Dyn Games Appl
https://doi.org/10.1007/s13235-018-0268-4

Stochastic Differential Games: A Sampling Approach via FBSDEs

Ioannis Exarchos[1] · **Evangelos Theodorou**[1] · **Panagiotis Tsiotras**[1]

Abstract The aim of this work is to present a sampling-based algorithm designed to solve various classes of stochastic differential games. The foundation of the proposed approach lies in the formulation of the game solution in terms of a decoupled pair of forward and backward stochastic differential equations (FBSDEs). In light of the nonlinear version of the Feynman–Kac lemma, probabilistic representations of solutions to the nonlinear Hamilton–Jacobi–Isaacs equations that arise for each class are obtained. These representations are in form of decoupled systems of FBSDEs, which may be solved numerically.

Keywords Stochastic differential games · Forward and backward stochastic differential equations · Numerical methods · Iterative algorithms

1 Introduction

The origin of differential games dates back to the work of Isaacs [28]. Isaacs provided a framework for the treatment of differential games for two strictly competitive players. He associated the solution of a differential game with the solution to a HJB-like equation, namely its min-max extension, also known as the Isaacs (or Hamilton–Jacobi–Isaacs, HJI) equation. This equation was derived heuristically under the assumptions of Lipschitz continuity of the cost and the dynamics, in addition to the assumption that both of them are *separable* in terms of the maximizing and minimizing controls. Berkovitz [5] addressed differential games using standard variational techniques, a framework which was later adopted by Bryson et al.

✉ Ioannis Exarchos
exarchos@gatech.edu

Evangelos Theodorou
evangelos.theodorou@gatech.edu

Panagiotis Tsiotras
tsiotras@gatech.edu

[1] Department of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0150, USA

Published online: 11 June 2018

[25] to treat a special case of differential games, namely games of pursuit and evasion. A treatment of the stochastic extension of differential games was first provided by Kushner and Chamberlain [34]. The authors of this reference provide a general definition of stochastic differential games and derive the underlying PDE, which is similar to the one derived by Isaacs, adjusted by a term owing to stochastic effects. They also present sufficient conditions for the existence of a saddle point and propose a finite difference scheme as a numerical procedure to solve the game. A series of papers exist investigating conditions for existence and uniqueness of a value function in stochastic two-player zero-sum games, see for example [8,10,20,24].

Despite the plethora of existing theoretic work in the area of differential games, the algorithmic part has received significantly less attention. However, since the game formulation leads to a HJI nonlinear PDE, several methods addressing the solution of general nonlinear PDEs should be mentioned here. A most notable example is the seminal paper investigating convergence of general numerical schemes by Barles and Souganidis [2]. Due to the inherent difficulty in solving such problems, most of the efforts were focused on particular types of PDEs, such as the Hamilton–Jacobi–Bellman (HJB) PDE in optimal control theory; therein, several different methods and approaches exist, including [3,35,42] for deterministic control problems, while a stochastic setting is considered in [22,26,27]. General PDE cases, that also include the HJI PDE, are furthermore treated within the framework that arises through the interplay between PDEs and forward and backward stochastic differential equations (FBSDEs). This interplay is exploited in both ways; for low-dimensional problems and smooth FBSDE coefficients, the PDE problem is solved in lieu of the FBSDE problem, in order to obtain a solution to the latter (see for example [13,18,40,41,43]). In contrast, for higher-dimensional cases, the FBSDE problem is solved in order to obtain a PDE solution. The literature in this latter case is very extensive, and a comprehensive coverage of it within the limits of this introduction is rather difficult; we refer the reader to [6,56] for the most established numerical schemes. (More details on their error analysis can be found in [7,21].) The case of quadratic growth (see Remark 2 in Sect. 4.2) is also treated in [9,11,32,37]. A comprehensive treatment on the literature of FBSDEs is also included in the recently published book by Zhang [57]. With respect to these results, the fundamental difference in our proposed approach is the iterative character of our numerical scheme. Most previous publications focus on the theoretic properties of the schemes and overlook their weaknesses during the actual implementation. One particular such weakness is obtaining enough sample trajectories in the vicinity of the game/state space that encompasses the sought-for optimal trajectory solution. In fact, without some form of iterative accuracy improvement through importance sampling, as it is the case in our method, none of these previously existing algorithms can be successfully applied in problems of more complex structure/dynamics.

As far as algorithms that are specifically tailored, or are suitably modified, for differential games are concerned, apart from results addressing special cases of differential games (such as linear games with quadratic penalties, e.g., [14]), only but a few numerical approaches have been suggested in the past, notably the Markov chain approximation method [33,51]. In general, however, these numerical procedures have found only limited applicability due to the "curse of dimensionality." Very recently, another specific class of minimax control trajectory optimization methods has been proposed, based on the foundations of *differential dynamic programming* (DDP) [44,45,52].

In this work, we focus on stochastic differential games in which the control effort of each player is penalized using either an $\mathcal{L}^2$ type or $\mathcal{L}^1$ type of norm. In engineering, those penalties are often related to *energy*, or *fuel* expenditure, respectively. By and large, the literature on optimal control deals with the minimization of a performance index which penalizes control

energy. Such $\mathcal{L}^2$ minimization problems have been studied extensively, in both deterministic and stochastic settings [3,15,26,29,42,53]. Their widespread use is due to the fact that they simplify the associated Hamilton–Jacobi–Bellman equation by imposing a structure that facilitates desirable properties for optimization.

While $\mathcal{L}^2$ minimization can be useful in addressing several optimal control problems in engineering (e.g., preventing engine overheating, avoiding high frequency control input signals), there are practical applications in which the control input is bounded (e.g., due to actuation constraints), and the $\mathcal{L}^1$ norm is a more suitable choice to penalize. These problems are also called *minimum fuel* problems, due to the nature of the running cost, which involves an integral of the absolute value of the input signal. Minimum fuel control appears as a necessity in several settings, especially in spacecraft guidance and control [12,50], in which fuel is a limited resource. The notion of $\mathcal{L}^1$-optimal control is also tightly related to *Maximum Hands-Off control* [46,47]. The distinguishing characteristic of a hands-off controller is that it tries to retain a zero control input value over an extended time interval. Thus, the objective of "maximum hands-off" control is to accomplish a specific task while applying zero input for the longest time duration possible. The "hands-off" property, especially in a discrete context, is equivalent to *sparsity* of the signal, i.e., minimizing the total length of intervals over which the signal takes nonzero values. The relationship between $\mathcal{L}^1$-optimality and the "hands-off", or sparsity, property is shown in [46,47].

In this paper, we propose an algorithm for solving stochastic differential games, which employs a nonlinear Feynman–Kac-type formula. The algorithm is a sampling-based scheme based on the theory of forward and backward stochastic differential equations (FBSDEs) and their connection to backward PDEs [39,55]; a probabilistic representation of the HJI partial differential equation is obtained in the form of a system of FBSDEs, which is then simulated by employing linear regression.

The paper is organized as follows: In Sect. 2, we provide the mathematical formulation of the game, with a presentation of the associated HJI equations that arise for each cost form following in Sect. 3. Section 4 introduces some necessary background on FBSDE theory, the nonlinear Feynman–Kac lemma, and demonstrates how the HJI equations of Sect. 3 satisfy the requirements for this lemma, under a certain decomposability condition. This allows for a probabilistic representation of the solution to the HJI equation using FBSDEs. Section 5 generalizes the framework to address games in which the time duration is not fixed and specified a priori; instead, the game terminates as soon as a particular game state has been reached (also known as *terminal surface*), with an upper bound on the duration of the game. Section 6 presents the numerical scheme employed in the process of solving FBSDEs, while Sect. 7 introduces importance sampling in order to refine the results in an iterative manner. A few remarks concerning the various error sources in the numerical scheme are given in Sect. 8. Finally, Sect. 9 offers a few application examples, while conclusions are presented in Sect. 10.

2 Game Formulation

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete, filtered probability space on which a p-dimensional standard Brownian motion W_t is defined, such that $\{\mathcal{F}_t\}_{t\geq 0}$ is the natural filtration of W_t augmented by all $\mathbb{P}$-null sets. Consider a differential game-theoretic setting, in which the expected game payoff is defined by the functional

$$P(u(\cdot), v(\cdot);\ \tau, x_\tau) = \mathbb{E}\left[g(x(T)) + \int_\tau^T q(t, x(t)) + L^u(u(t)) - L^v(v(t))\mathrm{d}t \right], \tag{1}$$

where $T > \tau \geq 0, T$ is a fixed time of termination (games in which the duration is not fixed a priori will be addressed in Sect. 5) , and $x \in \mathbb{R}^n$ represents the game state vector. The minimizing player seeks to minimize the payoff by controlling the vector $u \in \mathcal{U} \subset \mathbb{R}^\nu$, while the maximizing player seeks to maximize the payoff (1) by controlling the vector $v \in \mathcal{V} \subset \mathbb{R}^\mu$. The functions $g(\cdot)$ and $q(\cdot)$ represent a terminal payoff and a state-dependent running payoff, respectively, while $L^u(\cdot)$ and $L^v(\cdot)$ represent the penalties paid by the minimizing and maximizing player, respectively. It is assumed that the payoff functional is either of $\mathcal{L}^2$ type (minimum energy) or of $\mathcal{L}^1$ type (minimum fuel), that is, the functions L^u and L^v satisfy either one of the following two forms:

$$\mathcal{L}^2: \quad L(s) = \frac{1}{2} s^\top R s,$$
$$\mathcal{L}^1: \quad L(s) = p^\top |s|,$$

where R is a positive definite real-valued matrix, p is a vector of positive weights, and $|\cdot|$ denotes the element-wise absolute value. The game state obeys the dynamics of a stochastic controlled system which is represented by the Itô stochastic differential equation (SDE)

$$\begin{cases} \mathrm{d}x(t) = f(t, x(t))\mathrm{d}t + G(t, x(t))u(t)\mathrm{d}t + B(t, x(t))v(t)\mathrm{d}t + \Sigma(t, x(t))\mathrm{d}W_t, \\ t \in [\tau, T], \qquad x(\tau) = x_\tau, \end{cases} \tag{2}$$

in which $\mathrm{d}W_t$ are standard Brownian motion increments.

2.1 Standing Assumptions

Herein we assume that $g : \mathbb{R}^n \to \mathbb{R}, q : [\tau, T] \times \mathbb{R}^n \to \mathbb{R}, f : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n, G : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times\nu}, B : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times\mu}$ and $\Sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times p}$ are deterministic functions, that is, they do not depend explicitly on $\omega \in \Omega$. Furthermore, in order to guarantee existence and uniqueness of solutions to (2), and a well-defined payoff (1), the following conditions hold:

1. The functions g, q, f, G, B and Σ are continuous with respect to time t (in case there is explicit dependence), Lipschitz (uniformly in t) with respect to the state variables, and satisfy standard growth conditions over the domain of interest; namely, for a function $f(t, x)$, the last condition imposes that there exists $C \in \mathbb{R}_+$ such that $\|f(t, x)\| \leq C(1 + \|x\|)$ for all (t, x) in the domain of f.
2. The square-integrable[1] processes $u : [0, T] \times \Omega \to \mathcal{U} \subset \mathbb{R}^\nu$ and $v : [0, T] \times \Omega \to \mathcal{V} \subset \mathbb{R}^\mu$ are $\{\mathcal{F}_t\}_{t\geq 0}$-adapted, which essentially translates into the control inputs being non-anticipating, i.e., relying only on past and present information.
3. If the control penalty for the maximizing or minimizing player is of the $\mathcal{L}^2$ type, then $\mathcal{U}$ and/or $\mathcal{V}$ can be any compact subsets of $\mathbb{R}^\nu$ and $\mathbb{R}^\mu$, respectively. Otherwise, for an $\mathcal{L}^1$ type of penalty, the respective domain is a compact subset of the form $\mathcal{U} = [-u_1^{\min}, u_1^{\max}] \times [-u_2^{\min}, u_2^{\max}] \times \cdots \times [-u_\nu^{\min}, u_\nu^{\max}]$, with $u_i^{\min} \geq 0, u_i^{\max} > 0$, and similarly for $\mathcal{V}$. Note that the assumption about the signs of $u_i^{\min}$ and $u_i^{\max}$ is without loss of generality. The subsequent analysis can be performed for any $u_i^{\min} < u_i^{\max}$ regardless of their sign. In this setting, $p^\top |s|$ represents a positively weighted summation of the

[1] A process H_s is called square-integrable if $\mathbb{E}\left[\int_t^T H_s^2 \mathrm{d}s\right] < \infty$ for any $T > t$.

element-wise absolute values of the control input. If the "fuel consumption" penalty is to be applied on all control channels equally, then p reduces to a vector of ones. Note that one could also consider a time-/state-dependent weight vector $p(t, x)$, without modifying the analysis significantly.

The intuitive idea behind the game-theoretic setting is the existence of two players of conflicting interests. The first player controls u and wishes to minimize the payoff P over all choices of v, while the second player wishes to maximize P over all choices of u of his opponent. At any given time, the current state is known to both players, and instantaneous switches in both controls are permitted, rendering the problem difficult to solve, in general.

3 The Value Function and the HJI Equation

Given any initial condition (τ, x_τ), the goal is to solve the game of conflicting control actions u, v that maximize (1) under all admissible non-anticipating strategies assigned to $v(\cdot)$, while minimizing (1) over all admissible non-anticipating strategies assigned to $u(\cdot)$. The structure of this problem, due to the form of the dynamics and cost at hand, satisfies the Isaacs condition[2] [19,20,28,49], and the payoff is a saddle point solution to the following terminal value problem of a second-order partial differential equation, known as the Hamilton–Jacobi–Isaacs (HJI) equation

$$\begin{cases} V_t + \inf\limits_{u\in\mathcal{U}} \sup\limits_{v\in\mathcal{V}} \left\{ \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top(f + Gu + Bv) + q + L^u(u) - L^v(v) \right\} = 0, \\ (t, x) \in [0, T) \times \mathbb{R}^n, \qquad V(T, x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{3}$$

Herein, V_x and V_{xx} denote the gradient and the Hessian of V, respectively. The term within the brackets is the Hamiltonian. Depending on the form of $L^u(u)$ and $L^v(v)$, we distinguish three cases; a) both cost terms are of $\mathcal{L}^2$ type, b) both terms are of $\mathcal{L}^1$-type, and c) mixed $\mathcal{L}^2$, $\mathcal{L}^1$-type cost terms. We shall investigate each case separately in what follows.

3.1 Case I: $\mathcal{L}^2$–$\mathcal{L}^2$

Let $L^u(u) = \frac{1}{2}u^\top R_u u$ and $L^v(v) = \frac{1}{2}v^\top R_v v$, with u and v taking values in $\mathcal{U} \subset \mathbb{R}^\nu$ and $\mathcal{V} \subset \mathbb{R}^\mu$, respectively. Assuming that the optimal controls lie in the interiors of $\mathcal{U}$ and $\mathcal{V}$, we may carry out the infimum and supremum operations in (3) explicitly, by taking the gradient of the Hamiltonian with respect to u and v and setting it equal to zero to obtain

$$\begin{aligned} R_u u + G^\top(t, x)V_x(t, x) &= 0, \\ -R_v v + B^\top(t, x)V_x(t, x) &= 0. \end{aligned}$$

Therefore, for all $(t, x) \in [0, T] \times \mathbb{R}^n$, the optimal controls are given by

$$u^*(t, x) = -R_u^{-1}G^\top(t, x)V_x(t, x), \tag{4}$$

$$v^*(t, x) = R_v^{-1}B^\top(t, x)V_x(t, x). \tag{5}$$

[2] The Isaacs condition renders the viscosity solutions of the upper and lower value functions equal, thus making the order of maximization/minimization inconsequential.

Inserting the above expression back into HJI equation (3) and suppressing function arguments for notational brevity, we obtain the equivalent characterization

$$\begin{cases} V_t + \dfrac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + q - \dfrac{1}{2}V_x^\top\left(GR_u^{-1}G^\top - BR_v^{-1}B^\top\right)V_x = 0, \\ (t,x) \in [0,T)\times\mathbb{R}^n, \qquad V(T,x) = g(x), \quad x\in\mathbb{R}^n. \end{cases} \tag{6}$$

3.2 Case II: $\mathcal{L}^1$–$\mathcal{L}^1$

Let $L^u(u) = p_u^\top|u|$ and $L^v(v) = p_v^\top|v|$, with u and v taking values in $\mathcal{U} = [-u_1^{\min}, u_1^{\max}]\times[-u_2^{\min}, u_2^{\max}]\times\cdots\times[-u_\nu^{\min}, u_\nu^{\max}]$, and $\mathcal{V} = [-v_1^{\min}, v_1^{\max}]\times[-v_2^{\min}, v_2^{\max}]\times\cdots\times[-v_\mu^{\min}, v_\mu^{\max}]$, respectively. Then, HJI equation (3) can be written as

$$\begin{cases} V_t + \inf\limits_{u\in\mathcal{U}}\sup\limits_{v\in\mathcal{V}}\Big\{\dfrac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + \big(V_x^\top G + p_u^\top \mathrm{D}(\mathrm{sgn}(u))\big)u \\ \qquad\qquad + \big(V_x^\top B - p_v^\top \mathrm{D}(\mathrm{sgn}(v))\big)v + q_0\Big\} = 0, \quad (t,x)\in[0,T)\times\mathbb{R}^n, \\ V(T,x) = g(x), \quad x\in\mathbb{R}^n, \end{cases} \tag{7}$$

in which $\mathrm{D}(x)\in\mathbb{R}^{n\times n}$ denotes the diagonal matrix with the elements of $x\in\mathbb{R}^n$ in its diagonal, and $\mathrm{sgn}(\cdot)$ denotes the signum function.

Again, we may carry out the infimum and supremum operations over u and v explicitly. To this end, let u_i be the i-th element of u and consider the following cases:

- Case $u_i > 0$, that is, $\mathrm{sgn}(u_i) = +1$. Then, if $(V_x^\top G)_i + (p_u^\top)_i > 0$, the Hamiltonian is minimized for $u_i = -u_i^{\min} \le 0$, which leads to a contradiction. On the other hand, if $(V_x^\top G)_i + (p_u^\top)_i < 0$, the Hamiltonian is minimized for $u_i = u_i^{\max} > 0$, which is consistent with the hypothesis.
- Case $u_i < 0$, that is, $\mathrm{sgn}(u_i) = -1$. This is a valid case if $-u_i^{\min}$ is strictly less than zero. Then, if $(V_x^\top G)_i - (p_u^\top)_i < 0$, the Hamiltonian is minimized for $u_i = u_i^{\max} > 0$ which leads to a contradiction. On the other hand, if $(V_x^\top G)_i - (p_u^\top)_i > 0$, the Hamiltonian is minimized for $u_i = -u_i^{\min} < 0$, which is consistent with the hypothesis.

Thus, the optimal control law for the minimizing player is given by

$$u_i^* = \begin{cases} u_i^{\max}, & \left(V_x^\top G\right)_i < -\left(p_u^\top\right)_i \\ -u_i^{\min}, & \left(V_x^\top G\right)_i > \left(p_u^\top\right)_i \\ 0, & -\left(p_u^\top\right)_i < \left(V_x^\top G\right)_i < \left(p_u^\top\right)_i \end{cases}, \qquad i = 1,\ldots,\nu, \tag{8}$$

namely the optimal control law turns out to be *bang-off-bang* control. A similar analysis for the supremum yields the optimal control law for the maximizing player:

$$v_i^* = \begin{cases} v_i^{\max}, & \left(V_x^\top B\right)_i > \left(p_v^\top\right)_i \\ -v_i^{\min}, & \left(V_x^\top B\right)_i < -\left(p_v^\top\right)_i \\ 0, & -\left(p_v^\top\right)_i < \left(V_x^\top B\right)_i < \left(p_v^\top\right)_i \end{cases}, \qquad i = 1,\ldots,\mu, \tag{9}$$

Remark 1 We note that the control laws given by (8)–(9) are not uniquely defined whenever $(V_x^\top G)_i = -(p_u^\top)_i$ or $(V_x^\top G)_i = (p_u^\top)_i$ (and similarly for v), as any value in $[0, u_i^{\max}]$ and $[-u_i^{\min}, 0]$, respectively, attains the same infimum value in (7). A problem in which either one of these equalities is satisfied over a non-trivial time interval is a *singular* fuel-optimal problem [1]. In this work, we shall assume that the minimum fuel problem is *normal*, in the sense that the aforementioned equalities are not satisfied over a non-trivial time interval.

We may insert the optimal control laws (8)–(9) back into HJI equation (7), to obtain the equivalent expression

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}\left(V_{xx}\Sigma\Sigma^\top\right) + V_x^\top f + q \\ \qquad + \sum_{i=1}^{\nu} \min\left\{(V_x^\top G + p_u^\top)_i u_i^{\max},\ 0,\ -(V_x^\top G - p_u^\top)_i u_i^{\min}\right\} \\ \qquad + \sum_{i=1}^{\mu} \max\left\{(V_x^\top B - p_v^\top)_i v_i^{\max},\ 0,\ -(V_x^\top B + p_v^\top)_i v_i^{\min}\right\} = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad V(T,x) = g(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{10}$$

that is, the min and max operations are performed over three values for each control channel.

3.3 Case III: Mixed $\mathcal{L}^2$–$\mathcal{L}^1$

As it is evident from the previous two cases, each player's optimality analysis is done independently. Thus, we may combine the analysis performed in the two previous cases and consider a third case in which one player pays a $\mathcal{L}^2$-type penalty, while the other pays an $\mathcal{L}^1$ type. For example, the case in which the minimizing player is subject to an $\mathcal{L}^2$-type penalty, while the maximizing player is subject to an $\mathcal{L}^1$ type would yield control laws (4) and (9) for the minimizing and maximizing player, respectively, while the HJI equation would assume the form

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + q - \frac{1}{2}V_x^\top G R_u^{-1} G^\top V_x \\ \qquad + \sum_{i=1}^{\mu} \max\left\{(V_x^\top B - p_v^\top)_i v_i^{\max},\ 0,\ -(V_x^\top B + p_v^\top)_i v_i^{\min}\right\} = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad V(T,x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{11}$$

Expressions for the case in which the penalty-type assignment is switched between the two players are also readily available.

4 A Solution Representation via a Feynman–Kac Formula

The cornerstone of the proposed approach is the nonlinear Feyman–Kac lemma, which establishes a close relationship between stochastic differential equations (SDEs) and second-order partial differential equations (PDEs) of parabolic or elliptic type. Specifically, this lemma demonstrates that solutions to a certain class of PDEs can be represented by solutions to SDEs or systems of forward and backward stochastic differential equations (FBSDEs). There is a plethora of similar theoretic results in the literature, all of which are referred to as Feynman–Kac-type formulas, since the earliest result of this type was due to Feynman and Kac (see

[30]). In this work, we propose to employ a nonlinear Feynman–Kac-type formula, which links the solution of a nonlinear PDE to a system of FBSDEs. In what follows, we will briefly review the theory of forward and backward processes and then present the nonlinear Feynman–Kac formula.

4.1 The Forward and Backward Process

The forward process is defined as the square-integrable, $\{\mathscr{F}_s\}_{s\geq 0}$-adapted process X_s,[3] which, for any given initial condition $(t, x) \in [0, T] \times \mathbb{R}^n$, satisfies the Itô FSDE

$$\begin{cases} \mathrm{d}X_s = b(s, X_s)\mathrm{d}s + \Sigma(s, X_s)\mathrm{d}W_s, & s \in [t, T], \\ X_t = x. \end{cases} \tag{12}$$

In the literature, forward process (12) is referred to as the *state process*. We denote the solution to FSDE (12) as $X_s^{t,x}$, wherein (t, x) are the initial condition parameters. The solution is then written in integral form as

$$X_s^{t,x} = x + \int_t^s b(\tau, X_\tau^{t,x})\mathrm{d}\tau + \int_t^s \Sigma(\tau, X_\tau^{t,x})\mathrm{d}W_\tau, \quad s \in [t, T], \tag{13}$$

wherein the second integral of the right-hand-side is the Itô integral and τ is a dummy variable of integration.

The associated backward process is the square-integrable, $\{\mathscr{F}_s\}_{s\geq 0}$-adapted pair (Y_s, Z_s) defined via a BSDE satisfying a terminal condition

$$\begin{cases} \mathrm{d}Y_s = -h(s, X_s^{t,x}, Y_s, Z_s)\mathrm{d}s + Z_s^\top \mathrm{d}W_s & s \in [t, T], \\ Y_T = g(X_T). \end{cases} \tag{14}$$

The function $h(\cdot)$ is referred to as *generator* or *driver*. The initial condition parameters (t, x) implicitly define the solution of the BSDE due to the terminal condition $g(X_T^{t,x})$, and thus we will similarly use the notation $Y_s^{t,x}$ and $Z_s^{t,x}$ for the solution associated with a particular initial condition parameter (t, x). The integral form of (14) is simply

$$Y_s^{t,x} = g\left(X_T^{t,x}\right) + \int_s^T h\left(\tau, X_\tau^{t,x}, Y_\tau^{t,x}, Z_\tau^{t,x}\right)\mathrm{d}\tau - \int_s^T \left(Z_\tau^{t,x}\right)^\top \mathrm{d}W_\tau, \quad s \in [t, T]. \tag{15}$$

Whenever the forward SDE does not depend on Y_s or Z_s, the resulting FBSDEs are *decoupled*.

The difficulty in dealing with BSDEs is that, in contrast to FSDEs, and due to the presence of a terminal condition, integration must be performed backwards in time, i.e., in a direction opposite to the evolution of the filtration. If we do not impose the solution to be adapted (i.e., non-anticipating, obeying the evolution direction of the filtration), we must require new definitions such as the *backward Itô integral* or, more generally, the so-called *anticipating stochastic calculus* [39]. In this work, we will restrict the analysis to adapted—non-anticipating—solutions. As shown in [39], an adapted solution is obtained if the *conditional expectation* of the process is back-propagated, by setting $Y_s \triangleq \mathbb{E}[Y_s|\mathscr{F}_s]$. In a sense, systems of FBSDEs describe two-point boundary value problems involving SDEs, with the extra requirement that their solution is adapted to the forward filtration.

The following lemma states that, under the assumptions of Sect. 2.1, imposed on b, Σ, g, and h, the adapted solution (Y, Z) can be written as deterministic functions of time and the state process [16]:

[3] While X is a function of s and ω, we shall use X_s for notational brevity.

Lemma 1 (The Markovian Property)*: Under the assumptions of Sect. 2.1, there exist deterministic functions $V : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ and $d : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, such that the solution $(Y^{t,x}, Z^{t,x})$ of BSDE (14) is*

$$Y_s^{t,x} = V\left(s, X_s^{t,x}\right), \quad Z_s^{t,x} = \Sigma^\top(s, X_s^{t,x}) d\left(s, X_s^{t,x}\right), \tag{16}$$

for all $s \in [t, T]$.

4.2 The Nonlinear Feynman–Kac Lemma

The following lemma, which links the solution of a class of PDEs to that of FBSDEs, can be proven by an application of Itô's formula (see [16,39,55]):

Lemma 2 (Nonlinear Feynman–Kac)*: Consider the Cauchy problem*

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma(t,x)\Sigma^\top(t,x)) + V_x^\top b(t,x) + h(t,x,V,\Sigma^\top(t,x)V_x) = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad V(T,x) = g(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{17}$$

wherein the functions Σ, b, h and g satisfy mild regularity conditions (see Remark 2). Then (17) admits a unique (viscosity) solution $V : [0, T] \times \mathbb{R}^n \to \mathbb{R}$, which has the following probabilistic representation:

$$V(t,x) = Y_t^{t,x}, \qquad \forall (t,x) \in [0,T] \times \mathbb{R}^n, \tag{18}$$

wherein (X_s, Y_s, Z_s) is the unique adapted solution of FBSDE system (12), (14). Furthermore,

$$(Y_s^{t,x}, Z_s^{t,x}) = \left(V(s, X_s^{t,x}),\ \Sigma^\top(s, X_s^{t,x}) V_x(s, X_s^{t,x})\right), \tag{19}$$

for $s \in [t, T]$, and if (17) admits a classical solution, then (18) provides that classical solution.

Remark 2 Concerning the regularity conditions of Lemma 2, [55] requires the functions Σ, b, h and g to be continuous, Σ and b to be uniformly Lipschitz in x, and h to be Lipschitz in (y, z), uniformly w.r.t (t, x). However, the nonlinear Feynman–Kac lemma has been extended to cases in which the driver is only continuous and satisfies a quadratic growth in z; see References [9,11,32,37]. Concerning existence of solutions to the HJI equation in this case, see [10].

Remark 3 The viscosity solution is to be understood in the sense of $V(t,x) = \lim_{\varepsilon \to 0} V^\varepsilon(t,x)$, uniformly in (t, x) over any compact set, where V^ε is the classical solution of the non-degenerate PDE

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma_\varepsilon(t,x)\Sigma_\varepsilon^\top(t,x)) + V_x^\top b_\varepsilon(t,x) + h_\varepsilon(t,x,V,\Sigma_\varepsilon^\top(t,x)V_x) = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad V(T,x) = g_\varepsilon(x), \quad x \in \mathbb{R}^n, \end{cases}$$

in which $\Sigma_\varepsilon, b_\varepsilon, h_\varepsilon$ and g_ε are smooth functions that converge to Σ, b, h and g uniformly over compact sets, respectively, and $\Sigma_\varepsilon(t,x)\Sigma_\varepsilon^\top(t,x) \geq \varepsilon I + \Sigma(t,x)\Sigma^\top(t,x)$ for al (t, x).

By comparing the PDEs in Sects. 3.1, 3.2 and 3.3 with Cauchy problem (17), we may conclude that Lemma 2 can be applied to each HJI equation of these sections under a certain decomposability condition:

Assumption 1 There exist matrix-valued functions $\Gamma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{p\times\nu}$ and $\Lambda : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{p\times\mu}$ such that $G(t, x) = \Sigma(t, x)\Gamma(t, x)$ and $B(t, x) = \Sigma(t, x)\Lambda(t, x)$ for all $(t, x) \in [0, T] \times \mathbb{R}^n$.

Assumption 1 restricts the range of G and B to be a subset of the range of Σ and therefore excludes cases wherein a channel containing control input does not contain noise. Notice however that the converse is allowed. Under Assumption 1, the HJI equations in Sects. 3.1, 3.2 and 3.3 (Eqs. 6, 10 and 11, respectively) satisfy Cauchy problem (17) standard form. We readily obtain the following SDE coefficients $b(\cdot)$ and $h(\cdot)$:

$$b(t, x) \equiv f(t, x) \tag{20}$$

and

$$\text{Case I}: \quad h(t, x, z) \equiv q - \frac{1}{2} z^\top \left(\Gamma R_u^{-1} \Gamma^\top - \Lambda R_v^{-1} \Lambda^\top \right) z, \tag{21}$$

$$\begin{aligned} \text{Case II}: \quad h(t, x, z) \equiv q &+ \sum_{i=1}^{\nu} \min \left\{ \left(z^\top \Gamma + p_u^\top \right)_i u_i^{\max},\ 0,\ -(z^\top \Gamma - p_u^\top)_i u_i^{\min} \right\} \\ &+ \sum_{i=1}^{\mu} \max \left\{ (z^\top \Lambda - p_v^\top)_i v_i^{\max},\ 0, -\left(z^\top \Lambda + p_v^\top \right)_i v_i^{\min} \right\}, \end{aligned} \tag{22}$$

$$\begin{aligned} \text{Case III}: \quad h(t, x, z) \equiv q &- \frac{1}{2} z^\top \Gamma R_u^{-1} \Gamma^\top z \\ &+ \sum_{i=1}^{\mu} \max \left\{ \left(z^\top \Lambda - p_v^\top \right)_i v_i^{\max},\ 0, -\left(z^\top \Lambda + p_v^\top \right)_i v_i^{\min} \right\}. \end{aligned} \tag{23}$$

The (viscosity) solution of PDEs (6), (10) or (11) are thus obtained by simulating the FBSDE systems given by (12) and (14) per definitions (20) and (21), (22) or (23), respectively. Notice that the drift of the forward process corresponds to that of the *uncontrolled* ($u = v = 0$) system dynamics.

5 Games Without a Fixed Time of Termination

The game formulation presented in Sect. 2 assumes that the game has a fixed, prespecified duration. Nevertheless, in many games this is not the case; rather, the game terminates when a particular state (or set of states) is reached. Since the presented approach is a sampling-based method, allowing the game to continue without imposing an upper bound on its duration may yield trajectory samples that have an infinite time duration, and thus cannot be simulated. However, we may combine the two aforementioned game formulations to obtain a *fixed final time/first exit* problem, in which the game terminates as soon as the state of termination has been reached, *or* a fixed time duration has passed, whichever event occurs first. The fixed final time in this case essentially acts as an upper bound on the duration of the game. To formulate the problem in this setting, let $\mathcal{G}$ be the domain of the game space and let $\partial\mathcal{G} \in C^1$ be its boundary, the crossing of which signals game termination, i.e., $\partial\mathcal{G}$ represents the *terminal surface*. We may replace the payoff in (1) by

$$P(u(\cdot), v(\cdot); x_0) = \mathbb{E}\left[\Psi(\mathcal{T}, x(\mathcal{T})) + \int_0^{\mathcal{T}} q(t, x(t)) + L^u(u(t)) - L^v(v(t))\mathrm{d}t\right], \quad (24)$$

in which $\mathcal{T}$ and $\Psi(\cdot)$ are defined as follows:

$$\mathcal{T} \triangleq \min\{\tau_{\text{exit}}, T\}, \text{ with } \tau_{\text{exit}} \triangleq \inf\{s \in [0, T] : x(s) \in \partial\mathcal{G}\}, \quad (25)$$

that is, τ_{exit} is the first hitting time at which a trajectory reaches the boundary $\partial\mathcal{G}$, and

$$\Psi(t, x) \triangleq \begin{cases} g(x), & (t, x) \in \{T\} \times \mathcal{G}, \\ \psi(t, x), & (t, x) \in [0, T) \times \partial\mathcal{G}. \end{cases} \quad (26)$$

Here, $g(\cdot)$ is the terminal payoff of (1), while $\psi(\cdot)$ is a function assigning a terminal payoff for time instants $t < T$, whenever the trajectories hit the terminal surface. Following the same procedure as in Sect. 2, and under Assumption 1, the resulting HJB PDE is [19]

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top b + h(t, x, \Sigma^\top V_x) = 0, & (t, x) \in [0, T) \times \mathcal{G}, \\ V(T, x) = g(x), & x \in \mathcal{G}, \\ V(t, x) = \psi(t, x), & (t, x) \in [0, T) \times \partial\mathcal{G}, \end{cases} \quad (27)$$

in which $b(\cdot)$ and $h(\cdot)$ may take any of the forms given by Eqs. (20)–(23). The corresponding FBSDEs that yield a probabilistic solution to this problem are [55]

$$\begin{cases} \mathrm{d}X_s = b(s, X_s)\mathrm{d}s + \Sigma(s, X_s)\mathrm{d}W_s, & s \in [t, \mathcal{T}] \\ X_t = x. \end{cases} \quad (28)$$

and

$$\begin{cases} \mathrm{d}Y_s = -h(s, X_s, Z_s)\mathrm{d}s + Z_s^\top \mathrm{d}W_s, & s \in [t, \mathcal{T}], \\ Y_{\mathcal{T}} = \Psi(X_{\mathcal{T}}). \end{cases} \quad (29)$$

6 Obtaining a Numerical Solution to FBSDE Systems

FBSDEs have received a lot of attention in the literature, and their solution has been studied, by and large, independently from their connection to PDEs. Several results appear within the field of mathematical finance, and a few generic numerical schemes have been proposed [4,6,36]. In this paper, we employ a scheme proposed in previous work by the authors [17], which capitalizes on the regularity present whenever systems of FBSDEs are linked to PDEs.

6.1 Time Discretization of FBSDEs

On a time grid $\{t = t_0 < \cdots < t_N = T\}$ for the interval $[t, T]$, we denote by $\Delta t_i \triangleq t_{i+1} - t_i$ the $(i+1)$-th interval of the grid (which can be selected to be constant) and $\Delta W_i \triangleq W_{t_{i+1}} - W_{t_i}$ the $(i+1)$-th Brownian motion increment; here, ΔW_i is simulated as $\sqrt{\Delta t_i}\xi_i$, where $\xi_i \sim \mathcal{N}(0, I)$. We also denote $X_i \triangleq X_{t_i}$ for notational brevity. The simplest scheme for the FSDE is the *Euler–Maruyama* scheme [31]:

$$\begin{cases} X_{i+1} \approx X_i + b(t_i, X_i)\Delta t_i + \Sigma(t_i, X_i)\Delta W_i, & i = 0, \ldots, N-1, \\ X_0 = x. \end{cases} \quad (30)$$

Note that several higher-order schemes exist that can be used alternatively. The interested reader is referred to [31] for a detailed presentation.

The backward process is discretized by introducing the notation $Y_i \triangleq Y_{t_i}$ and $Z_i \triangleq Z_{t_i}$. Recalling that adapted BSDE solutions impose a back-propagation of conditional expectations, i.e., $Y_s \triangleq \mathbb{E}[Y_s|\mathscr{F}_s]$ and $Z_s \triangleq \mathbb{E}[Z_s|\mathscr{F}_s]$, Eq. (14) is approximated by

$$Y_i = \mathbb{E}[Y_i|\mathscr{F}_{t_i}] \approx \mathbb{E}[Y_{i+1} + h(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})\Delta t_i|X_i], \tag{31}$$

for $i = N-1, \ldots, 0$. We note that the term $Z_i^\top \Delta W_i$ in (14) vanishes in the last equality because of the conditional expectation. (ΔW_i is zero mean and independent of Z_i, X_i.) Furthermore, $\mathscr{F}_{t_i}$ is replaced by X_i in (31) by virtue of Lemma 1. In light of Eq. (19), the Z-process in (14) corresponds to the term $\Sigma^\top(s, X_s^{t,x})V_x(s, X_s^{t,x})$. Thus, we may write

$$\begin{aligned} Z_i &= \mathbb{E}[Z_i|\mathscr{F}_{t_i}] = \mathbb{E}[\Sigma^\top(t_i, X_i)\nabla_x V(t_i, X_i)|X_i] \\ &= \Sigma^\top(t_i, X_i)\nabla_x V(t_i, X_i). \end{aligned} \tag{32}$$

The above expression requires knowledge of the solution on a neighborhood x at time t_i, that is, $V(t_i, x)$. The back-propagation is initialized at

$$Y_T = g(X_T), \qquad Z_T = \Sigma(T, X_T)^\top \nabla_x g(X_T), \tag{33}$$

assuming that $g(\cdot)$ is differentiable almost everywhere. Several ways exist to numerically approximate the conditional expectation in (31). In this work, we employ the least squares Monte Carlo (LSMC) method, which is briefly reviewed in what follows.

6.2 Conditional Expectation Approximation and Its Application

The least squares Monte Carlo (LSMC) method for approximating conditional expectations was initially introduced within the field of financial mathematics by Longstaff and Schwartz [38]. Given two square-integrable random variables X and Y for which we can sample M independent copies of (X, Y) pairs, the method enables estimating conditional expectations of the form $\mathbb{E}[Y|X]$, based on the principle that the conditional expectation of a random variable is a function of the variable on which it is conditioned on: $\mathbb{E}[Y|X] = \phi^*(X)$, where ϕ^* is the solution to the infinite-dimensional minimization problem

$$\phi^* = \arg\min_{\phi} \mathbb{E}[|\phi(X) - Y|^2], \tag{34}$$

and ϕ ranges over all measurable functions with $\mathbb{E}[|\phi(X)|^2] < \infty$. In practice, this problem is substituted by a finite-dimensional approximation by decomposing $\phi(\cdot) \approx \sum_{i=1}^k \varphi_i(\cdot)\alpha_i = \varphi(\cdot)\alpha$, with $\varphi(\cdot)$ being a row vector of k predetermined basis functions and α a column vector of constants, and replacing the expectation operator with its empirical estimator [23]. Thus, one obtains the following least squares problem:

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}^k} \frac{1}{M} \sum_{j=1}^{M} |\varphi(X^j)\alpha - Y^j|^2, \tag{35}$$

wherein (X^j, Y^j), $j = 1, \ldots, M$ are independent copies of (X, Y). The final expression for the LSMC estimator is simply $\mathbb{E}[Y|X = x] = \phi^*(x) \approx \varphi(x)\alpha^*$. In order to numerically solve FBSDEs, the LSMC method is used to approximate the conditional expectation in Eq. (31) for each time step. We begin by sampling M independent trajectories of the FSDE, $\{X_i^m\}_{i=1,\ldots,N}, m = 1, \ldots, M$. The BSDE numerical scheme is initialized at the terminal

time T and is iterated backwards along the time grid. At any given time step t_i, M pairs of data (Y_i^m, X_i^m)[4] are available, on which linear regression is performed to estimate the conditional expectation of Y_i as a function of x at the time step t_i. Thus, we obtain an approximation of the Value function V at time t_i, which is valid for the neighborhood of the state space that has been explored by the sample trajectories at that time instant, since $V(t_i, x) = \mathbb{E}[Y_i | X_i = x] \approx \varphi(x)\alpha_i$. The Y_i^m sample values calculated before the regression are then replaced by their projection: $Y_i^m = \varphi(X_i^m)\alpha_i$. Finally, the associated values for Z_i are obtained by taking the gradient with respect to x on the choice of basis functions (assuming that they are differentiable almost everywhere), as follows:

$$Z_i^m \approx \Sigma(t_i, X_i^m)^\top \nabla_x \varphi(X_i^m)\alpha_i. \tag{36}$$

The aforementioned process is repeated for $t_{i-1}, \ldots, t_0$. The proposed algorithm is then summarized as

$$\begin{cases} \text{Initialize} : Y_T = g(X_T), \quad Z_T = \Sigma(T, X_T)^\top \nabla_x g(X_T), \\ \alpha_i = \arg\min_\alpha \dfrac{1}{M} \left\| \Phi(X_i)\alpha - \left(Y_{i+1} + \Delta t_i h(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1}) \right) \right\|^2, \\ Y_i = \Phi(X_i)\alpha_i, \quad Z_i^m = \Sigma(t_i, X_i^m)^\top \nabla_x \varphi(X_i^m)\alpha_i, \end{cases} \tag{37}$$

where $m = 1, \ldots, M$ and the matrix Φ contains all basis function evaluations for all data points X_i.[5] The minimizer in (37) can be obtained by directly solving the normal equation, or by performing gradient descent. The algorithm output is essentially the collection of α_i's, i.e., the basis function coefficients for each time instant, allowing for an approximation of the value function which is valid for the area of the state space explored by the FSDE.

7 Iterative Schemes Based on Importance Sampling

The method presented thus far suffers from a significant practical limitation. Specifically, an approximation to the HJI PDE solution is obtained, which is accurate for those areas of the game space that are visited by trajectories of uncontrolled dynamics (Eq. 12), i.e., trajectories in which the players do not apply any control input. Nevertheless, the optimal, saddle point trajectory may lie on a different area of the state space, an area which uncontrolled trajectories might not access. In this case, simply extrapolating the locally obtained solution to cover the rest of the game space may not be accurate. This is a practical limitation, because in theory, if one could sample infinitely many trajectories, they would cover the entire game space, thus eliminating this issue. While generating infinitely many sample trajectories is not a practical solution, the issue can be effectively addressed if one is given the ability to alter the drift term of these sampled trajectories. By changing the drift during sampling, we can essentially select the area of the game space for which the obtained solution is accurate.

In previous work by the authors [17], a scheme involving a drift term modification has been constructed through Girsanov's theorem on the change of measure [30,48]. Indeed, the system of FBSDEs given by Eqs. (12) and (14) is equivalent, in the sense explained below, to one with modified drift

[4] Here, Y_i^m denotes the quantity $Y_{i+1}^m + \Delta t_i h(t_{i+1}, X_{i+1}^m, Y_{i+1}^m, Z_{i+1}^m)$, which is the Y_i^m sample value before the conditional expectation operator has been applied.

[5] Whenever the m index is not present, the entirety with respect to this index is to be understood.

$$\begin{cases} \mathrm{d}\tilde{X}_s = [b(s,\tilde{X}_s) + \Sigma(s,\tilde{X}_s)K_s]\mathrm{d}s + \Sigma(s,\tilde{X}_s)\mathrm{d}W_s, \quad s \in [t,T], \\ \tilde{X}_t = x, \end{cases} \tag{38}$$

along with the compensated BSDE

$$\begin{cases} \mathrm{d}\tilde{Y}_s = [-h(s,\tilde{X}_s,\tilde{Y}_s,\tilde{Z}_s) + \tilde{Z}_s^\top K_s]\mathrm{d}s + \tilde{Z}_s^\top \mathrm{d}W_s, \quad s \in [t,T], \\ \tilde{Y}_T = g(\tilde{X}_T), \end{cases} \tag{39}$$

for any measurable, bounded and adapted process $K_s : [0,T] \to \mathbb{R}^p$. We note that this equivalence is not path-wise, since different paths are realized by both the forward as well as the backward process, under the modified drift dynamics. Nevertheless, the solution at starting time t, i.e., (Y_t, Z_t), remains the same. The formal proof which involves Girsanov's theorem on the change of measure can be found in [17]. Additionally, one can explain why the modified system of FBSDEs (38) and (39) may substitute the original FBSDE system by examining the associated PDEs. Indeed, the FBSDE problem defined by (38) and (39) corresponds to the PDE problem

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top(b + \Sigma K) + h(t,x,V,\Sigma^\top V_x) - V_x^\top \Sigma K = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad V(T,x) = g(x), \end{cases} \tag{40}$$

which is identical to PDE (17), since the term $V_x^\top \Sigma K$ is both added and subtracted. Therefore, the FBSDEs, while being different, are associated with the same PDE problem.

To implement importance sampling for the problem at hand, we may apply any nominal controls $\bar{u}, \bar{v}$ in the game dynamics given by (2), which, in light of the definition of $\Gamma(\cdot)$ and $\Lambda(\cdot)$ in Assumption 1, exhibit the form

$$\mathrm{d}x(t) = [f(t,x(t)) + \Sigma(t,x(t))\,(\Gamma(t,x(t))\bar{u}(t) + \Lambda(t,x(t))\bar{v}(t))]\mathrm{d}t + \Sigma(t,x(t))\mathrm{d}W_t. \tag{41}$$

By comparison with the forward process (38), we thus have

$$K_s = \Gamma(s,X_s)\bar{u}(s) + \Lambda(s,X_s)\bar{v}(s), \qquad s \in [t,T], \tag{42}$$

while $b(s,X_s) \equiv f(s,X_s)$ per (20). We note that, while the nominal controls $\bar{u}, \bar{v}$ may be any controls, we are especially interested in the case in which we assign controls calculated at a previous run of the algorithm. Thus, one arrives at an iterative scheme, in which a more accurate approximation of the saddle point solution is obtained at each iteration. On the time grid of Sect. 6, we define $K_i \triangleq K_{t_i}$ and use the Euler–Maruyama scheme. Similarly, the most straightforward way to embed importance sampling in the backward process is to simply define

$$\tilde{h}(s,x,y,z,k) \triangleq h(s,x,y,z) - z^\top k, \tag{43}$$

and utilize the discretized scheme of Sect. 6 using $\tilde{h}$ instead of h. The modified procedure is summarized in Algorithm 1.

Note that one can also terminate this algorithm when the evaluation of payoff (1) in successive iterations does not exhibit significant change, in lieu of a predetermined number of iterations N_{it}.

Algorithm 1 NFK-FBSDE Algorithm with Importance Sampling

Input: Initial condition x_0, initial control inputs $\bar{u}$, $\bar{v}$ if available (otherwise zero or random), terminal time T, number of Monte Carlo samples M, number of iterations N_{it}.
Output: Basis function coefficients for the value function, α_i.

procedure NFK_FBSDE(x_0, $\bar{u}$, $\bar{v}$, T, M, γ, N, N_{it})
  Assign M control inputs $\bar{u}$, $\bar{v}$ using either initial, zero, or random values, to generate collections $\mathscr{U}_c$, $\mathscr{V}_c$.
  Sample a collection $\mathscr{X}$ of M state trajectories by applying discretization (30) on equation (41), using the control sequences of $\mathscr{U}_c$, $\mathscr{V}_c$;
  for $1 : N_{\text{it}}$ **do**
    Using $\mathscr{X}$ and $\mathscr{U}_c$, repeat backward scheme (37) for $N-1$ time steps, using $\tilde{h}$ of equation (43) to obtain α_i for each time step $i = 0, \ldots, N-1$;
    Sample M new trajectories per discretization (30), using the calculated optimal controls u^* and v^*, to replace the trajectories of $\mathscr{X}$, and evaluate payoff (1);
  end for
  return α_i.
end procedure

8 Error Analysis

The main sources of errors in the proposed numerical algorithm consist of (a) the time discretization scheme, and (b) the LSMC method of approximating conditional expectations. The time discretization error in most schemes in the literature decreases at a rate $\sqrt{N}$, where N is the number of (equidistant) time steps [36]. Convergence of the LSMC method of approximating conditional expectations (Sect. 6.2) is straightforward: the error obtained by projecting the unknown function ϕ in (34) to a set of basis functions vanishes as their number tends to infinity, thereby spanning the entire space in which the original unknown function lies. Furthermore, the empirical estimator in (35) of the expectation operator converges as the Monte Carlo samples tend to infinity; the error in this case is inversely proportional to the square root of the number of realizations, $\sqrt{M}$ [54]. Note that the PDE-FBSDE problem equivalence, illustrated by the nonlinear Feynman–Kac lemma (Sect. 4), being exact does not introduce any errors. Similarly, the importance sampling component which is based on Girsanov's theorem (Sect. 7) is also exact; numerical differences appear only for finite number of samples. Proving the overall convergence of the proposed discretization scheme however is more involved and is part of our ongoing research.

9 Simulation Results

To demonstrate the algorithm's performance, we present simulations performed on two different systems. First, a linear system is presented in Sect. 9.1, in which the game can have either a fixed duration, or is terminated when the game state reaches a particular value, as described in Sect. 5. The algorithm is able to address both cases. Next, a stochastic differential game of a double integrator plant is presented in Sect. 9.2.

9.1 An $\mathcal{L}^2$–$\mathcal{L}^2$ Linear System Example With/Without Fixed Time of Termination

We consider the $\mathcal{L}^2$–$\mathcal{L}^2$ payoff form and simulate the algorithm for $\mathrm{d}x = (0.5x + 0.5u + 0.25v)\mathrm{d}t + 0.5\mathrm{d}w$, with $q(t,x) = 0$, $R_u = 1$, $R_v = 2$, $x(0) = 1$, and $g(x(T)) = 10x^2(T)$, thus penalizing deviation from the origin at the time of termination, T. We distinguish two cases: in case (a), the time of termination is specified a priori to be $T = 2$, whereas in case

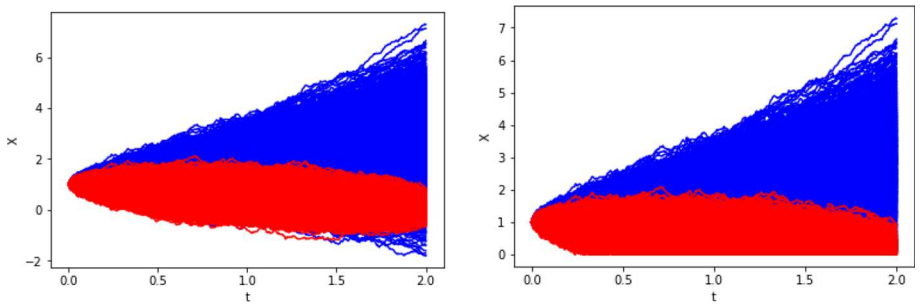

Fig. 1 Trajectories of the system of the. Uncontrolled sample trajectories are depicted in blue while optimally controlled trajectories in red. **a** The game terminates after a fixed, prespecified time duration $T = 2$, with a payoff of 3.66. **b** The game terminates as soon as the state $x = 0$ is reached, or a time duration of $t = 2$ has passed, whichever event occurs first. The payoff value is 3.01 (Color figure online)

(b), the game terminates when the x-axis ($x = 0$) is crossed, or the time reaches $t = 2$, whichever event occurs first. Two thousand trajectories were generated on a time grid of $\Delta t = 0.005$. The use of importance sampling for this problem is not necessary. Figure 1 shows the results for both cases (a) and (b), depicting uncontrolled sample trajectories in blue and optimally controlled trajectories in red, after the optimal controls u and v have been applied. The respective mean payoff values are 3.66 and 3.01, respectively.

9.2 The $\mathcal{L}^1$–$\mathcal{L}^2$ Double Integrator

In this section, we consider a stochastic differential game inspired by the fuel-optimal double integrator problem. The deterministic, one-player case of this problem offers a closed form solution; see [1, Ch. 8–6]. Specifically, the deterministic problem reads as follows: Given the system equations

$$\dot{x}_1(t) = x_2(t), \quad \dot{x}_2(t) = u(t), \quad |u(t)| \leq 1, \tag{44}$$

we wish to find the control which forces the system from an initial state (x_{10}, x_{20}) to the goal state $(0, 0)$, and which minimizes the fuel

$$J = \int_0^T |u(t)| \, \mathrm{d}t, \tag{45}$$

where T is a fixed (i.e., prespecified) response time. Existence of solutions is guaranteed if T satisfies a number of conditions depending on the values of the initial state. For an initial state (x_{10}, x_{20}) in the upper right quadrant of the plane, the condition reads $T \geq x_{20} + \sqrt{4x_{10} + 2x_{20}^2}$ and guarantees the existence of a unique solution. The corresponding fuel-optimal control sequence is $\{-1, 0, +1\}$, in which the control switching times t_1 and t_2 are

$$t_1 = 0.5\left(T + x_{20} - \sqrt{(T - x_{20})^2 - 4x_{10} - 2x_{20}^2}\right), \tag{46}$$

$$t_2 = 0.5\left(T + x_{20} + \sqrt{(T - x_{20})^2 - 4x_{10} - 2x_{20}^2}\right), \tag{47}$$

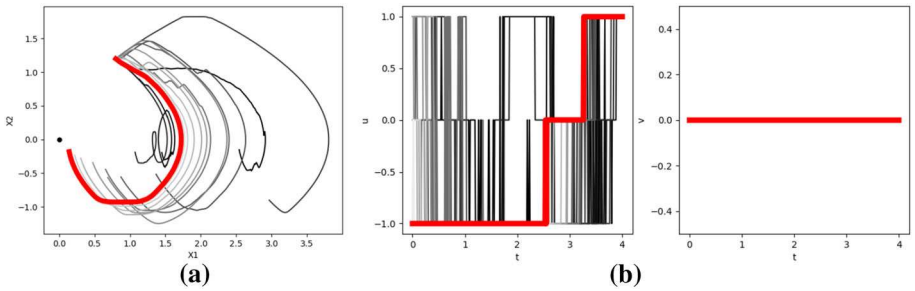

Fig. 2 Simulation results for $\beta = 10^{-8}$. **a** The mean of the controlled system trajectories of each iteration (grayscale) and after the final iteration (red). The black dot represents the origin. **b** The minimizing and maximizing control input for the mean system trajectory for each iteration (colored) and after the final iteration (black). We see that the optimal minimizing control sequence $\{-1, 0, +1\}$ is finally recovered (Color figure online)

that is,

$$u^*(t) = \begin{cases} -1, & t \in [0, t_1), \\ 0, & t \in [t_1, t_2), \\ 1, & t \in [t_2, T]. \end{cases} \tag{48}$$

Herein, we consider a stochastic differential game based on this problem, in which the minimizing player has a control restricted to $|u(t)| \leq 1$ and pays an $\mathcal{L}^1$ penalty, while the maximizing player has no control constraints and pays an $\mathcal{L}^2$ penalty. The dynamics are given by

$$\mathrm{d}x_1(t) = x_2(t)\,\mathrm{d}t, \quad \mathrm{d}x_2(t) = (u(t) + \beta\, v(t))\,\mathrm{d}t + \sigma\,\mathrm{d}w(t), \quad |u(t)| \leq 1, \tag{49}$$

i.e., stochasticity enters in form of perturbations in the control input channel. Here, β is a constant, the assigned value of which we may vary. An alternative stochastic counterpart could feature noise in the first channel as well. The payoff functional is given by:

$$P = \mathbb{E}\left[10\,(x_1^2(T) + x_2^2(T)) + \int_0^T |u(t)| - 2.5\,v^2(t)\,\mathrm{d}t\right]. \tag{50}$$

For the purposes of simulation, 3,000 trajectories were generated on a time grid of $\Delta t = 0.01$, with $\sigma = 0.1, T = 4$ and $(x_{10}, x_{20}) = (0.8, 1.2)$. The proposed algorithm was run for 50 iterations, using importance sampling. We run the algorithm for a very small value of β, (e.g., $\beta = 10^{-8}$), to investigate whether the solution of the stochastic differential game resembles the solution of the deterministic optimal control problem, in particular with respect to the optimal control sequence $\{-1, 0, +1\}$. Indeed, as shown in Fig. 2b, this optimal control sequence is recovered. Increasing $\beta = 0.1$, Fig. 3a depicts the mean of the controlled trajectories in phase-plane after each iteration of the algorithm (gray scale). The trajectory that corresponds to the final iteration is marked in red. Figure 3b depicts the payoff mean $\pm$ 3 standard deviations per iteration of the algorithm. Interestingly enough, the optimal minimizing control now differs, as shown in Fig. 3c.

10 Conclusions

In this work, we have presented a sampling-based algorithm designed to solve various classes of stochastic differential games, namely games in which the dynamics are affine in control,

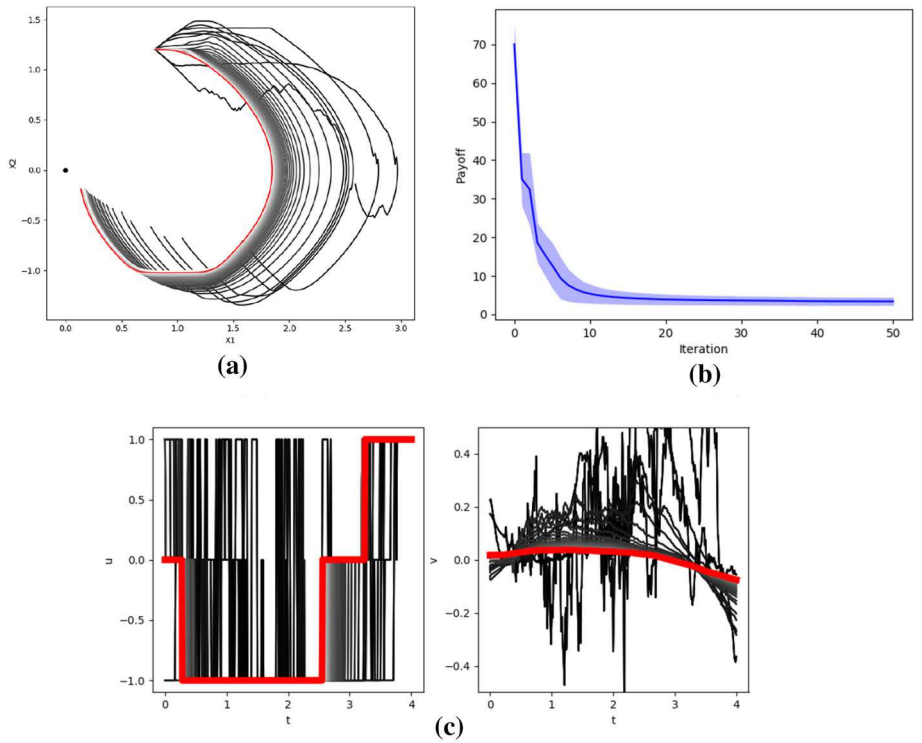

Fig. 3 Simulation results for $\beta = 0.1$. **a** The mean of the controlled system trajectories of each iteration (grayscale) and after the final iteration (red). The black dot represents the origin. **b** Cost mean $\pm$ 3 standard deviations per iteration. **c** The minimizing and maximizing control input for the mean system trajectory for each iteration (colored) and after the final iteration (black). We see that the optimal minimizing control sequence has now changed (Color figure online)

and the payoff involves either an $\mathcal{L}^2$ or an $\mathcal{L}^1$ type of control penalty. The time duration of the game can either be fixed and specified a priori, or the game may terminate as soon as a terminal surface in the game space is crossed (along with an upper bound on the maximum time duration). The cornerstone of the proposed approach is the formulation of the problem in terms of a decoupled pair of forward and backward stochastic differential equations (FBSDEs). This is done by means of a nonlinear version of the Feynman–Kac lemma, which suggests a probabilistic representation of the solution to the nonlinear Hamilton–Jacobi–Isaacs equations that arise for each class, expressed in the form of this decoupled system of FBSDEs. This system of FBSDEs can then be simulated using linear regression. In our paper, we embed these techniques within an iterative scheme based on Girsanov's theorem on the change of measure, to effectively address stochastic differential games. The applicability of the proposed method is demonstrated by means of numerical examples.

Acknowledgements Funding was provided by Army Research Office (W911NF-16-1-0390) and National Science Foundation (CMMI-1662523).

References

1. Athans M, Falb P (2007) Optimal control—an introduction to the theory and its applications. Dover Publications Inc, New York
2. Barles G, Souganidis P (1991) Convergence of approximation schemes for fully nonlinear second order equations. Asymptot Anal 4(3):271–283
3. Beard R, Saridis G, Wen J (1997) Galerkin approximation of the generalized Hamilton–Jacobi–Bellman equation. Automatica 33(12):2159–2177
4. Bender C, Denk R (2007) A forward scheme for backward SDEs. Stoch Process Appl 117:1793–1812
5. Berkovitz L (1961) A variational approach to differential games. RAND Corporation Report
6. Bouchard B, Touzi N (2004) Discrete time approximation and Monte Carlo simulation of BSDEs. Stoch Process Appl 111:175–206
7. Bouchard B, Elie R, Touzi N (2009) Discrete-time approximation of BSDEs and probabilistic schemes for fully nonlinear PDEs. Radon Ser Comput Appl Math 8:91–124
8. Buckdahn R, Li J (2008) Stochastic differential games and viscosity solutions of Hamilton–Jacobi–Bellman–Isaacs equations. SIAM J Control Optim 47(1):444–475
9. Chassagneux JF, Richou A (2016) Numerical simulation of quadratic BSDEs. Ann Appl Probab 26(1):262–304
10. Da Lio F, Ley O (2006) Uniqueness results for second-order Bellman-Isaacs equations under quadratic growth assumptions and applications. SIAM J Control Optim 45(1):74–106
11. Delbaen F, Hu Y, Richou A (2011) On the uniqueness of solutions to quadratic BSDEs with convex generators and unbounded terminal conditions. Annales de l'Institut Henri Poincarè, Probabilitès et Statistiques 47(2):559–574
12. Dixon M, Edelbaum T, Potter J, Vandervelde W (1970) Fuel optimal reorientation of axisymmetric spacecraft. J Spacecr Rockets 7(11):1345–1351
13. Douglas J, Ma J, Protter P (1996) Numerical methods for forward-backward stochastic differential equations. Ann Appl Probab 6:940–968
14. Duncan T, Pasik-Duncan B (2015) Some stochastic differential games with state dependent noise. In: 54th IEEE conference on decision and control, Osaka, Japan, December 15–18
15. Dvijotham K, Todorov E (2013) Linearly solvable optimal control. In: Lewis FL, Liu D (eds) Reinforcement learning and approximate dynamic programming for feedback control, pp 119–141. https://doi.org/10.1002/9781118453988.ch6
16. El Karoui N, Peng S, Quenez MC (1997) Backward stochastic differential equations in finance. Math Finance 7:1–71
17. Exarchos I, Theodorou E (2018) Stochastic optimal control via forward and backward stochastic differential equations and importance sampling. Automatica 87:159–165
18. Fahim A, Touzi N, Warin X (2011) A probabilistic numerical method for fully nonlinear parabolic PDEs. Ann Appl Probab 21(4):1322–1364
19. Fleming W, Soner H (2006) Controlled Markov processes and viscosity solutions, 2nd edn. Stochastic modelling and applied probability. Springer, Berlin
20. Fleming W, Souganidis P (1989) On the existence of value functions of two player zero-sum stochastic differential games. Indiana University Mathematics Journal, New York
21. Gobet E, Labart C (2007) Error expansion for the discretization of backward stochastic differential equations. Stoch Process Appl 117:803–829
22. Gorodetsky A, Karaman S, Marzouk Y (2015) Efficient high-dimensional stochastic optimal motion control using tensor-train decomposition. In: Robotics: science and systems (RSS)
23. Györfi L, Kohler M, Krzyzak A, Walk H (2002) A distribution-free theory of nonparametric regression. Springer series in statistics. Springer, New York
24. Hamadene S, Lepeltier JP (1995) Zero-sum stochastic differential games and backward equations. Syst Control Lett 24:259–263
25. Ho Y, Bryson A, Baron S (1965) Differential games and optimal pursuit-evasion strategies. IEEE Trans Autom Control 10:385–389
26. Horowitz MB, Burdick JW (2014) Semidefinite relaxations for stochastic optimal control policies. In: American control conference, Portland, June 4–6 pp 3006–3012
27. Horowitz MB, Damle A, Burdick JW (2014) Linear Hamilton Jacobi Bellman equations in high dimensions. In: 53rd IEEE conference on decision and control, Los Angeles, California, USA, December 15–17
28. Isaacs R (1965) Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization. Willey, New York
29. Kappen HJ (2005) Linear theory for control of nonlinear stochastic systems. Phys Rev Lett 95:200201
30. Karatzas I, Shreve S (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York

31. Kloeden P, Platen E (1999) Numerical solution of stochastic differential equations, vol 23 of Applications in Mathematics, Stochastic modelling and applied probability, 3rd edn. Springer, Berlin
32. Kobylanski M (2000) Backward stochastic differential equations and partial differential equations with quadratic growth. Ann Probab 28(2):558–602. https://doi.org/10.1214/aop/1019160253
33. Kushner H (2002) Numerical approximations for stochastic differential games. SIAM J Control Optim 41:457–486
34. Kushner H, Chamberlain S (1969) On stochastic differential games: sufficient conditions that a given strategy be a saddle point, and numerical procedures for the solution of the game. J Math Anal Appl 26:560–575
35. Lasserre JB, Henrion D, Prieur C, Trelat E (2008) Nonlinear optimal control via occupation measures and LMI-relaxations. SIAM J Control Optim 47(4):1643–1666
36. Lemor JP, Gobet E, Warin X (2006) Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. Bernoulli 12(5):889–916
37. Lepeltier JP, Martìn JS (1998) Existence for BSDE with superlinear-quadratic coefficient. Stoch Int J Probab Stoch Process 63(3–4):227–240
38. Longstaff FA, Schwartz RS (2001) Valuing American options by simulation: a simple least-squares approach. Rev Financ Stud 14:113–147
39. Ma J, Yong J (1999) Forward-backward stochastic differential equations and their applications. Springer, Berlin
40. Ma J, Protter P, Yong J (1994) Solving forward-backward stochastic differential equations explicitly—a four step scheme. Probab Theory Relat Fields 98:339–359
41. Ma J, Shen J, Zhao Y (2008) On numerical approximations of forward-backward stochastic differential equations. SIAM J Numer Anal 46(5):2636–2661
42. McEneaney WM (2007) A curse-of-dimensionality-free numerical method for solution of certain HJB PDEs. SIAM J Control Optim 46(4):1239–1276
43. Milstein GN, Tretyakov MV (2006) Numerical algorithm for forward-backward stochastic differential equations. SIAM J Sci Comput 28(2):561–582
44. Morimoto J, Atkeson C (2002) Minimax differential dynamic programming: An application to robust biped walking. In: Advances in neural information processing systems (NIPS), Vancouver, British Columbia, Canada, December 9–14
45. Morimoto J, Zeglin G, Atkeson C (2003) Minimax differential dynamic programming: Application to a biped walking robot. In: IEEE/RSJ international conference on intelligent robots and systems, Las Vegas, NV, 2: 1927–1932, October 27–31
46. Nagahara M, Quevedo DE, Nešić D (2016) Maximum hands-off control: a paradigm of control effort minimization. IEEE Trans Autom Control 61(3):735–747
47. Nagahara M, Quevedo DE, Nešić D (2013) Maximum hands-off control and L^1 optimality. In: 52nd IEEE conference on decision and control, Florence, Italy, December 10–13, pp 3825–3830
48. Øksendal B (2007) Stochastic differential equations—an introduction with applications, 6th edn. Springer, Berlin
49. Ramachandran KM, Tsokos CP (2012) Stochastic differential games. Atlantis Press, Paris
50. Seywald H, Kumar RR, Deshpande SS, Heck ML (1994) Minimum fuel spacecraft reorientation. J Guid Control Dyn 17(1):21–29
51. Song Q, Yin G, Zhang Z (2008) Numerical solutions for stochastic differential games with regime switching. IEEE Trans Autom Control 53:509–521
52. Sun W, Theodorou EA, Tsiotras P (2015) Game-theoretic continuous time differential dynamic programming. In: American Control Conference, Chicago, July 1–3, pp 5593–5598
53. Theodorou EA, Buchli J, Schaal S (2010) A generalized path integral control approach to reinforcement learning. J Mach Learn Res 11:3137–3181
54. Xiu D (2010) Numerical methods for stochastic computations—a spectral method approach. Princeton University Press, Princeton
55. Yong J, Zhou XY (1999) Stochastic controls: hamiltonian systems and HJB equations. Springer, New York
56. Zhang J (2004) A numerical scheme for BSDEs. Ann Appl Probab 14(1):459–488
57. Zhang J (2017) Backward stochastic differential equations. Probability theory and stochastic modelling. Springer, Berlin