main result of Buckdahn, Li and Quincampoix [5, Thm. 4.1]. Their proof of this statement is based on discrete time approximations. Roughly speaking, given a partition π of the interval $[0,T]$, they define controls to be stochastic processes $u(s),\ z(s)$, which are constant on each interval I_j of partition π. The upper and lower values of the discrete game are defined and pass to the limit when the mesh size of π goes to zero.The same result about existence and uniqueness also follows from our Theorem 5.2, which concerns what we call a Law of Large Numbers game associated with the problem.

The propose of this section is to introduce another approximation of the viscosity solution $v(t,x)$ in (4.1) by the value functions $V^\pi(t,x)$ of certain discrete time stochastic games where π denotes a partition of the time interval $[t,T]$. The method is based in the following construction, originally due to Souganidis [23]. For stochastic differential games a similar construction was used in [15, Sec. 2]. We describe the argument briefly, omitting some proofs, for completeness.

First, define the family of operators $G(t,\tau)$ as follows. For $t < \tau$ define the operator G on the set $C_b^{0,1}(\mathbb{R}^d)$ of bounded, Lipschitz continuous functions on $\mathbb{R}^d$ by

$$G(t,\tau)\varphi(x) = val_{u,z}\left\{\varphi(x_\tau) + \int_t^\tau L(s,x_s,u,z)ds\right\}, \tag{4.3}$$

for $0 \leq t \leq \tau \leq T$. Here x_s corresponds to the solution of (3.1) with initial condition $x_t = x$, when the first player chooses the probability measure $\mu \in \mathcal{P}(U)$ with constant realizations $u_s = u$ according to this distribution and, similarly, the second player chooses a probability measure $\nu \in \mathcal{P}(Z)$ with constant realizations $z_s = z$ along the interval $[t,\tau]$. Notice that the symbol on the right side of (4.3) should be understood as the value of the two-player, zero sum static game over $U \times Z$ with payoff

$$P = \varphi\left(x + \int_t^\tau f(r,x_r,u,z)dr\right) + \int_0^\tau L\left(s, x + \int_t^s f(r,x_r,u,z)dr, u, z\right)ds.$$

This operator maps the set $C_b^{0,1}(\mathbb{R}^d)$ into itself, and it can be verified that for each φ in $C_b^{0,1}(\mathbb{R}^d)$, we have that

$$\lim_{\tau\to t}\frac{G(t,\tau)\varphi(x) - \varphi(x)}{\tau - t} = H(\varphi_x(x),x,t), \tag{4.4}$$

with $H(p,x,t)$ as in (4.2).

Throughout $\pi = \{0 = t_0 < t_1 < \cdots < t_N = T\}$ denotes a partition of $[0,T]$, with $\|\pi\| = \max_j(t_{j+1} - t_j)$. Define recursively backward in time

$$V^\pi(t,x) = \begin{cases} g(x),\ t = T \\ \prod_{j=i}^{N-1} G(t_j, t_{j+1}) g(x), \text{ if } t = t_i < T. \end{cases} \tag{4.5}$$

In the Appendix we will see that V^π is the value function for the following discrete time dynamic game. This is called a verification theorem. The state at the initial time $t = t_i$ is $x = x_{t_i}$, and for simplicity we denote the state at time t_j by $x_j = x_{t_j}$, for $j = i, \ldots, N$. For $t_j \le s < t_{j+1}$ the minimizer chooses $u_s = u_j$ and the maximizer chooses $z_s = z_j$, where u_j is a $U-$valued random variable with distribution μ_j and z_j is a $Z-$valued random variable with distribution ν_j. The probability measures $\mu_j,\ \nu_j,\ j = i, \ldots, N-1$, are chosen according to decision strategies (see Appendix A), and the corresponding solution of (3.1) is denoted by x_s, which is constructed piecewise on each subinterval of the partition π. This defines a discrete time stochastic game, in which the random inputs for the minimizer and maximizer players are independently chosen on the subintervals, and knowing the initial state x_i and also the previous realizations of the controls $u_k,\ z_k$ for $i \le k < j$. The game payoff is

$$J(t,x;\mu.,\nu.) = \mathbb{E}\left[\sum_{j=i}^{N-1} \int_{t_j}^{t_{j+1}} L(s, x_s, u_s, z_s) ds + g(x_N)\right]. \tag{4.6}$$

A more precise formulation of such discrete time games is given in the Appendix. Standard arguments show that $V^\pi(t,x)$, as defined in (4.5), is indeed the value function:

$$V^\pi(t,x) = val J(t,x;\mu.,\nu.),$$

where as before *val* denotes the game value. Moreover, there is a discrete time dynamic programming principle. For $i < j < N$ with $t = t_i$ and $x = x_i$,

$$V^\pi(t_i, x_i) = val\ \mathbb{E}\left[\int_{t_i}^{t_j} L(x_s, u_s, z_s) ds + V^\pi(t_j, x_j)\right]. \tag{4.7}$$

In particular, a one-step dynamic programming principle holds

$$V^\pi(t_j, x) = val_{u,z}\left\{V^\pi(t_{j+1}, x_{t_{j+1}}) + \int_{t_j}^{t_{j+1}} L(s, x_s, u, z) ds\right\}, \tag{4.8}$$

with $V^\pi(T,x) = g(x)$. Using the dynamic game value interpretation on V^π, described in detail in the Appendix, we can derive important properties of this function.

Lemma 4.3. (1) *There exists M such that for all $x, \tilde{x} \in \mathbb{R}^d$,*

$$|V^\pi(t_i, x) - V^\pi(t_i, \tilde{x})| \le M\|x - \tilde{x}\|. \tag{4.9}$$

(2) *For $t = t_i,\ \tau = t_j$,*

$$|V^\pi(t,x) - V^\pi(\tau,x)| \le K(\tau - t),$$

for some constant K.

In the remainder of this section an increasing sequence of partitions π is chosen and we show that V^π tends to a limit v as the mesh size of the partition $\|\pi\| \to 0$.

Theorem 4.4. $\lim_{\|\pi\|\to 0} V^\pi(t,x) = v(t,x)$, *uniformly on compact sets, where v is the unique bounded, uniformly continuous viscosity solution to the Isaacs PDE (4.1).*

Proof. We shall argue proving that any possible limit of V^π is a viscosity solution of (4.1). Once this is proved the theorem follows applying a comparision principle [14, Thm. II.9.1] and using Lemma 4.3. Let v be the locally uniform limit of V^π through some sequence $\|\pi\| \to 0$, and note that thanks to Lemma 4.3 this convergence is uniform in compact sets. We shall prove that v is a viscosity solution of (4.1). We present here only the proof that v is subsolution, since the arguments to prove that it is supersolution are similar.

Let $w : [0,T) \times \mathbb{R}^d \to \mathbb{R}$ be a smooth function such that $v - w$ has a strict local maximum at $(\bar{t}, \bar{x})$. We shall prove that at $(\bar{t}, \bar{x})$ we have the inequality

$$w_t + H(Dw(t,x), x, t) \geq 0, \tag{4.10}$$

with H as in (4.2). Hence, there exist $(t_\pi, x_\pi) \to (\bar{t}, \bar{x})$, as $\|\pi\| \to 0$, with $t_\pi \in \pi$, such that $V^\pi - w$ attains a local maximum at (t_π, x_π). Then, $t_\pi = t_j$ for some j and, when $\|\pi\|$ is small enough, for the corresponding solution x_s of (3.1) with initial condition x_π and any constant realizations $u \in Y$, $z \in Z$, chosen with constant control $\mu = \mu_j$ and $\nu = \nu_j$, we have that

$$V^\pi(t_{j+1}, x_{t_{j+1}}) - w(t_{j+1}, x_{t_{j+1}}) \leq V^\pi(t_\pi, x_\pi) - w(t_\pi, x_\pi).$$

Using the discrete time dynamic programming principle and the previous inequality we have that

$$w(t_\pi, x_\pi) \leq val_{u,z}\, \mathbb{E}\left[\int_{t_\pi}^{t_{j+1}} L(x_s, u_s, z_s) ds + w(t_{j+1}, x_{t_{j+1}})\right].$$

Finally, applying the fundamental theorem of calculus to $w(t_{j+1}, x_{t_{j+1}}) - v(t_\pi, x_\pi)$ and using (4.4), we get (4.10) taking the limit when $\|\pi\| \to 0$. $\square$

We call the function $v(t,x)$ in Theorem 4.4 the continuous time value function for the mixed strategy differential game. It is of historical interest to note that $v(t,x)$ coincides with the function $\tilde{v}(t,x)$ which was obtained by a similar discretization technique in [11]. That paper appeared in 1964, before the Crandall-Lions theory of viscosity solutions was invented. In [11, Lemma 2] $\tilde{v}(t,x)$ was shown to be the "vanishing viscosity" limit as $a \to 0$ of the classical solution to the second order parabolic PDE with an added term a times the Laplacian in x of v in (4.1), and with the same boundary data $g(x)$ at time T. Hence $\tilde{v}(t,x)$ is a viscosity solution of (4.1), and a standard uniqueness result implies that $v = \tilde{v}$.

Remark 4.5. In [19] Kaise and Sheu describe a different model in which there is a game value between the upper and lower values. It involves time discretizations in which the minimizer and maximizer have an information advantage in alternating time intervals of the discretization. In a recent paper [17] another approach based on the Perron's method was used to have a characterization of the intermediate value functions for these type of games.

5. Law of Large Numbers (LLN) Differential Game

First, recall that $\mathcal{P}(U)$ and $\mathcal{P}(Z)$ denote the spaces of probability measures on the $\sigma-$algebra of Borel subsets of U and Z, respectively. The control spaces of each player are compact subsets of Euclidean spaces, with $U \subset \mathbb{R}^{m_1}$ and $Z \subset \mathbb{R}^{m_2}$. As before, these spaces of probability measures are endowed with the weak topology; recall that this topology is compatible with the Prohorov metric on $\mathcal{P}(U)$ (or $\mathcal{P}(Z)$).

Now we introduce the (relaxed) controls for the LLN game, which are useful for the analysis of this game. The controls for the players are measure-valued functions, i.e. $\mu.$ and $\nu.$ take values in $\mathcal{P}(U)$ and $\mathcal{P}(Z)$, respectively. The vector valued function f in (3.1) and the cost function L in (3.2) can be extended as follows, for $\mu \in \mathcal{P}(U)$ and $\nu \in \mathcal{P}(Z)$:

$$\begin{aligned} \bar{f}(t,x;\mu,\nu) &= \int_U \int_Z f(t,x;u,z)\mu(du)\nu(dz), \\ \bar{L}(t,x;\mu,\nu) &= \int_U \int_Z L(t,x;u,z)\mu(du)\nu(dz). \end{aligned}$$

The state dynamics $\bar{x}_s$ of the LLN game satisfies

$$\frac{d}{ds}\bar{x}_s = \bar{f}(s,\bar{x}_s;\mu_s,\nu_s), \quad \bar{x}_t = x, \tag{5.1}$$

and the game payoff is

$$\bar{P}(t,x;\mu.,\nu.) = \int_t^T \bar{L}(s,\bar{x}_s;\mu_s,\nu_s)ds + g(\bar{x}_T). \tag{5.2}$$

The analogous version of the *Isaacs minimax condition* (3.6) for this game can be written, replacing F in (3.5) by

$$\bar{F}(p,x,t;\mu,\nu) = \bar{f}(t,x;\mu,\nu)\cdot p + \bar{L}(t,x;\mu,\nu).$$

Since $\bar{F}(p,x,t;\mu,\nu)$ is bilinear and weakly continuous in $(\mu,\nu) \in \mathcal{P}(U) \times \mathcal{P}(Z)$, it follows that the Isaacs condition is satisfied for the LLN game and, therefore, *the upper and lower value* of this game are the same, and the common value is denoted by $V_M(t,x)$. We use the subscript M for mixed strategies.

Remark 5.1. If there is no maximizing control $z.$ in the model, the LLN game becomes a deterministic "relaxed" control problem, with control μ_s. Time discretizations, like those in Section 6 below, led to 'chattering control" approximations. In the deterministic control literature, chattering controls were described deterministically rather by independent random sampling over time intervals of a partition as in Section 4. See [3, Chapter 4], [4, Chapter 3].

Theorem 5.2. *The mixed strategy value function $V_M(t,x)$ is the unique bounded, Lipschitz continuous viscosity solution of the Isaacs PDE (4.1), with boundary condition $V_M(T,x) = g(x)$.*

This is a special case of Theorem 11.6.1 in [14], as applied to the LLN differential game, with control spaces $\mathcal{P}(U)$ and $\mathcal{P}(Z)$. What we call LLN games were also considered in [2, Sec. 8.2], for differential games on an infinite time horizon with discounted cost functions.

6. Approximately Optimal Markov Strategies I

The dynamic programming/PDE approach to optimal stochastic control and differential games involves the idea of controls which are functions of time and state variables. These are often called Markov control strategies. For technical reasons, we cannot develop a theory in terms of optimal Markov control strategies. Instead, we use a notion of approximately optimal Markov strategies which was considered in [12][13]. It is based on similar concepts in [15] and in stochastic control theory [14, Section 4.7]. In this section we apply some of those results to the LNN game, with control variables μ and ν.

We define an admissible control $\mu. : [t,T] \to \mathcal{P}(U)$ (respectively $\nu. : [t,T] \to \mathcal{P}(Z)$) for Player I (resp. Player II) on $[t,T]$ as a measurable function, with respect to the σ-algebra $\mathcal{B}(\mathcal{P}(U))$, taking values in $\mathcal{P}(U)$ (resp. $\mathcal{P}(Z)$). The set of admissible controls for Player I (resp. II) is denoted by $\mathcal{U}^{\mathcal{P}}(t)$ (resp. $\mathcal{Z}^{\mathcal{P}}(t)$). For every $\mu. \in \mathcal{U}^{\mathcal{P}}(t)$ and $\nu. \in \mathcal{Z}^{\mathcal{P}}(t)$ there is a solution to the ODE (5.1) with given initial data. We identify the controls which are equal almost everywhere in the interval $[t,T]$.

An Elliott-Kalton strategy for the maximizing Player II beginning at time t is a mapping β from $\mathcal{U}^{\mathcal{P}}(t)$ into $\mathcal{Z}^{\mathcal{P}}(t)$ provided that for each $s \in [t,T]$, if $\mu. = \tilde{\mu}.$ almost everywhere in $[t,s]$, then $\beta(\mu.) = \beta(\tilde{\mu}.)$ a.e. in $[t,s]$. The set of these strategies is denoted as $\Delta^{\mathcal{P}}_{EK}(t)$. The set of Elliott-Kalton strategies $\alpha : \mathcal{Z}^{\mathcal{P}}(t) \to \mathcal{U}^{\mathcal{P}}(t)$ for the minimizing Player I can be defined in a similar way, and is denoted by $\Gamma^{\mathcal{P}}_{EK}(t)$.

Next we present the definition of approximately Markov strategies.

Definition 6.1. (a) $\alpha \in \Gamma^{\mathcal{P}}_{EK}(t)$ is an *approximately Markov strategy* for the minimizing player if there exists a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_N = T\}$ of $[0,T]$ with $t = t_i$ and Borel measurable functions $\phi_j : \mathbb{R}^d \to \mathcal{P}(U)$, for $j = i, i+1, \ldots, N-1$, such that

$$\alpha(\nu.)_s = \phi_j(\bar{x}_{t_j}), \;\; t_j \le s < t_{j+1}.$$

(b) $\beta \in \Delta^{\mathcal{P}}_{EK}(t)$ is an *approximately Markov strategy* for the maximizing player if there exists such a partition π and Borel measurable functions $\eta_j : \mathbb{R}^d \to \mathcal{P}(Z)$ such that

$$\beta(\mu.)_s = \eta_j(\bar{x}_{t_j}), \;\; t_j \le s < t_{j+1}.$$

Since the Isaacs minimax condition holds for the LNN game, [15, Prop. 2.3] can be used to get approximating optimal Markov strategies for both players, as is outlined in the proof of Theorem 6.3.

We recall briefly the discretization method of Fleming and Soner [14, Section 11.8] applied to the LNN game described above. Let $\pi = \{0 = t_0, t_1, \ldots, t_N = T\}$ be a partition of $[0,T]$, with $\|\pi\| = \max_j(t_{j+1} - t_j)$. For the initial time $t = t_i$, define the set of measurable functions

$$\mathcal{U}^{\mathcal{P}}_\pi = \mathcal{U}^{\mathcal{P}}_\pi(t) = \{\mu. : [t,T] \to \mathcal{P}(U) : \mu_s = \mu_{t_j}, \text{ for } s \in [t_j, t_{j+1}), j = i, \ldots, N-1\}.$$

and the upper value function $\bar{V}^\pi_+(t,x)$ as

$$\bar{V}^\pi_+(t,x) = \sup_{\beta \in \Delta^{\mathcal{P}}_{EK}} \inf_{\mu. \in \mathcal{U}^{\mathcal{P}}_\pi} \bar{P}(t,x;\mu.,\beta(\mu).),$$

where $\Delta_{EK}^{\mathcal{P}} = \Delta_{EK}^{\mathcal{P}}(t)$ denotes the set of Elliott-Kalton strategies, with $\bar{P}$ as in (5.2). The corresponding lower value function $\bar{V}_{-}^{\pi}$ is defined similarly.

Now, since the LLN game satisfies the Isaacs condition, $\bar{V}_{+}^{\pi}$ and $\bar{V}_{-}^{\pi}$ tend to V_M as $\|\pi\| \to 0$, uniformly in compact sets [14, Thm. 8.1, p. 391]. Note that in the above definition of $\bar{V}_{+}^{\pi}$, the control $\mu_s = \mu_{t_j}$ is constant on each interval $I_j = [t_j, t_{j+1})$ of the partition π, but $\nu_s = \beta(\mu.)_s$ is not required to be constant. Later in the section, we consider slightly different versions of the discrete time formulation in which both μ_s and ν_s are constant on I_j. See Lemma 6.5 and its proof.

Remark 6.2. Historical Note. The definition of discrete time upper value function in [14, (11.8.1)] is essentialy due to Nisio [21, Section 4.2.2]. Related definitions of upper and lower value functions were given in [16, Chapter 1] also in terms of partitions of the time interval $[t, T]$. However, Friedman allowed both u and z to be functions of time on each subinterval. In Nisio's approach, the weaker player chooses a constant control on each subinterval. For the upper value, the weaker player is the minimizer.

In [12, Section 3] the concept of *saddle point property* was introduced for differential games. For the LLN game, this is defined as follows. Given $\varepsilon > 0$ a strategy α^ε is called $\varepsilon-$optimal for the minimizer if

$$\sup_{\nu.\in\mathcal{Z}^{\mathcal{P}}(t)} \bar{P}(t, x; \alpha^\varepsilon(\nu.), \nu.) \leq V_M(t, x) + \varepsilon, \tag{6.1}$$

with $\mathcal{Z}^{\mathcal{P}}(t)$ the set of all measurable $\mathcal{P}(Z)-$valued paths $\nu.$ on $[t, T]$. Similarly, a strategy β^ε is $\varepsilon-$optimal for the maximizer if

$$\inf_{\mu.\in\mathcal{U}^{\mathcal{P}}(t)} \bar{P}(t, x; \mu., \beta^\varepsilon(\mu.)) \geq V_M(t, x) - \varepsilon, \tag{6.2}$$

The *saddle point property* holds if such a pair $(\alpha^\varepsilon, \beta^\varepsilon)$ exists for every $\varepsilon > 0$.

Theorem 6.3. *For every $\varepsilon > 0$, there exist $\varepsilon-$optimal strategies $(\alpha^\varepsilon, \beta^\varepsilon)$ which are approximately Markov.*

Proof. The existence of an approximately Markov α^ε is a consequence of the following result [15, formula (2.4)]:

Given $\varepsilon > 0$ and a partition π, there exists α^ε of the desired form

$$\alpha^\varepsilon(\nu.) = \phi_j^\varepsilon(\bar{x}_{t_j}), \quad \text{for} \quad t_j \leq s < t_{j+1},$$

such that for initial time t and state x,

$$\sup_{\nu.\in\mathcal{Z}^{\mathcal{P}}(t)} \bar{P}(t, x; \alpha^\varepsilon(\nu.), \nu.) \leq \bar{V}_{+}^{\pi}(t, x) + \varepsilon/2. \tag{6.3}$$

Recall from Definition 6.1 that we consider only partitions π such that $t = t_i$ for some i.

In [15] the lower differential game value is considered instead of the upper value. To obtain (6.3) from [15, formula (2.4)] the payoff $\bar{P}$ should be replaced by $-\bar{P}$. The construction used in [12] to obtain the functions ϕ_j^ε is very similar to the one described later in this section for a fully discretised version of the LLN game.

Since $\bar{V}_+^\pi$ tends to V_M as $\|\pi\| \to 0$ uniformly on compact sets, (6.1) holds for $\|\pi\|$ sufficiently small.

For the maximizer, $\varepsilon-$optimal β^ε which is approximately Markov is obtained in the same way, by considering the lower discrete time value function $\bar{V}_-^\pi$. $\square$

Remark 6.4. We have used in an essential way the fact that the Isaacs minimax condition holds for the LLN game. Hence both $\bar{V}_+^\pi$ and $\bar{V}_-^\pi$ tend to the same limit V_M as $\|\pi\| \to 0$. In the definition of saddle point property for differential games which do not satisfy the Isaacs condition, the definition of approximately Markov strategy in [12, Section 3] for the stronger player must be changed. The stronger player chooses on each time interval $I_j = [t_j, t_{j+1})$ a control based on both the game state at time t_j and the opponent's control choices on this interval.

Fully discretized LLN game. The remainder of this section is in preparation for Section 7. In the discussion above, the LLN game has been partially discretized. Discrete times t_j are considered, but the dynamics of the LLN game state are still described by the differential equations (5.1). Let us now consider the following fully discretized version of the LLN game.

In order to simplify the calculations, we assume in the rest of this section and in Section 7 that there is no running cost, i.e. $L \equiv 0$. The general case is reduced to this one by considering an *augmented state* $(\bar{x}_s, \bar{x}'_s)$ of dimension $d+1$, where

$$\frac{d}{ds} x'_s = \bar{L}(s, \bar{x}_s, \mu_s, \nu_s), \quad x'_t = 0,$$

and new terminal cost $g(\bar{x}_T) + \bar{x}'_T$.

We shall continue working with a partition π as above. However, we now take controls which are constant (for both the minimizer and maximizer) on each interval $I_j = [t_j, t_{j+1})$ of π. Thus, $\mu_s = \mu_j, \ \ \nu_s = \nu_j$ for $s \in I_j$. The states $y_j = y_{t_j}$ satisfy the dynamics

$$\begin{cases} y_{j+1} = \bar{F}_j(y_j, \mu_j, \nu_j), \quad j = i, i+1, \ldots, N, \quad \text{with} \\ \bar{F}_j(y, \mu, \nu) = y + (t_{j+1} - t_j)\bar{f}(t_j, y, \mu, \nu), \end{cases} \tag{6.4}$$

and initial state $y_i = y_{t_i} = x$. This is an Euler-type scheme, which corresponds to the discrete version of the dynamics (5.1). The game payoff associated to this game is defined as $g(y_N)$. Abusing the notation, let $W_i(x) = \bar{W}_+^\pi(t_i, x)$ be the upper game value function. It is not difficult to verify that these functions are uniformly Lipschitz continuous, with constant Λ and, moreover, the dynamic programming equation holds:

$$\begin{cases} W_j(y_j) = \min_{\mu_j} \max_{\nu_j} W_{j+1}\left[\bar{F}_j(y_j, \mu_j, \nu_j)\right], \\ W_N(y_N) = g(y_N). \end{cases} \tag{6.5}$$

In this fully discretized upper game, the maximizer chooses ν_j knowing the opponent's choise of μ_j. Similarly, in the fully discretized lower game the minimizer chooses μ_j knowing ν_j. The lower value function $\bar{W}_-^\pi$ satisfies the recursive equation corresponding to (6.5) with *minmax* replaced by *maxmim*. As in previous arguments, each $\bar{W}_-^\pi(t_j, \cdot)$ is Lipschitz with constant not depending on j.

Lemma 6.5. *$\bar{W}_+^\pi$ and $\bar{W}_-^\pi$ tend to V_M as $\|\pi\| \to 0$, uniformly on compact sets.*

Proof. In the partially discretized LLN game defined earlier in this section, let $\tilde{V}_+^\pi$ and $\tilde{V}_-^\pi$ denote the upper and lower game values when both μ_j and ν_j are required to be constant on I_j. The *sup* in the definition of $\bar{V}_+^\pi$ is decreased when ν_s is required to be constant on each interval I_j. Hence $\tilde{V}_+^\pi \le \bar{V}_+^\pi$. Similarly, $\bar{V}_-^\pi \le \tilde{V}_-^\pi$. Hence by Theorem 6.3, $\tilde{V}_-^\pi$ and $\tilde{V}_+^\pi$ also tend to V_M as $\|\pi\| \to 0$, uniformly on compact sets.

Given $\mu. \in \mathcal{U}_\pi^{\mathcal{P}}(t)$ and $\nu. \in \mathcal{Z}_\pi^{\mathcal{P}}(t)$, let $\bar{x}_j = \bar{x}_{t_j}$ and $y_j = y_{t_j}$ where $\bar{x}_s$ satisfies (5.1) and y_j satisfies (6.4), with $\bar{x}_i = y_i = x$. It is easily shown that $|y_N - \bar{x}_N| \le C\|\pi\|$, for some constant C. Hence,

$$|g(y_N) - g(\bar{x}_N)| \le C\Lambda_g\|\pi\|,$$

where Λ_g is a Lipschitz constant for g.

Therefore,

$$|W_\pm^\pi(t,x) - \tilde{V}_\pm^\pi(t,x)| \le C\Lambda_g\|\pi\|.$$

□

To conclude this section, let us sketch a (rather standard) construction to describe approximately optimal Markov strategies for the fully discretized LLN game. We construct strategies for the minimizer, based on the upper value function $\bar{W}_+^\pi$. Approximately optimal Markov strategies for the maximzer, based on $\bar{W}_-^\pi$, are constructed in exactly the same way.

Given $\delta > 0$, partition $\mathbb{R}^d$ into Borel sets $A_1,\ A_2, \dots$ each of diameter less than δ. For each $l = 1, 2, \dots$ choose $\xi_l \in A_l$ and

$$\mu_{j,l} \in \arg\min_{\mu_j} \left[\max_{\nu_j} W_{j+1}(\bar{F}_j(\xi_l, \mu_j, \nu_j))\right].$$

By (6.5), for all ν_j,

$$W_{j+1}[\bar{F}_j(\xi_l, \mu_{j,l}, \nu_j)] \le W_j(\xi_l),$$

and the functions $W_j,\ W_{j+1}$ are Lipschitz.

Given $\epsilon > 0$, let $\phi_j^\epsilon(y_j) = \mu_{j,l}$ if $y_j \in A_l$, where δ is chosen small enough such that

$$W_{j+1}[\bar{F}_j(y_j, \phi_j^\epsilon(y_j), \nu_j)] \le W_j(y_j) + \epsilon(t_{j+1} - t_j), \quad \text{for all } \nu_j. \tag{6.6}$$

Since we have assumed running cost $L = 0$, the game payoff for initial time $t = t_i$ and state $x = x_i$ is

$$\bar{P}(t, x; \mu., \nu.) = g(y_N).$$

An induction argument gives

$$\sup_{\nu.} \bar{P}(t, x; \alpha^\epsilon(\nu.), \nu.) \le W^\pi(t,x) + \epsilon(T - t), \tag{6.7}$$

where $\alpha^\epsilon(\nu.)_j = \phi_j^\epsilon(y_j)$. Note that the above display is analogous to (6.1), with $\varepsilon = \epsilon(T - t)$.

Remark 6.6. (1) By discretizing the dynamics (5.1) of the LLN game, solving this ODE stepwise on each interval I_j (with constants controls $\mu_j,\ \nu_j$) is avoided. This procedure should be more convenient for computing numerical solutions to the corresponding dynamic programming equation (6.5).

(2) The spaces of probability measures $\mathcal{P}(U)$ and $\mathcal{P}(Z)$ are infinite dimensional, except when U and Z are finite sets. However, in some cases, U and Z can be replaced by finite sets. In particular, suppose that U, Z are subsets of Euclidean spaces and that $f(t,x,u,z)$ and $L(t,x,u,z)$ are polynomials in u,z with coefficients depending on the state variable x. Then $\bar{f}$ and $\bar{L}$ depend only on the corresponding moments of μ and ν. In general, it may be difficult to describe the constraints on the set of admissible moments, except in special cases.

7. Approximately Optimal Markov Strategies II

In this section we consider a modification of the discretized stochastic game model in Section 4. The choice of game strategies involves two steps. In the first step, an approximately Markov strategy is chosen. It is based on the discretized version of the Law of Large Numbers game considered at the end of Section 6. We call a partition π considered at the first step a "coarse" partition of the time interval $[t,T]$. At the second step, each subinterval of π is divided into a large number m of subintervals of equal length, which form a "fine" partition π_m of $[t,T]$. The players control choices $u_s,\ z_s$ are constants on each subinterval of π_m, obtained by independent random sampling from the probability measures on U and Z previously chosen at step 1.

At an intuitive level, in a mixed strategy differential game, u_s and z_s are to be chosen continuously in time s by random sampling from time-varying measures μ_s and ν_s which are chosen by the players. It seems reasonable to expect that μ_s and ν_s should vary with time much more slowly than the controls $u_s,\ z_s$ which result from repeated random sampling. The modified discrete time model considered here attempts to mimic this intuitive idea.

As in the previous sections, let the subintervals of the coarse partition π be denoted by $I_j = [t_j, _{j+1})$, for $j = i,\ i+1, \cdots, N-1$. The refinement π_m is obtained by dividing each I_j into m subintervals $I_{j,k}$ of equal length, for $k = 1, 2, \ldots, m$, with $\|\pi_m\| = \frac{\|\pi\|}{m}$ tending to zero as $m \to \infty$. In the approach presented next we require that the probability measures chosen by the players on the subintervals $I_{j,k}$ do not depend on k, i.e. $\mu_{j,k} = \mu_j$ and $\nu_{j,k} = \nu_j$. Then the controls chosen by the players are samples $u_s = u_{j,k}$ and $z_s = z_{j,k}$ of those measures, which are constants on the subintervals $I_{j,k}$, for $k = 1, \cdots, m$, of I_j. In other words, m independent random samples are taken from μ_j and ν_j, on each subinterval I_j of the partition π.

Given an initial time $t = t_i$ and state $x = x_i$, let x_s^m be the solution of the original ODE (3.1),

$$\frac{d}{ds} x_s^m = f(s, x_s^m, u_s, z_s), \tag{7.1}$$

with $u_s = u_{j,k},\ z_s = z_{j,k}$ for $s \in I_{j,k}$. Then x_s^m is the state of the modified stochastic game at time s. Let $x_j^m = x_{t_j}^m$. Then x_j^m is the state at step j for the discrete time stochastic game described at the end of the Appendix.

As in Section 6, we take running cost $L = 0$. Hence the stochastic game payoff in (4.6) becomes $J = \mathbb{E}[g(x_N^m)]$. We can write

$$\begin{aligned} x_{j+1}^m &= x_j^m + \sum_{k=1}^m \int_{I_{j,k}} f(s, x_s^m, u_{j,k}, z_{j,k}) ds \\ &= x_j^m + \sum_{k=1}^m \int_{I_{j,k}} f(t_j, x_j^m, u_{j,k}, z_{j,k}) ds + \lambda_j (t_{j+1} - t_j), \end{aligned} \tag{7.2}$$

where $\|\lambda_j\| \leq C_1 \|\pi\| (t_{j+1} - t_j)$ for some constant C_1.

We now recall from the Appendix the formulation of these modified stochastic games. Let $\mathcal{F}_j^m$ be the σ-algebra generated by the controls $u_{l,k},\ z_{l,k}$ for $i \leq l < j, \quad k = 1, \ldots, m$, which can be represented as realizations of random variables $\eta_{l,k}(\omega)$ and $\zeta_{l,k}(\omega)$, respectively. Notice that conditioned on $\mathcal{F}_j^m$, the random variables $\eta_{j,k},\ \zeta_{j,k}$ are independent with distributions $\mu_j,\ \nu_j$, and the controls $\mu_j,\ \nu_j$ are themselves random and $\mathcal{F}_j^m$ measurable. Instead of using the functions F_j^m in (A.9), write (7.2) as

$$\left\{ \begin{array}{l} x_{j+1}^m = \bar{F}_j(x_j^m, \mu_j, \nu_j) + [K_{m,j} + \lambda_j]\,(t_{j+1} - t_j), \quad \text{with} \\ K_{m,j} = \frac{1}{m} \sum_{k=1}^m \left[f(t_j, x_j^m, u_{j,k}, z_{j,k}) - \bar{f}(t_j, x_j^m, \mu_j, \nu_j) \right]. \end{array} \right. \tag{7.3}$$

Here $\bar{F}_j$ is defined as in (6.4) for the discretized LLN game.. When conditioned on $\mathcal{F}_j^m$, the random variable $K_{m,j}$ has mean zero and covariance matrix $\frac{1}{m}\Sigma_{m,j}$, where $\|\Sigma_{m,j}\| \leq C_2^2$, for some constant C_2.

We now consider Markov control strategies for the modified discrete-time stochastic game. As in Definition 6.1, Markov strategies for the minimizer and maximizer are sequences of Borel measurable $\mathcal{P}(U)-$valued and $\mathcal{P}(Z)-$valued functions ϕ_j and ψ_j on $\mathbb{R}^d$. Let $\mathcal{M}_N$ (respectively $\mathcal{N}_N$) denote the set of all sequences of $\mathcal{P}(U)-$valued (respectively $\mathcal{P}(Z)-$valued) random variables $\mu_\cdot = (\mu_i, \cdots, \mu_{N-1})$, $\nu_\cdot = (\nu_i, \cdots, \nu_{N-1})$ such that μ_j and ν_j are $\mathcal{F}_j^m-$measurable. A Markov strategy for the minimizer induces a mapping from $\mathcal{N}_N$ into $\mathcal{M}_N$ as follows. Given $\nu_\cdot \in \mathcal{N}_N$, let $\mu_j = \phi_j(x_j^m)$. Since x_j^m is $\mathcal{F}_j^m-$measurable, $\mu_\cdot \in \mathcal{M}_N$. Similarly, a Markov strategy for the maximizer induces a mapping from $\mathcal{M}_N$ into $\mathcal{N}_N$, by taking $\nu_j = \psi_j(x_j^m)$.

We next consider the functions $W_j(y_j) = \bar{W}_+^\pi(t_j, y_j)$ defined recursively by (6.5). Note that (6.5) holds for all $y_j \in \mathbb{R}^d$, and not merely when $\{y_j\}$ is a sequence satisfying (6.4). In particular, we will take $y_j = x_j^m$. This is not the same as taking $y_j = \bar{x}_j$ in Section 6.

The next lemma considers Markov strategies for the minimizer which are nearly optimal. Given $\epsilon > 0$ we choose ϕ_j^ϵ such that (6.6) holds for all $y_j \in \mathbb{R}^d$ and $\nu_j \in \mathcal{P}(Z)$. The functions W_j are Lipschitz, with Lipschitz constant Λ.

Lemma 7.1. *There exist constants C_1, C_2 such that*

$$\sup_{\nu_\cdot \in \mathcal{N}_N} \mathbb{E}\left[g(x_N^m)\right] \leq W_i(x_i^m) + \left[\epsilon + \Lambda(C_1\|\pi\| + \frac{C_2}{m^{1/2}})\right](T - t), \tag{7.4}$$

when the minimizer chooses $\mu_j = \mu_j^* = \phi_j^\epsilon(x_j^m)$.

Proof. Since W_{j+1} is Lipschitz with constant Λ, we have from (6.6) and (7.3)

$$\begin{aligned} W_{j+1}(x^m_{j+1}) &\leq W_{j+1}(\bar{F}_j(x^m_j, \mu^*_j, \nu_j) + \Lambda\,(|K_{m,j}| + |\lambda_j|)\,(t_{j+1} - t_j) \\ &\leq W_j(x^m_j) + [\epsilon + \Lambda\,(|K_{m,j}| + |\lambda_j|)](t_{j+1} - t_j), \end{aligned} \tag{7.5}$$

where $|\lambda_j| \leq C_1\|\pi\|(t_{j+1} - t_j)$.

Since

$$\mathbb{E}[|K_{m,j}| \parallel \mathcal{F}^m_j] \leq \mathbb{E}[|K_{m,j}|^2 \parallel \mathcal{F}^m_j]^{1/2} \leq \frac{C_2}{m^{1/2}},$$

taking expectations on both sides of (7.5) we get

$$\mathbb{E}[W_{j+1}(x^m_{j+1}) \parallel \mathcal{F}^m_j] \leq W_j(x^m_j) + \left[\epsilon + \Lambda(C_1\|\pi\| + \frac{C_2}{m^{1/2}})\right](t_{j+1} - t_j). \tag{7.6}$$

Since $W_N(x^m_N) = g(x^m_N)$, by induction on j we get (7.4). □

In exactly the same way there exists a Markov strategy $\psi^\epsilon_j,\ \ j = i, \ldots, N-1$, for the maximizer such that

$$\inf_{\mu_\cdot \in \mathcal{M}_N} \mathbb{E}\,[g(x^m_N)] \geq W_i(x^m_i) - \left[\epsilon + \Lambda(C_1\|\pi\| + \frac{C_2}{m^{1/2}})\right](T - t). \tag{7.7}$$

From Lemmas 6.5 and 7.1 we obtain the following main result. As in (4.6) with $L = 0$, we denote the payoff for the modified stochastic game by $J(t, x; \mu_\cdot, \nu_\cdot) = \mathbb{E}[g(x^m_N)]$.

Theorem 7.2. *Given $\varepsilon > 0$, there exist positive constants ϵ, b and integer m_0 such that the following saddle point property holds, provided that $\|\pi\| \leq b$ and $m \geq m_0$. Choose ϕ^ϵ_j, ψ^ϵ_j as in (7.4) and (7.7), and let $\mu^*_j = \phi^\epsilon_j(x^m_j)$, $\ \nu^*_j = \psi^\epsilon_j(x^m_j)$. Then*

$$\sup_{\nu_\cdot \in \mathcal{N}_N} J(t, x; \mu^*_\cdot, \nu_\cdot) \leq V_M(t, x) + \varepsilon$$

and

$$\inf_{\mu_\cdot \in \mathcal{M}_N} J(t, x; \mu_\cdot, \nu^*_\cdot) \geq V_M(t, x) - \varepsilon.$$

Appendix A. Discrete Time Markov Stochastic Games

In this section we describe in general form discrete time mixed strategy Markov games, motivated by the stochastic game associated with the function V^π defined in (4.5), which is obtained as a particular case of these games. Here we simplify the notation, writing i instead of $t_i \in \pi$ and x_j by x_{t_j}.

Let us assume that the dynamics of the state variable x_j are given by the difference equations

$$\begin{cases} x_{j+1} = F_j(x_j, u_j, z_j), \text{for } j = i, i+1, \cdots, N-1 \\ x_i = x, \end{cases} \tag{A.1}$$

with $x \in \mathbb{R}^d,\ u_j \in U,\ z_j \in Z$; the control sets U and Z are compact subsets of metric spaces. Here u_j and z_j are randomly (simultaneous and independently) chosen with probability distributions μ_j and ν_j, respectively. These probability distributions play the role of controls for each player, and we assume that at time step j, both players know the previous realizations $u_k,\ z_k$ for $i \leq k < j$. The game payoff, for initial time $t = t_i$ and state $x = x_i$, is

$$J(t, x; \mu., \nu.) = \mathbb{E}\{\sum_{j=i}^{N-1} L_j(x_j, u_j, z_j) + g(x_N)\}. \tag{A.2}$$

We shall assume that $F_j,\ L_j$ are continuous and Lipschitz in the state variable x, uniformly with respect to (u, z).

Remark A.1. For the stochastic game model in Section 4, $x_{j+1} = x_{t_{j+1}}$, where x_s is the solution to (3.1) with $x_j = x_{t_j}$ as initial data and $u_s = u_j,\ z_s = z_j$. This defines F_j:

$$F_j(x_j, u_j, z_j) = x_j + \int_{t_j}^{t_{j+1}} f(s, x_s, u_j, z_j) ds,$$

and the "running cost" function L_j is:

$$L_j(x_j, u_j, z_j) = \int_{t_j}^{t_{j+1}} L(s, x_s, u_j, z_j) ds.$$

In order to give a rigorous definition to the value of this game, we first fix the initial condition x_i, and define the information vector available to the players before taking decisions at time j as follows. For $j = i$, only x_i is known. For $i < j < N$, the information vector is $h_j = (u_\cdot^j, z_\cdot^j)$, where

$$u_\cdot^j = (u_i, u_{i+1}, \ldots, u_{j-1}), \quad z_\cdot^j = (z_i, z_{i+1}, \ldots, z_{j-1}).$$

Let $\mathcal{H}_j$ denote the set of all such h_j. For the stochastic game model, we choose $\mathcal{H}_N$ as the "canonical" sample space, with elements $\omega = h_{N-1}$, and let $\mathcal{F}_j = \sigma(h_j)$ denote the $\sigma-$algebra generated by the information available to both players at step j. A *decision strategy* $\eta = (\eta_i, \ldots, \eta_{N-1})$ for the minimizer is a $\mathcal{P}(U)-$valued discrete time stochastic process such that η_j is $\mathcal{F}_j-$measurable. Decision strategies ζ for the maximizer are defined similarly, as $\mathcal{P}(Z)-$valued stochastic processes.

Given an information vector h_j, the controls u_j, z_j are chosen by independent random sampling from the probability measures $\mu_j = \eta_j(h_j),\ \nu_j = \zeta_j(h_j)$. Thus, for Borel sets $A \in \mathcal{B}(U),\ B \in \mathcal{B}(Z)$,

$$Pr\,[u_j \in A,\ z_j \in B \parallel h_j] = \mu_j(A)\nu_j(B). \tag{A.3}$$

A pair (η, ζ) of decision strategies, together with the initial state x, uniquely determines a probability measure $\mathbb{P}_x$ on $\mathcal{H}_N$ such that $\mathbb{P}_x[x_i = x] = 1$. We also denote the payoff J in (A.2) by $P(t, x; \eta, \zeta) = J(t, x; \mu., \nu.)$, where $\mu_j = \eta_j(h_j)$ and $\nu_j = \zeta_j(h_j)$ as above. We also note that in (A.2)

$$\mathbb{E}\,[L_j(x_j, u_j, z_j) \parallel h_j] = \int_U \int_Z L_j(x_j, u, z)\nu_j(dz)\mu_j(du).$$

We say that a decision strategy η is a Markov control strategy for the minimizer if there exists a finite sequence of Borel measurable functions on $\mathbb{R}^d$ taking values in $\mathcal{P}(U) : \phi_i,\ \phi_{i+1}, \ldots, \phi_{N-1}$, such that $\eta_j(h_j) = \phi_j(x_j)$. Similarly, ζ is a Markov control strategy for the maximizer if $\zeta_j(h_j) = \psi_j(x_j)$ where ψ_j is Borel measurable and $\mathcal{P}(Z)-$valued.

To define the value of this stochastic game, we first define functions $W(i,x)$ recursively backward in time, in an analogous way as $V^\pi(t,x)$ was defined in (4.5):

$$\begin{cases} W(j,x_j) = val_{u_j,z_j}\left[W(j+1,x_{j+1}) + L_j(x_j,u_j,z_j)\right] \\ W(N,x_N) = g(x_N), \end{cases} \tag{A.4}$$

where x_{j+1} satisfies (A.1). Observe that the *value* appearing in the right side of (A.4) (and similarly in (4.8)) is in fact a conditional expectation with respect to the vector information $h_j \in \mathcal{H}_j$ available at each period of time j:

$$W(j,x_j) = val_{u_j,z_j}\left[\mathbb{E}(W(j+1,x_{j+1}) \parallel h_j) + L_j(x_j,u_j,z_j)\right]. \tag{A.5}$$

Moreover, we note that the sequence of functions defined as $W_j(x_j) = W(j,x_j)$, following the backward recursion (A.4), are Lipschitz continuous functions of x_j for $j = i,\ldots,N$. For the stochastic game model in Section 4, $W_j(x_j)$ is the same as $V^\pi(t_j,x_j)$.

We call $W(i,x_i)$ the value $val(J)$ of the discrete time stochastic game with payoff $P(t,x;\zeta,\eta) = J(t,x;\mu.,\nu.)$. The following can be proved in a way very similar to the discussion of the saddle point property in Section 6.

Theorem A.2. *For every $\varepsilon > 0$, there exists a pair of decision strategies η^ε, ζ^ε such that*

$$\sup_\zeta P(t,x;\eta^\varepsilon,\zeta) \quad \leq \quad W(i,x_i) + \varepsilon \tag{A.6}$$

$$\inf_\eta P(t,x;\eta,\zeta^\varepsilon) \quad \geq \quad W(i,x_i) - \varepsilon. \tag{A.7}$$

In fact, η^ε and ζ^ε can be obtained from Markov control policies: $\mu_j = \phi_j^\varepsilon(x_j)$, $\nu_j = \psi_j^\varepsilon(x_j)$, $j = i,\ldots,N-1$

The proof of Theorem A.2 uses standard properties of conditional expectations with respect to the $\sigma-$algebras $\mathcal{F}_j$, as well as a construction similar to that at the end of Section 6 to obtain ϕ_j^ε and ψ_j^ε.

Remark A.3. When $W_j(\cdot) = V^\pi(t_j,\cdot)$, the Markov control policy formulation does not give the uniform Lipschitz estimate (4.9), which does not depend on π.To avoid this difficulty we introduced decision strategies, which depend not on the current state but on information about the past controls. Summarizing, the formulation of V^π presented above as a stochastic game value, gives the desired uniform Lipschitz constant M in (4.9).

Equation (A.4) can be regarded as a one step dynamic programming principle for the discrete time stochastic game. More generally, there is the following multistep dynamic programming principle:

$$W(i,x_i) = val_{\eta,\zeta}\ \mathbb{E}\left[\sum_{l=i}^{j-1} L_l(x_l,u_l,z_l) + W(j,x_j)\right]. \tag{A.8}$$

Modified stochastic games. In Section 7 we considered a modified version of the discrete time stochastic games in Section 4, with running cost $L = 0$ to simplify notations. The game dynamics take the form

$$x_{j+1}^m = F_j^m(x_j^m, u_j^m, z_j^m), \tag{A.9}$$

where F_j^m is defined by the right side of (7.2) with $u_j^m = (u_{j,1}, \ldots, u_{j,m})$, $z_j^m = (z_{j,1}, \ldots, z_{j,m})$. The information vector at time step j is $h_j^m = \{u_{l.k},\ z_{l,k} \parallel i \leq l < j,\ k = 1, \ldots, m\}$. Let $\mathcal{H}_j^m$ be the space of all such h_j^m. The canonical sample space in now $\mathcal{H}_N^m$.

Given sets $A_k,\ B_k$ in $\mathcal{B}(U),\ \mathcal{B}(Z)$ respectively for $k = 1, \ldots, m$, let $A^m = A_1 \times \cdots \times A_m,\ B^m = B_1, \times \cdots \times B_m$. Corresponding to (A.3) we require the conditional independence condition

$$Pr\left[u_j^m \in A^m,\ z_j^m \in B^m \parallel h_j^m\right] = \prod_{k=1}^{m} [\mu_j(A_k)\nu_j(B_k)].$$

This corresponds to choosing $u_{j,k},\ z_{j,k}$ by random sampling from the distributions $\mu_j,\ \nu_j$, independently conditioned on $\mathcal{F}_j^m = \sigma(h_j^m)$. The discussion of the modified stochastic game then continues as for the case $m = 1$ described above.

References

1. Aubin, J. P.: *Mathematical Methods of Game and Economic Theory*, North-Holland, Amsterdam, 1979.
2. Bardi, M.: Capuzzo-Dolcetta I.: *Optimal control and viscosity solutions of Hamilton–Jacobi–Bellman equations*, Birkhauser, Boston, 1997.
3. Berkovitz, L. D.: *Optimal Control Theory*, Springer-Verlag, New York, 1974.
4. Berkovitz, L. D. and Medhin, N. G.: *Nonlinear Optimal Control Theory*, CRC Press-Taylor and Francis, 2013.
5. Buckdahn, R., Li, J., and Quincampoix, M.: Value function of differential games without Isaacs condition: An approach with nonanticipative mixed strategies, *Internat. J. Game Theory* **42** (2013), 989–1020.
6. Buckdahn, R., Li, J., and Quincampoix, M.: Value in mixed strategies for zero sum stochastic differential games without the Isaacs condition, *Ann. Probab.* **42** (2014), 1724–1768.
7. Crandall, M. G. and Lions, P. L.: Viscosity solutions of Hamilton–Jacobi equations, *Trans. Amer. Math. Soc.* **277** (1983), 1–42.
8. Elliott, R. J. and Kalton, N. J.: The existence of value in differential games, *Memoirs Amer. Math. Soc.* No. 126, 1972.
9. Ethier, S. N. and Kurtz, T. G.: *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York, 1986.
10. Evans, L. C. and Souganidis, P. E.: Differential games and representation formulas for solutions of Hamilton–Jacobi–Isaacs equations, *Indiana Univ. Math. J.* **33** (1984), 773–797.
11. Fleming, W. H.: The convergence problem for differential games II, in: *Advances in Game Theory, Ann. of Math. Studies* **52** (1964), 195–210. Princeton, New Jersey, Princeton Univ. Press.
12. Fleming, W. H. and Hernández-Hernández, D.: On the value of stochastic differential games, *Comm. Stochastic Analysis* **5** (2011), 341–351.
13. Fleming, W. H. and Hernández-Hernández, D.: Strategies for differential games, in: *Advances in Statistics, Probability and Actuarial Science* (2012), 89–104, Editors: S. Cohen and K. Siu, Scientific World.
14. Fleming, W. H. and Soner, H. M.: *Controlled Markov Processes and Viscosity Solutions.* Springer-Verlag, New York, Second Edition 2006.
15. Fleming, W. H. and Souganidis, P. E.: On the existence of value functions of two player, zero sum stochastic differential games, *Indiana Univ. Math. J.* **38** (1989), 293–314.
16. Friedman, A.: *Differential Games*, Wiley, 1971.
17. Hernández-Hernández, D. and Sirbu, M.: Zero-sum stochastic differential games without the Isaacs condition: random rules of priority and intermediate Hamiltonians, *Arxiv*, (2016).
18. Isaacs, R.: *Differential Games*, Wiley, New York, 1965.

19. Kaise, H. and Sheu, S.-J.: Differential games of inf-sup type and Isaacs equations, *Applied Math Optim.* **52** (2005), 1–22.
20. Krasovskii, N. N. and Subbotin, A. I.: *Game-Theoretic Control Problems*, Springer-Verlag, New York, 1988.
21. Nisio, M.: *Stochastic Control Theory*, Springer, New York, 2015.
22. Sirbu, M.: On martingale problems with continuous-time mixing and values of zero-sum games without Isaacs condition, *SIAM J. on Control and Optim.* **52** (2014), 2877–2890.
23. Souganidis, P. E.: Approximation schemes for viscosity solutions of Hamilton–Jacobi equations with applications to differential games, *J. Nonlinear Anal. T.M.A.* **9** (1985), 217–257.

Wendell H. Fleming: Division of Applied Mathematics, Brown University, Box F, Providence, R.I. 02912, USA

E-mail address: wendell_fleming@brown.edu

Daniel Hernández–Hernández: Centro de Investigación en Matemáticas, Apartado Postal 402, Guanajuato, Gto. 36000, México

E-mail address: dher@cimat.mx